IJSGS FUGUSAU VOL. 8 (2) JULY, 2022 WEBSITE: https://fugus-ijsgs.com.ng

IJSGS

ISSN: 2488-9229
FEDERAL UNIVERSITY
GUSAU-NIGERIA

INTERNATIONAL JOURNAL OF SCIENCE FOR GLOBAL SUSTAINABILITY

Variable Selection with Convex and Non-Convex Penalized Likelihood Models Using Rainfall Data

[1]Usman M., [1]Doguwa S. I. S., [2]Alhaji B. B. and [1]Dikko H, G.

[1]Department of Statistics, Ahmadu Bello University Zaria, Nigeria
[2]Department of Mathematics, Nigerian Defense Academy Kaduna, Nigeria

Corresponding Author's Email & Phone No.: ummohammed67@yahoo.com; 08035940877

Received on: April, 2022 **Revised and Accepted on:** June, 2022 **Published on:** July 2022

ABSTRACT

Accurate estimate of rainfall is very important for effective use of water resources and optimal planning of water structure in a day-to-day activity of life. Variable selection is an important aspect in penalization for the estimation of accurate outcome. Traditional variable selection such as stepwise and subset selections are usually used which can be computationally expensive and ignore stochastic errors in the variable selection process. Penalized likelihood methods are applied to select the important variables which can be used for accurate predictions. In this study, penalized likelihood approach is applied to select variables and estimate coefficients simultaneously. Some of penalized penalty functions were used to produce sparse solutions. From the results obtained the penalty functions produce the important variables that influence the total rainfall. Lasso model produces Four (4) important variables, Elastic net produces Two (2) important variables while SCAD produces only One (1) variable as important. This indicates that Lasso model is more complex than SCAD model. The results also show that SCAD penalty function out performed Ridge, Lasso and Elastic net. Based on the RMSE criteria, Ridge regression performed less compared to the other models.

Keywords: Ridge, Lasso, Elastic Net, SCAD, Variable selection

1.0 INTRODUCTION

Rainfall by definition is a liquid in the form of droplets that has condensed from the atmospheric water vapor and then became heavy enough to fall under gravity. If a region under study is blessed with a fertile land, and if there is enough rainfall and other supportive agricultural factors are met, then the region would have a bumper harvest. Rainfall is the most essential aspect in a farming system as it determines the accessibility of soil needed for maximum yield (Niles *et al*. 2015). Rainfall is the major climate resources that can be used as an index of climate change (Adhikary *et al*, 2016). The evaporated liquid condenses in the cold air, forming moisture-filled rain clouds (Wiltz, 2018). Rainfall plays an important role in the hydrological cycle which controls our water supplies and water disasters. In Tabari, (2020), the hydrological cycle is expected to intensify with global warming, which likely increases the intensity of extreme precipitation events and the risk of flooding. Knowing the nature and characteristics of rainfall, we can conceptualize and predict its effects in runoff, infiltration, evapotranspiration, and water yield (Tang *et al*, 2005). Hydrologic cycle is a formation of larger process of rain and snow. Living things including people require water to live. According to the National Climatic Data Center, the wettest place in the world with the highest average amount of yearly rainfall is Lloro, Colombia with a whopping 523.6 inches. The highest-ranked American location is Mt. Waialae, Hawaii with an average of 460 inches per year. The driest location in the world is also in South America is Arica, Chile with an average annual rainfall of 0.03 inches (Wiltz, 2018). Excess rainfall can cause floods and large number of properties and crops can be damaged. Also, deficiency in rainfall can cause drought and decline in crop yield. Rainfall is the primary source of agriculture in most areas of the world. It has been the most favourable source of water for agricultural purposes in Nigeria. Rainfall can release large amounts of latent heat that have tremendous impact on weather and climate in an area. This result in inadequate rainfall measurements or the existing climatic models to forecast the amount of heat being released by those models are not accurate.

Recently, the impact of climatic change with respect to delay at the commencement of rainfall as well as excessive rainfall has affected the planting date of different crops across the country. This may lead to food insecurity if appropriate measures are not put in place. The effect of climate change has been experiences in many parts of the world of which Nigeria is not left out of this, especially in their agricultural sector (Garuba *et al*, 2021). The reality of climate change and frequency of its unpleasant consequences constitute significant threats to human lives across different regions of the world. The adverse outcome of climate change has necessitated global concerns and efforts at mitigating its effects as well as advocacy for measures that would restrict human actions that induce climate change (Ani *et al*, 2022). The activities of farmers have been greatly changed due to climatic change which resulted in the delay in the onset of rainfall especially in Northern Nigeria, thus delay in the planting activities as well as maturity and harvesting of crops, while excess of the rainfall in Southern Nigeria results to flooding and landslides (APS report, 2019/2020/2021). The evidence of high intra-seasonal and inter-seasonal variability in rainfall, increase in extreme weather events such as drought and flood were presented in those studies of (Borhara *et al.*, 2020; Gebrechorkos *et al.*, 2018). Panga *et al*, (2021) observed that, for a decade now, the trends of rainfall instability and its impacts has been a crucial climatic problem facing different nations. The scenario has been linked directly or indirectly with global warming, which pose its impacts to different sectors particularly agriculture and tourism whose contribution is vital to any economy of a country. Rain water is very important for both environment and its inhabitants. Agricultural activities largely depend on precipitation and its availability. Therefore, the ability to forecast future rainfall at specific time is key for environmental decision and agricultural development.

In regression analysis, the researcher is concern with the discovery of the key predictor variables for achieving better prediction of the response variable. Therefore, identifying the potential predictor variables for knowledge discovery and boosting the predictive power of the model are very beneficial (Zou and Zhang, 2009). In practice, a large number of predictor variables can increase the variance of the fitted model, and selecting several predictor variables may result in unpredictable output or biased results. In other words, incorporating more predictor variables in the model may cause high variation in the least-squares fit, which, in turn, results in overfitting the model, and hence, it yields a poor forecast for the future (James *et al*, 2013). Variable selection aimed at choosing the "optimum" subset of predictors. According to Desboulets (2018), when building a statistical model, the question of which variables to include often arises. Practitioners in statistics have now at their disposal a wide range of technologies to solve these problems. If a model is to be used for prediction, time and money can be saved by measuring only the necessary predictors. Choosing which predictor variables amongst many potential ones to be included in a model is one of the central challenges in regression analysis. Traditionally, stepwise selection and subset selection are the main methods of variable selection for many years. Unfortunately, these methods are unstable and ignore the stochastic errors projected by the selection process (Sarpong *et al*, 2016). Variable selection methods in linear regression are grouped into two categories: sequential selection methods, such as forward selection, backward elimination and stepwise regression; and penalized regression methods, also known as shrinkage or regularization methods, including the Lasso, Elastic net, and their modifications and combinations. Sequential selection methods are easy to interpret but are a discrete search process in which variables are either included in or excluded from the model. Penalization techniques for variable selection in regression models are alternatives to sequential selection methods that are more continuous and do not suffer as much from this variability. They are becoming increasingly popular because they are able to perform variable selection while simultaneously estimating the coefficients in the model (Wang, 2020).

To obtain stability, higher prediction accuracy and computational efficiency, penalized regression methods are applied as alternative to sequential variable selection methods. Penalization methods set some regression coefficients that are not important to zero, and this can increase the prediction accuracy thus, means squared error can be improved.

In this study, penalized regression method such as model selection, can be of help, to find a parsimonious model for identifying significant factors associated with the amount of rainfall. In penalization or regularization method, a small bias is introduced into a loss function as a penalty, to achieve greater reduction in the variance of the estimates, which then improves the mean squared error.

2.0 MATERIALS AND METHODS

Meteorological section of Institute for Agricultural Research (IAR), Ahmadu Bello University, Samaru Zaria is responsible for the collection of meteorological data of the institute. The data used in this study were obtained from (IAR). The daily rainfall data from the period of January to December 2021 were used and the data consist of 365 samples. The following are the variables used in the study: relative humidity (10am), relative humidity (4pm), air temperature (10am), air temperature (4pm), earth temperature (10am), earth temperature (4pm), wind direction (10am), wind direction (4pm), wind speed, open evaporation, sunshine and total rainfall as the response variable.

The following are the penalized regression methods applied in this study. These are Ridge regression, Least absolute shrinkage and selection operator (Lasso), Elastic net and Smooth clipped absolute deviation (SCAD).

2.1 Smooth Clipped Absolute Deviation (SCAD)

A number of variable selection methods have been proposed involving nonconvex penalty functions. These methods include the smoothly clipped absolute deviation (SCAD) penalty as one of the nonconvex methods. SCAD penalty proposed by Fan and Li (2001) is known to satisfy the three desirable requirements (unbiasedness, sparsity, continuity) for good penalty function. They also show that the performance of SCAD is expected to be as good as that of the oracle estimator as the sample size increases. With the oracle properties possess by SCAD, the first benefit is that the non-zero coefficients will be determined easily with probability converging to one for all predictors (Yousef, 2019). The problem of variable selection arose in the context of multiple linear regression with continuous predictor variables, when the number of predictor variables is larger than the observation.

Suppose we are given the observations that are independently and identically distributed (*iid*), (x_i, y_i), for $i = 1, 2, \ldots, n$ and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$. Now the model is given as

$$y_i = x_i\beta + \varepsilon_i \qquad (1)$$

$$i = 1, 2, \ldots, n$$

where ε_i the error terms are $i.i\,d\;N(0;\sigma^2)$, and the coefficient vector is $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$. With sparse models, the predictor variable (p) can be categorized into two classes, the important ones whose corresponding coefficients are non-zero and the less important ones whose coefficients are zero. The SCAD method as compared to the classical variable selection methods such as subset selection, has two advantages. The first one is that, variable selection with SCAD method is continuous and therefore more stable than the subset selection, which is a discrete and non-continuous process. Secondly, the SCAD method is computationally feasible for large dataset. In contrast, computation in subset selection is mathematical intensive and not feasible when the predictor variable is large.
Consider the following objective function.

$$\hat{\beta} = \arg\min \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + P_\lambda(\beta) \qquad (2)$$

$i = 1, 2, \ldots, n$, the number of observations, $j = 1, 2, \ldots, p$, the number of predictors

The SCAD penalty is given as

$$P_\lambda(\beta) = \begin{cases} \lambda|\beta|, & \text{if } |\beta| \le \lambda \\ \dfrac{2a\lambda|\beta| - \beta^2 - \lambda^2}{2(a-1)}, & \text{if } 0 < |\beta| \le a\lambda \\ \dfrac{\lambda^2(a+1)}{2}, & \text{if } |\beta| > a\lambda \end{cases} \quad (3)$$

and $\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2$ is the ordinarily least square part of equation (2)

The penalty function can also be expressed in its derivative form, with respect to the regression coefficient (β) and the rationale behind the penalty can again be understood by considering its derivative given as:

$$P'_\lambda(\beta) = \begin{cases} \lambda, & \text{if } |\beta| \le \lambda \\ \dfrac{a\lambda - \beta}{a-1}, & \text{if } 0 < |\beta| \le a\lambda \\ 0, & \text{if } |\beta| > a\lambda \end{cases} \quad (4)$$

Equation (4) can also be expressed as follows:

$$P'_\lambda(\beta) = \lambda I(\beta \le \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)} I(\beta > \lambda) \quad (5)$$

where a is constant, and $a > 2$ usually taken to be $a = 3.7$ as suggested by Fan and Li (2001) and $\beta > 0$ improves the properties of some penalty functions. $I =$ indicator function, β is the coefficient and λ is the tuning parameter (or amount of shrinkage), that can be estimated by the method of cross validation.

2.2 Ridge Regression

Ridge regression estimator proposed by Hoerl and Kennard (1970) has been introduced as an alternative to the ordinary least squares estimator (OLS) in the presence of multicollinearity. Ridge is one of the convex penalty functions, that shrinks the coefficients of the correlated predictor variable equally toward zero. In Ridge regression, variances are reduced in exchange with some bias of the OLS estimators. The estimator shrinks the OLS estimators using the L_2 norm of the coefficients. The loss function of Ridge regression method is shown to be

$$\hat{\beta}_{Ridge} = \arg\min \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta^2_j \quad (6)$$

where L_2 penalty is $\lambda\sum_{j=1}^{p}\beta^2_j$, β is the regression coefficient and λ controls the amount of shrinkage of the coefficients and $\lambda > 0$ is a penalty parameter. When $\lambda = 0$, equation (6), corresponds to least squares regression. Now for real valued function *f*, with domain S, arg min [$f(\beta)$] $\in Sf(\beta)$ is the set of elements in *S* that achieve the global minimum in *S*. Using matrix algebra, we can rewrite equation (6) in matrix form as:

$$\hat{\beta}_{(Ridge)} = (X^TX + \lambda I)^{-1} X^TY \quad (7)$$

where $\boldsymbol{I}$ is the $p \times p$ identity matrix. Adding λ to the diagonal of X^TX makes the problem nonsingular even with the multi-collinearity in the data.

2.3 Least Absolute shrinkage and Selection Operator (Lasso) Regression

Lasso proposed by Tibshrini (1996) is also a regularization method that shrinks the loss of absolute error sum of square and constraints the sum of absolute value of the coefficients. The Lasso estimator is another convex penalty function which uses the L_1 norm to penalize the least square in order to obtain sparse solution.
The loss function of Lasso regression method is shown to be

$$\hat{\beta}_{Lasso} = \arg\min \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \qquad (8)$$

where L_1 penalty is $\lambda\sum_{j=1}^{p}|\beta_j|$, λ is the Lasso penalty parameter which penalizes the sum of the absolute values of the regression coefficients.

$$MSE(\lambda) = (Y - \widehat{Y}_{(\lambda)})^T (Y - \widehat{Y}_{(\lambda)}) / \mathrm{n}$$

To obtain the $\boldsymbol{\beta}_{lasso}$ in matrix form, we need to minimise equation (8) with respect to $\boldsymbol{\beta}$:
This involves differentiating the equation with respect to $\boldsymbol{\beta}$ and setting the derivative to zero in order to obtain the system of equations:

$$(X^T X)\beta + 0.5\lambda\beta^* = X^T Y \qquad (9)$$

where $\boldsymbol{\beta}^*$ is defined as,

$$\boldsymbol{\beta}^* = \begin{pmatrix} \beta_0 / |\beta_0| \\ \beta_1 / |\beta_1| \\ . \\ . \\ \beta_p / |\beta_p| \end{pmatrix} \qquad (10)$$

Clearly equation (9) is not in closed form, so iterative method has to be used to determine the Lasso estimate of $\boldsymbol{\beta}$. Using the Newton Raphson Algorithm, and for a given $\lambda > 0$ and initial value of β_0, one can run the iteration in equation (11).

$$\boldsymbol{\beta_t} = \boldsymbol{\beta}_{t-1} - (X^T X + 0.5\lambda \boldsymbol{G}_{t-1})^{-1} (X^T X \boldsymbol{\beta}_{t-1} + 0.5\lambda\boldsymbol{\beta}^*_{t-1} - X^T Y) \qquad (11)$$

where the $p+1$ by $p+1$ diagonal matrix $\boldsymbol{G}$ is defined as,

$$\boldsymbol{G} = diag\left(\left(\frac{1-\beta_0^{-2}}{|\beta_0|}\right), \left(\frac{1-\beta_1^{-2}}{|\beta_1|}\right), \left(\frac{1-\beta_2^{-2}}{|\beta_2|}\right), \ldots\ldots, \left(\frac{1-\beta_p^{-2}}{|\beta_p|}\right)\right) \qquad (12)$$

and t =1, 2, 3,…, until convergence is achieved. A possible convergence criterion could be to stop the iteration whenever:

$$(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})^T(\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}) / (\boldsymbol{\beta}_{t-1})^T(\boldsymbol{\beta}_{t-1}) < 10^{-6} \qquad (13)$$

and take $\widehat{\boldsymbol{\beta}}_{lasso} = \boldsymbol{\beta}_{t-1}$ for the given λ.
To obtain the optimal λ we run equation (11) for a range of λ values for example (λ = 0.01, 0.02, 0.03,…,1) and choose the optimal $\hat{\beta}_{(lasso)}$ that minimises the mean squared error MSE(λ), while the corresponding value of λ gives the λ_{opt}

(14)

and

$$\widehat{Y}_{(\lambda)} = X\hat{\beta}_{(lasso)} \qquad (15)$$

where λ_{opt} is the value of λ in which MSE(λ) attains the global minimum.

The Lasso estimator reduces the variability of the estimates by shrinking some coefficients, and produces interpretable models by shrinking some other coefficients to exactly zero. These properties make the Lasso a highly popular method in variable selection.

2.4 Elastic net Regression

Elastic net is also, a convex penalty function introduced by Zou and Hastie (2005), to extend the Lasso by improving some of its limitations, especially with respect to the variable selection. The method produces a regression model that is penalized with both the Ridge regression penalty term of L_1 - norm and Lasso regression penalty term of L_2 - norm. The consequence of this is to effectively shrink coefficients just like in Ridge regression and to set some coefficients to zero as in Lasso. The L_1 - norm part of the penalty generates a sparse model by shrinking some regression coefficients exactly to zero. The L_2 - norm part of the penalty removes the limitation on the number of selected variables, encourages grouping effect, and stabilizes the L_1 regularization path (Park and Konishi, 2015).

In this situation, Elastic net not only selects variables, but may also perform better than Lasso with observations that are collinear. Liu and Li (2017) used an efficient Elastic net with regression coefficients method to select the significant variables of the spectrum data. Steele *et al*, (2018) analyzed the electronic patient health records for predicting patient mortality with Elastic net method. Furthermore, Elastic net is particularly useful in cases where the number of predictor variables (p) in datasets are much larger than the number of observations (*n*). In such cases, Lasso is not capable of selecting more than '*n*' predictors but the Elastic net has this capability (Frank and Matthias, 2019).
Elastic net minimises the loss function and the estimated parameter vector is given by

$$\hat{\beta}_{Elastic} = \arg\min\left[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\left[(1-\alpha)\sum_{j=1}^{p}\beta_j^2 + \alpha\sum_{j=1}^{p}|\beta_j|\right]\right] \quad (16)$$

where
λ = tuning parameter
a = weight that determines how much should be given to Lasso or Ridge regression,

such that $0 \le \alpha \le 1$, where $a = 0$, $\hat{\beta}_{Elastic}$ becomes $\hat{\beta}_{Ridge}$ and where $\alpha = 1$, $\hat{\beta}_{Elastic}$ becomes $\hat{\beta}_{Lasso}$.

It is known that the Ridge regression penalty shrinks the coefficients of correlated predictors towards each other, while the Lasso regression penalty tends to select one of the coefficients and discard the others. The Elastic net penalty mixes the two penalties when the predictors are correlated in groups. An alpha (a) = 0.5 tends to either select or leave out the entire group of the variables.

The matrix form of (16) does not have a closed form and its solution can only be obtained iteratively using:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} - (X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}(X^TX\boldsymbol{\beta}_{t-1} + \lambda\{\alpha\boldsymbol{\beta}_{t-1} + 0.5(1-\alpha)\boldsymbol{\beta}^*_{t-1}\} - X^TY) \quad (17)$$

where $t = 1, 2, \ldots$ and the diagonal matrix $\boldsymbol{G}$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are as defined in equation (11). For a given α such that $0 < \alpha < 1$ and a given initial value $\boldsymbol{\beta}_0$, one runs equation (17) for each value of λ (λ = 0.01, 0.02, …,1) say, until convergence is achieved based on the convergence criterion defined in equation (13), and then compute the MSE(λ). The λ_{opt} is obtained as that value of λ that returns the minimum MSE (λ). Having obtained λ_{opt} we can use it to re-run equation (17) with values of α say (α =0.01, 0.02,…,0.99) and for each α we estimate MSE(α).
The optimal α is determined as the one, the α returns the smallest MSE (α). Both the optimal α and λ are then used in equation (17) to re-run the iteration until convergence. The converged $\hat{\boldsymbol{\beta}}$ is the elastic net estimate $\hat{\boldsymbol{\beta}}_{elastic}$ of $\boldsymbol{\beta}$. The variance of the estimate is then computed as,

$$Var(\hat{\beta}_{elastic}) = \sigma^2\, diag[(X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}X^TX(X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}] \quad (18)$$

where σ^2 is defined as

$$\sigma^2 = (Y - \hat{Y}_{(\lambda)})^T(Y - \hat{Y}_{(\lambda)})/(n-p) \quad (19)$$

and

$$\hat{Y}_{(\lambda)} = X\hat{\beta}_{(Elastic)} \quad (20)$$

Cross Validation (CV) criterion is used to determine the optimum lambda (λ_{opt}) value. By cross validation method, different values of lambda (λ) will be generated from the analysis, with their corresponding means square errors (MSE). The lambda (λ) value with the minimum MSE will be the λ_{opt}. The identification of λ_{opt} produces the most parsimonious model i.e a model with few non – zero coefficients. To compare the prediction performance of the models, Root Mean Square Error (RMSE) criterion was used.

3.0 RESULTS AND DISCUSSION

The dataset was used to assess the performance of the model and also the number of variables selected. The dataset was divided into two subsets. The first subset is the training set ($n_{train} = 256$ samples) that was used in estimating the models, while the second subset is the testing set (n_{test}= 109 samples), which was used for assessing the prediction accuracy of the estimated models. We used cross validation method to select the λ_{opt} which was used for shrinkage and estimation of the regression coefficients. To estimate the coefficients in Ridge model we used λ_{opt} = 1.542733. In Ridge regression, all the variables were included in the model, since the model does not perform variable selection as displays in Table 3.1

Table 3.1: Coefficient Estimate with Ridge Regression

Variable	Coefficient
(Intercept)	8.537348578
Relative_Humidity_10am	0.018204825
Relative_Humidity_4pm	0.016838141
Air_Temp_Max.	-0.006400661
Air_Temp_Min.	0.049235794
Earth_Temp_10am	-0.057764855
Earth_Temp_4pm	-0.059309156
Wind_Speed	-0.002888592
Wind_Direction_10am	-0.139587146
Wind_Direction_4pm	0.012553632
Open_Evap	-0.069005492
Sunshine_Hours	-0.250264196

To investigates the influence of lambda (λ) on the estimated coefficients for all variables, we plot the trace of coefficient against log Lambda (λ) as depicted in Figure 3.1

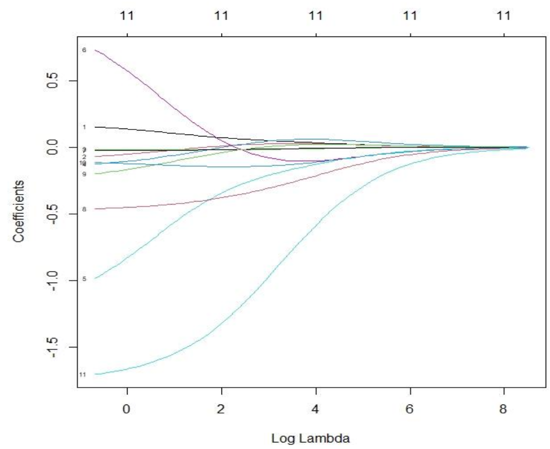

Figure 3.1: Coefficient estimates of βi for Ridge regression

The upper part of the plot in Figure 3.1 indicates the 11 number of coefficients estimates for a given value of log λ. From Figure 3.1, in Ridge regression this number is constant for all the values and equals the number of predictor variables in the data.

The plot displays Ridge regression coefficients against the shrinkage penalty lambda (λ). Each colour line represents a trace of the value in the coefficient for a single predictor variable. The left part of the plot shows the coefficient estimates, and lambda (λ) starts shrinking the parameter estimates at a different rate moving towards right-hand side. At the right-hand side of the plot all estimates become zero. Although, Ridge regression shrinks the coefficient estimates close to zero, even with large values, the model does not produce estimates that are exactly equal to zero, because the method includes all predictor variables and the penalty lambda (λ) will never reach exactly zero.

Table 3.2 indicates the non-zero coefficient estimates of Lasso regression obtained with the aid of λ_{opt} using cross validation method. The λ_{opt} = 1.063346. Lasso method selects 4 non-zero coefficients and shrinks the other coefficients to zero as displays in Table 3.2

Table 3.2: Coefficient Estimate of Lasso Regression

Variable	Coefficient
Intercept	6.806265279
Relative_Humidity_10am	0.080466055

Relative_Humidity_4pm	-
Air_Temp_Max.	-
Air_Temp_Min.	-
Earth_Temp_10am	-
Earth_Temp_4pm	-
Wind_Speed	-0.003762919
Wind_Direction_10am	-0.006880016
Wind_Direction_4pm	-
Open_Evap	-
Sunshine_Hours	-0.911241688

The plot of the trace of coefficient against log Lambda (λ) for Lasso regression is displays in Figure 3.2

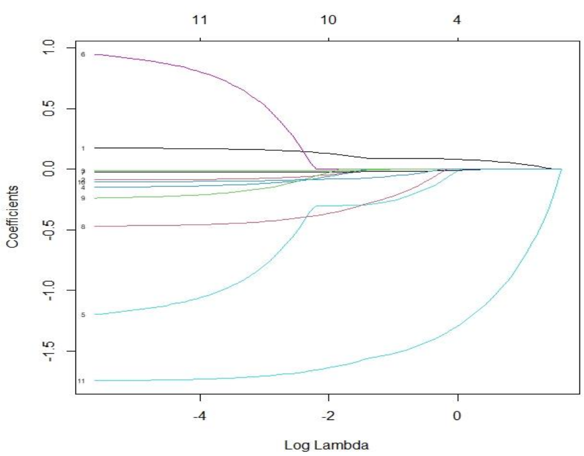

Figure 3.2: Coefficient estimates of βi for Lasso regression

Figure 3.2 indicates the trace of Lasso regression coefficients versus log lambda (λ). The penalty strength lambda (λ) decreases from left to right, shrinking the coefficients toward zero and other coefficients to exactly zero. In Lasso regression, the model also tends to shrink the regression coefficients to zero as λ increases. It can be seen from the plot in Figure 3.2, the Lasso regression shrink the regression coefficients to zero as λ increases. The values in the upper part of the plot in Figure 3.2 indicates the number of non-zero coefficients in the regression model. For instance, when $\log\lambda = -4$ we have eleven (11) non-zero coefficients and when $\log\lambda$ increased to -2 we get ten (10) non-zero coefficients in the model. That is to say, the Lasso regression shrinks some unimportant variables to zero thus, performs variable selection when λ is large enough.

Table 3.3 indicates the non-zero coefficient estimate of Elastic net regression with λ_{opt} = 2.561603. Elastic net method selects 2 non-zero coefficients and shrinks the other coefficients to zero as displays in Table 3.3.

Table 3.3: Coefficient Estimate of Elastic net Regression

Variable	Coefficient
Intercept	5.89671162
Relative_Humidity_10am	0.07252912
Relative_Humidity_4pm	-
Air_Temp_Max.	-
Air_Temp_Min.	-
Earth_Temp_10am	-
Earth_Temp_4pm	-
Wind_Speed	-
Wind_Direction_10am	-
Wind_Direction_4pm	-
Open_Evap	-
Sunshine_Hours	-0.80395374

The plot of the trace of coefficient against log Lambda (λ) for Elastic net regression is displays in Figure 3.3

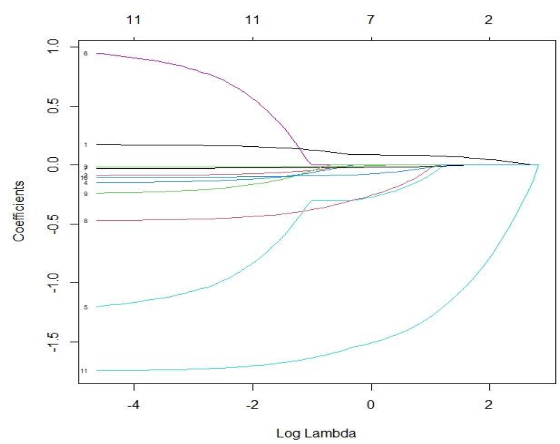

Figure 3.3: Coefficient estimates of βi for Elastic net regression

Elastic net regression model also tends to shrink the regression coefficients to zero as λ increases. In Figure 3.3, Elastic net regression shrink the regression coefficients to zero as λ increases. From Figure 3.3, the values in the upper part of the plot shows the non-zero coefficients in the regression model. For instance, when $\log\lambda = -2$ we have 11 non-zero coefficients and when $\log\lambda$ increased to 2 we get only 2 non-zero coefficients in the model. Also, Elastic net regression performs variable selection when λ is large enough.

Table 3.4 indicates the non-zero coefficient estimate of SCAD regression using the λ_{opt} = 0.023616. SCAD method selects only 1 non-zero coefficient and shrinks the other coefficients to zero as displays in Table 3.

Table 3.4: Non-zero coefficient estimate of SCAD regression

Variable	Coefficient
Intercept	-0.06377521
Relative_Humidity_10am	0.09418830
Relative_Humidity_4pm	-
Air_Temp_Max.	-
Air_Temp_Min.	-
Earth_Temp_10am	-
Earth_Temp_4pm	-
Wind_Speed	-
Wind_Direction_10am	-
Wind_Direction_4pm	-
Open_Evap	-
Sunshine_Hours	-

The plot of the trace of coefficient against log Lambda (λ) for SCAD regression is displays in Figure 3.4

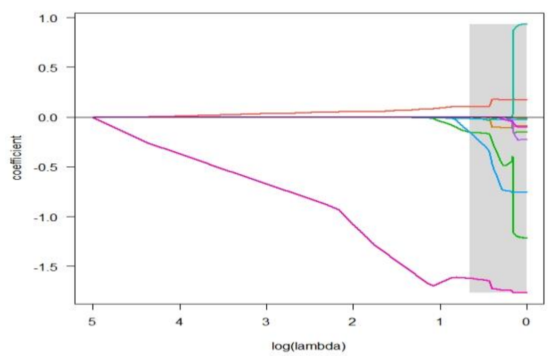

Figure 3.4: Coefficient estimates of βi for SCAD regression

SCAD model also tends to shrink the regression coefficients to zero as λ increases. Also, SCAD regression performs variable selection when λ is large enough.

Table 3.5: Summary of the coefficients for the four regression models

Variable	Coefficient			
	Ridge	Lasso	Elastic net	SCAD
(Intercept)	8.537348578	6.806265279	5.89671162	-0.06377521
Relative_Humidity_10am	0.018204825	0.080466055	0.07252912	0.09418830
Relative_Humidity_4pm	0.016838141	-	-	
Air_Temp_Max.	-0.006400661	-	-	
Air_Temp_Min.	0.049235794	-	-	
Earth_Temp_10am	-0.057764855	-	-	
Earth_Temp_4pm	-0.059309156	-	-	
Wind_Speed	-0.002888592	-0.003762919	-	
Wind_Direction_10am	-0.139587146	-0.006880016	-	
Wind_Direction_4pm	0.012553632	-	-	
Open_Evap	-0.069005492	-	-	
Sunshine_Hours	-0.250264196	-0.911241688	-0.80395374	

Table 3.5 portrays the regression coefficients for the four models. Ridge regression produced the coefficients of all the 11 variables including the intercept, because the model does not perform the variable selection. Lasso returned four predictor variables (Relative humidity-10am, wind speed, wind direction -10am and sunshine hours) that are important for prediction. Elastic net produced only two variables (Relative humidity-10am and sunshine hours) as important variables. Lastly, SCAD produced only one variable (Relative humidity-10am) as important variable.

4.0 GOODNSS OF FIT

The criteria used to select the best model for the forecast is the Root Means Square Error (RMSE). The lower the value of RMSE the better and 0 means the model is perfect, which may not be feasible in practice. From Table 3.6, with RMSE criteria, SCAD has the best performance compared to Ridge, Lasso and Elastic net. Lasso has a RMSE value of (51.44227) which is greater than the SCAD RMSE value of (1.74077) and also, Lasso has more variables to be included in the model. Ridge, Lasso and Elastic net performance are almost the same as displayed in Table 3.6 Based on the complexity of the model, SCAD produce a simpler model than the other three models. Ridge with the highest RMSE of (51.69574) performed less compared to the other three penalty functions. The results obtained in this study are almost the same with the results obtained by Fan and Li (2001) which indicate that, SCAD method has the best performance in selecting significant variables with simple and almost unbiased model.

Table 3.6: Model Goodness of Fit Test

Penalty Function	RMSE
Ridge	51.69574
Lasso	51.44227
Elastic net	51.51351
SCAD	1.74077

5.0 CONCLUSION

Variables that influence the total rainfall have been assessed based on the four penalized penalty functions (Ridge, Lasso, Elastic net and SCAD). SCAD model produces sparse solution by selecting the least variable. Ridge regression performed less based on RMSE criteria. Lasso produces more complex model by selecting variables more than Elastic net and SCAD. The SCAD method was found to be effective and provide better fit to the data. The model can also be used for rainfall forecasting for the study area.

6.0 REFERENCES

Adhikary, S. K., Yilmaz, A. G., & Muttil, N. (2016). Optimal design of rain gauge network in the MiddleYarra River catchment, Australia. Hydrological Processes,29(11), 2582–2599

Ani K. J., Anyika V. O. &Mutambara E. (2022). “The impact of climatic change on food and human society in Nigeria ”*international Journal of Climatic Change Strategies and Management* Vol. 14 No. 2, pp 148-167 http://doi.org/10.1108IJCCSM-11-2020-0119

Annual Performance Survey (APS) Report (2019/2020/2021). National Agricultural Extension and Research Liason Service.

Borhara K., Pokharel B., Bean B., Deng L. & Wang S. Y. S. (2020). On Tanzania’s Precipitation Climatology, Variability, and Future Projection. Climate, 8(2), pp 34.

Desboulets L. D. D. (2018). A Review on Variable Selection in Regression Analysis. CNRS, EHESS, Centrale Marseille, AMSE, Aix-Marseille University,

Fan, J. & Li, R. (2001). Variables selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Society Association, 96:1344 -1360.

Frank E. & Matthias D. (2019). High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection.

Garuba, H.S., Mukhtar, M. K., Ugama, G.N., Jamila,A.,Aliyu, M. K. & Yahaya, G. (2021). Evaluation of Rainfall Trends Due to Climate Change: A Case Study of Samaru, Zaria. FUDMA Journal of Sciences (FJS). DOI: https://doi.org/10.33003/fjs-2021-0504-807.

Gebrechorkos S. H., Hülsmann S., & Bernhofer C. (2018). Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. International Journal of Climatology, 39(1), 18–30. https://doi.org/10.1002/joc.5777

Hoerl, A. E. and Kennard, R. W.(1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. Vol. 12, pp. 55-67

James G., Witten D., Hastie T. & Tibshirani R. (2013) An Introduction to Statistical Learning, Springer, vol. 112, p. 18, New York, NY, USA.

Liu, W. & Li, Q. (2017). An efficient elastic net with regression coefficients method for variable selection of spectrum data.

Niles M. T; Lubell M. & Brown M. (2015) how limiting factors drive agricultural adaptation to climate change. Agric Ecosyst Environ 200, pp 178–185

Panga, P. A., Nyimvua S. & , Mbalawata (2021). Forecasting Rainfall in Tanzania Using Time Series Approach. International Journal of

Scientific and Innovative Mathematical Research Vol. 8 No. 4. doi.org/10.20431/2347-3142.0804002 pp 14 – 27

Park, H. & Konishi, S. (2015). Robust logistic regression modelling via the elastic net Type regularization and tuning parameter selection Journal of Statistical Computation and Simulation, 86(7): 1- 12.

Sarpong S. A., Nuamah N. N. N., Avuglah R. K., Seungyoung O. and Youngjo L. (2016). Performance Assessment of Penalized Variable Selection Methods Using Crop Yield Data International Journal of Probability and Statistics, 5(1)DOI:10.5923/j.ijps.20160501.02

Steele, A.J.; Cakiroglu, S.A.; Shah, A.D.; Denaxas, S.C.; Hemingway, H.; Luscombe, N.M. (2018). Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease.

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. Sci Rep 10, 13768 https://doi.org/10.1038/s41598-020-70816-2

Tang Z, Engel B. A., Lim K. J., Pijanowski B. C. & Harbor J. (2005) Minimizing the impact of urbanization on long term runoff. J Am Water Resour Assoc 41, 1347–59
009.

Tibshirani R.(1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B, Vol. 58, pp. 267-288

Wang Y. (2020). A Survey of Methods in Variable Selection and Penalized Regression. SAS Institute Inc.

Wiltz J. (2018). Sciencing https://sciencing.com/importance-of-rain-water-5275473.html

Yousef M. A. (2019). Two-Stage SCAD Lasso for Linear Mixed Model Selection. Bowling Green State University USA. Unpublished PhD Statistics Thesis

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net.

Journal of Royal Statistical Society B, 67(2):301-320.

Zou H. & Zhang H. H (2009) “On the Adaptive Elastic-net with a diverging number of parameters,” Annals of Statistics, vol. 37, no. 4, pp. 1733–1751, 2