

Efficient Model Selection For Moisture Ratio Removal Of Seaweed Using Hybrid Of Sparse And Robust Regression Analysis

Anam Javaid^{1 2*}, Mohd. Tahir Ismail¹, M.K.M. Ali¹



* Corresponding Author

1 School of Mathematical Sciences, Universiti Sains Malaysia majidkhanmajaharali@usm.my,
m.tahir@usm.my

2 Department of statistics, The Women University Multan, Pakistan, anamjavaid7860@yahoo.com

Abstract

The Internet of things (IoT) consisted of physical devices networks such as sensors, home appliances, electronics, and software's. It enables us to collect and exchange data in several fields. After data collection from IoT, variable selection is considered a major problem because many variables are involved in real life datasets. The current study focused on large data analysis of the problem of model selection, including interaction terms. The dataset used in this study is taken from solar drier with moisture ratio removal (%) as dependent variable while ambient temperature, chamber temperature, collector temperature, chamber relative humidity, ambient relative humidity, and solar radiation as independent variables. LASSO with Huber M, LASSO with Hampel M and LASSO with Bisquare M are used in this study. Comparison of techniques are made with ridge regression and OLS (ordinary least square) after multicollinearity test and coefficient test. MAPE (mean absolute percentage error) for the efficient selected model is used to forecast with its minimum possible value. As a result, the hybrid model of LASSO with Bisquare-M is considered as the efficient model with its minimum MAPE value. Thus, the resulting model with the selected variables can be used to predict Moisture Ratio Removal (%) to determine seaweed drying behavior.

Key Words: Shrinkage estimator, Robust estimators, Ridge regression, MAPE

Mathematical Subject Classification: 92B15, 62J07, 62F07, 62P10

1. Introduction

The Internet of Things (IoT) involves different types of software, electronics, sensors, and home appliances from which data can be connected and exchanged. Using these systems, the computer-based system can be designed to improve economic benefits, efficiency and results in the reduction of human exertions by Malavade and Akulwar (2017). Using IoT devices, data can be collected from the sensor to determine factors such as humidity, air speed, temperature, and water irrigation. The population is on the rise and there are nearly 797 million people facing food insecurity problems, which are considered to be a major problem in the agricultural sector Ahmed et al., (2017). Seaweed is the product used to manufacture food, the medical and manufacturing industries in the fields of agriculture. Seaweed appears to be a potential source for renewable energy, but it can also be converted to energy such as biofuel and gas. Disa et al., (2011) considered the carrageenan to be the main cause of seaweed extraction.

Ali et al., (2015) subsequently, in their study, carrageenan was examined as the main product used in cosmetics, per food, human food and non-food and meat binder, etc. Numerous forecasting models have been developed from the study of previous research using empiric or analytical solutions from Neitsch et al. (2011).

One of the techniques for model selection is OLS (Ordinary Least Square), but do not perform well in case of certain conditions are violated Zuur et al., (2009). Multiple regression was used by Rischbeck et al., (2016) to study mildly drought-stressed field and had a multicollinearity problem due to many variables in the model. If there is multicollinearity among variables then the variance and covariance for the estimate of OLS increases, making the estimates as inaccurate Gujarati (2004). Mendelsohn and Dinar (2003) used linear and quadratic Regression analysis for agriculture, water and climate with interactions terms. One of the uses of the linear regression analysis

was the estimation of the structural economic model for the increase in agricultural productivity by Pender et al. (2004). To address the multicollinearity problem, Giacalone and Panarello (2017) introduced L_p norm estimation methods. One contribution is the use of analysis of variance (ANOVA) in the assessment of the yield factor used by Neitsch et al., (2011). Also, Xu and Ying (2010) made variable selection using median regression with LASSO-type penalty. Afterwards, Zhao et al., (2012) investigated wavelet-based LASSO for the purpose of regressing function scalars. The simulation data and the real dataset were used for the investigation of asymptotic convergence and for the performance of the finite sample. Zhang et al., (2016) used LASSO, Adaptive LASSO, Adaptive LASSO II, Multitask LASSO, reweighted LASSO for quantitative trait loci analysis (eQTL) in the study of human cancer-related factors. According to Shariff and Ferdaos (2017), the Ridge regression is one of the methods used in the case of multicollinearity, but its results are affected in the case of outliers. For data analysis, outliers may have a negative impact on the results. Gad and Qura (2016) reviewed different types of robust methods in the case of outliers. Some practical lower bound (LB) and some upper bound (UB) were proposed by Midi et al., (2011) in the case of high Leverage Collinearity Influential Measure (HLCIM) and are considered to be essential measure for the detection of multicollinear degrees. Model selection was also made by different researchers as Abdullah et al., (2008) used 8Sc explained in Table 2 to obtain the best model among all possible models, similar to Yahaya et al., (2013) presented a logistic regression model using 8SC. Later on, Javaid et al., (2019) conducted the study regarding the efficient model selection for the moisture ratio removal of seaweed using solar drier. For that purpose, multiple linear regression analysis was used up to four phases. The efficient model selection was made with the significant factors selected in the analysis. Total of 32 models were entertained in that study up to third level interaction term. While the hybrid models of the LASSO and robust regression analysis is presented in the current study. All possible models containing variables with their interaction terms is analyzed. Efficient model selection was made with 8SC and MAPE was used to predict the best model for moisture ratio removal by Javaid et al., (2019). While in this study, total of 192 models are entertained. The hybrid model selection is made by using LASSO and robust estimators up to fifth order interaction terms. The large dataset is used in the study so that the results can be more efficient in term of MSE.

2. Methodology

Multiple Regression

Stuart (2011) explained the regression analysis as a widely used technique with x as an independent variable, y as a dependent variable and ε as an error term. The design matrix X can then be defined as the vector Y and ε as follows.

$$X = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \text{ and } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Thus, the classic linear model is considered as $Y = X\beta + \varepsilon$ where the aim for the least square estimation is to minimize of (1) for getting the estimates for β .

$$\begin{aligned} \sum_{i=1}^n \varepsilon_i^2 &= \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \\ \frac{\partial}{\partial \beta} (\sum_{i=1}^n \varepsilon_i^2) &= \frac{\partial}{\partial \beta} (Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta) \\ &= 0 - X^T Y - X^T Y + 2(X^T X)\beta \end{aligned} \tag{1}$$

Thus, the estimate of least square $\hat{\beta}$ is solution to

$$X^T X \hat{\beta} = X^T Y$$

As it minimizes $\hat{\varepsilon}^T \hat{\varepsilon} = \sum_{i=1}^n r_i^2$ so in case of $X^T X$ as nonsingular, the estimates of the least square can be directly estimated from the data.

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Ridge Regression

Ridge Regression is used in the case of a Multicollinearity problem as according to Miller (2002), a biased estimator $b(d)$ can be obtained by using all variables as follows.

$$b(d) = (X'X + dI)^{-1}X'y$$

where d can be considered as scalar of the positive values range and the predictors should be standardised with zero means and sum of squares for any columns of X elements as one. This means that the correlation matrix should be replaced instead of the $X'X$ and afterwards, a plot of b can obtain (d) against d called the ridge trace. According to them, ridge regression justification may be the same as for subset selection in case of bias variance trade off and the bias amount may be small in case of small d and the variance reduction may be substantial. Typically, this type of application occurs in the case of ill-conditioned matrix. The singular-value decomposition (s.v.d) can be seen for X having n rows and k columns when viewing the ridge regression situation.

$$X_{n \times k} = U_{n \times k} \Lambda_{k \times k} V_{k \times k}$$

With U and V column as normalized and orthogonal as $U'U = I$ and also $V'V = I$ where Λ as diagonal matrix having a diagonal as a singular value. Then, with λ_i as $i = 1, 2, \dots, k$ as the singular values with a smaller value as variance but with an addition of d , the contribution can be significantly reduced.

$$X'X = V'\Lambda^2V$$

s.v.d with a biased estimates of regression coefficients as

$$b(d) = V'(\Lambda^2 + dI)^{-1}\Lambda U'y$$

with λ_i as $i = 1, 2, \dots, k$ as the singular values with a smaller value as variance but with an addition of d , the contribution can be significantly reduced

LASSO Regression

Tibshirani (1996) first introduced *LASSO* which can select coefficients β which can be minimized,

$$\begin{aligned} & (y - X\beta)(y - X\beta)' + \lambda \sum_{j=1}^p |\beta_j| \\ & = (y - X\beta)(y - X\beta)' \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq s \end{aligned} \quad (2)$$

where s and λ are non-negative regularization parameters. In *LASSO*, L_1 constraints explain coordinates vertices and the edge polo type where some of the coordinate values are zero. Solutions for *LASSO* are commonly found on the polo type vertex or on the edges of the polo type, so that *LASSO* can be called a variable selection method where the coefficient of shrinkage to zero can remove variables from the model. Later on Hastie et al., (2009) defined the *LASSO* as the shrinkage method like the ridge regression, with subtle but with important differences. They defined the *LASSO* estimates as

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2$$

$$\text{Subject to } \sum_{j=1}^p |\beta_j| \leq t.$$

Just like the ridge regression, the constant β_0 can be parameterized by doing standardization to the predictors. $\hat{\beta}_0$ solution is \bar{y} . According to them, the *LASSO* problem can also be written in an equivalent lagrangian form.

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \}$$

It can be noted that like the ridge regression problem, penalty L_2 is only replaced by *LASSO* penalty L_1 . *LASSO* penalty does not make a linear solution in y_i and there will be no closed form solution as in ridge regression. *LASSO* computing solution is a quadratic programming issue. There are different efficient algorithm paths that can be used for the computation of the entire solution path as for the variation of λ containing the same computational cost as for the ridge regression. Because of this constraint's nature, since t will be sufficiently small, some of the coefficients will be exactly zero. so Therefore, *LASSO* can make a continuous selection of

subsets. If t is selected as larger than $t_0 = \sum_1^p |\hat{\beta}_j|$ where $\hat{\beta}_j = \hat{\beta}_j^{ls}$, then estimates of LASSO are $\hat{\beta}_j$'s. On the other hand, for $t = t_0/2$ say then, in that case, the coefficients of the least square are reduced by about 50% on average. Like the size of the subset in the selection of the variable subset or in the ridge regression of the penalty parameter, t should be chosen in an adaptive way to minimization the expected prediction error estimation.

Robust Regression

Robust regression involves different kinds of estimation. The most used are M estimators.

Draper and Smith (1998) state that minimizing $\sum_{i=1}^n \rho(\frac{\epsilon_i}{s})$ for finding an estimation of M, a partial differentiation is required with respect to each parameter p resulting in a system of p equations where x is a dependent variable, ψ is an error term and s is a standard deviation.

$$\sum_{i=1}^n x_{ij} \psi(\frac{y_i - x_i^T \beta}{s}) = \sum_{i=1}^n x_{ij} \psi(\frac{\epsilon_i}{s}) = 0, \quad j = 1, 2, \dots, p \quad (3)$$

Where $\psi(u) = \frac{\partial \rho}{\partial u}$ called as a score function and the weight function can be defined as,

$$w(u) = \frac{\psi(u)}{u}$$

results from $w_i = w(\frac{\epsilon_i}{s})$ for $i = 1, 2, \dots, n$, with $w_i = 1$ if $\epsilon_i = 0$, substituting into (3) the results are

$$\begin{aligned} \sum_{i=1}^n x_{ij} w_{ij} \frac{\epsilon_i}{s} &= \sum_{i=1}^n x_{ij} w_i (y_i - x_i^T \beta) \frac{1}{s} = 0 \quad j = 1, 2, \dots, p \\ \sum_{i=1}^n x_{ij} w_i (y_i - x_i^T \beta) &= 0, \quad j = 1, 2, \dots, p \\ \sum_{i=1}^n x_{ij} w_i x_i^T \beta &= \sum_{i=1}^n x_{ij} w_i y_i \quad j = 1, 2, \dots, p \end{aligned} \quad (4)$$

Since $s \neq 0$, defining the weight matrix $W = \text{diag}(\{w_i: i = 1, 2, \dots, n\})$ as follows

$$W = \begin{pmatrix} w_1 & . & 0 \\ & w_2 & \\ 0 & . & w_n \end{pmatrix}$$

yields the following matrix form of (4)

$$\begin{aligned} X^T W X \beta &= X^T W Y \\ \hat{\beta} &= (X^T W X)^{-1} X^T W Y \end{aligned} \quad (5)$$

It is very similar to the solution of the least square estimator, but with the introduction of a weight matrix to reduce the outlier's influence. Usually, in contrast to the least squares, equation (4) is not possible to calculate the M-estimation directly from the dataset, W depends on the residuals, that depend on the estimation. In fact, an initial estimate and iterations are required, to finally converge on W and M-estimate for β . Iterative procedure IRLS is used to identify the M-estimates of regression. Stuart (2011) defined the functions for the least square and M estimators as follows.

Table 1: Weight Function Used for Different Regression Methods

	Objective Function $\rho(u)$	Score function $\psi(u)$	Weight Function $w(u) = \frac{\psi(u)}{u}$
a)Least Squares	$\frac{1}{2}u^2 \quad -\infty \leq u \leq \infty$	u	1
b)Huber (1973) a>0	$\begin{cases} \frac{1}{2}u^2 & \text{if } u < a \\ a u - \frac{1}{2}a^2 & \text{if } u \geq a \end{cases}$	$\begin{cases} u & \text{if } u < a \\ a \operatorname{sign} u & \text{if } u \geq a \end{cases}$	$\begin{cases} 1 & \text{if } u < a \\ \frac{a}{ u } & \text{if } u \geq a \end{cases}$
c)Hampel, a,b,c>0	$\begin{cases} \frac{1}{2}u^2 & \text{if } u < a \\ a u - \frac{1}{2}a^2 & \text{if } a \leq u < b \\ a\frac{c u - \frac{1}{2}u^2}{c-b} - \frac{7a^2}{6} & \text{if } b \leq u \leq c \\ \frac{a(b+c-a)}{6} & \text{Otherwise} \end{cases}$	$\begin{cases} u & \text{if } u < a \\ a \operatorname{sign} u & \text{if } a \leq u < b \\ a\frac{c \operatorname{sign} u - u}{c-b} & \text{if } b \leq u \leq c \\ 0 & \text{Otherwise} \end{cases}$	$\begin{cases} 1 & \text{if } u < a \\ \frac{a}{ u } & \text{if } a \leq u < b \\ 0 & \text{Otherwise} \end{cases}$
d)Tukey Bisquare, a>0	$\begin{cases} \frac{a^2}{6}\left(1 - \left(1 - \frac{u^2}{a^2}\right)^3\right) & \text{if } u \leq a \\ \frac{1}{6}a^2 & \text{if } u > a \end{cases}$	$\begin{cases} u\left(1 - \left(\frac{u}{a}\right)^2\right)^2 & \text{if } u < a \\ 0 & \text{if } u > a \end{cases}$	$\begin{cases} \left(1 - \left(\frac{u}{a}\right)^2\right)^2 & \text{if } u \leq a \\ 0 & \text{if } u > a \end{cases}$

Value for typical tuning constant of Huber is $a = 1.345$ for 95% relative efficiency, For Hampel is $a = 2$, $b = 4$ and $c = 8$ and for Tukey's Bisquare is $a = 4.685$ results in 95% relative efficiency Stuart (2011).

To data analysis, the following phases are performed in order to obtain efficient model selection using different regression analysis techniques.

Phase I – All Possible Models

According to Khuneswari et al., (2008), all possible models are a prerequisite for determining the best model and can be derived using the following formula,

$$N = \sum_{j=1}^k j \binom{k}{j} \quad (6)$$

where N is considered to be the number of all possible models, k represents the total number of independent variables and $j=1, 2, \dots, k$.

All possible models will be used and *OLS*, *LASSO* and *Ridge* will be applied to them after the procedure will be moved to next phase.

There was no missing value on the dataset. For all six independent variables, 5% of the data was randomly selected. Mean Absolute Percentage Error (MAPE) was calculated using the dataset to be used in Phase 3. for the forecasting of the best model for the selected model, two tests will be performed for OLS, *i.e.* the multicollinearity test and the coefficient test.

Phase 2- Selected Models

For the forecasting of the best model from the selected model, two tests will be performed for OLS, *i.e.* the multicollinearity test and the coefficient test. For the coefficient test, 5% of the significance level is used. After these tests have been carried out, a list of selected models will be obtained in the OLS regression analysis, while only significant variables will be retained for the selected models in ridge and LASSO regression.

Multicollinearity test

Multicollinearity occurs when the model contains collinear predictors. In this case, results may not be valid Joseph (2009). Thus, it is important to handle this type of problem in data analysis before proceeding.

Coefficient test

Ramanathan (2002) defined the coefficient test, a test used for each independent variable to check whether it differs significantly from zero or not. That is, the following hypothesis is made for it.

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Where β_j denotes the coefficient of variable in the model for $j = 1, 2, \dots, k$ and the *t* test will be used for testing at a 5% level of significance.

Phase 3- The Best Model

After the selected models have been obtained, the best selection of the models is made between the selected models. Eight selection criteria were identified for this purpose by Ali et al., (2017). In this study, the eight selection criteria is used for the best model selection. The list of formulas for 8SC can be shown as in Table 2.

Table 2: List of 8SC

AIC: $\left(\frac{SSE}{n}\right) (e)^{2(k+1)/n}$ (Akaike, 1969)	RICE: $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{2(k+1)}{n}\right)\right]^{-1}$ (Rice, 1984)
FPE: $\left(\frac{SSE^2}{n}\right) \frac{n + (k+1)}{n - (k+1)}$ (Akaike, 1974)	SCHWARZ: $\left(\frac{SSE}{n}\right) n^{(k+1)/n}$ (Schwarz, 1978)
GCV: $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{k+1}{n}\right)\right]^{-2}$ (Golub et al., 1979)	SGMASQ: $\left(\frac{SSE}{n}\right) \left[1 - \left(\frac{k+1}{n}\right)\right]^{-1}$ (Ramanathan, 2002)
HQ: $\left(\frac{SSE}{n}\right) (\ln n)^{2(k+1)/n}$ (Hannan and Quinn, 1979)	SHIBATA: $\left(\frac{SSE}{n}\right) \frac{n + 2(k+1)}{n}$ (Shibata, 1981)

where

n = total number of observations

$k + 1$ = estimated parameters numbers (including constant)

SSE = sum of squares error.

Phase 4-Goodness of Fit

Gujarati (2004) has defined some assumptions for the least square estimate as no perfect multicollinearity or correct identification of the model. The goodness of fitness test was defined in Ramanathan (2002) as ensuring that the model fits well into the dataset. For this phase, the 5% dataset maintained will be used for model efficiency calculation using the MAPE value and then the residuals will be calculated using the difference between the actual and observed values using the best model obtained in Phase 2 so that the residual test is analyzed using the randomness and normality test. Also, a nonparametric test like a run test will be used to check the random pattern of observations. For normality check assumptions, some supporting documents such as Sharpio Wilk test and Kolmogrov Smirnov test used for this purpose. Afterwards, scatter plot, histogram, and box plot for residuals for supporting evidence are obtained.

3. Result and Discussion

Data Collection and Procedure

The data used in this study was taken from a seaweed dryer for four days-using V-groove hybrid solar dryers. Seven variables are used in this study as Moisture Ratio content (%) as the dependent variable with six independent variables as Ambient Temperature (x_1), Chamber Temperature (x_2), Collector Temperature (x_3), Chamber Relative Humidity (x_4), Ambient Relative Humidity (x_5) and Solar radiation (x_6) variables of Temperature from solar drier. The data for four days consisted of a total of 1924 observations. Where 1826 observations were taken for the purpose of the analysis and 98 observations were kept for the purpose of calculating the MAPE value to forecast. From 16 March 2017 to 19 March 2017, the data was taken for each minute from 9 am to 5 pm as solar radiation better works at that time. Total of 8 hours dataset is taken per day. So, overall, it was 32 hours dataset used in the study. For, the minutes, it was thus total of 1924 observations were obtained. The list of all possible models are kept in Table 3 for six variables taken in the analysis.

Table 3: All Possible Models for six independent variables

Number of variables	single	Interact					Total
		1st	2 nd	3 rd	4 th	5 th	
1	6	-	-	-	-	-	6
2	15	15	-	-	-	-	30
3	20	20	20	-	-	-	60
4	15	15	15	15	-	-	60
5	6	6	6	6	6	-	30
6	1	1	1	1	1	1	6
Total Models	63	57	42	22	7	1	192
Model ID	M1-M63	M64-M120	M121-M163	M164-M185	M186-M191	M192	

By using ordinary least square, all possible models are calculated as shown in Phase I. After this, the selected models are obtained through a multicollinearity test and a coefficient test on all possible models to be moved to Phase II. The selected models have thus obtained their respective SSEs. M192.0.0 was the original model with all possible interaction terms included in it and 46 variables were removed from the original model after the multicollinearity test was carried out. Following this step, the model became M192.46.0 and after that coefficient test was conducted, then the final model was obtainable in the form of M192.46.3, which indicates that 3 variables were removed in the coefficient test. Grouping is carried out based on the same variables in the model, e.g., M144.7.0 and M175.10.0 have the same variables so that they are kept in the same group. There are now 145 models left out of 192 models after grouping. These final models have no problem with multicollinearity, and all the variables included are also significant. The procedure shall be applied to all possible models for obtaining the selected models. The best model is selected using eight selection criteria.

For OLS, the minimum value was found for M183.36.3 and the MAPE value was obtained using the reserved data. It was determined to be MAPE = 9.781944.

Stuart (2011) explained that the least square estimates cannot perform well in the case of outliers or in the case of normality assumptions are not met. Therefore, the performance of ridge regression is better than OLS in the case of highly collinear predictors in the model. Estimates derived from the ridge regression are biased, but the mean square error of ridge estimator is smaller than the OLS estimators of Hoerl and Kennard (1970). Since the data used in this study also suffers from a multicollinearity problem so next technique applied is the ridge regression on all possible models for the purpose of comparison and the selected models were obtained.

Ridge regression is a type of regularized regression analysis so that at least two variables should be included in the model starting with M5.0.0 containing two variables Hocking (1976). Ridge regression has the ability to make the correlation coefficient as a reduction to zero but cannot make it exactly zero because it contains the L2 penalized least square penalty term Ogutu et al., (2012) so that the best model selection in Phase 2, 8SC has been used and the selection for the best model is obtained with a minimum value among all models. The best model selected with its coefficient was obtained.

OLS and ridge regression do not have the ability to select the model as they contain all variables in the model so modified LASSO has been performed for handling outliers in the dataset for the purpose of selecting crucial variables among all variables including their interaction terms. Modified LASSO was performed using three robust M estimator methods, i.e., Huber's M estimator, Hampel's M estimator, and Bisquare's M estimator.

For the Sparse regression analysis, LASSO is performed on all possible models and the results have been analyzed. The list of 183 models were obtained by grouping the same models as in OLS. Robust methods are applied to these resultant 183 models to obtain the selected models in the modified LASSO. 151 models were obtained as selected models using the Huber M estimator method for the above 183 models obtained from LASSO. Coefficient test is performed at 0.05 a level of significance for removal the insignificant variables from the models in Huber's M estimation. Thus, after performing the coefficients tests, only 151 models are left in the Huber's M estimation method. Eight selection criteria for the best model selection were performed on these 151 models and results were observed. The best model M192.27.11 with SSE as 63759 was observed after using 8SC. The second method used is the

Hampel's M estimator, which is applied to selected models of LASSO. The final selected model obtained from Hampel with LASSO was observed as M192.27.10 with SSE 63269. The third Robust method used in this study was the Tukey's Bisquare method. 144 models were selected after grouping the model and the best model was observed as M192.27.13 with its SSE as 64743

Table 4: MAPE for Final Selected Models from all Methods

Selected Model	Technique used	No of Variables	SSE	MAPE
M183.36.3	OLS	16	75921	9.782
M183.0.0	Ridge Regression	52	86405.7	10.771
M190.27.11	LASSO with Huber M	25	63886	9.0208
M190.27.10	LASSO with Hampel	26	63269	9.118
M190.27.13	LASSO with Bisquare	23	64743	8.968

Table 4 shows that LASSO with Bisquare's M provides a good fit than other methods. From the methods used in Table 4, OLS uses equal weights $1/n$ for all observations. Ridge regression MAPE is high as it can deal only with multicollinearity, not with outliers. If we compare with OLS, there is a 15.85% decrease in SSE in Huber's M, there is a 16.66% decrease in Hampel's M, a 14.72% decrease in LASSO with Bisquare's M, while in ridge there is a 12.13% increase in SSE compared to OLS. If the number of variables is observed, ridge regression contains more variables than others, but cannot perform well in the case of outliers. In the same way, OLS has a very small number of variables in the final model due to the performance of the multicollinearity and coefficient test. The SSE for OLS is higher than the technique proposed. The proposed technique has almost the same number of variables in the model with a slight difference between SSE and MAPE. Robust estimators can deal with outliers. The combination of LASSO and Bisquare provides good results from the above hybrid models as opposed to other techniques. Standardized residual plots are observed as supporting documents for this purpose. Standardized residual graph has been observed for each method and can be observed as in Figure 1-5.

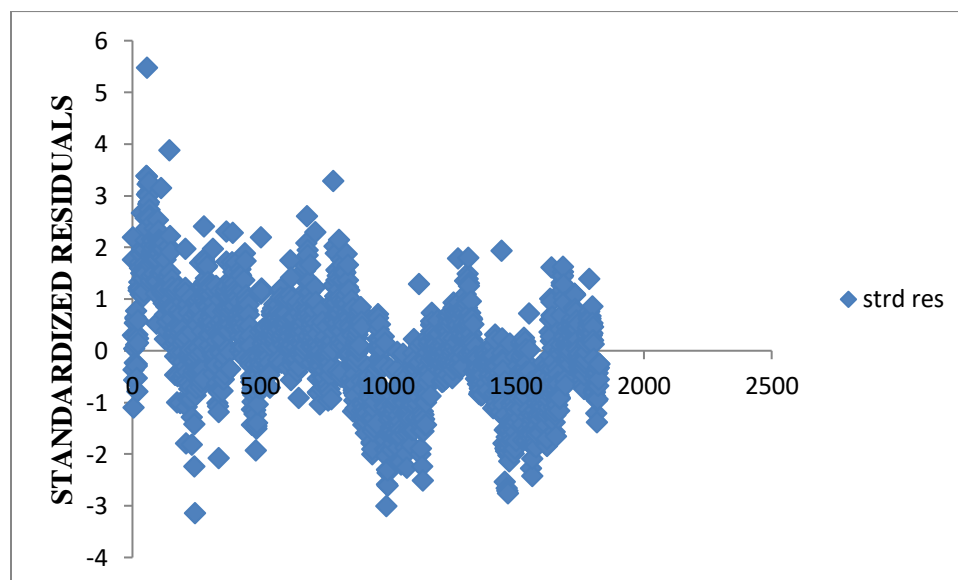


Figure 1. Standardized residual for OLS

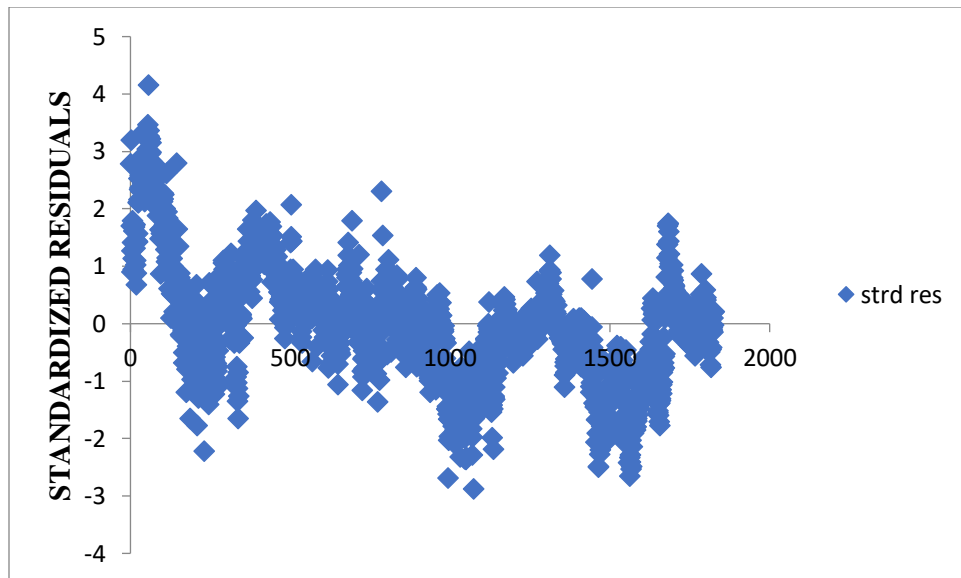


Figure 2. Standardized Residual by using Ridge Regression

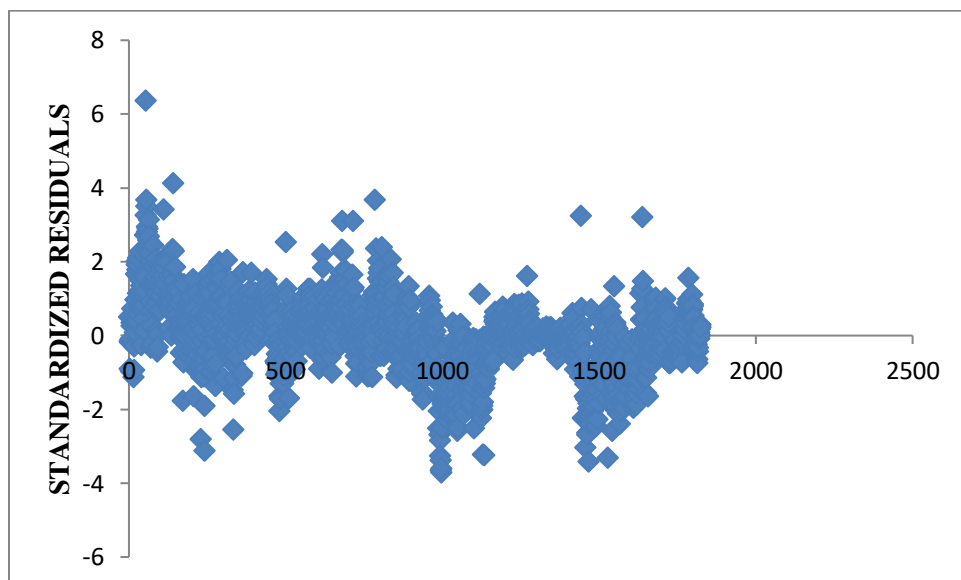


Figure 3. Standardized Residual Using Huber's M after LASSO

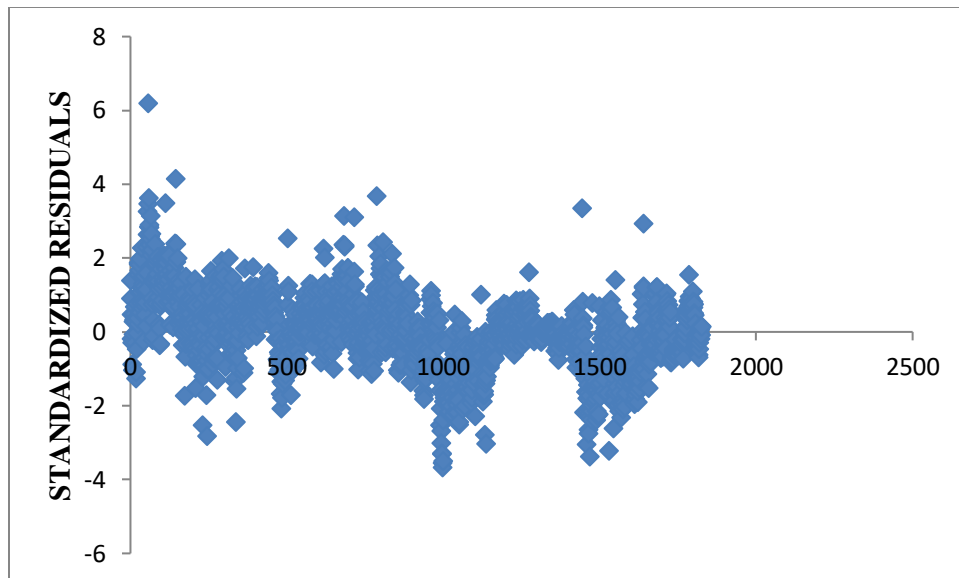


Figure 4. Standardized Residual by using Hampel's M after LASSO

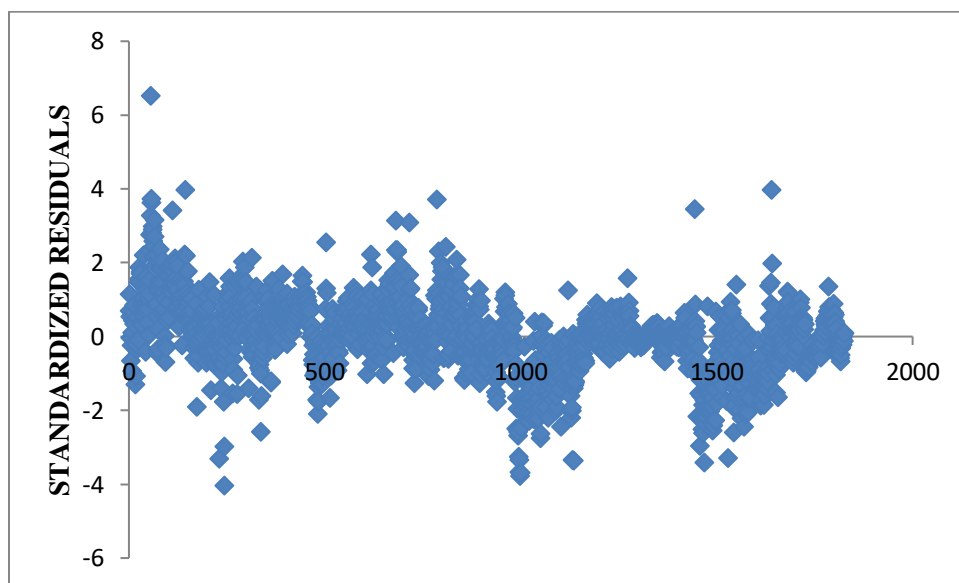


Figure 5. Standardized Residual by Using Bisquare's M after LASSO

From the figures 1- 5 above, outliers pattern can be observed. OLS and ridge have more outliers than the technique being proposed. Although the OLS model is selected after multicollinearity and coefficient test, it still contains more outliers. It is because OLS cannot deal with outliers, while ridge can also deal with multicollinearity only, thus, compared to the proposed technique, there are more outliers. in real life dataset, it is not always a good option to remove outliers. the outliers outside the sigma limits are observed in Table 5.

Table 5: Outliers outside the sigma limits

Technique	$\mu \pm \sigma$	$\mu \pm 2\sigma$	$\mu \pm 3\sigma$
OLS	30.668	4.326	0.712
Ridge	28.423	6.024	0.602
LASSO Huber M	13.801	5.531	1.2596

LASSO Hampel M	26.506	5.312	1.150
LASSO Bisquare M	26.068	5.476	1.258

For outliers, in standardized residuals plot, observations outside the sigma limits are observed in Table 5. It is noted that although the ridge regression and OLS has a smaller number of outliers in three sigma limits, but the SSE value was very much high for ridge regression analysis. So, the ridge regression cannot be considered as a good technique. While LASSO with Bisquare M has some observations outside three sigma limits, but due to the use of robust regression analysis, outliers are not affected on its coefficients and the model. That is the advantage of the proposed technique that LASSO can have only significant variables in the model. While robust estimators can deal with outliers. Thus, the proposed model can deal with the multicollinearity and outliers' issue in the real-life dataset.

4. CONCLUSION

Solar drier has its importance in the field of agriculture and aquaculture. There are many factors involved in solar drier efficiency, so it is important to identify those factors. Different variables affecting moisture ratio removal have been observed with their interaction effects by considering a large data. Hybrid of LASSO and robust regression is proposed for identification of significant factors including the interaction terms. The results show that the combination of LASSO and Bisquare's M provides better results compared to other methods. The value of SSE for the hybrid of LASSO and Bisquare's M is found be 8.9689, that is less than 10%. Thus the resulting selected model is now ready to predict the moisture ratio removal of seaweed with its selected significant factors.

Acknowledgment

The author would like to thank to "The Women University Multan" and "University Sains Malaysia" for this research project.

References

1. Abdullah, N., Hajjibok, Z., & J.B, N. J. (2008). Multiple Regression Models of the Volumetric Stem Biomass Multiple Regression Models of the Volumetric Stem Biomass, 7(7), 492–502.
2. Ahmed, U. I., Ying, L., Bashir, M. K., Abid, M., & Zulfiqar, F. (2017). Status and determinants of small farming households' food security and role of market access in enhancing food security in rural Pakistan. *PLoS ONE*, 12(10), 1–15.
3. Ali M.K.M, A. Fudholi, M. S. Muthuvalu, J. Sulaiman and S. M. Yasir (2017) November. Implications of drying temperature and humidity on the drying kinetics of seaweed. *AIP Conference Proceedings*.1905 (1)
4. Ali, M. K. M., Sulaiman, J., Yasir, S. M., & Ruslan, M. (2015). The effectiveness of sauna technique on the drying period and kinetics of seaweed *Kappaphycus alvarezii* using solar drier. *Advances Environ Agri Sci*, 1, 86-95.
5. Dissa, A. O., Bathiebo, D. J., Desmorieux, H., Coulibaly, O., & Koulidiati, J. (2011). Experimental characterisation and modelling of thin layer direct solar drying of Amelie and Brooks mangoes. *Energy*, 36(5), 2517-252
6. Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons.
7. Gad, A. M., & Qura, M. E. (2016). Regression Estimation in the Presence of Outliers : A Comparative Study. *International Journal of Probability and Statistics*, 5(3), 65–72.
8. Giacalone, M., Panarello, D., & Mattera, R. (2018). Multicollinearity in regression: an efficiency comparison between L p-norm and least squares estimators. *Quality & Quantity*, 52(4), 1831-1859.
9. Gujarati, D. N. (2004). *Basic Econometrics*. The McGraw– Hill Companies.
10. Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *Springer series in statistics*.
11. Hocking, R. R. (1976). A Biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1-49.
12. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
13. Javaid, A., Muthuvalu, M. S., Sulaiman, J., Ismail, M. T., & Ali, M. K. M. (2019, December). Forecast the moisture ratio removal during seaweed drying process using solar drier. In *AIP Conference Proceedings* (Vol. 2184, No. 1, p. 050016). AIP Publishing LLC.
14. Joseph, A. B. *Multiple Alliant International University*. 1, 3-5

15. Khuneswari, G., H.J. Zainodin, G. Darmesah, & S.H. Sim (2008) *Malaysia Journal for Mathematical sciences*, **4**, 190-195.
16. Malavade, V. N., & Akulwar, P. K. (2017). Role of IoT in Agriculture. In *IOSR Journal of Computer Engineering* 1(13), 56–57
17. Mendelsohn, R., & Dinar, A. (2003). Climate, Water, and Agriculture. *Land Economics*, 79(3), 328–341.
18. Midi, H., Bagheri, A., & Imon, A. H. M. R. (2011). A Monte Carlo simulation study on high leverage collinearity-enhancing observation and its effect on multicollinearity pattern. *Sains Malaysiana*, 40(12), 1437–1447.
19. Miller, A. (2002). *Subset selection in regression*. Chapman and Hall/CRC.
20. Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). *Soil and water assessment tool theoretical documentation version 2009*. Texas Water Resources Institute.
21. Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012, December). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In *BMC proceedings* 6(2).
22. Pender, J., Nkonya, E., Jagger, P., Sserunkuuma, D., & Ssali, H. (2004). Strategies to increase agricultural productivity and reduce land degradation: evidence from Uganda. *Agricultural economics*, 31(2-3), 181-195.
23. Ramanathan, R. (2002) *Introductory Econometrics with application*. 5th edition.:Thomson Learning Ohio, South Western, United States
24. Rischbeck, P., Elsayed, S., Mistele, B., Barmeier, G., Heil, K., & Schmidhalter, U. (2016). Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. *European Journal of Agronomy*, 78, 44–59.
25. Shariff, N. S. M., & Ferdaos, N. A. (2017). An Application of Robust Ridge Regression Model in the Presence of Outliers to Real Data Problem. *Journal of Physics: Conference Series PAPER*, 890(1).
26. Stuart, C. (2011). *Robust regression. Guide to Statistics*.
27. Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society . Series B (Methodological)*, 58(1), 267–288.
28. Xu, J., & Ying, Z. (2010). Simultaneous Estimation and Variable Selection in Median Regression using Lasso-type penalty. *Annals of the Institute of Statistical Mathematics*, 62(3), 487–514.
29. Yahaya, A. H., Norianai, A., & Jubok, Z. H. (2013). *UTM*, 29(1), 11–16.
30. Zhang, K., Zhe, S., Cheng, C., Wei, Z., Chen, Z., Chen, H., Jiang, G., Qi, Y., Ye, J. (2016). Annealed Sparsity via Adaptive and Dynamic Shrinking. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 1325–1334.
31. Zhao, Y., Ogden, R. T., & Reiss, P. T. (2012). Wavelet-based LASSO in Functional Linear Regression. *Journal of Computational and Graphical Statistics*, 21(3), 600–617.
32. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Limitations of linear regression applied on ecological data. In *Mixed effects models and extensions in ecology with R* .11-33.