Pertanika J. Sci. & Technol. 28 (2): 609 - 625 (2020)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Comparison of Sparse and Robust Regression Techniques in Efficient Model Selection for Moisture Ratio Removal of Seaweed using Solar Drier

Anam Javaid[1,2], Mohd. Tahir Ismail[1] and Majid Khan Majahar Ali[1]*

[1]School of Mathematical Sciences, Universiti Sains Malaysia 11800 USM, Penang, Malaysia
[2]Department of Statistics, The Women University Multan, L.M.Q. Road, Multan, Pakistan

ABSTRACT

Solar drier is considered to be an important product used in the internet of things (IoT). It is used to dry different kinds of products used in agriculture or aquaculture. There are many factors that have different effects on the drying of items in the solar drier. The current study focused on the removal of the moisture ratio in the drying process for seaweed using solar drier. For this purpose, a dataset containing 1924 observations was used to study the effect of six different independent variables on the dependent variable. Moisture ratio removal (%) was considered to be dependent variable with ambient temperature, chamber temperature, collector temperature, chamber relative humidity, ambient relative humidity and solar radiation as independent variables. All possible models were used in the analysis till fifth order interaction terms. Hybrid model of LASSO with bisquare M was proposed for efficient selection of the model. The procedure based on four phases was used for efficient model selection and a comparison was made with other existing sparse and robust regression techniques. The result indicates that the proposed technique is better than other existing techniques in terms of mean squared error (MSE) and mean absolute percentage error (MAPE).

Keywords: All possible models, LASSO, model selection, robust, seaweed, selection criteria

INTRODUCTION

In agriculture field, the items based on the Internet of thing (IoT) help to reduce human effort as a kind of narrow band model was proposed by Klaina et al. (2018). Using IoT, different factors such as humidity, air speed, temperature and irrigation of water were determined by Gondchawar & Kawitkar

ARTICLE INFO

Article history:
Received: 05 Octobe 2019
Accepted: 15 January 2020
Published: 15 April 2020

E-mail addresses:
anamjavaid7860@yahoo.com (Anam Javaid)
m.tahir@usm.my (Mohd. Tahir Ismail)
majidkhanmajaharali@usm.my (Majid Khan Majahar Ali)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

(2016). There is a large increase in population, so there is need to produce more food to meet the demand of the population (Rockström et al., 2009). There are many steps involved in the food production process from seeding to harvesting (Yan, 2011). One of the most important step is the drying (Ali et al., 2014). Seaweed is considered to be the most widely used product in the aquaculture or agiculture sector (Dissa et al., 2011). In agriculture, mathematically there are many models produced for the purposes of forecasting (Neitsch et al., 2011). Ordinary least square (OLS) is one of the technique used in model selection, but it suffers from limitations in case of certain conditions violated (Zuur et al., 2009). Mendelsohn and Dinar (2003) used the linear and quadratic regression analysis with the interaction terms of factors. Giacalone et al. (2018) had introduced L_p norm estimation methods. These simple methods have no model selection capability, so that Xu and Ying (2010) performed the selection of variables using the median regression with least absolute shrinkage and selection operator (LASSO) type penalty. The presence of outlier is also considered a major data problem because removing such observations is not always a good solution, so there are robust methods used to deal with these types of observations, as Gad and Qura (2016) reviewed different types of robust methods for outliers. According to Shariff and Ferdaos (2017), tikhonov regularisation (ridge regression) is considered to be one of the methods used in the case of multicollinearity, but its results are affected in the case of outliers. In robust regression, many types of estimates are available as Susanti et al. (2014) compared maximum likelihood type estimators (M estimators), modified M estimates (MM) and estimators of scale (S) of maize production data, but M estimators are preferred by the majority of researchers as their advantages were demonstrated by Sinova and Van Aelst (2018). It was based on tukey bisquare and were compared with hampel loss function. Another method of robust ridge regression was provided by Shariff and Ferdaos (2017) in case of both multicollinearity and outliers problem. For the model selection purpose, eight selection criteria (8SC) were used by Ali et al. (2017). From all possible models the 8SC was used by Zainodin et al. (2011) in model selection problem. It can be seen from previous research that there are different methods such as OLS, ridge and LASSO with robustic approach have been used, but no studies have been conducted using a combination of LASSO and robust with 8SC for all possible models, including interaction terms. Therefore, the contribution is the use of a newly-developed hybrid model of LASSO and robust with all possible models. The best choice of these models is made by means of 8SC that can be used further to choose the efficient model to forecast.

METHODS

This study used hybrid of LASSO and robust regression.The details of the methodology used are discussed as follows.

LASSO

Tibshirani (1996) proposed a new sparse estimation method called "LASSO" that minimised the sum of squares subject to a restriction that the sum of absolute value of the coefficient was less than the constant value. This kind of constraint has the capacity of sparsness, as some coefficients will be exactly zero, so the resulting model would have a better interpretation. LASSO has the property of both subset selection and ridge regression analysis simultaneously. It is a very general method as it can be applied in different statistical methods such as in extension of generalised linear model and in tree based models. According to Zhang et al. (2016) if there is a response vector $Y = [y_1, y_2, \ldots, y_n]$ and the predictors $X \in R^{n \times D}$, Then in case of without generality of data loss. LASSO, a sparse regression method has the ability to resolve the following problem.

$$min \, ||Y - X\beta||_2^2 + \lambda || \beta ||_1$$

Where $\beta \in R^{D \times 1}$ is considered as vector of regression coefficient. L_1 norm regularization has the ability to provide as sparse solution so that the model can be easily interpreted.

M Estimator

Draper and Smith (1998) stated that for finding the maximum likelihood type estimate (M estimate), it is required to minimize the term $\sum_{i=1}^{n} \rho(\frac{\varepsilon_i}{s})$, where ε_i is the error term of i^{th} observation and s is an estimate of the scale. For this purpose, a partial differentiation is used with respect to each parameter p which results in a system of p equations.

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

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

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

yeilds $w_i = \left(\frac{\varepsilon_i}{s}\right)$ for $i = 1, 2, \ldots, n,$ with $w_i = 1$ if $\varepsilon_i = 0$, substituting into (1) the results are

$$\sum_{i=1}^{n} x_{ij} \, w_i \frac{\varepsilon_i}{s} = \sum_{i=1}^{n} x_{ij} w_i \, (y_i - x_i^T \beta) \frac{1}{s} = 0 \quad j = 1,2,\ldots,p$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij} \, w_i (y_i - x_i^T \beta) = 0, \qquad j = 1,2,\ldots,p$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij} \, w_i x_i \beta = \sum_{i=1}^{n} x_{ij} w_i y_i \qquad j = 1,2,\ldots,p \qquad (2)$$

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

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

yields the following matrix form of (2)

$$X^T W X \beta = X^T W Y$$

$$\Rightarrow \hat{\beta} = (X^T W X)^{-1} X^T W Y \qquad (3)$$

It is very similar to least square estimator solution, but the introduction of weight matrix reduce the outlier's influence. Usually, contrasting to the least squares, (3) can not be used directly in calculation of M estimation 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 an M estimation for *β*. An iterative procedure called as iteratively reweighted least squares (IRLS) is used to identify M regression estimates. There are three types of M estimators commonly used as huber M, hampel M and tukey bisquare M with different weighting function (Stuart, 2011). In the present study, bisquare M is used for making the hybrid model with LASSO.

All possible models are considered in this study. Efficient model selection is made in four phases. The purpose of the study is also to highlight the importance of interaction terms so the variables till fifth order interaction terms are considered. All the assumptions regarding the random pattern of observations, homogeneity of variances and autocorrelations are fulfilled.

Phase I – All Possible Models

Ali et al. (2017) stated all possible models using the formula

$$N = \sum_{j=1}^{k} j \begin{pmatrix} k \\ C \\ j \end{pmatrix} \qquad (4)$$

Using (4), all possible models are calculated with LASSO and with all other existing techniques used in this study. Total number of all possible models for six independent variables untill fifth order interactions can be observed as in Table 1.

Table 1
Number of all possible models

No of variables single		**Interact**					
		1st	2nd	3rd	4th	5th	Total
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	

A total of 5% of the dataset is stored for forecasting purposes and the mean absolute percentage error (MAPE) value is used for forecasting estimates (Ali et al., 2017) in phase 4. The MAPE is calculated using (5).

$$MAPE = \frac{100}{N}\left(\frac{\sum_{i=1}^{j}|A_i - E_i|}{A_i}\right) \quad i=1,2,\ldots,j \qquad (5)$$

Where
A = Actual value of dependent variable (y)
E = Expected value ($\hat{y}$)
N = Number of fitted points

Phase 2- Selected Models

After calculating all possible models, the next step is to obtain the selected models (Zainodin et al., 2011). For this purpose, bisquare M was applied to LASSO selected models at a 5% significance level. Significant factors had been observed. From other techniques, significant factors were also observed at a 5% level of significance.Only one non significant variable was removed at one time and the procedure rerun again. The procedure continued until all the variables remaining in the model were significant.

Phase 3 - The Best Model

The next step was to get the best model after a list of selected models was obtained. 8SC were defined for this purpose by (Zainodin et al., 2011). 8SC formula can be displayed as shown in Table 2. By using mentioned formulas in Table 2, Akaike information criterion *(AIC*), *RICE*, Final prediction error (*FPE*), SCHWARZ(*SBC*), Generalized cross validation

(*GCV*), Sigma square (*SGMASQ*), Hannan-Quinn information criterion (*HQ*) and *SHIBATA* were calculated. The efficient model was obtained on the basis of minimum value obtained from all mentioned criteria.

Table 2
Formula used for 8SC

AIC: $\left(\frac{SSE}{n}\right)e^{\frac{2(k+1)}{n}}$ (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:* $\left(\frac{SSE}{n}\right)n^{(\frac{k+1}{n})}$ (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: $\left(\frac{SSE}{n}\right)(lnn)^{\frac{2(k+1)}{n}}$ (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 square of error

Phase 4 - Goodness of Fit

The goodness of fit test was performed on the final models selected in phase 3 to check the efficiency of the selected model. 5% dataset kept in phase 1 was used for MAPE calculation using (5) for this purpose. Other supporting evidence, such as scatter plot, histogram and residual box plot, was obtained for supporting evidence.

RESULT AND DISCUSSION

Data Collection and Procedure

The data used in this study were taken from a seaweed drier for four days using V-groove hybrid solar dryer. Seven variables were used in this study with moisture ratio content(%) as dependent variable while six independent variables include 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). Moisture ratio content (%) was basically decreasing with the time passed by. So the time effect was already in study of the percentage decrease in moisture ratio. Significance of interaction terms had also been observed in this study. So, x_{12} represents the interaction between x_1 and x_2. Similarly all other interactions are presented in this study. The four days of data were taken from a total of 1924 observations where 1826 observations were made for the purposes of analysis and 98 observations were kept for the purpose of prediction by calculating MAPE value. The data for each second was collected from 8 a.m. to 5 p.m. from 16 March 2017 to 19 March 2017. All possible models for the six independent variables were calculated from Table 1. On these 192 models, LASSO was applied and 183 models were obtained. Models with the same number of variables were stored in the same group. The tukey bisquare M estimator was applied to these 183 models at a 5% significance level in phase 2. As a result, 144 models were obtained from the application of the M estimator. Out of the 144 selected models, an efficient selection of the models was made on the basis of 8SC in phase 3. The minimum value for 8SC were found for model M192.27.13 meaning that M192.0.0 was original model where 27 variables were removed in LASSO and 13 were removed in bisquare M, thus the final model became M192.27.13 with SSE as 64743. The results obtained from 8SC are observed in Table 3.

Table 3
Results for 8SC using LASSO with bisquare M

Model number	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M7.1.0=M9.1.0=M64.1.1	156.01	156.01	156.01	156.36	156.01	156.95	155.84	156.01
M8.1.0=M22.1.0	154.55	154.55	154.55	155.06	154.55	155.95	154.29	154.55
M10.1.0	122.73	122.73	122.73	123.14	122.73	123.84	122.52	122.73
M11.1.0=M25.1.0=M30.1.0=M68.1.0	149.71	149.71	149.71	150.21	149.71	151.07	149.46	149.71
M12.1.0=M69.1.0=M73.1.1	154.47	154.47	154.47	154.81	154.47	155.41	154.30	154.47
M13.1.0	166.42	166.42	166.42	166.79	166.42	167.43	166.24	166.42
M15.1.0=M20.1.0=M21.1.0=M36.1.0 =M37.1.0	147.52	147.52	147.52	147.84	147.52	148.41	147.35	147.52
M16.1.0=M32.1.1	151.53	151.53	151.53	152.03	151.53	152.91	151.28	151.53
M17.1.0=M33.1.1=M40.1.1=M54.1.2	123.95	123.95	123.95	124.36	123.95	125.07	123.74	123.95
M18.1.0=M34.1.0=M39.1.0=M75.1.0	151.66	151.66	151.66	152.17	151.66	153.04	151.41	151.66
M19.1.0=M41.1.1	94.35	94.35	94.35	94.66	94.35	95.20	94.19	94.34
M23.1.1=M66.1.0	153.88	153.88	153.88	154.39	153.88	155.28	153.63	153.88
M24.1.0	121.56	121.56	121.56	122.10	121.56	123.03	121.29	121.56
M26.1.0=M42.1.1=M49.1.1=M58.1.2	149.49	149.49	149.49	150.16	149.49	151.31	149.16	149.49
M27.1.0=M50.1.1	113.33	113.33	113.33	113.83	113.33	114.70	113.08	113.33
M28.1.0=M44.1.0	151.22	151.22	151.22	151.89	151.22	153.06	150.89	151.22
M29.1.0	85.01	85.01	85.01	85.39	85.01	86.04	84.83	85.01
M31.1.0	119.26	119.26	119.26	119.79	119.26	120.71	119.00	119.26
M35.1.0	77.65	77.65	77.65	78.00	77.65	78.59	77.48	77.65
M38.1.0=M56.1.0	92.06	92.06	92.06	92.47	92.06	93.17	91.85	92.05
M43.1.0=M59.1.1	111.99	111.99	111.99	112.62	111.99	113.70	111.69	111.99
M45.1.0	73.22	73.22	73.22	73.63	73.22	74.33	73.02	73.22
M46.1.1	147.40	147.40	147.40	148.06	147.40	149.19	147.08	147.40

Table 3 *(continue)*

Model number	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M47.1.0	118.40	118.40	118.40	119.06	118.40	120.20	118.08	118.40
M48.1.0=M61.1.0	87.00	87.00	87.00	87.49	87.01	88.33	86.77	87.00
M51.1.0	86.99	86.99	86.99	87.48	86.99	88.31	86.75	86.99
M52.1.0=M62.1.0	77.68	77.68	77.68	78.11	77.68	78.86	77.47	77.68
M53.1.1	149.88	149.88	149.88	150.55	149.88	151.70	149.55	149.88
M55.1.0	77.69	77.69	77.70	78.13	77.70	78.88	77.48	77.69
M57.1.0	70.96	70.96	70.97	71.44	70.97	72.26	70.73	70.96
M60.1.0	72.80	72.80	72.80	73.28	72.80	74.13	72.56	72.79
M63.1.0	71.29	71.29	71.29	71.84	71.29	72.81	71.01	71.28
M65.1.0=M83.1.2	154.53	154.53	154.53	154.87	154.53	155.46	154.36	154.53
M67.1.0=M81.1.1	122.50	122.50	122.50	122.91	122.50	123.61	122.30	122.50
M70.1.0	162.41	162.41	162.41	163.13	162.41	164.38	162.05	162.41
M71.1.0	151.27	151.27	151.27	151.95	151.27	153.11	150.94	151.27
M72.1.0	150.36	150.36	150.36	150.86	150.36	151.72	150.11	150.36
M74.1.0	121.72	121.72	121.72	122.13	121.72	122.83	121.52	121.72
M76.1.0	85.23	85.23	85.23	85.61	85.23	86.26	85.04	85.23
M77.1.0	177.37	177.37	177.37	177.96	177.37	178.98	177.08	177.37
M78.1.0	134.59	134.59	134.60	135.05	134.60	135.82	134.37	134.59
M79.1.0	151.11	151.11	151.11	151.95	151.11	153.41	150.70	151.11
M80.1.1	148.18	148.18	148.18	149.00	148.18	150.43	147.77	148.17
M82.1.0	135.42	135.42	135.42	136.48	135.42	138.31	134.90	135.42
M84.1.0	108.30	108.30	108.30	108.78	108.30	109.61	108.06	108.30
M85.1.1	128.35	128.35	128.35	129.21	128.35	130.69	127.93	128.34
M86.1.0=M128.1.0	76.47	76.47	76.47	76.90	76.47	77.63	76.26	76.47
M87.1.1	112.07	112.07	112.07	112.82	112.08	114.12	111.71	112.07
M88.1.0=M130.1.0	88.49	88.49	88.49	89.08	88.49	90.10	88.20	88.49
M89.1.1	152.41	152.41	152.41	153.09	152.41	154.26	152.08	152.41
M90.1.1	106.50	106.50	106.50	107.09	106.50	108.11	106.20	106.49
M91.1.1	124.02	124.02	124.02	124.85	124.02	126.28	123.61	124.01
M92.1.0	67.85	67.85	67.85	68.30	67.85	69.09	67.63	67.85
M93.1.0	127.73	127.73	127.73	128.58	127.73	130.06	127.31	127.72
M94.1.0	97.75	97.75	97.75	98.52	97.75	99.84	97.38	97.75
M95.1.0	74.94	74.94	74.94	75.28	74.94	75.85	74.78	74.94
M96.1.0	109.06	109.06	109.06	109.79	109.06	111.05	108.70	109.06
M97.1.0=M139.1.0	91.72	91.72	91.72	92.44	91.72	93.68	91.37	91.72
M98.1.0	62.66	62.66	62.66	63.15	62.67	64.00	62.42	62.66
M99.1.0	146.74	146.74	146.74	148.21	146.74	150.78	146.02	146.73
M100.1.0	99.99	99.99	99.99	100.66	99.99	101.82	99.66	99.99
M101.1.1	127.89	127.89	127.89	129.32	127.89	131.80	127.19	127.88
M102.1.1	62.73	62.73	62.73	63.36	62.74	64.46	62.42	62.73
M103.1.1	112.64	112.64	112.64	113.52	112.64	115.04	112.21	112.63
M104.1.0	84.49	84.49	84.49	85.06	84.49	86.04	84.22	84.49
M105.1.0	70.77	70.77	70.77	71.41	70.78	72.50	70.46	70.77
M106.1.0	104.31	104.31	104.31	105.36	104.31	107.18	103.80	104.30
M107.1.1	80.91	80.91	80.91	81.72	80.91	83.13	80.51	80.90
M108.1.1	51.05	51.05	51.05	51.45	51.06	52.14	50.86	51.05
M109.1.0	51.86	51.86	51.87	52.50	51.87	53.61	51.55	51.86
M110.1.1	107.75	107.75	107.75	108.59	107.75	110.05	107.33	107.74
M111.1.0	84.28	84.28	84.28	85.13	84.28	86.60	83.86	84.27
M112.1.0	50.86	50.86	50.86	51.37	50.86	52.26	50.61	50.85
M113.1.0	60.10	60.10	60.10	60.63	60.10	61.56	59.83	60.09

Table 3 *(continue)*

Model number	**AIC**	**FPE**	**GCV**	**HQ**	**RICE**	**SCHWARZ**	**SGMASQ**	**SHIBATA**
M114.1.0	46.50	46.50	46.50	47.23	46.50	48.50	46.14	46.49
M115.1.1	103.16	103.16	103.16	104.31	103.16	106.32	102.59	103.15
M116.1.2= M118.1.3	77.95	77.95	77.96	78.56	77.96	79.62	77.66	77.95
M117.1.2	45.86	45.86	45.86	46.32	45.86	47.12	45.64	45.86
M119.1.3	47.17	47.17	47.18	47.70	47.18	48.62	46.92	47.17
M120.1.1	40.99	40.99	40.99	41.63	40.99	42.76	40.67	40.98
M121.1.1	152.90	152.90	152.91	153.59	152.91	154.76	152.57	152.90
M122.1.1	146.88	146.88	146.88	147.70	146.88	149.11	146.48	146.88
M123.1.0	125.34	125.34	125.34	126.04	125.34	127.24	125.00	125.34
M124.1.0	126.99	126.99	126.99	128.12	126.99	130.09	126.43	126.98
M125.1.0	158.61	158.61	158.61	159.49	158.61	161.02	158.17	158.61
M126.1.1	107.59	107.59	107.59	108.19	107.59	109.23	107.30	107.59
M127.1.0	124.74	124.74	124.74	125.86	124.74	127.79	124.20	124.73
M129.1.0	104.95	104.95	104.95	105.77	104.95	107.19	104.55	104.95
M130.1.1	154.50	154.50	154.50	155.36	154.50	156.84	154.07	154.49
M131.1.0	114.90	114.90	114.90	115.67	114.91	117.00	114.53	114.90
M132.1.0	118.18	118.18	118.18	119.23	118.18	121.07	117.66	118.17
M133.1.0	68.28	68.28	68.28	68.81	68.28	69.74	68.02	68.28
M134.1.0	157.02	157.02	157.03	158.25	157.03	160.38	156.42	157.02
M135.1.0= M137.1.0	94.67	94.67	94.67	95.51	94.67	96.98	94.26	94.66
M136.1.0	66.69	66.69	66.69	67.22	66.70	68.12	66.44	66.69
M138.1.0	108.23	108.23	108.24	109.20	108.24	110.88	107.76	108.23
M140.1.1	61.03	61.03	61.03	61.44	61.03	62.15	60.83	61.03
M142.1.2	135.41	135.41	135.42	136.93	135.42	139.56	134.68	135.41
M143.1.1	94.92	94.92	94.92	95.88	94.92	97.53	94.45	94.92
M144.1.0	127.21	127.21	127.21	128.35	127.21	130.32	126.65	127.20
M145.1.2	59.86	59.86	59.86	60.46	59.86	61.51	59.56	59.85
M146.1.2=M166.1.2	103.30	103.30	103.30	104.22	103.30	105.82	102.85	103.30
M147.1.0=M167.1.0	82.53	82.53	82.54	83.18	82.54	84.30	82.22	82.53
M148.1.2	47.26	47.26	47.27	47.90	47.27	49.01	46.96	47.26
M149.1.3	100.26	100.26	100.26	101.27	100.26	103.02	99.77	100.25
M150.1.0	80.22	80.22	80.22	80.94	80.23	82.18	79.87	80.22
M151.1.1	50.84	50.84	50.84	51.35	50.84	52.23	50.59	50.83
M152.1.2	50.85	50.85	50.85	51.36	50.85	52.25	50.60	50.85
M153.1.2	103.96	103.96	103.97	105.01	103.97	106.83	103.45	103.96
M154.1.1=M174.1.1	83.56	83.56	83.56	84.40	83.56	85.86	83.15	83.55
M155.1.1	50.11	50.11	50.11	50.67	50.11	51.64	49.84	50.11
M156.1.1	56.83	56.83	56.83	57.33	56.83	58.21	56.58	56.82
M157.1.3	42.07	42.07	42.07	42.83	42.08	44.15	41.71	42.07
M158.1.1	97.01	97.01	97.01	98.31	97.02	100.59	96.38	97.00
M159.1.2	67.64	67.64	67.65	69.16	67.66	71.85	66.91	67.63
M161.1.2	47.15	47.15	47.15	47.62	47.15	48.45	46.92	47.15
M162.1.2	45.63	45.63	45.63	46.24	45.63	47.31	45.33	45.62
M163.1.5	46.47	46.47	46.47	47.14	46.47	48.32	46.14	46.46
M164.1.4	40.36	40.36	40.37	40.95	40.37	41.98	40.08	40.36
M165.1.1	138.21	138.21	138.22	139.91	138.22	142.88	137.38	138.20
M168.1.1	99.96	99.96	99.96	100.86	99.97	102.40	99.53	99.96
M169.1.2	124.32	124.32	124.32	125.57	124.33	127.74	123.71	124.32
M170.1.2	59.71	59.71	59.71	60.25	59.72	61.17	59.45	59.71
M171.1.3	46.73	46.73	46.73	47.36	46.73	48.45	46.42	46.73
M172.1.1	102.12	102.12	102.12	103.26	102.13	105.25	101.56	102.11

Table 3 *(continue)*

Model number	AIC	FPE	GCV	HQ	RICE	SCHWARZ	SGMASQ	SHIBATA
M173.1.1	74.11	74.11	74.11	75.11	74.12	76.84	73.63	74.10
M175.1.1	51.83	51.83	51.83	52.29	51.83	53.09	51.60	51.82
M176.1.1	50.43	50.43	50.43	51.16	50.44	52.45	50.07	50.42
M177.1.1	100.69	100.69	100.70	101.59	100.70	103.15	100.25	100.69
M178.1.0	55.97	55.97	55.97	56.66	55.97	57.86	55.63	55.96
M179.1.3	44.97	44.97	44.97	45.88	44.98	47.48	44.53	44.96
M180.1.1	61.03	61.03	61.03	62.06	61.04	63.85	60.53	61.02
M181.1.6	66.31	66.31	66.32	67.35	66.32	69.17	65.81	66.30
M182.1.4	42.40	42.40	42.40	43.16	42.40	44.49	42.03	42.39
M183.1.7	41.29	41.29	41.29	42.21	41.30	43.85	40.84	41.28
M184.1.7	43.37	43.37	43.37	44.15	43.38	45.51	42.99	43.36
M185.1.3	39.88	39.88	39.89	41.18	39.90	43.52	39.25	39.86
M186.1.1	47.85	47.85	47.86	48.77	47.86	50.37	47.41	47.85
M187.1.2	97.87	97.87	97.87	98.96	97.87	100.86	97.33	97.86
M188.1.3	66.66	66.66	66.67	67.94	66.67	70.17	66.05	66.65
M189.1.4	42.73	42.73	42.74	43.50	42.74	44.85	42.36	42.73
M190.1.7	41.21	41.21	41.22	42.42	41.23	44.58	40.63	41.20
M191.1.3	44.34	44.34	44.35	45.19	44.35	46.68	43.93	44.33
M192.1.14	36.61	36.61	36.62	37.60	36.62	39.36	36.13	36.60

The minimum value for M192.1.14 represented the efficient model obtained in phase 3. For LASSO, package *glmnet* was used and for bisquare M estimator, library *MASS* was used for the purpose of analysis in R software. The coefficients were obtained by means of the R software and can be observed as in (6).

$$\begin{aligned}
&\text{M192.1.13} = \hat{Y} = 2.350e^{+02}\text{-}3.1975e^{+00}x_2\text{ -}9.921e^{+00}x_3\text{ -}1.005e^{+01}x_5\\
&+ 1.572e^{-01}x_{13}+2.210e^{-01}x_{15}+2.453e^{-03}x_{26}+3.991e^{-01}x_{35}+2.738e^{-02}x_{45}\\
&\text{-}2.477e^{-02}x_{46}+1.224e^{-02}x_{56}+1.589e^{-03}x_{123}+1.150e^{-03}x_{124}\text{-}1.128e^{-02}x_{135}\\
&+8.000e^{-04}x_{245}+1.185e^{-04}x_{256}\text{-}3.980e^{-04}x_{346}+2.162e^{-04}x_{356}+8.468e^{-05}x_{1235}\\
&+2.203e^{-05}x_{1246}\text{-}1.642e^{-05}x_{1256}\text{-}3.053e^{-06}x_{12345}\text{- } 6.217e^{-08}x_{12356}\\
&+1.298e^{-07}x_{13456} \qquad\qquad\qquad\qquad\qquad (6)
\end{aligned}$$

Crucial variables with their respective coefficients can be observed from the above model. From (6) onwards, the importance of interaction terms can be observed in the form of significant variables. MAPE is found with (5) and was obtained as 8.97 for this efficient model using the proposed hybrid model.

Comparison with other existing sparse and robust regression techniques was carried out to verify the efficiency of the proposed technique. Using all other existing techniques, the final model was obtained in a similar way based on four phases in this study. The results of mean squared error (MSE) and for MAPE were observed using the *R* software with the numbers of variables left in a model (*k*). The proposed technique was compared with a variety of other existing sparse and robust techniques. Least trimmed square (LTS), modified M estimator (MM), estimators of scale (S estimator) were applied using R software. Elastic net LTS (E. Net LTS), elastic net with S estimator (E. Net S) using the *pense* package in *R* is applied. M step after elastic net S estimator (M ENet S) using the *Mpense* package in *R* software was performed by default. The purpose was to compare the final selected model among all possible models obtained from all techniques in phase 3.

Table 4
Comparison of proposed method with other existing methods

Selected Model	Technique	(*k*)	MSE	MAPE
M192.1.13	**LASSO with bisquare M**	**23**	**35.46**	**8.968**
M192.1.0	LTS	63	100766.703	72.58
M192.10.0	MM	53	37.88	9.313
M192.1.0	S	63	3627.52	70.76
M192.2.0	E.net LTS	59	44.47	10.06
M192.2.0	E.net S	57	42.84	8.994
M192.3.0	M E.net S	27	87.53	14.373

Table 4 shows the results of the different techniques used in this study. The number of variables can be observed in all techniques with their respective MSE and MAPE in Table 4. M192.2.0 in E.net LTS shows that after performing the method in two steps, all the variables remain in the model as significant. Similarly other models are presented in this way.LTS and S estimator have the highest mean square error. Because of the trimmed observations in LTS, LTS cannot be considered as a good method in forecasting (Alma, 2011). The detailed behavior for the observation pattern for LTS and S estimator is analyzed in Figure 2 and Figure 4. Clearly, the proposed technique is better than all other techniques. The minimum value of MSE was found to be for LASSO with bisquare M in comparison with other techniques. So, on the basis of minimum MSE value, LASSO with bisquare M is preffered than other existing methods for forecasting. It has significant number of variables with minimum MSE and MAPE value as compared to other existing techniques. Although the number of variables in other techniques are higher than the proposed technique but MSE and MAPE is also high in comparison. So, the proposed technique is the best selection for forecasting the model as compared to others.

For the purpose of observing outliers outside the sigma limits, standardized residuals plots are observed for each final model.

Outliers outside 2 sigma limit can be observed from Figure 1-7. The percentage of outliers is obtained based on number of observations outside the 2 sigma limit. The percentage of outliers outside 2 sigma limit in each technique is observed in Table 5.

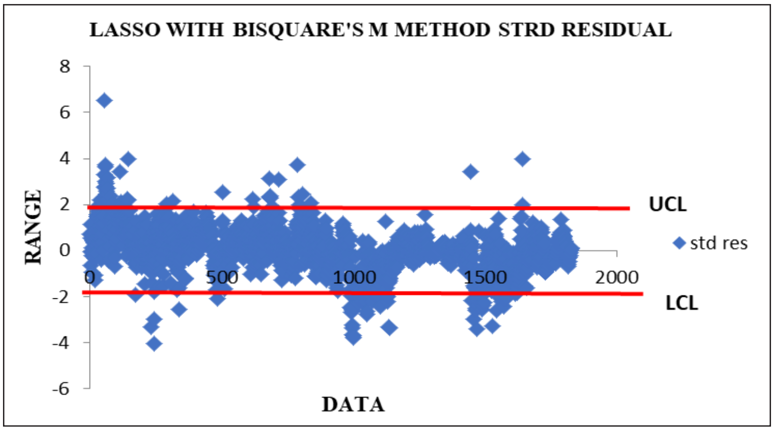

Figure 1. Standardized residual by for LASSO with bisquare M

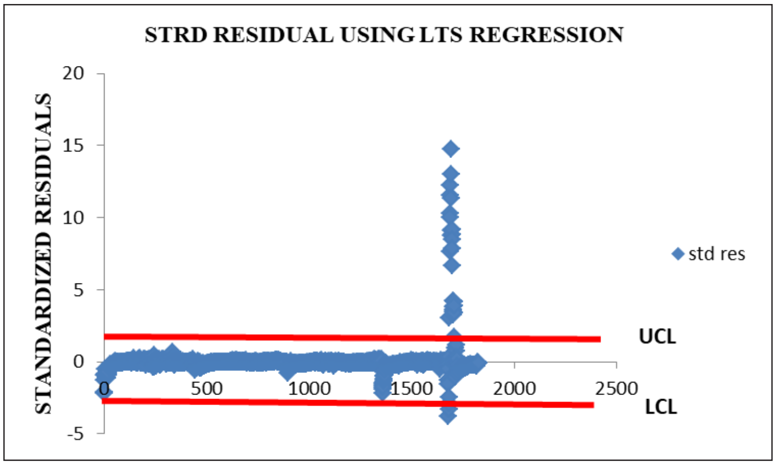

Figure 2. Standardized residual for LTS regression

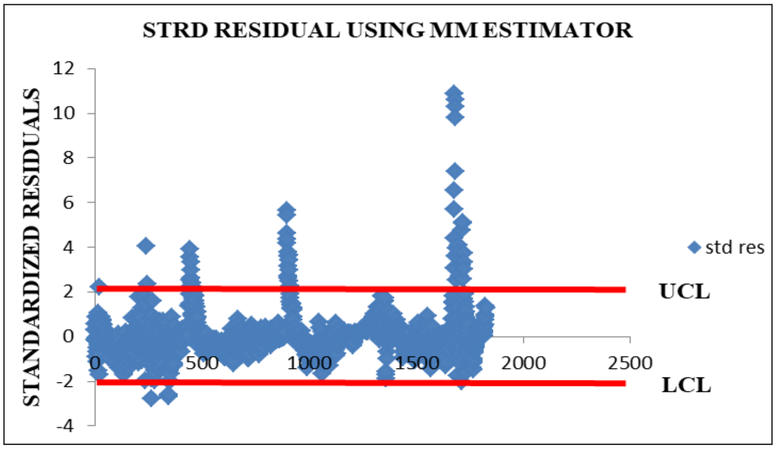

Figure 3. Standardized residual for MM estimator

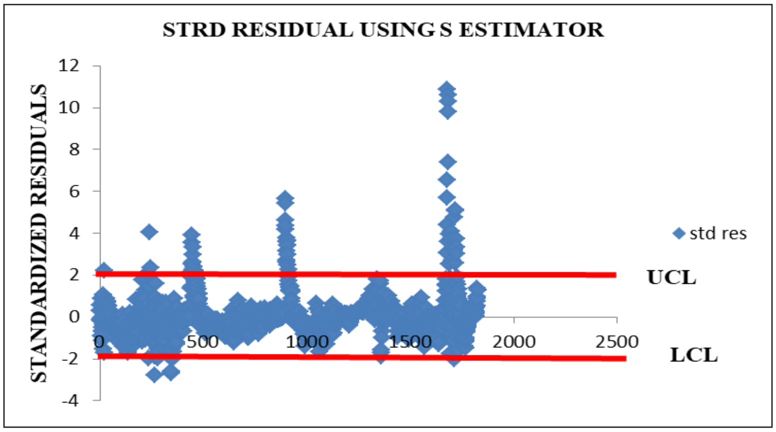

Figure 4. Standardized residual for S estimator

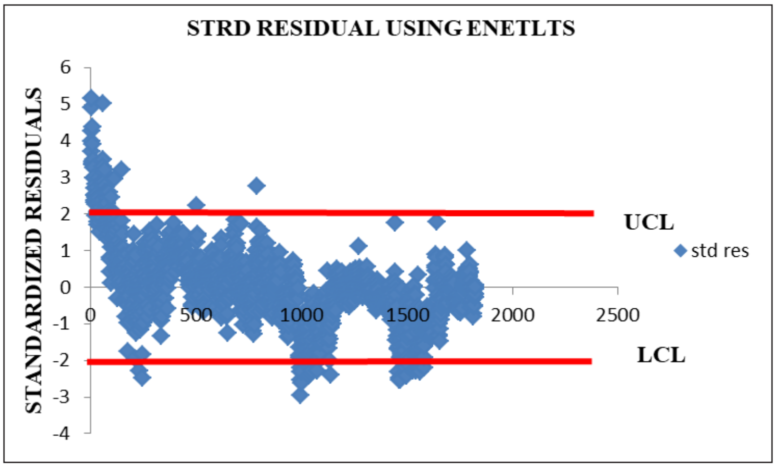

Figure 5. Standardized residual for E.net LTS estimator

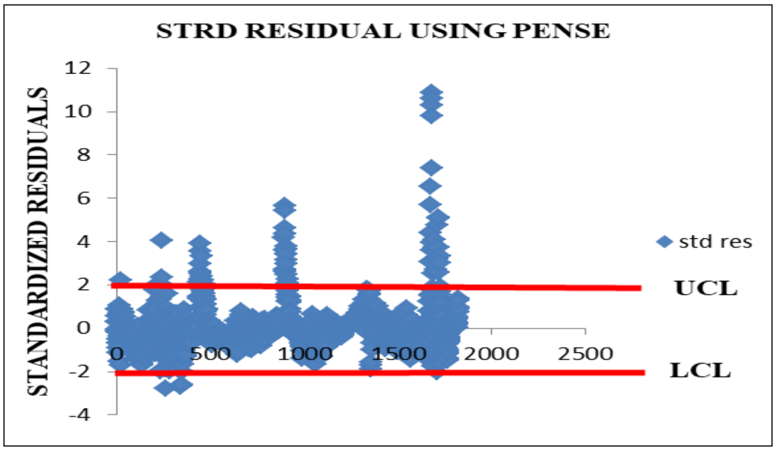

Figure 6. Standardized residual for E.net S estimator

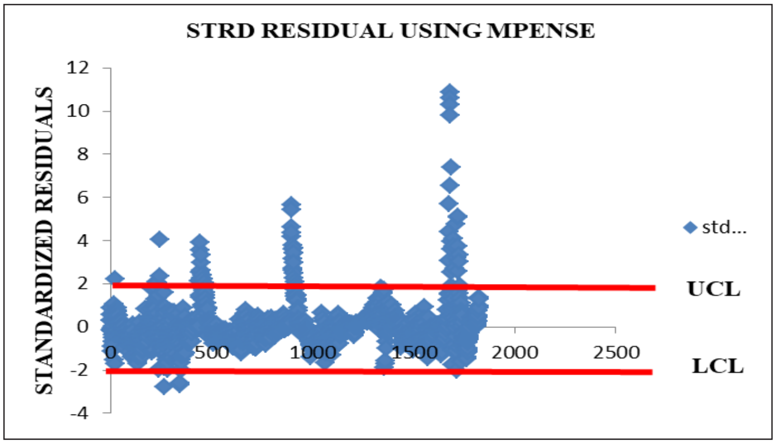

Figure 7. Standardized residual for M.Enet S estimator

Table 5
Percentage of outliers outside 2 sigma limits

Selected Model	Method	$\mu \pm 2\sigma$
M192.1.13	**LASSO with bisquare M**	**5.48%**
M192.1.0	LTS	1.70%
M192.10.0	MM	6.08%
M192.1.0	S	3.51%
M192.2.0	E.net LTS	5.92%
M192.2.0	E. net S	6.24%
M192.3.0	M E.net S	6.24%

There are 5.48% observations as outliers in the proposed hybrid model. The outlier's percentage in LTS and S estimator is lower than the proposed method. But due to high MSE value, LTS and S estimators cannot be considered as suitable for forecasting. In this study, real dataset is used so exclusion of outlier observations is not a good option. The pattern in the proposed technique is random, while more outliers are in a positive direction in all other techniques. There are fewer outliers in the LTS regression, but the MAPE for LTS is 72.58 using (5) with a very high MSE value. The *pense* and *Mpense* packages show almost the same behaviour in E.net S and M E.net S estimators respectively. E.net LTS also shows random pattern, but taking the more observations as outliers than the proposed hybrid model. The MSE and MAPE are also high as compared to the proposed hybrid model.

The estimates are not influenced from outliers in the proposed hybrid model due to the use of robust estimator. On the basis of minimum MSE and MAPE, the proposed hybrid model is preferred than other existing methods. The model obtained from LASSO with bisquare is therefore ready to forecast the moisture ratio removal (%) of seaweed with higher efficiency.

CONCLUSION

The results show that LASSO with bisquare M model provides the best model as compared to other existing methods used in the analysis.The selection of efficient model need to deal with all possible models with the interaction terms. The significance of interaction terms highlight the importance of interactions in the real life dataset. The proposed hybrid model is found to be better in term of MSE and MAPE value in comparison to other existing methods. The pattern of observations is also found to be random in graph of standardized residuals. So, the proposed hybrid model of LASSO and bisquare M can therefore be used for the efficient selection of the model including the interaction terms in it. The model is prepared to predict the moisture ratio removal (%) of seaweed with its crucial factors involving interaction terms. For the future work, the developed procedure based on four phases can also be used in efficient model selection for any other field of study.

ACKNOWLEDGEMENT

The author would like to thank the Women University Multan Pakistan and University Sains Malaysia for the support in this research project

REFERENCES

Ali, M. K., Ruslan, M. H., Muthuvalu, M. S., Wong, J., Sulaiman, J., & Yasir, S. M. (2014). Mathematical modelling for the drying method and smoothing drying rate using cubic spline for seaweed Kappaphycus Striatum variety Durian in a solar dryer. *AIP Conference Proceedings 1602*(1),113-120.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics, 21*(1),243-247.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Ali M. K. M., Fudholi, A., Muthuvalu, M., S., Sulaiman, J., & Yasir, S. M. (2017). Implications of drying temperature and humidity on the drying kinetics of seaweed. *AIP Conference Proceedings, 1905*(1), 1-7.

Alma, O. G. (2011). Comparison of Robust Regression Methods in Linear Regression. *International Journal of Contemporary Mathematical Sciences, 6*(9), 409-421.

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.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). New York, USA: John Wiley & Sons.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215-223.

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.

Giacalone, M., Panarello, D., & Mattera, R. (2018). Multicollinearity in regression: an efficiency comparison between L p-norm and least squares estimators. *Quality and Quantity, 52*(4), 1831-1859.

Gondchawar, N., & Kawitkar, P. R. S. (2016). IoT based smart agriculture. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(6), 838-842.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological), 41*(2), 190-195.

Klaina, H., Vazquez Alejos, A., Aghzout, O., & Falcone, F. (2018). Narrowband characterization of near-ground radio channel for wireless sensors networks at 5G-IoT bands. *Sensors*, *18*(8), 2428-2442.

Mendelsohn, R., & Dinar, A. (2003). Climate, water, and agriculture. *Land Economics*, *79*(3), 328-341.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., & Williams, J. R. (2011). *Soil and water assessment tool theoretical documentation version 2009*. Texas, USA: Texas Water Resources Institute.

Ramanathan, R. (2002). *Introductory Econometrics with application* (5th Ed.): South Western, USA: Harcourt College Publishers.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics, 12*(4), 1215-1230.

Rockström, J., Falkenmark, M., Karlberg, L., Hoff, H., Rost, S., & Gerten, D. (2009). Future water availability for global food production: the potential of green water for increasing resilience to global change. *Water Resources Research*, *45*(7), 1-16.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

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, 890*(1), 1-7.

Shibata, R. (1981). An optimal selection of regression variables. *Biometrika*, *68*(1), 45-54.

Sinova, B., & Van Aelst, S. (2018). Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function. *International Journal of Approximate Reasoning*, *93*, 219-237

Stuart, C. (2011). *Robust regression*. Durham, England: Durham University.

Susanti, Y., Pratiwi, H., Sulistijowati H., S., & Liana, T. (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Apllied Mathematics*, *91*(3), 349-360.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267-288.

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.

Yan-E, D. (2011, March 28-29). Design of Intelligent Agriculture Management Information System based on IoT. In *4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011* (Vol. 1, pp. 1045-1049). Shenzhen, Guangdong, China.

Zainodin, H. J., Noraini, A., & Yap, S. J. (2011). An alternative multicollinearity approach in solving multiple regression problem. *Trends in Applied Sciences Research*, *6*(11), 1241-1255.

Zhang, K., Zhe, S., Cheng, C., Wei, Z., Chen, Z., Chen, H., … & Ye, J. (2016, August 13-17). Annealed sparsity via adaptive and dynamic shrinking. In *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* - *KDD '16* (pp. 1325-1334). San Francisco, California, USA.

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* (pp. 11-33). New York, NY: Springer.