Vol. 20, No. 2, pp. 213-223 (2020)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in

Research Article

Multistage Mustard Yield Estimation Based on Weather Variables using Multiple Linear, LASSO and Elastic Net Models for Semi Arid Region of India

ANANTA VASHISTH* AND ARAVIND K.S.

Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi-110 012

ABSTRACT

Multistage mustard yield estimation was done at vegetative, flowering and grain filling stage of the crop. Daily weather data during crop growing period as well as mustard yield data for the period of 1984-2019 for IARI, New Delhi were used for developing model by stepwise multiple linear regression (SMLR), least absolute shrinkage and selection operator (LASSO) and elastic net techniques. Analysis was carried out by fixing 70% of the data for calibration and remaining dataset for validation. Mustard crop estimation at vegetative, flowering and grain filling stage were done for *Rabi* 2018-19 and 2019-20. On examining these models the value of RMSE and nRMSE was found to be lowest for Elastic Net followed by LASSO and SMLR model. Percentage deviation of estimated yield by observed yield was ranged between 4.53 to 18.93%, 1.86 to 13.87% and 0.75 to 14.89% during vegetative, flowering, and grain filling stage respectively. On the basis of percentage deviation and model accuracy Elastic net model was found best followed by LASSO and SMLR.

Key words: Weather variables, Stepwise multiple linear regression, Least absolute shrinkage and selection operator, Elastic net, Yield estimation

Introduction

Mustard is the most important oilseed crop grown in *Rabi* season in north-west part of India, semiarid to sub-tropic regions having well defined wet and dry winter season. Weather parameters like maximum temperature, minimum temperature, relative humidity, rainfall, etc. have a great impact on crop yield .The pre-harvest forecast of crop yield is likely to provide valuable information in regard to storage, import, export, industries and government for advanced planning. Many techniques have been developed to forecast crop production. In traditional methods, crop cutting experiments were widely used for crop yield forecast at different regions. The relationship between weather variables and yield of the crop can be estimated though different statistical methods. For achieving effective crop yield forecast based on weather variables, models are required to be calibrated and validated with the historical data. Dutta *et al*. (2001) reported good accuracy pre-harvest district wise rice yield prediction for Bihar by utilizing weather data. Agrawal *et al*. (2012) have developed forecast models for wheat yield in Kanpur district using discriminant function analysis of weakly data on weather variables. Garde *et al*. (2015) used multiple linear regression technique and discriminant function analysis for estimating wheat productivity for the district of Varanasi in eastern Uttar Pradesh. He reported that stepwise

*Corresponding author,
Email: ananta.iari@gmail.com

multiple linear techniques can be used successfully for pre-harvest wheat crop yield forecast, which are more consistent in performance on zone or state level. Different weather variables were used for generating weighted and un-weighted weather indices and these indices were used for developing multiple linear regression yield forecast model (Agrawal and Mehta, 2007; Chauhan *et al*., 2009). Das *et al*., (2018) reported that LASSO, Elastic Net and SMLR developed by using long term weather variables found to be the best models on the basis of coefficient of determination, root mean square error and normalized root mean square error for rice crop estimation in west coast of India. Tibshirani (1996) formulated a powerful variable selection LASSO (Least Absolute Shrinkage and Selection Operator) method that performs two main tasks viz. regularization and feature selection in order to enhance the prediction accuracy and interpretability of the statistical model. Regularization techniques are used to prevent statistical over fitting in a predictive model. The aim of this study is to do multistage mustard yield estimation by SMLR, LASSO and Elastic Net models and evaluate the performance of these models in order to enhance accuracy of crop yield forecast.

Materials and Methods

Mustard yield data as well as daily weather data such as maximum and minimum temperature, morning and evening relative humidity, rainfall from 1984 to 2019 were collected from IARI, New Delhi during mustard crop growing period. Maximum temperature, minimum temperature, rainfall, morning and evening relative humidity were arranged for three different stages viz. vegetative (40th to 52nd SMW), flowering (40th to 4th SMW) and grain filling (40th to 8th SMW) stage separately. Daily weather data was converted into simple and weighted composite weather indices. Summation of individual weather variable or interaction of two weather variable at a time were used for generating simple weather indices, sum product of individual weather variable or interaction of weather variables and its correlation with adjusted crop yield were resulted with weighted weather indices. Computation of simple and weighted weather indices were based on following formula.

Simple weather indices

$$Zij = \sum_{w=1}^{m} Xiw \qquad \ldots(1)$$

$$Zii'j = \sum_{w=1}^{m} Xiw\, Xi'w \qquad \ldots(2)$$

Weighted weather indices

$$Zij = \sum_{w=1}^{m} r^{\,j}iw\, Xiw \qquad \ldots(3)$$

$$Zii'j = \sum_{w=1}^{m} r^{\,j}ii'wXiw\, Xi'w \qquad \ldots(4)$$

Where,

Xiw / $Xii'w$ = value of i^{th}/i'^{th} weather variable in wth week.

$r^{j}iw$ / $r^{j}ii'w$ = correlation coefficient of yield with i^{th} weather variable or product of i^{th} or i'^{th} weather variable in wth week.

m = week at which forecast done.

P = number of variables

Combination of weather variables for weather indices, generated are presented in Table 1. 70% of data were used for model calibration and remaining 30% were used for the validation of models. Model for estimating the mustard yield at vegetative, flowering and grain filling stage was developed by SMLR, LASSO and Elastic Net techniques.

Stepwise multiple linear regression (SMLR)

Weather indices developed by maximum and minimum temperature, rainfall, morning and evening relative humidity were used for developing model. Impact of important weather indices were determined by adopting Stepwise regression technique. Using different weather variables, appropriate weighted and un-weighted weather indices are generated and multiple linear regression forecasting models was developed (Kumar *et al*., 1999). SMLR used for pre-harvest wheat crop yield estimation because of its more

Table 1. Simple and weighted weather indices used for developing model

	Simple weather indices					Weighted weather indices				
	Tmax	Tmin	Rainfall	RH I	RH II	Tmax	Tmin	Rainfall	RH I	RH II
Tmax	Z10					Z11				
Tmin	Z120	Z20				Z121	Z21			
Rainfall	Z130	Z230	Z30			Z131	Z231	Z31		
RH I	Z140	Z240	Z340	Z40		Z141	Z241	Z341	Z41	
RH II	Z150	Z250	Z350	Z450	Z50	Z151	Z251	Z351	Z451	Z51

consistent performance and applicability at zone or state level (Garde *et al.,* 2015). Feature selection helps to attain selection of best regression variables and thereby good interpretable results among independent variables (Singh *et al*., 2014).

Least absolute shrinkage and selection operator (LASSO)

Least Absolute Shrinkage and Selection Operator (LASSO) is a model selection technique proposed by Tibshirani (1996). LASSO models are used to overcome the shortcomings of ordinary least square (OLS) and ridge regression. Though residual mean square error can minimize by OLS, it has low biasness and large variance, reduces the prediction accuracy. With large number of predictors, smaller subset selection exhibit stronger effect on interpretation of data. Subset selection is discrete and variable process, regressors are either retained or eliminated from the model in order to provide better interpretable model. LASSO estimators are used for consistent regression coefficient and automatic variable selection. Continuous shrinkage of some coefficients by imposing L1 penalty and others to zero, hence it helps to retain some good features of both subset selection and ridge regression.

Suppose we have data (x_i, y_i), i=1,2,……, N, where $x_i = (x_{il},\ldots\ldots\ldots\ldots, x_i p)^T$ are the predictor variables and y_i are the yield responses. In the usual regression, assuming that the observations are independent or y_i's are conditionally independent on given x_i's. We assume

$$(\hat{\alpha}, \hat{\beta}) = \arg\min\left\{\sum_{i=1}^{N}\left(y_i - \alpha - \sum_j \beta_j x_{ij}\right)^2\right\}$$

Subject to

$$\sum_j \left|\beta_j\right| \le t \qquad \ldots(5)$$

Here t ≥ 0 is a tuning parameter, which controls the amount of shrinkage is applied to the estimates. Let $\beta_j|$, the full least square estimates and $t_0 = \Sigma\beta_j|$,. Values of t< t will cause shrinkage of solution towards 0, and some coefficients may be exactly equal to 0. Lasso gives sparse interpretable model with excellent prediction accuracy. An alternate formulation of lasso to solve penalised likelihood problem is,

$$\min_{\beta} \frac{1}{n}\left(y - X\beta\right)^T \left(y - X\beta\right) + \lambda \sum_{j=1}^{d} |\beta_j| \qquad \ldots(6)$$

Both the formulas are equivalent in sense, for any given $\lambda \in (0,\infty)$, there exists t ≥ 0 such that the two problem have same solution and vice-versa.

Elastic net

Elastic net penalises the size of regression coefficients based on both L1 norm and L2 norm penalty. L1 norm used to generate sparse model, L2 penalty removes the limitation on the number of selected variables, encourage grouping effect, stabilises the L1 regularization path.

Suppose the data set has n number of observation with p number of variables or predictors, the yield or response can be expressed as $y=(y_1,\ldots\ldots,y_n)^T$, j= 1,……., p are the predictors

$$\sum_{i=1}^{n} y_i = 0, \ \sum_{i=1}^{n} x_{ij} = 0 \ \text{ and } \ \sum_{i=1}^{n} x_{ij}^2 = 1 \qquad \ldots(7)$$

For any fixed non negative λ_1 and λ_2, elastic net is,

$$L(\lambda_1, \lambda_2, \beta) = | y - X\beta |^2 + \lambda_2 | \beta |^2 + \lambda_1 | \beta |_1 \quad \ldots(8)$$

$$| \beta |^2 = \sum_{j=1}^{p} \beta_j^2$$

$$| \beta |_1 = \sum_{j=1}^{p} | \beta_j |$$

$$\hat{\beta} = \arg\min_{\beta} \{L(\lambda_1, \lambda_2, \beta)\} \quad \ldots(9)$$

Where, β is the elastic net estimator, minimiser of equation.

This equation is explained as penalised least square method. Let $\alpha = \lambda_2 / (\lambda_1 + \lambda_2)$, on solving the equation

$$\hat{\beta} = \arg\min_{\beta} | y - X\beta |^2 \quad \ldots(10)$$

$$(1-\alpha) | \beta |_1 + \alpha | \beta |^2 \leq t \quad \ldots(11)$$

Subject to, for some value of *t*.

We call the function $(1- \alpha)\ \beta_1 + \alpha\beta_2$ the elastic net penalty, which is convex combination of lasso and ridge penalty.

R (version 3.6.0) software was used for model development by LASSO and Elastic Net and SPSS (Version 16.0) was used for model development by SMLR. Lasso and elastic net have two parameters, lambda and alpha, these values need to be optimized and selected by minimizing average mean square error in cross validation. The tuning parameter t value was set as 1 for lasso and 0.5 for elastic net. The size of this penalty, referred to as L2 (or Euclidean) norm, can take on a wide range of values, which is controlled by the tuning parameter λ. When λ=0 there is no effect and our objective function equals the normal OLS regression objective function of simply minimizing SSE. The LASSO penalty is an alternative to the ridge penalty that follows L1 norm. Switching to the LASSO penalty not only improves the model but it also conducts automated feature selection. When a data set has many features, lasso can be used to identify and extract those features with the largest (and most consistent) variables by keeping λ=1 (smaller λ values will retain more and least important variables within the model). Elastic net is combination of both penalties i.e., ridge and lasso penalties, it enables effective regularization via the ridge penalty with the feature selection characteristics of the LASSO penalty. Any λ value between 0 to 1 will perform elastic net. When λ=0.5 it performs an equal combination of penalties whereas λ= <0.5 will have a heavier ridge penalty applied and λ= >0.5 will have a heavier LASSO penalty. The R package implementing regularized linear models is glmnet. For mustard yield estimation 'glmnet' package in R (version 3.6.0) was used for developing model by LASSO and Elastic Net.

Mustard crop yield estimation at vegetative, flowering and grain filling stage was done for Rabi 2018-19 and 2019-20. Percentage deviation of yield estimation by SMLR, LASSO and Elastic Net model done at different growth stage by observed yield was calculated.

Statistical test

Model performance during calibration and validation was observed on the basis of root mean square error (RMSE), normalized mean square error (nRMSE) and percent Deviation.

Root mean square error (RMSE)

This is often used to measure the difference between estimated values from the model and actual observed values from the experiment that is being modeled. By this test, model performance during the calibration as well as validation period can be determined. It is also helpful in comparing individual model performance with that of other predictive models.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Pi - Oi)^2} \quad \ldots(12)$$

Where RMSE is absolute root mean square error, Pi is the predicted value, O_i is the observed value and N is the number of observations

Normalized mean square error (nRMSE)

Normalized mean square error expressed in percentage, values close to zero indicates better model performance. nRMSE is a measure (%) of the relative difference of estimated versus observed data. The prediction is considered excellent with the nRMSE <10%, good if 10–20%, fair if 20–30%, poor if >30%.

If Pi, O_i, N and M are notated as predicted value, observed value, number of observations and mean of observed value, nRMSE can be written as the formula given below.

$$nRMSE = \frac{100}{M} * \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Pi - Oi)^2} \qquad \ldots(13)$$

Percent Deviation

Percent Deviation is calculated using following formula:

$$\%Deviation = \frac{Pi - Oi}{Oi} * 100 \qquad \ldots(14)$$

Pi is the predicted value and O_i is the observed value

Where Pi is the predicted value, O_i is the observed value, N is the number of observations.

Results and Discussion

Mustard Yield estimation at vegetative stage by SMLR, LASSO and Elastic net model

Model for estimating mustard yield at vegetative stage by different techniques for IARI, New Delhi have been developed using long term crop yield data as well as long period daily weather data from 40th to 52th standard meteorological week for IARI, New Delhi. Model performance developed by SMLR, LASSO and Elastic Net techniques for mustard yield estimation during calibration at vegetative stage are presented in Fig 1. The coefficient of determination (R^2) was significant at 1% probability level for all developed models. Value of coefficient of determination R^2 for models developed by different techniques for estimating the mustard crop yield at vegetative stage was 0.85% for model developed by SMLR techniques, 0.89% for modal developed by LASSO techniques and 0.90% for modal developed by Elastic net techniques. The RMSE value during calibration was lowest for Elastic Net modal (82.2 kg/ha) followed by LASSO (145.2 kg/ha) and SMLR (169.2 kg/ha). During calibration nRMSE value was < 10% for all three models having lowest value 4.47% for Elastic Net followed by 7.89% for LASSO and 9.2% for SMLR model. Model performance developed by SMLR, LASSO and Elastic Net techniques for mustard yield estimation during validation at vegetative stage are presented in Fig 2. The RMSE value during validation was lowest for Elastic Net modal (274.3 kg/ha) followed by LASSO (325.2 kg/ha) and SMLR (334.2 kg/ha). During validation nRMSE value was lowest 10.6% for Elastic Net followed by 12.51% for LASSO and 12.86% for SMLR model.

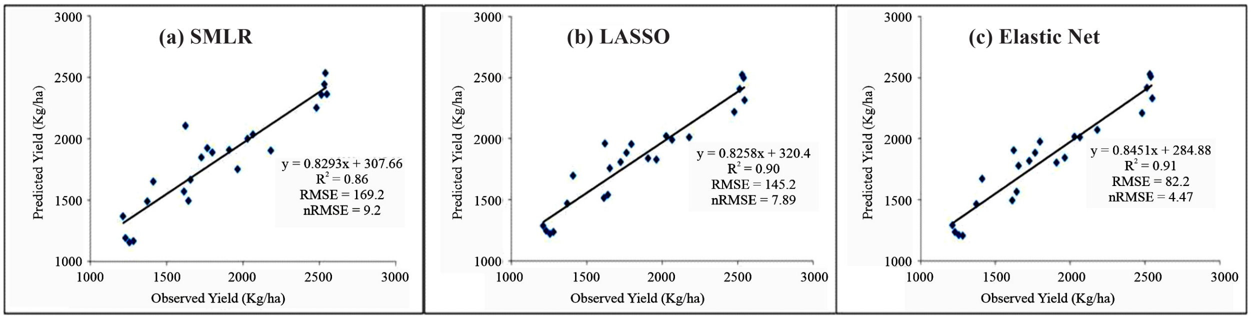

Fig. 1. Performance during calibration of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at vegetative stage

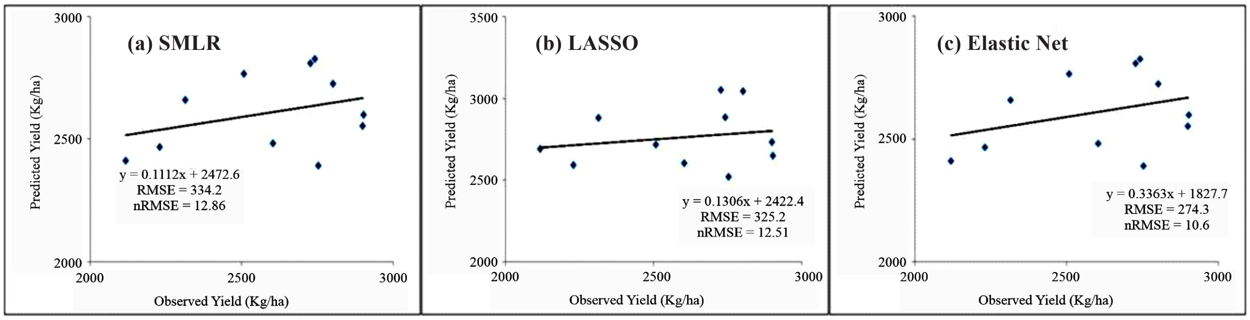

Fig. 2. Performance during validation of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at vegetative stage

The most important weather parameter identified by SMLR for mustard yield estimation at vegetative stage are time and Z21 (weighted minimum temperature). For LASSO most important weather parameter are time, Z21 (weighted minimum temperature), Z120 (maximum temperature*minimum temperature), Z121 (weighted maximum temperature*minimum temperature), Z141 (weighted maximum temperature*morning relative humidity) and Z151 (weighted maximum temperature*evening relative humidity). For Elastic Net most important weather parameters are time, Z11 (weighted maximum temperature), Z21 (weighted minimum temperature) and Z241 (weighted minimum temperature*morning relative humidity). Equations developed by SMLR, LASSO and Elastic Net model for mustard crop yield estimation at vegetative stage are given in table 2.

Percentage deviation of estimated yield for mustard crop done at vegetative stage by observed yield for IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in table 2. During *Rabi* 2018-19 the percentage deviation was lowest 9.49% for Elastic Net model followed by 11.99% for LASSO and 13.33% for SMLR modal respectively. During Rabi 2019-20 percentage deviation of estimated yield by observed yield was 4.53% for Elastic net, 12.99% for LASSO and 18.93% for SMLR model respectively. Lowest value of percentage deviation was observed by Elastic Net followed by LASSO and SMLR in both the year. Dutta *et al*. (2001) had developed district wise yield model for rice in Bihar using meteorological data and concluded that models were able to predict pre-harvest crop yield with good accuracy. Tibshirani (1996) proposed the method LASSO for shrinkage and selection for regression and generalized regression

Table 2. Mustard Yield estimation at vegetative stage by SMLR, LASSO and Elastic Net model during *Rabi* 2018-19 and 2019-20

Model	Model equation	Estimated yield (kg/ha)		Observed yield (kg/ha)		% Deviation	
		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
SMLR	y=1854.46+61.14*time+ Z21*91.35	3088.8	3266.5	2725.6	2746.6	13.33	18.93
LASSO	y=2907.71+59.93*time+ 53.78*Z21-0.16* Z120 + 0.07* Z121 + 0.04*Z141+ 0.17*Z151	3052.4	3103.6	2725.6	2746.6	11.99	12.99
Elastic Net	y=1709.85+43.87*time+ 3.49*Z11+26.21*Z21+ 0.05*Z241	2984.4	2871.1	2725.6	2746.6	9.49	4.53

problems. He reported that LASSO does not focus on subsets but rather it defines a continuous shrinking operation that can produce coefficient that is exactly to zero. Kumar *et al.* (2019) evaluate the performance of stepwise and Lasso regression technique in variable selection and development of wheat forecast model for crop yield using weather data and wheat yield for the period of 1984-2015 for IARI, New Delhi. He reported that performance of Lasso regression is better than stepwise regression.

Mustard yield estimation at flowering stage by SMLR, LASSO and Elastic net model

For developing model for estimating mustard yield at flowering stage by different techniques for IARI, New Delhi, long period daily weather data from 40th to 4th standard meteorological week as well as long term crop yield data for IARI, New Delhi were used. Model performance developed by SMLR, LASSO and Elastic Net techniques for mustard yield estimation during calibration at flowering stage are shown in Fig 3. Value of coefficient of determination R^2 for models developed by different techniques for estimating the mustard crop yield at flowering stage was 0.87% for model developed by SMLR and LASSO techniques and 0.88% for modal developed by Elastic net techniques. The RMSE value during calibration was lowest for Elastic Net modal (159.7 kg/ha) followed by SMLR (164.6 kg/ha) and LASSO (172.7 kg/ha). During calibration nRMSE value was < 10% for all three models having lowest value 8.68% for Elastic Net followed by 8.95% for SMLR and 9.39% for LASSO model. Model performance developed by SMLR, LASSO and Elastic Net techniques for mustard yield estimation during validation at flowering stage are presented in Fig 4. The RMSE value during validation was lowest for Elastic Net modal (253.3 kg/ha) followed by LASSO (294.5 kg/ha) and SMLR (311.6 kg/ha). During validation nRMSE value was lowest 9.74% for Elastic Net followed by 11.33% for LASSO and 11.99% for SMLR model.

The most important weather parameters identified by SMLR for mustard yield estimation at flowering stage are time and Z21 (weighted minimum temperature). For LASSO most important weather parameter are time, Z21 (weighted minimum temperature) and Z151 (weighted maximum temperature*evening relative humidity). For Elastic Net most important weather parameters are time, Z11 (weighted maximum temperature) and Z21 (weighted minimum temperature). Equation developed by SMLR, LASSO and Elastic Net model for mustard crop yield estimation at flowering stage are given in table 3.

Percentage deviation of estimated yield for mustard crop done at flowering stage by observed yield for IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in table 3. During *Rabi* 2018-19 the percentage deviation was lowest 5.73% for Elastic Net model followed by 8.76% for LASSO and 12.33% for SMLR modal respectively. During *Rabi* 2019-20 percentage deviation of estimated yield by observed yield was 1.86% for Elastic net, 9.29% for LASSO and 13.87% for SMLR model respectively. Lowest value of percentage deviation was observed by Elastic Net followed by LASSO and SMLR in both the year. Percentage deviation has lower

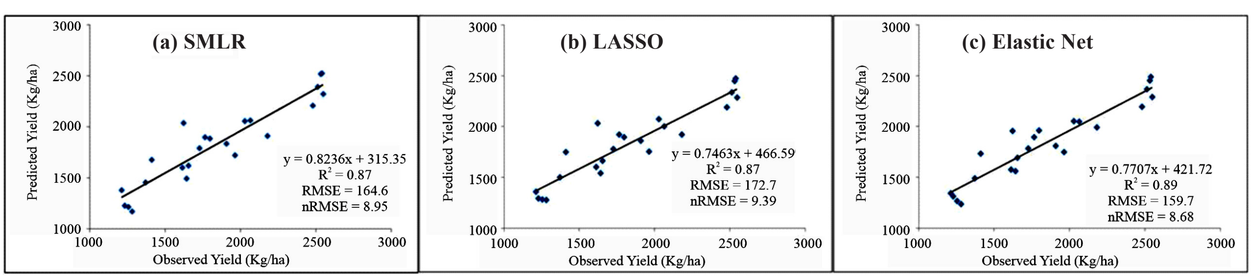

Fig. 3. Performance during calibration of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at flowering stage

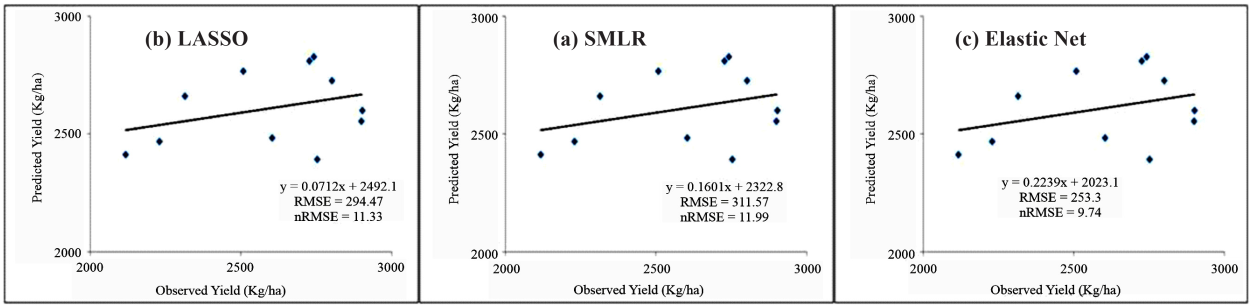

Fig. 4. Performance during Validation of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at flowering stage

Table 3. Mustard yield estimation at flowering stage by SMLR, LASSO and Elastic Net model during *Rabi* 2018-19 and 2019-20

Model	Model equation	Estimated yield (kg/ha)		Observed yield (kg/ha)		% deviation	
		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
SMLR	y=1933.2+57.17*time+ 83.8*Z21	3061.7	3127.6	2725.6	2746.6	12.33	13.87
LASSO	y=1845.58+52.66*time+ 58.32*Z21+0.04*Z151	2964.5	3001.8	2725.6	2746.6	8.76	9.29
Elastic Net	y=1739.53+42.04*time+ 1.95*Z11+36.38*Z21	2881.9	2797.7	2725.6	2746.6	5.73	1.86

value at flowering stage as compared to corresponding value at vegetative stage. Vashisth *et al*. (2018) reported that percentage deviation of estimated yield by actual yield of maize crop done at flowering stage and at grain filling stage was 10.3 and 7.1% by weather based statistical model.

Mustard yield estimation at grain filling stage by SMLR, LASSO and Elastic Net model

Model for estimating mustard yield at grain filling stage was developed by SMLR. LASSO and Elastic Net techniques for IARI, New Delhi using long period daily weather data from 40th to 8th standard meteorological week as well as long term crop yield data for IARI, New Delhi. Performances of the model developed for mustard yield estimation at grain filling stage during calibration are shown in Fig 5. Value of coefficient of determination R^2 for models developed for estimating the mustard crop yield at grain filling stage was 0.87% for SMLR model followed by 0.88% for LASSO model and 0.90% for Elastic net modal. The RMSE value during calibration was lowest for Elastic Net modal (150.7 kg/ha) followed by SMLR (159.7 kg/ha) and LASSO (165.7 kg/ha). During calibration nRMSE value was < 10% for all three models having lowest value 8.19% for Elastic Net followed by 8.68% for SMLR and 9.01% for LASSO model. Model performance developed by SMLR, LASSO and Elastic Net techniques for mustard yield estimation during validation at grain filling stage are given in Fig 6. The RMSE value during validation was lowest for Elastic Net modal (253.2 kg/ha) followed by LASSO (274.5 kg/ha) and SMLR (306.7 kg/ha). During validation nRMSE value was lowest 9.74% for Elastic Net followed by 10.56% for LASSO and 11.80% for SMLR model.

The most important weather parameters identified by SMLR for mustard yield estimation at grain filling stage are time and Z21 (weighted minimum temperature). For LASSO most important weather parameter are time, Z11 (weighted maximum temperature), Z21 (weighted

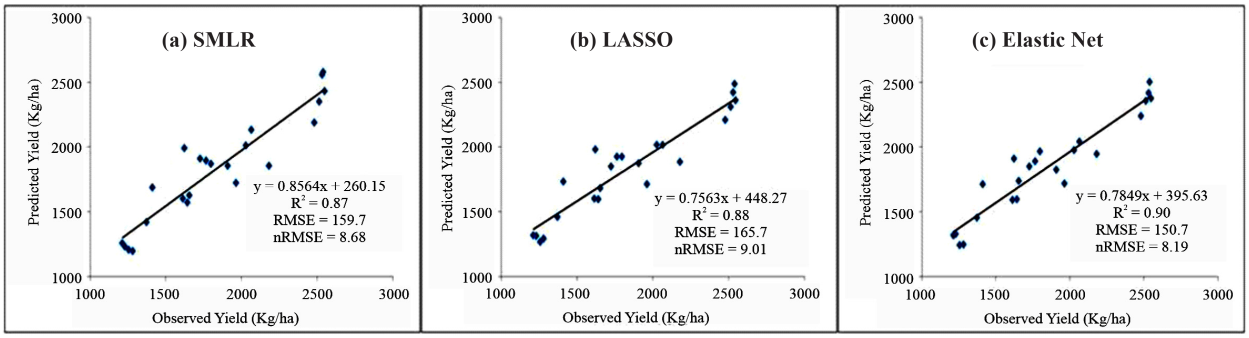

Fig. 5. Performance during calibration of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at grain filling stage

minimum temperature) and Z241 (weighted minimum temperature*morning relative humidity). For Elastic Net most important weather parameters are time, Z11 (weighted maximum temperature), Z21 (weighted minimum temperature) and Z241 (weighted minimum temperature*morning relative humidity).. Equation developed by SMLR, LASSO and Elastic Net model for mustard crop yield estimation at grain filling stage are given in table 4.

Percentage deviation of estimated yield for mustard crop done at grain filling stage by observed yield for IARI, New Delhi during *Rabi* 2018-19 and 2019-20 are shown in table 4. During *Rabi* 2018-19 the percentage deviation was lowest 3.13% for Elastic Net model followed by 4.24% for LASSO and 9.94% for SMLR modal respectively. During *Rabi* 2019-20 percentage deviation of estimated yield by observed yield was 0.75% for Elastic net, 9.74% for LASSO and 14.89% for SMLR model respectively. Lowest value of percentage deviation was observed by Elastic Net followed by LASSO and SMLR in both the year. Vashisth *et al*., (2014) reported that percentage deviation of observed yield by estimated yield done at forty-five days before harvest by weather based statistical model was found to be 10.7, 5.7 and 8.53 respectively during the period of 2011-12, 2012-13 and 2013-14. Similarly, the percentage deviation of yield estimation done at 25 days before harvest by weather based statistical model was 9.7, 7.0 and 8.29 respectively. In our study based on percentage deviation of estimated yield by observed yield done at tillering, flowering and grain filling stage, Elastic Net is giving better results followed by LASSO and SMLR. Kumar *et al.* (2019) used Stepwise and LASSO regression variable selection techniques and weather parameters for developing regression forecast model forty-five days before harvest. Based on forecast model results he found that stepwise forecast model over fit, whereas LASSO

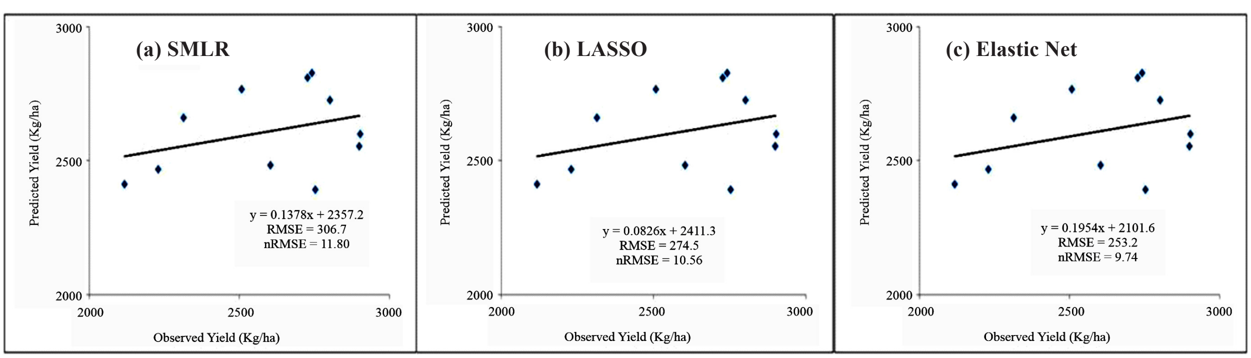

Fig. 6. Performance during validation of model developed using (a) SMLR, (b) LASSO and (c) Elastic Net models for mustard yield estimation at grain filling stage

Table 4. Mustard yield estimation at grain filling stage by SMLR, LASSO and Elastic Net model during *Rabi* 2018-19 and 2019-20

Model	Model equation	Estimated yield (kg/ha)		Observed yield (kg/ha)		% deviation	
		2018-19	2019-20	2018-19	2019-20	2018-19	2019-20
SMLR	y=1656.66+57.15*time+ 78.95*Z21	2996.5	3155.4	2725.6	2746.6	9.94	14.89
LASSO	y=2342.81+50.57*time+ 15.67*Z11+39.05*Z21+ 0.13*Z241	2841.3	3014.2	2725.6	2746.6	4.24	9.74
Elastic Net	y=2268.23+39.76*time+ 15.86*Z11+30.39*Z21+ 0.03*Z241	2810.9	2767.2	2725.6	2746.6	3.13	0.75

performs better fit model. Also, the per cent error by LASSO regression model was less than Stepwise regression. He inferred that LASSO variable selection method performed better than stepwise. Model developed using different methods using weather parameters had lower value of nRMSE and root mean square error (RMSE) for the yield forecast done by the model at grain filling stage as compared to flowering and tillering stage. This indicates better performance of the model at the grain filling stage. This work is line of the pre-harvest forecast models for several crops based on time series data on crop yield and weekly data on weather variables developed by various research workers (Pandey *et al*., 2014; Azfar *et al*., 2015; Yadav *et al*., 2015). Das *et al*. (2018) used long-term weather data and six different statistical methods for determination of rice yield prediction. Based on Friedman test overall ranking he reported that LASSO (2.63) and Elastic Net (3.07) were the best model.

Conclusions

On evaluation of overall performance of different models used for multistage mustard crop yield estimation Elastic Net models have lowest value of nRMSE and RMSE followed by LASSO and SMLR model. Percentage deviation of estimated yield by observed yield at different growth stage, Elastic net has least value followed by LASSO and SMLR model. From this study it may be concluded that Elastic Net, LASSO and SMLR model based on weather parameters can be used for district level yield estimation at different crop growth stage and Elastic Net performed best among all the three model followed by LASSO and SMLR model.

References

Agrawal, R., Chandrahas and Aditya, K. 2012. Use of discriminant function analysis for forecasting crop yield. *Mausam* **36(3)**: 455-458.

Agrawal, R. and Mehta, S.C. 2007. Weather based forecasting of crop yields, pests and diseases-IASRI models. *Journal of Indian Society of Agricultural Statistics* **61(2)**: 255-263.

Azfar, M., Sisodia, B.V.S., Rai, V.N. and Devi, M. 2015. Pre-harvest forecast models for rapeseed & mustard yield using principal component. *Mausam* **4**: 761-766.

Chauhan, V.S., Shekh, A.M., Dixit, S.K., Mishra, A.P. and Kumar, S. 2009. Yield prediction model of rice in Bulsar district of Gujarat. *Journal of Agrometeorology* **11(2)**: 162-168.

Das, B., Nair, B., Reddy, V.K. and Venkatesh, P. 2018. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology* **62(10)**: 1809-1822.

Dutta, S., Patel, N.K. and Srivastava, S.K. 2001. District wise yield models of rice in Bihar based on water requirement and meteorological data. *Journal of the Indian Society of Remote Sensing* **29**(3): 175-182.

Garde, Y.A., Dhekale, B.S. and Singh, S. 2015. Different approaches on pre harvest forecasting of wheat yield. *Journal of Applied and Natural Science* **7**(2): 839-843.

Kumar R., Gupta B.R.D., Athiyaman, B., Singh K.K. and Shukla R.K. 1999. Stepwise regression technique to predict pigeon pea yield in Varanasi district. *Journal of Agrometeorology* **1**(2): 183-186.

Kumar, S., Attri, S.D. and Singh, K.K. 2019. Comparison of Lasso and stepwise regression technique for wheat yield predication. *Journal of Agrometeorology* **21**(2): 188-192.

Pandey, K.K., Rai, V.N. and Sisodia, B.V.S. 2014. Weather variables based rice yield forecasting models for Faizabad district of eastern UP. *International Journal of Agricultural and Statistical Sciences* **10**(2): 381-385.

Singh, R.S., Patel, C., Yadav, M.K. and Singh, K.K. 2014. Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *Journal of Agrometeorology* **16**: 199-202.

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

Vashisth, Ananta, Singh R. and Choudary, Manu 2014. Crop yield forecast at different growth stage of wheat crop using statistical model under semi arid region. *Journal of Agro ecology and Natural Resource Management*: 1-3.

Vashisth, Ananta, Goyal, Avinash and Roy, Debasish 2018. Pre harvest maize crop yield forecast at different growth stage using different model under semi arid region of India. *International Journal of Tropical Agriculture* **36**(4): 915-920.

Yadav, R.R. and Sisodia, B.V.S. 2015. Predictive models for Pigeon-pea yield using weather variables. *International Journal of Agricultural and Statistical Sciences* **11**(2): 462-472.

Received: September 12, 2020; Accepted: November 26, 2020