Journal of Agrometeorology

ISSN : 0972-1665
Vol. No. 24 (1) : 18-25 (March 2022)

https://journal.agrimetassociation.org/index.php/jam

Research Paper

Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models

K. S. ARAVIND[1], ANANTA VASHISTH[1*], P. KRISHANAN[1] and B.DAS[2]

[1]ICAR-Indian Agricultural Research Institute, New Delhi, India
[2]ICAR-Central Coastal Agricultural Research Institute, Goa, India
Corresponding Author e-mail: ananta.iari@gmail.com

ABSTRACT

Wheat yield production is largely attributed by weather parameters. Model developed by multiple linear, neural network and penalised regression techniques using weather data have the potential to provide reliable, timely and cost-effective prediction of wheat yield. Wheat yield data and weather parameter during crop growing period (46th to 15th SMW) for more than 30 years were collected for study area and model was developed using stepwise multiple linear regression (SMLR), principal component analysis (PCA) in combination with SMLR, artificial neural network (ANN) alone and in combination with PCA, least absolute shrinkage and selection operator (LASSO) and elastic net (ENET) techniques. Analysis was carried out by fixing 70% of the data for calibration and remaining dataset for validation. On examining these models, LASSO and elastic net are performing excellent having nRMSE value less than 10 % for four out of five location and good for one location, because of prevention in over fitting and reducing regression coefficient by penalization.

Keywords: Wheat yield prediction, stepwise multiple linear regression, principal components analysis, artificial neural network, least absolute shrinkage and selection operator, elastic net

Wheat (*Triticum aestivum*) is second most consumed important staple food grain after rice, grown widely in the northern part of India. Wheat crop is thermo-sensitive in nature. Adverse changes in the weather parameters affect the crop growth and development and shows declining trends in yield. Crop yield forecast is essential in regard to storage, import, export and improves the decisions of government planning and policy-making to manage the produce. In traditional methods, crop cutting experiments were widely used for crop yield forecast. Models provide alternative methods for crop yield prediction. These methods are fast, cost effective and give the understanding about the factors which affect the crop yield. Statistical method is widely used for crop yield prediction using weather data (Lobell and Burke, 2010; Shi *et al*., 2013). Singh *et al*. (2014) used weather indices such as minimum temperature and maximum temperature, rainfall, relative humidity for forecasting of rice and wheat yield in nine districts of eastern UP by stepwise regression. Garde *et al*. (2015) reported that model developed by weather indices along with incorporation of technical and statistical indicators was found to be best as compared to model developed based on only weather indices for the wheat yield forecast. Sisodia *et al*. (2017) reported that pre-harvest forecast of wheat in Faizabad district of Uttar Pradesh, based on biometric characters for both early and late sowing varieties showed best result in principle component analysis model. Vashisth *et al*. (2014) reported that percentage deviation of forecasted yield of wheat crop from actual yield using statistical model for forecast done at forty-five days before harvest was 10.7, 5.7 and 8.53 respectively and for forecast done at 25 days before harvest was 9.7, 7.0 and 8.29 respectively during *Rabi* 2011-12, 2012-13 and 2013-14. Kumari *et al*. (2016) indicated that weather-based forecasting of pigeon pea yield in Varanasi region by artificial neural network model was best among other regression models. Azfar (2015) showed the effectiveness of PCA considering all weather indices including interaction indices as regressors was best reliable forecast model for mustard and rapeseed compared to other models. Highest predictive accuracy for district level mustard yield in Haryana was found in stepwise multiple regression and PCA by the inclusion of crop condition term along with the weather parameters (Verma *et al*., 2016). Das *et al*. (2018) reported that out of six multivariate models developed using long term weather variables for rice crop in West coast of India, LASSO, Elastic net and SMLR was found to be the best.

Article info - DOI: https://doi.org/10.54386/jam.v24i1.1002
Received: 29 September 2021; Accepted: 6 January 2022; Published online: 11 February 2022

To overcome the various challenges in crop yield prediction, in the present investigation, models was developed using SMLR, PCA-SMLR, ANN, PCA-ANN, LASSO and Elastic Net techniques for improving the accuracy of wheat yield prediction.

MATERIALS AND METHODS

Data collection

Weather data viz maximum temperature, minimum temperature, rainfall, morning and evening relative humidity, sunshine hours during crop growing period of wheat as well as wheat yield data were collected from Hisar (1985-2017), Ludhiana (1971-2017), Amritsar (1972-2017), Patiala (1972-2016) and IARI, New Delhi (1985-2018). Weather parameter range during wheat growing period for different locations is given in Table1. Data analysis was done after converting daily weather data into simple and weighted composite indices. 70% of the total dataset for each location were used for the calibration of the model and remaining 30% were used for validation of the model.

Development of wheat yield prediction model using different techniques

Simple and weighted weather indices are developed for each station. 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 and weighted weather indices used for developing model are given in Table 2.

Simple weather indices:

$$Zij = \sum_{w=1}^{m} Xiw$$

$$Zii'j = \sum_{w=1}^{m} Xiw\, Xi'w$$

Weighted weather indices:

$$Zij = \sum_{w=1}^{m} r^{\,j}iw\; Xiw$$

$$Zii'j = \sum_{w=1}^{m} r^{j}ii'w\; Xiw\; Xi'w$$

Where,

Xiw/ Xii′ w = value of th/th weather variable in th week.

r ʲ*iw/r* ʲ*ii′* w = correlation coefficient of yield with *i*th weather variable or product of th or *i′*th weather variable in th week.

m = week at which forecast done.

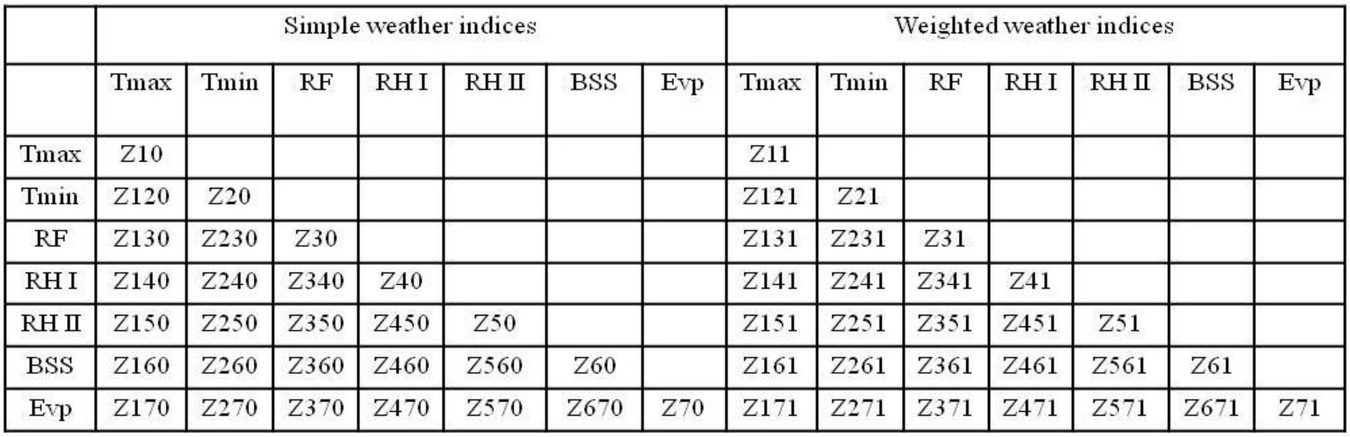
P = number of variables

Impact of important weather indices were determined by stepwise multiple linear regression technique and using different simple and weighted weather indices wheat yield prediction models was developed. Stepwise multiple linear regression-principal component analysis (SMLR-PCA) is a combination of feature selection and selection method used for the data analysis. Principal components scores or factors are calculated from the data analysis which is used as an input variable for stepwise multiple linear regression. PCA is a multivariate technique used for data reduction and reduce multicolinearity problems, transforms original set of correlated variables in to a new set of uncorrelated variables. Principal components (PCs) were selected based on their Eigen values. Eigen values more than 1 condition can able to describe more than 90 percent variability in the data. PCA scores were used as input for SMLR analysis. Artificial neural network consists of many artificial neurons that are connected together to network architecture specifically. Neural network has various architectures to approximate any linear function such as feed forward network, feedback network, lateral network etc. ANN composed of three layers namely, input layer, hidden layer and output layer. Multilayer perceptron (MLP) technique is one of the popular neural network types. This network interpreted as a form input-output model, with weights and threshold (biases) as free parameters of the model. By learning process, it attains optimized weighted value of variables, and it tries to produce the output based on the corresponding input provided. The main objective of the neural network is to produce its own output having reduced discrepancies with target output value, which will help to transform the input into meaningful output. In Principal component analysis-Artificial neural network (PCA-ANN) techniques data analysis were done through combination of feature selection. Principle components scores or factors are calculated from the data analysis which is used as an input variable for ANN. Least Absolute Shrinkage and Selection Operator (LASSO) is a model selection technique. Lasso models are used to overcome the shortcomings of ordinary least square (OLS) and ridge regression. 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 reduce multicollinearity and retain some good features of both subset selection and ridge regression. With large number of predictors, smaller subset selection exhibit stronger effect on interpretation of data. Subset selection is discrete and variable process, repressor are either retained or eliminated from the model in order to provide better interpretable model. 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. Alpha and beta are the two model parameters, need to be optimized by minimizing average mean square error in cross validation. Tuning parameter alpha values set in LASSO and Elastic Net were 1 and 0.5. “glmnet” package in R software was used to solve LASSO and ENET.

Table 1: Weather parameter range during wheat growing period for different location

Location	Latitude and Longitude	Tmax (°C)	Tmin (°C)	RF (mm)	RH I (%)	RH II (%)	BSS (hrs)	Evp (mm/day)
Hisar	29.1492° N, 75.7217° E	22.2-25.5	6.3-9.7	17.5-226.2	79.5-88.6	30.4-55.1	5.9-8.5	2.2-4.2
Ludhiana	30.9010° N, 75.8573° E	20.9-24.5	6.3-10.7	4.4-262.3	87.2-95.2	33.2-58.6	5.4-8.9	-
Amritsar	31.6340° N, 74.8723° E	18.7-24.7	5.5-9.5	20.8-361.9	79.5-97.4	47.3-68.8	-	-
Patiala	30.3398° N, 76.3869° E	22.2-26.3	8.6-12.0	56.5-1062.6	73.3-93.1	42.5-66.3	-	-
IARI, New Delhi	28.6377° N, 77.1571° E	23.9-26.6	8.2-11.2	16.7-315.8	79.1-93.1	35.1-63.1	4.6-8.1	2.9-4.9

Table 2: Simple and weighted weather indices used for developing model



	Simple weather indices							Weighted weather indices						
	Tmax	Tmin	RF	RH I	RH II	BSS	Evp	Tmax	Tmin	RF	RH I	RH II	BSS	Evp
Tmax	Z10							Z11						
Tmin	Z120	Z20						Z121	Z21					
RF	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		
BSS	Z160	Z260	Z360	Z460	Z560	Z60		Z161	Z261	Z361	Z461	Z561	Z61	
Evp	Z170	Z270	Z370	Z470	Z570	Z670	Z70	Z171	Z271	Z371	Z471	Z571	Z671	Z71

Performance of statistical models were estimated by calculating R^2, Root mean square error (RMSE), normalized root mean square error (nRMSE) and percentage deviation using the following formula.

$$R^2=1-\Sigma(y_i-\hat{y})^\wedge 2/\Sigma(y_i-\bar{y})2$$

Where, $\Sigma(y_i-\hat{y})2$ = sum squared regression error, = sum squared total error.

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

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

Percentage Deviation= $(P_i\text{-}O_i)$ *100/ O_i

Where RMSE is root mean square error, nRMSE is normalized root mean square error Pi is the predicted value, O_i is the observed value, N is the number of observations and M is the mean of observed value. Model performs excellent having nRMSE value less than 10%, good having nRMSE value between 10-20%, fair having nRMSE value between 20-30%.

RESULTS AND DISCUSSION

Performance of wheat yield prediction model developed by different techniques

Yield prediction models for wheat crop have been developed using long term crop yield data as well as long period daily weather data during crop growing period (46th to 15th standard meteorological week) for respective location. The coefficient of determination (R^2) was significant at 1% probability level for all the locations. Performance of models was categorized based on value of RMSE and nRMSE during validation and are presented hereunder for different locations.

Hisar

Performance of the model developed using different techniques for wheat yield prediction of Hisar is shown in the Table 3. During calibration the models had the value of coefficient of determination R^2 between 0.75 for PCA-SMLR to 0.96 for PCA-ANN. The RMSE value during calibration ranged between 82.0 for PCA-ANN to 217.2 kg ha^{-1} for PCA-SMLR. During calibration lowest value of nRMSE was found for PCA-ANN (2.15 %) followed by ANN (3.49 %), SMLR (3.75%), LASSO (4.27%), elastic net (4.41%) and PCA-SMLR (5.69%). During validation RMSE value ranged between 313.9 for SMLR and 586.3 kg ha^{-1} for PCA-SMLR. Based on nRMSE values during validation, the model predictions were excellent for SMLR (3.75 %), LASSO (4.27%), Elastic net

Table 3: Performance of the model developed using different techniques for wheat yield prediction of Hisar

Name of the Model	Equation	During calibration			During validation	
		R^2 ($p < 0.01$)	RMSE (kg ha^{-1})	nRMSE (%)	RMSE (kg ha^{-1})	nRMSE (%)
SMLR	y=3801.076+48.44×time-7.72×Z271-6.61×Z561-47.67×Z41	0.88	143.4	3.75	313.9	3.75
PCA-SMLR	y=3339.38+49.08×time-220.64×PC1; No of PC's: 8	0.75	217.2	5.69	586.3	13.33
PCA-ANN	No of hidden Neurons : 2, No of PC's: 8	0.96	82.0	2.15	493.0	11.01
ANN	No of hidden Neurons : 10	0.91	133.2	3.49	409.5	9.15
LASSO	y=4909.55+43.76×time-93.82×Z71-0.24×Z241-0.79×Z261-0.23×Z471-3.24×Z561	0.89	162.9	4.27	162.9	4.27
ENET	y=4670.21+25.2×time-20.83×Z71-0.04×Z151-0.1×Z260-0.55×Z261-0.04×Z270-0.01×Z451-0.73×Z471-0.4×Z561-0.4×Z671	0.89	168.3	4.41	220.4	5.01

Table 4: Performance of the model developed using different techniques for wheat yield prediction of Ludhiana

Name of the Model	Equation	During calibration			During validation	
		R^2 ($p < 0.01$)	RMSE (kg ha^{-1})	nRMSE (%)	RMSE (kg ha^{-1})	nRMSE (%)
SMLR	y=1944.40+56.35×time+1.60×Z151×4.17×Z361	0.92	159.5	4.30	486.5	10.11
PCA-SMLR	y=2751.846+58.702×time+188.670×PC2; No of PC's: 7	0.86	208.2	5.64	318.5	6.66
PCA-ANN	No of hidden Neurons : 1, No of PC's: 7	0.84	312.8	8.00	580.4	12.11
ANN	No of hidden Neurons : 9	0.90	286.4	7.27	1070.6	22.36
LASSO	y=1265.03+52.49×time+0.7×Z251+0.44×Z151+0.29 ×Z141+0.10×Z361	0.93	182.9	4.73	454.5	9.54
ENET	y=1485.58+36.79×time+7.39×Z21+6.16×Z41+0.62× Z121+0.58×Z251+0.45×Z151+0.08×Z141	0.94	175.3	4.53	372.6	7.81

(5.01%) and ANN (9.15 %), good for PCA-ANN (11.01%) and PCA-SMLR (13.33 %). The most important weather parameter identified using SMLR was Z41, Z271 and Z561. For PCA-SMLR, time was found the most important parameter influencing the crop yield followed by PC1. For developing wheat yield prediction model using ANN techniques, the Z variates were taken as inputs. The optimum number of hidden neurons was 10. For wheat yield prediction model using PCA feature extraction method followed by ANN, PCA factors were generated and time along with PCA factors were considered for developing PCA-ANN models. The number of PCs and optimum number of hidden neurons ranged between 8 and 2. Numbers of hidden neurons were less in PCA-ANN compared to ANN model. LASSO model has the characteristics of automatic variable selection, reducing multicolinearity and minimized residual mean square error. Wheat yield predication done for Hisar using LASSO and elastic net, weather indices Z71, Z241, Z261, Z471 and Z561 have negative influence on yield for LASSO and Z71, Z151, Z260, Z261, Z270, Z451, Z471, Z561 and Z671have negative influence on yield for Elastic Net. On the basis of RMSE and nRMSE value during validation of models developed using different techniques for wheat crop prediction for Hisar SMLR performed best having followed by LASSO, Elastic Net, ANN, PCA-ANN and PCA-SMLR.

Ludhiana

Performance of the model developed using different techniques for wheat yield prediction of Ludhiana is shown in the Table 4. During calibration models had the value of coefficient of determination R^2 ranged between 0.84 for PCA-ANN to 0.94 for elastic net. The RMSE value during calibration was lowest for SMLR (159.5 kg ha^{-1}) followed by Elastic Net (175.3 kg ha^{-1}), LASSO (182.9 kg ha^{-1}), PCA-SMLR (208.2 kg ha^{-1}), ANN (286.4 kg ha^{-1}) and PCA-ANN (312.8 kg ha^{-1}). During calibration nRMSE value was ranged between 4.3 % for SMLR to 8.0 % for PCA-ANN. During validation RMSE value ranged between 318.5 kg ha^{-1} for PCA-SMLR to 1070.6 kg ha^{-1} for ANN. Based on nRMSE values during validation, the model predictions were excellent for PCA-SMLR (6.66 %), Elastic Net (7.80 %) and LASSO (9.54%), good for SMLR (10.11%) and PCA-ANN (12.11%), fair for ANN having nRMSE value 22.36%. The most important weather parameter identified using SMLR for Ludhiana was Z151 and Z361. For PCA-SMLR model, time was found the most important parameter influencing the crop yield followed by PC2. For developing wheat yield prediction model using ANN techniques, optimum number of hidden neurons was 9. For PCA-ANN models, number of PCs and optimum number of hidden neurons was 7 and 1. Using LASSO the

Table 5: Performance of the model developed using different techniques for wheat yield prediction of Amritsar

Name of the Model	Equation	During calibration			During validation	
		R^2 ($p < 0.01$)	RMSE (kg ha^{-1})	nRMSE (%)	RMSE (kg ha^{-1})	nRMSE (%)
SMLR	y=851.64+70.85×-time+1.89×Z121+0.68×Z131+0.094×Z250	0.95	150.7	4.79	573.8	12.85
PCA-SMLR	y=2121.18+71.76×time+154.21×PC2; No of PC's: 6	0.91	203.3	6.46	529.0	11.85
PCA-ANN	No of hidden Neurons : 1, No of PC's: 6	0.92	245.6	7.56	606.3	13.66
ANN	No of hidden Neurons : 7	0.81	366.3	11.28	853.6	19.23
LASSO	y=820.52+62.44×time+0.79×Z121+0.50×Z131+0.042×Z150+0.038×Z151+0.00067×Z50	0.94	174.8	5.56	427.4	9.58
ENET	y=1031.07+51.59×TIME+4.59×Z21+1.74×Z31+0.044×Z121+0.25×Z131+0.038×Z150+0.33×Z151+0.01×Z341	0.94	169.3	5.38	423.9	9.50

Table 6: Performance of the model developed using different techniques for wheat yield prediction of Patiala

Name of the Model	Equation	During calibration			During validation	
		R^2 ($p < 0.01$)	RMSE (kg ha^{-1})	nRMSE (%)	RMSE (kg ha^{-1})	nRMSE (%)
SMLR	y=-1667.55+87.15×time +0.16×Z141+6.96×Z20	0.95	160.2	4.77	931.2	20.03
PCA-SMLR	y=1953.22+95.96×time+91.43×PC1; No of PC's: 5	0.94	193.1	5.75	743.5	15.99
PCA-ANN	No of hidden Neurons : 1, No of PC's: 5	0.93	264.4	7.67	749.4	16.14
ANN	No of hidden Neurons : 6	0.92	245.3	7.12	365.6	7.87
LASSO	y=-2376.45 +79.34×time +25.15×Z11 +7.82×Z20 +19.68×Z21 +1.53×Z41-0.25×Z120 +0.007×Z130 +0.07×Z141	0.98	111.9	3.28	772.6	16.66
ENET	y=-1048.63 +64.63 ×time +1.82×Z11 +13.55×Z21 +1.48×Z41 +0.30×Z51+0.11×Z141+0.018×Z241+0.0005×Z451	0.98	109.3	3.2	740.0	15.96

most influencing weather parameter for wheat yield predication was Z251, Z151, Z141, Z361 and using Elastic net, the most influencing weather parameter for wheat yield predication was Z21, Z41, Z121, Z251, Z151 and Z141. On the basis of RMSE and nRMSE value during validation of models developed using different techniques for wheat crop prediction for Ludhiana PCA-SMLR performed best followed by Elastic Net, LASSO, SMLR, PCA-ANN and ANN.

Amritsar

Performance of the model developed using different techniques for wheat yield prediction of Amritsar is shown in the Table 5. During calibration value of coefficient of determination R^2 for model developed by different techniques was between 0.81 for ANN to 0.95 for SMLR. The RMSE value during calibration was between 150.7 kg ha^{-1} for SMLR to 366.3 kg ha^{-1} for ANN. The value of nRMSE was lowest for SMLR (4.79%) followed by Elastic Net (5.38%), LASSO (5.56%), PCA-SMLR (6.46%), PCA-ANN (7.56%) and ANN (11.28%). During validation RMSE value was lowest for Elastic net (423.9 kg ha^{-1}) followed by LASSO (427.4 kg ha^{-1}), PCA-SMLR(529.0 kg ha^{-1}), SMLR (573.8 kg ha^{-1}), PCA-ANN (606.3 kg ha^{-1}) and ANN (853.6 kg ha^{-1}). Based on nRMSE values during validation, the model predictions were excellent for elastic net (9.50 %) and LASSO (9.58 %), good for PCA-SMLR (11.85%), SMLR (12.85%), PCA-ANN (13.66%) and ANN (19.23%). The most important weather parameter identified using SMLR for wheat prediction of Amritsar was Z121, Z131 and Z250. For PCA-SMLR model, time was found the most important parameter influencing the crop yield followed by PC2. For developing wheat yield prediction model using ANN techniques, optimum number of hidden neurons was 7. For PCA-ANN models, number of PCs and optimum number of hidden neurons was 6 and 1. Using LASSO the most influencing weather parameter for wheat yield predication of Amritsar was Z121, Z131, Z150, Z151 and Z50. Using Elastic net, the most influencing weather parameter for wheat yield predication of Amritsar was Z21, Z31, Z121, Z131, Z150, Z151 and Z341.

On the basis of RMSE and nRMSE value during validation of models developed using different techniques for wheat crop prediction for Amritsar Elastic Net performed best followed by LASSO, PCA-SMLR, SMLR, PCA-ANN and ANN.

Table 7: Performance of the model developed using different techniques for wheat yield prediction of IARI, New Delhi

Name of the Model	Equation	During calibration			During validation	
		R^2 ($p < 0.01$)	RMSE (kg ha^{-1})	nRMSE (%)	RMSE (kg ha^{-1})	nRMSE (%)
SMLR	$y=2968.63+47.33\times time+0.26\times Z341$	0.83	136.9	4.04	382.7	9.06
PCA-SMLR	$y=2780.2+51.39\times time +88.27\times PC3$; No of PC's: 10	0.8	157.6	4.65	260.6	6.2
PCA-ANN	No of hidden Neurons : 2, No of PC's: 10	0.85	161.8	4.7	618.4	14.58
ANN	No of hidden Neurons : 8	0.90	122.9	3.57	656.8	15.48
LASSO	$y=2623.84+36.56\times time+0.0054\times Z10-0.10\times Z50-1.84\times Z60-0.21\times Z70-90.9\times Z71 +0.008\times Z120 +0.12\times Z131+0.31 \times Z141+0.12\times Z151-0.1\times Z160 +0.02\times Z240+0.44 \times Z241+2.46\times Z261+5.42\times Z271-0.19\times Z360-0.0004\times Z450 +0.005\times Z470-0.085\times Z560+1.1\times Z670$	0.98	45.8	1.35	258.0	6.11
ENET	$y=3350.1+19.84\times time-Z60+0.24Z121-0.016\times Z160+ 0.75\times Z171 -0.22\times Z360-0.0002\times Z460+0.053\times Z471$	0.95	79.1	2.33	351.9	8.33

Patiala

Performance of the model developed using different techniques for wheat yield prediction of Patiala is shown in the Table 6. During calibration value of coefficient of determination R^2 for model developed by different techniques was lowest for ANN (0.92) followed by PCA-ANN (0.93), PCA-ANN (0.94), SMLR (0.95), LASSO (0.98) and Elastic net (0.98). The RMSE value during calibration was between 109.3 kg ha^{-1} for elastic net 264.4 kg ha^{-1} for PCA-ANN. The value of nRMSE was lowest for Elastic net (3.2 %) followed by LASSO (3.28%), SMLR (4.77%), PCA-SMLR (5.75%), ANN (7.12%) and PCA-ANN (7.67%). During validation RMSE value was lowest for ANN (365.6 kg ha^{-1}) followed by Elastic net (740.0 kg ha^{-1}), PCA-SMLR (743.5 kg ha^{-1}), PCA-ANN (749.4 kg ha^{-1}), LASSO (772.6 kg ha^{-1}), and SMLR (931.2 kg ha^{-1}). Based on nRMSE values during validation, the model predictions were excellent for ANN (7.87 %), good for elastic net (15.96%), PCA-SMLR (15.99%), ANN (16.14%), LASSO (16.66 %) and SMLR (20.03%). The most important weather parameter identified using SMLR for wheat prediction of Patiala was Z141 and Z20. For PCA-SMLR model, time was found the most important parameter influencing the crop yield followed by PC1. For developing wheat yield prediction model using ANN techniques, optimum number of hidden neurons was 6. For PCA-ANN models, number of PCs and optimum number of hidden neurons was 5 and 1. Using LASSO the most influencing weather parameter for wheat yield predication of Patiala was Z11, Z20, Z21, Z41, Z130 and Z141. Z120 has negative influence on wheat yield. Using Elastic net, the most influencing weather parameter for wheat yield predication of Patiala was Z11, Z21, Z41, Z51, Z141, Z241 and Z451. On the basis of RMSE and nRMSE value during validation of models developed using different techniques for wheat crop prediction for Patiala ANN performed best followed by Elastic Net, PCA-SMLR, PCA-ANN, LASSO and SMLR.

IARI, New Delhi

Performance of the model developed using different techniques for wheat yield prediction of IARI, New Delhi is shown in the Table 7. During calibration value of coefficient of determination R^2 for model developed by different techniques was between 0.80 for PCA-SMLR to 0.98 for LASSO. The RMSE value during calibration was lowest 45.8 kg ha^{-1} for LASSO followed by 79.1 kg ha^{-1} for elastic net, 122.9 kg ha^{-1} for ANN, 136.9 kg ha^{-1} for SMLR, 157.6 kg ha^{-1} for PCA-SMLR and 161.8 kg ha^{-1} for PCA-ANN. During calibration all models developed by different techniques have nRMSE values less than 10 % with lowest value 1.35 % for LASSO followed by 2.33 % for elastic net, 3.57 % for ANN, 4.04% for SMLR, 4.65% for PCA-SMLR and 4.7% for PCA-ANN. During validation RMSE value was lowest for LASSO (258.0 kg ha^{-1}) followed by PCA-SMLR (260.6 kg ha^{-1}), Elastic net (351.9 kg ha^{-1}), SMLR (382.7 kg ha^{-1}), PCA-ANN (618.4 kg ha^{-1}) and ANN (656.8 kg ha^{-1}). Based on nRMSE values during validation, the model predictions were excellent for LASSO (6.11 %), PCA-SMLR (6.2%), elastic net (15.96%) and SMLR (9.06%), good for PCA-ANN (14.58%) and ANN (15.48%). The most important weather parameter identified using SMLR for wheat prediction of IARI, New Delhi was Z341. For PCA-SMLR model, time was found the most important parameter influencing the crop yield followed by PC3. For developing wheat yield prediction model using ANN techniques, optimum number of hidden neurons was 8. For PCA-ANN models, number of PCs and optimum number of hidden neurons was 10 and 2. Using LASSO the most influencing weather parameter for wheat yield prediction of IARI, New Delhi was Z10, Z120, Z131, Z141, Z151, Z240, Z241, Z261, Z271, Z470, Z670 while Z50, Z60, Z70, Z71, Z160, Z360, Z450, Z560 has negative influence on wheat yield prediction. Using Elastic net, the most influencing weather parameter for wheat yield predication was Z121, Z171, Z471, while Z60, Z160, Z360, Z460 has negative influence on wheat yield prediction. On the basis of RMSE and nRMSE value during validation of models developed using different techniques for wheat crop prediction for IARI, New Delhi LASSO performed best followed by PCA-SMLR, Elastic Net, SMLR, PCA-ANN and ANN.

In our study, the performance based on RMSE and nRMSE during validation of different models for wheat crop prediction of different locations showed that Elastic Net and LASSO performed excellent for Hisar, Ludhiana, Amritsar, IARI, New Delhi and good for Patiala. Tibshirani (1996) proposed the method LASSO for shrinkage and selection for regression and generalized regression 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. PCA-SMLR performed excellent for Ludhiana and IARI, New Delhi, good for Hisar, Amritsar and Patiala. SMLR model performed excellent for Hisar and IARI, New Delhi, good for Ludhiana and Amritsar, fair for Patiala. PCA-ANN model performed good for all the five districts. ANN model performed excellent for Hisar and Patiala, good for Amritsar and IARI, New Delhi and fair for Ludhiana. The range of RMSE and nRMSE of the model developed was superior in PCA-ANN as compared to ANN for Ludhiana, Amritsar and Patiala. This result is in line with previous findings of Suleiman *et al*. (2016) while comparing ANN and PCA-ANN for predicting roadside particulate matter. Singh *et al.* (2014) used eighteen years weather data and yield data of rice and wheat for nine districts of Eastern Uttar Pradesh for developing yield prediction equations. They indicated that models explained 51 to 79 percent variations for rice yield and 65 to 92 percent variations for wheat yield in different districts. The performance of ANN was good during calibration while it was the worst model during validation which indicated over fitting. The overall ranking based on RMSE and nRMSE value during validation revealed that LASSO and Elastic net is performing best as compared to other models. Our result is in line with previous findings reported by (Das *et al*., 2018). They used six models SMLR, PCA-SMLR, ANN, PCA-ANN, LASSO and Elastic Net for prediction of rice yield based on weather parameters for west cost of India and he found that LASSO performed best followed by Elastic Net. LASSO and Elastic Net showed good performance due to the prevention of over fitting of model and reducing the magnitude of regression coefficient with feature selection by penalization decreases the model complexity. These penalised models give better computational advantage over SMLR or ANN as the features with zero coefficients can be eliminated from the model. The feature selection algorithms like LASSO, Elastic Net and SMLR performed better than methods utilising all the weather indices like ANN as feature selection reduces over fitting and avoids multicollinearity present in the dataset. Vashisth and Aravind (2020) reported that on the basis of percentage deviation and model accuracy Elastic Net model was found best followed by LASSO and SMLR for multistage mustard yield estimation done at vegetative, flowering and grain filling stage during *Rabi* 2018-19 and 2019- 20. 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. They reported that performance of LASSO regression is better than stepwise regression.

CONCLUSION

In the present study six models were developed for prediction of wheat yield for five different locations using long term weather data. Results showed that LASSO and Elastic Net performed excellent for Hisar, Ludhiana, Amritsar, IARI New Delhi and good for Patiala. PCA-SMLR performed excellent for Ludhiana and IARI, New Delhi, good for Hisar, Amritsar and Patiala. SMLR performed excellent for Hisar and IARI, New Delhi, good for Ludhiana and Amritsar, fair for Patiala. PCA-ANN performed good for all the five districts. ANN performed excellent for Hisar and Patiala, good for Amritsar and IARI, New Delhi and fair for Ludhiana. Hence out of six different models, Elastic Net and LASSO was found to be the best model followed by PCA-SMLR, SMLR, PCA-ANN and ANN respectively for wheat yield prediction.

ACKNOWLEDGEMENT

The first author acknowledges PG school, Indian Agricultural Research Institute, New Delhi for providing fellowship. Authors acknowledge the research facilities provided by the Director and Head, Division of Agricultural Physics, Indian Agricultural Research Institute, New Delhi.

Conflict of Interest Statement: The author(s) declare(s) that there is no conflict of interest.

Disclaimer: The contents, opinions and views expressed in the research article published in Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

REFERENCES

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.

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. *Int. J. Biometeorol.*, ٦٢(10), 1809-1822.

Garde, Y. A., Dhekale, B. S. and Singh, S. (2015). Different approaches on pre harvest forecasting of wheat yield, *J. Appli. Natural Sci., 7* (2): 839 – 843.

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

Kumari, P., Mishra, G.C. and Srivastava, C.P. (2016). Statistical models for forecasting pigeon pea yield in Varanasi region. *J Agrometeorol.,* 18(18): 306–310.

Lobell, D. B. and Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agri. Forest Meteorol.,* 150:1443–1452.

Shi W, Tao F and Zhang Z (2013). A review on statistical models for identifying climate contributions to crop yields. *J. Geogr. Sci.,* 23:567–576.

Singh, R. S., Patel, C., Yadav, M. K., and Singh, K. K., (2014). Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *J. Agrometeorol*., 16:199– 202.

Sisodia, B. V. S. and Rai, V. N. (2017). An application of principal component analysis for pre-harvest forecast model for wheat crop based on biometrical characters, *Int. Res. J. Agri. Econo.Stat.*, 8(1): 83-87.

Suleiman, A., Tight, M.R. and Quinn, A.D. (2016) Hybrid neural networks and boosted regression tree models for predicting roadside particulate matter. *Environ. Model Assess* 21:731–750.

Tibshirani, R. (1996). Regression shrinkage and selection via LASSO**.** *J. Roy Stat. Soc. B*, 58:267–288.

Vashisth, Ananta and Aravind, K.S. (2020). Multistage Mustard Yield Estimation Based on Weather Variables using Multiple Linear, LASSO and Elastic Net Models for Semi Arid Region of India. *J. Agri. Physi.,* 20(2): 213-223.

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. *J. Agroecol. Natural Resour. Manage.,*1(1): 1-3.

Verma, U., Piepho, H. P. and Goyal, A. (2016). Role of climatic variables and crop condition term for mustard yield prediction in Haryana. *Int. J. Agric. Stat. Sci*., 12:45–51.