MAUSAM, **74**, 3 (July 2023), 833-846

MAUSAM

DOI : https://doi.org/10.54302/mausam.v74i3.1923
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 633.11 : 004.032.26 : 551.509.314

Multi stage wheat yield estimation using multiple linear, neural network and penalised regression models

K. S. ARAVIND, ANANTA VASHISTH, B. DAS* and P. KRISHANAN

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

**ICAR-Central Coastal Agricultural Research Institute, Goa – 403 402, India*

(*Received 7 September 2021, Accepted 22 November 2022*)

e mail : ananta.iari@gmail.com

सार — भारत के उत्तरी भाग में लगभग 26 Mha क्षेत्र में चावल के बाद गेहूँ दूसरा सबसे अधिक खाया जाने वाला खाद्यान्न है। अधिकतम तापमान, न्यूनतम तापमान, सापेक्ष आर्द्रता, वर्षा, तेज धूप के घंटे, वाष्पीकरण आदि जैसे मौसम परिवर्ती फसल की उपज पर बहुत प्रभाव डालते हैं। मौसम आधारित फसल -पूर्व उपज का अनुमान विपणन , मूल्य निर्धारण, आयात-निर्यात, नीति निर्माण आदि का निर्णय लेने में सहायक है। गेहूं की फसल उगाने की अवधि के दौरान गेहूं की उपज और दैनिक मौसम आँकड़े 35 वर्ष से अधिक समय से हिसार, लुधियाना, अमृतसर, पटियाला और IARI, नई दिल्ली से एकत्र किए गए। गेहूँ की उपज का अनुमान 46वैं से 4वैं, 46वैं से 8वैं और 46वैं से 11वैं मानक मौसम विज्ञान सप्ताह के मौसम परिवर्ती पर विचार करके फसल की दौजी लगने, फूल लगने और दाने भरने की अवस्था में किया गया। मॉडल को परिवर्तीणबद्ध तरीके से मल्टीपल लीनियर रिग्रेशन (एसएमएलआर), प्रिंसिपल कंपोनेंट एनालिसिस इन कॉम्बिनेशन एसएमएलआर (पीसीए-एसएमएलआर), आर्टिफिशियल न्यूरल नेटवर्क (एएनएन) और प्रिंसिपल कंपोनेंट एनालिसिस (पीसीए-एएनएन) के संयोजन में कम से कम निरपेक्ष संकोचन और चयन ऑपरेटर (LASSO) और इलास्टिक नेट (ENET) तकनीक का उपयोग करके विकसित किया गया था। अंशांकन के लिए 70% डेटा और सत्यापन के लिए शेष डेटासेट तय करके विश्लेषण किया गया। परिणामों से पता चला है कि दौजी लगने, फूल लगने और दाने भरने के परिवर्तीण के दौरान प्रेक्षित उपज द्वारा अनुमानित उपज का प्रतिशत विचलन क्रमशः -0.1 से 25.6, 0.9 से 22.8, -0.7 से 22.5% के बीच था। प्रतिशत विचलन और मॉडल सटीकता के आधार पर इलास्टिक नेट और LASSO मॉडल को बेहतर पाया गया और इसका उपयोग विभिन्न फसल विकास परिवर्तीणों में जिला स्तर पर गेहूं की फसल की उपज के अनुमान के लिए किया जा सकता है।

ABSTRACT. Wheat is the second most consumed staple food grain after rice, cultivated in nearly 26 Mha areas in the northern part of India. Weather variables like maximum temperature, minimum temperature, relative humidity, rainfall, bright sunshine hours, evaporation etc. have a great impact on crop yield. Weather based pre harvest crop yield estimation is helpful for deciding marketing, pricing, import-export, policy making etc. Wheat yield and daily weather data during wheat crop growing period were collected from Hisar, Ludhiana, Amritsar, Patiala and ICAR-ICAR-IARI, New Delhi for more than 35 years. Wheat yield estimation was done at tillering, flowering and grain filling stage of the crop by considering weather variables from 46 to 4^{th}, 46 to 8^{th} and 46 to 11^{th} standard meteorological week (SMW). Model was developed using stepwise multiple linear regression (SMLR), Principal component analysis in combination with SMLR (PCA-SMLR), Artificial Neural Network (ANN) alone and in combination with principal components analysis (PCA-ANN), 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. Results showed that percentage deviation of estimated yield by observed yield was ranged between -0.1 to 25.6, 0.9 to 22.8, -0.7 to 22.5% during tillering, flowering, and grain filling stage respectively. On the basis of percentage deviation and model accuracy Elastic net and LASSO model was found better and can be used for district level wheat crop yield estimation at different crop growth stage.

Key words – Multistage wheat yield, SMLR, PCA-SMLR, PCA-ANN, LASSO and Elastic net.

1. Introduction

Wheat yield in different geographical areas is highly related to the spatial variability of weather. Weather is dynamic, continuous and multi-dimensional, these unfavorable properties make weather estimation a formidable challenging task for the meteorologists. For proper and efficient planning and policy making, crop

TABLE 1

Data used for model calibration, validation and yield estimation for different location

Years	Hisar	Ludhiana	Amritsar	Patiala	ICAR-IARI, New Delhi
Model Calibration	1985-2008	1971-2003	1972-2003	1971-2003	1985-2008
validation	2009-2017	2004-2016	2004-2016	2004-2016	2009-2017
Yield estimation	2018	2017	2017	2017	2018

yield forecast is a vital tool, which helps to manage excess production (Dutta *et al*., 2001). In traditional methods, crop cutting experiments were widely used for crop yield estimation at different regions. The relationship between weather variables and yield of the crop can be estimated through 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 estimation for Bihar by utilizing weather data. (Garde *et al*., 2015) used multiple linear regression (MLR) technique and discriminant function analysis for estimating wheat productivity for the district of Varanasi in eastern Uttar Pradesh. He reported that stepwise 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). (Azfar *et al*., 2015) developed the model using principal component analysis of weekly data on weather variables for developing rapeseed and mustard yield forecast model for Faizabad district of UP. (Verma *et al*., 2016) used statistical modelling approaches, *viz*., multiple linear regression and principal component analysis for developing zonal weather models for district-level mustard yield estimation in Haryana. (Kumari *et al*., 2016) evaluated the performance of artificial neural network (ANN) by comparing it with multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) Model for forecasting yield of pigeon pea for Varanasi region of Uttar Pradesh using 27 years of data (1985-86 to 2011-12). The performance of the model was assessed by root mean squared error (RMSE). As compared to both linear model, ANN was found to be best suitable model having lowest RMSE with forecasted yield during the year 2012-13 for Varanasi region. (Emamgholizadeh *et al*., 2015) used two methods namely arficial neural network (ANN) and multiple regression model (MLR) for estimating the seed yield of sesame from readily measurable plant characters. (Das *et al*., 2018) used long-term weather data and six different statistical methods for determination of rice yield estimation. Based on Friedman test overall ranking he reported that LASSO (2.63) and Elastic Net (3.07) were the best model. In present study model was developed using SMLR, PCA-SMLR, ANN, PCA-ANN, LASSO and Elastic Net technique for improving the accuracy of multi stage wheat yield estimation for different district of north India.

2. Materials and method

Wheat yield data as well as weather data during wheat crop growing period were collected from Hisar (1985-2018), Ludhiana (1971-2017), Amritsar (1972-2017), Patiala (1971-2017) and ICAR-IARI, New Delhi (1985-2018). Maximum temperature, minimum temperature, rainfall, morning and evening relative humidity, sunshine hours were arranged for three different stages, *viz*., tillering (46^{th} to 4^{th} SMW), flowering (46^{th} to 8^{th} SMW) and grain filling (46^{th} to 11^{th} SMW) for each station separately and analysis was done after converting daily weather data into weighted and unweighted composite weather data. 70% of data were used for model calibration and remaining 30% were used for the validation of models.Data used for model calibration, validation and yield estimation for different districts are given in Table 1. Simple and weighted weather indices were developed for each district during different stage under consideration. 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:

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

TABLE 2

Simple and weighted weather indices used for developing model

	Simple weather indices							Weighted weather indices						
	Tmax	Tmin	RF	RH I	RH II	SSH	EVP	Tmax	Tmin	RF	RH I	RH II	SSH	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		
SSH	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

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

Weighted weather indices:

$$Zij = \sum_{w=1}^{m} r^{j} iwXiw$$

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

where,

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

m = week at which forecast done.

P = number of variables

Combination of weather variables for weather indices, generated are presented in Table 2.

For developing crop yield forecast model at different crop growing stage following six techniques were used.

2.1. *Stepwise multiple linear regression (SMLR)*

Weather indices developed by maximum and minimum temperature, rainfall, morning and evening relative humidity, bright sunshine hour was 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. SMLR used for pre-harvest wheat crop yield estimation because of its more 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).

2.2. *Stepwise multiple linear regression-principal component analysis (PCA-SMLR)*

It is a combination of feature selection and selection method used for the data analysis. Principle 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 multicollinearity problems, transforms original set of correlated variables in to a new set of uncorrelated variables. Principal components (PCs) were selected based on their eigenvalues; eigenvalues more than 1 conditions can able to describe more than 90 percent variability in the data. PCA scores were used as input for SMLR analysis.

2.3. *Artificial neural network (ANN)*

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,

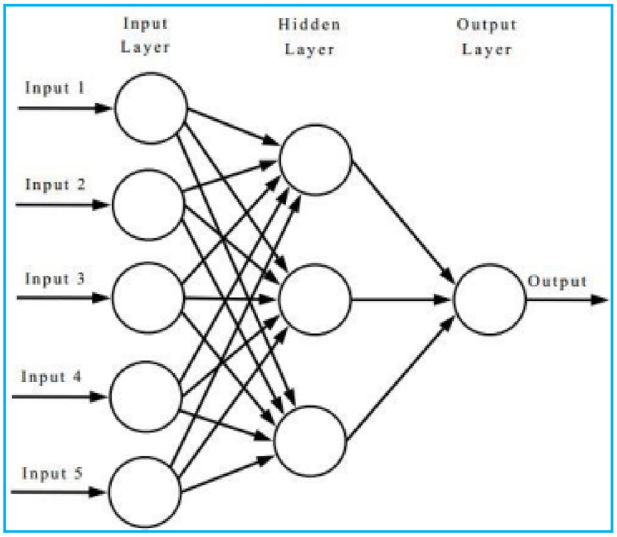

Fig. 1. Schematic representation of the ANN model

hidden layer and output layer. Multilayer perception (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. Schematic representation of the ANN model is show in Fig. 1.

2.4. *Principal component analysis - artificial neural network (PCA-ANN)*

In this technique 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.

2.5. *Least absolute shrinkage and selection operator (LASSO)*

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, regressors are either retained or eliminated from the model in order to provide better interpretable model.

2.6. *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.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 Elastic Net.

2.7. *Model accuracy*

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

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

where, $\sum (y_i - \hat{y})^2$ = sum squared regression error, $\sum (y_i - \bar{y})^2$ =sum squared total error.

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

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

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

where, *Pi*, O_i, *N* and M are predicted value, observed value, number of observations and mean of observed value.

Percentage Deviation= (*Pi*-*Oi*) *100/ *Oi*

where *Pi* and *Oi* are predicted and observed yield.

TABLE 3

Performance of wheat yield estimation using Stepwise Multiple Linear Regression (SMLR) model at different growth stage for different districts

Name of the District	Equation	Model Performance during calibration				Model Performance during validation		
		R^2	RMSEc (kg/ha)	*n*RMSEc (%)	MSEc (kg/ha)	RMSEv (kg/ha)	*n*RMSEv (%)	MSEv (kg/ha)
At Tillering stage								
Hisar	y=4934.37+58.71×time-46.93×Z271-31.01×Z671-0.018×Z450	0.87	167.2	4.38	27946	513.7	11.67	263867
Ludhiana	y=572.57+64.4×time+0.58×Z141-0.38×Z360	0.93	166.7	4.31	27806	703.8	14.76	495349
Amritsar	y=1711.25+79.13×time+1.25×Z131×4.09×Z121	0.93	209.2	6.45	43777	713.4	16.14	508968
Patiala	y=981.37+82.87×time-2.11×Z231-0.74×Z141	0.95	160.7	4.78	25824	562.3	12.09	316170
ICAR-IARI, New Delhi	y=2762.32+47.70×time+2.46×Z231	0.9	111.1	3.27	12330	172.2	4.08	29656
At Flowering stage								
Hisar	y=4890.15+56.75×time-108.82×Z11	0.75	209.7	5.49	43987	379.12	8.6	143732
Ludhiana	y=1568.1+66.08×time+0.88×Z151+0.48×Z141	0.95	142.3	3.68	20244	573.92	12.1	329384
Amritsar	y=1958.43+79.98×time+0.86×Z131	0.89	241.2	7.43	58202	699.85	15.8	489790
Patiala	y=685.15+82.68×time-0.58×Z141-1.13×Z231	0.96	157.6	4.69	24825	589.16	12.7	347110
ICAR-IARI, New Delhi	y=2656.43+46.19×time+5.19×Z231 +0.64×Z251-36×Z31	0.94	81.6	2.4	6660	201.65	4.8	40663
At grain filling stage								
Hisar	y=3801.07+48.44×time-7.72 ×Z271-6.61×Z561-47.67×Z41	0.9	143.4	3.76	20572	313.9	7.13	98521
Ludhiana	y=1758.2+65.52×time+0.73×Z151+0.49×Z141	0.95	137.8	3.57	19000	553.1	11.6	305931
Amritsar	y=1751.46+74.29×time+0.76×Z131+2.42×Z121	0.94	181.6	5.59	32997	610.9	13.76	372930
Patiala	y=265.6+78.92×time-0.55×Z141-0.99×Z231	0.96	148.3	4.41	21999	570.1	12.26	324957
ICAR-IARI, New Delhi	y=2588.91+39.75×time+0.71×Z251+0.96×Z231	0.92	98.2	2.89	9639	263.8	6.25	69580

3. Results

3.1. *Performance of wheat yield estimation using stepwise multiple linear regression model (SMLR) at different growth stage for different districts of north India*

Model performance developed by SMLR techniques for wheat yield estimation at tillering stage are presented in Table 3. During validation value of nRMSE ranged between 4.08 to 16.14%. The performance of model developed by SMLR techniques was excellent for ICAR-IARI, New Delhi with *n*RMSE value 4.08 %, good for Hisar, Ludhiana, Amritsar and Patiala with *n*RMSE value 11.67, 14.76, 16.14 and 12.09 % respectively. The percentage deviation of estimated yield done at tillering stage by observed yield was lowest for Patiala (-0.12 %) followed by ICAR-IARI, New Delhi (3.09 %), Hisar (7.17 %), Ludhiana (14.11 %) and Amritsar (19.71 %) respectively.

Model performance developed by SMLR techniques for wheat yield estimation at flowering stage are shown in Table 3. Value of *n*RMSE during validation ranged between 4.77 to 15.77 %. Model performance was excellent for ICAR-IARI, New Delhi and Hisar with *n*RMSE value 4.77 and 8.62 % respectively, good for Ludhiana, Amritsar and Patiala having *n*RMSE value 12.04, 15.77 and 12.67 % respectively. The percentage deviation of estimated yield done at flowering stage by

TABLE 4

Performance of wheat yield estimation using Principal Component Analysis- Stepwise Multiple Linear Regression (PCA-SMLR) model at different growth stage for northern part of India

Name of the District	Equation	No. of PC's	Model Performance during calibration				Model Performance during validation		
			R^2	RMSEc (kg/ha)	*n*RMSEc (%)	MSEc (kg/ha)	RMSEv (kg/ha)	*n*RMSEv (%)	MSEv (kg/ha)
At tillering stage									
Hisar	y=3124.08+62.94×time	9	0.67	264.2	6.92	69823	527.9	12	278731
Ludhiana	y=2740.64+68.89×time+117.37×PC6-77.45×PC3	7	0.91	190.8	5.04	36412	807.1	16.93	651378
Amritsar	y=2007.58+79.60×time+107.12×PC1	7	0.91	236.2	7.28	55814	683.2	15.39	466817
Patiala	y=2230.35+79.82×time-109.07×PC4-135.22×PC5	6	0.95	162.1	4.83	26283	532.6	11.45	283673
ICAR-IARI, New Delhi	y=2872.8+42.43×time+166.1×PC2+84.18×PC4	6	0.89	117.2	3.45	13733	195.0	4.62	38021
At flowering stage									
Hisar	y=3124.08+62.94×time	9	0.67	264.2	69823	6.92	527.9	12	278731
Ludhiana	y=2823.12+63.48×time+119.84×PC5+98.98×PC4 +96.04×PC6-66.69×PC2	6	0.95	146.1	21354	3.78	646.5	13.56	417911
Amritsar	y=1974.36+82.03×time	7	0.87	264.8	70103	8.16	757.4	17.06	573670
Patiala	y=2194.93+81.94×time-97.86×PC4-129.17×PC5	6	0.94	171.4	29378	5.1	562.4	12.02	316350
ICAR-IARI, New Delhi	y=2790.29+48.52×time+187.13×PC2 +119.51×PC5+55.66×PC6	6	0.89	103.2	10658	3.04	182.7	4.33	33368
At grain filling stage									
Hisar	y=3369.38+49.08×time-220.64×PC1	8	0.78	217.2	5.69	47158	586.3	13.32	343736
Ludhiana	y=2819.15+63.42×time+138.17×PC5 +104.87×PC3+87.17×PC6	6	0.94	172.2	4.45	29656	628.7	13.18	395276
Amritsar	y=2171.68+70.55×time+157.59×PC2+104.61×PC4	6	0.93	202.5	6.23	40994	532.0	11.99	282981
Patiala	y=2223.31+82.21×time-94.96×PC3 -151.84×PC6+84.83×PC4	6	0.96	152.7	4.55	23314	669.1	14.39	447735
ICAR-IARI, New Delhi	y=2924.12+40.28×time+146.83×PC4 +91.88×PC1+144.38×PC3	6	0.92	95.3	2.81	9080	184.1	4.36	33878

observed yield was lowest for Patiala (-1.79 %) followed by ICAR-IARI, New Delhi (4.46 %), Hisar (7.73 %), Ludhiana (8.11 %) and Amritsar (20.56 %). Negative sign here indicates under estimation of yield.

Model performance developed by SMLR techniques for wheat yield estimation at grain filling stage are shown in Table 3. Value of *n*RMSE during validation ranged between 6.25 to 13.76 %. The performance of SMLR model was excellent for ICAR-IARI, New Delhi and Hisar with *n*RMES value 6.25 and 7.13 % respectively, good for Ludhiana, Amritsar and Patiala with *n*RMSE value 11.6, 13.76 and 12.26 % respectively. The percentage deviation of yield estimation at grain filling stage by observed yield was lowest for ICAR-IARI, New Delhi (-2.37 %) followed by Patiala (-3.09 %), Hisar (6.51 %), Ludhiana (7.21 %) and Amritsar (13.27 %).

3.2. *Performance of wheat yield estimation using principal component analysis-stepwise multi linear regression (PCA-SMLR) model at different growth stage for different districts of north India*

Performance of wheat yield estimation using PCA-SMLR model at different growth stage for different districts of north India are shown in Table 4. PCA feature extraction method followed by SMLR is used for developing wheat yield estimation model. Principal components (PCs) were selected based on their

TABLE 5

Performance of wheat yield estimation using Artificial Neural Network (ANN) model at different growth stage for northern part of India

Name of the District	No. of hidden neurons	Model Performance during calibration				Model Performance during validation		
		R^2	RMSEc(kg/ha)	*n*RMSEc(%)	MSEc(kg/ha)	RMSEv(kg/ha)	*n*RMSEv(%)	MSEv(kg/ha)
At tillering stage								
Hisar	6	0.81	216.5	5.68	46885	623.9	13.93	389276
Ludhiana	7	0.84	317.0	8.1	100470	655.1	13.67	429117
Amritsar	3	0.90	314.2	9.34	98690	606.6	13.75	368012
Patiala	4	0.96	200.5	5.81	40180	485.2	10.45	235429
ICAR-IARI, New Delhi	9	0.87	144.5	4.20	20872	495.8	11.63	245847
At flowering stage								
Hisar	8	0.87	171.5	4.5	29405	588.4	13.15	346262
Ludhiana	9	0.75	404.2	10.33	163345	937.1	19.56	878081
Amritsar	7	0.81	395.7	11.76	156555	498.5	11.30	248532
Patiala	9	0.95	209.9	6.09	44041	483.7	10.41	233937
ICAR-IARI, New Delhi	6	0.89	122.9	3.57	15117	460.9	10.81	212456
At grain filling stage								
Hisar	10	0.91	133.2	17729	3.49	409.5	9.15	167723
Ludhiana	12	0.59	430.1	184926	11.00	598.8	12.50	358621
Amritsar	9	0.86	377.7	142642	11.23	920.8	20.88	847891
Patiala	9	0.93	229.6	52712	6.66	446.4	9.61	199273
ICAR-IARI, New Delhi	8	0.88	138.7	19249	4.03	659.8	15.48	435310

eigenvalues; eigenvalues more than 1 conditions can able to describe more than 90 percent variability in the data of selected districts. Number of PCs retained in the model was ranged between 6 and 9. PCA factors along with time as a variable were taken as input variables. The performance of PCA-SMLR model for wheat yield estimation at tillering stage was excellent for ICAR-IARI, New Delhi with *n*RMSE value 4.62 %, good for Hisar, Ludhiana, Amritsar and Patiala with *n*RMSE value 12.0, 16.93, 15.39 and 11.45 % respectively. Percentage deviation of yield estimation at tillering stage by observed yield was lowest for Patiala (-1.81 %) followed by ICAR-IARI, New Delhi (-2.19 %), Hisar (7.84 %), Amritsar (21.63 %) and Ludhiana (25.67 %).

Performance of the model developed by PCA-SMLR techniques for wheat yield estimation at flowering stage are presented in Table 4. Number of PCs retained in the model was ranged between 6 and 9. Time was the most important variable affecting the crop yield followed by PC2, PC4, PC5 and PC6. The most influential weather parameter identified using PCA-SMLR was PC2, PC4, PC5 and PC6 for Ludhiana, PC2, PC5 and PC6 for ICAR-IARI, New Delhi, for Patiala PC4 and PC5 respectively. Value of *n*RMSE during validation ranged between 4.33 to 17.06 %. The performance of PCA- SMLR model for yield estimation as flowering stage was excellent for ICAR-IARI, Delhi having *n*RMSE value 4.33 %, good for Hisar, Patiala, Ludhiana and Amritsar with *n*RMSE value 12.0, 12.02, 13.56 and 17.06 % respectively. The percentage deviation of wheat yield estimation at flowering stage by observed yield was lowest for ICAR-IARI, New Delhi (0.91 %) followed by Patiala (-4.58 %), Hisar (7.84 %), Ludhiana (22.7 %) and Amritsar (22.86 %).

Performance of the model for wheat yield estimation at grain filling stage is shown in Table 4. Model retained 6 to 8 PC's. Time followed by PC1 to PC6 was the most important variable affecting the crop yield. PC1 was important variable for Hisar, PC3, PC5 and PC6 for Ludhiana, PC2 and PC4 for Amritsar, PC3, PC4 and PC6

TABLE 6

Performance of wheat yield estimation using Principal Component Analysis -Artificial Neural Network (PCA-ANN) model at different growth stage for northern part of India

Name of the District	No. of hidden neurons	No. of PC's	Model Performance during calibration				Model Performance during validation		
			R^2	RMSEc(kg/ha)	*n*RMSEc(%)	MSEc(kg/ha)	RMSEv(kg/ha)	*n*RMSEv(%)	MSEv(kg/ha)
At tillering stage									
Hisar	1	9	0.83	187.6	4.92	35213	601.0	13.42	361153
Ludhiana	1	7	0.88	262.2	6.7	68744	487.4	10.17	237559
Amritsar	2	7	0.88	361.4	10.75	130624	663.4	15.04	440139
Patiala	2	6	0.95	211.3	6.13	44656	648.0	13.95	419891
ICAR-IARI, New Delhi	2	6	0.86	142.6	4.14	20340	537.2	12.6	288616
At flowering stage									
Hisar	2	9	0.91	144.4	3.79	20863	536.7	11.99	288025
Ludhiana	3	6	0.89	278.3	7.11	77423	441.5	9.21	194940
Amritsar	1	7	0.87	346.5	10.3	120041	626.4	14.2	392402
Patiala	1	6	0.92	233.2	6.79	54378	625.9	13.48	391713
ICAR-IARI, New Delhi	2	6	0.76	188.1	5.46	35378	640.8	15.03	410573
At grain filling stage									
Hisar	2	8	0.96	82.0	2.15	6722	493.0	11.01	243059
Ludhiana	3	6	0.85	291.1	7.45	84751	424.8	8.86	180481
Amritsar	2	6	0.89	352.7	10.49	124418	664.3	15.06	441308
Patiala	3	6	0.96	177.3	5.14	31435	515.8	11.11	266039
ICAR-IARI, New Delhi	2	6	0.88	131.5	3.82	17284	578.6	13.57	334743

for Patiala, PC1, PC3 and PC4 for ICAR-IARI, New Delhi, respectively and establish most influence on wheat yield. Value of *n*RMSE during validation was ranged between 4.36 to 14.39 %. The performance of PCA-SMLR model was excellent for ICAR-IARI, New Delhi having *n*RMSE value 4.36 %, good for Amritsar (11.99 %), Ludhiana (13.18 %), Hisar (13.32 %) and Patiala (14.39%) respectively. Percentage deviation of estimated yield at grain filling stage by observed yield was lowest for ICAR-IARI, New Delhi (-2.29 %) followed by Hisar (4.97 %), Patiala (-5.02 %), Amritsar (17.37 %) and Ludhiana (22.55%) respectively.

3.3. *Performance of wheat yield estimation using artificial neural network (ANN) model at different growth stage for different districts of north India*

Performance of wheat yield estimation using ANN model at different growth stage for different districts of north India are shown in Table 5. For development of ANN model, the *Z* variates were taken as inputs. The value of coefficient of determination R^2 for models developed for estimating yield at tillering stage was between 0.81 to 0.96. The optimum number of hidden neurons varied between 3 and 9. The RMSE during calibration was lowest for ICAR-IARI, New Delhi (144.4 kg/ha) followed by Patiala (200.5 kg/ha), Hisar (216.5 kg/ha), Amritsar (314.2 kg/ha) and Ludhiana (316.9 kg/ha). Performance of ANN model for yield estimation at tillering stage was good with *n*RMSE value 13.93, 13.67, 13.75, 10.45 and 11.63 % for Hisar, Ludhiana, Amritsar, Patiala and ICAR-IARI, New Delhi respectively. Percentage deviation of yield estimation done at tillering stage by observed yield was lowest for ICAR-IARI, New Delhi (4.65 %) followed by Ludhiana (-10.98 %), Hisar (-13.95 %), Patiala (-16.41 %) and Amritsar (-18.78 %).

Performance of ANN model for yield estimation done at flowering stage was good having value of *n*RMSE< 20 % for all the districts. The optimum number of hidden neurons varied between 6 and 9. Percentage deviation of yield estimation done at flowering stage by observed yield was lowest for ICAR-IARI, New Delhi (-1.26 %) followed by Hisar (-12.66 %), Ludhiana (-13.01 %), Patiala (-16.60 %) and Amritsar (-19.58 %).

TABLE 7

Performance of wheat yield estimation using Least Absolute Shrinkage and Selection Operator (LASSO) model at different growth stage for northern part of India

Name of the District	Equation	Model Performance during calibration				Model Performance during validation		
		R^2	RMSEc (kg/ha)	*n*RMSEc (%)	MSEc (kg/ha)	RMSEv (kg/ha)	*n*RMSEv (%)	MSEv (kg/ha)
At tillering stage								
Hisar	y=3567.03+38.95×time-0.12×Z151-3.56×Z261-23.15×Z271-8.12×Z671	0.82	222.2	5.82	49346	298.3	6.78	88965
Ludhiana	y=1413.17+56.24×time+0.35×Z141+0.38×Z151-0.014×Z360+0.48×Z361	0.93	192.9	4.99	37207	493.0	10.34	243010
Amritsar	y=2103.62+71.72×time+1.67×Z121+0.83×Z131+0.24×Z141	0.92	224.4	6.91	50360	546.9	12.32	299143
Patiala	y=279.11+77.25×time-6.98×Z41-0.23×Z141-1.59×Z231	0.96	166.6	4.89	27766	488.6	10.54	238759
ICAR-IARI, New Delhi	y=2827.93+43.69×time+2.91×Z31+0.11×Z121+0.0004 ×Z131+1.54×Z231+0.33×Z251	0.91	107.6	3.18	11582	174.6	4.13	30475
At flowering stage								
Hisar	y=4631.35+33.91×time-9.14×Z11-8.2×Z71-3.24×Z171-6.92×Z261-0.61×Z471-3.96×Z671	0.86	193.3	5.06	37353	291.8	6.63	85136
Ludhiana	y=1303.06+59.81×time+7.41×Z41+0.28×Z141 +0.47×Z151+0.81×Z361	0.95	151.2	3.91	22871	498.1	10.45	248054
Amritsar	y=2423.41+70.85×time+2.68×Z41+0.60×Z121 +0.29×Z131+0.15×Z141+0.28×Z151	0.9	242.1	7.46	58622	527.1	11.87	277782
Patiala	y=799.39+81.19×time-6.87×Z11-1.02×Z41-0.45×Z141+0.013×Z230-0.91×Z231	0.96	158.0	4.63	24961	578.8	12.48	334975
ICAR-IARI, New Delhi	y=2608.26+40.3×time+0.55×Z41 +1.16×Z231+0.60×Z251	0.93	97.5	2.87	9512	220.7	5.22	48708
At grain filling stage								
Hisar	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	4.26	26553	289.8	6.58	83972
Ludhiana	y=1156.61+61.25×time+5.93×Z41+0.007×Z140+0.35×Z141 +0.48×Z151+0.65×Z361	0.95	136.7	3.54	18698	510.3	10.7	260416
Amritsar	y=1967.35+69.26×time+1.15×Z31+0.55×Z131 +0.15×Z151	0.94	198.4	6.11	39367	511.6	11.53	261724
Patiala	y=626.19+78.5×time-1.94×Z11-2.31×Z21-0.005×Z41-0.45×Z141-0.78×Z231+0.004×Z240	0.96	151.7	4.45	23022	567.7	12.24	322272
ICAR-IARI, New Delhi	y=2489.89+36.6×time+2.97×Z11+0.63×Z41 +0.62×Z231+0.63×Z251	0.92	102.8	3.03	10574	263.0	6.23	69174

ANN Model performance for yield estimation during grain filling stage was excellent for Hisar and Patiala with *n*RMSE value 9.15 and 9.61 % respectively, good for Ludhiana and ICAR-IARI with *n*RMSE value 12.5 and 15.48 respectively, fair for Amritsar with *n*RMSE value 20.88 %. Percentage deviation of yield estimation done by ANN model at grain filling stage by observed yield was lowest for ICAR-IARI, New Delhi (3.79 %)followed by Ludhiana (-10.17 %), Hisar (-11.12%), Patiala (-16.68 %) and Amritsar (-20.20 %).

3.4. *Performance of wheat yield estimation using principal component analysis -artificial neural network (PCA-ANN) model at different growth stage for different districts of north India*

Performance of wheat yield estimation done using PCA-ANN model at different growth stage for different districts of north India are shown in Table 6. Based on the applied *Z* variates, PCA factors were generated. Time along with PCA factors were considered for developing

PCA-ANN models. R^2 for developed models during calibration for estimating yield at tillering stage was between 0.83 to 0.95. The number of PCs found between 6 and 9. The optimum number of hidden neurons varied between 1 and 2. During validation value of *n*RMSE varied between 10.17 to 15.04 %. Performance of PCA-ANN model for yield estimation at tillering stage was good having *n*RMSE value <20 % for all the district of north India. Percentage deviation of yield estimation done at tillering stage by observed yield was lowest for ICAR-IARI, New Delhi (4.94 %) followed by Hisar (7.84 %), Ludhiana (-9.35 %), Patiala (-16.92 %) and Amritsar (-20.09 %) respectively.

During validation performance of PCA-ANN model for yield estimation at flowering stage was excellent for Ludhiana with *n*RMSE value 9.21 % and good for Hisar, Patiala, Amritsar and ICAR-IARI, New Delhi with *n*RMSE value 11.99, 13.48, 14.2 and 15.03 % respectively. Percentage deviation of yield estimation done at flowering stage by observed yield was lowest for ICAR-IARI, New Delhi (-6.98 %) followed by Ludhiana (-10.35 %), Hisar(-11.03 %), Patiala (-15.04 %) and Amritsar (-19.22 %) respectively.

The performance of PCA-ANN model for yield estimation at grain filling stage was found excellent for Ludhiana having *n*RMSE value 8.86 % and good for Hisar, Patiala, ICAR-IARI, New Delhi and Amritsar having *n*RMSE value 11.01, 11.11, 13.57 and 15.06 % respectively.Validation of dataset at grain filling stage showed RMSE value ranged between 424.8 to 664.3 kg/ha, and *n*RMSE value between 8.86 to 15.06 %. The percentage deviation of yield estimation done at grain filling stage by observed yield was lowest for ICAR-IARI, New Delhi (6.66 %) followed by Ludhiana (-10.05 %), Hisar(-10.41%), Patiala (-15.49 %) and Amritsar (-20.09 %) respectively.

3.5. *Performance of wheat yield estimation using least absolute shrinkage and selection operator (LASSO) model at different growth stage for different districts of north India*

Performance of wheat yield estimation using LASSO model at different growth stage for different districts of north India are shown in Table 7. LASSO technique was used to minimize residual mean square error. The value of coefficient of determination R^2 during calibrationfor models developed for wheat yield estimation at tillering stage was ranged between 0.82 to 0.96. The *Z* variates were having positive influence on yield using LASSO except Z151, Z261, Z271 and Z671 for Hisar, Z360 for Ludhiana, Z41, Z141 and Z231 for Patiala. RMSE during validation was ranged between 174.6 to 546.9 kg/ha. Performance of LASSO model was excellent for ICAR-IARI, New Delhi and Hisar with *n*RMSE value 4.13 and 6.78 % respectively, good for Ludhiana, Amritsar and Patiala with *n*RMSE value 10.34, 12.32 and 10.54 % respectively. Percentage deviation of estimated yield done by LASSO model at tillering stage by observed yield was lowest for Patiala (-2.38 %) followed by Hisar (-2.85 %), ICAR-IARI, New Delhi (4.02 %), Ludhiana (10.65 %), and Amritsar (14.81 %) respectively.

Performance of LASSO model for yield estimation at flowering stage was excellent for ICAR-IARI, New Delhi and Hisar having *n*RMSE value 5.22 and 6.63 % respectively, good for Ludhiana, Amritsar and Patiala having *n*RMSE value 10.45, 11.87 and 12.48 % respectively. The value of coefficient of determination R^2 during calibration for models developed was ranged between 0.86 to 0.96. Percentage deviation of estimated yield done by LASSO model at flowering stage by observed yield was lowest for Hisar (-1.64 %) followed by Patiala (-2.06 %), ICAR-IARI, New Delhi (9.07 %), Ludhiana (9.78 %), and Amritsar (12.09 %) respectively.

Performance of LASSO model for wheat yield estimation at grain filling stage was excellent <10 % for ICAR-IARI, New Delhi and Hisar with *n*RMSE value 6.23 and 6.58 % respectively, good for Ludhiana, Amritsar and Patiala with *n*RMSE value 10.7, 11.53 and 12.24 % respectively. Percentage deviation of yield estimation done by LASSO model at grain filling stage by observed yield was lowest for Patiala (-3.06%) followed by ICAR-IARI, New Delhi (3.99 %), Hisar (-6.47 %), Ludhiana (6.78 %) and Amritsar (10.84 %) respectively.

3.6. *Performance of wheat yield estimation using Elastic Net model at different growth stage for different districts of north India*

Performance of wheat yield estimation using Elastic Net model at different growth stage for different districts of north India are shown in Table 8. Value of coefficient of determinant R^2 for models developed at tillering stage during calibration was between 0.85 to 0.96. The mean square error and *n*RMSE for validation was ranged between 42181 kg/ha to 247297 kg/ha and 4.86 to 10.97 % respectively. Performance of Elastic Net model for yield estimation done at tillering stage was excellent for ICAR-IARI, New Delhi, Hisar and Patiala with *n*RMSE value of 4.86, 6.67 and 8.85 % respectively, good for Ludhiana and Amritsar with *n*RMSE value of 10.43 and 10.97 % respectively. Percentage deviation of yield estimation done by Elastic Net at tillering stage by observed yield was lowest for ICAR-IARI, New Delhi (3.57 %) followed by Ludhiana (5.48 %), Amritsar (5.81 %), Hisar (-9.98%) and Patiala (12.15 %) respectively.

TABLE 8

Performance of wheat yield estimation using Elastic Net model at different growth stage for northern part of India

Name of the District	Equation	Model Performance during calibration				Model Performance during validation		
		R^2	RMSEc (kg/ha)	*n*RMSEc (%)	MSEc (kg/ha)	RMSEv (kg/ha)	*n*RMSEv (%)	MSEv (kg/ha)
At tillering stage								
Hisar	y=3773.58+21.54×time-0.03×Z151-3.85 ×Z261-0.21×Z670-4.47×Z671	0.85	197.3	5.16	38908	293.5	6.67	86136
Ludhiana	y=48.27+45.78×time+1.67×Z11+14.53×Z41+0.14×Z121 +0.24×Z141+0.45×Z151-0.019×Z360+0.58×Z361	0.94	163.4	4.22	26690	497.3	10.43	247297
Amritsar	y=2458.48+57.28×time+0.75×Z11+2.66×Z31 +0.30×Z131+0.34×Z141+0.25×Z151	0.92	216.3	6.66	46790	487.1	10.97	237237
Patiala	y=-581.95+61.41×time-0.71×Z31+0.81×Z40-10.57 ×Z41-0.91×Z231-0.02×Z341-0.004×Z351	0.96	167.9	4.93	28200	410.6	8.85	168584
ICAR-IARI, New Delhi	y=2882.08+34.95×time+3.9×Z31+2.08×Z51+0.1×Z131 +0.29×Z231+0.22×Z251+0.03×Z341+0.01×Z351	0.91	114.0	3.36	12998	205.4	4.86	42181
At flowering stage								
Hisar	y=4496.69+19.12×time-31.56×Z71-1.74×Z171-0.12×Z260-2.66×Z261-0.23×Z270-0.35×Z471-4.09×Z671	0.86	189.2	4.96	35804	293.2	6.66	85949
Ludhiana	y=-72.71+44.42×time+1.23×Z11+0.23×Z31+19.59×Z41 +0.09×Z121+0.013×Z140+0.12×Z141+0.04×Z151+0.32×Z361	0.96	137.9	3.56	19008	502.3	10.53	252275
Amritsar	y=2107.26+54.41×time+3.56×Z11+0.18×Z141+0.40×Z151	0.91	235.4	7.25	55432	446.8	10.06	199621
Patiala	y=274.62+67.02×time-22.2×Z11-5.24×Z41+0.01×Z140-0.21×Z141-0.4×Z231-0.016×Z341-0.016×Z351	0.96	159.8	4.69	25549	498.6	10.75	248562
ICAR-IARI, New Delhi	y=2708+32.98×time+1.78×Z51+0.66×Z231+0.45×Z251 +0.008×Z351+0.006×Z451	0.93	99.3	2.92	9851	233.8	5.54	54662
At grain filling stage								
Hisar	y=4670.21+25.20×time-20.83×Z71-0.04×Z151-0.10 ×Z260-0.55×Z261-0.04×Z270-0.01×Z451-0.73 ×Z471-0.4×Z561-0.44×Z671	0.89	168.3	4.41	28338	220.4	5.01	48567
Ludhiana	y=138.20+52.49×time+1.75×Z11+11.03×Z41+0.05×Z121+0.022 ×Z140+0.25×Z141+0.23×Z151+0.043×Z161+0.62×Z361	0.97	113.0	2.92	12769	565.0	11.85	319259
Amritsar	y=2143.59+56.86×time+0.60×Z121+0.44×Z151+0.39Z231	0.94	193.1	5.95	37272	452.6	10.19	204856
Patiala	y=2700.12+28.8×time+1.21×Z51+0.34×Z231+0.48×Z251 +0.0008×Z341 +0.016×Z351+0.0027×Z451	0.92	107.5	3.17	11546	287.7	6.81	82794
ICAR-IARI, New Delhi	y=-222.55+65.46×time-1.77×Z31-3.43×Z41+0.012×Z140-0.28×Z141-0.33×Z231-0.024×Z341-0.009×Z351	0.96	149.0	4.37	22204	477.6	10.3	228111

During calibration model developed at flowering stage had R^2 value between 0.86 to 0.96. The Root mean square error during validation was ranged between 233.8 to 502.3 kg/ha. Performance of Elastic Net model for yield estimation at flowering stage was found excellent for ICAR-IARI, Delhi and Hisar with *n*RMSE value 5.54 and 6.66 %, good for Amritsar, Ludhiana and Patiala with *n*RMSE value 10.06, 10.53 and 10.75 % respectively. Percentage deviation of yield estimation done by Elastic Net at flowering stage by observed yield was lowest for ICAR-IARI, New Delhi (1.63 %) followed by Amritsar (3.90 %), Hisar (-6.32%), Ludhiana (6.38 %) and Patiala (-7.70 %) respectively.

At grain filling stage coefficient of determination R^2 during calibration was between 0.89 to 0.97. Performance of Elastic Net model for wheat yield estimation at grain filling stage was excellent for Hisar and Patiala with *n*RMSE value 5.01 and 6.81 % respectively, good for Amritsar, ICAR-IARI, New Delhi and Ludhiana with

TABLE 9

Percentage deviation of estimated yield by observed yield for different district using different models at different growth stage

District	SMLR	PCA-SMLR	ANN	PCA-ANN	LASSO	Elastic Net
At Tillering stage						
Hisar	7.17	7.84	-13.95	7.84	-2.85	-9.98
Ludhiana	14.11	25.67	-10.98	-9.35	10.65	5.48
Amritsar	19.71	21.63	-18.78	-20.09	14.80	5.81
Patiala	-0.12	-1.81	-16.41	-16.92	-2.38	12.15
ICAR-IARI, New Delhi	3.09	-2.19	4.65	4.94	4.01	3.57
At Flowering stage						
Hisar	7.73	7.84	-12.66	-11.03	-1.64	-6.32
Ludhiana	8.12	22.76	-13.01	-10.35	9.78	6.38
Amritsar	20.56	22.86	-19.58	-19.22	12.99	3.90
Patiala	-1.79	-4.58	-16.60	-15.04	-2.06	-7.70
ICAR-IARI, New Delhi	4.46	0.91	-1.26	-6.98	9.07	1.63
At Grain filling stage						
Hisar	6.51	4.97	-11.12	-10.41	-6.47	6.29
Ludhiana	7.21	22.55	-10.17	-10.05	6.78	4.93
Amritsar	13.27	17.37	-20.20	-20.09	10.84	5.05
Patiala	-3.09	-5.02	-16.68	-15.49	-3.06	-2.09
ICAR-IARI, New Delhi	-2.37	-2.29	3.79	6.66	3.99	-0.74

*n*RMSE value 10.19, 10.30 and 11.85 respectively. Percentage deviation of yield estimation done by Elastic Net model at grain filling stage by observed yield was lowest for ICAR-IARI, New Delhi (-0.74 %) followed by Patiala (-2.09 %), Ludhiana (4.93 %), Amritsar (5.05%) and Hisar (6.29 %) respectively.

3.7. *Performance of different model for different districts of north India*

Percentage deviation of estimated yield done by different model at different growth stage by observed yield is given in Table 9. At tillering stage, Elastic Net had percentage deviation < 10% for all districts except 12.15 % for Patiala. LASSO had percentage deviation < 5% for Patiala, Hisar ICAR-IARI, New Delhi and < 15 % for Ludhiana and Amritsar. SMLR model had percentage deviation < 5 % for Patiala, ICAR-IARI, New Delhi and Hisar, <15 % for Ludhiana, <20 % for Amritsar. PCA-SMLR had percentage deviation < 5% for Patiala, ICAR-IARI, New Delhi and Hisar, 21.63 % for Amritsar and 25.67 % for Ludhiana. ANN model had percentage deviation < 5 % for ICAR-IARI, New Delhi and < 20 % for Hisar, Ludhiana, Amritsar and Patiala. PCA-ANN model had percentage deviation <10% for ICAR-IARI, New Delhi, Hisar and Ludhiana, 16.41 % for Patiala, 20.09 % for Amritsar.

At flowering stage, Elastic Net had percentage deviation < 10% for all five districts, LASSO had percentage deviation < 10 % for four districts and 12.99 % for Amritsar. SMLR model had percentage deviation < 10 % for four district and 20.56 % for Amritsar. PCA-SMLR had percentage deviation < 10 % for three districts and < 25% for two districts. ANN had percentage deviation < 10 % for ICAR-IARI, New Delhi, < 15 % for Hisar, Ludhiana and <20 % for Amritsar and Patiala. PCA-ANN model had percentage deviation < 10 % for ICAR-IARI, New Delhi <15% for Hisar, Ludhiana and Patiala, <20% for Amritsar.

At grain filling stage Elastic Net had percentage deviation < 10% for all five districts, LASSO had percentage deviation < 10 % for four districts and 10.8 %

for Amritsar. SMLR had percentage deviation < 10 % for four districts and 13.27 % for Amritsar.

PCA-SMLR had percentage deviation < 10 % for Hisar, Patiala and ICAR-IARI, New Delhi, 17.4% for Amritsar and 22.6 % for Ludhiana. ANN model had percentage deviation < 10 % for ICAR-IARI, New Delhi and < 15 % for Hisar, Ludhiana and <20 % for Amritsar and Patiala. PCA-ANN model had percentage deviation < 10 % for ICAR-IARI, New Delhi <15% for Hisar, Ludhiana and <20 % for Amritsar and Patiala.

In our study wheat yield estimation done at different crop stage using different models had percentage deviation between-0.1 to 19.7% by SMLR, -1.8 to 25.7 % by PCA-SMLR, 4.7 to -18.8 % by ANN, 4.9 to -20.1 % by PCA-ANN, -2.4 to 14.8 by LASSO and 3.6 to 12.2 % by Elastic Net at tillering stage for different districts respectively. At flowering the percentage deviation was between-1.8 to 20.6 % by SMLR, 0.9 to 22.9 % by PCA-SMLR, -1.3 to -19.6 % by ANN, -6.9 to -19.2 % by PCA-ANN, -2.1 to 12.9 % by LASSO and 1.6 to -7.0 % by Elastic Net and at grain filling stage between-2.4 to 13.3 % by SMLR, -2.3 to 22.6 % by PCA-SMLR, 3.8 to -20.2 by ANN, 6.7 to -20.1 % by PCA-ANN, -3.1 to 10.8 by LASSO and -0.7 to 6.3% by Elastic Net for different districts respectively. (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. (Singh *et al*., 2014) reported that statistical models based on weather indices can successfully simulate multi-stage yield forecast of wheat at mid-season and at pre-harvest for Amritsar, Bhatinda and Ludhiana districts. This model is simple, does not require any sophisticated statistical tools, and can be used satisfactorily for district, agro-climatic zone and state level forecasting.

In our study based on percentage deviation of estimated yield by observed yield done at tillering, flowering and grain filling stage, LASSO and Elastic Net is giving better results followed by SMLR, PCA-ANN, ANN and PCA- ANN. (Kumar *et al*., 2019) used Stepwise and LASSO regression variable selection techniques 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 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. (Vashisth and Aravind, 2020) reported that Elastic Net, LASSO and SMLR model based on weather parameters can be used for multistage mustard yield estimation and Elastic Net performed best among all the threemodels followed by LASSO and SMLR model.

The developed weather-based models using different methods showed that results of multi stage in season yield forecasting are closer to observed yield at the pre-harvest stage as compared to flowering and tillering stage. Model developed using different methods using weather parameters had lower value of *n*RMSE 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). (Vasthisth *et al.,* 2014) reported that the percentage deviation of observed yield by simulated wheat crop yield forecast done at forty-five days and twenty-five days before harvest using statistical model was less than 10 %. (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. (Azfar *et al.*, 2015) reported that the model developed using principal component analysis with six weather variables (maximum and minimum temperature, Morning and evening relative humidity, wind velocity and sun shine hour) for mustard yield was found to be most appropriate for providing yield forecast one and half months before the harvest.

4. Conclusions

On evaluation of overall performance of different empirical models used for five major wheat growing districts of north India at different crop growth stages, LASSO and Elastic Net models were found to be best for multistage wheat yield estimation. On the basis of percentage deviation of estimated yield by observed yield, estimation accuracy at different growth Elastic net and LASSO model was better followed by SMLR model. From this study it may be concluded that Elastic Net and LASSO model can be used for district level wheat yield forecast at different crop growth stage.

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Agrawal, R. and Mehta, S. C., 2007, “Weather based forecasting of crop yields, pests and diseases-IASRI models”, *J. Indian Soci. Agricul. Statis.*, **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 estimation model of rice in Bulsar district of Gujarat", *J. Agrometeorol*., **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 estimation of rice yield based on weather parameters for west coast of India", *Int. J Biometeorol*., **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", *J. Indian Soci. Remote Sens.*, **29**, 3, 175.

Emamgholizadeh, S., Parsaeian, M. and Baradaran, M., 2015, "Seed yield estimation of sesame using artificial neural network", *Euro J. Agron*., **68**, 89-96.

Garde, Y. A., Dhekale, B. S. and Singh, S., 2015, "Different approaches on pre harvest forecasting of wheat yield", *J Appli. Nat 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**, 306-310.

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. *Int. J Agri. Stat. Sci*., **10**, 2, 381-385.

Singh, A. K., Vashisth, Ananta, Sehgal,V. K, Goyal, A., Pathak, H. and Parihar, S.S., 2014, "Development of Multi Stage District Level Wheat Yield Forecast Models", *J Agri. Physic.*, **14**, 2, 189-193.

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.

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. Agroecolo. Nat. Res. Manag*., 1-3.

Vashisth, Ananta, Goyal, A. and Roy, D., 2018, "Pre harvest maize crop yield forecast at different growth stage using different model under semi arid region of India", *Int. J. Tropi. Agri.*, **36**, 4, 915-920.

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. Physic.*, **20**, 2, 213-223.

Verma, U., Piepho, H. P. and Goyal, A., 2016, "Role of climatic variables and crop condition term for mustard yield estimation in Haryana", *International Journal of Agricultural and Statistical Sciences*, **12**, 45-51.

Yadav, R. R. and Sisodia, B. V. S., 2015, "Predictive models for Pigeon-pea yield using weather variables", *Int. J. Agri.Statis. Sci*., **11**, 2, 462-472.