International Journal of Biometeorology (2018) 62:1809–1822
https://doi.org/10.1007/s00484-018-1583-6

ORIGINAL PAPER

Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India

Bappa Das[1] · Bhakti Nair[1] · Viswanatha K. Reddy[1] · Paramesh Venkatesh[1]

Received: 6 April 2018 / Revised: 19 June 2018 / Accepted: 8 July 2018 / Published online: 24 July 2018

Abstract

Rice is generally grown under completely flooded condition and providing food for more than half of the world's population. Any changes in weather parameters might affect the rice productivity thereby impacting the food security of burgeoning population. So, the crop yield forecasting based on weather parameters will help farmers, policy makers and administrators to manage adversities. The present investigation examines the application of stepwise multiple linear regression (SMLR), artificial neural network (ANN) solely and in combination with principal components analysis (PCA) and penalised regression models (e.g. least absolute shrinkage and selection operator (LASSO) or elastic net (ENET)) for rice yield prediction using long-term weather data. The R^2 and root mean square error (RMSE) of the models varied between 0.22–0.98 and 24.02–607.29 kg ha^{-1}, respectively during calibration. During validation with independent dataset, the RMSE and normalised root mean square error (nRMSE) ranged between 21.35–981.89 kg ha^{-1} and 0.98–36.7%, respectively. For evaluation of multiple models for multiple locations statistically, overall average ranks on the basis of R^2 and RMSE of calibration; RMSE and nRMSE of validation were calculated and non-parametric Friedman test was applied to check the significant difference among the models. The ranking of the models revealed that LASSO (2.63) was the best performing model followed by ENET (3.07) while PCA-ANN (4.19) was the worst model which was found significant at $p < 0.001$. The reason behind good performance of LASSO and ENET is that these models prevent overfitting and reduce model complexity by penalising the magnitude of coefficients. Then, pairwise multiple comparison test was performed which indicated LASSO as the best model which was found similar to SMLR and ENET. So, for prediction of rice yield, these models can very well be utilised for west coast of India.

Keywords Stepwise multiple linear regression · Artificial neural network · Least absolute shrinkage and selection operator · Elastic net · Rice yield prediction · Weather data

Introduction

Rice is principal food grain crop of India which occupies about 43.50 m ha area with the production of 104.32 million tons (Government of India, Ministry of Agricuture and Farmers Welfare: Deparment of Agriculture Cooperation, and Welfare 2016). As rice is mainly cultivated under flooded conditions, any change in climate which leads to reduction in water availability might impact the productivity and production to great extent (Bhuvaneswari et al. 2014). From past few decades, rice production and productivity in India have shown a remarkable growth. In this regard, crop yield forecasting is essential for proper planning and policy-making to manage the excess produce (Dutta et al. 2001). There are mainly two types of approaches to forecast crop yield: crop simulation and empirical statistical models (Bocca and Rodrigues 2016). Crop simulation models are process-based and input data-intensive. Though crop simulation models are precise, hardly these models can be applied to large spatiotemporal scales due to unavailability of sufficient input data. On the other hand, empirical statistical models are simple and require less

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00484-018-1583-6) contains supplementary material, which is available to authorized users.

✉ Bappa Das
bappa.iari.1989@gmail.com

[1] Indian Council of Agricultural Research-Central Coastal Agricultural Research Institute, Goa, India

input data. So, statistical models using crop yield and weather data by means of simple regression techniques have been broadly used as a common alternative to process-based models (Lobell and Burke 2010; Shi et al. 2013). Though applicability of statistical models are limited beyond the space and time of the regression, can offer many insights about historical yield and weather interactions and can be utilised to update the other kinds of models (Lobell and Burke 2010; Lobell et al. 2011; Basso et al. 2013). For successful crop yield forecasting based on weather information, statistical models should be first calibrated and tested using historical dataset. 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. Pandey et al. (2015) determined the individual and joint effect of weather variables on rice yield of eastern Uttar Pradesh. They found that individually, sunshine (hour) is more important for yield forecasting followed by wind velocity and rainfall (with $R^2 = 67.57$, 48.63 and 46.74%, respectively). The combined effect of weather variables like rainfall and wind velocity ($R^2 = 82\%$), rainfall and sunshine hour ($R^2 = 63\%$) and wind velocity and sunshine hour ($R^2 = 53.8\%$) were also found important for crop yield modelling. Similar to this, Rai et al. (2013) also developed forecasting model for rice using multiple regression technique and reported that model developed by joint effect of weather variables had given best yield prediction results. Singh et al. (2014) developed weather-based statistical crop yield prediction model of rice and wheat for eastern Uttar Pradesh. They found that models were able to explain 51 to 79% variability for rice yield while it was 65 to 92% for wheat yield.

Most of the studies in the past have used multiple linear regressions (MLRs) to develop statistical crop yield prediction model (Rai et al. 2013; B S Dhekale 2014; Dhekale et al. 2014; Kumar et al. 2014). But MLR suffers from over-fitting when the number of samples is less than the number of predictors and multicollinearity, when the independent variables are correlated (Verma et al. 2016). To overcome these problems, feature selection (e.g. stepwise multiple linear regression (SMLR), least absolute shrinkage and selection operator (LASSO) or elastic net (ENET) method) or feature extraction (e.g. principal component analysis) statistical techniques can be used (Das et al. 2017). Though in few studies PCA has been used in conjunction with MLR (Azfar et al. 2015; Verma et al. 2016; Annu et al. 2017), comparison of the performance of feature selection, feature extraction and combination of both the methods for forecasting the crop yield is scarce. In this context, our study has found scope to develop and select a statistical forecasting model for rice using various regression techniques for west coastal region of India with the objectives (i) to develop district-wise crop yield prediction models using different multivariate models and (ii) to evaluate the predictive performance of the developed models.

Material and Method

Data collection

Time series data of rice yield (*Oryza sativa* L.) for western coastal districts of India for 33 years (1983 to 2015) has been obtained from State Department of Agriculture, Department of Agriculture, Cooperation, Ministry of Agriculture, Government of India. Location details of district selected from western coastal zone are presented in Fig. 1. West coastal regions belong to monsoon-type climate with short dry winter season (Am, except Trivandrum which belongs to class Aw) according to Koeppen's classification, experience monsoon rainfall from June to September. Due to favourable climate conditions, rice crop is generally cultivated during the *kharif* (rainy) season.

Daily weather data were collected from India Meteorological Department (IMD) for 1983 to 2015 (33 years). Solar radiation data were obtained from NASA's Prediction of Worldwide Energy Resources (NASA/POWER; power.larc.nasa.gov). Data gaps were filled using maximum likelihood estimation which operate by estimating a set of parameters that maximise the probability of getting the estimates from the sample data that is analysed, and it provides a deterministic result (Collins et al. 2001). The data on five weather variables namely maximum and minimum temperature (Tmax and Tmin, °C), relative humidity (RH, %), solar radiation (SRAD, MJ m^{-2} day^{-1}) and rainfall (mm) for 20 weeks of the crop cultivation, which includes 23rd standard meteorological week (SMW) to 43rd SMW had been used in the study. Daily data of Tmax, Tmin, RH and SRAD had been converted into its weekly average values while weekly sum of rainfall has been considered (Fig. 2). Out of the 33-year data, 29-year data were used for model calibration while remaining 4 years data were used for model validation.

Weather indices calculation

Two types of weather indices were developed for each weather variable, i.e. simple and weighted weather indices. Simple weather indices were generated by summing the individual or interaction of weather parameters by taking weekly two weather variables at a time. Weighted weather indices were generated from sum product of individual or interaction of weather variables and it is in correlation with crop yield. The formula for computation of simple and weighted weather indices are given below.

Simple weather indices:

$$Z_{ij} = \sum_{w=1}^{m} X_{iw} \tag{1}$$

$$Z_{ii'j} = \sum_{w=1}^{m} X_{iw} X_{i'w} \tag{2}$$

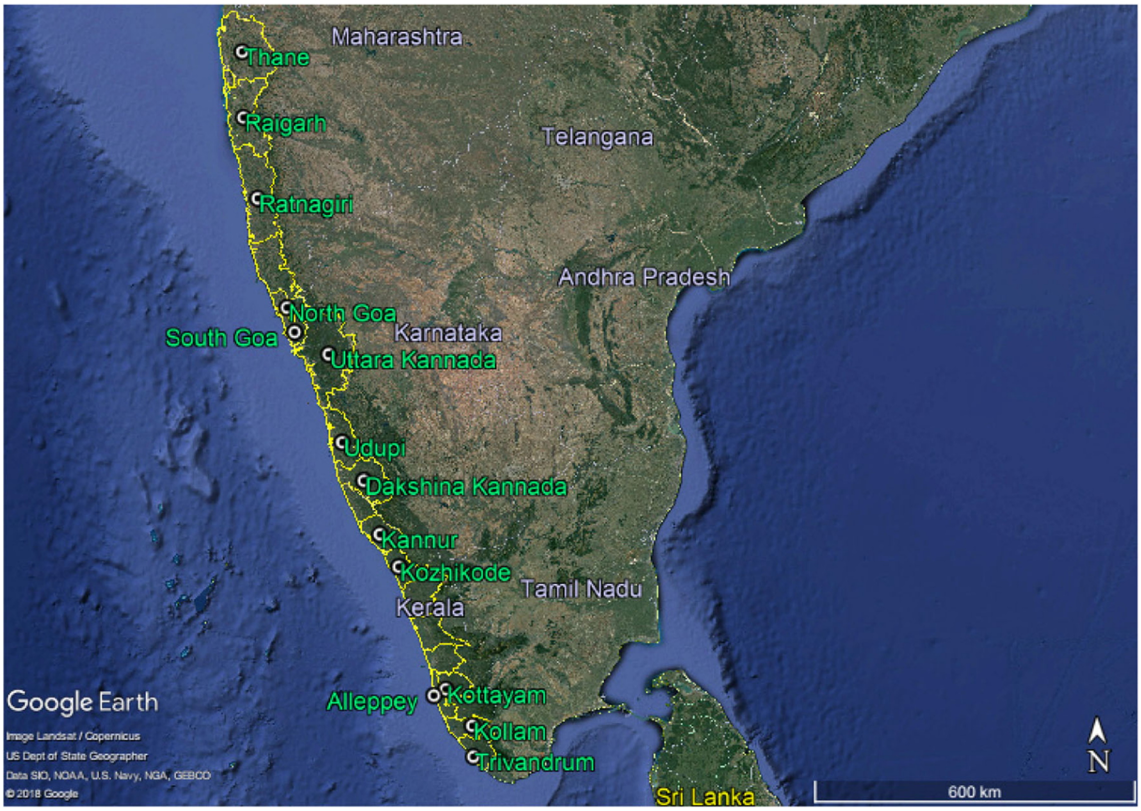

Fig. 1 Location of the selected districts (created using GoogleEarth)

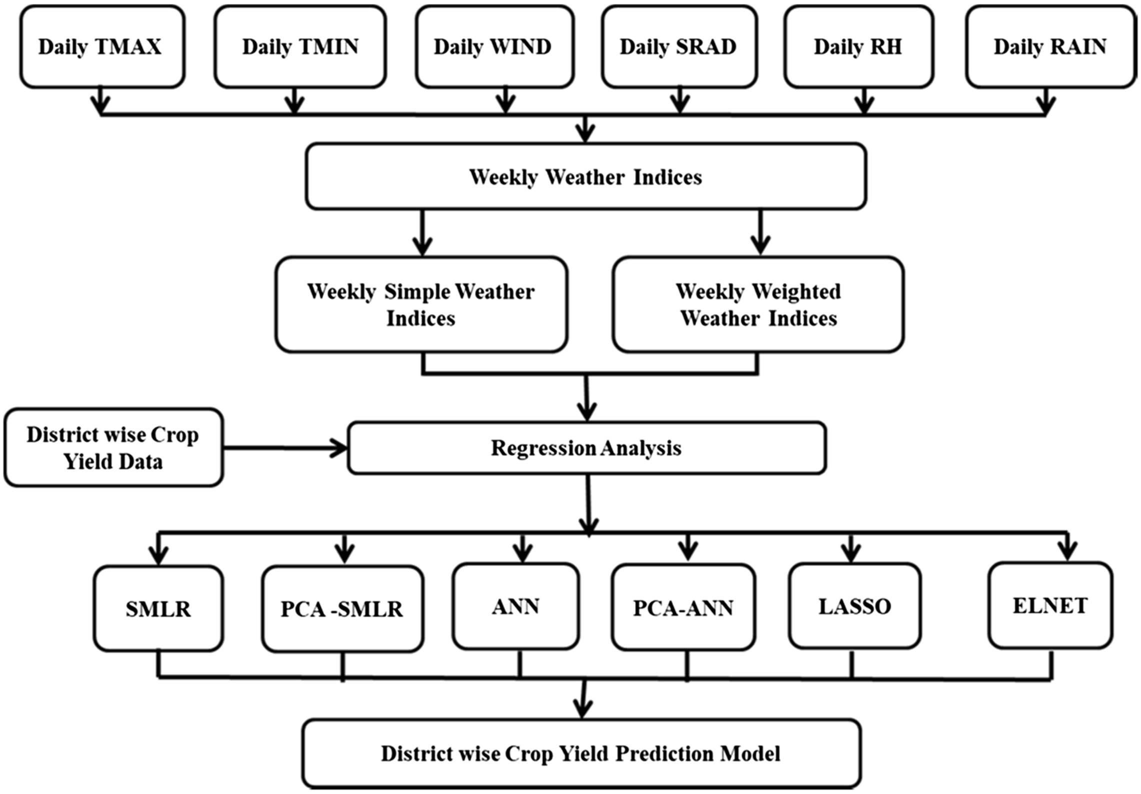

Fig. 2 Flowchart representing stages in model preparation. SMLR, PCA, ANN, LASSO and ENET denote stepwise multiple linear regression, principal component analysis, artificial neural network, least absolute shrinkage and selection operator and elastic net

Weighted weather indices:

$$Z_{ij} = \sum_{w=1}^{m} r_{iw}^{j} X_{iw} \quad (3)$$

$$Z_{ii'j} = \sum_{w=1}^{m} r_{ii'w}^{j} X_{iw} X_{i'w} \quad (4)$$

where

$X_{iw}/X_{ii'w}$ value of ith/i'th weather variable under study in wth week,
$r^{j}_{iw}/r^{j}_{ii'w}$ correlation coefficient of yield with ith weather variable/product of ith and i'th weather variables in wth week
m week of forecast
p number of weather variables used

The formation of weather variables for weather indices, thus, generated are presented in Table 1.

Multivariate techniques

To develop good crop yield prediction model, seven different types of multivariate analysis techniques are used in this study. Details of those models are given as follows:

Principal component analysis

In our study, we have performed principal component analysis (PCA) on all 42 weather indices calculated for each district. All the input variables were normalised by subtracting the minimum from each value and divide by the range, (x − min)/(max − min) before PCA analysis. As per the benchmarks set by Brejda et al. (2000) the principal components (PCs) with eigenvalues more than 1 were only considered. PCA was performed to avoid the over-fitting due to high dimension and large interdependency among independent variables. The first PC interprets maximum variability present in the data, and each subsequent component interprets remaining variability (Sharma et al. 2008).

Artificial neural network

In the present study, we have used three layers namely input, hidden and output feed-forward artificial neural network (ANN). Each layer consists of neurons or nodes interconnected with each other. The number of nodes in input and output layer is fixed by the dataset used. The main problem in the implementation ANN is to find the optimum number of hidden neurons or nodes. We have selected the number of hidden nodes by 'train' function of the 'caret' package using the method 'nnet' with 10-fold cross-validation in R software (Kuhn 2008). In the present study, all 42 indices were used as inputs whereas yield was the response variable (Fig. 3).

Table 1 Simple and weighted weather indices

Parameter	Simple weather indices	Weighted weather indices
Tmax	Z10	Z11
Tmin	Z20	Z21
Wind	Z30	Z31
SRAD	Z40	Z41
RH	Z50	Z51
Rain	Z60	Z61
Tmax*Tmin	Z120	Z121
Tmax*Wind	Z130	Z131
Tmax*SRAD	Z140	Z141
Tmax*RH	Z150	Z151
Tmax*Rain	Z160	Z161
Tmin*Wind	Z230	Z231
Tmin*SRAD	Z240	Z241
Tmin*RH	Z250	Z251
Tmin*Rain	Z260	Z261
Wind*SRAD	Z340	Z341
Wind*RH	Z350	Z351
Wind*Rain	Z360	Z361
SRAD*RH	Z450	Z451
SRAD*Rain	Z460	Z461
RH*Rain	Z560	Z561

Stepwise multiple linear regression

Multiple linear regression (MLR) is the standard and simplest approach for development of calibration models. But its application for dataset with independent variables greater sample number is not always successful (Balabin et al. 2011). However, feature selection in the form of stepwise MLR (SMLR) gives good results over large dataset. Stepwise regression procedure was adopted for selection of the best regression variable among many independent variables (Singh et al. 2014).

Principal components analysis-stepwise multiple linear regression and principal components analysis-artificial neural network

PCA followed by SMLR or ANN is the combination of feature extraction and selection method for data analysis. To overcome multicollinearity problem among weather variables, PC scores were used as regressors for SMLR and ANN to develop the crop yield models (Verma et al. 2016). PCA decomposes the original data matrix X into two matrices P and T as

$$\mathrm{X} = \mathrm{TP}^{\mathrm{T}}$$

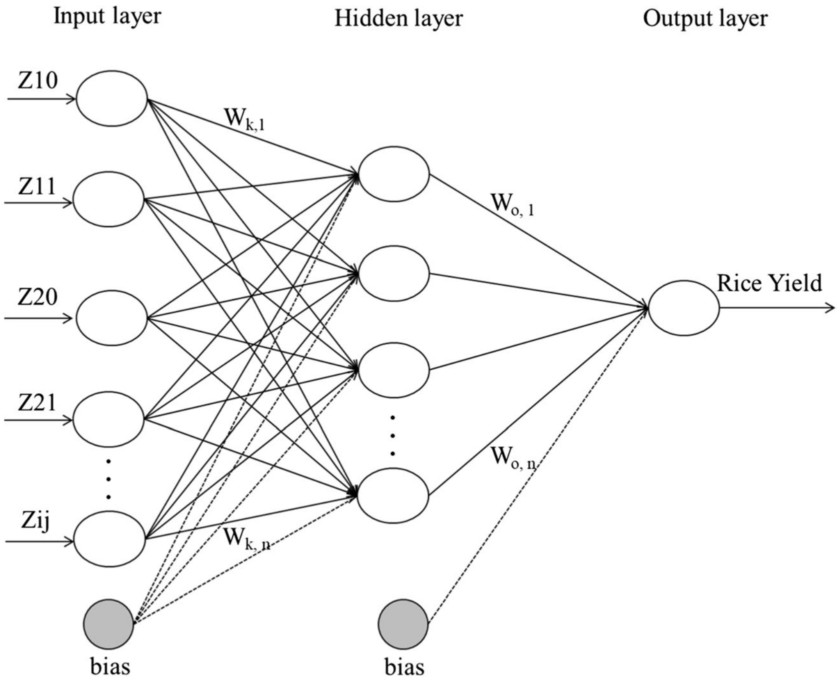

Fig. 3 Schematical representation of the ANN used in the study. Zij indicates weather indices, $W_{k,\ n}$ and $W_{o,\ n}$ are hidden-input and output connection weights

The matrix P is usually referred to as loadings matrix and the matrix T as score matrix which are orthogonal to each other. The superscript T indicates transpose of a matrix. Loadings are linear combinations of the original variables. The matrix T contains the original data in the rotated coordinate system.

Least absolute shrinkage and selection operator and elastic net

Least absolute shrinkage and selection operator (LASSO) and elastic net (ENET) are two shrinkage regression methods used for handling multicollinearity. These methods deal with multicollinearity by penalising the magnitude of regression coefficients. LASSO and ENET have two parameters namely lambda and alpha which needs to be optimised. The optimal lambda values for LASSO and ENET were selected by minimising the average mean square error in leave-one-out cross-validation (Piaskowski et al. 2016). The other tuning parameter alpha was set at 1 for LASSO and 0.5 for the ENET. In the present study, 'glmnet' package was used for LASSO and ENET implementation in R software (Friedman et al. 2009). The glmnet solves the following problem:

$$\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} w_i l\left(y_i, \beta_0-\beta^T x_i\right) + \lambda\left[(1-\alpha)\|\beta\|_2^2/2+\alpha\|\beta\|_1\right] \tag{5}$$

over a grid of values of λ covering the entire range. Here, l(y, η) is the negative log-likelihood contribution for observation i; e.g. for the Gaussian case, it is $1/2(y-\eta)^2$. The ENET penalty is controlled by α and bridges the gap between LASSO (α = 1, the default) and ridge (α = 0). The tuning parameter λ controls the overall strength of the penalty (Hastie and Qian 2014).

Model performance

For testing the performance of developed statistical forecasting models, R^2, root mean square error (RMSE) and normalised root mean square error (nRMSE) were calculated using the following formula:

$$R^2 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}\left(M_i-\overline{M}\right)\left(O_i-\overline{O}\right)}{\sigma_M\sigma_O}\right)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_i-M_i\right)^2} \tag{7}$$

$$nRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_i-M_i\right)^2} \times \frac{100}{\overline{O}} \tag{8}$$

M_i: model output; $\overline{M}$ and σ_M: mean and standard deviation of model output, respectively; O_i: observations; $\overline{O}$ and σ_O: mean and standard deviation of observations, respectively. R^2 values close to 1 and RMSE close to 0 indicate better model performance. According to nRMSE, the model is considered excellent, good, fair and poor when the values ranged < 10%, 10–20%, 20–30% and > 30%, respectively (Jamieson et al. 1991).

Comparison of multivariate models

For multiple datasets, the models should be evaluated using statistical hypothesis testing (Demšar 2006; Garcia and Herrera 2008; Soares and Anzanello 2018). For evaluation of models, non-parametric tests are preferred as outputs of multivariate models do not follow any probability distribution (Soares and Anzanello 2018). So, we have used non-parametric Friedman test for testing the significant difference among the models. If the test was found significant, then pairwise multiple comparison test was performed to identify the best model (Demšar 2006). The models were ranked on the basis of R^2, RMSE of calibration and validation (RMSEC and RMSEV) and nRMSE of validation, and average ranks across the districts were calculated to identify the best performing model. The average ranks were used for Friedman test followed by pairwise multiple comparison test.

Results

Summary statistics of yield data

The summary statistics of yield data pertaining to west coastal districts of India over the years 1983 to 2015 is presented in Supplementary Table 1. Maximum yield from the collected data was found in Alleppey district of Kerala (3247.99 kg ha^{-1}) and minimum yield was observed in Uttar Kannada district (1111.87 kg ha^{-1}) of Karnataka. The standard deviation of the yield across the districts varied between 264.98 and 456.47 kg ha^{-1}. The normality of the yield data was tested using normal Q–Q plot and Jarque-Bera test (Fig. 4). The yield data were found to be normally distributed as indicated by Jarque-Bera test (p value > 0.05) for all the districts except Udupi (p value = 0.02). The normal Q–Q plot also confirmed the normality thereby fulfilling the basic assumption of parametric models (MLR, LASSO and ENET).

Rice yield forecasting models

Stepwise multiple linear regression model

The yield prediction models developed using SMLR are shown in Table 2. The coefficient of determination (R^2) was significant at 1% probability level for all the districts of the west coastal zone of India. R^2 RMSE ranged between 0.62 (Alleppey) to 0.94 (Kozhikode) and 67.70 kg ha^{-1} (Kozhikode) to 253.01 kg ha^{-1}(Alleppey). The most influential weather parameter identified using SMLR was temperature followed by SRAD, RH and wind as identified by decoding the Z variates. All the selected Z variates were having significant positive influence ($p < 0.05$) on rice yield except Z20, Z260 and Z160 for Udupi, Dakshina Kannada and Kottayam districts, respectively. During validation, the highest RMSE was recorded in

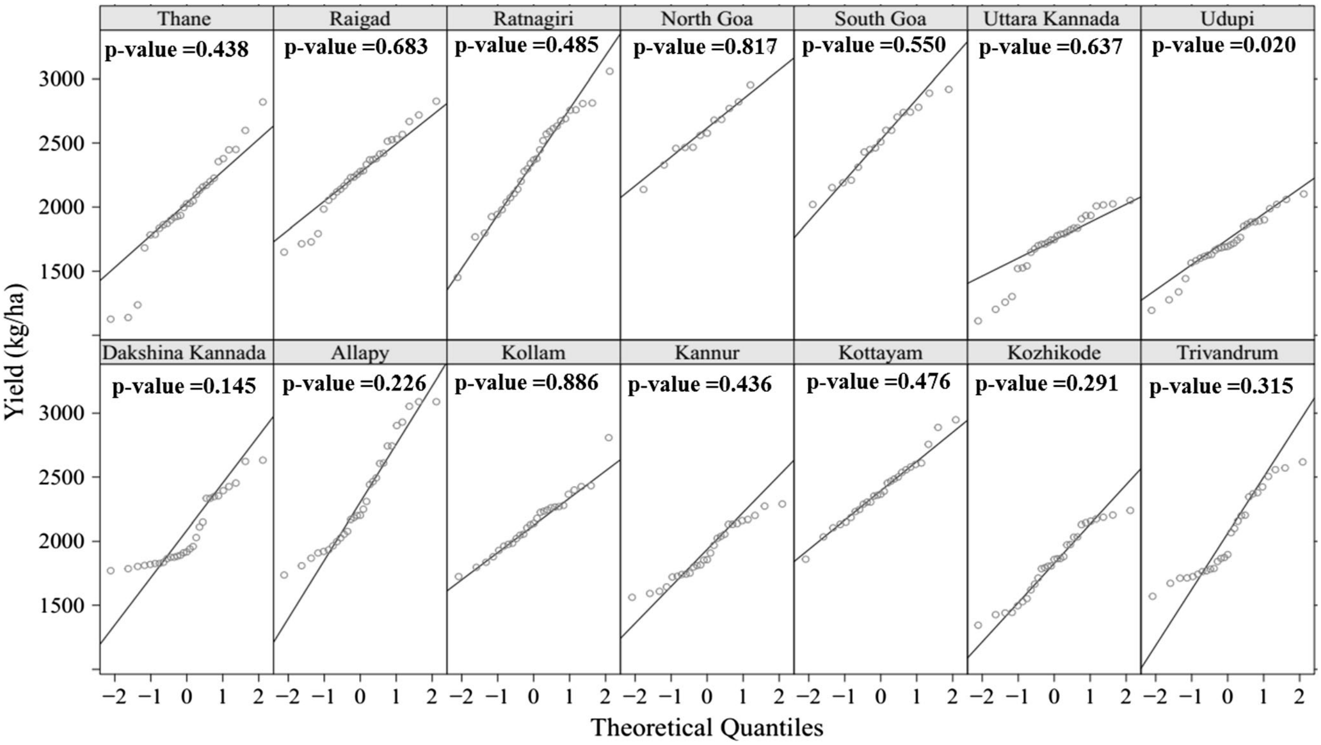

Fig. 4 Normal Q–Q plot for rice yield in 14 west coastal districts

Table 2 The yield prediction models for different districts of West Coast developed using SMLR

Sl. No.	Districts	Equation	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV (%)
1.	Raigad	$Y = -1508.19 + (0.291 \times Z120^{**}) + (2.447 \times Z241^{**})$	0.78	136.47	151.74	5.71
2.	Thane	$Y = 5194.129 + (163.790 \times Z21^{**}) + (1.490 \times Z41^{**})$	0.65	226.97	67.79	2.74
3.	Ratnagiri	$Y = 3304.736 + (26.464 \times Time^{**}) + (1.203 \times Z141^{**})$	0.81	162.95	91.43	3.09
4.	North Goa	$Y = 1334.24 + (1.35 \times Z41^{**})$	0.62	148.19	420.43	14.71
5.	South Goa	$Y = 490.37 + (162.73 \times Z11^{**})$	0.84	94.09	439.55	15.78
6.	Uttara Kannada	$Y = 2192.766 + (100 \times Z21^{*}) + (80.897 \times Z31^{*}) + (0.611 \times Z241^{*})$	0.80	109.70	331.28	15.44
7.	Udupi	$Y = 10{,}599.760 - (17.845 \times Z20^{**}) + (101.815 \times Z31^{**}) + (0.647 \times Z141^{**})$	0.80	107.57	511.86	20.01
8.	Dakshina Kannada	$Y = 4512.547 + (12.201 \times Time^{**}) + (70.592 \times Z21^{**}) + (1.786 \times Z121^{**}) + (1.276 \times Z241^{**}) - (0.005 \times Z260^{**})$	0.93	73.82	453.21	16.94
9.	Alleppey	$Y = -319.043 + (1.621 \times Z151^{**}) + (3.127 \times Z341^{**})$	0.62	253.01	645.48	21.26
10.	Kozhikode	$Y = 2200.645 + (25.544 \times Time^{**}) + (12.886 \times Z51^{**})$	0.94	67.70	179.43	7.42
11.	Kannur	$Y = 3170.280 + (26.376 \times Time^{**}) + (37.499 \times Z21^{**}) + (0.012 \times Z450^{*})$	0.98	33.39	66.63	3.05
12.	Kottayam	$Y = 3329.476 + (24.641 \times Time^{**}) - (0.008 \times Z160^{**}) + (0.181 \times Z361^{**}) - (0.080 \times Z120^{*})$	0.88	89.00	364.88	12.21
13.	Kollam	$Y = 1788.134 + (25.640 \times Time^{**})$	0.75	114.80	267.71	10.57
14.	Trivandrum	$Y = 3506.844 + (29.367 \times Time^{**}) + (0.601 \times Z241^{**})$ - $(0.378 \times Z120^{**}) + (1.480 \times Z141^{*})$	0.93	84.16	227.44	9.01

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

*Significance at 5% level; **significance at 1% level

Alleppey (645.48 kg ha^{-1}) while the lowest was found in Kannur district (66.63 kg ha^{-1}). According to nRMSE values computed during validation, the model predictions were excellent for Raigad, Thane, Ratnagiri, Kozhikode, Kannur and Trivandrum while it was good for North Goa, South Goa, Uttara Kannada, Dakshina Kannada, Kottayam and

Table 3 The yield prediction models for different districts of West Coast developed using PCA-SMLR

Sl. No.	Districts	No. of PCs	Equation	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV (%)
1.	Raigad	8 (92.87)	$Y = 2008.317 + (16.893 \times Time^{**}) + (130.831 \times PC1^{**})$	0.66	170.55	158.45	5.96
2.	Thane	9 (93.77)	$Y = 2071.509 + (182.517 \times PC6^{**}) + (185.429 \times PC2^{**}) + (265.752 \times PC1^{**})$	0.69	212.09	290.86	11.82
3.	Ratnagiri	7 (97.56)	$Y = 1858.34 + (32.69 \times Time^{**}) + (103.05 \times PC2^{*})$	0.77	176.99	68.07	2.30
4.	North Goa	6 (96.35)	$Y = 2629.791 + (161.194 \times PC3^{**})$	0.54	165.39	489.99	17.15
5.	South Goa	6 (94.48)	$Y = 2594.547 + (158.643 \times PC3^{**}) - (20.927 \times Time^{*}) + (74.872 \times PC2^{*})$	0.86	88.80	602.00	21.62
6.	Uttara Kannada	7 (93.89)	$Y = 1723.037 + (159.402 \times PC2^{**}) + (123.384 \times PC3^{**})$	0.74	124.61	408.81	18.91
7.	Udupi	8 (94.05)	$Y = 1718.268 + (159.685 \times PC1^{**}) + (97.018 \times PC4^{**}) + (63.885 \times PC6^{*}) + (62.837 \times PC8^{*}) + (55.414 \times PC5^{*})$	0.81	106.13	788.05	31.21
8.	Dakshina Kannada	7 (94.74)	$Y = 1761.73 + (19.84 \times Time^{**}) + (81.36 \times PC4^{*}) + (50.56 \times PC5^{*})$	0.82	114.39	981.89	36.70
9.	Alleppey	8 (96.08)	$Y = 1845.286 + (32.787 \times Time^{**})$	0.45	301.55	258.92	8.34
10.	Kozhikode	7 (95.09)	$Y = 1437.888 + (28.191 \times Time^{**}) + (40.335 \times PC4^{*})$	0.92	71.84	154.56	6.47
11.	Kannur	7 (96.06)	$Y = 1529.010 + (26.755 \times Time^{**}) + (37.338 \times PC6^{**}) + (21.395 \times PC5^{*})$	0.95	37.02	77.32	3.54
12.	Kottayam	7 (97.03)	$Y = 2117.768 + (19.999 \times Time^{**}) + (78.815 \times PC3^{**}) + (59.670 \times PC2^{*})$	0.82	107.87	301.36	10.09
13.	Kollam	7 (97.56)	$Y = 1788.134 + (25.640 \times Time^{**})$	0.75	114.80	267.71	11.71
14.	Trivandrum	6 (93.79)	$Y = 1517.627 + (37.662 \times Time^{**})$	0.83	131.34	185.34	7.34

Values in parenthesis indicates percentage variability explained by respective number of PCs

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

*Significance at 5% level; **significance at 1% level, respectively

Table 4 The yield prediction models for different districts of West Coast developed using ANN

Sl. No.	Districts	No. of hidden neurons	R_c^2 ($p < 0.01$)	RMSE (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV
1.	Raigad	09	0.75	132.37	292.49	11.00
2.	Thane	08	0.85	154.48	439.54	17.87
3.	Ratnagiri	05	0.98	60.19	386.67	13.07
4.	North Goa	04	0.94	24.02	680.07	23.80
5.	South Goa	03	0.95	51.46	462.32	16.60
6.	Uttara Kannada	09	0.71	142.22	297.01	13.73
7.	Udupi	06	0.69	138.25	630.40	24.97
8.	Dakshina Kannada	09	0.96	54.53	573.99	21.40
9.	Alleppey	03	0.84	168.48	794.29	25.57
10.	Kozhikode	12	0.93	68.56	529.93	34.76
11.	Kannur	02	0.92	83.71	304.27	19.95
12.	Kottayam	03	0.84	185.87	412.96	13.82
13.	Kollam	10	0.70	132.05	748.04	25.04
14.	Trivandrum	05	0.93	113.81	380.45	15.01

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

Kollam districts. The performance of the developed models was fair for Udupi and Alleppey districts with nRMSE values of 20.01 and 21.26%, respectively.

Principal components analysis-stepwise multiple linear regression model

PCA feature extraction method followed by SMLR is used for rice yield forecasting. The number of principal components (PCs) selected according to the eigenvalues more than 1 conditions were able to explain more than 90% variability present in the dataset for all the districts (Table 3). The number of PCs retained ranged between 6 and 9. Only Z variates were taken into consideration for PCA score generation. However, during model development, the PCA scores with time were taken as input variables. The R^2 was maximum for Kannur (0.95) with RMSE of 37.02 kg ha^{-1} and minimum for Alleppey (0.45) with RMSE of 301.55 kg ha^{-1} during calibration. We

Table 5 The yield prediction models for different districts of West Coast developed using PCA-ANN

Sl. No.	Districts	No. of hidden neurons	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV
1.	Raigad	02	0.75	142.23	440.80	16.58
2.	Thane	04	0.74	198.04	144.72	5.88
3.	Ratnagiri	02	0.89	121.91	314.49	10.63
4.	North Goa	01	0.22	244.44	187.96	6.58
5.	South Goa	03	0.83	118.59	137.31	4.93
6.	Uttara Kannada	03	0.96	51.00	449.11	20.77
7.	Udupi	01	0.81	113.13	760.31	30.11
8.	Dakshina Kannada	02	0.94	68.94	618.22	12.69
9.	Alleppey	02	0.64	248.57	440.28	14.17
10.	Kozhikode	02	0.91	91.87	521.99	33.68
11.	Kannur	02	0.87	94.22	159.74	7.32
12.	Kottayam	03	0.79	112.02	377.90	12.65
13.	Kollam	02	0.78	112.36	689.79	23.09
14.	Trivandrum	01	0.89	139.88	426.72	16.84

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

Table 6 The yield prediction models for different districts of West Coast developed using LASSO

Sl. No.	Districts	Equation	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV
1	Raigad	$Y = -2282.131 + (Z10 \times 0.118) + (Z11 \times 48.148) + (Z251 \times 0.497) + (Z451 \times 0.329) + (Z461 \times 0.0007)$	0.84	123.93	203.76	7.66
2	Thane	$Y = 1488.088 + (Time \times 1.550) + (Z11 \times 22.283) + (Z21 \times 105.390) - (Z60 \times 0.008) + (Z131 \times 0.374) + (Z151 \times 0.391) + (Z231 \times 0.903) + (Z251 \times 0.165) - (Z351 \times 0.009) + (Z451 \times 0.427) + (Z461 \times 0.041) - (Z560 \times 0.0007)$	0.81	171.63	62.71	2.55
3	Ratnagiri	$Y = 106.884 + (Z21 \times 7.561) + (Z31 \times 6.015) + (Z51 \times 1.685) + (Z240 \times 0.045) + (Z250 \times 0.035) + (Z251 \times 1.101) + (Z260 \times 0.004) + (Z261 \times 0.017)$	0.98	37.57	273.17	9.24
4	North Goa	$Y = -2553.866 + (Z11 \times 12.947) + (Z41 \times 0.386) + (Z51 \times 9.607) + (Z121 \times 0.443) + (Z261 \times 0.0001) + (Z341 \times 0.091) - (Z450 \times 0.022) - (Z460 \times 0.012) + (Z461 \times 0.019)$	0.87	92.04	58.09	2.03
5	South Goa	$Y = 11{,}190.36 - (Z10 \times 11.460) + (Z11 \times 164.510) - (Z50 \times 0.025) - (Z120 \times 0.107) - (Z130 \times 0.064) - (Z150 \times 0.0003) - (Z251 \times 0.017) + (Z341 \times 0.843) - (Z361 \times 0.036) + (Z451 \times 0.033) + (Z460 \times 0.0005)$	0.97	51.67	44.68	1.60
6	Uttara Kannada	$Y = 2118.033 + (Z11 \times 4.446) + (Z21 \times 69.313) + (Z31 \times 62.055) + (Z41 \times 0.036) + (Z141 \times 0.362) + (Z161 \times 0.008)$	0.81	109.18	375.58	17.37
7	Udupi	$Y = 5731.894 + (Z11 \times 17.392) - (Z20 \times 5.446) + (Z21 \times 62.986) + (Z31 \times 78.727) + (Z41 \times 0.113) + (Z141 \times 0.111) + (Z361 \times 0.031) + (Z461 \times 0.002) + (Z561 \times 0.001)$	0.83	103.25	842.59	32.04
8	Dakshina Kannada	$Y = 4601.416 + (Time \times 9.726) + (Z21 \times 66.756) + (Z51 \times 13.839) + (Z121 \times 1.741) + (Z151 \times 0.033) + (Z241 \times 0.984) - (Z260 \times 0.003) - (Z261 \times 0.0007) - (Z360 \times 0.003) + (Z450 \times 0.0004)$	0.94	69.96	695.05	25.91
9	Alleppey	$Y = 1242.784 + (Time \times 9.690) + (Z11 \times 9.994) + (Z151 \times 0.887) + (Z241 \times 0.083) + (Z341 \times 1.327) + (Z451 \times 0.035)$	0.68	250.73	620.03	19.96
10	Kozhikode	$Y = 1982.949 + (Time \times 21.772) + (Z41 \times 0.019) + (Z51 \times 11.363) + (Z141 \times 0.002) + (Z251 \times 0.026) + (Z561 \times 0.003)$	0.94	64.16	228.70	9.45
11	Kannur	$Y = 2746.798 + (23.246 \times Time) + (Z11 \times 2.757) - (Z20 \times 0.228) + (Z21 \times 41.673) + (Z51 \times 0.614) + (Z121 \times 0.091) + (Z151 \times 0.003) - (Z251 \times 0.130) - (Z260 \times 0.0002) + (Z340 \times 0.012) + (Z450 \times 0.007) - (Z451 \times 0.00004) + (Z461 \times 0.009)$	0.97	45.76	48.47	2.22
12	Kottayam	$Y = 3249.939 + (Time \times 27.885) - (Z10 \times 0.101) - (Z151 \times 0.075) + (Z231 \times 0.282) + (Z241 \times 0.677) - (Z460 \times 0.002) - (Z461 \times 0.0136)$	0.84	607.29	515.37	17.25
13	Kollam		0.81	109.18	146.68	6.42

Table 6 (continued)

Sl. No.	Districts	Equation	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV
		Y = 1263.614 + (Time × 13.540) + (Z51 × 0.301) + (Z161 × 0.0016) + (Z251 × 0.188) + (Z361 × 0.016) + (Z451 × 0.163)				
14	Trivandrum	Y = 3769.22 + (Time × 28.570) − (Z10 × 4.556) + (Z21 × 3.170) + (Z61 × 0.750) + (Z251 × 0.045) + (Z351 × 0.024) + (Z451 × 0.081)	0.87	158.20	175.60	6.96

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

observed that time was the most important variable affecting the crop yield followed by PC2 and PC3. RMSE during validation ranged between 68.07 kg ha^{-1} (Ratnagiri) and 981.89 kg ha^{-1} (Dakshina Kannada). The developed models performed excellent for Raigad, Ratnagiri, Alleppey, Kozhikode, Kannur and Trivandrum with nRMSE of 5.96, 2.30, 8.34, 6.47, 3.54 and 7.34%, respectively during validation while the performance was good for Thane (11.82%), Uttara Kannada (18.91%), North Goa (17.15%), Kottayam (10.09%) and Kollam (11.71%) districts. For Udupi and Dakshina Kannada, the prediction was poor with nRMSE of 31.21 and 36.70%, respectively.

Artificial neural network and principal component analysis-artificial neural network model

For development of ANN model, the Z variates were taken as inputs whereas, for PCA-ANN model, the PCA scores generated from PCA analysis applied on Z variates were used. The optimum number of hidden neurons varied between 1 and 12. The number of hidden neurons was less in case of PCA-ANN as compared to ANN as the number of inputs was much less in PCA-ANN. The predictive performance of the models as indicated by R^2 and RMSE during calibration varied between 0.69–0.98 and 24.02–185.87 kg ha^{-1} for ANN (Table 4) and between 0.22–0.96 and 51–248.57 kg ha^{-1} for PCA-ANN (Table 5). However, during validation with independent dataset, the RMSE and nRMSE ranged between 292.49 to 794.29 kg ha^{-1} and 11 to 34.76% for ANN and between 137.31 to 760.31 and 4.93 to 33.68% for PCA-ANN. The performance of ANN model was found good for Raigad, Thane, Ratnagiri, South Goa, Uttara Kannada, Kannur, Kottayam and Trivandrum; fair for North Goa, Udupi, Dakshina Kannada, Alleppey and Kollam and poor for Kozhikode with respect to nRMSE of validation while for PCA-ANN, the performance was excellent for Thane, North Goa, South Goa and Kannur; good for Raigad, Ratnagiri, Uttara Kannada, Dakshina Kannada, Alleppey, Kottayam, Trivandrum; fair for Kollam and poor for Udupi and Kozhikode. For none of districts, the performance was found excellent while using standalone ANN during validation. The range of RMSE and nRMSE during validation was found superior in PCA-ANN as compared to ANN unlike during calibration which indicted overfitting when using ANN alone. This result is in line with previous findings of Suleiman et al. (2016) while comparing ANN and PCA-ANN for predicting roadside particulate matter but differs with Kumari et al. (2016) while comparing MLR, autoregressive integrated moving average (ARIMA) and ANN model for predicting pigeon pea yield in Varanasi region.

Table 7 The yield prediction models for different districts of West Coast developed using ENET

Sl. No.	Districts	Equation	R_c^2 ($p < 0.01$)	RMSEC (kg ha^{-1})	RMSEV (kg ha^{-1})	nRMSEV
1	Raigad	$Y = -422.228 + (\text{Time} \times 1.753) + (Z11 \times 25.835) + (Z41 \times 0.029) + (Z121 \times 0.113) + (Z141 \times 0.029) + (Z241 \times 0.352) + (Z251 \times 0.291) + (Z451 \times 0.137) + (Z461 \times 0.0007)$	0.84	119.60	246.31	9.26
2	Thane	$Y = 1887.276 + (Z11 \times 13.075) + (Z21 \times 39.389) + (Z41 \times 0.149) + (Z121 \times 0.184) + (Z141 \times 0.149) + (Z151 \times 0.1527) + (Z241 \times 0.064) + (Z251 \times 0.124) + (Z341 \times 0.711) + (Z451 \times 0.124) + (Z461 \times 0.0077)$	0.76	197.90	240.38	9.77
3	Ratnagiri	$Y = -384.2118 + (Z20 \times 0.943) + (Z21 \times 32.915) + (Z51 \times 8.355) + (Z240 \times 0.0325) + (Z250 \times 0.019) + (Z251 \times 0.590) + (Z260 \times 0.002) + (Z261 \times 0.016) + (Z341 \times 0.019) + (Z361 \times 0.011) + (Z451 \times 0.031) + (Z460 \times 0.001) + (Z461 \times 0.005) + (Z560 \times 0.0001)$	0.98	35.59	348.48	11.78
4	North Goa	$Y = 1786.423 + (Z21 \times 0.479) + (Z121 \times 0.324) + (Z241 \times 0.310) - (Z460 \times 0.0006) + (Z461 \times 0.014)$	0.72	200.74	134.18	4.70
5	South Goa	$Y = 7979.20 - (Z10 \times 7.036) + (Z11 \times 92.507) + (Z41 \times 0.023) - (Z50 \times 0.014) + (Z121 \times 0.522) + (Z141 \times 0.029) - (Z150 \times 0.013) + (Z261 \times 0.0001) - (Z340 \times 0.048) + (Z341 \times 0.969) + (Z451 \times 0.0068)$	0.90	71.63	95.20	3.42
6	Uttara Kannada	$Y = 2030.744 + (Z21 \times 32.519) + (Z31 \times 22.065) + (Z41 \times 0.128) + (Z61 \times 0.065) + (Z141 \times 0.128) + (Z161 \times 0.003) + (Z231 \times 0.085) + (Z241 \times 0.139) + (Z351 \times 0.189) + (Z451 \times 0.001) + (Z561 \times 0.0007)$	0.81	112.13	389.40	18.01
7	Udupi	$Y = 5378.858 + (Z11 \times 16.053) - (Z20 \times 4.733) + (Z21 \times 51.568) + (Z31 \times 26.329) + (Z41 \times 0.0812) + (Z121 \times 0.079) + (Z131 \times 0.399) + (Z141 \times 0.081) + (Z231 \times 0.590) + (Z241 \times 0.081) + (Z351 \times 0.212) + (Z361 \times 0.035) + (Z451 \times 0.016) + (Z461 \times 0.002) + (Z561 \times 0.001)$	0.94	61.65	830.11	31.56
8	Dakshina Kannada	$Y = 3053.82 + (\text{Time} \times 8.293) + (Z11 \times 7.822) + (Z21 \times 47.213) + (Z41 \times 0.219) + (Z51 \times 8.463) - (Z60 \times 0.0054) + (Z61 \times 0.045) + (Z212 \times 0.762) + (Z141 \times 0.221) + (Z151 \times 0.063) + (Z241 \times 0.186) - (Z260 \times 0.0008) + (Z451 \times 0.038) + (Z461 \times 0.0013) + (Z561 \times 0.0005)$	0.93	71.66	731.83	27.28
9	Alleppey	$Y = 1417.182 + (\text{Time} \times 6.504) + (Z151 \times 0.534) + (Z241 \times 0.014) + (Z341 \times 0.690) + (Z451 \times 0.020)$	0.69	243.03	669.61	21.56
10	Kozhikode	$Y = 844.422 + (\text{Time} \times 14.334) + (Z11 \times 2.557) + (Z41 \times 0.020) + (Z51 \times 9.149) + (Z61 \times 0.096) + (Z141 \times 0.020) + (Z251 \times 0.095) + (Z361 \times 0.024) + (Z561 \times 0.002)$	0.94	65.34	317.48	13.12
11	Kannur	$Y = 2134.676 + (\text{Time} \times 19.967) + (Z11 \times 4.689) + (Z21 \times 28.185) + (Z340 \times 0.021) + (Z450 \times 0.0005) + (Z461 \times 0.014)$	0.97	42.12	21.35	0.98
12	Kottayam	$Y = 2512.320 + (\text{Time} \times 14.457) + (Z41 \times 0.020) + (Z141 \times 0.019) + (Z231 \times 0.043) + (Z241 \times 0.364) + (Z351 \times 0.0009)$	0.77	143.16	95.20	3.42
13	Kollam	$Y = 1535.133 + (\text{Time} \times 8.093) + (Z51 \times 1.636) + (Z451 \times 0.112)$	0.79	155.98	118.60	5.19
14	Trivandrum	$Y = 1577.922 + (\text{Time} \times 16.015) + (Z51 \times 3.599) + (Z61 \times 0.086) + (Z161 \times 0.0024) + (Z241 \times 0.019) + (Z251 \times 0.133) + (Z261 \times 0.0003) + (Z361 \times 0.032) + (Z451 \times 0.028) + (Z561 \times 0.00001)$	0.91	97.31	253.14	10.03

R_c^2 coefficient of determination of calibration, *RMSEC* root mean square error of calibration, *RMSEV* root mean square error of validation, *nRMSEV* normalised root mean square error of validation

Table 8 Descriptive statistics of climatic variables

	Tmean (°C)	Wind (m s^{-1})	SRAD (MJ m^{-2} day^{-1})	RH (%)	Rainfall (mm year^{-1})
Mean	27.1	4.7	17.3	85.5	2734.8
Maximum	32.4	11.0	26.9	94.1	3636.7
Minimum	22.9	1.2	1.5	38.8	1790.3
Standard deviation	1.1	1.7	3.8	4.9	544.2
CV (%)	4.0	37.1	21.8	5.7	19.9

Table 9 Multiple pairwise comparisons of the multivariate models using Friedman's aligned ranks post hoc test followed by Bergmann and Hommel dynamic correction of *p* values

	SMLR	PCA-SMLR	LASSO	ENET	ANN	PCA-ANN
SMLR	–	0.111	1.000	1.000	0.111	0.054
PCA-SMLR	0.111	–	0.009	0.087	1.000	1.000
LASSO	1.000	0.009	–	1.000	0.009	0.002
ENET	1.000	0.087	1.000	–	0.087	0.037
ANN	0.111	1.000	0.009	0.087	–	1.000
PCA-ANN	0.054	1.000	0.002	0.037	1.000	–

Least absolute shrinkage and selection operator and elastic net

The models developed using LASSO and ENET and subsequent validation of the developed models is presented in Tables 6 and 7. Maximum R^2 was found for the Ratnagiri district (0.98) with RMSE 37.57 kg ha^{-1} and the minimum R^2 was recorded for Alleppey district (0.68) with RMSE 250.73 kg ha^{-1}. The Z variates were having positive influence on yield using LASSO except Z351 for Thane; Z450 and Z460 for North Goa; Z10, Z50, Z120, Z130, Z150, Z251 and Z361 for South Goa; Z260, Z261 and Z360 for Dakshina Kannada; Z20, Z251, Z260 and Z451 for Kannur; Z10, Z151, Z460 and Z461 for Kottayam and Z10 for Trivandrum while for ENET, these were Z460 for North Goa; Z10, Z150 and Z340 for South Goa; Z20 for Udupi and Z60 and Z260 for Dakshina Kannada district. The most important meteorological variable included for LASSO model was RH followed by SRAD and Tmin while it was SRAD followed by Tmin, RH, Rain, Tmax and wind, respectively for ENET model. Validation of LASSO model revealed that the predictions were excellent for Raigad, Thane, Ratnagiri, North Goa, South Goa, Kozhikode, Kannur, Kollam and Trivandrum; good for Uttara Kannada, Alleppey and Kottayam and fair and poor for Dakshina Kannada and Udupi, respectively with respect to nRMSE. The minimum nRMSE was recorded for South Goa district (1.60%) and the highest was observed for Udupi district (32.04%). For ENET model, the performance according to nRMSE was found excellent for Raigad, Thane, North Goa, South Goa, Kannur, Kottayam and Kollam; good for Ratnagiri Uttara Kannada and Kozhikode, Trivandrum and fair for Dakshina Kannada and Alleppey while it was poor for Udupi district.

Discussion

Effect of weather parameters on rice yield

Weather parameters have profound influence on rice yield. The effect of temperature on rice yield has been reported extensively (Nyang'Au et al. 2014; Sánchez et al. 2014; Jagadish et al. 2015; Sridevi and Chellamuthu 2015; Cai et al. 2016; Shi et al. 2016; Talla et al. 2017). The mean weekly temperature of the study region during the rice-growing season ranged from 22.9 to 32.4 °C (Table 8) which is very much within the optimum temperature required for rice growth (15–18 to 30–33 °C, (Nishiyama 1976). But sometimes, the maximum temperature exceeded 35 °C and these extreme temperatures have destructive effect on rice growth and yield (Yoshida 1981; Sun and Huang 2011). Temperature affects the crop yield by changing the rate of photosynthesis, respiration, spikelet sterility and length of growing season (Wassmann et al. 2009; Krishnan et al. 2011; Rai et al. 2012; Akinbile et al. 2015). Higher temperature found to decrease duration of crop life cycle thereby shortens the grain filling period which leads to lower crop yield and grain quality. Solar radiation has a positive impact on rice yield by directly affecting the biomass accumulation (Akinbile et al. 2015). Reduction in solar radiation particularly during reproductive and ripening stage leads to reduction in yield (Rai et al. 2012). Higher RH has negative influence on crop yield as higher humidity causes reduction in evapotranspiration thereby lowering the cooling effect due to evaporation (Matsui et al. 2007; Wassmann et al. 2009). High RH also causes incidence of pest and diseases which leads to crop yield reduction. Higher vapour pressure deficit during anthesis will lead to reduction of panicle temperature due to transpirational cooling which helps in reducing the high-temperature-induced spikelet sterility (Matsui et al. 2007). Wind indirectly affects the crop yield by changing the vapour pressure deficit and transpirational cooling. The selection of solar radiation as an important variable affecting the rice yield using LASSO and ENET is in line with previous studies (Zhang et al. 2010; Yang et al. 2015; Oguntunde et al. 2018). The annual average rainfall in the region varied between 1790.3 and 3636.7 mm (Table 8), out of which 56.2 to 94.9% of rainfall received during June to September. So, there is sufficient rainfall throughout the rice-growing season in the region. However, high rainfall during flowering and ripening stage may reduce pollination and cause lodging which may lead to decline in yield and quality (Yang et al. 2015).

Cross-comparison of the models

The ranking of the models on the basis of R^2 and RMSE of calibration revealed that LASSO was the best performing model followed by ENET while PCA-SMLR was the worst model. The order of performance of the model during calibration was as follows: LASSO (2.52) > ENET (2.82) > ANN (3.16) > SMLR (3.61) > PCA-ANN (4.20) > PCA-SMLR (4.70). The models were also ranked using RMSE and nRMSE of validation which was found as follows: SMLR (2.55) > LASSO (2.75) > ENET (3.32) > PCA-SMLR (3.38) > PCA-ANN (4.18) > ANN (4.82). The performance of ANN was good during calibration while it was the worst model during validation which indicated over fitting as it uses all the 42 Z variates as input. The overall ranking based on R^2 and RMSE of calibration; RMSE and nRMSE of validation revealed the order as LASSO (2.63) > ENET (3.07) > SMLR (3.08) > ANN (3.99) > PCA-SMLR (4.04) > PCA-ANN (4.19). The reason behind the better performance of LASSO and ENET is that these models penalise the magnitude of coefficients with feature selection. Penalisation prevents overfitting and reduces model complexity by making some of the coefficients zero, which is equivalent to the particular feature being excluded from the model. It provides great computational advantage over SMLR or ANN as the features with zero coefficients can simply be ignored. The feature selection algorithms like LASSO, ENET 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. During validation, the performance of combination of feature extraction and feature selection methods like PCA-SMLR and feature extraction with neural network was found poor. This may be due to the fact that PCA does not consider the dependent variable during transformation of input variables. On the other hand, in the present study, the components with large variances were retained while those with small variances were rejected with the assumption that components with small variance have very little predictiveness in the regression which may not be true always (Jolliffe 1982). For evaluation of multiple models statistically, overall average ranks were calculated and non-parametric Friedman test was applied to check the significant difference among the models. Non-parametric tests are preferred as outputs of multivariate models do not follow any probability distribution. The Friedman test was found significant at $p < 0.001$ which indicated the presence of significant difference among the models. Then, the Friedman's aligned ranks post hoc test followed by Bergmann and Hommel dynamic correction of p values was performed for pairwise multiple comparison. The results indicated LASSO as the best model which was found similar to SMLR and ENET (Table 9). All other multivariate models did not revealed any significant difference among them during multiple pairwise comparison.

Conclusions

In the present investigation, six different multivariate models were compared for prediction of rice yield using long-term weather variables and the results revealed that LASSO model can be used for west coast of India. It was also found that the performance of SMLR and ENET were at par with LASSO. So, these models can also be very well utilised for rice yield forecasting for the studied region.

Acknowledgements The India Meteorological Department is duly acknowledged for providing weather data of different stations. This work was supported by the Indian Council of Agricultural Research under Institute project at ICAR-Central Coastal Agricultural Research Institute, Old Goa, Goa, India. The authors are thankful to the reviewer whose comments improved the quality of this paper.

References

Akinbile CO, Akinlade GM, Abolude AT (2015) Trend analysis in climatic variables and impacts on rice yield in Nigeria. J Water Clim Chang 6:534. https://doi.org/10.2166/wcc.2015.044

Annu, Sisodia BVS, Rai VN (2017) An application of principal component analysis for pre- harvest forecast model for wheat crop based on biometrical characters. Int Res J Agric Econ Stat 8:83–87. https://doi.org/10.15740/HAS/IRJAES/8.1/83-87

Azfar M, Sisodia BVS, Rai VN, Devi M (2015) Pre-harvest forecast models for rapeseed & mustard yield using principal component. Mausam 4:761–766

B S Dhekale PKS and TPU (2014) Weather based pre-harvest forecasting of rice at Kolhapur (Maharashtra). Trends Biosci 7:39–41

Balabin RM, Lomakina EI, Safieva RZ (2011) Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. Fuel 90:2007–2015. https://doi.org/10.1016/j.fuel.2010.11.038

Basso B, Cammarano D, Carfagna E (2013) Review of crop yield forecasting methods and early warning systems. In: Intergovernmental panel on climate change (ed) climate change 2013 - the physical science basis. Cambridge University Press, Cambridge, pp 1–30

Bhuvaneswari K, Geethalaxmi V, Lakshmanan A et al (2014) Climate change impact assessment and developing adaptation strategies for rice crop in western zone of Tamil Nadu. J Agrometeorol 16:38–43

Bocca FF, Rodrigues LHA (2016) The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Comput Electron Agric 128:67–76. https://doi.org/10.1016/j.compag.2016.08.015

Brejda JJ, Moorman TB, Karlen DL, Dao TH (2000) Identification of regional soil quality factors and indicators I. Central and southern high plains. Soil Sci Soc Am J 64:2115–2124

Cai C, Yin X, He S et al (2016) Responses of wheat and rice to factorial combinations of ambient and elevated CO_2 and temperature in FACE experiments. Glob Chang Biol 22:856–874. https://doi.org/10.1111/gcb.13065

Collins LM, Schafer JL, Kam C-M (2001) A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol Methods 6:330–351. https://doi.org/10.1037/1082-989X.6.4.330

Das B, Sahoo RN, Pargal S et al (2017) Comparison of different uni- and multi-variate techniques for monitoring leaf water status as an indicator of water-deficit stress in wheat through spectroscopy. Biosyst Eng 160:69–83. https://doi.org/10.1016/j.biosystemseng.2017.05.007

Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7:1–30

Dhekale BS, Mahdi SS, Dalvi TP, Sawant PK (2014) Forecast models for groundnut using meteorological variables in Kolhapur, Maharashtra. J Agrometeorol 16:238–239

Dutta S, Patel NK, Srivastava SK (2001) District wise yield models of rice in Bihar based on water requirement and meteorological data. J Indian Soc Remote Sens 29:175–181

Friedman J, Hastie T, Tibshirani R (2009) glmnet: Lasso and elastic-net regularized generalized linear models. R Package Version 2009:1

Garcia S, Herrera F (2008) An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. J Mach Learn Res 9:2677–2694

Government of India, Ministry of Agricuture and Farmers Welfare: Deparment of Agriculture Cooperation, and Welfare (2016) India Annu Rep 2016–17

Hastie T, Qian J (2014) Glmnet vignette. http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html. Last access: 19 June 2018

Jagadish SVK, Murty MVR, Quick WP (2015) Rice responses to rising temperatures - challenges, perspectives and future directions. Plant Cell Environ 38:1686–1698. https://doi.org/10.1111/pce.12430

Jamieson PD, Porter JR, Wilson DR (1991) A test of the computer simulation model ARCWHEAT1 on wheat crops grown in New Zealand. F Crop Res 27:337–350. https://doi.org/10.1016/0378-4290(91)90040-3

Jolliffe IT (1982) A note on the use of principal components in regression. Appl Stat 31:300. https://doi.org/10.2307/2348005

Krishnan P, Ramakrishnan B, Reddy KR, Reddy VR (2011) High-temperature effects on rice growth, yield, and grain quality. Adv Agron 111:87–206

Kuhn M (2008) Building predictive models in R using caret package. J Stat Softw 28:1–26

Kumar N, Pisal RR, Shukla SP, Pandye KK (2014) Regression technique for South Gujarat. MAUSAM 65:361–364

Kumari P, Mishra GC, Srivastava CP (2016) Statistical models for forecasting pigeonpea yield in Varanasi region. J Agrometeorol 18(18): 306–310

Lobell DB, Burke MB (2010) On the use of statistical models to predict crop yield responses to climate change. Agric For Meteorol 150: 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

Lobell DB, Schlenk er W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. Science 333:616–620. https://doi.org/10.1126/science.1204531

Matsui T, Kobayasi K, Yoshimoto M, Hasegawa T (2007) Stability of rice pollination in the field under hot and dry conditions in the Riverina region of New South Wales, Australia. Plant Prod Sci 10:57–63. https://doi.org/10.1626/pps.10.57

Nishiyama I (1976) Effects of temperature on the vegetative growth of rice plants. Clim Rice 159–185

Nyang'Au WO, Mati BM, Kalamwa K et al (2014) Estimating rice yield under changing weather conditions in Kenya using ceres rice model. Int J Agron 2014:1–12. https://doi.org/10.1155/2014/849496

Oguntunde PG, Lischeid G, Dietrich O (2018) Relationship between rice yield and climate variables in Southwest Nigeria using multiple linear regression and support vector machine analysis. Int J Biometeorol 62:459–469. https://doi.org/10.1007/s00484-017-1454-6

Pandey KK, Rai VN, Sisodia BVS, Singh SK (2015) Effect of weather variables on rice crop in eastern Uttar Pradesh, India. Plant Arch 15: 575–579

Piaskowski JL, Brown D, Campbell KG (2016) Near-infrared calibration of soluble stem carbohydrates for predicting drought tolerance in spring wheat. Agron J 108:285–293. https://doi.org/10.2134/agronj2015.0173

Rai YK, Ale BB, Alam J (2012) Impact assessment of climate change on paddy yield: a case study of Nepal agriculture research council (NARC), Tarahara, Nepal. J Inst Eng 8:147–167. https://doi.org/10.3126/jie.v8i3.5941

Rai KK, N P V, Bharti BVS, SA K (2013) Pre -harvest forecast models based on weather variable. Adv Biores 4:118–122

Sánchez B, Rasmussen A, Porter JR (2014) Temperatures and the growth and development of maize and rice: a review. Glob Chang Biol 20: 408–417. https://doi.org/10.1111/gcb.12389

Sharma KL, Grace JK, Mandal UK et al (2008) Evaluation of long-term soil management practices using key indicators and soil quality indices in a semi-arid tropical Alfisol. Aust J Soil Res 46:368–377. https://doi.org/10.1071/SR07184

Shi W, Tao F, Zhang Z (2013) A review on statistical models for identifying climate contributions to crop yields. J Geogr Sci 23:567–576. https://doi.org/10.1007/s11442-013-1029-3

Shi W, Yin X, Struik PC et al (2016) Grain yield and quality responses of tropical hybrid rice to high night-time temperature. F Crop Res 190: 18–25. https://doi.org/10.1016/j.fcr.2015.10.006

Singh RS, Patel C, Yadav MK, Singh KK (2014) Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. J Agrometeorol 16:199–202

Soares F, Anzanello MJ (2018) Support vector regression coupled with wavelength selection as a robust analytical method. Chemom Intell Lab Syst 172:167–173. https://doi.org/10.1016/j.chemolab.2017.12.007

Sridevi V, Chellamuthu V (2015) Impact of weather on rice—a review. Int J Appl Res 1:825–831

Suleiman A, Tight MR, Quinn AD (2016) Hybrid neural networks and boosted regression tree models for predicting roadside particulate matter. Environ Model Assess 21:731–750. https://doi.org/10.1007/s10666-016-9507-5

Sun W, Huang Y (2011) Global warming over the period 1961-2008 did not increase high-temperature stress but did reduce low-temperature stress in irrigated rice across China. Agric For Meteorol 151:1193–1201. https://doi.org/10.1016/j.agrformet.2011.04.009

Talla A, Swain DK, Tewari VK, Biswal MP (2017) Significance of weather variables during critical growth stages for hybrid rice production in subtropical India. Agron J 109:1891–1899. https://doi.org/10.2134/agronj2017.01.0052

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

Wassmann R, Jagadish SVK, Sumfleth K et al (2009) Regional vulnerability of climate change impacts on Asian rice production and scope for adaptation. Adv Agron 102:91–133

Yang L, Qin Z, Tu L (2015) Responses of rice yields in different rice-cropping systems to climate variables in the middle and lower reaches of the Yangtze River, China. Food Secur 7:951–963. https://doi.org/10.1007/s12571-015-0497-y

Yoshida S (1981) Fundamentals of rice crop science. International Rice Research Institute, Los Banos

Zhang T, Zhu J, Wassmann R (2010) Responses of rice yields to recent climate change in China: an empirical assessment based on long-term observations at different spatial scales (1981–2005). Agric For Meteorol 150:1128–1137. https://doi.org/10.1016/j.agrformet.2010.04.013