Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 25 (3) : 440-447 (September - 2023)
https://doi.org/10.54386/jam.v25i3.2194
https://journal.agrimetassociation.org/index.php/jam

Research paper

Development of groundnut yield forecasting models in relation to weather parameters in Andhra Pradesh, India

K. NIRMAL RAVI KUMAR[1], ANURAG SATPATHI[2*], M. JAGAN MOHAN REDDY[3], PARUL SETIYA[2] and AJEET SINGH NAIN[2]

[1]*Acharya NG Ranga Agricultural University, Lam- 522034. Guntur (Dist), Andhra Pradesh, India*
[2]*Department of Agrometeorology, College of Agriculture, G.B. Pant University of Agriculture and Technology, Pantnagar-263145, Uttarakhand, India*
[3]*Extension Education Institute, Professor Jayashankar Telangana State Agricultural University, Rajendranagar 500030, Hyderabad, India*
Corresponding Author email: anuragsatpathi50@gmail.com*

ABSTRACT

Groundnut is a key oilseed crop in the world and India is one of the largest groundnuts producing country in terms of area and yield. Keeping that in view, five models were developed for five districts of Andhra Pradesh to forecast the groundnut yield *viz*., Stepwise Multiple Linear Regression (SMLR), Ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ELNET) and Artificial Neural Network (ANN). The historical data on the weather parameters are obtained from NASA POWER web portal and groundnut yields for these districts of the state during both Kharif and Rabi seasons obtained through Season and Crop Report, Government of Andhra Pradesh for the period, 2001 to 2020. In total 30 weather indices were generated through five weather variables. The assessment of models was done by fixing 75 % of the data for calibration and left 25 % data for validation. The findings inferred that based on the values of R^2, RMSE, nRMSE and EF, Ridge regression, ELNET and ANN models showed better performance for Ananthapur, Chittoor and Kadapa districts and SMLR and LASSO models showed better performance for Kurnool and Nellore districts during both Kharif and Rabi seasons at calibration and validation stages.

***Keyword**:* Groundnut, Stepwise multiple linear regression, Ridge regression, Least absolute shrinkage and selection operator, Elastic net, Artificial neural network.

Agriculture is one of the greatest susceptible sectors to climatic change as weather variables involving temperature and rainfall are direct inputs into agriculture production system (Saravanakumar *et al.,* 2022). According to Jasna *et al.,* (2014), crop yields in India will decline by 4.5 to 9.0 % as a result of weather abnormalities. Given that agriculture accounts for around 16 % of India's Gross Domestic Product (GDP) and due to weather abnormalities, reduction in crop yields indicates a cost of climate change to be roughly at 1.5 % of GDP annually. In view of changing weather variables, it is difficult to predict food security issues, and this will certainly increase pressure on agriculture.

India is in leading position for groundnut acreage and is the second leading producer of groundnut in the world with 1.01 million tonnes with a productivity of 1816 kg/ha in 2020-21. In Andhra Pradesh, 62.17 % of total working population is dependent on agriculture and allied activities (Agricultural Statistics at a Glance, 2020-21). As around 46% of gross sown area in Andhra Pradesh is under rainfed condition. Andhra Pradesh contributed 7% to total groundnut production in India (2020-21). Groundnut contributes around 87% of acreage and 91% of production to total oilseeds in Andhra Pradesh (ANGRAU Groundnut Outlook Report, 2021). It is cultivated in one or more time of year, but nearly 80 % of land and production comes from Kharif season (June-October) under various agro-climatic zones and remaining 20 % from Rabi season (November-March).

The fluctuations in weather variables due to climate

Article info - DOI: https://doi.org/10.54386/jam.v25i3.2194
Received: 14 April 2023; Accepted: 08 June 2023; Published online : 31 August 2023

change has exerting adverse influences on crop yields and productivity. The weather variables *viz.*, precipitation, maximum and minimum temperatures, relative humidity, wind speed etc., are on constant rise in Andhra Pradesh during 1956-2010 (Season and Crop Report, Government of Andhra Pradesh). As agriculture is largely dependent on the above weather parameters, and there is excessive use of natural resource reserves, but with inadequate coping mechanisms, the State of Andhra Pradesh is especially vulnerable to changing climate.

The fluctuations in the above weather variables may exert adverse impact on the groundnut yields. This calls for prediction of groundnut yields in tune with weather variability in different districts of Andhra Pradesh. Although, several researchers have developed statistical and machine learning models for various crops including groundnut , there is a need to compare the performance of well-developed conventional statistical methods and ML approaches in the domain of crop yield due to the fact that the performance of statistical methods differs regionally (spatially), Hence, it is important to study the accuracy of different models for predicting groundnut separately for each study area (Sridhara *et al.,* 2023). Accordingly, different yield forecasting models have been employed in Andhra Pradesh region to enable the farmers and other stakeholders to employ the best-fitted model to utilize weather information in a timely manner for successful planning and groundnut farming decision-making.

MATERIALS AND METHODS

The top five cultivating districts in Andhra Pradesh in terms of area during both Kharif and Rabi seasons *viz.*, Ananthapur, Chittoor, Kadapa, Kurnool and Nellore are purposively selected for this study. These districts together account for around 98 and 93% of annual groundnut area and production respectively in Andhra Pradesh during 2020-21 (Season and Crop Report, 2020-21

The requisite daily weather data of five weather variables *viz.* maximum temperature (^{o}C), minimum temperature (^{o}C), rainfall (mm), relative humidity (%) and wind speed (kmph) were collected from NASA POWER web portal (https://power.larc.nasa.gov/data-access-viewer/) for the period, 2001 to 2020. Recent studies aimed to evaluate the performance of NASA POWER data (Aboelkhair *et al.,* 2019; Rodrigues and Braga, 2021). Those studies showed that there is a significant agreement between NASA POWER reanalysis and observed data for most weather parameters (mostly air temperature and solar radiation). However, it is noteworthy that when the daily weather variables of NASA power data were aggregated and compared with 10-day time scale by Monteiro *et al.,* (2017), a strong improvement of statistical indices was obtained. In this study also, the daily weather data were converted into weekly aggregated variables, which justifies the use of NASA power dataset. Time-Series data on groundnut yields for the selected districts were gathered from various issues of Season and Crop Report, Directorate of Economics and Statistics, Government of Andhra Pradesh. Since the variability in groundnut yield (time-series) data over a long period of time can be driven by factors like changes in technology, climate variability and so on, and this can lead to non-stationary trends, the variability should be reduced before further analysis is carried out. (Sridhara *et al*., 2020; Das *et al.*, 2020). So, the yield statistics were "detrended" before the analysis had been performed. The daily weather readings were used to figure out the weekly average. These weekly average values are then used to figure out weighted and unweighted weather indices. The weighted and unweighted weather indices are computed from the subsequent equations given by Ghosh *et al*., (2014).

$$Z_{ij} = \sum_{w=1}^{n} X_{iw}\ , \qquad Z_{ii'j} = \sum_{w=1}^{n} X_{iw}X_{i'w} \quad (1)$$

$$Z_{ij} = \sum_{w=1}^{n} r_{iw}^{j}\, X_{iw}, \qquad Z_{ii'j} = \sum_{w=1}^{n} r_{ii'w}^{j}\, X_{iw}X_{i'w} \quad (2)$$

Here, n is the week of forecast, is the value of ith/ i'th weather variable and is the value of correlation coefficient of detrended yield with ith weather variables product of ith and i'th weather variables in wth week. The above process is repeated, and 30 weather indices were generated as shown in Table 1. These weather indices alongside the crop harvest data were used to form the forecast models by utilizing various multivariate methods. Out of the total dataset, 75% data are used for calibration purpose, and remaining 25% data are used for the validation of the employed models (Uno *et al*., 2005; Li *et al*., 2017, Montaseri *et al*., 2018). In the present study, to develop all the models and to carry out analysis, SPSS (version IBM SPSS Statistics 20) and R-studio software (version 4.1.2) were used. Details of all models are discussed as follows:

Multivariate techniques employed

***Stepwise multiple linear regression (SMLR)**:* Multiple linear regression (MLR) is the most common and easiest model. But it doesn't always work well for datasets with a large number of independent variables (Balabin *et al.,* 2011). This model helps to select best predicting variable(s) among many independent variables (Singh *et al*., 2014; Das *et al*., 2018; Das *et al*., 2020).

***Least absolute shrinkage and selection operator (LASSO)**:* This method works to lower the Mean Squared Error (MSE) and, as a result, reduces the number of predictors that are needed to create a model. It brings the regression coefficients closer to zero by using

Table 1: Unweighted and weighted weather indices for development of multivariate models

Parameters	Unweighted Weather indices					Weighted Weather indices				
	Tmax	Tmin	RF	RH	WS	Tmax	Tmin	RF	RH	WS
Tmax	Z_{10}					Z_{11}				
Tmin	Z_{120}	Z_{20}				Z_{121}	Z_{21}			
RF	Z_{130}	Z_{230}	Z_{30}			Z_{131}	Z_{231}	Z_{31}		
RH	Z_{140}	Z_{240}	Z_{340}	Z_{40}		Z_{141}	Z_{241}	Z_{341}	Z_{41}	
WS	Z_{150}	Z_{250}	Z_{350}	Z_{450}	Z_{50}	Z_{151}	Z_{251}	Z_{351}	Z_{451}	Z_{51}

a penalty term called L1-norm, which is the sum of the absolute coefficients. For LASSO the loss is defined as:

$$r_{iw}^{j}/r_{i'w}^{j} \quad (3)$$

where, is the independent variable, β is the corresponding coefficient and λ is the L1 norm penalty.

***Ridge regression**:* Ridge regression causes the regression coefficients to shrink so that factors that have a negligible impact on the outcome have their coefficients near to zero. The reduction of the coefficients is accomplished by penalising the regression model with a term known as L2-norm, which is the sum of the squared coefficients (Zou and Hastie, 2005). Here, the loss is defined as:

$$L_1 = \sum(\widehat{Y}_i - Y_i)^2 + \lambda\sum|\beta| \quad (4)$$

where represents the independent variable, β represents the coefficient associated with it, and λ represents the L۲ norm penalty.

***Elastic net (ELNET) regression**:* The ELNET model has features of both LASSO and ridge regressions i.e., it considers both the L1 and L2 norms. (Hoerl and Kennard, 1970). This causes some coefficients to shrink and some coefficients to be set to zero. Therefore, it reduces the impact of various features without eliminating them completely (Cho *et al*., 2009).

$$L_2 = \sum(\widehat{Y}_i - Y_i)^2 + \lambda\sum\beta^2 \quad (5)$$

where, represents the independent variable, β represents the corresponding coefficient and λ represents the penalty.

***Artificial Neural Network (ANN)**:* The way in which ANN operates is quite comparable to the way in which the biological neural process of the human brain operates. ANN is a machine learning technique that has three layers: input, hidden and output. The input layer serves as a source for data, that is then passed on to the output layer via a hidden layer (Kaul *et al*., 2005). The number of input layer nodes depends on the number of independent predictors.

Evaluation matrices

Statistical measures such as Mean Biased Error (MBE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (nRMSE), and Evaluation Factor (EF) are employed to select the most appropriate model. The following are the formulas for the above-mentioned metrics:

$$R^2 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y}_i)(\hat{y}_i-\hat{\bar{y}}_i)}{\sigma_y\sigma_{\hat{y}}}\right)^2,$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}},$$

$$\text{nRMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} * \frac{100}{\bar{y}_i},$$

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i),$$

$$\text{EF} = \left[1 - \frac{\sum_{1=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}\right]$$

Here, is the actual value and is the predicted value for i=1, 2,……, n. and are the standard deviation of observed and estimated produces respectively. and is the mean of observed and estimated produces respectively. The value of R^2 and EF close to 1 and the value of RMSE near ‘0’ denote better model performance. The developed models were compared based on the value of R^2, as $R^2 > 0.90$, excellent, R^2 = 0.90-0.75, good, R^2 = 0.75-0.50, fair and $R^2 < 0.50$, poor (Terra et al., 2015). The model is judged in terms of excellent, good, fair and poor based on the value of nRMSE that remains between 0-10%,10-20%, 20-30% and >30%, respectively. Moreover, positive value of MBE shows over estimation and negative value indicates underestimation of the fitted model.

RESULTS AND DISCUSSION

SMLR

The findings from SMLR (Table 2) at calibration stage revealed that R^2 values ranged between 0.681 (Kadapa) to 0.921 (Kurnool) during Kharif season and between 0.592 (Chittoor) to 0.934 (Kurnool) during Rabi season. The R^2 and EF values are higher for Kurnool and Nellore districts during both Kharif and Rabi seasons. Similarly, for both these districts, during Kharif season, the RMSE values (0.018 t ha^{-1} and 0.004 t ha^{-1} respectively) are minimum and nRMSE values (4.5% and 3.3% respectively) showed ‘excellent’ performance. In Rabi season, RMSE values are minimum for Nellore (0.011 t ha^{-1}) and Kurnool (0.027 t ha^{-1}) and nRMSE for these two districts again showed ‘excellent’ performance (< 10%). At calibration stage, the model performance is ‘fair’ in Kharif season (nRMSE ranged between 20-30%) and ‘poor’ in Rabi season (nRMSE > 30%) for Ananthapur, Chittoor and Kadapa districts. Even during validation stage, similar findings were perceived and performance is found to be ‘excellent’ for Kurnool and Nellore districts (nRMSE < 10%) during both Kharif and Rabi season, unlike other districts, (nRMSE > 30%). The MBE values for both Kharif and Rabi seasons, at calibration stage showed little over estimation (0.072 to 0.684 in Kharif season and 0.108 to 0.496 in Rabi season) and at validation stage showed little under estimation (-0.692 to -0.114 in Kharif season and -0.516 to -0.114 in Rabi season).

Ridge regression

Regarding the prediction of groundnut yields during Kharif season, at calibration stage (Table 3), maximum R^2 was noticed for Ananthapur district (0.961) by RMSE of 0.243 t ha^{-1} followed by Chittoor (0.818 & 0.186 t ha^{-1} respectively) and Kadapa (0.815 & 0.133 t ha^{-1} respectively) districts, while the least R^2 was recorded for Nellore (0.446) with RMSE of 0.255 t ha^{-1}. The EF values range from 0.712 (Kadapa) to 0.892 (Ananthapur). During validation, highest R^2 was observed for Kadapa (0.916) with RMSE of 0.543 t ha^{-1}, and the lowest R^2 was recorded for Nellore (0.526) with RMSE of 0.615 t ha^{-1}. Further, the nRMSE statistic showed that the ridge regression model performance was found to be ‘good’ for Ananthapur (10.8%), Chittoor (10.7%) and Kadapa (18.1%), ‘fair’ for Kurnool (28.1%), while model performed ‘poor’ for Nellore (43.8%). However, regardless of good model performance at the calibration phase, the MBE values for all the regions showed underestimation of the groundnut produce at the validation phase.

Table 2: Results obtained through SMLR model

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF_{cal}	R^2_{val}	$RMSE^2_{val}$	$nRMSE_{val}$	MBE_{val}
Kharif									
Ananthapur.	0.745	0.176	24.832	0.072	-1.691	0.191	0.452	57.041	-0.298
Chittoor	0.832	0.161	21.279	0.269	0.009	0.127	0.627	42.291	-0.259
Kadapa	0.681	0.267	20.268	0.684	-2.628	0.239	0.298	64.214	-0.483
Kurnool	0.921	0.018	4.482	0.267	0.892	0.821	0.059	9.328	-0.371
Nellore	0.917	0.004	3.297	0.182	0.907	0.863	0.026	8.219	-0.692
Rabi									
Ananthapur.	0.814	0.189	38.158	0.185	-1.529	0.247	0.235	42.158	-0.314
Chittoor	0.592	0.251	34.153	0.496	-1.812	0.317	0.472	41.219	-0.329
Kadapa	0.621	0.176	61.248	0.158	-1.296	0.527	0.661	50.183	-0.516
Kurnool	0.934	0.027	9.962	0.216	0.801	0.924	0.042	9.821	-0.215
Nellore	0.905	0.011	4.210	0.108	0.846	0.915	0.019	5.127	-0.114

Table 3: Results obtained through Ridge regression model

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF_{cal}	R^2_{val}	$RMSE^2_{val}$	$nRMSE_{val}$	MBE_{val}
Kharif									
Ananthapur.	0.961	0.243	6.695	-7.394e-17	0.892	0.902	0.405	10.751	-0.649
Chittoor	0.818	0.186	7.418	-6.162e-17	0.790	0.832	0.422	10.711	-0.717
Kadapa	0.815	0.133	7.891	-4.581e-17	0.712	0.916	0.543	18.096	-0.523
Kurnool	0.521	0.132	6.231	-0.312e-17	0.736	0.617	0.713	28.138	-0.442
Nellore	0.446	0.255	7.982	-2.162e-17	0.807	0.526	0.615	43.828	-0.671
Rabi									
Ananthapur.	0.825	0.222	7.294	-6.325e-17	0.775	0.910	0.459	8.658	-0.446
Chittoor	0.821	0.129	5.486	-5.367e-17	0.793	0.821	0.129	9.160	-0.939
Kadapa	0.780	0.257	6.882	-3.218e-17	0.790	0.780	0.257	7.849	-0.897
Kurnool	0.591	0.345	8.771	-2.002e-17	0.629	0.697	0.909	9.362	-0.857
Nellore	0.549	0.171	9.466	-3.027e-17	0.577	0.741	0.542	9.727	-0.521

Regarding Rabi season, similar findings are obtained regarding R^2, RMSE and nRMSE in the calibration stage. However, in the validation stage, R^2 is highest (0.910) and RMSE value (0.459 t ha^{-1}) was lowest for Ananthapur followed by Chittoor with R^2 of 0.821 and RMSE of 0.129 t ha^{-1}; and Kadapa with R^2 of 0.780 and RMSE of 0.257 t ha^{-1}. It is interesting that the ridge regression model performance was 'excellent' across all the selected districts with nRMSE values of less than 10 %. The EF ranged from 0.577 for Nellore to 0.793 for Chittoor district. Even during Rabi season, the MBE values for all the locations showed underestimation of the crop yield at the validation phase compared to the calibration phase.

LASSO regression

For Kharif season (Table 4), the findings from LASSO regression at calibration stage revealed that the value of R^2 was highest for Nellore (0.931) followed by Kurnool (0.929) and lowest for Chittoor (0.901). Again, Nellore observed with lowest RMSE value of 0.119 t ha^{-1} followed by Kurnool (0.195 t ha^{-1}) and Chittoor (0.211 t ha^{-1}). During validation phase, the value of R^2 ranged from 0.431 to 0.906. The maximum value of R^2 was again observed for Kurnool (0.906) with RMSE value of 0.128 t ha^{-1}, while the minimum R^2 was observed for Ananthapur (0.431) with RMSE of 0.365 t ha^{-1}. The nRMSE values at calibration phase showed that the model performance was found to be 'excellent' for all the selected regions. On the contrary, at the validation stage, performance of LASSO model was 'excellent' for Kurnool (7.2%) and Nellore (8.4%), 'good' for Chittoor (19.7%), 'fair' for Ananthapur (20.2%) and 'poor' for Kadapa (39.1%) district. The EF ranged from 0.811 for Kadapa to 0.918 for Nellore. Though the MBE values for all the locations indicated reliable model performance at calibration phase, it resulted in under estimation during validation phase for groundnut yield in all the districts, except Ananthapur. Even during Rabi season, the findings from LASSO regression are almost similar both at calibration and validation stages. The values pertaining to R^2, RMSE, nRMSE and EF are more satisfactory for Kurnool and Nellore compared to other districts. Again, compared to calibration stage, the model resulted in underestimation during validation stage for all the districts.

Table 4: Results obtained through LASSO model

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF_{cal}	R^2_{val}	$RMSE^2_{val}$	$nRMSE_{val}$	MBE_{val}
				Kharif					
Ananthapur	0.916	0.246	6.695	1.207e-17	0.856	0.431	0.365	20.171	0.001
Chittoor	0.901	0.211	4.418	4.627e-17	0.889	0.711	0.484	19.711	-0.025
Kadapa	0.920	0.338	3.891	2.207e-17	0.811	0.612	0.243	39.096	-0.323
Kurnool	0.929	0.195	2.143	0.116e-17	0.913	0.906	0.128	7.153	-0.112
Nellore	0.931	0.119	3.982	3.201e-17	0.918	0.892	0.126	8.434	-0.221
				Rabi					
Ananthapur.	0.902	0.403	8.234	1.108e-17	0.893	0.810	0.691	28.658	-0.446
Chittoor	0.756	0.516	5.486	3.112e-17	0.799	0.763	0.758	14.160	-0.339
Kadapa	0.680	0.317	2.882	4.261e-17	0.716	0.623	0.882	27.849	-0.297
Kurnool	0.913	0.115	3.571	6.362e-17	0.919	٠,٨٤١	0.209	3.362	-0.157
Nellore	0.821	0.165	3.466	6.226e-17	0.902	0.836	0.242	3.727	-0.321

Table 5: Results obtained through ELNET model

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF_{cal}	R^2_{val}	$RMSE^2_{val}$	$nRMSE_{val}$	MBE_{val}
				Kharif					
Ananthapur.	0.949	0.101	2.584	8.716e-16	0.940	0.914	0.738	16.481	-0.718
Chittoor	0.928	0.113	1.096	2.526e-16	0.887	0.817	0.474	14.070	-0.214
Kadapa	0.910	0.133	3.891	1.329e-16	0.915	0.864	1.343	16.096	-0.596
Kurnool	0.605	0.314	6.157	3.824e-16	0.889	0.679	0.679	23.563	-0.327
Nellore	0.913	0.289	5.411	2.175e-16	0.871	0.596	1.072	43.828	-0.269
				Rabi					
Ananthapur.	0.916	0.186	1.626	1.596e-16	0.924	0.873	0.932	11.682	-0.783
Chittoor	0.892	0.138	1.249	4.528e-16	0.826	0.726	0.629	17.256	-0.326
Kadapa	0.880	0.257	2.671	3.697e-16	0.815	0.802	0.982	11.319	-0.417
Kurnool	0.767	0.306	5.327	2.158e-16	0.843	0.507	0.846	46.297	-0.617
Nellore	0.819	0.328	4.196	1.249e-16	0.637	0.613	0.913	36.297	-0.315

ELNET model

From Table 5, the findings from ELNET model revealed that values of R^2, RMSE, nRMSE and EF are satisfactory for Ananthapur, Chittoor and Kadapa compared to Kurnool and Nellore districts during Kharif season. For calibration dataset, the results showed that highest and lowest R^2 were recorded for Ananthapur (0.949) and Kurnool (0.605) districts respectively. The RMSE values ranged between 0.101 t ha^{-1} (Ananthapur) to 0.314 t ha^{-1} (Kurnool) and nRMSE ranged between 1.1 % (Chittoor) to 6.2 % (Kurnool). During the validation stage, Ananthapur again recorded highest R^2 value (0.914) and lowest value was recorded for Nellore (0.596). The RMSE ranged from 0.474 t ha^{-1} (Chittoor) to 1.343 t ha^{-1} (Kadapa) and the values of nRMSE ranged between 14.1 % (Chittoor) to 43.8 % (Nellore). The EF is near to 1 for all the locations indicating 'good' model performance. From the MBE values, it can be inferred that the model performance is underestimated at the validation stage compared to calibration stage across all the districts. In Rabi season, similar findings are obtained for Ananthapur, Chittoor and Kadapa districts with respect to R^2, RMSE, nRMSE and EF across both calibration and validation stages. Once again, the fitted model is found efficient (EF), as the values approach close to 1 across all the districts, except for Nellore (0.637) in calibration stage. Further, the fitted model is underestimated at the validation stage compared to calibration stage for all the selected districts.

ANN model

The results from ANN model (Table 6) during Kharif season are found satisfactory for Ananthapur, Chittoor and Kadapa compared to other two districts in terms of R^2, RMSE, nRMSE and EF values. The model prediction for Ananthapur and Chittoor districts revealed higher R^2 values of 0.911 and 0.903 respectively and with RMSEs of 0.160 t ha^{-1} and 0.172 t ha^{-1} respectively for calibration dataset. The values of nRMSE for all the districts are less than 10 %, suggesting 'excellent' model performance. Similarly, the EF ranged from 0.811 for Nellore to 0.945 for Ananthapur. As the MBE values are near to zero for all the locations, the model showed 'excellent' performance during calibration stage, unlike validation phase for the estimation of groundnut produce in all the districts. For Rabi season, similar findings are noticed in terms of R^2, RMSE, nRMSE and EF values for Ananthapur, Chittoor and

Table 6: Results obtained through ANN model

Districts	R^2_{cal}	$RMSE_{cal}$	$nRMSE_{cal}$	MBE_{cal}	EF_{cal}	R^2_{val}	$RMSE^2_{val}$	$nRMSE_{val}$	MBE_{val}
Kharif									
Ananthapur.	0.911	0.160	5.469	0.081	0.945	0.926	0.128	7.240	-0.416
Chittoor	0.903	0.172	1.782	0.042	0.926	0.816	0.199	3.438	-0.339
Kadapa	0.852	0.184	9.747	0.033	0.847	0.912	0.086	2.456	-0.192
Kurnool	0.710	0.469	7.236	0.011	0.813	0.637	0.358	12.436	-0.287
Nellore	0.741	0.389	4.411	0.007	0.811	0.561	0.672	23.828	-1.071
Rabi									
Ananthapur.	0.893	0.183	4.383	0.024	0.952	0.893	0.131	3.243	0.893
Chittoor	0.769	0.104	8.403	0.079	0.945	0.916	0.173	7.785	0.769
Kadapa	0.880	0.157	2.882	0.019	0.890	0.823	0.282	7.849	0.880
Kurnool	0.656	0.587	12.563	0.236	0.879	0.647	0.795	15.516	0.656
Nellore	0.690	0.466	13.651	0.279	0.812	0.511	0.623	16.657	0.690

Table 7: Cross comparison of developed models based on R^2 values during calibration

Districts	SMLR	Ridge	LASSO	ELNET	ANN
Kharif season					
Ananthapur	Fair	Excellent	Excellent	Excellent	Excellent
Chittoor	Good	Good	Excellent	Excellent	Excellent
Kadapa	Fair	Good	Excellent	Excellent	Good
Kurnool	Excellent	Fair	Excellent	Fair	Fair
Nellore	Excellent	Poor	Excellent	Excellent	Fair
Rabi season					
Ananthapur	Good	Good	Excellent	Excellent	Good
Chittoor	Fair	Good	Good	Good	Good
Kadapa	Fair	Good	Fair	Good	Good
Kurnool	Excellent	Fair	Excellent	Good	Fair
Nellore	Excellent	Fair	Good	Good	Fair

Kadapa districts, unlike other two districts. Both for calibration and validation datasets, the values of nRMSE are less than 10 % for the above three districts indicating excellent model performance, unlike Kurnool and Nellore. Further, the model found underestimated in the validation stage compared to calibration stage for all the selected districts in terms of MBE values.

A cross comparison of all models for all locations during both seasons based on R^2 values was made and represented in Table 7. Based on the R^2 values during calibration stage LASSO and ELNET were among the best performing models. The performance of LASSO during calibration stage was found to be excellent for all the districts during kharif season, while during rabi season it was excellent for Ananthapur and Kurnool. The performance of the ELNET model was also excellent and good during both season for all location except Kurnool in kharif season, where it was fair., It can be concluded from the present findings that ANN, ELNET and LASSO regression were the best models for Ananthapur, Chittoor and Kadapa districts compared to SMLR and Ridge regression models. So, it can be concluded that these models are potential tool for forecasting groundnut yields in both Kharif and Rabi seasons in Ananthapur, Chittoor and Kadapa districts. These findings are similar to Sridhara *et al.,* (2020) where ANN was best model for yield prediction with a good model fit, efficacy, and lower error values. For Kurnool and Nellore districts during both Kharif and Rabi seasons in Andhra Pradesh SMLR and LASSO performed excellent, while performances of other models were good to poor. These findings are also in line with the findings of Setiya *et al.,* (2022), Satpathi *et al.,* (2023) and Aravind *et al.,* (2022), as these the studies concluded that ANN, LASSO and ELNET models achieved better yield prediction accuracy compared to conventional models. It is worth noticing that the performance of the models can vary according to the datasets used and locations. Hence, it is not necessary for a model to always perform best for all datasets and locations.

CONCLUSION

Findings from this study revealed that LASOO, ELNET and ANN models based on the values of R^2, RMSE, nRMSE and EF showed better performance for Ananthapur, Chittoor and Kadapa districts during both Kharif and Rabi seasons. On the contrary, in terms of above parameters, SMLR and LASSO models performed better for Kurnool and Nellore districts during both Kharif and Rabi seasons. So, these models should be employed accordingly for forecasting the groundnut yields in both the seasons. Overall, according to the performance the selected models may be ranked in the order of ANN>ELNET>LASSO>Ridge>SMLR across both Kharif and Rabi seasons

ACKNOWLEDGEMENT

Authors are thankful to NASA (National Aeronautics and Space Administration) for providing weather data and Directorate of Economics and Statistics, Government of Andhra Pradesh for providing historical data on groundnut yield. Authors thank the Acharya N.G. Ranga Agricultural University and G.B. Pant University of Agriculture and Technology for providing the facility throughout the research.

Funding: No external funding is involved

Conflict of Interests: The authors declare that there is no conflict of interest related to this article.

Data availability: To be provided on request.

Authors contribution: **K. N. Ravi Kumar**: Investigation, Formal analysis, Writing-original draft; **A. Satpathi**: Writing-review and editing; **M. J. M.Reddy**: Visualization, Editing; **P. Setiya**: Conceptualization, Methodology; **A. S. Nain**: Supervision, Editing.

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

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

REFERENCES

Aboelkhair, H., Morsy, M. and El Afandi, G. (2019). Assessment of agroclimatology NASA POWER reanalysis datasets for temperature types and relative humidity at 2 m against ground observations over Egypt. *Adv. Space Res.*, 64(1): 129-142.

Agricultural Statistics at a Glance, (2020-21). Government of India. Available online: https://eands.dacnet.nic.in/PDF/Agricultural%20Statistics%20at%20a%20Glance%20-%202021%20(English%20version).pdf (Accessed on 12 April 2023).

ANGRAU (Acharya N.G. Ranga Agricultural University), Groundnut Outlook Report. January to December, 2021. Crop Outlook Reports of Andhra Pradesh. Available online: https://angrau.ac.in (Accessed on 8 March 2023).

Aravind, K., Vashisth, A., Krishanan, P. and Das, B. (2022). Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *J. Agrometeorol.,* 24(1): 18-25. https://doi.org/10.54386/jam.v24i1.1002

Balabin, R. M., Lomakina, E. I. and Safieva, R. Z. (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(5): 2007-2015.

Cho, S., Kim, H., Oh, S., Kim, K. and Park, T. (2009). Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. *BMC Proceed.,* 3(7)**:** 1-6

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

Das, B., Nair, B., Arunachalam, V., Reddy, K. V., Venkatesh, P., Chakraborty, D. and Desai, S. (2020). Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India. *Int. J. Biometeorol*., 1-13.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K.K. and Rathore, L.S. (2014). Development of crop yield forecast models under FASAL a case study of *kharif* rice in West Bengal. *J. Agrometeorol.,* 16(1):1-8. https://doi.org/10.54386/jam.v16i1.1479

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics.*, 12(1):55-67.

Jasna, V. K., Som, S., Burman, R. R., Padaria, R. N. and Sharma, J. P. (2014). Socio economic impact of climate resilient technologies. *Int. J. Agric. and Food Sci. Technol.*, 5(3): 185-190.

Kaul, M., Hill, R. L. and Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agric. Syst*., 85(1): 1-18.

Li, X., Sha, J. and Wang, Z.L. (2017). Chlorophyll-A Prediction of lakes with different water quality patterns in China based on hybrid neural networks. *Water.*, 9: 524.

Montaseri, M., Zaman Zad Ghavidel, S. and Sanikhani, H. (2018). Water quality variations in different climates of Iran: Toward modeling total dissolved solid using soft computing techniques. *Stoch. Environ. Res. Risk Assess*., 32: 2253–2273.

Monteiro, L. A., Sentelhas, P. C. and Pedra, G. U. (2017). Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation. *Int. J. Climatol.*, 38(3): 1571-1581.

Rodrigues, G. C. and Braga, R. P. (2021). Evaluation of NASA POWER reanalysis products to estimate daily weather variables in a hot summer Mediterranean climate. *Agron.*, 11(6): 1207.

Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S. and Singh, S. (2023). Comparative Analysis of Statistical and

Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. *Sustain.*, 15(3): 2786.

Saravanakumar, V., Lohano, H. D. and Balasubramanian, R. (2022). A district-level analysis for measuring the effects of climate change on production of rice: evidence from Southern India. *Theor. Appl. Climatol.*, 150(3-4): 941-953.

Season and Crop Report, various issues, Directorate of Economics and Statistics, Government of Andhra Pradesh. Available online: https://des.ap.gov.in/jsp/social/SEASONANDCROP REPORT202021.pdf (Accessed on 12 April 2023)

Setiya, P., Satpathi, A., Nain, A. S., and Das, B. (2022). Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning techniques. *J. Agrometeorol.*, 24(3): 255-261. https://doi.org/10.54386/jam.v24i3.1571

Singh, A. K., Vashisth, A., Sehgal, V. K., Goyal, A., Pathak, H. and Parihar, S. (2014). Development of Multi Stage District Level Wheat Yield Forecast Models. *J. Agric. Phys.,* 14(2): 189-193.

Sridhara, S., Ramesh, N., Gopakkali, P., Das, B., Venkatappa, S. D., Sanjivaiah, S. H. and Elansary, H. O. (2020). Weather Based Neural Network, Stepwise Linear and Sparse Regression Approach for Rabi Sorghum Yield Forecasting of Karnataka, India. *Agron.*, 10(11): 1645.

Sridhara, S., Manoj, K. N., Gopakkali, P., Kashyap, G. R., Das, B., Singh, K. K. and Srivastava, A. K. (2023). Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. *Int. J. Biometeorol*., 67(1): 165-180.

Terra, F. S., Demattê, J. A. and Rossel, R. A. V. (2015). Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. *Geoderma,* 255: 81-93.

Uno, Y., Prasher, S. O., Lacroix, R., Goel, P. K., Karimi, Y., Viau, A. and Patel, R. M. (2005). Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Comput. Electron. Agric.*, 47:149–161.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *J. R. Statist. Soc. B.,* 67 (2): 301–305.