The Pharma Innovation Journal 2021; 10(12): 963-966

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; 10(12): 963-966

www.thepharmajournal.com
Received: 13-09-2021
Accepted: 19-10-2021

Abhinaya D
Research Scholar, Department of Physical Sciences & Information Technology, AEC & RI, TNAU, Coimbatore, Tamil Nadu

Patil SG
Assistant Professor, Department of Physical Sciences & Information Technology, AEC & RI, TNAU, Coimbatore, Tamil Nadu

Djanaguiraman M
Assistant Professor, Department of Crop Physiology, AC & RI, TNAU, Coimbatore, Tamil Nadu

Gunasekaran
Associate Professor, Department of Mathematics, PSG College of Arts and Science, Coimbatore, Tamil Nadu

Corresponding Author:
Abhinaya D
Research Scholar, Department of Physical Sciences & Information Technology, AEC & RI, TNAU, Coimbatore, Tamil Nadu

Prediction of groundnut yield using principal component analysis of weather parameters

Abhinaya D, Patil SG, Djanaguiraman M and Gunasekaran

Abstract
The use of principal component analysis in the development of statistical models for crop yield forecasting has been demonstrated. This study employed time series data on groundnut yield and weekly weather variables, such as minimum and maximum temperature, relative humidity, wind speed, and rainfall, from 1990 to 2019 during kharif season in the Erode district of Tamil Nadu. Weekly data on weather variables was used to create weather indices (Agrawal *et al*., 1983). Four models were created with principal component analysis as Independant variables which also includes time trend and groundnut yield as dependant variable. The model performance was measured using Adjusted R-squared (adj R^2) and Root Mean Squared Error (RMSE) as goodness-of-fit criteria. On the basis of adj R^2 and RMSE, model 1 which includes all the calculated weather indices, was found to be the best suited model with high adj R^2 (65.51%) and least RMSE (254.7343). Hence, this model can be used to forecast groundnut yield for the studied region.

Keywords: Groundnut yield prediction, principal component analysis, weather indices

Introduction
Crop yield prediction is a critical responsibility for decision-makers at the national and regional levels who needs to make quick decisions. Farmers may use an accurate crop yield prediction model to assist them determine what to produce and when to grow it. Yield prediction also helps growers and farmers make better managerial and financial decisions. To increase national food security, policymakers rely on accurate estimates to make timely import and export choices.
Technological advancements and weather fluctuation have an impact on crop productivity. Technological variables improve yield gradually over time, hence years or other time units can be used to investigate the total influence of technology on yield. Crop yield prediction is one of precision agriculture's most difficult challenges, and numerous models have been suggested and validated so far. Among them the notable ones were the regression models developed by Fisher (1924) [7], Hendricks and Scholl (1943), Agrawal *et al*. (1980, 83, 86 & 2001) [1, 3, 4, 5], Jain and Singh (1980) [1] and application of discriminant function analysis of weather indices attempted by Rai and Chandrahas (2000), Agrawal *et al*. (2012) [5]. Forecast models based on principal components of biometrical characters have also been developed by Aneja and Chandrahas (1984), Chandrahas and Prem Narain (1993), Jain *et al*. (1985) [8]. Application principal component analysis using weather indices for forecasting crop yield had been done by Yadav *et al*., (2014) [6]. Hence, in the present paper, an attempt has been made to develop suitable statistical models for forecasting of groundnut yield during kharif season in Erode district of Tamil Nadu using principal component analysis of weather indices of weather variables.

Materials and Methods
Data Collection and Study area
The study has been administered for Erode district of Tamil Nadu, India, which is situated at 11.5246° N latitude / 77.4702° E longitudes. Time series data of groundnut yield (*Arachis hypogea* L) for Erode district of Tamil Nadu over a span of 29 years (1991 to 2019) has been collected from Season and Crop Report and Department of Economics and Statistics (DES). Daily weather data were collected from Agro Climate Research Centre, TNAU. The data on five weather variables namely maximum temperature (Tmax, ºC), minimum temperature (Tmin, ºC), relative humidity (RH (%)), windspeed (kph), and rainfall (mm) for a total of 18 standard meteorological week (SMW) has been used in the study.

Daily data of weather parameters had been converted into its weekly average whereas weekly sum of rainfall has been considered. Out of the 29 year data, 25 years were used as training dataset while the remaining 4 years were used as testing dataset.

Statistical Methodology
The whole 18-week data from the 23rd SMW to the 41st SMW of a year was used to create weighted and unweighted weather indices of weather variables, as well as their interactions. The weighted indices are a weekly weighted average of the weather variables, with the weights being the correlation coefficients between the de-trended yield and the respective weather data. The weather variables are simply averaged throughout the weeks to get the unweighted indexes. Similarly, using the product of weather variables, the unweighted and weighted indices of interactions between weather variables were generated (taking two at a time). In all, 30 indices (15 weighted and 15 unweighted) were generated, including 5 weighted weather indices and 10 weighted interaction indices, as well as 5 unweighted weather indices and 10 unweighted interaction indices. The following formula was used to calculate these weather indices and interaction indices.

$$Z_{ij} = \frac{\sum_{w=1}^{n} r_{iw}^{j} x_{iw}}{\sum_{w=1}^{n} r_{iw}^{j}} \quad Z_{ii'j} = \frac{\sum_{w=1}^{n} r_{ii'w}^{j} x_{iw} x_{i'w}}{\sum_{w=1}^{n} r_{ii'w}^{j}}$$

Where Z_{ij} is unweighted (for j = 0) and weighted (for j = 1) weather indices for i^{th} weather variable and $Z_{ii'j}$ is the un-weighted (for j = 0) and weighted (for j = 1) weather indices for interaction between i th and ii'th weather variables. X_{iw} is the value of the i th weather variable in w th week, $r_{iw}/r_{ii'w}$ is correlation coefficient between yield adjusted for trend effect and value of i th weather variable/product of i th and ii'th weather variable in w th week, n is the number of weeks considered in developing the indices and p is number of weather variables used.

Model 1: All 30 indices were employed in principal component analysis in this approach, and the first five principal components were chosen as regressors in the development of the forecasting model since they explained 94.58% of total variance. The form of the model fitted is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \beta_3 PC_3 + \beta_4 PC_4 + \beta_5 PC_5 + \beta_6 T + e \quad (1)$$

Where, y is de-trended crop yield, β_i's (i = 0, 1, 2,..,6) are model parameters; pc1, pc2, …, pc5 are first five principal components, T is the trend variable and e is error term assumed to follow independently N (0, σ2).

Model 2: Five weighted and unweighted weather indices of five weather variables were employed in this technique. The first four principal components were determined as the most important in terms of loading and explained almost 89.30% of the overall variance using principal component analysis. As a result, the first three major components were employed as explanatory variables in the building of the forecasting model. The form of model fitted is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \beta_3 PC_3 + \beta_4 PC_4 + \beta_5 T + e \quad (2)$$

Model 3: Five unweighted weather indicators and ten unweighted interactions were employed in this technique. The forecasting model used the first four major components as regressors since they explained 93.70% of total variance. The form of the model fitted is as follows:

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \beta_3 PC_3 + \beta_4 PC_4 + \beta_5 T + e \quad (3)$$

Model 4: Five weighted weather indices and ten weighted interactions were employed in this technique. Since the first four main components explained around 91.70% of overall variation, they were employed as predictors in the forecasting model. The form of the model fitted is as follows

$$Y = \beta_0 + \beta_1 PC_1 + \beta_2 PC_2 + \beta_3 PC_3 + \beta_4 PC_4 + \beta_5 T + e \quad (4)$$

Model Performance Metrics
Coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used to evaluate the performance of the generated statistical models. They were calculated using the following formula:

$$R^2 = \frac{\text{Regression sum of square}}{\text{Total sum of square}} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{\bar{Y}}_i)^2} \quad (0 \le R^2 \le 1) \ (5)$$

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2]^{\frac{1}{2}} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (7)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (8)$$

Y_i = actual value
$\hat{Y}_I$= Model output
R^2 towards 1 and RMSE towards 0 indicates better performance of the developed models. Also lesser the MAE and MAPE values, better fit the model is.

Results and Discussion
The descriptive statistics of the weather data are presented in the Table 1. The association between the weather variables and yield was studied using Karl Pearson coefficient of correlation and it was shown in Fig.1. It is observed that a strong positive and significant correlation exists between maximum and minimum temperature and a negative significant correlation was observed between maximum temperature and rainfall. The groundnut yield was observed to be positively associated with wind speed and rainfall, whereas negative relationship exists between yield and variables such as relative humidity, maximum and minimum temperature. The yield forecasting models developed using the principal component analysis of weather indices were presented in Table 2 along with their values of R^2 adj. In model 1, except fourth principal component, all others including time trend (T) have shown significant effect on groundnut yield.

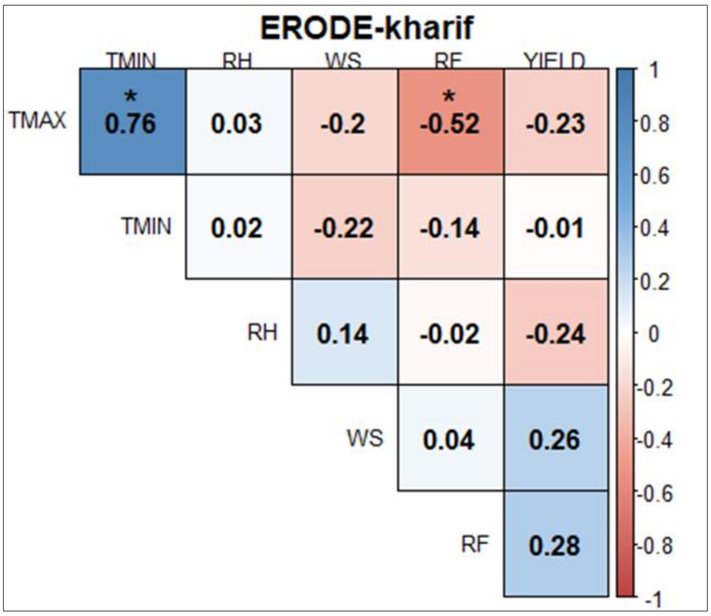


Fig 1: Correlation plot of studied Weather variables with groundnut yield

Table 1: Descriptive statistics of weather parameters

Parameters	Mean	Maximum	Minimum	SD	CV
Tmax (^{0}C)	30.83	34.98	27.59	1.15	3.75
Tmin (^{0}C)	21.82	25.50	20.05	0.61	2.78
RH (%)	65.40	83.60	46.05	4.49	6.91
RF (mm)	21.77	117.37	0.00	15.06	70.26
WS (kph)	3.17	7.80	0.91	0.93	29.02

Only first and second principal component has shown significant effect in model 2. In model 3, only first principal components have shown significant effect while first and third principal components along with time trend (T) have shown significant effects on groundnut yield in model 4. The value of R^2 adj has been found to be maximum of about 65% in model 1 followed by about 60.42% in model 2. Using these forecast models the forecast values of groundnut yield for the years 2015-16, 2016-17,2017-18 and 2018- 19 were obtained and the results are presented in Table 3.

Table 2: Groundnut yield forecast models

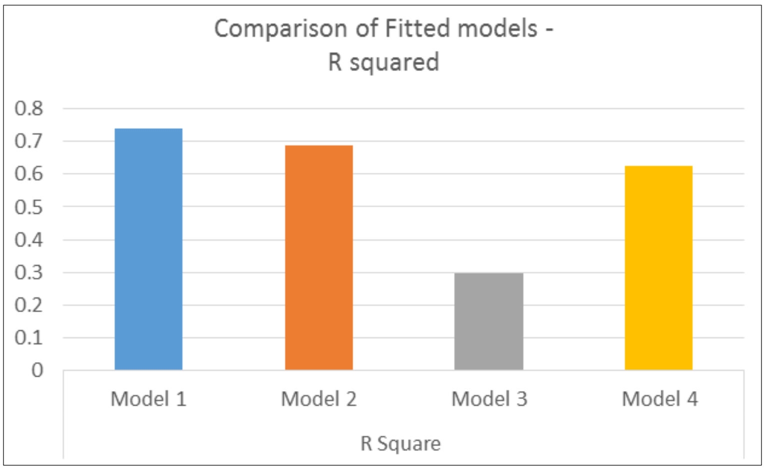
Model	Forecast equation	R^2	R^2_{adj}
1.	Yield = 537.99 -108.67 pc1 + 50.78 pc2 + 100.23 pc3 -10.63 pc4 -229.79 pc5 -36.96 T (**) (***) (.) (**) () (***) (**)	0.7413	0.6551
2.	Yield = 263.80+ 145.34 pc1 -224.51pc2 + 41.11 pc3 + 18.47pc4 -15.70T (***) (***)	0.6866	0.6042
3.	Yield = 154.388+ 77.939 pc1 -73.122 pc2 + 102.548 pc3 -19.056 pc4 -7.174T (.)	0.2978	0.1131
4.	Yield = 417.107 + 171.816 pc1 -4.136 pc2 + 103.096 pc3 -16.170 pc4 -26.221T (.) (***) (.) (***)	0.6254	0.5268

Note: Symbols in brackets denotes significant at '***' $P< 0.001$, '**' $P< 0.01$, '*'$P< 0.05$, '**.**' $P< 0.1$.

The model 1 is found to be with lower RMSE (254.7343) and high adj R^2 (0.6551) followed by model 2 (0.6042) (Fig. 2). On the other hand, model 3 performs poorly in comparison with other models. The Tables (2 and 3) reveal that the model 1 is the most appropriate one followed by the model 2 for the forecast of the groundnut yield during kharif season in Erode district of Tamil Nadu which is in accordance with the models developed by applying principal component analysis on weather variables for wheat yield forecasting by Yadav *et al.*, (2014) [6]. Hence, the model 1 and model 2 are recommended for groundnut yield forecasting of the studied region.

Table 3: Actual and Predicted groundnut yield

Year	Actual yield	Forecasted yield			
		Model 1	Model 2	Model 3	Model 4
2015-16	710	856	881	1436	934
2016-17	2067	1934	2244	1996	2157
2017-18	1995	1942	2105	2281	2251
2018-19	1557	1121	1919	1972	1677
RMSE (Kg/ha)		254.7343	260.3689	404.0041	278.1985



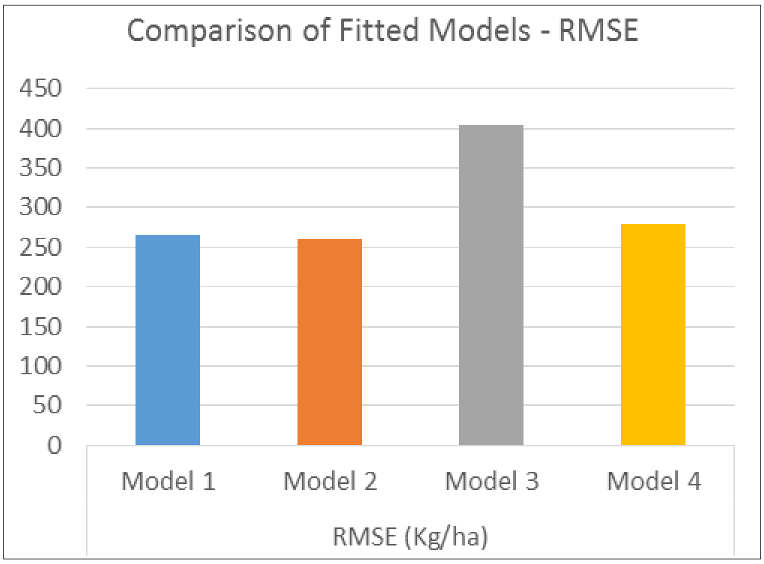


Fig 2: Comparison of the developed statistical models using R^2 and MAPE values

Conclusion

Application of Principal Component Analysis (PCA) on weather variables for predicting crop yield produces improved results since PCA converts the set of correlated variables into non-correlated components. In our study, the PCA model including all the weighted and unweighted indices performs better comparing to models containing either weighted or unweighted indices. Hence, the model 1 which utilizes the principal components of both weighted and unweighted indices as regressors can be used to forecast the groundnut yield during kharif season for Erode district of Tamil Nadu.

Acknowledgement

The required weather data and groundnut yield data were provided by the Agro Climate Research Centre, TNAU, and the Department of Economics and Statistics, Tamil Nadu and are duly acknowledged.

References

1. Agrawal R, Jain RC, Jha MP. Forecasting of rice yield using climatic variables, Indian Journal of Agricultural Science 1980;50(9):680-684.
2. Agarwal R, Jain RC, Jha MP, Singh D. Joint effects of weather variables on rice yield, Mausam, 1983;34(2):189-194.
3. Agrawal R, Jain RC, Jha MP. Models for studying rice crop-weather relationship, Mausam, 1986;37(1):67-70.
4. Agrawal R, Jain RC, Mehta SC. Yield forecast based on weather variables and agricultural inputs on agro-climatic zone basis. Indian Journal of Agricultural Science. 2001;71(7):487-90.
5. Agrawal R, Aditya K. Use of discriminant function analysis for forecasting crop yield., Mausam 2012;63(3):455-458.
6. Yadav RR, Sisodia BV, Kumar S. Application of principal component analysis in developing statistical models to forecast crop yield using weather variables. Mausam 2014;65(3):357-60.
7. Fisher RA. III. The influence of rainfall on the yield of wheat at Rothamsted. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character 1925;213(402-410):89-142.
8. Jain RC, Agrawal R, Jha MP. Use of growth indices in yield forecast. Biometrical journal 1985;27(4):435-9.
9. Manandhar S, Dev S, Lee YH, Winkler S, Meng YS. Systematic study of weather variables for rainfall detection. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium IEEE 2018, 3027-3030.