Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2021; 10(2): 595-600
Received: 18-12-2020
Accepted: 05-02-2021

Harithalekshmi V
Department of Agricultural Meteorology, Kerala Agricultural University, Vellanikkara, Kerala, India

Ajithkumar B
Department of Agricultural Meteorology, Kerala Agricultural University, Vellanikkara, Kerala, India

Arjun Vysakh
Department of Agricultural Meteorology, Kerala Agricultural University, Vellanikkara, Kerala, India

Corresponding Author:
Ajithkumar B
Department of Agricultural Meteorology, Kerala Agricultural University, Vellanikkara, Kerala, India

Rice yield forecasting using principal component regression and composite weather variables

Harithalekshmi V, Ajithkumar B and Arjun Vysakh

Abstract
Reliable and timely forecasting helps policy makers and farmers to carry out proper crop planning. Weather based crop forecasting is having a paramount importance, since more than 50% of crop loss is due to uncertainties in weather. In this study two weather based statistical models, Principal component regression and composite weather variables were used to forecast yield of two different rice varieties *i.e* Jaya and Kanchana. Weather data were collected from principal Agromet observatory, Vellaikkara. Rice yield data during *kharif* season were collected from the experiment plot maintained in Agricultural Research Station, Mannuthy for a period of 2013 to 2019. Both PCR and CWV models explains the interactive effect of weather variables. The goodness of fit of these models were tested using t test. Calculated value of t was found to be less than that of t- critical value in both the models. Hence it was found out that the predicted yield was similar to that of actual yield. A comparison of model performance was done by estimating mean absolute percentage error (MAPE). The MAPE value calculated for PCR model in Jyothi was 0.76% and 6.19% and for CWV model the MAPE values were 3.69 for Jyothi and 2.95 for Kanchana. Eventhough by comparing MAPE value, PCR model was found to be performing better for the variety Jyothi and models based on composite weather variables performed better in the variety Kanchana. Both the models showed a good performance since the error percentage was in the acceptable limit of ±10%.

Keywords: Principal component regression, composite weather variables, crop planning

Introduction
Accurate and timely crop yield forecasting is imperative for proper crop planning of the country. Among the different input factors in the rice production, weather parameters have a huge impact. Uncertainty in weather variables may cause a large deviations in crop yield (Biswas, 2017). Weather based models give a reliable crop yield forecast, and also helps in forewarning of pest and diseases (Agrawal and Mehta, 2007) [1]. Considering the importance of rice production in Kerala, various studies have been conducted to evaluate different weather based models to forecast rice yield. Ravindran *et al*. (2018) [10] compared different weather based models for forecasting rice yield in central zone of Kerala. During Early periods, studies have been conducted to evaluate the effect of individual weather parameters on the crop yield (Fisher, 1924, Huda *et. al*., 1975 and Jain *et. al*., 1980) [4, 6, 7]. Compared to the individual effect, combined effect of weather factors have more influence on crop yield (Biswas *et al*., 2017). Models based on Principal component analysis and composite weather variables generates indices which explain the combined effect of weather variables. Haritharaj (2019) [5] validated different weather based statistical model *viz*., based on weekly weather variables, fortightly weather variable, crop stage wise weather variables and composite weather variables. Among these models, model based on composite weather variables performed better. She also carried out Principal component analysis. Components obtained from Principal components analysis were used to fit regression equations and the observed and predicted yields were in good agreement. The present study was undertaken to evaluate the performance of Principal component analysis and composite weather variables in predicting rice yield.

Materials and Methods
Data collection
The rice yield data during *kharif* season (June 20th plantig) for the period of 2013-2019 were collected from the experiment plot at Agricultural research station, Mannuthy (10^{0} 32’ N latitude and 76^{0} 20’ E longitude) maintained under All India Coordinated Research Project on Agrometeorology. Two short duration rice varieties namely Jyothi and Kanchana were used for this study. Weather data which includes maximum temperature (Tmax), minimum temperature (Tmin), bright sunshine hours (BSS), Forenoon and after noon relative humidity

(RH1 and RH2), rain fall (RF), wind speed (WS) and pan evaporation (Epan) were collected from the principal agromet observatory, Vellanikkara.

Principal component Regression

Principal component regression (PCR) is a type of regression analysis, which considers principal components (PC) as independent variables, instead of adopting original variables. The PCs are the linear combination of the original variables which can be obtained by Principal component analysis (PCA). PCA is known as the variable reduction method. It may be useful to transform the original set of variables to a new set of uncorrelated variables called principal components (Agrawal and Mehta, 2007) [1]. In this study the value of each weather variables experienced in each phenophase of the crops i.e transplanting to active tillering (P1), active tillering to panicle initiation (P2), panicle initiation to booting (P3), booting to heading (P4), heading to 50% flowering (P5) and 50% flowering to physiological maturity (P6) for a period of 2013-2019 was calculated. Hence in total 46 predictors were there in each year. Due to large dimensionality of predictor variables, it may be computationally unstable. Hence, PCA was performed to reduce the dimensionality of the predictor variables (Sooryanarayana and Mistry, 2016). PCA was done with the help of R statistical software. Each Principal component is obtained by the sum of the product of the loading value obtained by PCA and the value of the weather variable. Hence a Principal component is a linear combination of variables. First few components accounts for most of the variability in the original data. The components which explains most of the variability were selected by eign values. An eigenvalue is the amount of variance that a particular component contributes to the total variance. This corresponds to the equivalent number of variables that the component represents. Screeplots were drawn by plotting eigenvalues against their ordinal numbers. The scree plot was used in deciding the number of factors to retain in factor analysis, and was proposed by Cattell (1966) [3]. A break or a leveling of the slope of the plotted line or bars were noted. The Principal components which lies above the breaking point in a scree plot were selected for further analysis (Kanyongo, 2005) [8]. The selected PCs were used as predictors to fit linear regression model to predict rice yield.

Composite weather variables

Indian agricultural statistical research institute has developed a model which express the effect of weather variables on yield in the wth on yield is a linear function of respective correlation coefficients between yield and weather variables. Weather indices generated based on composite weather variable methods were depicted in Table 1. In this method two types of weather variables were generated *i.e* non weighted weather variable and weighted weather variable. Non weighted weather indices were calculated by taking sum of the weekly weather variables experienced in the crop season. Correlation coefficient between yield and the weather variables experienced during the respective week were calculated. Weighted indices were calculated as the sum product of this correlation coefficient and value of weather variable.

$Zi0 = \sum_{i=1}^{p} \sum_{j=1}^{n} xij$

Zi0 = Non weighted weather indices

Zi1 = Sum product (aij, xij)

Zi1 = Weighted weather indices

aij= Correlation coefficient between yield and ith weather variable in jth week.

xij = ith weather variable in jth week

combined effect of weather variables were calculated as follows

$Zii'0 = \sum_{i=1}^{p} \sum_{j=1}^{n} xij \times xi'j'$

Zii'0 = Non weighted weather indices

Zii'1 = Sum product $(aii'j, xij \times xi'j')$

aii'j= Correlation coefficient between yield and two weather variables

Zii'1 = Weighted weather indices

The calculated weather indices were used as independent variables and yield was considered as dependent variable. Stepwise regression was used to fit model with R statistical software.

Evaluation of the model performance using t- test

Two sample t test assuming equal variance were carried out to evaluate the performance of the model. Null hypothesis formulated depicted no difference between predicted and actual value. Based on the t value and t critical value, null hypothesis was rejected or accepted.

Comparison of model using mean absolute percentage error

The accuracy of both the models were compared by calculating Mean Absolute Percentage Error (MAPE). The difference between actual and predicted yield were calculated. The absolute value of this calculation was summed for every year and it was divided by the number of years under study. The formula for calculating MAPE is give as,

$$\mathrm{MAPE} = \frac{\sum \frac{|\text{Actual yield} - \text{Preicted yield}|}{\text{Actual yield}}}{\text{Number of observation}} \times 100$$

MAPE was calculated for each model under each variety and the model which shows least MAPE was selected as the best model.

Results and Discussion

Principal components and weather indices were generated for both the varieties and were used to fit regression models. The Principal component analysis were carried out and loading values of different weather variables in each Principal components were obtained for Jyothi and Kanchana (Table 2). The variance explained by Principal components were plotted as scree plot. Scree plot drawn for the varieties Jyothi and Kanchana were given as Fig. 1 and 2 respectively. Based on the cumulative variance calculated (Table 3) first five PCs contributed 94.0% variance in Jyothi and 94.9% in Kanchana. Principal components were calculated by taking the sum of the product of values of weather variables and corresponding loading values. The selected Principal components were used to fit linear regression models.

The model developed For Jyothi was,

Yield = 4597.538+-17.692 x X1+14.49 x X2+-13.418 x X3+2.582 x X4+4.783 x X5

The model developed For Kanchana was,

Yield =3260.842+14.897 x X1+42.118 x X2+4.548 x X3+6.622 x X4+11.633 x X5

(Principal component, X = Loading value of weather variables (Table:2) x Value of weather variables)

X1 to X5 denotes Principal component 1 to 5

Weather indices were also calculated based on composite weather variables, these weather indices were used to fit regression models.

The regression model fitted for Jyothi was
Yield = 29060.671+2.87 x Z131
Z131= Weighted weather indices (maximum temperature x forenoon relative humidity)
The regression model fitted for Kanchana was
Yield=40673.779+407.653 x Z11+0.596 x Z570-0.153 x Z340-0.565 x Z231+0.221 x Z871
Z11 = Weighted maximum temperature indices
Z570= Non weighted indices (wind speed x rainfall)
Z340= Non weighted indices (forenoon relative humidity x afternoon relative humidity)
Z231= Weighted indices (minimum temperature x forenoon relative humidity)
Z871= Weighted indices (evaporation x rainfall)
Weighted value of the influence of maximum temperature and forenoon relative humidity was found to be influencing yield in case of Jyothi. Maximum temperature, combination of wind speed and rainfall, afternoon relative humidity and forenoon relative humidity, minimum temperature and forenoon relative humidity and combination of pan evaporation and rain were found to influence yield in case of Kanchana. The predicted and actual yield of both the varieties bases on both these models were depicted in Table 4.

Evaluation of model performance using t test

t test was done to evaluate model performance. The results were depicted in Table 5. t calculated value was less than t critical value in all the cases therefore null hypothesis was accepted. Hence the predicted yield and actual yield of both the varieties simulated by both the models were found to be same.

Comparison of models using MAPE

Mean absolute percentage error of each model for both the varieties were calculated and values were given in Table 6. The MAPE calculated for PCR were 0.76% and 6.19% for Jyothi and Kanchana respectively whereas for CWV model the MAPE values were 3.69% and 2.95% for Jyothi and Kachana respectively. As the error calculated by these models were with in acceptable limit *i.e* ±10% all these models can be used to forecast the yield. To find the best model, the MAPE values of both the models were compared. CWV model were found to be suitable for Kanchana and PCR model was found to be suitable for Jyothi. Haritharaj (2019) [5] compared both PCR model and CWV models in predicting rice yield in central zone of Kerala. Both the models performed well for the rice variety Jyothi in central zone of Kerala with an error percentage, 0.137 for PCR and 33.43% for CWV method. She also found out that PCR model performed better than CWV model in predicting yield of the variety Jyothi. Khair *et al.* (2017) [9] also used MAPE to calculate forecasting error.

Table 1: Weather indices generated based on composite weather variable methods

	Non weighted								Weighted							
No.	1 2 3 4 5 6 7 8								1 2 3 4 5 6 7 8							
	Tmax	**Tmin**	**RH1**	**RH2**	**WS**	**BSS**	**RF**	**Epan**	**Tmax**	**Tmin**	**RH1**	**RH2**	**WS**	**BSS**	**RF**	**Epan**
Tmax	Z10								Z11							
Tmin	Z120	Z20							Z121	Z21						
RH1	Z130	Z230	Z30						Z131	Z231	Z31					
RH2	Z140	Z240	Z340	Z40					Z141	Z241	Z341	Z41				
WS	Z150	Z250	Z350	Z450	Z50				Z151	Z251	Z351	Z451	Z51			
BSS	Z160	Z260	Z360	Z460	Z560	Z60			Z161	Z261	Z361	Z461	Z561	Z61		
RF	Z170	Z270	Z370	Z470	Z570	Z670	Z70		Z171	Z271	Z371	Z471	Z571	Z671	Z71	
Epan	Z180	Z280	Z380	Z480	Z580	Z680	Z780	Z80	Z181	Z281	Z381	Z481	Z581	Z681	Z781	Z81

Zi0 = Non weighted weather indices, Zi1= Weighted weather indices, Zii'0 = Non weighted weather indices (Combination), Zii'1 = Weighted weather indices (Combination) (i,i'= number assigned to each weather variables (1 to 8)

Table 2: Loading values of weather variables obtained after Principal component analysis of Jyothi and Kanchana

	Tmax P1	**Tmax P2**	**Tmax P3**	**Tmax P4**	**Tmax P5**	**Tmax P6**	**Tmin P1**	**Tmin P2**	**Tmin P3**	**Tmin P4**	**Tmin P5**	**Tmin P6**	**RH1 P1**	**RH1 P2**	**RH1 P3**	**RH1 P4**
Jyothi																
X1	0.117	-0.082	0.161	0.199	0.174	0.110	0.042	0.050	0.230	0.025	0.191	0.216	-0.093	0.063	-0.097	-0.144
X2	-0.206	-0.073	-0.062	-0.063	0.176	-0.241	-0.121	-0.123	0.030	0.192	0.159	-0.027	0.260	0.025	0.074	0.007
X3	0.126	-0.290	-0.164	0.156	0.064	0.091	0.255	0.141	-0.021	0.175	-0.063	-0.112	-0.090	0.253	0.172	-0.070
X4	-0.090	0.069	-0.218	0.134	-0.140	-0.066	-0.031	0.189	-0.039	0.153	0.119	0.118	0.004	-0.162	0.254	-0.260
X5	0.001	-0.140	0.027	-0.074	0.020	-0.053	-0.249	-0.174	-0.207	-0.205	-0.032	-0.085	-0.042	0.234	0.161	0.046
Kanchana																
X1	0.074	0.197	0.200	0.192	0.053	0.111	0.071	0.169	0.188	0.179	-0.002	0.183	-0.077	-0.131	-0.168	-0.176
X2	0.271	0.102	-0.036	-0.026	0.038	-0.153	0.195	0.142	0.049	0.102	0.189	-0.040	-0.260	-0.080	-0.021	0.065
X3	-0.081	-0.037	-0.029	-0.165	-0.399	-0.123	0.043	0.077	0.149	0.012	-0.299	0.108	0.054	-0.065	0.047	0.141
X4	0.070	-0.023	0.107	0.061	-0.034	0.285	-0.072	-0.102	-0.156	-0.203	-0.134	-0.135	-0.041	-0.026	-0.006	-0.135
X5	-0.105	-0.125	0.316	0.012	0.070	-0.062	-0.066	-0.181	0.040	-0.048	0.022	-0.070	0.078	0.247	-0.258	-0.030

Table 2 (Contd): Loading values of weather variables obtained after Principal component analysis Jyothi and Kanchana

	RH1 P5	**RH1 P6**	**RH2 P1**	**RH2 P2**	**RH2 P3**	**RH2 P4**	**RH2 P5**	**RH2 P6**	**RF P1**	**RF P2**	**RF P3**	**RF P4**	**RF P5**	**RF P6**	**BSS P1**	**BSS P2**
Jyothi																
X1	0.010	-0.022	-0.136	0.022	-0.210	-0.115	-0.231	-0.080	0.009	-0.100	-0.195	-0.127	-0.178	-0.195	-0.127	-0.178
X2	-0.168	0.242	0.213	0.060	-0.083	-0.119	-0.106	0.078	0.249	0.009	-0.093	-0.154	-0.143	-0.137	-0.154	-0.143
X3	-0.151	0.056	-0.141	0.304	0.110	-0.147	-0.009	0.073	0.035	0.313	0.191	0.134	0.026	0.025	0.134	0.026

X4	0.238	-0.077	-0.002	-0.043	0.131	-0.105	0.064	-0.312	-0.179	-0.006	0.042	-0.152	0.104	-0.093	-0.152	0.104
X5	-0.016	-0.119	0.012	-0.218	0.010	-0.317	-0.057	-0.185	-0.113	-0.105	0.016	0.129	-0.168	-0.147	0.129	-0.168
Kanchana																
X1	0.122	-0.094	-0.014	-0.097	0.212	0.196	0.199	0.099	-0.165	-0.022	0.203	0.144	0.203	0.234	-0.029	0.197
X2	-0.149	0.203	0.226	0.016	-0.014	-0.014	-0.124	0.114	0.153	-0.002	-0.020	-0.055	-0.114	-0.053	-0.251	-0.036
X3	0.207	-0.100	0.194	-0.310	-0.125	0.152	-0.056	-0.100	-0.067	-0.340	-0.200	-0.181	0.024	-0.052	-0.155	0.205
X4	0.162	0.019	0.019	-0.029	-0.154	0.139	-0.155	0.262	0.188	0.055	-0.080	0.154	-0.162	0.058	-0.059	-0.078
X5	-0.056	-0.125	-0.099	-0.107	-0.082	-0.103	-0.105	-0.201	-0.126	-0.106	-0.012	0.277	-0.120	-0.158	0.153	0.063

Table 2 (Contd): Loading values of weather variables obtained after Principal component analysis Jyothi and Kanchana

	BSS P3	BSS P4	BSS P5	BSS P6	WS P1	WS P2	WS P3	WS P4	WS P5	WS P6	Epan P1	Epan P2	Epan P3	Epan P4	Epan P5	Epan P6
Jyothi																
X1	0.188	0.175	0.230	0.117	-0.126	-0.167	-0.121	-0.059	-0.107	-0.146	0.199	-0.091	0.214	0.136	0.125	0.112
X2	0.052	0.192	0.097	-0.116	0.118	0.043	0.232	0.221	0.216	0.136	-0.033	0.168	0.043	0.113	0.146	-0.141
X3	-0.126	0.063	0.054	-0.053	0.064	-0.061	0.123	-0.016	0.022	0.039	0.024	-0.184	-0.174	-0.165	0.224	-0.024
X4	-0.188	0.077	-0.041	0.287	0.020	-0.102	0.040	0.226	0.136	0.247	0.098	0.116	-0.036	0.070	0.076	0.265
X5	0.026	0.060	0.076	0.024	-0.277	-0.342	0.022	0.040	0.159	0.005	-0.270	-0.055	-0.039	-0.294	-0.023	-0.086
Kanchana																
X1	-0.164	-0.239	-0.203	-0.098	-0.125	0.168	0.037	-0.051	-0.005	-0.115	-0.169	0.067	-0.206	-0.151	-0.115	-0.079
X2	0.001	0.067	0.162	-0.180	-0.130	0.108	0.273	0.221	0.185	0.184	-0.152	0.189	-0.060	0.012	0.184	-0.215
X3	0.140	-0.023	-0.049	0.081	-0.210	-0.016	-0.107	0.034	-0.093	-0.183	0.023	0.190	0.157	0.228	-0.183	0.053
X4	0.234	-0.075	-0.048	-0.231	0.022	0.029	-0.072	-0.239	-0.236	-0.092	-0.031	-0.135	0.108	-0.009	-0.092	-0.223
X5	0.093	-0.002	0.124	0.036	-0.169	-0.274	-0.121	0.083	-0.013	-0.005	-0.324	-0.182	-0.107	-0.271	-0.005	0.043

X1 to X5 = Principal components

Table 3: Variance percentage and cumulative variance of each principal component in Jyothi and Kanchana

Variety	Principal components	Variance percent	Cumulative variance percent
Jyothi	X1	32.0	32.0
	X2	24.1	56.1
	X3	17.4	73.5
	X4	13.7	87.2
	X5	6.8	94.0
Kanchana	X1	33.4	33.4
	X2	23.7	57.0
	X3	17.2	74.3
	X4	14.4	88.7
	X5	6.2	94.9

Table 4: Actual and predicted yield of rice varieties based on both the models

Year	Jyothi			Kanchana		
	Actual	PCR	CWV	Actual	PCR	CWV
2013	5622.5	5677.133	5657.836	6355	6618.188	6302.275
2014	4775	4702.851	4623.81	5160	4593.834	5106.093
2015	4380	4413.518	4172.883	4810	5274.205	4755.145
2016	5782.5	5769.137	5842.362	6155	6119.515	6101.076
2017	5590	5555.791	5222.226	6150	5714.677	6096.617
2018	4376.75	4419.369	4706.577	3756	4130.073	3702.076
2019	3609	3598.083	3688.949	6720	6655.278	5747.076

PCR- Principal Component regression CWV- Composite weather variable

Table 5: Results of t test to check the performance of the model

	Jyothi		Kanchana	
	PCR	CWV	PCR	CWV
t- value	$4.35*10^{-5}$	0.073	$6.03*10^{-5}$	0.347
t critical value	2.178	2.178	2.178	2.178

PCR- Principal Component regression CWV- Composite weather variable

Table 6: MAPE calculated for different models

	Jyothi		Kanchana	
	PCA	CWV	PCA	CWV
MAPE	0.766	3.69	6.19	2.95

PCR - Principal component regression, CWV - Composite weather variable

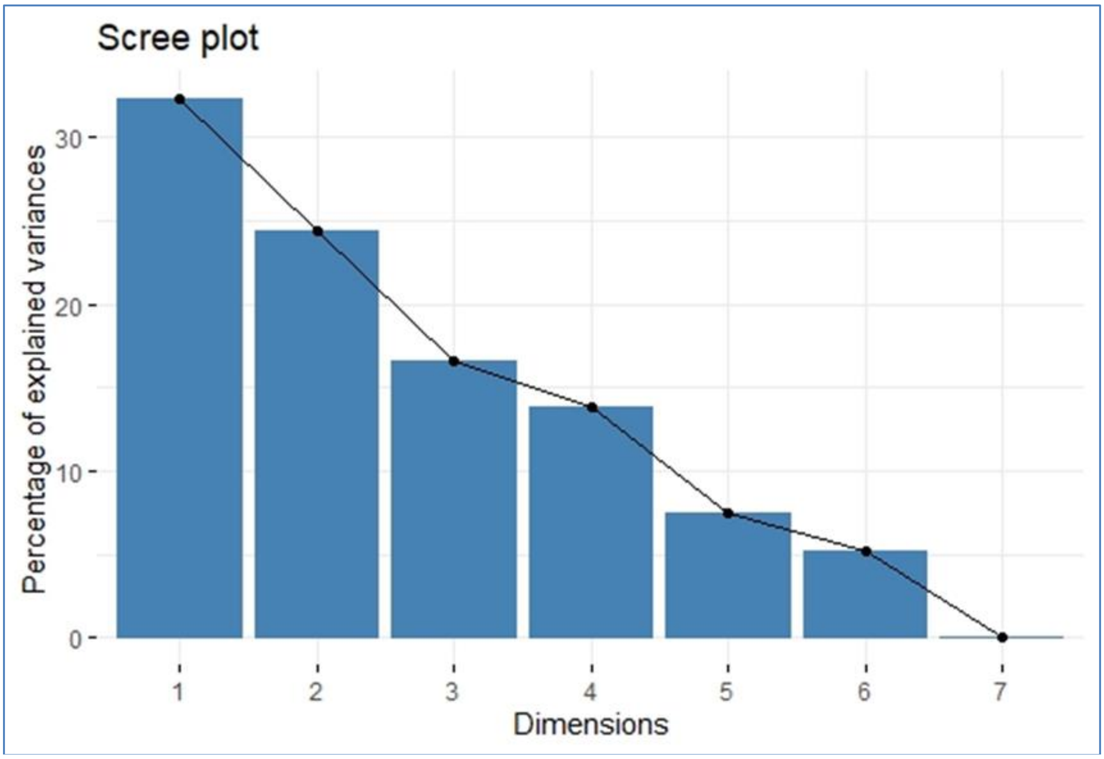

Fig 1: Variability expressed by each principal components - scree plot (Jyothi)

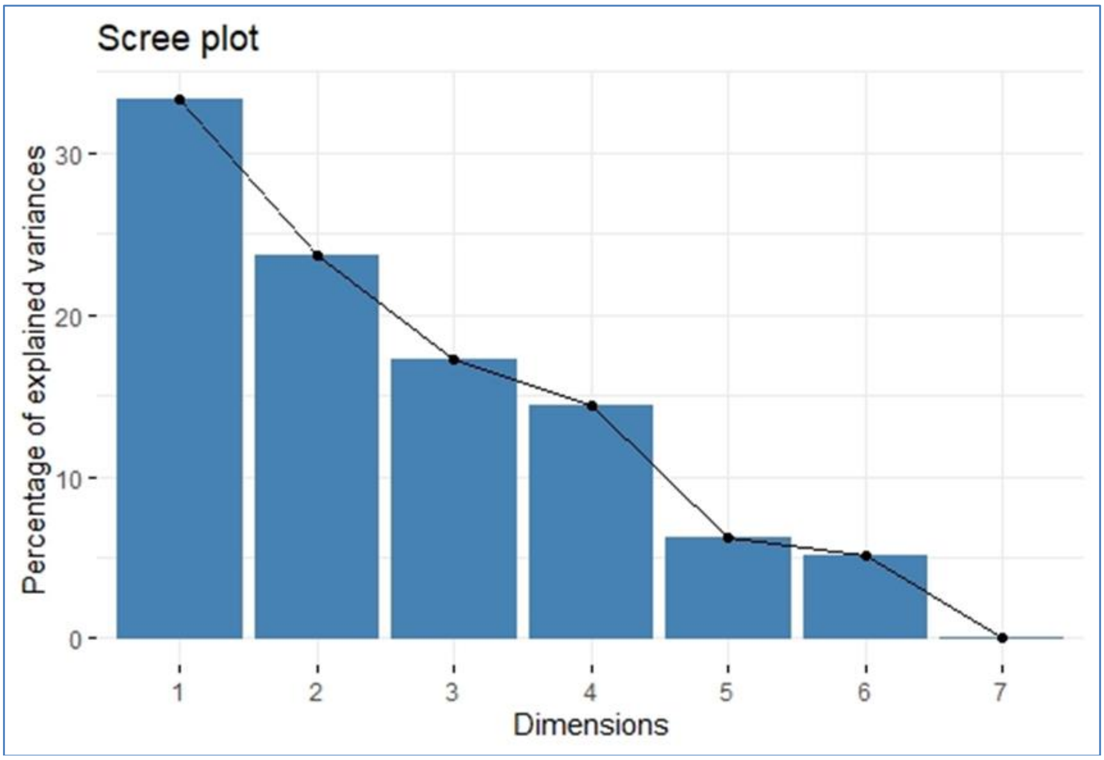

Fig 2: Variability expressed by each principal components - Scree plot (Kanchana)

Conclusion
Yield forecasting models were developed for the rice varieties Jyothi and Kanchana using principal component regression and composite weather variable method. The goodness of fit for both the models were evaluated by t test and predicted yield was found to be in good agreement with actual yield. Mean absolute error percentage was also calculated for the comparison of models. Based on the MAPE value in Jyothi PCR model and in Kanchana CWV model were found to be suitable. Error percentage of both these models were within the permissible level, hence both these models can be used as an effective tool for yield forecasting in Kerala.

References
1. Agrawal R, Mehta SC. Weather based forecasting of crop yields, pests and diseases-IASRI models. J. Ind. Soc. Agrl. Statist 2007;61(2):255-263.
2. Agrawal R. Forecasting techniques in crops. Indian Agricultural Statistics Research Institute 2006,13p.
3. Cattell RB. The scree test for the number of factors. Multivariate Behavioral Res 1966;1:245-276.

4. Fisher RA. The influence of rainfall on the yield of wheat at Rothamsted. Royal Scociety (London) Phil. Trans. Ser. B 1924;213:89-142.
5. Haritharaj S. Crop Weather Relationship of Rice Varieties Under Different Growing Environments. MSc. (Ag) thesis, Kerala Agricultural University, Thrissur 2019,199p.
6. Huda AKS, Ghildyal BP, Jain RC. Contribution of climatic variables in predicting rice yield. Agric. Meteorol 1975;15:71-86.
7. Jain RC, Agarwal R, Jha MP. Effects of climatic variables on rice yield and its forecasts. Mausam 1980;31(4):591-596.
8. Kanyongo GY. Determining the correct number of components to extract from a principal components analysis: a monte carlo study of the accuracy of the scree plot 4(1):120-133
9. Khair U, Fahmi H, Al Hakim S, Rahim R. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. In J. Phy.: Conference Series 2017;930:12
10. Ravindran A. Comparison of Different Weather-Based Models for Forecasting Rice Yield in Central Zone of Kerala. MSc. (Ag) thesis, Kerala Agricultural University, Thrissur 2018,213p