International Journal of Agriculture, Environment and Biotechnology
Citation: IJAEB: 10(1): 149-162, February 2017
DOI: 10.5958/2230-732X.2017.00015.8

AGRICULTURE STATISTICS

Weather based modeling for forecasting area and production of mango in Karnataka

Santosha Rathod[1*] and G.C. Mishra[2]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012, India
[2]Department of Farm Engineering, IAS, BHU, Varanasi-221005, India

*Corresponding author e-mail: santosha.rathod@icar.gov.in

Paper No.: 563 Received: 03-02-2016 Accepted: 08-02-2017

Abstract

A timely and reliable system of forecasting of mango area and production well in advance is of prime importance to farmers and other people who are dependent on horticultural sector. In this study the crop yield forecast models have been developed by considering time series dataon the area and production of mango crop in Karnataka. Daily data on weather variables and yearly data on other exogenous variables of Karnataka state are considered under this study. Weighted and unweighted indices are developed based on the considered weather variables and these indices are further used as independent variables in the regression model. In this study the stepwise regression analysis and ARIMA models were used to forecast the area and production of mango in Karnataka. The empirical study reveals that the weather based stepwise regression model performed better than the ARIMA model for forecasting area and production of mango in Karnataka.

Highlights:

- The weather based stepwise regression model performed better as compare to univariate ARIMA model for forecasting area and production of Mango in Karnataka

Keywords: Mango, Weather indices, Stepwise regression, ARIMA, Forecast

Mango (*Mangifera indica* L.) is the major fruit crop of India, which is considered as the king of fruits. Mango fruits are being utilized at all stages of its development, both in its immature and mature state, raw fruits are used for making chutney, pickles and juices. Mango occupies 22 per cent of the total area under fruits consists of 1.2 million hectares, with production of 11 MT in the world. In India,mangocrop occupies 34.9 and 20.7 percentage of total area and total fruit production of the country (NHB data base 2014-15). Although, India is the largest mango producing country inthe world production, the export of fresh fruit is limited only to cultivars like Alphonso and Dashehari. India's contribution in the world mango market is about 15 percent and accounts to 40 percent of the total fruit export from the country. Horticulture sector in Karnataka is contributing nearly about 10 percent to state GDP. Among the fruit crops Mango(*Mangiferaindica*) occupies an area of 175.40 ha (000'ha)with a production of 1646.50 tones (000'tones)in the state. The demand for mango fruit is growing annually and the requirement is not meeting with increasing rate of production. So, there is good scope for increasing the area and productivity of mango in the country. This can be achieved by making necessary policy implications. Forecasting is used to provide asupport to decision-making and in planning for the future effectively and efficiently.

The crop yield is mainly influenced by the factors like weather parameters and the input variables.

The effect of weather on crop growth varies with growth period of the crop. The influence of weather parameters on crop yield depends on the magnitude and distribution of the weather variables over crop growth period. In forecasting approach for crop production utilizing information on both weather parameters and input variables is advantageous. For accurate forecasting, long term data on weather parameters and input variables are required but practically obtaining long term time series data is very difficult. Therefore to overcome this problem one can build the model with less number of parameters at the same time we have to consider the pattern or the distribution over the entire crop growth period.

Approaches based on various weather based regression analysis capture the effect of climate variables on crop yields was proposed by Agarwal *et al.* (1986), Yang *et al.* (1992), Dixon *et al.* (1994), Garde *et al.* (2012), Rathod *et al.* (2012) Kandiannan *et al.* (2002), and Tannura *et al.* (2008) observed that the explanatory power of the multiple regression models are much better and they express how weather conditions and crop yield are related to one another. Thus, the accuracy of the multiple regression approach is much better than the simulation approach and it is also much easier to deal with a multiple regression approach as compared to a simulation approach. In most of the previous studies linear regression was used to forecast the area and production. However if we consider the number variables, it leads to over fitting of the model and may leads to the problems of multicoliniarity. To overcome these problems stepwise regression is applied in this study. To compare the forecasting performance of the regression model, ARIMA model is also applied in this study as a benchmark model. Sarika *et al.* (2011) used, ARIMA model for modelingand forecasting India's pigeon pea production. Suresh *et al.* (2011) used ARIMA model for forecasting sugarcane area, production and productivity in Tamilnadu state of India. Mishra and Singh (2013) forecasted prices of groundnut oil in Delhi by ARIMA methodology and Artificial Neural Networks. Kumari *et al.* (2014) used the ARIMA model for prediction of rice yield of India. Naveena *et al.* (2014) forecasted coconut production of India using ARIMA methodology.

With these backgrounds efforts, has been made to develop the weather based regression forecasting model and ARIMA model to forecast the area and production of mangoin Karnataka. In the next section, the data description and detailed description of the models used is explained. The weather index development methodology is explained in section 3. Results and discussion is reported in section 4 and finally the concluding remarks are given in section 5.

Material and methods

Data Description

Yearly data on area ('000 ha) and production ('000 MT) of mango crop from 1985 to 2014 were collected from data base of National Horticulture Board (NHB) and www.indiastat.com. Daily data on weather variables (Table 1) were obtained from www.iniawaterportal.org a secondary website of India meteorological department and some weather information was also collected from http://globalweather.tamu.edu. Annual data on other exogenous variables (Table 1) were collected from "Agricultural Statistics at a Glance 2014-15", report published by Department of Economics and Statistics, Karnataka. In ARIMA modeling we used only univariate time series as data on exogenous variables for longer period is unavailable. For regression analysis i.e. weather based forecasting, data from 1985 to 2011 were used for model building and 2012 to 2014 were used to check the forecasting performance of the model. For ARIMA modeling data from 1980 to 2011 were used for model building and 2012 to 2014were used for model validation.

Regression model

The Multiple Linear Regression (MLR) models are applied when two or more independent variables areinfluencing the dependent variable. It uses as few variables or all variables for prediction as necessary to get a reasonably accurate forecast. The MLR model is expressed as follows:

$$Y = b_0+b_1 X_1+ \ldots, b_2 X_2+ b_n X_n+e_t \qquad (1)$$

Where, is the intercept, are the coefficients representing the contribution of the independent variables on the dependent variable Y and error at time t is *i.i.d.* with zero mean and finite variance (Drapper and Smith 1966).

Stepwise Regression Analysis

An important issue in regression modeling is the selection of explanatory variables which are really influencing the dependent variable. There are many methods for selection, stepwise regression analysis is frequently used variable selection algorithm in regression analysis. This is a modification of forward selection, in which at each step all independent variables entered into the model previously are reassessed by their partial F-statistics. An independent variable added at an earlier step may now be redundant because of the relationships between it and latest variable entered in the model (Montgomery*et al.*2003). The predictor variables finally selected by the stepwise algorithm were included in the final model (Eqn.1).

ARIMA model building

Generally ARIMA model (Box and Jenkins 1970), denoted as ARIMA (p,d,q), is expressed as follows;

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \qquad (2)$$

Where,

$\phi(B) = 1-\phi_1 B-\phi_2 B^2 - \cdots - \phi_p B^p$ (Autoregressive parameter) (3)

$\theta(B) = 1-\theta_1 B-\theta_2 B^2 - \cdots - \theta_p B^q$ (Moving average parameter) (4)

ε_t =white noise or error term, d= differencing term,

B=Backshift operator i.e. $B^a Y_t = Y_{t-a}$

The ARIMA model building consists of three stages, *viz.* Identification, estimation and diagnostic checking. Parameters of this model are experimentally selected at the identification stage. Identification of d is necessary to make a non-stationary time series to stationary. A statistical test can by employed to check the existence of stationarity, known as the test of the unit-root hypothesis. Popularly Augmented Dickey Fuller (ADF) test is utilize to test the stationarity. At the estimation stage the parameters are estimated by employing iterative least square or maximum likelihood techniques. The efficacy of the selected model is then tested by diagnostic checking stage by employing Ljung-Box test. If the model is found to be insufficient, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration.

Weather indices

Suppose that we have daily data d on p variables, now new weather variables and interaction components can be generated with respect to each of the weather variables using the below mentioned procedure (Agarwal *et al.* (1986). In order to study the individual effect of each weather variables, two new variables from each variable can be generated as follows:

Let be the value of the i^{th} weather variable at the d^{th} day, is the simple correlation coefficient between weather variable at the d^{th} day and yield over a period of k years. The generated variables are given as follows;

$$Zij = \frac{\Sigma_{d=1}^{n} r_{id}{}^{j} x_{id}}{\Sigma_{d=1}^{n} r_{id^{j}}}, j = 0,1 \qquad (5)$$

For j=0 we have unweighted generated variable as:

$$Zij = \frac{\Sigma_{d=1}^{n} x_{id}}{n} \qquad (6)$$

And weighted generated variables as:

$$Z_{i1} = \frac{\Sigma_{d=1}^{n} r_{id} x_{id}}{n\Sigma_{d=1}^{n} r_{id}} \qquad (7)$$

After calculating these indices, which are again used as independent variables in regression models (Eqn.1). Weighted (Index 2) as well as unweighted (Index 1) weather indices were constructed using daily weather variables.

Result and discussion

Regression analysis of mango area time series

Regression analysis was carried out to understand the influence of weather variables and other variables on mango area of Karnataka. The variables considered under study are listed in table 1. Weather indices for each weather variable viz., Weighted (Index 2) as well as unweighted (Index 1) index have been developed, and then the total number of independent variables increases, in this study the total number of independent variables becomes 21. Firstly the multiple linear regression analysis was carried out by considering all the independent variables. The R^2 of MLR model obtained in table 2 shows that all independent variables considered in the study explains 99.40 percent of the variation in dependent variable. Though the R^2 of MLR model is very high but, if we look at significance of the variables (Table 3), most of the variables in the models are non-significant and variance inflating factor (VIF) is also very high. This clearly indicates the multicollinearity problem among the independent variables. To overcome the same, one of the measure is to drop the unexplained variables (Gujarati*et al.* 2013), the dropping of variable can be done through the stepwise regression analysis.

Hence, stepwise regression analysis was carried out to fit the model. The detailed summary of stepwise regression is given in table 4. The stepwise regression for mango area time series was completed in five steps, the maximum R^2 obtained was 98.7 percent and it was increased in each step. The unexplained or non-significant variables are dropped from the model so that we can get maximum error degrees of freedom. In this stepwise regression analysis we obtained totally five significant independent variables (Table 4) as compare to MLR model (Table 3) in which only one variable was significant. The results of stepwise regression (Table 5) shows that the variable like mangoproduction (X13) significantly contributes in increasing the area, which explains 97 percent of variation in the model. Based on this result we can say that as production in previous year increases, the farmers are more willing to go for the same crop. The variable like number of irrigation pump sets (X21), the rural road length (X20), number of regulated markets (X19) and precipitation (X7) increases the mango area also increases. Based on the results obtained one can say that marketing infrastructure like regulated markets and transportation facility are very important, as these factors increases the area of the mango crop also increases. Based on these significant factors considered, the next step is to forecast the area of mango crop in Karnataka. Performance of stepwise regression model for forecasting mango area of Karnataka was observed in both training and testing data set is given in table 16 and in 17 respectively. The observed versus fitted plot of mango area time series is also depicted in figure 1.

Table 1: Variables considered for regression analysis for forecasting mango area time series

Notation	Variables	Units
	Dependent Variable	
Y	**Mango Area**	**Thousand Hectares**
	Independent Variables	
X_1	Maximum Temperature (Index 1)	Degree Celsius (^{0}C)
X_2	Maximum Temperature (Index 2)	Degree Celsius (^{0}C)
X_3	Minimum Temperature (Index 1)	Degree Celsius (^{0}C)
X_4	Minimum Temperature (Index 2)	Degree Celsius (^{0}C)
X_5	Relative Humidity (Index 1)	Fraction
X_6	Relative Humidity (Index 2)	Fraction
X_7	Precipitation (Index 1)	Millimeter (mm)
X_8	precipitation (Index 2)	Millimeter (mm)
X_9	Wind Speed (Index 1)	Miles per second (mps)
X_{10}	Wind Speed (Index 2)	Miles per second (mps)
X_{11}	Solar Radiation (Index 1)	Megajoules per square meter (MJ/m^2)
X_{12}	Solar Radiation (Index 2)	Megajoules per square meter (MJ/m^2)
X_{13}	Mango production	Thousand Million Tons
X_{14}	Avg. size of operational holdings	Hectares

X_{15}	Area Sown	Hectares
X_{16}	Net area irrigated	Hectares
X_{17}	Fertilizer distribution	Tons
X_{18}	Argil. credit cooperative societies	Numbers
X_{19}	Regulated markets	Numbers
X_{20}	Rural road length	Kilo meters (Kms)
X_{21}	No. of I.P. sets	Numbers

Table 2: Multiple linear regression ANOVA for mango area time series

Source of variation	Sum of Squares	DF	Mean Square	F	Probability	
Regression	35576.17	21	1694.10	41.64	<0.001	0.994
Residual	203.42	5	40.68			
Total	35779.59	26				

Table 3: Multiple linear regression analysis of mango area time series

Variable	Coefficient	Std. Error	t test	Probability	VIF
Constant	212.982	273.934	0.777	0.462	
X_1	-5.972	7.555	-0.790	0.455	25.989
X_2	-0.881	4.891	-0.180	0.862	45.363
X_3	7.279	9.990	0.729	0.490	53.867
X_4	-2.528	2.547	-0.993	0.354	13.303
X_5	-233.872	198.943	-1.176	0.278	24.327
X_6	96.619	67.758	1.426	0.197	24.613
X_7	0.841	1.937	0.434	0.677	9.021
X_8	0.354	0.667	0.531	0.612	8.061
X_9	14.771	27.201	0.543	0.604	5.709
X_{10}	-7.843	21.259	-0.369	0.723	30.378
X_{11}	-1.641	6.536	-0.251	0.809	20.926
X_{12}	0.475	4.095	0.116	0.911	67.341
X_{13}	0.097	0.016	6.164	0.002	38.127
X_{14}	-18.129	25.829	-0.702	0.505	54.771
X_{15}	0.003	0.030	-0.311	0.765	10.878
X_{16}	0.007	0.009	-0.394	0.705	72.720
X_{17}	0.002	0.017	-0.713	0.499	6.457
X_{18}	0.010	0.003	0.556	0.596	22.799
X_{19}	0.031	0.153	0.202	0.845	62.988
X_{20}	0.003	0.375	-0.008	0.994	22.386
X_{21}	0.009	0.010	0.822	0.438	41.185

Table 4: Stepwise Regression ANOVA for mango area time series

Model	Source of variation	DF	Sum of Squares	Mean Square	F	Prob.		Adj.
1	Regression	1	34723.58	34723.58	822.05	<0.0001	0.970	0.969
	Residual	25	1056.01	42.24				
	Total	26	35779.59					
2	Regression	2	35051.56	17525.78	577.75	<0.0001	0.980	0.978
	Residual	24	728.02	30.33				
	Total	26	35779.59					
3	Regression	3	35159.14	11719.71	434.45	<0.0001	0.983	0.981
	Residual	23	620.45	26.98				
	Total	26	35779.59					
4	Regression	4	35270.32	8817.58	380.91	<0.0001	0.986	0.983
	Residual	22	509.27	23.15				
	Total	26	35779.59					
5	Regression	5	35335.14	8817.58	416.62	<0.0001	0.987	0.984
	Residual	21	444.45	21.16				
	Total	26	35779.59					

Table 5: Stepwise Regression Analysis for mango area time series

Variable	Coefficient	Std. Error	t test	Probability	VIF
Constant	-55.373	15.98	3.465	0.0021	
X13	0.073	0.004	16.35	<0.0001	4.292
X21	0.151	0.038	3.933	0.0007	5.423
X20	0.434	0.123	3.495	0.0021	6.988
X19	0.001	0.006	2.289	0.0325	2.721
X7	0.648	0.310	2.090	0.0489	8.061

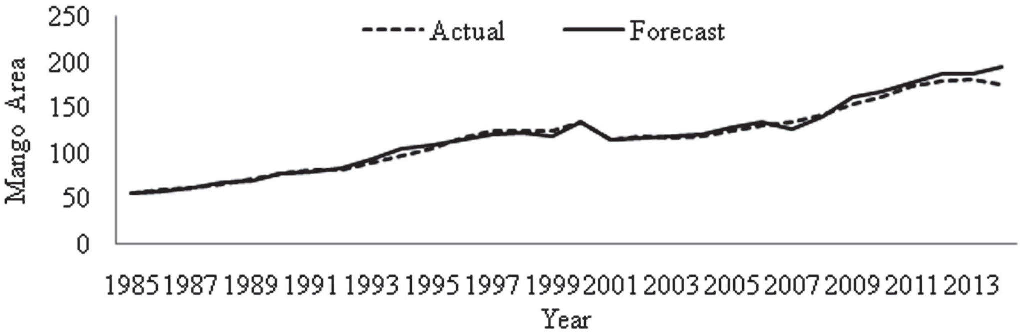

Fig.1: Actual v/s fitted plot of mango area time series using stepwise regression analysis

Regression Analysis of mango production time series

As discussed in section 4.1 the weather indices and other exogenous variables are considered as independent variables in Regression analysis. The variables considered under study are listed in table 6. The R^2 of MLR model obtained in table 7 shows that all independent variables considered in the study explains 99.60 percent of the variation in dependent variable. Though the R^2 of MLR model is very high but, if we look at significance of the variables (Table 8), most of the variables in the models are non-significant and VIF is also very high, which indicates that there is a multicollinearity among the independent variables. To overcome the multicollinearity problem, the stepwise regression analysis was carried out to fit the model. The detailed summary of stepwise regression is given in table 9. The stepwise regression for mango area time series was completed in five steps, the maximum R^2 obtained was 98.7 percent and it was increased in each step. In this stepwise regression analysis, we obtained totally five significant independent variables (Table 9) as compare to MLR model (Table 8) in which only two variables are significant. The results of stepwise regression (Table 10) shows that the variable like mango area (X13) significantly contributes in increasing the production, which explains maximum variation in the model. Based on this result we can say that as area increases, the production also increases and also as the net irrigated area increases production also increases. The climatic variable like minimum temperature (X4) and solar radiation (X11) also significantly influence the changes in production. As temperature decreases the production decreases because low temperature in mango leads to powdery mildew attack. Number of regulated markets (X19) also significantly influence the mangoproduction. Performance of stepwise regression model for forecasting mango production of Karnataka was observed in both training and testing data set is given in table 18 and in 19 respectively. The observed versus fitted plot of mango area time series is given in figure 2.

Table 6: Variables considered for regression analysis for forecasting mango production time series

Notation	Variables	Units
	Dependent Variable	
Y	MangoProduction	Million Tons
	Independent Variables	
X_1	Maximum Temperature (Index 1)	Degree Celsius (^{0}C)
X_2	Maximum Temperature (Index 2)	Degree Celsius (^{0}C)
X_3	Minimum Temperature (Index 1)	Degree Celsius (^{0}C)
X_4	Minimum Temperature (Index 2)	Degree Celsius (^{0}C)
X_5	Relative Humidity (Index 1)	Fraction
X_6	Relative Humidity (Index 2)	Fraction
X_7	Precipitation (Index 1)	Millimeter (mm)
X_8	precipitation (Index 2)	Millimeter (mm)
X_9	Wind Speed (Index 1)	Miles per second (mps)
X_{10}	Wind Speed (Index 2)	Miles per second (mps)
X_{11}	Solar Radiation (Index 1)	Megajoules per square meter (MJ/m^2)
X_{12}	Solar Radiation (Index 2)	Megajoules per square meter (MJ/m^2)
X_{13}	Mango area	Thousand Hectares
X_{14}	Avg. size of operational holdings	Hectares
X_{15}	Production Sown	Hectares
X_{16}	Net area irrigated	Hectares
X_{17}	Fertilizer distribution	Tons
X_{18}	Argil. credit cooperative societies	Numbers

X_{19}	Regulated markets	Numbers
X_{20}	Rural road length	Kilo meters (Kms)
X_{21}	No. of I.P. sets	Numbers

Table 7: Multiple linear regression ANOVA of mango production time series

Source of variation	Sum of Squares	DF	Mean Square	F	Probability	
Regression	4490031	21	213811	60.695	0.0001	0.996
Residual	17538.30	5	3507.66			
Total	4507569.00	26				

Table 8: Multiple linear regression analysis of mango production time series

Variable	Coefficient	Std. Error	t test	Probability	VIF
Constant	-1878.570	2387.204	-0.787	0.457	
X_1	76.942	73.377	1.049	0.329	28.432
X_2	6.838	44.017	0.155	0.881	61.625
X_3	-90.371	87.598	-1.032	0.337	48.039
X_4	38.187	21.405	1.784	0.118	10.894
X_5	2963.283	1821.383	1.627	0.148	23.650
X_6	-1862.938	694.771	-2.681	0.031	20.114
X_7	-19.443	16.917	-1.149	0.288	7.977
X_8	-8.313	7.014	-1.185	0.275	8.637
X_9	-63.014	276.128	-0.228	0.826	6.823
X_{10}	-56.156	242.617	-0.231	0.824	45.901
X_{11}	-27.479	64.326	-0.427	0.682	23.508
X_{12}	24.078	45.239	0.532	0.611	82.133
X_{13}	8.427	1.345	6.263	0.003	25.852
X_{14}	139.977	236.860	0.591	0.573	53.421
X_{15}	0.001	0.005	0.207	0.842	10.753
X_{16}	0.009	0.016	0.559	0.594	77.258
X_{17}	0.003	0.011	0.271	0.794	5.890
X_{18}	-0.132	0.148	-0.891	0.403	19.926
X_{19}	0.665	1.377	0.483	0.644	58.775
X_{20}	0.001	0.001	0.840	0.429	22.509
X_{21}	0.006	0.120	0.050	0.962	42.522

Table 9: Stepwise Regression ANOVA for mango production time series

Model	Source of variation	DF	Sum of Squares	Mean Square	F	Prob.		Adj.
1	Regression	1	4374531.21	4374531	887.81	<0.0001	0.970	0.969
	Residual	25	133037.72	5321.51				
	Total	26	4507568.93					
2	Regression	2	4406463.42	2203231	522.99	<0.0001	0.980	0.978
	Residual	24	101105.51	4212.73				
	Total	26	4507568.93					
3	Regression	3	4422655.62	1474218	399.31	<0.0001	0.983	0.981
	Residual	23	84913.31	3691.88				
	Total	26	4507568.93					
4	Regression	4	4437116.88	1109279	346.39	<0.0001	0.986	0.983
	Residual	22	70452.05	3202.37				
	Total	26	4507568.93					
5	Regression	5	4473711.00	894742.20	554.95	<0.0001	0.987	0.984
	Residual	21	33857.93	1612.28				
	Total	26	4507568.93					

Table 10: Stepwise Regression Analysis for mango production time series

Variable	Coefficient	Std. Error	t test	Probability	VIF
Constant	-648.20	356.28	-1.82	0.081	
X13	9.85	0.89	11.02	0.001	4.29
X16	2.80E-07	2.80E-06	3.69	0.001	5.42
X4	-112.87	47.25	-2.39	0.026	5.99
X19	-1.05	0.47	-2.24	0.035	2.72
X11	34.26	15.43	2.22	0.036	1.11

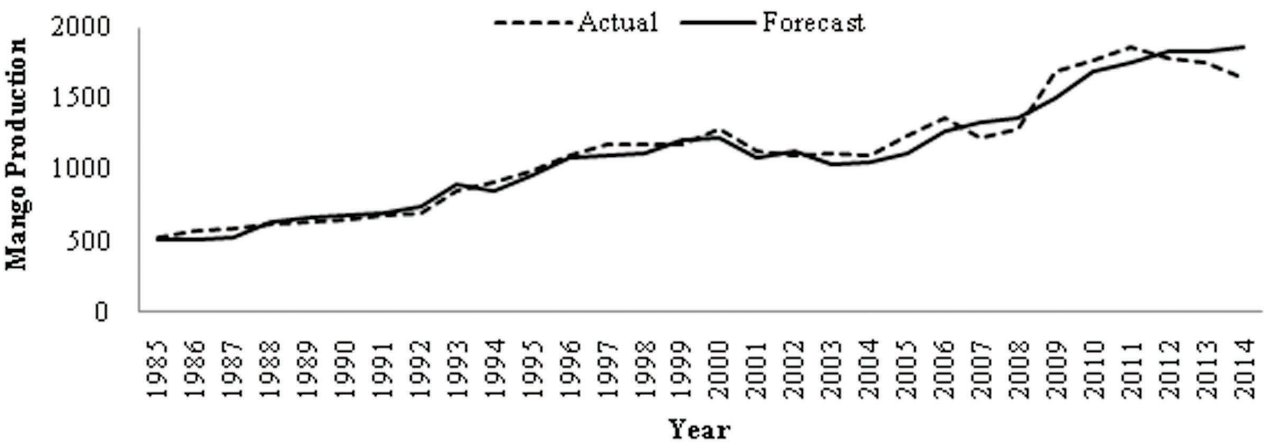

Fig.2: Actual v/s fitted plot of mango production time series using stepwise regression analysis

ARIMA Model for forecasting mango area time series of Karnataka

The time series plot (Fig. 3) of mango area time series reveals that there is a positive trend over time which indicates the time series non-stationary (Fig. 4) in nature. Which is again confirmed by results of Augmented Dickey-Fuller unit root test is given in table 11, which indicates the actual series is nonstationary but after the second differencing the series becomes stationary (Table 12). Finally, the ARIMA (1 2 0) was found adequate for considered time series and parameter estimates of the same are given in table13. Autocorrelation check for residuals obtained from ARIMA model of Mango area time series indicates the residuals found to be non-autocorrelated as probability of chi-square is 0.3122. Further the model performance in training and testing data set is given in table 16 and 17, and the Actual v/s ARIMA fitted plot of mango area time series is depicted in figure 5.

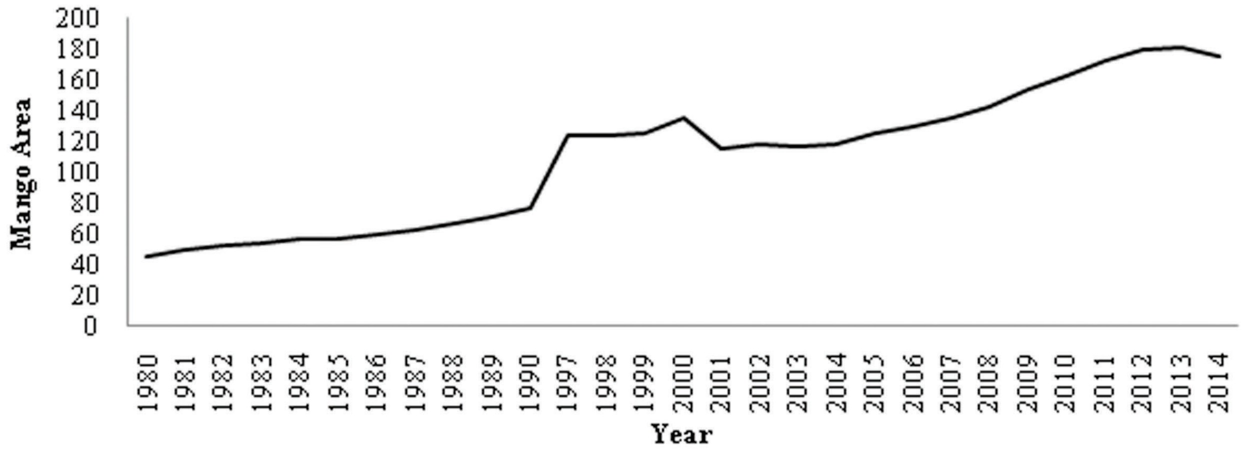

Fig. 3. Time series plot of mango area of Karnataka

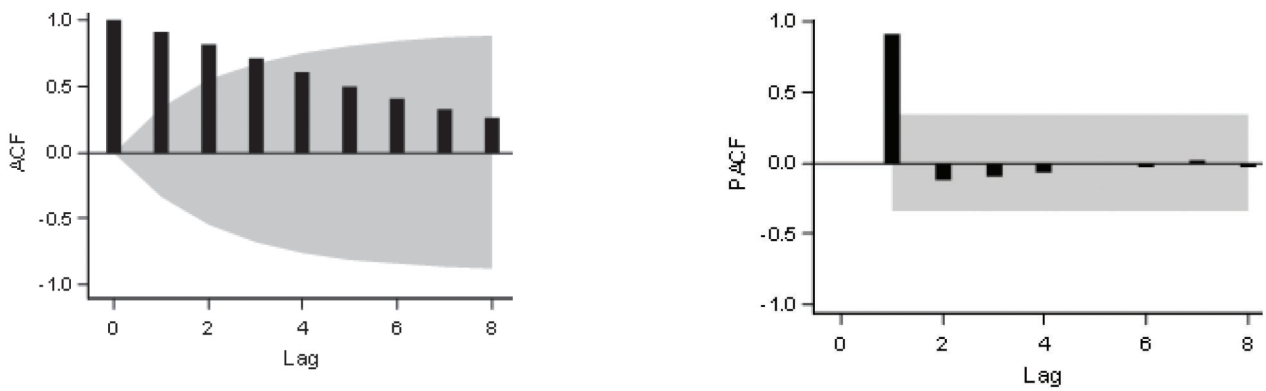

Fig. 4: ACF and PACF plots for mango area time series

Table 11: Stationary test of mango area time series

ADF test statistic				PP test statistic			
Single mean	With trend	Probability		Single mean	With trend	Probability	
		Single mean	With trend			Single mean	With trend
-1.03	62.91	0.71	0.019	-0.20	-9.43	0.92	0.049

Table 12: Stationary test of second differenced mango area(2)time series

ADF test statistic				PP test statistic			
Single mean	With trend	Probability		Single mean	With trend	Probability	
		Single mean	With trend			Single mean	With trend
95.76	95.51	0.0010	0.0001	-66.96	-69.55	0.001	<0.001

Table 13: Parameter estimation of ARIMA (1 2 0) by Maximum Likelihood Estimation method for mango area time series

Parameter	Estimate	Standard Error	t Value	Approx. Pr > |t|	Lag
AR1,1	-0.55	0.14	-3.71	0.0002	1

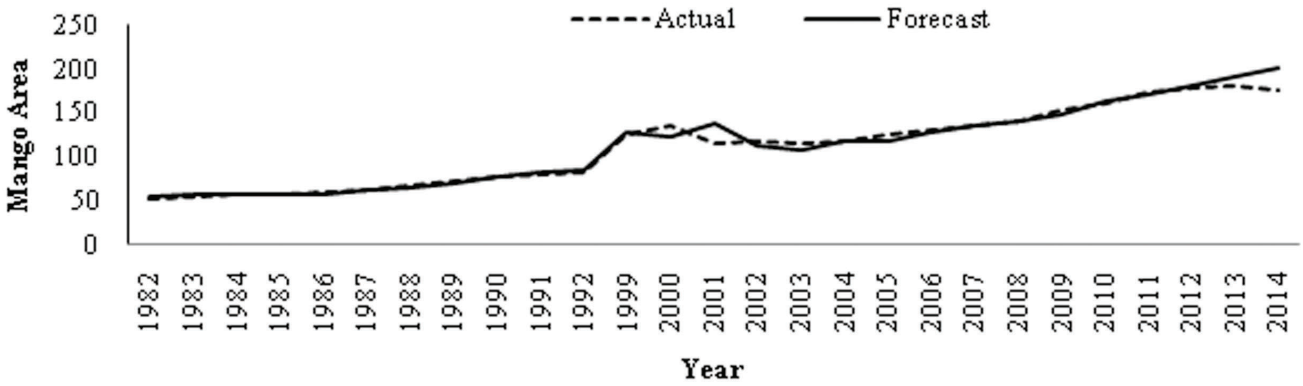

Fig. 5: Actual v/s ARIMA fitted plot of mango area time series

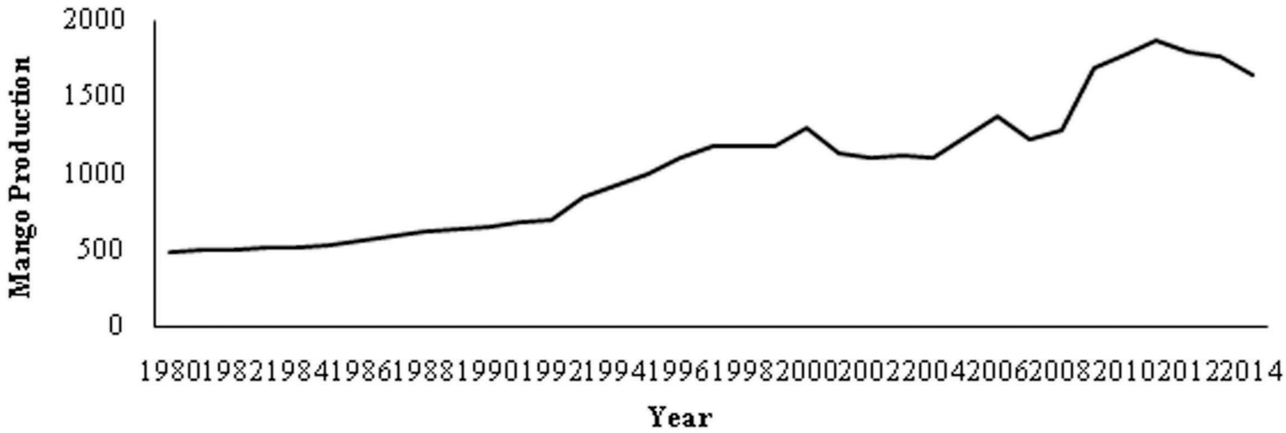

Fig. 6. Time series plot of mango production of Karnataka

ARIMA Model for forecasting mango production time series of Karnataka

The time series plot (Fig. 2) of mango production time series indicates the considered time series is stationary, which is again confirmed by results of Augmented Dickey-Fuller unit root test is given in table 14, which indicates the series is stationary. Finally, ARIMA (1 0 0) model was found adequate for considered time series and parameter estimates of the same are given in table 15. Autocorrelation check for residuals obtained from ARIMA model of Mango Production time series indicates the residuals found to be non-autocorrelated as probability of chi-square is 0.57. Further the model performance in training set and testing data set is given in table 18 and 19 and the actual v/s ARIMA fitted plot of mango production time series is given in figure 6.

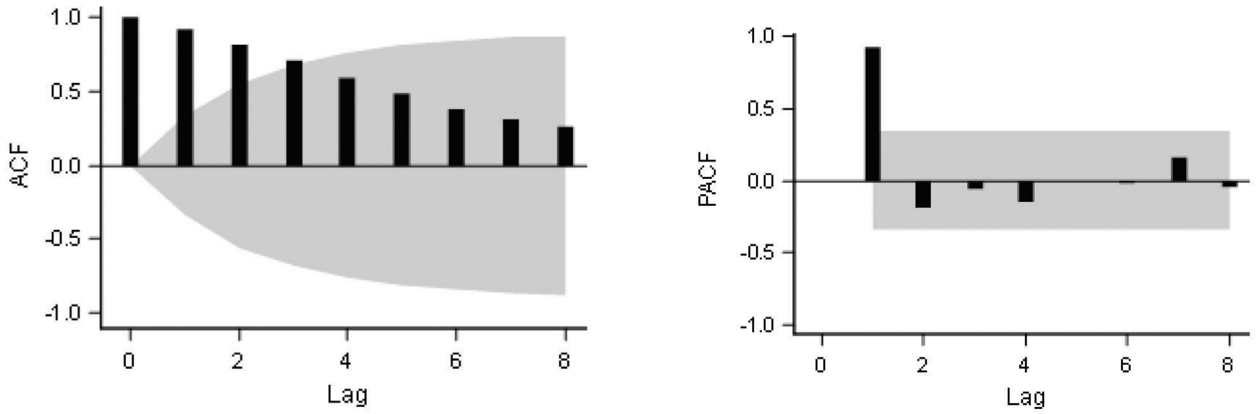

Fig. 7: ACF and PACF plots for mango production time series

Table 14: Stationary test of mango production time series

ADF test statistic				PP test statistic			
Single mean	With trend	Probability		Single mean	With trend	Probability	
		Single mean	With trend			Single mean	With trend
-0.59	10.66	0.048	0.039	-0.6182	-10.039	0.0501	0.049

Table 15: Parameter estimation of ARIMA (100) by Maximum Likelihood Estimation method for mango production time series

Parameter	Estimate	Standard Error	t Value	Approx. Pr > |t|	Lag
MU	1055.5	625.15	1.69	0.0914	0
AR1,1	0.98	0.03	26.76	<0.0001	1

Table 16: Comparison of forecasting performance of all models for mango area time series in training data set

Criteria	Stepwise Regression	ARIMA
MSE	14.55	36.60
RMSE	3.81	6.05
MAPE	2.56	3.50

Table 17: Comparison of forecasting performance of all models for mango area ('000 ha) time series in testing data set

Year	Actual	Forecast	
		Stepwise Regression	ARIMA
2012	178.80	185.73	181.73
2013	180.50	190.25	192.25
2014	175.40	199.38	202.38
	MSE	239.32	291.58
Criteria	RMSE	15.47	17.08
	MAPE	7.65	7.85

Table 18: Comparison of forecasting performance of all models for mango production time series in training data set

Criteria	Stepwise Regression	ARIMA
MSE	5208.30	23913.59
RMSE	72.16	154.64
MAPE	5.78	10.16

Table 19: Comparison of forecasting performance of all models for mango production ('000 MT) time series in testing data set

Year	Actual	Forecast	
		Stepwise Regression	ARIMA
2012	1795.10	1833.21	1857.89
2013	1755.60	1839.32	1847.64
2014	1646.50	1867.52	1837.52
	MSE	16012.05	16300.86
Criteria	RMSE	126.54	127.67
	MAPE	6.70	6.78

For mango area and production time series of Karnataka, the above mentioned model has been fitted and forecasting performance has been assessed in terms of their prediction ability measured by model errors under both training and testing data set. Performance of the stepwise regression analysis was better as compare to ARIMA model in both training (Table 16 and 18) and testing (Table 17 and 19) data set for mango area and production time series of Karnataka as MSE, RMSE and MAPE of stepwise regression is lower compare to ARIMA model in both training and testing data set. The reason for better performance of regression model may be due to consideration of exogenous variables in stepwise regression model, where as in ARIMA model only univariate series has been considered due to non-availability of data on exogenous variables for longer period of time.

Conclusion

Based on the results obtained in this study one can conclude that the stepwise regression analysis for forecasting area and production of mango in Karnataka performed better as compared to the univariate ARIMA model, the reason could be consideration of exogenous variables. Further, as a future study the ARIMA model performance can be improved by considering time series data on exogenous variables for modeling and

References

Agrawal, R., Jain, R.C. and Jha, M.P. 1986. Models for studying rice crop weather relationship. Mausam, 37(1): 67-70.

Box, G.E.P and Jenkins, G. 1970. Time series analysis, Forecasting and control, Holden-Day, San Francisco, CA.

Dixon, B.L., Hollinger., S.E., Garcia, P. and Tirupattur, V. 1994. Estimating Corn Yield Response Models to Predict Impact of Climate Change, *Journal of Agricultural and Resource Economics*, 19(1): 58-68.

Draper, N.R. and Smith, H. 1998. Applied Regression Analysis, 3rd Edition, John Wiley & Sons, New York.

Garde, Y.A., Singh, S., Mishra, G.C. and Singh, T. 2012. Weather Based Pre-Harvest Forecasting of Wheat at Ghazipur (U.P.), *International Journal of Agricultural Sciences*, 8(2): 325-328.

Gujarati, D.N., Porter, D.C. and Gunasekar, S. 2013. Basic Econometrics (Fifth Edition), Tata McGraw-Hill Education Pvt. Ltd, ISBN 10: 0071333452 / ISBN 13: 9780071333450

Kandiannan, K., Karithikeyan, R., Krishnan, R., Kailasam, C. and Balasubramanian, T.N. 2002. A Crop-Weather Model for Prediction of Rice Yield Using an Empirical- Statistical Technique, *Journal of Agronomy and Crop Science*, 188: 59-62.

Karnataka at a Glance 2014-15, Department of Economics and Statistics, Government of Karnataka.

Kumari, P., Mishra, G.C., Pant, A.K., Shukla, G. and Kujur, S.N. 2014. Autoregressive Integrated Moving Average (ARIMA) approach for prediction of rice (*Oryza sativa* L) yield in India. *The Bioscan*, 9(3): 1063-1066.

Mishra, G.C. and Singh, A. 2013. A study on forecasting, prices of groundnut oil in Delhi by ARIMA methodology and Artificial Neural Networks. *AGRIS online papers in Economics and Informatics*, 5: 25-34.

Montgomery, D.C., Peck, E.A. and Vining, G. 2003, Introduction to Linear Regression Analysis, 3rd Ed. John Wiley and Sons (Asia) Pvt. Ltd.

Naveena, K., Rathod, S., Shukla, G. and Yogish, K.J. 2014. Forecasting of coconut production in India: A suitable time series model. *International Journal of Agricultural Engineering*, 7(1): 190–193.

National Horticultural Board (NHB) Data Base (2014-15). Current scenario of Horticulture in India.

Rathod, S., Surendra, H.S., Munirajappa, R. and Murthy, K.B. 2012. Statistical Assessment on the Factors Influencing Agriculture Diversification in Different Districts of Karnataka. *Environment and Ecology*, 30 (3A): 790-794.

Sarika, Iquebal, M.A. and Chattopadhyay, C. 2011. Modelling and forecasting of pigeonpea (*Cajanus cajan*) production using autoregressive integrated moving average methodology. *Ind. J. Agric. Sci.*, 81(6): 520-523.

Suresh, K.K. and Priya, S.R.K. 2011. Forecasting sugarcane yield of Tamilnadu using ARIMA models. *Sugar Tech.*, 13(1): 23-26.

Tannura, M.A., Irwin, S.H. and Good, D.L. 2008. Weather technology and corn and soybean yield in the U.S. Corn Belt, Marketing and Outlook Research Report 2008–2011, Department of Agricultural and ConsumerEconomics, University of Illionis, Illionis.

Yang, S.R., Koo, W.W. and Wilson, W.W. 1992. Heteroscedasticity in crop yield models, *Agricultural and Resource Economics*, 17(1): 103-109.