Yield Forecasting in Cardamom (*Elettaria cardamomum* Maton) Plantations Using Principal Component Regression

Priya P. Menon and V. V. Radhakrishnan

Research Journal of Agricultural Sciences
An International Journal

P- ISSN: 0976-1675
E- ISSN: 2249-4538

Volume: 13
Issue: 01

Res. Jr. of Agril. Sci. (2022) 13: 211–213

CARAS

Res. Jr. of Agril. Sci. (Jan-Feb 2022) 13(1): 211–213

www.rjas.org

ISSN: 0976-1675 (P)
ISSN: 2249-4538 (E)

Full Length Research Article

Yield Forecasting in Cardamom (*Elettaria cardamomum* Maton) Plantations Using Principal Component Regression

Priya P. Menon*[1] and V. V. Radhakrishnan[2]

Received: 12 Oct 2021 | Revised accepted: 16 Jan 2022 | Published online: 07 Feb 2022

ABSTRACT

A study was undertaken to develop a model for forecasting the yield in cardamom plantations. Eleven biometrical characters namely leaves per tiller, tiller height, vegetative buds per clump, bearing tillers per clump, tillers per clump, capsules per raceme, racemes per panicle, panicles per clump, panicle length, seeds per capsule and internodal length were recorded from the plants. The actual yield (Y) of individual plants was also recorded and used as the dependent variable for analysis. Principal component regression analysis was used for estimating the regression coefficients instead of regressing the independent variables. The yield forecasting model developed using principal component regression exhibited a precision of about 87.6% precision.

Key words: Cardamom, Elettaria cardamomum, Principal component regression, Yield forecasting

Cardamom or '*Elettaria cardamomum* Maton' rightly called, as Queen of Spices is one of the most exotic and highly priced spices. Yield forecasting is very essential to device marketing strategies of agricultural crops. Various forecasting methods were developed for annual crops [1] studies were also made in a few perennial crops viz. cocoa [2], coconut [3], cashew [4], clove [5], rubber [6] and coconut [7]. Yield forecasting model in cardamom plantations under intensive management was developed [8]. In all the above models that use biometrical characters for yield forecasting, non-independence or multicollinearity among the regressors arise thereby making the forecasting model less precise. Principal component regression analysis can be used to overcome the problems arising due to multicollinearity and the determination of the best model to predict the dependent variable, yield. Hence attempts were made to estimate the yield in cardamom based on biometrical observations using principal component regression.

MATERIALS AND METHODS

Ninety accessions of cardamom were selected at random from a well-managed plantation in Udumbanchola taluk of Idukki district, Kerala. The farm is situated at 9°53' N latitude, 77°09' E longitude and 1068 m above mean sea level. Data on biometrical characters were recorded from ninety selected plants for three consecutive years from 2017 to 2020. The data for the 3 years were pooled and the resultant data was used for the analysis. Eleven biometrical characters namely leaves per tiller (x_1), tiller height, (x_2), vegetative buds per clump (x_3), bearing tillers per clump (x_4), tillers per clump(x_5), capsules per raceme(x_6), racemes per panicle(x_7), panicles per clump (x_8), panicle length (x_9), seeds per capsule (x_{10}) and internodal length (x_{11}) were recorded from the selected plants. The actual yield (Y) of individual plants was also recorded and used as the dependent variable for the analysis.

The data was checked for multivariate normality and was found true [9]. The data being a set of correlated variables, the principal component analysis was conducted on the correlation matrix. Principal component regression analysis [10] is a type of regression analysis that uses the principal components for estimating the regression coefficients. i.e., instead of regressing the independent variables principal components are used. Hence the problems of non-independence among the regressors do not arise. Thus, principal component regression analysis can be used to overcome disturbance of multicollinearity and for the determination of the best equation to predict the dependent variable. Usually, principal component regression analysis is a three-step regression analysis. First step is to run a principal component analysis on the explanatory variables. Then run an Ordinary Least Square (OLS) regression on the selected components that are most correlated with the dependent variables. Finally, the parameters of the model are computed that corresponds to the input variables (explanatory variables).

RESULTS AND DISCUSSION

The first 6 principal components explain around 94 per cent variation of the data set (Table 1). The most important

* **Priya P. Menon**
priya@maharajas.ac.in

1 Department of Statistics, Maharaja's College, Ernakulam, Kochi - 682 011, Kerala, India

2 Department of Botany, University of Calicut, Thenhipalam - 673 635, Kerala, India

variables (Table 2) to explain 41.441 per cent variation of the data set by the first component were tillers per clump and bearing tillers. The first principal component (z_1) was influenced positively by those variables.

Table 1 Total variance explained

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	Percent of variance	Cumulative percent	Total	Percent of variance	Cumulative percent
1	4.558	41.441	41.441	4.558	41.441	41.441
2	1.878	17.076	58.517	1.878	17.076	58.517
3	1.184	10.763	69.280	1.184	10.763	69.280
4	1.111	10.099	79.379	1.111	10.099	79.379
5	0.890	8.091	87.470	0.890	8.091	87.470
6	0.713	6.485	93.955	0.713	6.485	93.955
7	0.336	3.058	97.013			
8	0.202	1.840	98.853			
9	0.079	0.718	99.571			
10	0.032	0.295	99.866			
11	0.015	0.134	100.000			

Table 2 Component matrix

	Components					
	1	2	3	4	5	6
Leaves/tiller	0.542	0.406	-0.291	-0.234	-0.055	0.552
Tiller height	0.706	0.292	-0.115	-0.369	0.093	0.228
Vegetative buds	-0.219	0.247	-0.033	0.551	0.742	0.181
Bearing. tillers	0.912	-0.243	-0.073	0.029	0.040	-0.175
Tillers/clump	0.920	-0.316	-0.060	0.093	0.074	-0.110
Capsules/raceme	0.668	0.275	-0.516	0.244	0.003	-0.216
Racemes/panicle	0.378	-0.703	0.438	-0.069	0.171	0.297
Panicles/clump	0.912	-0.298	-0.077	0.087	0.074	-0.117
Panicle length	0.655	0.336	0.650	-0.045	0.071	-0.047
Internodal length	0.351	0.798	0.417	0.052	-0.059	-0.198
Seeds/capsule	0.262	-0.017	0.119	0.728	-0.527	0.303

The second component (z_2) which explained 17.076 per cent variation of data set was mostly influenced by the variables intermodal length and racemes per panicle. The third component (z_3) explained 10.763 per cent variation of the data set and was mainly influenced by the variables panicle length and capsules per raceme. The fourth component (z_4) explained 10.099 per cent variation of the data set and was mainly influenced by the variables seeds per capsule and vegetative buds. The fifth component (z_5) explained 8.091 percent variation of the dataset and was mainly influenced by the variable's vegetative buds and seeds per capsule. The sixth component (z_6) explained 6.485 per cent variation of the data set and was influenced by the variables leaves per tiller and seeds per capsule. R is the multiple correlation coefficient which explains how strongly the components are related to the dependent variable. Here, the large value of R (Table 3) indicates that the components are highly related to the dependent variable.

Table 3 Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.971[a]	0.942	0.938	0.08257

Table 4 ANOVA Table

Model	Sum of Squares	Degrees of freedom	Mean Sum of Squares	F	Sig.
Regression	9.188	6	1.531	224.601	0.000**
Residual	0.566	83	0.007		
Total	9.754	89			

**Significant at 1% level of significance

Table 5 Regression coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t value	Sig.
		B	Std. Error	Beta		
	Constant	-0.976	0.159		-6.124	0.000(**)
	z_1	0.035	0.003	1.385	11.580	0.000(**)
	z_2	-0.025	0.004	-0.563	-6.114	0.000(**)
1	z_3	0.004	0.001	0.452	3.224	0.002(**)
	z_4	0.012	0.005	0.410	2.240	0.028(*)
	z_5	0.007	0.005	0.129	1.270	0.208(ns)
	z_6	0.017	0.004	0.525	4.527	0.000(**)

**Significant at 1% level; *Significant at 5% level; NS- Not significant

From (Table 3-4) it is observed that the fitted regression model is good as the value of R is 0.971 and the model is statistically significant. As it was observed that the component z_5 is not significant (Table 5) in this model, the model was revised by dropping the fifth component. From (Table 6-7) it is observed that. the fitted regression model is good as the value of R is 0.97 and the model is statistically significant.

Table 6 Model summary

Model	R	R Square	Adjusted R Square	Std. Error of the estimate
1	0.970[a]	0.941	0.937	0.08287

Table 7 ANOVA Table

Model	Sum of squares	Degrees of freedom	Mean sum of square	F	Sig.
Regression	9.177	5	1.835	267.248	0.000**
Residual	0.577	84	0.007		
Total	9.754	89			

**Significant at 1% level

Table 8 Regression coefficient

Model		Unstandardized coefficients		Standardized coefficients	t value	Sig.
		B	Std. Error	Beta		
1	Constant	-1.003	.159		-6.329	.000**
	z_1	.038	.002	1.488	16.894	.000**
	z_2	-.027	.004	-.612	-7.293	.000**
	z_3	.004	.001	.490	3.563	.001**
	z_4	.009	.005	.324	1.900	.045*
	z_5	.018	.004	.539	4.654	.000**

**Significant at 1% level; *Significant at 5% level

In the revised model all the principal components z_1, z_2, z_3, z_4 and z_6 are all statistically significant (Table 8), So our fitted model is:

$$Y=-1.003 + 0.038\, z_1 - 0.027\, z_2 + 0.004\, z_3 + 0.009\, z_4 + 0.018\, z_6$$

Finally, after the adjustments for unstandardized coefficients, the final fitted model in terms of the input variables is:

$$Y = -1.003 + .0097x_1 + .0040x_2 + .0025x_3 + .0044x_4 + .0081x_5 - .0036x_6 + 0.0166\, x_7 + 0.0093\, x_8 + 0\ .0012\, x_9 - 0.00\, x_{10} + 0.0163\, x_{11}$$

This model was verified in the field with twelve promisi--ng clones which revealed a mean precision of 87.6 per cent.

CONCLUSION

The price behavior in cardamom is highly seasonal and hence early information on the production that can be expected is essential in deciding the market strategies. Further, the international prices and local production figures do have a direct bearing on the current year prices. Since harvest of cardamom continues for a longer period, recording the yield of individual plants may not be feasible all the time. In this context the model proposed is very relevant for estimating the production well in advance and thereby to make a market commitment for better returns for the produce.

LITERATURE CITED

1. Singh D, Jha ML, Areja KG. 1979. *Pre-harvest Forecasting of Crop Yield*. Souvenir Volume, Golden Jubilee of ICAR, IASRI. pp 85-106.
2. Vernon AJ. 1971. The estimation of annual yield of experimental plots of cocoa from sample picks. *Fiji Agric. Journal* 32(2): 41-46.
3. Arulraj S, Narasimhayya G, Haveri RR, Rajasulochana. 1979. Estimation of coconut yield based on snap observations". *In*: E. V. Nelliat, V. Ranganathan S, Vishveshwara SN Potti, Krishna Murthy Rao W (Eds) Proe. PLACROSYM. I. Indian Society for Plantation Crops. Kasaragod. pp 525.
4. George MV, Vijayakumar K, Amarnath CH. 1982. Pre-harvest forecasting of cashew yield based on biometrical characters. (Eds) *In*: KVA Bavappa, RD Iyer, P.K Das, MR Sethuraj. Proc. PLACROSYM V. Indian Society for Plantation Crops. Kasaragod. pp 41-45.
5. Martin PJ, Riley J. 1989. Indirect method for estimating clove tree yields. *Trop. Agriculture* 66(4): 361-364.
6. Reza G, Joseph OO, Joachim S. 2015. Rubber yield prediction by meteorological conditions using mixed models and multi-model inference techniques. *International Journal of Biometeorology* 59: 1747-1759.
7. Nayak SK. 2020. Analysis and high accuracy prediction of coconut crop yield production based on principal component analysis with machine learning models. *International Journal of Modern Agriculture* 9(4): 359-369.
8. Joseph T, Naidu R. 1992. Price structure of cardamom in India – an analysis. *Journal of Spices and Aromatic Crops* 1: 65-71.
9. Anderson TW. 2003. *An Introduction to Multivariate Statistical Analysis*. Third edition, Wiley series in probability and Statistics.
10. Bhuyan KC. 2008. *Multivariate Analysis and its Applications*. New Central Book Agency (p) Ltd, New York, John Wiley and Sons.