Indian Journal of Agricultural Sciences **84** (5): 665–8, May 2014/Short communication
https://doi.org/10.56093/ijas.v84i5.40495

Statistical models to identify significant crop-logging parameters influencing crop-yield across various growth stages in brinjal (*Solanum melongena*)

R HANUMANTHAIAH[1], R VENUGOPALAN[2], D M GOWDA[3], K PADMINI[4], V MANJUNATH[5], G B MALLIKARJUNA[6] and S THIPPESWAMY[7]

University of Agricultural Sciences, GKVK, Bangalore, Karnatka 560 065

Received: 7 November 2012; Revised accepted: 17 April 2014

Key words: Growth, Nitrogen, Phenophases, Phosphorus, Yield, Winter maize

Crop yield forecast poses a formidable challenge to the statistician in developing a most appropriate statistical technique/model satisfying optimum statistical properties. Such a reliable forecast, before harvest, is likely to provide valuable information to the farmers/researchers/policy makers on sales, storage, export, price fixation, grading, marketing to horticultural industries and also to the government for advance planning so as to ensure sustainable crop production in the years ahead. On contrary, researchers are interested to know explicitly by which stage of a standing crop, yield could be predicted more accurately and what are all the key indicators of crop yield (crop logging parameters) across all the stages along with the desired optimum values (bench mark values) of such identified significant indicators across important crop growth stages. Such information would facilitate for the horticulture production experts to incorporate suitable agro-management measures at the early crop stages.

Crop improvement research may also be benefited, as selection can be made in the early stages based on the identified significant crop-logging parameters leading to development of markers. Thus, crop logging is a systematic process in which the record of crop's progress from start till harvest as dictated by biometrical traits (such as plant height, girth, leaf length, breadth, leaf area, no of branches, plant spread etc) are monitored. In the present study, by adopting crop growth modeling as a tool of precision farming in brinjal, significant biometrical traits (along bench mark values) across crop growth stages which can explain the crop yield were identified.

This study utilized the data collected on brinjal (*Solanum melongena* L.) plants at farmers field at Kolar during year 2008 for yield along with eight yield attributing characters such as plant height (X_1), girth (X_2), plant spread (NS, EW) X_3 & X_4, no of branches (X_5) and leaves (X_6), leaf length, breadth and leaf area (X_7, X_8 and X_9).across four growth stages (26 DAP: Seedling stage; 52 DAP: Vegetative phase; 72 DAP: Flowering stage; 89 DAP: Fruiting phase). To understand about the yield-biometrical factors relation across growth stages, multiple linear regression models were constructed for each of the stages, as below.

$$Y = Xb + e$$

where Y is the response vector of order (n×1), X is a matrix of order (n×k+1) with n rows corresponding to observations and k is a parameter vector corresponds to number of independent variables, of order (k+1) × 1 and e is the error component of y assumed to be a random variable with zero mean and constant variance across the values of x, assumed to follow normal distribution.

The least squares estimator of the parameter vector β is

$$\hat{\beta} = b = (X'X)^{-1} X'Y \quad \text{with } V(b) = (X'X)^{-1}\sigma^2.$$

Diagnostics for influential data point are an important part of the regression model-builder's arsenal of tools. They are intended to offer the analyst insight about the data, and to signal which observations may deserve more scrutiny. Outliers were identified using cook's statistic and residuals. A detailed note about outliers and their role in regression analysis is discussed in Hocking and Pendleton (1983), Hocking (1996), and from application point of view in Venugopalan and Rawal (2011), which are being followed in the present work. Finally, Step-wise regression models were developed to identify significant crop-logging parameters, for all the four stages. Adequacy of the fitted models was tested based on the measure of co-efficient of determination (Kvalseth 1985) and based on Mallows Cp statistic (Mallows 1973). SAS V 9.3 software available at IIHR, Bangalore was used for development of models.

Model derived results of seedling stage data (Table 1) indicated that only three variables, viz. plant height, plant girth and plant spread north south could explain the variability in yield to the extent of 71.3 %. Optimum values

[1]e mail: hanumanthaiahragstat@gmail.com, Department of Agricultural Statistics, Applied Mathematics and Computer Science; [2]Senior Scientist (Agril. Statistics), [4]Senior Scientist (Horticulture), IIHR, Bangalore

Table 1 Results of statistical models for seedling and vegetation stage in brinjal with and without outliers

Data set	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b7	b8	R^2	Max. VIF	Eigen value	RMSE	Cp
Seedling stage														
All data (n=40)	11.52	-0.02	0.363	-0.002	-0.22	-0.02	0.02			0.455	1.64	0.008	0.5	
t-value	-0.75	-0.03	-0.15	-0.02	-0.06	-0.01	-0.009							
	15.370**	0.92	2.450**	0.09	3.320**	1.99	2.130*							
Delete outliers (31,32,33,36) n=36		11.36	-0.037	0.36	0.015	-0.14	-0.009	0.004			0.556	1.621	0.008	0.303
t-value	-0.28	-0.01	-0.06	-0.08	-0.03	-0.004	-0.004							
	39.580**	3.920**	6.820**	1.51	1.67	2.430**	0.589							
Delete outliers (1,30,34,39,40) n=31		11.37	-0.04	0.42	0.012	-0.052	-0.01	-0.002			0.766	1.578	0.009	0.184
t-value	-0.287	-0.01	-0.063	-0.008	-0.031	-0.004	-0.004							
	39.580**	3.920**	6.820**	1.51	1.679	2.430**	0.58							
Optimized model I	9.98		0.33							0.38	1	0.06	0.27	0.05
t-value	-1.36		-0.8											
	73.530**		4.167											
Optimized model II	11.13	-0.034	0.426			-0.011				0.71	1.284	0.012	0.189	2.22
t-value	-0.233	-0.009	-0.062			-0.004								
	47.670**	3.770**	6.820**			2.647**								
Vegetation stage														
All data (n=40)	10.034	0.036	0.249	-1.027	0.053	-0.011	-0.013	-0.018	-0.016	0.502	2.001	0.004	0.523	
t-value	-1.05	-0.017	-0.201	-0.007	-0.09	-0.01	-0.011	-0.034	-0.036					
	9.550**	2.190*	1.23	3.930**	0.589	1.281	0.997	0.526	0.441					
Delete outliers (29,33,34,37,40) n=35		9.54	0.004	0.252	-0.023	0.113	0.005	-0.109	0.004	0.004	0.721	2.159	0.004	0.33
t-value	-0.691	-0.012	-0.134	-0.005	-0.059	-0.008	-0.006	-0.025	-0.023					
	13.820**	3.760**	1.88	4.78**	1.911	0.667	2.868**	0.154	0.16					
Delete outliers (31) n=34	9.42	0.04	0.215	-0.021	0.136	-0.016	0.001	-0.001	0.008	0.83	2.167	0.004	0.23	
t-value	-0.499	-0.008	-0.097	-0.003	-0.043	-0.006	-0.005	-0.018	-0.017					
	18.860**	5.720**	2.210*	5.890**	3.178**	0.0141	3.330**	0.076	0.498					
Optimized model I	9.8	0.04		-0.024						0.64	1.059	0.012	0.31	0.96
	-0.373	-0.008		-0.004										
t-value	26.410**	5.880**		5.66**										
Optimized model II	10.1	0.05		-0.02	0.131		-0.02			0.78	1.949	0.008	0.24	3.17
	-0.378	-0.006		-0.003	-0.043		-0.004							
t-value	26.58**	7.409**		5.940**	3.120**		3.410**							

Table 2 Results of statistical models for flowering stage and fruiting stage in brinjal with and without outliers

Data set	b_0	b_1	b_2	b_3	b_4	b_5	b_6	R^2	Max.VIF	Eigen value	RMSE	Cp
Flowering stage												
All data (n=40)	9.68	0.02	0.344	0.001	-0.313	0.027	-0.043	0.458	1.725	0.004	0.577	
	-1.162	-0.014	-0.162	-0.008	-0.031	-0.007	-0.012					
t-value	8.330**	1.44	2.12*	0.024	2.38*	3.660**	3.601**					
Delete outliers (29,33,34,37,40) n=35		9.068	0.009	0.343	0.002	-0.232	0.026	-0.034	0.704	1.95	0.004	0.3
t-value	-0.637	-0.008	-0.1	-0.005	-0.08	-0.004	-0.007					
	14.22**	1.2	3.410**	0.392	2.904*	5.940**	5.130**					
Delete outliers (28,33) n=33	9.426	0.011	0.217	0.002	-0.181	0.027	-0.037	0.803	2.075	0.004	0.23	
t-value	-0.502	-0.006	-0.084	-0.004	-0.065	-0.003	-0.005					
	18.760**	1.853	2.600*	0.404	2.801*	7.870**	7.071**					
Optimized model I	8.81					0.02		0.33	1	0.012	0.39	1.67
	-0.453					-0.005						
t-value	19.45**					3.850**						
Optimized model II	9.928					0.03	-0.032	0.712	1.264	0.012	0.26	1.14
	-0.352					-0.004	-0.005					
t-value	28.230**					7.970**	6.204**					
Fruiting stage												
All data (n=40)	15.16	0.026	-0.27	-0.042	-0.011	0.011	0.018	0.81	1.828	0.001	0.38	
	-1.453	-0.006	-0.079	-0.006	-0.09	-0.009	-0.005					
t-value	10.430**	4.290**	3.530**	6.550**	1.301	1.152	3.330**					
Delete outliers (33,37,40) n=37	14.32	0.023	-0.0296	-0.034	-0.177	0.09	0.021	0.89	1.832	0.001	0.88	
t-value	-1.067	-0.005	-0.06	-0.005	-0.067	-0.007	-0.004					
	13.420**	4.990**	4.910**	6.940**	2.629*	1.28	5.140**					
Optimized model I	18.9		-0.53	-0.03				0.59	1.001	0.002	0.5	1.13
	-1.46		-0.093	-0.008								
t-value	12.880**		5.710**	4.210**								
Optimized model II	16.3		-0.43	-0.03			0.021	0.75	1.001	0.002	0.4	2.13
	-1.298		-0.077	-0.006			-0.005					
t-value	12.532**		5.59**	5.360**			4.563**					
Optimized model III	13.6	0.024	-0.33	-0.003			0.024	0.883	1.832	0.001	0.3	3.61
t-value	-1.154	-0.005	-0.064	-0.005			-0.004					
	11.780**	4.76**	5.078**	6.837**			6.577**					

of the identified significant crop-logging parameters indicated that to have a good hold on the crop growth at 26 DAP (seedling stage), 22 cm of plant height, 10 cm of plant girth and 50 cm north-south plant–spread, is ideal. Results of step-wise regression models of vegetative stage indicated that only four variables, viz. plant height, number of leaves and primary branches and plant spread east-west could explain the variability in yield to the extent of 78. Further optimum values of the identified significant crop-logging parameters indicated that to have a good hold on the crop growth at 52 DAP (vegetative stage), 47 cm of plant height, 58 leaves and 71 cm of east-west plant–spread, is ideal.

Results of optimized models for flowering stage (Table 2) indicated that only two variables, viz. plant spread north-south and plant spread east-west could explain the variability in yield to the extent of 71.20 % as against 80.3 %. Optimum values of the identified significant crop-logging parameters indicated that to have a good hold on the crop growth at 72 DAP (flowering stage), 93 cm of north-south plant–spread and 63 cm of east-west plant–spread. Results of optimized models for fruiting stage indicated that plant height and girth, number of leaves and plant spread east-west could explain the variability in yield to the extent of 88% as against 89% (when all the biometrical characters are included in the model). Further optimum values of the identified significant crop-logging parameters indicated that to have a good hold on the crop growth at 89 DAP (fruiting stage), 83 cm of plant height, 15 cm of girth, 180 leaves and 133 cm of east-west plant–spread, represent an ideo-type crop canopy structure in brinjal.

Before drawing final conclusion about the adequacy of the selected models for all the four crop growth stages, randomness and normality assumptions of model generated residuals were tested by using run and Shapiro-Wilk tests, respectively, and found satisfied, which also ensured the suitability of the selected model.

SUMMARY

Crop-logging models developed showed that brinjal (*Solamun melogena* L.) crop yield could be predicted well in advance as early as 26 DAP using three biometrical traits (plant height, plant girth and plant spread north south) to an extent of 71 %. As the DAP increases prediction of yield could be possible to an extent of 88%. Identification and removal of outliers in data set increased the prediction of models in the range of 31%, 33%, 34 and 8% respectively across four crop growth stages.

REFERENCES

Hocking R R 1996. *Method and Application of Linear Models.* John Wiley & Sons. New York.

Hocking R R and Pendleton O J.1983. The regression dilemma. *Communication in Statistic, Part A-Theory and Methods* **12**(5): 497–7.

Kvalseth T O. 1985. Cautionary note about R^2. *American Statistician* **39**: 279–85.

Mallows C L. 1973. Some Comments on C_p. *Technometrics* **15** (4): 661–75.

Venugopalan R and Rawal R D. 2011. Effect of outliers in statistical modelling for predicting the outbreak of anthracnose in grapes. *Indian Journal of Agricultural Sciences* **81** (10): 61–3.