

A statistical model for ascertaining the influence and reliability of weather parameters on incidence of blossom blight in mango (*Mangifera indica* L.)

R. Venugopalan, R. D. Rawal¹ and A. K. Saxena¹

Section of Economics and Statistics
Indian Institute of Horticultural Research
Hessaraghatta Lake Post, Bangalore-560 089, India
E-mail: venur@iihr.ernet.in

ABSTRACT

A statistical model was developed to study the influence and reliability of weather parameters on incidence of blossom blight in Mango (*Mangifera indica*) and subsequently to predict their incidence. Results showed that preceding week's weather variables viz., maximum and minimum temperature, evaporation, rainfall, morning and evening relative humidity and wind speed were found to collectively predict blossom blight incidence to the extent of 94.3 per cent. Further, as a measure of goodness-of-fit, the coefficient of determination (R^2) and mean squared error were used to evaluate the empirical model developed by using above variables. Validation test showed that the model developed using relative humidity at 07.30 h (X_3), evaporation (X_5) and wind speed (X_6) ($Y = 883.4 - 8.065 X_3 - 11.506 X_5 - 33.619 X_6$) could predict the incidence to the extent of 75.7%. This model is useful in determining the role of climatic factors in disease appearance and progression and devising suitable management strategy.

Keywords: Blossom blight, mango, coefficient of determination, climatic factors, model

INTRODUCTION

Mango (*Mangifera indica* L.) is the most important fruit crop grown across various climatic and soil conditions in India. Although mango is grown in large area, the productivity is much lower than the world average. There are many biotic and abiotic stresses, which are responsible for the low productivity. Among the biotic factors, diseases like powdery mildew, leaf spot and blossom blight are the most serious on mango and account for the major losses. The pathogens such as, *Colletotrichum glaeosporioides*, *Alternaria alternata* and *Pestalotiopsis mangiferae* are responsible for blossom blight. These pathogens can cause disease singly or in combination depending upon the weather conditions. In India, the occurrence and importance of blossom blight disease was reported during 1992 (Rawal, 1992). However, information on the influence of weather factors on disease outbreak is lacking. In view of this, a study was undertaken to find out the role of weather factors individually and in combination that lead to the disease incidence. Such work will be helpful to develop prediction equations so as to facilitate devising a suitable management strategy.

MATERIAL AND METHODS

Mango var. Totapuri orchards were surveyed during August to October 2005 (flowering period) to record the blossom blight disease initiation and its progression. Disease ratings were recorded at weekly interval by following 0-5 scale, where 0 = nil PDI; 1= 0 > PDI d ≥ 10; 2= 11 ≥ PDI ≤ 25; 3= 26 ≥ PDI ≤ 50; 4= 51 ≥ PDI ≤ 75 and 5 = e ≤ 76 Per cent Disease Intensity (PDI). Data thus recorded were converted to Percent Disease Index as per McKenny (1923). The weekly weather data such as maximum temperature (°C) (X_1), minimum temperature (°C) (X_2), relative humidity (%) (7.30 h) (X_3); X_4 : relative humidity (%) (14.30 h), (X_4), evaporation (mm) (X_5), wind speed (Kph) (X_6), rainfall (mm) (X_7) and number of rainy days (X_8) were collected from IIHR meteorological observatory for the same period. All the data were subjected to statistical analysis in order to assess the influence of abiotic factors on blossom blight incidence and subsequently for the development of disease prediction models as detailed below.

In order to assess the degree of linear association of each of the weather variables on blossom blight incidence over a time period, linear correlation coefficient was worked out. Further, with a view to understand the role of weather

¹Division of Plant Pathology

parameters on degree of disease incidence, a statistical model was developed. As a measure of goodness-of-fit, the value of the co-efficient of determination (R^2) was calculated (Kvalseth, 1985) as illustrated below:

$$R^2 = 1 - \frac{[\sum (Y_t - \hat{Y})^2]}{[\sum (Y_t - \bar{Y})^2]}$$

where Y_t represents PDI during the time period t .

However, inclusion of an additional independent variable into the selected candidate model always boosts the computed R^2 value. Hence, to ensure the statistical significance of the computed regression coefficients, these were subjected to t -test statistical analysis. Further, to test whether the regression models are robust against the basic assumption of regression approach, *viz.*, independent variables should not be related among themselves, commonly known as the problem of multi-collinearity, Variance Inflation Factor (VIF) was worked out. A value of VIF exceeding 10 indicates the presence of strong multi-collinearity among observations (Ryan, 1997). To select the significant weather parameters influencing the observed variability in blossom blight incidence, a step-wise regression procedure (Ryan, 1997), was employed as delineated below.

In step-wise regression, the final regression equation was developed stage by stage. During each stage, making use of F test, an independent variable (weather factor) would enter into the equation if the significance level of its F value is ≤ 0.05 , and would be removed if the significance level is ≥ 0.1 . This process was continued until all the variables were exhausted and are found significant, resulting in the final equation comprising only the significant weather variables. Further, as we are dealing with a set of sample observations to take inference about the whole population under study (variability in blossom blight incidence over time), in general, it is essential to perform a detailed residual (difference between observed and predicted PDI) analysis

(Venugopalan and Prajneshu, 1997) before flagging of the developed model for its universal validity. To this end, two important assumptions about the model generated residuals; *viz.* randomness and normality were tested by following one-sample run test and Shapiro-Wilk test, respectively (Agostid'no and Stephens, 1986).

RESULTS AND DISCUSSION

Linear correlation coefficient analysis among weekly blossom blight incidence with preceding week's weather parameters were worked out and presented in Table 1. Perusal of the results indicated the presence of highly significant correlation among blossom blight incidence with the averages of preceding week's weather variables *viz.*, relative humidity at 14.30 h ($r=0.60$) and wind speed ($r=0.69$). Among the intra-class linear correlation coefficients based on preceding week's weather parameters, maximum temperature had shown a significant positive correlation with wind speed ($r=0.65$); relative humidity at 7.30h with evaporation ($r=-0.82$), with rainfall ($r=0.79$); relative humidity at 14.30h with wind speed ($r=-0.88$) indicating about the indirect effect of these factors on blossom blight incidence.

As a next step, statistical model was developed by regressing weekly blossom blight incidence with all the weather parameters of preceding week. Perusal of Table 2 indicates that the preceding week's weather variables were found to predict blossom blight incidence to the extent of 94.3%. Though the model developed resulted in considerably high R^2 value, some of the regression coefficients corresponding to weather parameters were only significantly related to blossom blight incidence as indicated by the t -test statistic value (being greater than 1.96). In addition to this, the model indicated the presence of strong multi-collinearity among weather variables as indicated by VIF (Variance Inflation Factor) value 23.18 (being >10 and eigen value being nearer to zero).

Table 1. Correlation coefficient (r) among weekly blossom blight incidence and weather parameters

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% damage (1)	1.00							
Maximum temperature (2)	-0.37	1.00						
Minimum temperature (3)	-0.41	-0.22	-1.00					
RH 7hrs (4)	0.31	-0.11	-0.22	1.00				
RH 13hrs (5)	0.60	-0.41	-0.17	0.37	1.00			
Wind speed (6)	-0.54	-0.12	0.25	-0.82	-0.33	1.00		
Evaporation (7)	-0.69	0.65	-0.02	-0.52	-0.88	0.40	1.00	
Rain fall (8)	0.10	0.09	0.19	0.79	0.05	-0.58	-0.13	1.00
No of rainy days (9)	0.31	-0.50	0.35	0.76	0.56	-0.51	-0.69	0.60

Note: bold values are significant at 5% level

Table 2. Results of statistical models with goodness of fit statistics

Model type	Statistical model (with Standard error of b _j)									R ² (%)	VIF
Full regression model (All weather parameters)	Y = 1866.4 - 31.1 X ₁ +30.4 X ₂ -19.9 X ₃ - 25.5 X ₄ +4.6 X ₅ -21.52 X ₆ +3.64 X ₇ -15.45 X ₈									94.3	23.2
	(18.3)	(20.3)	(7.7)	(4.8)	(8.8)	(33.7)	(1.6)	(8.03)			
	-stat -1.6	1.5	-2.6	0.95	-2.87	-0.63	2.25	-1.9			
Model using wind speed (X ₆), relative humidity 07.30 h (X ₃) evaporation (X ₅) and rainfall (X ₇)	Y = 1290.8 - 12.9X ₃ -12.7X ₅ -41.2X ₆ +1.57 X ₇									80.7	8.76
	(5.04)	(3.98)	(11.3)	(1.27)							
	t-stat	-2.56	-3.2	.15	-3.67						
Optimised model wind speed (X ₆), relative humidity 07.30 h (X ₃) evaporation (X ₅) only]	Y = 883.4 - 8.06 X ₃ -11.506 X ₅ -33.619 X ₆									75.7	3.56
	(3.3)	(4.02)	(9.84)								
	t-stat	-2.42	-2.86	-3.42							

Note: Values parenthesis are standard error of regression estimates bi

Accordingly, to eliminate this multi-collinearity problem, step-wise regression models were developed and the results are presented in Table 2. The results indicated that only four variables viz., wind speed, relative humidity at 07.30h, rainfall and evaporation could explain the variability in blossom blight incidence to the extent of 80.7 % as against 94.3% (when all weather variables included in the model). However, in another optimized model, rain fall was also eliminated, because the corresponding regression coefficient was not statistically significant as indicated by t-test statistical value 1.27 (being <1.96). Therefore, among all weather parameters tried, only three weather variables, viz. relative humidity at 07.30h, evaporation and wind speed could themselves collectively explain 75.7% of the variation in weekly blossom blight incidence, which is quite high while predicting a biological variable. Further, the regression coefficients corresponding to these variables in the final model were also statistically significant, as indicated by the t-statistic value, which in

absolute value exceeded 1.96, the critical region value. Also the VIF value computed for this model was well inside the acceptable limit (3.56<10.0) which further strengthens the statistical validity of the optimized model.

Before drawing final conclusion about the adequacy of the selected model, universal validity of the model was assessed by performing detailed residual analysis. The randomness assumption of the residuals tested using the one-sample run test resulted in the test statistic value as 0.009, which being less than 1.96, is well inside the critical region of normal table at 5% level of significance. The normality assumption of the model generated residuals tested using Shapiro-Wilk test resulted in the test statistic value as 0.931, which being less than 1.96, is well inside the critical region at 5% level of significance (Table 3). These two results further ensured the universal validity of the developed model. A graphic representation of the adequacy of the fitted models is presented in Fig. 1. The per cent error variation between observed and predicted values ranged from 0.83 to 20.9 (Table 4). The approach of this study is to develop a reasonable prediction model for mango (cv Totapuri) blossom blight using reliable and dependable weather variables which have direct influence on blossom blight incidence. Further, validation of optimized model (the observed incidence values were

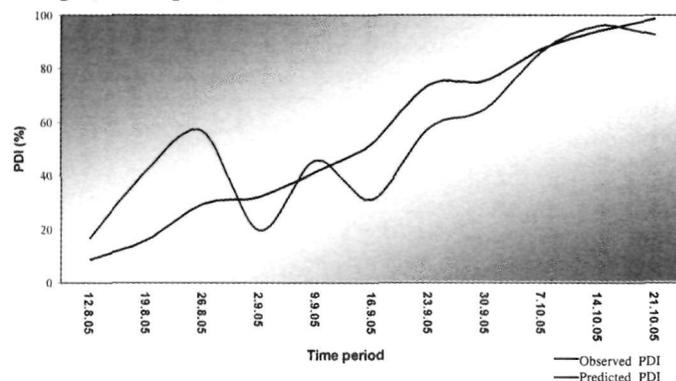
Table 3. Results of residual analysis for the optimized model

Test criterion	Residual assumption tested	Statistic value	Significance
Run test	Randomness	0.0091	P<0.005
Shapiro-wilk	Normality	0.931	P<0.005

Table 4. Calculated per cent error variation for predicted blossom blight incidences

Time in Week	Date of observation	Predicted blossom blight incidence (%)	Error variation between observed and blossom blight (%)
1	12.8.05	16.44	-7.76
2	19.8.05	41.68	-25.60
3	26.8.05	56.67	-27.30
4	2.9.05	19.49	12.50
5	9.9.05	45.78	-4.44
6	16.9.05	31.08	20.92
7	23.9.05	57.48	16.52
8	30.9.05	64.84	10.48
9	7.10.05	86.50	0.83
10	14.10.05	96.18	-2.17
11	21.10.05	92.63	6.03

Fig. 1. Statistical model for epidemiology of Blossom Blight in Mango (cv Totapuri)



compared to predicted values using optimized model), clearly indicated that the model could predict the incidence reasonably well.

By using the model developed in the present study, it is possible to work out mango (cv Totapuri) blossom blight incidence with minimum available data *viz.*, relative humidity at 07.30h, evaporation and wind speed. However, future studies to standardize variables for improving the precision of blossom blight incidence estimates will be envisaged with good variability in the data set. This model is useful in determining the role of climatic factors in disease appearance and progression and devising a suitable management strategy. Thus, we have used statistical modelling as a power tool for developing suitable disease forecasting models and also for optimizing factors influencing disease incidence.

ACKNOWLEDGEMENTS

The authors are grateful to the Director, Indian Institute of Horticultural Research, Hesseraghatta Lake PO,

Bangalore, India for providing facilities to carry out the work. The authors also thank the anonymous referee and the editor for their critical suggestions which led to substantial improvement in the quality of the paper.

REFERENCES

- Agostid'no R.B and Stephens. M.A. 1986. Goodness of Fit Techniques. Marcel Dekker, New York.576p
- Kvalseth, T. O. 1985. Cautionary note about R^2 . *Amer. Stat.*, **39**:279-85
- Mcknney, H. H., 1923. Influence of soil temperature and moisture on infection of wheat seedlings by *Helminthosporium sativum*. *J. Agric. Res.*, **26**:195-217
- Rawal, R.D. 1992. *Ann. Rep. 1991-92*. Indian Institute of Horticultural Research, Bangalore, India 204p
- Ryan, Thomas P. 1997. Modern Regression Methods. John Wiley and Sons Inc., New York. 515p
- Venugopalan, R and Prajneshu. 1997. A generalized allometric model for determining length-weight relationship. *Biometrical J.*, **39**:733-39.

(MS Received 3 April, 2006 Revised 12 June, 2006)