Journal of Biodiversity and Environmental Sciences (JBES)
ISSN: 2220-6663 (Print) 2222-3045 (Online)
Vol. 4, No. 5, p. 1-9, 2014
http://www.innspub.net

INNSPUB

RESEARCH PAPER

OPEN ACCESS

A statistical model to estimate potential yields in Persian walnut (*Juglans regia* L.)

Behnam Ojaghloo[1], Kourosh Vahdati[*1], Reza Amiri[2], Darab Hassani[3]

[1]*Department of Horticulture, College of Aburaihan, University of Tehran, Tehran, Iran*

[2]*Department of Agronomy and Crop Breeding Sciences, College of Aburaihan, University of Tehran, Tehran, Iran*

[3] *Department of Horticulture, Seed and Plant Improvement Institute, Karaj, Iran*

Article published on May 09, 2014

Key words: Frost, Insurance, Lateral bearing, Multiple regression, Trunk circumference.

Abstract

Winter and spring frost damage to fruit and nut trees is common so growers are interested in insuring their crop against losses. The purpose of this work was to develop a model that could be used for this purpose in walnut (*Juglans regia* L.) orchards. During 2010, 40 seedling walnut trees from Koleeim and 39 from Jezla, the two main walnut growing locations of Tarom, in Zanjan province in Iran, ranging in age between 10 and 50 years old, were selected at random. Trunk circumference (TC), trunk cross sectional area (TCA), canopy radius (CR), lateral bud fruiting (LBF), tree growth habit, average annual shoot growth (SHGPY), the ratio of external/internal fruit bearing shoots in the canopy, dichogamy, fruit set in the spring, fruit weight, kernel percentage, and number of fruits per tree, were recorded. Yield per tree was calculated using nut weight and number of nuts per tree. The effect of these traits as independent variables on nut yield as a dependent variable was studied using multiple linear regression models. The data from two sites were used separately and as a pooled set for analysis. Results were validated using a second set of data. The best-fit model employed trunk circumference and lateral bearing percentage as independent variables: Y= (0.12465 * TC) + (1.32889 * LTB),Y= (0.15071 * TC) + (0.55075 * LTB) and Y= (0.14049 * TC) + (0.93873 * LTB).Coefficients of determination were 0.8417, 0.8773, and 0.8526 for the two sites and the pooled data respectively.

***Corresponding Author:** Behnam Ojaghloo ✉ Ojaghloobehnam@yahoo.com

Introduction

Persian walnut (*Juglans regia* L.), widely cultivated for nut production, is an ancient species (Fjellstrom and Parfitt, 1994; Vahdati, 2000) originating in areas of Central Asia, including Iran. Winter and spring cold damage to fruit trees occurs frequently in many areas of the world and it is common practice for growers to insure crops against this damage. Cold can affects dormant shoots in the winter, current-year shoots which bear flowers in the spring, and even young developing fruits. Extent of any damage must be estimated by comparing the amount of fruit left on the tree to what would have been produced under normal conditions. It is difficult for insurers to estimate the amount of crop loss because it must be determined using variables which can be measured after the frost event and it must be performed quickly in the field without complex equipment.

A close relationship has been shown previously between crop load and specific phenotypic traits in fruit trees. For instance, tree canopy size per available area is related to the amount and efficiency of light interception and has been shown to be a good parameter to estimate potential yields in peach (Miranda and Royo, 2003a), pear (Miranda and Royo, 2003b), apple (Miranda and Royo, 2004) and Japanese plum (Miranda *et al.,* 2008) orchards. Also scaffold crop density was found to be correlated with spur density per scaffold and scaffold cross sectional area in sweet cherry (Santesteban *et al.,* 2008). Yield potential increases with tree size, but not linearly, since larger trees are less efficient (Faust, 1989). In walnut trees, lateral bearing habit is moderately correlated with early leafing, tree architecture, and precocity (Germain, 1990). Lateral flowering trees developed flowers earlier in the spring and had better yield potential than terminal bearers (Solar *et al.,* 2001). Forde and McGranahan (1996) reported a negative correlation between tree height and yield. In contrast, Atefi (1990) found a strong positive correlation of yield with tree diameter and height but a negative correlation with nut weight. Reported correlations of yield with protogynous or protandrous flowering habits are conflicting (Kornienko, 1974; Majacka, 1971). Correlations among various walnut traits have been also reported by Komanich (1980), Sharma (1996), Sholokhov (1974) and Atefi (1990).

Several mathematical models have been developed to estimate yield in fruit trees (Miranda and Royo, 2003a; Miranda and Royo, 2003b; Miranda and Royo, 2004; Miranda *et al.,* 2008; Santesteban *et al.,* 2008). Hassani *et al.* (2010a and 2010b) reported a mathematical model for walnut yield using trunk cross sectional area as the main yield predictor supplemented by coefficients for several additional traits such as tree spacing, predicted lateral bearing class and annual shoot growth that could be evaluated before leafing date and used to determine crop insurance. Trunk circumference, trunk cross sectional area, canopy radius, lateral bearing percentage, tree growth habit, and annual shoot growth are some additional traits that could be employed to estimate the yield potential of walnut trees. The objective of this study was to develop a regression model, using parameters which can be measured easily and rapidly, to estimate the full yield potential of walnut orchards.

Materials and methods

The study area

Data were collected during 2010 from seedling walnut trees, ranging between 10 and 50 years old. The trees were located in Koleeim (40 trees) and Jezla (39 trees), two main walnut growing villages in Tarom city of Zanjan province in Iran.

Plant material

Walnut trees were cultivated sporadically, singular or quasi-singular in these regions. The selected trees represented a broad range of growth condition (Table 1). Cultural practices, including irrigation, fertilization etc., were carried out as is normally done in the area.

Table 1. Trunk circumference, lateral bearing, canopy cross-sectional area, canopy radius, mean annual shoot growth, and yield per tree (mean ± standard deviation). In Koleeim, Jezla, and combined locations.

Characteristic	Koleeim		Jezla		Combined Locations	
	Mean	SD	Mean	SD	Mean	SD
Nº of sampled trees	40		39		79	
Trunk circumference (TC, cm)	156.3	55.47	170.07	85.72	163.1	71.88
Lateral bearing (LBF, %)	3	6.77	2.05	6.25	2.53	6.5
Canopy area (CA, m^2)	172.79	109.23	156.9	98.65	165	103.8
Trunk cross-sectional area (TCA, cm^2)	2184	1666	2873.1	3544	2524	2762
Canopy radius (CR, m)	7.08	2.24	6.77	2.05	6.93	2.14
Mean annual shoot growth (SHGPY, cm)	18.8	5.38	9.87	3.157	14.39	6.28
Yield per tree (Y, kg)	21.11	19.89	26.02	18.67	23.54	19.33

Measurement of traits

Trunk circumference was measured 30 cm above the soil surface and used to calculate trunk cross sectional area. Canopy radius was calculated as the average of the extent of canopy spread in four directions for each tree. Lateral bearing percentage was determined as the average of lateral buds of four one-years-old shoot in different directions of a tree, which bears flowers in April. Lateral bearing was scored at time of the end of female bloom and base the percent on possible flowering positions that produce a flower. Tree structure was scored as three categories - 1) spreading, 2) semi-spreading and 3) upright. Yearly shoot growth was calculated from the average length of 20 one-year-old (previous season's growth) shoots randomly selected from all sides of each tree. At harvest, nuts per tree were counted; and10 randomly-chosen nuts per tree were weighed to obtain mean nut weight (NW). Yield was calculated by multiplying mean fruit weight by nut number per tree (Table 2).

Table 2. Independent variables, with their corresponding acronyms and units used in this study.

Variables	Acronyms	Units
Trunk circumference	TC	Cm
Lateral bearing	LTB	%
Canopy area	CA	m2
Trunk cross-sectional area	TCA	cm2
Canopy radius	CR	m
Mean annual shoot growth	SHGPY	cm
Yield per tree	Y	kg

Simultaneously, in order to test the validity and applicability of the models for crop estimation, data were collected using the same methods from 11 trees in Koleeim and 20 in Jezla and data was pooled to produce a validation data set for the two villages combined.

Multiple linear regression models linking yield in estimation data sets were examined with the parameters that measured in the trees for each village and sum of the two villages. The models obtained were evaluated: yield was the dependent variable and the rest were independent variables. Independent variables, with their corresponding acronyms and units used in this study are shown in Table 2.

The relationships were determined by fitting polynomial regression models with the linear regression and the stepwise regression. Analysis of variance was performed to test quality of models for different villages and the sum of two villages (Table 3).

The validity of the models was tested using F test, coefficient of determination (R^2) and adjusted coefficient of determinations (adj R^2). Variance inflation factor (VIF) for each of the parameter estimates was calculated to detect inter-correlation between variables (Miranda and Royo, 2003a, b, Miranda *et al.,* 2008) (Table 3).

Table 3. Parameter estimates and statistics from the regression model for yield (Y, weight of fruit/tree) prediction.

Statistic or parameter estimate	**Koleeim**		**Jezla**		**Pooled data for both locations**	
	Estimation model	**Validation model**	**Estimation model**	**Validation model**	**Estimation model**	**Validation model**
Nº of trees sampled	40	11	39	20	79	31
Regression DF	2	2	2	2	2	2
Error DF	38	9	37	18	77	29
Total DF	40	11	39	20	79	31
Regression sum of squares (RSE)	27998	4946.8	34796	23642	62175	26977
Error sum of squares (SSE)	5264	326.7	4865	6488	10748	8426
Total sum of squares (TSE)	33262	5273.5	39661	30130	72923	35404
Mean square regression(MSR)	13999	2473.4	17398	11821	31087	13489
Mean square error (MSE)	138.5	36.3	131.5	360	139.6	290.55
F value	101.06**	68.51**	132.33**	32.79**	222.71**	46.42**
Regression coefficient for TC	0.12465**	0.09341**	0.15071**	0.17946**	0.14049**	0.14999**
Standardized regression coefficient for TC	0.7159	0.72497	0.8978	0.85936	0.82333	0.8
Variance inflation factor for TC	1.2464	1.08185	1.09242	1.20625	1.15	1.16349
Regression coefficient for LBF	1.32889**	1.71770**	0.55075**	0.27074**	0.93873**	0.69344**
Standardized regression coefficient for LBF	0.33786	0.47307	0.11234	0.0599	0.21429	0.15958
Variance inflation factor for LBF	1.2464	1.08185	1.09242	1.20625	1.15	1.16349
Partial regression coefficient for TC	0.7502	0.7312	0.8658	0.7817	0.8128	0.7401
Adjusted partial regression coefficient for TC	0.7438	0.7043	0.8623	0.7702	0.8104	0.7314
Partial regression coefficient for LBF	0.0915	0.2069	0.0115	0.003	0.0398	0.0219
Adjusted partial regression coefficient for LBF	0.0896	0.22	0.0084	0.0095	0.0384	0.0142
Coefficient of determination R^2	0.8417	0.9381	0.8773	0.7847	0.8526	0.762
Adjusted coefficient of determination (Adj R^2)	0.8334	0.9243	0.8707	0.7607	0.8488	0.7456

**Significant at 1% probability level

To investigate the estimated models validation and efficiency in this study, following approaches were used.

1. The estimation and validation models were statistically compared for each village and for the sum of two villages combined using Student's t test (Table 4).

2. Estimated yields (ŷ) were calculated using step-by-step regression of data collected for trunk circumference and lateral bearing for each village and the two combined. Then the means of the actual yield (Y) and the estimated yield, as calculated through fitted models, were compared using Student's test for paired observations (Table 5).

3. The simple correlation coefficient was obtained for actual average yield and estimated average yield as calculated by the estimation model for each village and for the combined sum of two villages (Table 6).

As if for comparison of the results of two sets of data, and to be more confident about the determined models in the locations that justify each other, the locations were considered separately in this study. However, the final estimation of parameters was accomplished by pooling data from two locations to develop the final model (Myers, 1990).

Table 4. Comparison of parameters used in the estimation and validation models for each location and for pooled data from both locations.

Location	Variables	Model	Coefficient	Variance of coefficient difference	Standard error	t_c
Koleeim	TC	Estimation	0.12465	0.00028	0.01676	1.86^{ns}
		Validation	0.09342			
	LTB	Estimation	1.32889	0.1784	0.4224	-0.92^{ns}
		Validation	1.7177			
Jezla	TC	Estimation	0.15071	0.06305	0.251	-0.11^{ns}
		Validation	0.17946			
	LTB	Estimation	0.55076	0.38182	0.6179	0.45^{ns}
		Validation	0.27074			
Combined locations	TC	Estimation	0.14049	0.00039	0.01999	-0.47^{ns}
		Validation	0.14999			
	LTB	Estimation	0.93873	0.22268	0.47189	0.52^{ns}
		Validation	0.69344			

ns = non-significant

Table 5. Comparison of the differences between means of the actual yield and the estimated yield for each location and for pooled data for the combined locations.

Model	Location	Difference in mean of estimated and actual yield	Standard deviation	Standard error	T
Estimation	Koleeim	-2.36	11.37	1.80	-1.31^{ns}
	Jezla	-0.74	11.29	1.81	-0.41^{ns}
	Pooled data	-1.75	11.60	1.30	-1.34^{ns}
Validation	Koleeim	0.025	5.71	1.72	0.01^{ns}
	Jezla	-4.097	18	4.02	-1.02^{ns}
	Pooled data	-3.29	16.42	2.95	-1.12^{ns}

ns = non-significant

Table 6. Mean, minimum, maximum, standard deviation, and correlation coefficients of actual and estimated yields of 40, 39 and 79 walnut seedling trees randomly sampled for estimation models and 11, 20 and 31 randomly sampled for validation models at each location and for combined data from the two locations.

Model	Location	Yield	Mean	Standard deviation	Min	Max	Correlation coefficient of actual and estimated yield
Estimation	Koleeim	actual	21.11	19.89	2.07	90.19	0.85**
		estimated	23.47	12.37	10.97	63.14	
	Jezla	actual	26.02	18.67	1.06	75.38	0.75**
		estimated	26.76	13.44	12.21	75.35	
	Combined	actual	23.54	19.33	1.06	90.19	0.77**
		estimated	25.29	12.22	11.38	70.24	
Validation	Koleeim	actual	18.72	11.9	6.33	48.83	0.88**
		estimated	18.7	10.5	8.4	48.83	
	Jezla	actual	27.11	28.5	1.85	98.5	0.84**
		estimated	31.21	14.8	13.1	61.73	
	Combined	actual	24.14	24.04	1.85	98.5	0.81**
		estimated	27.43	11.03	10.95	59.61	

**Significant at 1% probability level

Results and discussion

Stepwise regression

Stepwise regression showed that trunk circumference (TC) and lateral bearing (LTB) as independent variables had a significant effect on yield. The estimated statistical parameters for these for each location and for the pooled data for both locations are presented in Table 3. The effects of other variables were not significant. The variance inflation factors for the variables used were less than 10, so it indicates absence of colinearity (Neter *et al.,* 1996). Analysis of variance showed the regression was significant at the 1% level. Coefficients of determination (R^2) of the models were 0.84, 0.87 and 0.85 for Koleeim, Jezla and the combined locations. The R^2 of the validation data were 0.93, 0.78 and 0.86 for Koleeim, Jezla and the pooled data respectively. The adjusted R^2 values were quite similar for the estimation and validation models in each location, providing some assurance about the applicability of the model (Table 3) (Neter *et al.,* 1996). The standardized regression coefficients as direct effects of each parameter on the dependent variable showed that TC (Trunk circumference) had the greatest impact on yield (Table 3). This result is supported by previous works on peach, pear, apple (Miranda and Royo 2003a, b, Miranda and Royo 2004) and walnut (Atefi 1990, Hassani *et al* 2010). The highly significant correlation observed between lateral bearing and crop load is also consistent with previous reports (Amiri *et al.,* 2010, *Hansche et al.,* 1972 and Solar *et al.,* 2001).

Student's t test showed that the differences obtained in the regression coefficients of independent variables using data for model prediction and data for validation were not significant (Table 4), indicating that the validation model for each location strongly supported its related estimation model. Therefore, estimation models were considered reliable and could be used for prediction of yield in walnut trees.

Estimation of yield

The estimated yield using TC and LTB as independent variables and the actual yield are shown in Table 5. The Student's t test showed differences between means of actual and estimated yield (Y, ŷ) at each location and the combined location were not significant.

Correlation

The Table 6 shows the correlation coefficients of actual and estimated yields using estimation and validation data for each location and pooled data of both locations. These were significant at 0.01 statistical probability levels, indicating significant correlation between actual and estimated yields for each location and the pooled data.

To evaluate the predictive ability of the models, the estimates yields (ŷ) were plotted against the actual values (Y) (Fig. 1). The line of equality depicts the accuracy of the estimation versus actual value in both locations.

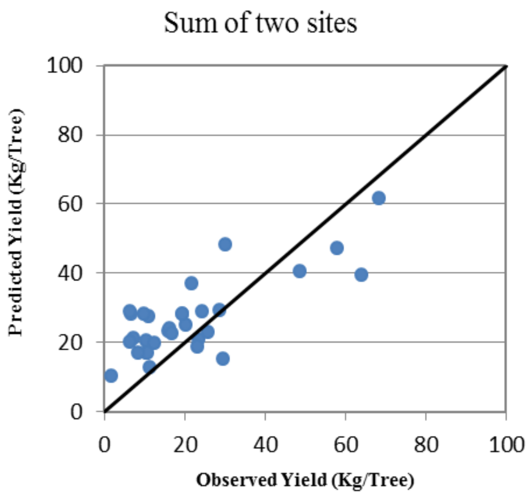

Fig. 1. Plot of the estimated yield per tree (ŷ), calculated from multiplying TC x LTB vs. the actual yield per tree (Y) for each tree in the validation data set for Koleeim, Jezla and Sum of two villages (Combined). The solid line is the line of equality on which all points would lie if the estimation method gave the true value for every tree.

The accuracy of this prediction method versus actual values was evaluated by means of a plot (Bland and Altman, 1986). Plotting values also allows investigation of possible relationships between measurement errors and true values. The plot in Fig. 2 indicates no obvious relationship between the predicted and actual yields. Lack of agreement between an estimation method and actual values can be determined by calculating the bias, estimated by the mean of the differences and their SDs. Fig. 2, shows the mean of the variances and the ±2SD interval. These values are referred to as the limits of agreement (Bland and Altman, 1986), and providing that the differences within them are not important, and the estimation method can be used to predict the yield.

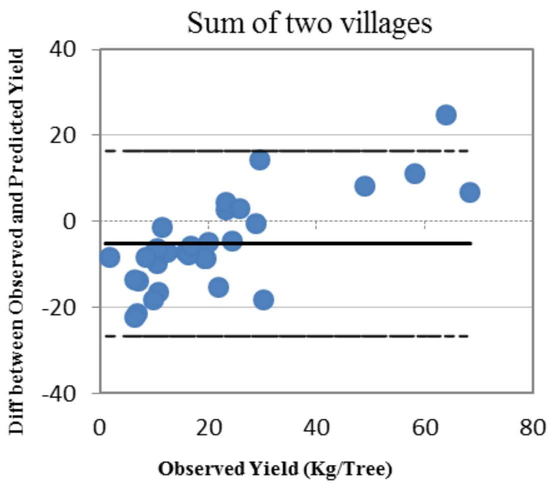

Fig. 2. The difference between estimated yields per tree (ŷ), calculated from multiplying TC × LTB vs. the actual yield per tree (Y) for each tree in the validation data set for Koleeim, Jezla and Sum of two villages (Combined). The solid Line is the mean of the differences. The broken lines are the limits of agreement, calculated as d±2SD, where d=the mean of the differences, and SD= the standard deviation of the differences.

We conclude that this multiple regression model, using trunk circumference and lateral bearing habit as independent variables, can be used to estimate the unrealized yield potential of sporadically, singular or quasi-singular cultivated Persian walnut trees that are prevalent in many countries. When contrasted with the actual measured yield, this can provide a useful tool for determining the quantity, and hence value, of crop loss.

References

Amiri R, Vahdati K, Mozaffari MR, Leslie CA. 2010. Correlations between Some Horticultural Traits in Walnut. Horticultural Sciences **45,** 1690– 1694.

Atefi J. 1990. Preliminary research of Persian walnut and correlation between pair characters. Acta Horticulture **284,** 97- 104.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1,** 307- 310.

Faust M. 1989. Physiology of temperate zone fruit trees. Wiley. N.Y.

Fjellstrom RG Parfitt DE. 1994. Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphisms. Genome **37,** 690–700.

Forde HI McGranahan GH. 1996. Walnuts, p. 241– 274. In: Janick, J. and N. Moore (eds.). Fruit breeding.Vol.3, Nuts. John Wiley and Sons, Inc, New York, NY.

Germain E. 1990. Inheritance of late lateral bud fruitfulness in walnut (*Juglans regia* L.), phenotypic correlation among some traits of the tree. Acta Horticulture. **284,** 125- 134.

Hansche PE, Beres V, Forde HI. 1972. Estimation of quantitative genetic properties of walnut and their implications for cultivar improvement. Journal of American Society Horticultural Sciences **97,** 279- 285.

Hassani D, Jafaraghaee M, Rezaee R, Eskandari S, Mozaffari MR. 2010a. Determination of the yield potential of walnut trees. Final Report of Project. Seed and Plant Improvement Institute, Karaj, Iran (in Persian).

Hassani D, Dastjerdi R, Jafaraghaee M. 2010b. Guideline for evaluation and determination of factors causing damages in walnut orchards.Final Report of Project. Seed and Plant Improvement Institute, Karaj, Iran (in Persian).

Komanich IG. 1980. Correlation between walnut fruit characters in a seed population. Bull. Academic Science RSS Mold **2,** 27– 30.

Kornienko NA. 1974. Types of dichogamy in walnut. P1. Br. Ab. 44, 452.

Loague K, Green RE. 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. Journal of Contaminant Hydrology **7,** 51- 73.

Marini RP. 1999. Estimating apple diameter from fruit mass measurements to time thinning sprays. Horticultural Technology **9,** 109– 113.

Majacka AD. 1971. Dichogamy and fertility in walnut. P1. Br. Ab. 41, 186.

Miranda C, Royo JB. 2003a. A statistical model to estimate potential yields in peach before bloom. Journal of American Society Horticultural Sciences **128,** 297– 301.

Miranda C, Royo JB. 2003b. Statistical model estimates potential yields in pear cultivars 'Blanquilla' and 'Conference' before bloom. Journal of American Society Horticultural Sciences **128,** 454–457. journal.ashspublications.org/content/128/4/452.abstract

Miranda C, Royo JB. 2004. Statistical model estimates potential yields in 'Golden Delicious' and

'Royal Gala' apples before bloom. Journal of American Society Horticultural Sciences **129,** 20–25.

Miranda C, Santesteban LG, Royo JB. 2008, Establishment of a Model to Estimate Crop Load on Japanese Plum (*Prunus salicina* Lindl.) before Bloom. Acta Horticulture **803,** 307- 312.

Myers RH. 1990. Classical and modern regression with applications. 2nd ed. PWSKent, Boston.

Neter J, Kutner MH, Nachtshein CJ. Wasserman W. 1996. Applied linear regression models. 3rd ed. Irwin. Homewood Ill.

Santesteban LG, Miranda C, Royo JB. 2008. Establishment of a Model to Estimate Crop Load on Sweet Cherry (*Prunus avium* L.) before Bloom. Acta Horticulture **803,** 313- 318.

Sharma OC. 1996. Variability in seedling trees of Persian walnut in Solan area of Himachal Pardesh. MSc thesis, Solan UHF University.

Sholokhov LV. 1974. Quality characters of fruit in walnut and correlation between them. Tr. Novocherkas. Inzha Melior. Inta **13,** 106– 112.

Solar A, Hudina M, Stampar F. 2001. Relationship between tree architecture, phonological data and generative development in walnut (*Juglans regia* L.). Acta Horticulture **544,** 275– 285.

Soltani A, Robertson MJ, Torabi B, Yousefi-Daz M, Sarparast R. 2006. Modelling seedling emergence in chickpea as influenced by temperature and sowing depth. Agriculture and Forest Meteorology **138,** 156– 167.

Vahdati K. 2000. Walnut situation in Iran. Nucis-Newsletter **9,** 32– 33.