Advances in Research

16(6): 1-7, 2018; Article no.AIR.43391
ISSN: 2348-0394, NLM ID: 101666096

Statistical Model for Yield Estimation of 'Gala Red Lum' Apples before Bloom in Northern India

Tabasum Mushtaq[1*], S. A. Mir[1], Nageena Nazir[1], T. A. Raja[1], A. H. Pandith[2], Khalid Rasool[2] and Mushtaq A. Lone[1]

[1]Division of Agricultural Statistics, Sher-e-Kashmir University of Agricultural Sciences and Technology (SKUAST-K), Srinagar, India.
[2]Division of Fruit Sciences, Sher-e-Kashmir University of Agricultural Sciences and Technology (SKUAST-K), Srinagar, India.

Authors' contributions

This work was carried out in collaboration between all authors. Author TM designed the study, performed the statistical analysis, wrote the protocol, and wrote the first draft of the manuscript. Authors SAM, NA and MAL managed the analyses of the study. Author TAR, AHP and KR managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AIR/2018/43391
Editor(s):
(1) Dr. Ibrahim Elbasyoni, Department of Agronomy and Horticulture, University of Nebraska, Lincoln, USA.
Reviewers:
(1) Elvis Djam Miensah, University of Ghana, Ghana.
(2) Kunio Takezawa, Institute for Agro-Environmental Sciences, NARO (NIAES), Japan.
(3) Mujahid Khan, SKN Agriculture University, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/26812

Original Research Article

Received 18 July 2018
Accepted 13 October 2018
Published 24 October 2018

ABSTRACT

In order to solve the problems of damage in crops due to natural disasters like frosts and hail storm, it is common to insure crops against the damage. After damage, crop loss must be evaluated by comparing what amount of crop is left with the amount that would have been obtained under normal conditions. Potential crop must be evaluated quickly through the use of measurements obtained at the beginning of the cycle of tree's growth. The objective of the study was to develop the best fitted model for estimating yield on the population of Gala Red Lum plants before bloom. The data used for this research were primary data collected from high density apple block SKUAST-Kashmir (HDP, Plate-1). The study was undertaken at experimental field of Division of Fruit Science, SKUAST-Kashmir, Srinagar, J&K, India, during the years 2015 and 2016. The measurements of various tree/fruit characteristics in Gala Red Lum were recorded. The developed model was validated using k-fold cross validation technique and bootstrap validation.

**Corresponding author: E-mail: tabumushtaq2010@gmail.com;*

Keywords: Yield estimation; regression models; validation.

1. INTRODUCTION

Jammu and Kashmir State being endowed with natural advantages of topography and climate with an enormous diversity of agro-climatic condition has immense scope for horticultural development. Apple ranks first covering 43.30 percent area and 80.18 percent production. During the last 30 years, yield of apple has shown an increase from 4.12 to 10.00 MT/ha in Jammu and Kashmir. Though it appears to be highest among the apple producing states in the country, yet it is far below the level achieved by advanced countries where productivity is of 50-60 MT/ha. Jammu and Kashmir government is now emphasising on use of high density plantation (HDP) system, which gives the fruit production within 2-3 years after plantation. From last few years, progressive farmers are switching over to high density apple plantation and the university SKUAST-K is following the pursuit as well to develop the new technologies so that farmers adapting HDP system is benefited. Natural disasters like spring frosts, hail storms etc. can cause damage to the crop, so it is common to insure crops against the damage. These natural calamities usually affect flowers or young developing fruits. It is necessary to evaluate the loss so that any damage must be estimated comparing the amount of fruit left on the tree to the number that would have been produced under normal conditions. Insurance companies face two main difficulties in their work: the study must be performed after the damage has occurred (i.e., must be performed using variables which can be measured after the damage) and it must be performed in the field, quickly and without complex equipment.

Tree yield is a function of the number of both fruit on the tree and their weight. The number of fruit depends on the flower density (number of flower clusters per TCA, [1]), the tree's size and percentage of set. The percent of fruit set decreases as the flower density increases, but this relationship is heavily influenced by climate [2,3]. When fruit set is too large for the tree's size, it becomes necessary to thin to ensure both return bloom and a good fruit development [4].

Typical fruit size is different for each cultivar [5], although it is also highly dependent on the number of fruit left on the tree [6,7,8], climate [9,10] and cultural techniques [11,12]. Fruit growth rate is faster for early cultivars but growth time is shorter, final size tends to be smaller [5].

Only tree size, plantation density and number of flower buds after pruning are available at the beginning of the active period to estimate yield in a simple and rapid way. Potential yield increase with tree size, although not linearly since bigger trees are less efficient [13,5,14]. Many studies have shown that tree size and trunk cross-sectional area (TCA) are closely related, so that TCA is used regularly to compare different plots vigor (TCA/ha), efficiency (kg/TCA), flower load (number of flowers/TCA), etc. [13,15,16,1,17,18]. Plant density also influences yield, since it affects the amount of light intercepted by the trees. Robinson and Lakso [19] found that the amount of light intercepted by a tree was correlated with TCA per available area (TCA/ha).TCA/ha has been shown to be a good parameter to estimate potential yields in peach, nectarine and pear orchards [8,20].

The productivity of an apple tree also depends on the bud load after pruning, i.e., the magnitude of flowering [2,21,1]. Some pomologists have proposed flower bud density (number of flower buds per TCA) as an index to express the magnitude of flowering [1]. Flower density per land area (number of flower buds/m^2) in pear cultivars 'Blanquilla' and 'Conference' Miranda and Royo [8] was found to be the most influential parameter in multiple regression models that allowed cluster set and cluster yield estimation.

The focus of the study was confined to the statistical evaluation of the biometrical characters in high density apple orchards and to develop model for 'Gala Red Lum' before bloom that can be employed by the users to predict yield, using parameters which can be easily measured before the beginning of the active period.

2. MATERIALS AND METHODS

2.1 Data Collection

Keeping in view the specific objective of present study, data on population of *Gala Red Lum* plants after the harvesting season were collected on the yield and biometrical characters of apple trees of the block namely (HDP, Plate-1) located at Shalimar campus of Division of Fruit Science, Sher-e-Kashmir University of Agricultural Science & Technology of Kashmir, Shalimar, Srinagar, Jammu & Kashmir in Northern India. The exotic variety '*Gala Red* Lum' of apple grafted on M-9 T337 rootstock was introduced by SKUAST-Kashmir in spring 2013 from an Italian nursery, GRIBA, Italy. The plant material was one year old with 3 plus feathers. These two year old trees with uniform size, vigor and bearing capacity were used for the study. All these trees received uniform cultural practices during the years under study as per the package of practices of SKUAST-Kashmir. The required data for the present study included only primary data. The orchard consisted of three blocks of '*Gala Red Lum*' having 49 trees in each block, with block spacing of 3 meters and tree spacing 1.5 meters. For development of yield estimation/ prediction model (before bloom) on the population of Gala Red Lum independent variables (parameters) used were trunk cross sectional area (TCA (cm^2) = (Trunk diameter/2)^2 × π), tree height (Height, m), canopy area (CA (m^2)= [(Canopy width + Canopy length)/4)]^2 × π, [22]), canopy volume (CV (m^3)= (Canopy area × canopy height)/2, [23]), number of primary branches (PB), branch density (BD = number of primary branches/TCA) and number of flower buds (FB). Primary data collected individually for two years (2015 and 2016) was pooled to average out the on/off bearing effect and a population of 270 plants was used for the development of models in *Gala Red Lum*. Models were first developed for population and then were tried on sample of 41 trees.

2.2 Model Building

Relationship between yield & biometrical characters of tree and fitting the regression models in biological studies is gaining a lot of importance in horticulture. Multiple regression equations play a vital role in the prediction of yield of fruit crops, thus an effort is made to fit a multiple linear regression equation considering apple yield as a dependent variable and biometrical characters as explanatory variables. To estimate the yield for an apple plot before bloom on the basis of various parameters of apple trees, multiple linear regression models linking parameters that could be measured before the beginning of the activity cycle were examined. The models used for the present study were linear, log-linear, single variable, two variable and multiple variable models.

SET	Linear models	SET	Log-linear models
Set A	$Y=\beta_0+\beta_1 X_1$	Set E	$LnY=\beta_0+\beta_1 LnX_1$
	$Y=\beta_0+\beta_1 X_1^2$		$LnY=\beta_0+\beta_1 LnX_1+\beta_2 LnX_2$
Set B	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2$		$LnY=\beta_0+\beta_1 LnX_1 X_2+\beta_2 LnX_2^2$
	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2^2$		
	$Y=\beta_0+\beta_1 X_1^2+\beta_2 X_2^2$		
Set C	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_1 X_2$		
	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_1^2 X_2$		
	$Y=\beta_0+\beta_1 X_1 X_2+\beta_2 X_2^2$		
Set D	$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_5 X_5$		

Many models were developed on the population of 'Gala Red Lum' having single variables, two variables and more than two variables. In all the models Yield (Y) was the dependent variable and the independent variables were trunk cross-sectional area (TCA), Height, canopy area (CA), canopy volume (CV), number of primary branches (PB), branch density (BD) and number of flower buds (FB). The relationships were evaluated by fitting the linear, log-linear and polynomial regression models with the linear regression procedure of R- software and a backward stepwise elimination option. The models were given the ranks on the basis of coefficient of determination (R^2), adjusted R^2 and AIC. The models with highest value of adjusted R^2 and least AIC value were selected for sample study. Residuals were analysed to determine the presence of outliers and non constant error variance. The best model in Gala Red Lum was Yield= TCA+CA+FB and showed the significance at 5% with the highest value of adjusted R^2 (0.9821).

2.3 Model Validation

The models that were tested on sample population were cross validated by using k-fold cross validation technique. In our study we used k=10. The data was partitioned into 10 subsamples. Of the 10 subsamples, a single subsample is retained as validation data for testing the model, and remaining (10-1) subsamples are used as training data. The cross validation was repeated 10 times with each of the 10 subsamples used exactly once as validation data. Various characteristics like root mean square error (RMSE), mean absolute error (MAE) and cross validated R^2 were obtained during the process. Adjusted R^2 obtained during the normal fitting was compared with the cross validated R^2. The root mean square error (RMSE) has been used as a standard metric to measure model performance. Mean absolute error (MAE) was also considered a better metric to choose a model. The process was done on the yield predicting models. R software was used for validation. The Bootstrap method has been used to quantify the uncertainty associated with a predictive model. It consisted of randomly selecting a sample of *n* observations from the original data set. This subset, called bootstrap data set has been used to evaluate the model. This procedure was repeated a large number of times and the standard error of the bootstrap estimate were calculated. The results provide an indication of the variance of the model performance. The sampling, in this case, was performed with replacement, which means that the same observation can occur more than once in the bootstrap set. Bootstrap with 100 resample was used to test the selected model.

3. RESULTS AND DISCUSSION

The cross validation results obtained are presented in (Table 1). R^2 and adjusted R^2 in the model at S. No. 4 is more as compared to other models. This model has the least value of RMSE (0.3190) and MAE (0.2510) and the highest value of cross validated R^2 (0.9900). Thus we select the model $YIELD = 2.43 + 1.16TCA + 2.46CA + 0.01FB$ for predicting yield (before bloom) in Gala Red Lum. Fig. 1 displays the fitting of the selected model. It is clear from the graph that the model shows the best fit. *k* -fold cross validation of the selected model is shown in (Table 2).

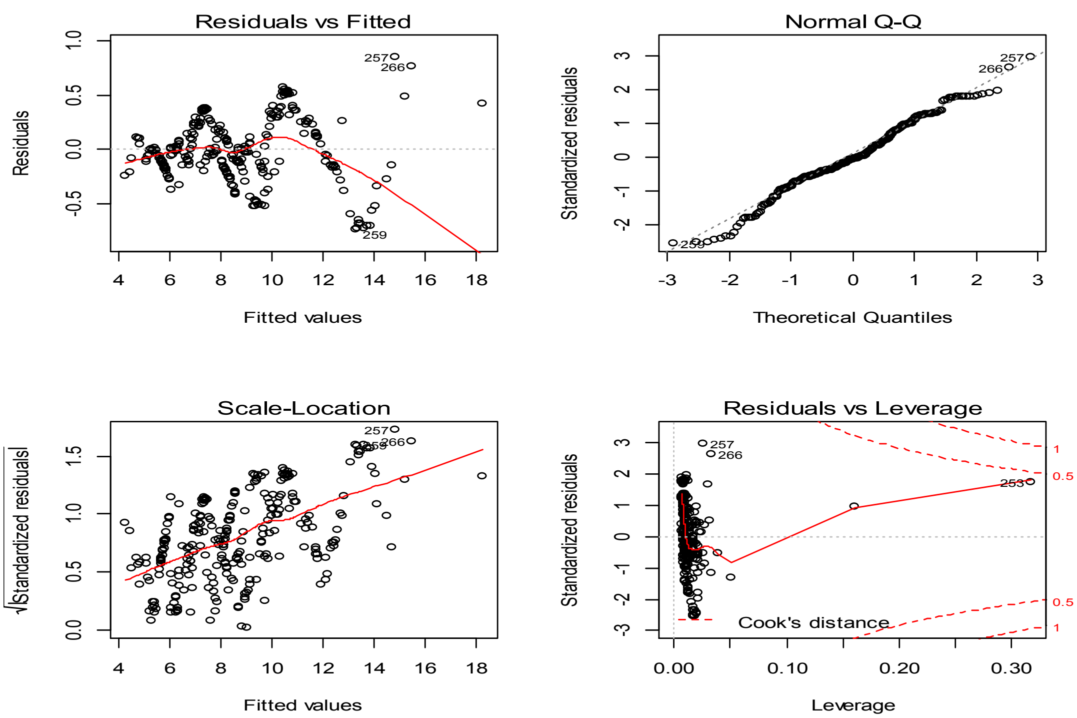

Fig. 1. Fitting of Y= TCA+CA+FB

Table 1. Cross validation of selected yield predicting models (before bloom) in Gala Red Lum

S. No.	Normal fitting			Cross validated		
	Model	**R^2**	**Adj.R^2**	**RMSE**	**MAE**	**CV R^2**
1	$Y = \beta_0+\beta_1TCA$	0.8779	0.8748	0.677	0.562	0.9110
2	$Y=\beta_0+\beta_1TCA+ \beta_2CA$	0.9794	0.9783	0.333	0.278	0.9880
3	$Y=\beta_0+\beta_1TCA+\beta_2CA+\beta_3TCA*CA$	0.9826	0.9812	0.3292	0.2882	0.9899
4	$Y=\beta_0+\beta_1TCA+\beta_2CA+\beta_3FB$	0.9835	0.9821	0.3190	0.2510	0.9900

Table 2. *k*-fold cross validation of selected yield predicting model Y= TCA+CA+FB

Folds	Obs.	Predicted	CV predicted	Actual	CV residual	SS	MS	Test set obs.
1	2	10.023	10.042	9.940	-0.102	0.17	0.04	4
	8	8.539	8.548	8.530	-0.018			
	19	8.189	8.204	7.990	-0.214			
	38	13.660	13.705	13.370	-0.335			
2	3	8.894	8.916	8.690	-0.226	0.86	0.17	5
	11	7.359	7.345	7.710	0.365			
	25	13.186	13.307	12.790	-0.517			
	33	13.091	13.209	12.610	-0.599			
	36	8.252	8.263	8.040	-0.223			
3	5	5.811	5.825	5.600	-0.225	0.55	0.14	4
	12	6.468	6.485	6.490	0.005			
	20	10.403	10.365	11.040	0.675			
	30	9.121	9.087	9.290	0.203			
4	4	7.346	7.326	7.680	0.354	0.34	0.08	4
	28	8.112	8.117	7.940	-0.177			
	35	10.229	10.226	10.630	0.404			
	40	9.547	9.544	9.420	-0.124			
5	6	5.539	5.549	5.460	-0.089	0.04	0.01	4
	21	12.541	12.532	12.430	-0.102			
	29	12.063	12.054	12.200	0.146			
	32	7.805	7.813	7.770	-0.043			
6	1	7.360	7.350	7.730	0.380	0.31	0.08	4
	18	8.784	8.769	8.680	-0.089			
	37	8.937	8.955	8.990	0.035			
	39	6.343	6.409	6.020	-0.389			
7	9	5.836	5.859	5.610	-0.249	0.10	0.02	4
	26	9.095	9.102	9.180	0.078			
	27	9.548	9.555	9.420	-0.135			
	34	12.295	12.276	12.370	0.094			
8	16	12.815	12.665	13.060	0.395	0.85	0.21	4
	17	9.357	9.369	8.950	-0.419			
	31	10.650	10.683	11.230	0.547			
	41	13.330	13.328	12.860	-0.468			
9	10	6.724	6.736	6.580	-0.156	0.53	0.13	4
	14	8.114	8.067	8.300	0.233			
	22	10.759	10.708	11.320	0.612			
	24	8.366	8.379	8.110	-0.269			
10	7	6.639	6.653	6.540	-0.113	0.49	0.12	4
	13	6.810	6.824	6.600	-0.224			
	15	7.030	7.044	7.200	0.156			
	23	10.389	10.359	10.990	0.631			

Overall MS = 0.103

Table 3. Bootstrap validation of yield predicting model (before bloom) in Gala Red Lum

Ordinary nonparametric bootstrap						
	Original	**Bias**	**S.E**	**RMSE**	**MAE**	**R^2**
t1*	2.4318	-0.02647	0.083	0.359	0.287	0.9780
t2*	1.1598	0.02455	0.081			
t3*	2.4557	-0.26022	0.689			
t4*	0.0117	0.00255	0.007			

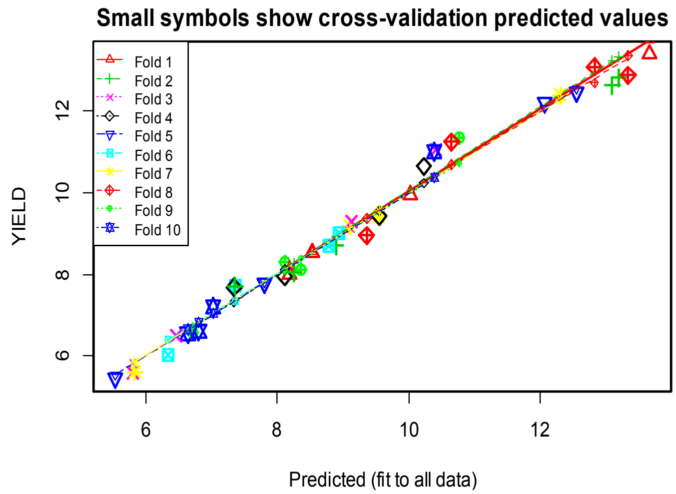

Fig. 2. k-fold cross validation of the model yield = 2.43+ 1.16TCA+ 2.46CA+0.01FB

Fig. 2 displays Cross validation of selected yield predicting model (before bloom) in Gala Red Lum. Here we can see that each fold is near the best fit which reveals the closeness of the actual values and predicted values. Thus from the plot also we conclude that the model $YIELD = 2.43 + 1.16\,TCA + 2.46\,CA + 0.01\,FB$ is the best fit model for predicting the yield (before bloom) in Gala Red Lum.

The output showed the average performance of the selected models across the 100 resample. RMSE (root mean squared error) and MAE (mean absolute error), represent two different measures of the model prediction error. Lower the RMSE and the MAE, the better the model. The R^2 represents the proportion of variation in the outcome explained by the predictor variables included in the model. Higher the R^2 better is the model. Table 3 provides the output of the model accuracy by bootstrap statistics of selected yield predicting model. The 'original' column corresponds to the regression coefficients. The associated standard errors are given in the column 'S.E', t_1 corresponds to the intercept and t_2, t_3, t_4 correspond to the independent variables of the selected yield predicting model (before bloom) in Gala Red Lum. It also depicts the corresponding values of RMSE (0.359), MAE (0.287) and R^2 (0.9780) for the selected model. The model showed good performance with less bias and hence can be used for predicting yield in 'Gala Red Lum' before bloom.

4. CONCLUSION

Trunk cross sectional area (TCA), canopy area (CA) and number of flower buds (FB) had the greatest influence and a positive effect in explaining Yield. It was observed that these parameters amongst various other parameters studied had the significant effect on the yield of '*Gala Red Lum*'. Predictive model developed based on TCA, CA and FB parameters forecasts the yield of '*Gala Red Lum'* apple efficiently before bloom. Hence provides useful tool for early forecast of yield and evaluation of '*Gala Red Lum*' crop losses due to natural calamities like hail storms etc.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Lombard PB, Callan FG, Dennis NE, et al. Towards a standardized nomenclature, procedures, values and units in determining fruit and nut tree yield performance. Hort Science. 1988;23:813-817.
2. Dennis FG. Apple. 1st ed. CRC Press, Boca Raton, Fla.p1-41. In: S.P. Monselise (ed.). CRC handbook of fruit set and development; 1986.
3. Tromp J, Borsboom O. The effect of autumn and spring temperature on fruit set and on the effective pollination period in apple and pear. Scientia Hort. Sci. 1994;124:468-477.
4. Wertheim SJ. Developments in chemical thinning of apple and pear. Plant Growth Regulat. 2000;31:85 -100.

5. Faust M. Physiology of temperate zone fruit trees. Wiley. N.Y.; 1989.
6. Chang LS, Iezzoni AF, Flore JA. Yield components in 'Montmorency' and 'Meteor, sour cherry. J. Amer. Soc. Hort. Sci. 1987;112:247-251.
7. Gu S, Lombard PB. Effect of crop density on vegetative growth, flowering, fruit size, and yield in 'Nijisseiki' pear trees. Acta Hort. 1994;367:271-277.
8. Miranda C, Royo JB. A statistical model to estimate potential yields in pear cultivars 'Blanquilla' and 'Conference' before bloom. Journal of American Society for Horticulture Science. 2003a;128:452-457.
9. Lakso AN, Corelli Grappadelli L, Barnard J, Goffinet MC. An expolinear model of the growth pattern of the apple fruit. J. Hort. Sci. 1995,70:389-394.
10. Warrington IJ, Fulton TA, Halligan EA, de Silva HN. Apple fruit growth and maturity are affected by early season temperatures. J. Amer. Soc. Hort. Sci. 1999,124:468-477.
11. Berman ME, Dejong TM. Water stress and crop load effects on fruit fresh and dry weights in peach (*Prunus persica*). Tree Physiol. 1996;16:859-864.
12. Naor A, Klein I, Doron I, Gal Y, Ben-David Z, Bravdo B. The effect of irrigation and crop load on stem water potential and apple fruit size. J. Hort. Sci. 1997;72:765-771.
13. Barritt BH, Konishi BS, Diley MA. Tree size, yield and biennial bearing relationships with 40 rootstocks and three scion cultivars. Acta Hort. 1997;451:105-112.
14. Robinson TL, Lakso AN, Carpenter SG. Canopy development yield and fruit quality of Empire and Delicious apple trees grown in four orchard production systems for ten years. Journal of American Society for Horticulture Science. 1991;116:179-187.
15. Carusso T, Inglese P, Sotile F, Marra FP. Effect of planting system on productivity, dry- matter partitioning and carbohydrate content in above-ground components of 'Flordaprince' peach trees. J. Amer. Soc. Hort. Sci. 1999;124:39-45.
16. Dennis FG, Masabni JG, Ketchie DO. Evaluating twenty eight strains of 'Delicious' apple in Michigan. J. Amer. Soc. Hort. Sci. 1996;121:988-995.
17. Strong D, Azarenko AN. Relationship between trunk cross-sectional area, harvest index, total tree dry weight and yield components of 'Starkspur Supreme Delicious' apple trees. J. Amer. Pom. Soc. 2000;54:22-27.
18. Westwood MN, Robersts AN. The relationship between trunk cross-sectional area and weight of apple trees. J. Amer. Soc. Hort. Sci. 1970;95:28-30.
19. Robinson TL, Lakso. Bases of yield and production efficiency in apple orchard systems. J. Amer. Soc. Hort. Sci. 1991;116:188-194.
20. Miranda C, Royo JB. A statistical model to estimate potential yields in peach before bloom. Journal of American Society for Horticulture Science. 2003b;128:297-301.
21. Kappel F. Yied component analysis of 'horrow Delight', 'Kieffer', and 'harvest Queen' pear. J. Amer. Soc. Hort. Sci. 1990;115:25-29.
22. Westwood MN. Temperate zone pomology. San Francisco, W. H. Freeman and Company; 1978.
23. Silbereisen R, Scherr F. Vergleichende Untersuchungen ü Wuchs, Ertrag und Fruchtqualität ausländisher Apfelsorten. 1. Folge. Obst und Garten. 1968;87(6):217-222.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/26812