Scientia Horticulturae 130 (2011) 404–409

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Sensitivity in the estimation of parameters fitted by simple linear regression models in the ratio of blueberry buds to fruits in Chile using percentage counting

Sonia Salvo [a,*], Carlos Muñoz [b], Julio Ávila [b], Jaime Bustos [c], Emilio Cariaga [d], Carolina Silva [e], Gabriel Vivallo [e]

[a] Department of Mathematics and Statistics, Universidad de La Frontera, Chile
[b] Department of Electrical Engineering, Universidad de La Frontera, Chile
[c] Department of Systems Engineering, Universidad de La Frontera, Chile
[d] Department of Mathematical and Physical Sciences, Universidad Católica de Temuco, Chile
[e] School of Agronomy, Universidad Católica de Temuco, Chile

ARTICLE INFO

Article history:
Received 23 February 2011
Received in revised form 22 June 2011
Accepted 28 June 2011

Keywords:
Blueberry
Estimation of *fruits per bud*
SLRM

ABSTRACT

Blueberry exporting is an important activity in Chile, with fresh blueberries commanding the highest prices and being among the most exported products to the European and North American markets. To maintain quality in the centres of consumption, farmers must continuously improve the logistics of harvesting and shipping the blueberries. Thus every year they must calculate the production of the orchard well in advance in order to hire staff and ensure the logistic cold chain. For this calculation they use a count of *flower buds* and a simple linear model of which the slope parameter represents the number of *fruits per bud*. However, due to the cost of the counting procedure, some producers count only a fraction of each plant (25%, 50% or 100%), and furthermore they do not know what effect the variety and productive age of the plants may have on the estimation. The objective of this work is to measure the impact of the cultivated variety, the age of the plant in productive years, and the percentage of *fruits* counted in estimating the parameter *fruits per bud*. The study involved monitoring 310 plants of different varieties and ages distributed in northern, central and southern Chile (over an area of approximately 700 km × 200 km). The parameter was estimated by fitting simple linear regression models (SLRM) as a function of the number of *fruits* and *flower buds*. To evaluate the impact on the parameter, the SLRM was fitted considering the variables observed in all the plants, by percentage counted, by variety and by variety-age of the plant. The major findings indicate significant differences in the estimation of the parameter, suggesting that in order to estimate *fruits per bud* the whole plant must be counted and its age and variety taken into account.

1. Introduction

In Chile, fresh blueberries are among the products with the greatest prospects for growth in volume and market share. The volume exported in 2010 was around 44,000 tonnes. The markets where demand is highest are Europe and North America (ODEPA, 2010). The continued increase in the supply of fresh blueberries obliges producers to reduce costs and be more efficient throughout the production, harvest and shipping processes (Bañados, 2006, 2009).

Once harvested the fruit undergoes a decay process, meaning that it must be consumed rapidly. For this reason shipping and conservation conditions have been improving progressively; for example the use of biodegradable packaging can extend the shelf-life (Almenar et al., 2008). The amount of decayed fruit can be reduced to 10% by treatment with ultraviolet-C light (Perkins-Veazie et al., 2008); to 7% with the use of oxygen and carbon dioxide controlled atmospheres (Schotsmans et al., 2007) and to between 4% and 6% using only high concentrations of oxygen (Zheng et al., 2008). At the same time, the detection and classification of diseases in the fruit using an electronic nose can help to avoid the propagation of fungi (Li et al., 2010). These techniques present improvements in the post-harvest process of the fruit; however the factor which has the greatest effect on fruit quality is the delay between pre-packing and storage (Jackson et al., 1999). To minimise this delay, strategic decisions need to be taken with respect to the number of pickers hired for harvesting, the quantity of clamshells for packing, and the contracts to ensure the cold chain in shipping the product. This makes very important to have a good prediction of the production of the orchard, while it is not easy to obtain (Swain et al., 2010).

* Corresponding author. Tel.: +56 45 325330; fax: +56 45 592801.
E-mail address: ssalvo@ufro.cl (S. Salvo).

0304-4238/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.scienta.2011.06.042

The normal method used for predicting the production of fruit trees is the ratio between the flower buds, the quantity of set fruit, and production. For example, research was done in apricots on the development and drop of flower buds when affected by a quantity of units of cold, and the effects of irrigation and stem length; it was found that the characteristic with the greatest effect was the genotype of the variety cultivated (Alburquerque et al., 2003). Studies have also been done on the ratio between flower bud density, drop of flower buds and setting of fruit in nine varieties of apricot, from which it is concluded that the principal effects on the quantity of fruits are due to the variety cultivated, rather than environmental conditions (Alburquerque et al., 2004); Ruiz and Egea (2008) showed that apricot flower drop has a negative effect on fruit setting. For almonds the flower density, as well as the set fruit, the fruit density and the productivity are strongly influenced by the variety and age of the plant (Kodad and Socias, 2006) demonstrating that production is influenced by the number of flower buds and the variety within the species, despite losses due to flower drop, pollination and diseases.

For blueberries, Chilean producers use a ratio between *flower buds* and *fruits*, which has not been formally studied. Relating these two variables generates the parameter *fruits per flower bud* for predicting the quantity of *fruits* and estimating the production of the plant, and thus the harvest of the orchard. Counting *flower buds* in blueberries is possible, since the plants measure between 0.5 and 2 m in height and the number per plant varies between 50 and 1000 buds. However in order to estimate the parameter *fruits per flower bud* the quantity of *fruits* must be counted. This is when the farmers count the whole plant, or in some cases, to reduce the costs associated with this procedure, they count only a fraction of the plant: 50% or 25%. However counting a fraction of the plant may affect the estimation of the parameter *fruits per flower bud*, and the effects of this practice have not yet been studied. At the same time it is a common practice among farmers to treat the result as a fixed parameter, independent of the variety of the plants and the number of years they have been in production. The impact of these factors on the estimation of the parameter has also not yet been studied in the literature.

The purpose of this work is to verify the ratio between *flower buds* and *fruits*, and to quantify the effect on the estimation of the parameter *fruits per flower bud* considering the partial counting of *fruits*, the variety cultivated and the years in production of the plant.

2. Materials and methods

2.1. Plants selected

The data were collected during the season August 2008 to January 2009 in 13 conventional commercial blueberry orchards located between the Metropolitan and Araucanía Regions, including the northern, central and southern zones of Chile, and covering an area of approximately 700 km × 200 km. Within this area we typically find varieties of highbush blueberries which are specially suited to their climate conditions. Among them Star and O'Neal (in the north), Duke and Legacy (center), and Bluecrop, Brigitta and Elliot (in the south) (Godoy et al., 2008). Climatic area is relevant given that harvest time window differs in one month for northern and center zones, while it is two months for the northern and southern zones.

The plants selected were chosen to represent a proportional stratified random sample, in which the strata considered were the number of plants per orchard, proportional to the variety and production age of the plants with a confidence level of 95%, sampling error of 5% and maximum variance of 0.25. Based on these considerations a sample size of 310 plants was obtained. Table 1 shows the distribution of the number of plants selected by variety and age.

Table 1
Distribution of plants by variety and production age.

	Age					
	2	3	4	5	7	Total
O'Neal	5	4			4	**13**
Star	6					**6**
Duke	36	13	9	8	19	**85**
Legacy		10				**10**
Elliot	15	7	18		37	**77**
Bluecrop		15	30	5		**50**
Briguitta	13	22	25	9		**69**
Total	**75**	**71**	**82**	**22**	**60**	**310**

The plants selected were inspected twice during the phenological states of budding and fruiting.

2.2. Buds campaign

Bud counting was done between August and October 2008, after pruning of the plants. The *flower buds* and *flowers* present in the whole plant were counted (exhaustive count) using a counting machine.

2.3. Fruits campaign

The fruits were counted between September 2008 and January 2009. During this period the number of *fruits* was also recorded using a counting machine. To quantify the effect of partial (non exhaustive) counting on the estimation of *fruits per flower bud* plants were counted randomly to 25%. Counts of 50% and 100% were done only in the case of small plants.

2.4. Analysis of parameter sensitivity

For the phenological stages of interest, the variables observed were grouped by *buds* (Eq. (1)) and *total fruits* (2) depending on the percentage counted. To calculate the total number of buds (1), a parameter of 8 flowers per bud was used, which is commonly accepted and used by farmers as a factor accounting for the overall production which reaches the packing process.

$$buds = \frac{flower\ \ buds + flowers}{8} \tag{1}$$

$$total\ \ fruits = k * fruits$$
$$k = \begin{cases} 1 : 100\%\, plant\ \ counted \\ 2 : 50\%\, plant\ \ counted \\ 4 : 25\%\, plant\ \ counted \end{cases} \tag{2}$$

In order to estimate the effect of the percentage counted on the estimation of the parameter *fruits per bud*, a simple linear regression model (SLRM) without intercept was fitted. The reason for not using an intercept is that if there are no *buds* on the plant it cannot bear *fruits* (3). To evaluate whether the model describes a linear ratio between the study variables it was considered that the Pearson linear correlation coefficient, *r*, must be close to 1. The SLRM assumes that there are no errors in the *bud* count.

$$total\ \ fruits = \beta * buds + \varepsilon \tag{3}$$

where

β: parameter to be estimated (represents the parameter of *fruits per flower bud*).

$\varepsilon \sim N(0,\sigma^2)$: random error to account for the variability of the *fruits*, which cannot be explained by the linear ratio between *buds* and *fruits*.

Parameter β is estimated by minimising the sum of the squared error (4).

$$\min \sum_{i=1}^{N} (total\ \ fruits_i - buds_i * \beta)^2 \quad (4)$$

The variability of the errors, which is the variance of the model, is determined by the mean squared error (*MSE*). The *MSE* is calculated according to (5).

$$MSE = \frac{1}{N-1} \sum_{i=1}^{N} (total\ \ fruits_i - buds_i * \hat{\beta})^2 \quad (5)$$

The SLRM is fitted for four cases:

i) *Fit for all observations*
Initially the SLRM was fitted considering all the observations to detect counting errors, extreme data and an estimation of the parameter β.
ii) *Fit for percentage counted*
The SLRM was fitted for each percentage counted (25%, 50% and 100%), to quantify the effect of partial counting on the estimation of parameters, and then eliminating the atypical values on the more distant residues ±1SD, ±2SD and ±3SD (SD: standard deviation).
iii) *Fit for variety*
The SLRM was fitted independently for each of the following varieties: O'Neal, Duke, Legacy, Elliot, Bluecrop and Briguitta. The Star variety was excluded from this analysis because there were very few plants in the sample (see Table 1).
iv) *Fit for variety-age of the plants*
The SLRM was fitted for variety-age without atypical values, to quantify the effects of these factors on the estimation of parameters. Variety-age categories with less than 10 observations were discarded from the analysis.

3. Results

3.1. Fit for all observations

Fig. 1a presents a dispersion diagram of the variables *buds* and *total fruits*, with a band of ±1.96SD marked. It should be noticed that the data present a funnel effect, in that the plants with a smaller number of buds present a better fit than plants with a larger number of buds. This may be due to the commission of significant counting errors in large plants, or else to the development of each individual plant. If the former is the case, it would suggest that it is very important to be rigorous in the counting procedures and, if possible, to distinguish between different counting methods as a function of the variability which they generate. Fig. 1a shows data which are distant from the trend, indicated by arrows. If data outside ±1.96SD are considered to be atypical, it may be observed that many data comply with that condition. This is a further indicator showing that counting a fraction of the plant is inadequate.

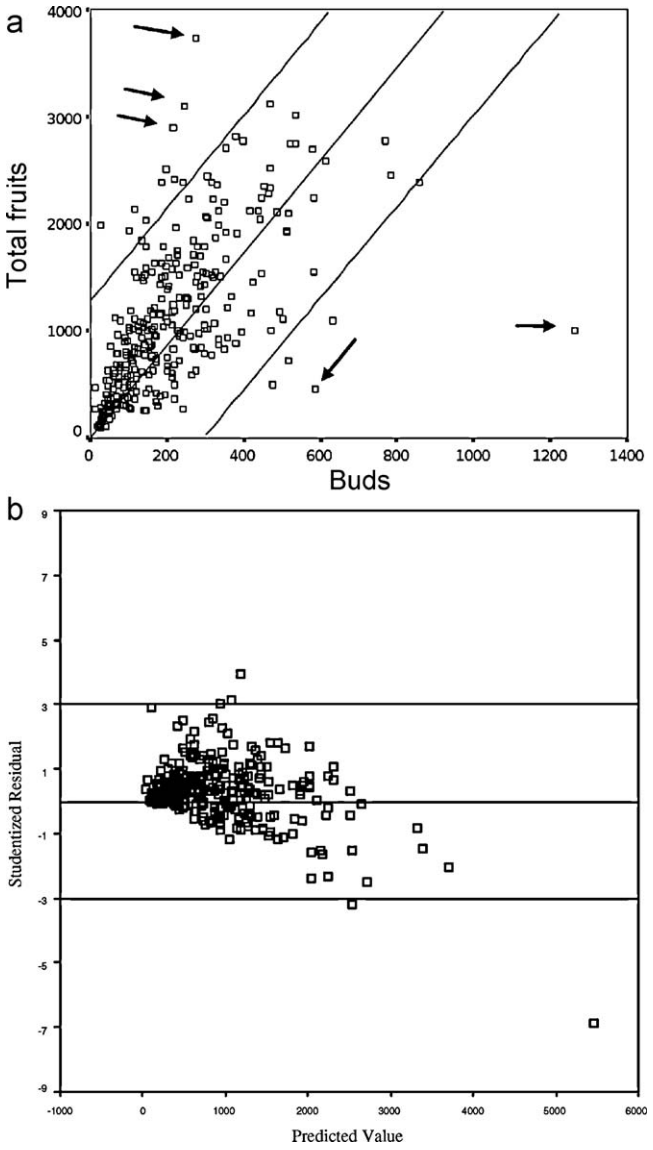

Fig. 1. Dispersion diagram (a) and Studentized Residual diagram (b) for all observations.

In Fig. 1b, the dispersion diagram of the Studentized residues shows the typical funnel shape, indicating that there is a problem in the variance of the error in the SLRM. As the value predicted by the model increases, so does the variance. This reaffirms the fact noted above, as the number of *buds* increases, so does the variability of the count.

The results of the fit of the SLRM indicate a $\hat{\beta}$ of 4.3 (statistically significant), a *MSE* of 421749.4 and an *r* of 0.87. The $\hat{\beta}$ representing

Table 2
Results of parameters estimated by the SLRM for each percentage counted and with elimination of atypical data.

	N			*r*			$\hat{\beta}$			*MSE*		
	25%	50%	100%	25%	50%	100%	25%	50%	100%	25%	50%	100%
All	228	65	27	0.91	0.80	0.97	5.1	2.0	2.5	369758.5	163557.5	13320.7
±3SD	223	64	27	0.93	0.87		5.2	2.6		283019.1	110395.9	
±2SD	214	63	16	0.95	0.87	0.97	5.2	2.6	2.4	20158.3	103594.7	9171.6
±1SD	174	47	12	0.97	0.92	0.97	5.3	2.4	2.5	86270.1	37904.2	5691.7

Note: in the case of ±3SD with 100% count no data were eliminated (*N* = 27 in both cases), so that the results are identical to "All" at 100%.

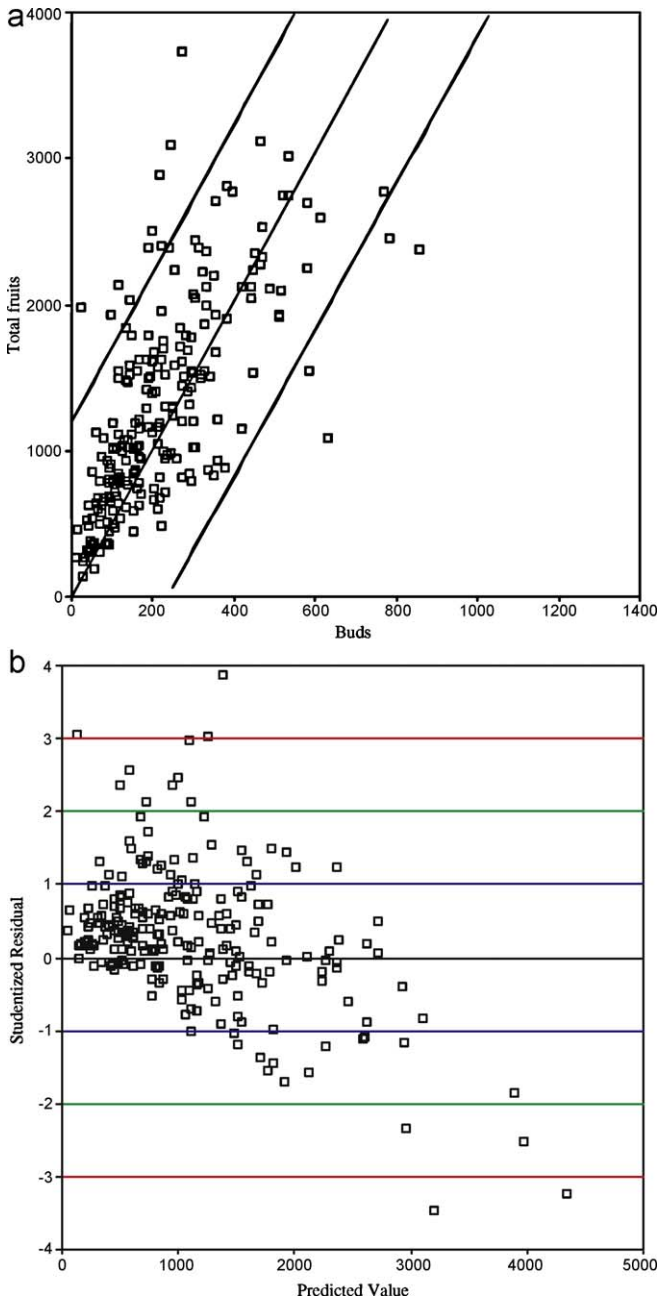

Fig. 2. (a) Dispersion and (b) Studentized residues graphs with percentage counting at 25%.

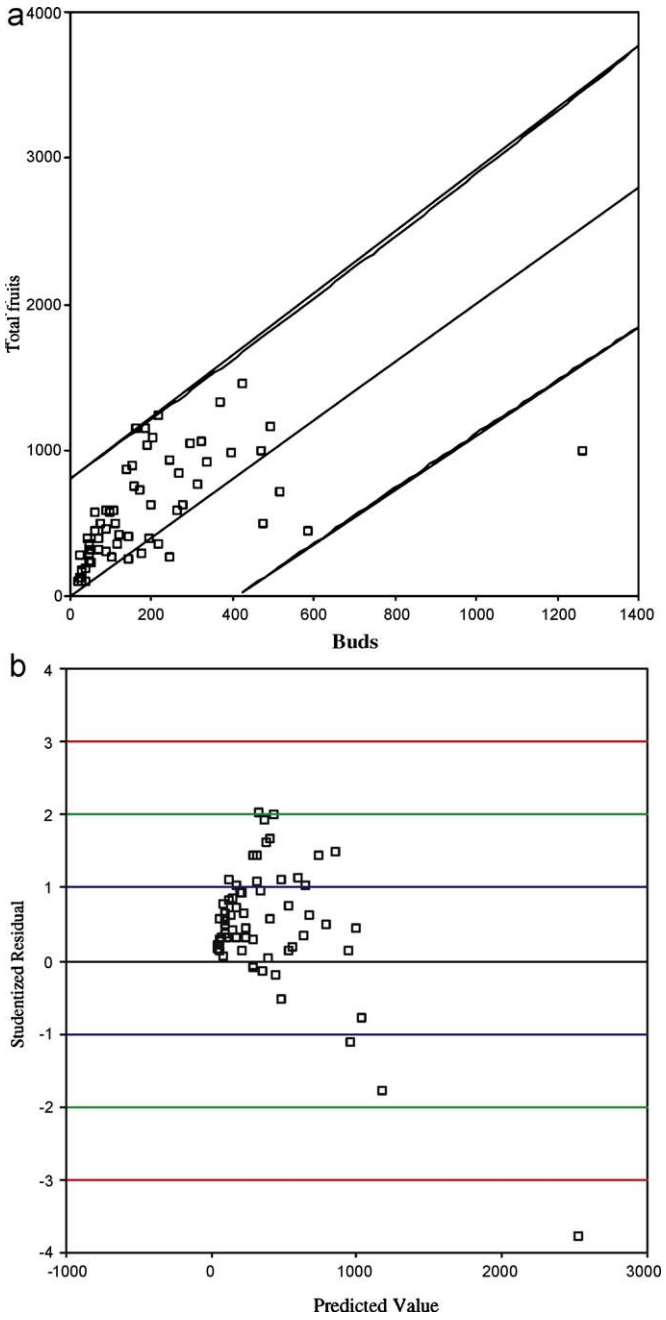

Fig. 3. (a) Dispersion and (b) Studentized residues graphs with percentage counting at 50%.

the *fruits per bud* was found to be very different from that used by the farmers (equal to 8). These results will be used for the comparison of fits of the models for percentage counted.

3.2. Fit for percentage counted

Figs. 2–4a and b present the dispersion diagrams of the variables *buds* and *total fruits* with a confidence band of ±1.96SD for the estimated model, and the corresponding Studentized residues graph with bands from ±1SD to ±3SD considering different counting percentages (as indicated in Section 2.3). In the counts at 25% and 50%, as the quantity of *buds* increases, so does the dispersion of the *fruits*. This occurs to a lesser extent when 100% of the plant is counted. Distant points which influence the slope of the estimated straight line are present when 25% and 50% are counted, indicating possible counting errors. The residues form a funnel in the case of the partial count and the predicted values are distant by more than 3SD when percentages of the plant are counted. On the other hand, when the whole plant is counted the majority of these values do not exceed 2SD.

Table 2 compares the parameters estimated from the models fitted for the percentage counted and with elimination of atypical data. Independent of the percentage counted, a linear ratio is maintained between the *buds* and the *total fruits*; all the *r* are statistically significant ($P < 0.01$) and positive. A significant change is observed in the value of $\hat{\beta}$ when the percentage counted is 25% and 50%, corresponding to smaller plants. The same does not occur in the behaviour of the *MSE*, which diminishes strongly as the percentage counted increases and when atypical data are eliminated.

When these fitted results are compared with those obtained in the first case, the *r* increased significantly, indicating that there is

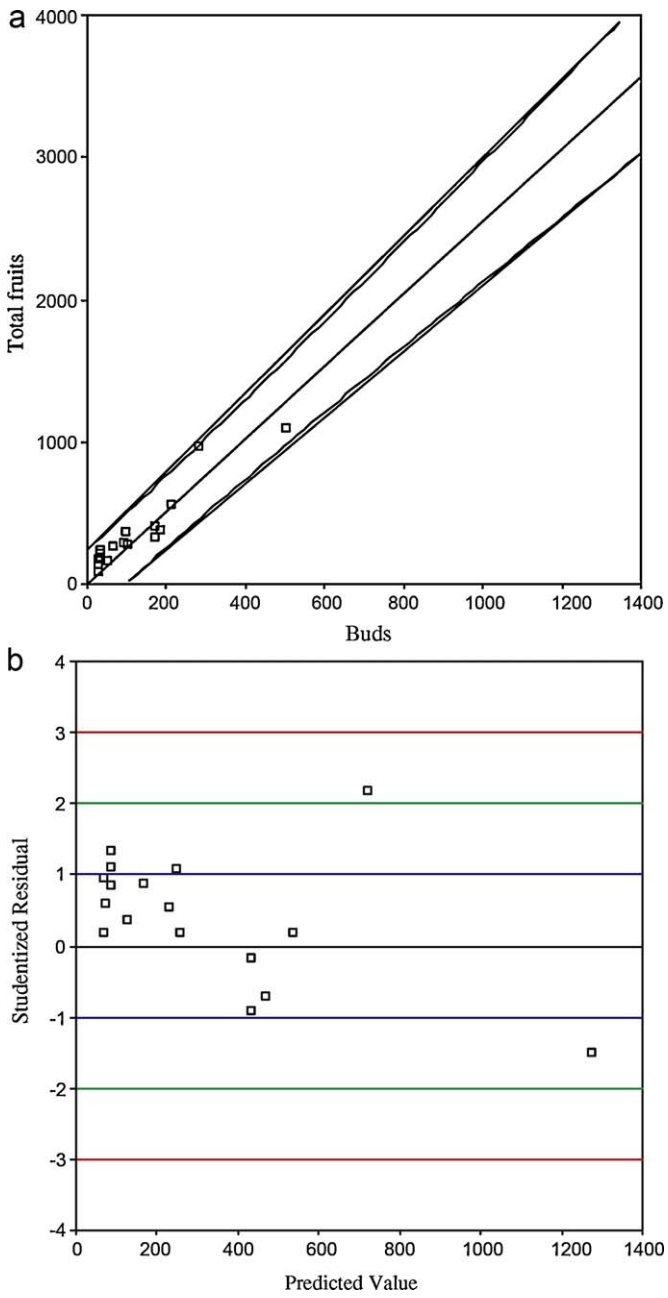

Fig. 4. (a) Dispersion and (b) Studentized residues graphs with percentage counting at 100%.

a better linear ratio between the variables *Buds* and *Total fruits*. This implies that the estimated parameter value is more reliable, except when the percentage counted was 50% and atypical data between 2SD and 3SD were considered, when it remained the same or diminished. The value of $\hat{\beta}$ increased when the percentage counted was 25%. In all other cases it diminished significantly. The MSE diminished significantly in all cases but continued high, which is an indicator that the fit values of the parameter are not necessarily right, despite the good linear ratio between the study variables. These findings suggest the importance of counting 100% of the plant and the relevance of the counting procedure.

3.3. Fit for variety

Table 3 shows a summary of the fit of the SLRM for *variety*. All the values of *r* are positive and statistically significant, and greater than 0.87, indicating that it is important to consider the variety of blueberry in order to obtain the number of *fruits per bud*. The values of $\hat{\beta}$ increased with respect to the value of 4.3, except for the variety Elliot; they were differentiated by variety, a measure of the phenotype expression of blueberries. This behaviour might suggest that the factor *variety* influences the ratio between *total buds* and *fruits*. In general, the values obtained are very different from those traditionally used by farmers.

3.4. Fit for variety-age

In the analysis by *variety-age* atypical values and categories with less than 10 observations were discarded. This means that the number of observations in *variety-age* does not correspond to the total of observations in *variety*. This excludes from the study the varieties for the northern zone as they are new orchards with new varieties being introduced in Chile.

Table 3 shows a summary of the fit of the SLRM for *variety-age*. All the values of *r* are positive and statistically significant, in some cases with a small increase or decrease in the value when the *variety-age* factor is compared to the variety; there is a large difference when these factors are not considered, in this case significantly less. This behaviour might suggest that the factor *variety and age* influences the ratio between *total buds* and *fruits*. A change is observed in the values of $\hat{\beta}$, in particular significant increases in the varieties Briguitta at 4 years, Duke at 2 years and Elliot at 4 years. The *MSE* also undergoes significant modifications, being either increased or diminished in the majority of cases. Thus the number of *fruits per bud* depends on the variety and age of the plant, signifying differences in the quantity of fruit produced by the plant.

Table 3
Results of fit of the SLRM by *variety* and *variety-age*.

Variety	Age	*N*		*r*		$\hat{\beta}$		*MSE*	
		Variety	*Variety-age*	*Variety*	*Variety-age*	*Variety*	*Variety-age*	*Variety*	*Variety-age*
Bluecrop	3	42	14	0.94	0.93	5.5	5.1	90250.7	29078.2
	4		24		0.94		5.4		126253.1
Briguitta	3	54	19	0.96	0.95	4.9	4.8	109866.3	149344.9
	4		22		0.98		5.2		62704.7
Duke	2	62	28	0.99	0.96	5.5	6.5	43102.6	37439.6
	3		11		0.97		5.1		19276.6
	7		11		0.99		5.5		45094.1
Elliot	4	52	18	0.92	0.94	3.8	4.9	174408.3	45557.1
	7		19		0.95		3.6		199612.2
Legacy	3	10	10	0.96	0.96	7.0	7.0	108920.1	108920.1

4. Discussion and conclusions

In this work it was found that the relationship between buds and total fruits is strong, as was established by Alburquerque et al. (2003, 2004) for apricots. Moreover, the results found by Kodad and Socias (2006) for almonds were verified for blueberries, in the sense that the genetic factor is decisive in the ratio between buds and total fruits, and that different values are obtained for fruits per flower bud if the age factor is considered.

When small plants – which have few buds – are counted, the counting error is less, while in large plants there are more important counting errors. This may be due to errors committed by the people who carry out this procedure, and the error is clearly apparent in the funnel effect observed in Figs. 2–4 in the Studentized residues graphs. This suggests the need to incorporate count checking mechanisms to reduce the errors committed by people.

At the same time, the percentage counted is also decisive in estimating the parameter *fruits per bud*. The estimation of this parameter diminishes significantly as the percentage counted increases and the *MSE* diminishes significantly when 100% of the plant is counted, indicating the high variability of the data. These findings suggest the importance of counting the whole plant. This, while appearing obvious, is not a common practice for farmers as they usually count only a fraction of a plant.

The results with the sample used indicate that in generating a production model for blueberries based on flower buds, the variety and age of the plant must be considered, with a 100% count. This suggests that farmers may be committing very significant errors by using a single factor of 8 fruits per bud. Indeed, in this study it was found that if a single factor of *fruits per bud* is sought for all the plants in the sample studied, which included the varieties Bluecrop, Briguitta, Duke, Elliot and Legacy, this factor is 4.3 fruits per flower bud. If the analysis is done by variety, values are obtained which range from 3.8 to 7 fruits per flower bud, and when the analysis considers the variety-age factor, the values obtained range from 3.6 to 7 fruits per flower bud. As it can be seen from the data, there is a strong interaction between varieties and age, however it seems convenient to develop a more conclusive study in a longer timeframe (and thus validating if this behaviour holds in consecutive seasons). All these results are very different from the 8 fruits per flower bud used by blueberry farmers, indicating that if orchard models are used which consider particular varieties and planting ages, better predictions can be generated than those currently made by farmers.

Acknowledgements

The authors wish to thank the producers associated with project INNOVA 07CN13PAT-213 “Formulation of a Predictive Model for Blueberry Production”.

References

Alburquerque, N., Burgos, L., Egea, J., 2003. Apricot flower bud development and abscission related to chilling, irrigation and type of shoots. Sci. Hortic. 98 (3), 265–276.
Alburquerque, N., Burgos, L., Egea, J., 2004. Influence of flower bud density, flower bud drop and fruit set on apricot productivity. Sci. Hortic. 102 (4), 397–406.
Almenar, E., Samsudin, H., Auras, R., Harte, B., Rubino, M., 2008. Postharvest shelflife extension of blueberries using a biodegradable package. Food Chem. 110 (1), 120–127.
Bañados, M., 2006. Blueberry production in South America. Acta Hort. 715, 165–172.
Bañados, M., 2009. Expanding blueberry production into non-traditional production areas: northern Chile and Argentina, Mexico and Spain. Acta Hort. 810, 439–444.
Godoy, C., Monterubbianesi, G., Tognetti, J., 2008. Analysis of highbush blueberry (*Vaccinium corymbosum* L.) fruit growth with exponential mixed models. Sci. Hortic. 115 (4), 368–376.
Jackson, E.D., Sanford, K.A., Lawrence, R.A., McRae, K.B., Stark, R., 1999. Lowbush blueberry quality changes in response to prepacking delays and holding temperatures. Postharvest Biol. Technol. 15 (2), 117–126.
Kodad, O., Socias, R., 2006. Influence of genotype, year and type of fruiting branches on the productive behaviour of almond. Sci. Hortic. 109 (3), 297–302.
Li, C., Krewer, G.W., Ji, P., Scherm, H., Kays, S.J., 2010. Gas sensor array for blueberry fruit disease detection and classification. Postharvest Biol. Technol. 55 (3), 144–149.
ODEPA, 2010. Rep. tec., boletín número 10.
Perkins-Veazie, P., Collins, J.K., Howard, L., 2008. Blueberry fruit response to postharvest application of ultraviolet radiation. Postharvest Biol. Technol. 47 (3), 280–285.
Ruiz, D., Egea, J., 2008. Analysis of the variability and correlations of floral biology factors affecting fruit set in apricot in a Mediterranean climate. Sci. Hortic. 115 (2), 154–163.
Schotsmans, W., Molan, A., MacKay, B., 2007. Controlled atmosphere storage of rabbit-eye blueberries enhances postharvest quality aspects. Postharvest Biol. Technol. 44 (3), 277–285.
Swain, K., Zaman, Q., Schumann, A., Percival, D., Bochtis, D., 2010. Computer vision system for wild blueberry fruit yield mapping. Biosystems Eng. 106 (4), 389–394.
Zheng, Y., Yang, Z., Chen, X., 2008. Effect of high oxygen atmospheres on fruit decay and quality in Chinese bayberries, strawberries and blueberries. Food Control 19 (5), 470–474.