Segmented regression models to estimate the optimal size of the experimental plot with sugar cane

Luz Elena Barrantes Aguilar  
Adrián González Estrada§  
Miguel Ángel Martínez Damián  
Ramón Valdivia Alcalá

División de Ciencias Económico Administrativas-Universidad Autónoma Chapingo. Carretera México-Texcoco km 38.5, Chapingo, Estado de México. CP. 56230. (luzbtes@gmail.com; ngel01@colpos.mx; ramvaldi@gmail.com).

Corresponding author: adrglez@prodigy.net.mx.

Abstract

The objective of this research was to estimate the optimal size of the experimental plot with sugar cane in the Brunca Region of Costa Rica. Segmented regression models were used with the data obtained from a uniformity test (40 rows of 84 meters long with a separation between each of 1.5 meters, for a total of 5 040 m$^2$ of experimental area), the work of the field was conducted at the El Porvenir farm in Perez Zeledon Costa Rica, between 2018 and 2019. The coefficients of the linear regression models with constant (LRP) and quadratic regression with constant (QRP) were statistically significant. It was concluded that the optimal plot size that minimizes the experimental error for the trials established in the region, should be in the range of 72 to 93 m$^2$.

Keywords: efficiency of agricultural research, minimization of experimental error, soil heterogeneity.

Reception date: May 2020  
Acceptance date: July 2020
Introduction

In Costa Rica there are seven cane-producing areas and thirteen sugar mills. According to the Annual Operational Plan (PAO) of DIECA (Directorate of Research and Extension of Sugar Cane) in 2018, approximately 22% of the current trials were in the Brunca Region. Among the most important research topics studied in that region are: variety evaluation, nutrient interaction and nutritional dosage, ripening agents, herbicides, and planting density.

The target variables are frequently field yield and agro-industrial yield. For this type of research in the region, three main plot sizes have been used: 5 rows of 10 m with a separation of 1.5 m (75 m$^2$), 6 rows of 9 meters with a separation of 1.5 m (81 m$^2$) or 5 rows of 8 m spaced 1.5 m (60 m$^2$), but the efficiency of these plot sizes has never been validated.

The size of the experimental plot with which the experimental error is minimized has not been investigated either. For this reason, the objective of this work is to find the optimal size of the experimental plot with sugar cane in the Brunca region of Costa Rica. Sripathi et al. (2017) group, within what they call experimental design factors, the size of the plot, the size of the block and the number of repetitions and affirm that in the yield trials the experimental error is sensitive to these factors, due to the fact that the research agronomy depends on the data recorded in the field trials.

Many factors influence the definition of the size of the experimental plot, the most important of which are: experimental error and associated costs. The size of the plot must allow satisfactory capture of all the heterogeneity of the soil, considered the main source of variability of the experimental plot. Cocco et al. (2009) compared the results obtained with different sizes and shapes of experimental plot in the strawberry crop planted in soil and under the hydroponic system. The cultivation in the soil presented greater experimental variability than the hydroponic cultivation.

The development of methods for the estimation of the optimal plot size has an important basis with the pioneering work of Smith (1938), who found a negative asymptotic relationship between the variance and the plot size. There are many proposals made since then. Current research was largely developed in Brazil and applied to various crops, for example: crotalaria (Crotalaria juncea L.) (Facco et al., 2017), eggplant (Solanum melongena) (Kryszczun et al., 2018), cherry tomato (Solanum lycopersicum L.) (Giacomini and Lucio, 2018), tomato (Solanum lycopersicum), string beans (Phaseolus vulgaris) and zucchini (Cucurbita pepo) (Schwertner et al., 2015), sunflower (Helianthus annus L.) (Santos et al., 2015), coffee (Coffea) (Mendes et al., 2016), taro (Colocasia esculenta) (Silva, 2014) and sweet potato (Ipomoea batatas) (Rodríguez et al., 2018).

Some recent work has also been done in India, with the cultivation of Indian mustard (Brassica juncea L.) (Khan and Tanwar, 2017). In the United States of America, with watermelon (Citrullus lanatus) (Boyhan, 2013). In Costa Rica, this type of research has been carried out for the cultivation of coffee in the early sixties (Páez, 1962) and for beans in the seventies (Mamani, 1971).
The most recent trials were carried out in the Bagaces area, Guanacaste, for rice (*Oryza sativa*) (Vargas and Navarro, 2014) and corn (*Zea mays*) (Vargas and Navarro, 2017) crops. For the cultivation of sugar cane in other countries, work has already been done on calculating the optimal size of the experimental plot. For example: Igue *et al.* (1991) in Brazil; Bose and Khanna (1939) in India and Palencia (1965) and Álvarez (1982) in Guatemala.

**Materials and methods**

**Segmented regression models**

Among the methodological proposals for estimating the optimal size of the experimental plot, are the segmented regression models. Paranaiba *et al.* (2009) propose to fit a linear regression model with constant (LRP) for the relationship between the coefficient of variation (CV) of the yield and the plot size (measured in basic experimental units (UEB)), as seen in the Figure 1, where for the values \( x_i \leq x_0 \) this relationship is described by a decreasing linear model to a certain point, after which it becomes constant. The CVP value that corresponds to the inflection point is obtained from an iterative process.

![Figure 1. Relationship between plot size and coefficient of variation for the linear segmentation method with constant.](image)

This segmented linear regression model can be expressed as: \( CV_x = \begin{cases} \beta_0 + \beta_1 x_i + \varepsilon_i, & \text{if } x \leq x_0; \\ CVP + \varepsilon_i, & \text{if } x > x_0, \end{cases} \)

where: \( CV_x \) is the coefficient of variation between the totals for plots with \( x_i \) basic units; \( CVP \) is the coefficient of variation at the point where the two segments meet; \( x \) is the parcel size measured in basic units; and \( \varepsilon_i \) is the error associated with \( CV_x \), supposedly normal and independently distributed, with zero mean and constant variance.

Since the two segments of (1) are equal at point \( x_0 \), then: \( \beta_0 + \beta_1 x_0 = CVP \) 2). Therefore, the optimal size of the experimental plot is given by: \( x_0 = \frac{CVP - \beta_0}{\beta_1} \) 3).
Another alternative within the possibilities of segmented models is the quadratic regression method with constant (QRP), described by Ferreira (2007) and applied by Mendes et al. (2016) for the estimation of the size of the experimental plot. With respect to the previous model, this method assumes a second degree polynomial form, instead of a linear form in the first segment (Figure 2).

![Figure 2](image_url)

**Figure 2. Relationship between the size of the plot and the coefficient of variation for the quadratic segmentation method with constant.**

For values $x_i \leq x_0$ the model is quadratic and for values $x_i > x_0$ it is constant. Similarly, the intercept between the segments determines the optimal size: $CV_x = \begin{cases} \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, & \text{if } x \leq x_0; \\ CVP + \epsilon_i, & \text{if } x > x_0, \end{cases}$  

(4). The point $x_0$ represents the junction point of the two segments and must be estimated together with the other parameters of the model. Since the curve must be continuous and smooth, the first derivatives with respect to $x$ in both segments must be the same for the point $x_0$. According to Ferreira (2007), this condition implies that: $\frac{\partial CV_x}{\partial x} = \beta_1 + 2\beta_2 x_i$  

(5). Once equalized to zero, equation (5) is solved for $x$ and after substituting $x$ for $x_0$, we obtain: $x_0 = \frac{-\beta_1}{2\beta_2}$  

(6). The value of the constant (CVP) can be obtained by substituting this value in equation (4): $CV_x = CVP = \beta_0 + \beta_1 x_0 + \beta_2 x_0^2 = \beta_0 + \beta_1 \left(\frac{-\beta_1}{2\beta_2}\right) + \beta_2 \left(\frac{-\beta_1}{2\beta_2}\right)^2 = \beta_0 - \frac{\beta_1^2}{4\beta_2}$  

(7).

**Information source and uniformity test**

The data from a sugarcane uniformity test carried out between 2018 and 2019 at the El Porvenir farm in Pérez Zeledón, Costa Rica, located at an altitude of 591 meters above sea level, were used. In 2017 an annual precipitation of 3673.8 mm was recorded and the average temperature was 23.3 °C, with a maximum of 34.5 °C and a minimum of 15.4 °C.

40 rows of 84 m long with a separation between rows of 1.5 m of the variety RB 99-381 were planted manually at a density of three jets. In total, there were 4 800 m$^2$ of useful plot, which was divided into 1 600 basic experimental units (UEB), each one 2 m long by 1.5 meters wide (3 m$^2$).
The cultivation practices that were carried out were the same as those applied to commercial plantations in the region. One month before sowing the ground was prepared. Sowing was done manually by placing three canes with a 15 cm overlap and 13 t ha$^{-1}$ of seed are used. For weed control, a chemical mixture of Pendimethalin 50 EC (3 L ha$^{-1}$) with Terbutylazine 50 SC (2 L ha$^{-1}$) was used in pre-emergence of the weed (bare soil). And in early post-emergence, a chemical mixture of Hexazinone 75 WG (0.5 kg ha$^{-1}$) with Diuron 80 WG (2 kg ha$^{-1}$).

As it was first cut cane, 140 kg of nitrogen, 140 kg of phosphorus, 167 kg of potassium, 35 kg of magnesium and 40 kg of sulfur per hectare were applied, distributed in three fertilizations. For the harvest, characteristics of topography, size and distribution of the sugarcane farms in the region are considered, according to which it is carried out manually or semi mechanized. The harvest was carried out manually on March 6 and 7, 2019 at 10 months of age from the plantation. The variable that was measured was field performance, taking the weight of each UEB measured in kilograms.

The basic experimental units were grouped into secondary units, of different shapes and sizes, taking as a requirement that these groupings of adjacent plots always use the entire experimental area, as described by Paranaiba et al. (2009), each of these secondary units was calculated the mean of the production, the variance and the coefficient of variation.

For the estimation of the optimal plot size, the methods of linear regression with constant (LRP) and quadratic regression with constant (QRP) were used. In the estimation of the parameters and the analysis of the information, the statistical package SAS version 9.3 and Python version 3 (2008) were used.

**Results and discussion**

At the time of harvest and weighing of the test, of the 1,600 basic experimental units an average production of 19.9 kg was obtained per basic experimental unit, with a standard deviation of 3.8 kg. The values ranged between 9.5 kg and 38 kg. As observed in Figure 3, 41% of the basic experimental units weighed between 18 and 22 kg.

![Figure 3. Histogram of the production obtained from the uniformity test.](image-url)
In total, the trial data were grouped into 20 sizes corresponding to 63 different shapes (Table 1).

Table 1. Secondary plot sizes and quantities into which trial data could be pooled.

<table>
<thead>
<tr>
<th>Number of secondary plots</th>
<th>Size</th>
<th>Shapes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UEB m²</td>
<td></td>
</tr>
<tr>
<td>1 600</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>800</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>400</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>320</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>200</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>160</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>100</td>
<td>16</td>
<td>48</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>64</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>50</td>
<td>32</td>
<td>96</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td>32</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>25</td>
<td>64</td>
<td>192</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>240</td>
</tr>
<tr>
<td>16</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>10</td>
<td>160</td>
<td>480</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
<td>600</td>
</tr>
<tr>
<td>5</td>
<td>320</td>
<td>960</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>1 200</td>
</tr>
<tr>
<td>2</td>
<td>800</td>
<td>2 400</td>
</tr>
</tbody>
</table>

The relationship between the coefficient of variation of each of these shapes and the size of the experimental plot can be seen in Figure 4 (for a better appreciation the abscissa axis was cut at 400), as expected the coefficient of variation decreases rapidly in the segment of small parcels and then, as the size increases, the coefficient of variation tends to decrease less than proportionally.
Figure 4. Relationship between the coefficient of variation (CV) of production and the size of the plot measured in UEB.

By applying the segmented linear regression model (LRP), the model parameters were obtained:

\[ CV_x = \begin{cases} 
12.61 - 0.43x & \text{if } x \leq 24.05; \\
2.17 & \text{if } x > 24.05, 
\end{cases} \]

The coefficient of determination was 61.62% and the \( F \) statistic of the analysis of variance for the significance test of the estimated parameters was also significant (\( p < 0.01 \)) (Table 2).

Table 2. Results of the linear regression model with constant (LRP) and of the quadratic regression model with constant (QRP) for the sugarcane trial.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LRP(^{(1)})</th>
<th>QRP(^{(1)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta}_0 )</td>
<td>12.6093 **</td>
<td>14.1172 **</td>
</tr>
<tr>
<td></td>
<td>(0.6728)</td>
<td>(0.8055)</td>
</tr>
<tr>
<td>( \hat{\beta}_1 )</td>
<td>-0.4338 **</td>
<td>-0.7654 **</td>
</tr>
<tr>
<td></td>
<td>(0.0537)</td>
<td>(0.1149)</td>
</tr>
<tr>
<td>( \hat{\beta}_2 )</td>
<td>0.0123 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td></td>
</tr>
<tr>
<td>( n )</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>( F )</td>
<td>114.3413</td>
<td>136.2508</td>
</tr>
<tr>
<td>( R^2_{\text{adjusted}} )</td>
<td>61.62</td>
<td>76.96</td>
</tr>
<tr>
<td>( W )(^{(2)})</td>
<td>0.9276</td>
<td>0.9309</td>
</tr>
<tr>
<td>( P )</td>
<td>2.17</td>
<td>2.2256</td>
</tr>
<tr>
<td>( x_0 )</td>
<td>24.0531</td>
<td>31.0742</td>
</tr>
</tbody>
</table>

\(^{(1)}\) = standard error in parentheses; ** = significant at 1%; * = significant at 5%; \(^{(2)}\) = shapiro-Wilk test statistics.

The estimated optimal size is approximately 25.05 UEB. Considering that each UEB is 3 m\(^2\), the optimal size is 72.16 m\(^2\) (Figure 5).
Based on the data to fit the quadratic regression model with constant, the estimators were obtained: 

\[
CV_x = \begin{cases} 
14.12 - 0.77x_i + 0.01x_i^2 & \text{if } x \leq 31.07; \\
2.23 & \text{if } x > 31.07, 
\end{cases}
\]  

In this case, the optimal size was 31.07 UEB, equivalent to 93.22 m\(^2\) plots (Figure 6).

This model has a coefficient of determination of 76.96% and an F statistic that is also significant \((p < 0.01)\) (Table 2). Plot size tends to be higher with the QRP method than with the LRP, due to the difference in fit. Peixoto et al. (2011) in trials with passion fruit attribute this tenure to the curvature of the model. Mendes et al. (2016) found a plot size for coffee of 10.53 UEB larger with the QRP method compared to the LRP method.
On the other hand, Silva et al. (2012) using a radish test, found a difference of 2.57 UEB in favor of the QRP method. In both cases, the Shapiro-Wilk normality test (Shapiro and Wilk, 1965) was applied to the errors of the model ($\varepsilon_i$), both for the LRP and for the QRP model, and the null hypothesis of normality was not rejected with ($p= 0.1998$) and ($p= 0.2181$), respectively. Using the segmented regression methods, plot sizes between 72 and 93 m$^2$ were found.

Álvarez (1982) recommended that the size of the sugarcane research plots range between 80 and 115 m$^2$ (5 or 6 rows of 8 or 9.6 m wide), with a length of 10 to 12 m and plots of 4 furrows of 12 m long (76.8 m$^2$) in cases where there is any technical or practical limitation (inputs, seed, land, etc.). This author applied the multiple regression analysis and the bivariate maximum curvature method, without finding large differences between the two methods and in both cases, with determination coefficients greater than 80%.

Although the results obtained here are similar to those found by Álvarez (1982), the comparison must be taken with reservations, because it is an investigation that was carried out in Guatemala and as the author indicates in the title of his work, its result is subject to the conditions of the Bulbuxya farm, which is located in the department of Suchitepequez in the south-western region of the Republic of Guatemala and the inference, beyond said zone, is not valid. Another important aspect is the validity of this type of investigation.

Also, regular updates are necessary. That work is from the early eighties and there are no records of any update. In Guatemala, Palencia (1965), used the method of Smith (1938) and calculated an optimal size of 27.24 m$^2$, with which he rounded and recommended a useful plot of 28 m$^2$ (two rows of 8 m long with a separation of 1.8 to 2 m), much smaller than in the previous case. The total plot would be 4 rows of 10 m long to exclude the two lateral rows and one meter at each end of the rows.

**Conclusions**

The optimal size of the experimental plot calculated with the linear regression method with constant was 72 m$^2$ and with the quadratic regression method with constant 93 m$^2$. If the variation of the other factors that were not included in this estimate is considered, then, in order to minimize the experimental error, it is recommended for yield trials that the size of the experimental plot with sugarcane should be between 72 and 93 m$^2$ and adjust the shape of the plot according to space availability.

It is recommended to replicate the methodology developed here in the other sugarcane regions.

**Acknowledgments**

To the Department of Research and Extension of Sugar Cane (DIECA) of the Industrial Agricultural League of Sugar Cane (LAICA) and CoopeAgri, RL for the resources and knowledge provided for the development of this research.
Cited literature


