Methodology to estimate the soil heterogeneity coefficient, the number of repetitions and the optimum plot size for common bean (*Phaseolus vulgaris* L.) research

Metodología para estimar el coeficiente de heterogeneidad del suelo, el número de repeticiones y el tamaño de parcela en investigaciones con frijol (*Phaseolus vulgaris* L.)

Víctor Manuel Mayor-Duran[1]*, *Matthew Blair* [2], *Jaime Eduardo Muñoz* [2]

[1]International Center for Tropical Agriculture (CIAT), AA 6713, (CIAT), Cali, Colombia. [2]Universidad Nacional de Colombia - Palmira, AA 237, Palmira, Valle, Colombia. *Corresponding author: v.m.mayor@cgiar.org, vmmayord@unal.edu.co

Abstract

With the aim to estimate the coefficient of soil heterogeneity (b) without performing tests of uniformity, we proposed a methodology to use data from yield experiments isolating the treatment effect of the response variable. To identify methodological issues and illustrate the statistical proceeding management, data from one of the yield trials conducted in common bean by the Andean Breeding Program of CIAT were used. The coefficient of soil heterogeneity was estimated using the law of variance of Smith (1938) and the equation proposed by Federer (1963). Values of 0.59 and 0.66 respectively were obtained. Finally, we used the methodology of Hatheway (1961) and the "b" value estimated based on Federer (1963) to find the best combination of plot size, number of repetitions and difference to be detected as a mean percentage.

Keys words: Experimentation, field research, law of variance, *Phaseolus vulgaris*, soil heterogeneity, statistical analysis, yield.

Resumen

Con el objeto de estimar el coeficiente de heterogeneidad del suelo (b), sin realizar ensayos de uniformidad, se propuso una metodología en la cual se usan datos provenientes de ensayos de rendimiento, aislando el efecto de tratamiento de la variable de respuesta. Para definir aspectos metodológicos y mostrar algunos resultados obtenidos en el manejo estadístico de la información, se utilizaron los datos de un ensayo de rendimiento de frijol común realizado en el Centro Internacional de Agricultura Tropical (CIAT) por el Programa de Mejoramiento de Frijol Andino. El coeficiente de heterogeneidad del suelo se estimó a partir de la ley de varianza de Smith (1938) y la ecuación propuesta por Federer (1963) encontrando valores de 0.59 y 0.66, respectivamente. Teniendo como referencia el valor de b estimado a partir de la metodología de Federer (1963) se recurrió a la metodología de Hatheway (1961) para encontrar la mejor combinación de tamaño de parcela, número de repeticiones y diferencia a detectar como porcentaje de la media.

Palabras clave: Análisis estadísticos, experimentación, heterogeneidad del suelo, investigación de campo, ley de varianza, *Phaseolus vulgaris*, rendimiento.

Introduction

Soil variability due to changes in physical, chemical and biological characteristics does not allow homogeneity with experimental aims (Escobar, 1982). This variability is the most influential factor on the experimental error since it has a big impact on crop production. The absence of suitable controls for experimental error affects results precision on the research (Rosello and Gorostiza, 1993). To characterize the tendency on experimental plots, blank assays are used. These assays consist of sowing a pure line or a cultivar that is managed with similar practices to find yield differences in each plot that are due mainly to the soil variability, using plants as biological indicator. (Baena *et al.*, 1977; Escobar, 1982). Knowing soil heterogeneity plot sizes, repetition number and difference in yield between different treatments can be selected; this reduces the experimental error and increases the trust on the results (Escobar, 1982). The main aim of this research was to develop a methodology to estimate the heterogeneity coefficient of soil, in order to ensure the experimental conditions and create a yield map with results coming from yield experiments on common bean (*Phaseolus vulgaris* L.).

Materials and methods

Location and soils

Methodological and statistical management of the results were taken from a yield experiment on common bean that was done in the International Center for Tropical Agriculture (CIAT) in Palmira (03° 31' N, 76° 18' O; 1001 MASL) during the first semester of 2009 (January-April). Weather was characterized by a variable maximum temperature between 23.8 and 33 °C and minimum between 17.4 and 21.3 °C. Total precipitation was 270.2 mm irregularly distributed. Total evapotranspiration was 408.6 mm which corresponds to intermittent drought (Figure 1). Two supplementary irrigations by gravity were done the first day and 15 days after sowing.

Soil of the experimental site is a Mollisol Aquic Haplustoll silty clay with pH 7.0, M.O. 13.2% , P 70.5 mg/kg, 0.7, 11.5, 7.2 cmol/kg of K, Ca, Mg, respectively, and 22.4 mg/kg S, C.I.C. of 28.4.

Genotypes used

64 common bean lines were evaluated in this experiment. They are bush type and of agricultural interest for their drought tolerance. Experimental design was complete randomized blocks with three replicates and plot size of 3.09 m x 0.6 m with 2 plots per genotype, for a total of 192 units in 3708 m². Sowing distance between plants was 7 cm for an approximate density of 200,000 plants/ha.

Statistical analysis of results

Having into account that the evaluated genotypes were different, it was necessary to remove the effect of each one of them and, ho-

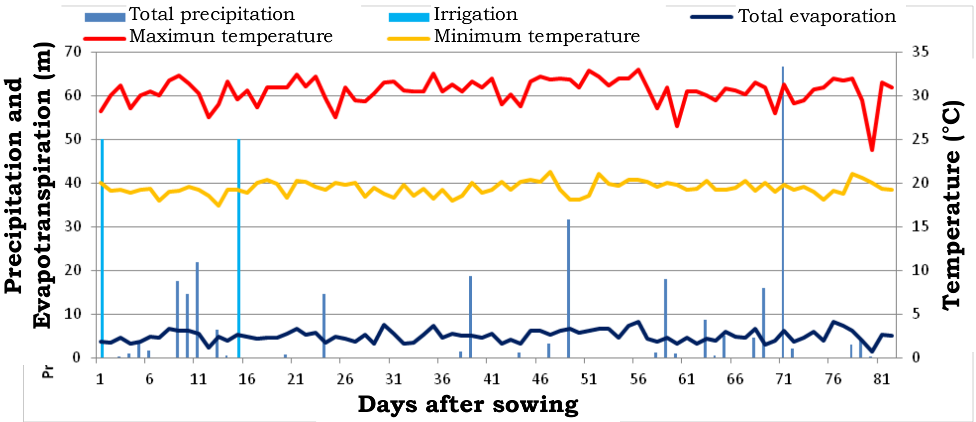

Figure 1. Weather conditions in the experimental location during January – April 2009.

mogenized the data obtained for further analysis as a uniformity assay. Next, it is presented an example used to analyze the experimental data based on a complete randomized block experimental design. Starting with the next relation:

$$Y_{ij} = \mu + G_i + B_j + \varepsilon_{ij} \qquad [\mathbf{1}]$$

where,

Yij = variable of response in the jth replication of the ith treatment

µ = general mean.

G_i = effect of genotype *i*.

B_j = effect of block *j*.

ε_{ij} = random error.

genotype effect should be estimated which is the difference between genotype average and general mean:

$$G_i = Y_{i.} - Y_{..} \qquad [\mathbf{2}]$$

Afterwards, the variable of response for each experimental unit is obtained, it is genotype effect free. With this value the production of ach plot is estimated like it has a uniform cultivar:

$$Y_{ij} - G_i = \mu + B_j + \varepsilon_{ij} \qquad [\mathbf{3}]$$

To construct the productivity map, the reference value taken was the standard deviation of yield for each plot as clustering criterion for homogeneous plots, meaning that, each one of the classes had an interval equal to the standard deviation of the unitary plots, in a range from 4.20 and 10.08 kg/plot (Figure 2).

Aiming to estimate the soil heterogeneity index (*b*) different arrays in size and shape were designed by summing up the yield of

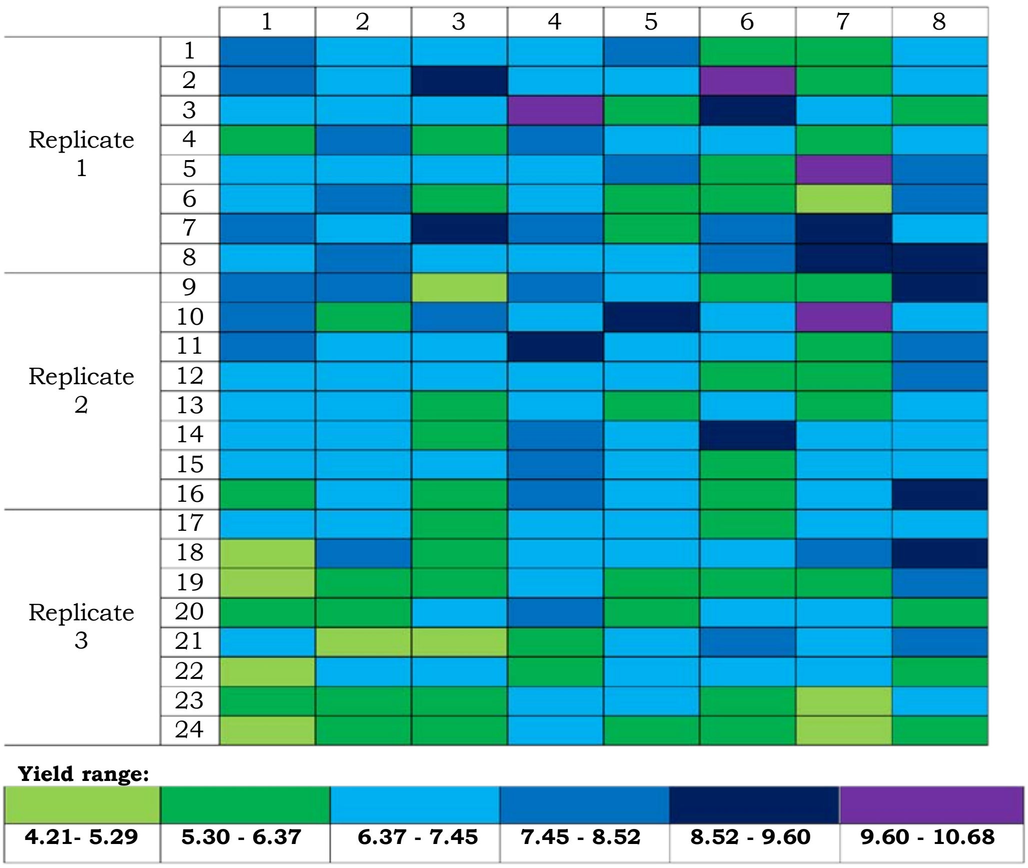

Figure 2. Productivity contour plot for yield (kg/plot) in 192 plots of 3.708 m^2 in the O_2 area of CIAT..

adjacent unitary plots. For each array the arithmetical mean, standard deviation, coefficient of variation and variance per unitary area were calculated and expressed in the following equation:

$$Vx_i = \frac{S^2x}{X^2} \qquad [4]$$

where, S^2x = variance between plots of X basic units, and X = number of basic units composing the plot.

When the plot size (X_i) is placed on the abscissa axis and variances per unitary area are plotted in the ordinate axis, the relation between the two variables is observed, that is the smith´s variance law (Smith, 1938).

$$Vx_i = \frac{V_1}{X^b} \qquad [5]$$

where,
V_1 = yield variance between unitary plots.
Vx_i = yield variance between secondary units, expressed per unit of area.
X_i = area of the secondary plots in the different arrays.
b = soil heterogeneity coefficient.

This expression can be linearized in logarithmic terms. Slope was proposed by Smith in 1938 as a measurement or index of soil heterogeneity.

$$\log Vx_i = \log V_1 - b \log X_i \qquad [6]$$

In some cases, heterogeneity coefficient (b) calculated by Smith´s equation can be higher than 1, which prevent a correct results interpretation because since b is an index of correlation between plot variability and soil heterogeneity coefficient, it is desired that it oscillates between 0 and 1. For this reason, Federer (1963) recommends that the soil heterogeneity index is estimated taking into account the variance logarithm and degrees of freedom associated with each array. Values of 0 correspond to totally homogeneous soils and when closer to 1, means a higher heterogeneity degree.

$$b = \frac{\sum(WiQiPi) - \sum(WiQi)\frac{\sum(WiPi)}{\sum Wi}}{\sum WiPi^2 - \frac{\sum(WiPi)^2}{\sum Wi}} \qquad [7]$$

where,
Qi = logarithm of yield variance per unit of area.
Pi = logarithm of the number of basic units in each plot size.
Wi = degrees of freedom associated with a given variance (number of plots with size Xi - 1).

From the soil heterogeneity coefficient it was estimated the optimal plot size; afterwards, a graph was built to determine the number of repetitions and values of difference between treatments means to be detected, this by using the Hatheway methodology (1961) expressed in the following equation:

$$X^b = \frac{2(t_1 + t_2)^2 * CV^2}{r * d^2} \qquad [8]$$

where,
X = optimal size of useful plot.
b = soil heterogeneity coefficient (weighted).
r = number of replicates.
d = difference needed to be detected between two treatments, expressed a mean percentage (+).
t_1 = 't' value on the Table for a given α level and (r-1) (t-1) degrees of freedom, being t = number of treatments.
t_2 = 't' value on the Table for (r-1) (t-1) degrees of freedom on one level.
α = 2(1 - p), where p is the probability estimated by the researcher to get a significant result.
CV = variation coefficient between unitary plots.

t_1 and t_2 values depend on the probability levels and degrees of freedom of the experimental error, which are chosen by the researcher according to the experiment. To calculate, researchers normally assume a significant level of 5%, the hope of getting significant differences in eight of ten experiments (P = 0.8) and assays with more than 14 degrees of freedom to estimate the experimental error. Under this concepts the $(t_1 + t_2)^2$ value is close to 9, which transforms the expression in:

$$X^b = \frac{18 * CV^2}{r * d^2} \qquad [9]$$

In the analysis done to define the optimal plot size and number of replicates, the experimental units were takes as X_i value. Each experimental unit corresponded to 3708 m^2. To develop the Hatheway's methodology (1961) the following variables were used (Ec. **9**): r = varied from 3 till 8; d = 5, 10, 15, 20, 25, 30, 35 and 40; b = estimated with Federer's methodology (1963); $2(t_1 + t_2)^2$ = 18; CV = variation coefficient between unitary plots; X = plot size in unities.

Gómez and Gómez (1984) proposed to estimate the soil heterogeneity coefficient using complete randomized blocks designs, divided plot or subdivided plots. For the complete randomized blocks design variance is estimated in the size plot –from the y block and d experimental unit- and for the subdivided plots design variance is estimated in the block, the main plot, the subplot and sub-subplot. Therefore, there is only two (complete randomized blocks) and four (subdivided plots) points to estimate the soil heterogeneity coefficient. In the methodology proposed by this study 30 points are used to estimate this coefficient, although they can be more if other plot arrays are considered in order to get a higher exactitude in the soil heterogeneity coefficient estimation.

Results and discussion

This study presented six classes or categories of homogeneous plots ranging from 4.20 to 10.08 kg/plot (Figure 2), with an interval length corresponding to the standard deviation of the unitary plots (1.08). In the Figure is appreciated the contour plot for yield showing clearly a high variation in terms of productivity in the different locations of the experimental area. A defined gradient in the area is not appreciated, showing the high heterogeneity degree on the soil. There is a slight trend in neighboring plots since they normally show a similar production.

From the practical point of view, knowing the productivity maps helps to prevent errors in future experiments where low or high production of a certain genotype could be the result of its location in the area of study. A productivity map allows the exclusion of certain areas and takes into account the fertility to determine experimental blocks to improve precision (Escobar *et al.*, 2006).

The soil heterogeneity coefficient was estimated by Smith´s variance law (1938) (Figure 3) and the equation proposed by Federer (1963) (Ec. **7**), the values were 0.59 and 0.66, respectively, which are high (Escobar, 1982). These values are similar, however, Federer´s method is more reliable due to its weighted coefficient, as such, it was selected as reference value to find the optimal plot size, number of replicates and the difference to detect compared to the mean.

In previous studies, Davis *et al.* (1981), Escobar (1982) and Escobar *et al.* (2006) found soil heterogeneity coefficients, in brackets, in CIAT plots cultivated with rice (0.64), cassava (0.87), climbing bean (0.87) and bush bean (0.75), indicating a high soil heterogeneity. The value obtained in this study (0.66)

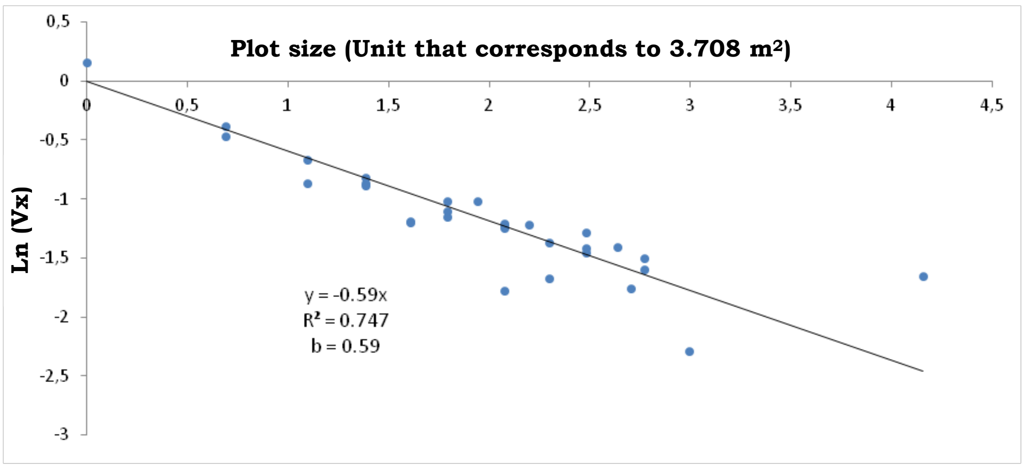

Figure 3. Lineal regression of the yield variance logarithm (*Vx*) in function of the size plot logarithm (*x*).

is similar to the previous ones, which indicates the effectiveness of the proposed method in this research.

The variation coefficient fluctuated between 6.38 and 15.78, with the highest value corresponding to the unitary plots (1 x 1) and the lowest to the 5 x 4 arrangement (Table 1). The fact that the lowest arrangement did not

Table 1. Statistical data from a common bean yield assay done in the O2 plot of CIAT.

Arrangement (no.)	**Arrangement**	**H**	**V**	**X_i**	**G.L.**	**Aver. (kg)**	**S^2**	**S**	**C.V**	***Vx*** (Ec. **4**)
1	1 x 1	1	1	1	191	6.84	1.16	1.08	15.78	1.16
2	1 x 2	1	2	2	95	13.69	2.49	1.58	11.52	0.62
3	1 x 3	1	3	3	47	20.35	3.78	1.94	9.55	0.42
4	1 x 4	1	4	4	47	27.37	6.69	2.59	9.45	0.42
5	1 x 5	1	5	5	23	33.89	7.57	2.75	8.12	0.30
6	1 x 6	1	6	6	23	40.70	11.37	3.37	8.29	0.32
7	1 x 7	1	7	7	23	47.43	17.63	4.20	8.85	0.36
8	1 x 8	1	8	8	23	54.74	18.69	4.32	7.90	0.29
9	2 x 1	2	1	2	95	13.69	2.72	1.65	12.05	0.68
10	2 x 2	2	2	4	47	27.37	6.55	2.56	9.35	0.41
11	2 x 3	2	3	6	23	40.41	11.84	3.44	8.52	0.33
12	2 x 4	2	4	8	23	40.70	10.78	3.28	8.07	0.17
13	2 x 5	2	5	10	11	67.78	25.22	5.02	7.41	0.25
14	2 x 6	2	6	12	11	81.41	33.40	5.78	7.10	0.23
15	2 x 7	2	7	14	11	94.86	47.79	6.91	7.29	0.24
16	2 x 8	2	8	16	11	109.49	56.94	7.55	6.89	0.22
17	3 x 1	3	1	3	63	20.53	4.60	2.14	10.44	0.51
18	3 x 2	3	2	6	31	41.06	12.93	3.60	8.76	0.36
19	3 x 3	3	3	9	15	61.06	23.93	4.89	8.01	0.30
20	3 x 4	3	4	12	15	82.12	34.79	5.90	7.18	0.24
21	4 x 1	4	1	4	47	27.37	7.05	2.66	9.70	0.44
22	4 x 2	4	2	8	23	54.74	19.14	4.37	7.99	0.30
23	4 x 3	4	3	12	11	81.41	39.92	6.32	7.76	0.28
24	4 x 4	4	4	16	11	109.49	51.54	7.18	6.56	0.20
25	5 x 1	5	1	5	31	34.92	7.54	2.75	7.86	0.30
26	5 x 2	5	2	10	15	69.83	18.69	4.32	6.19	0.19
27	5 x 3	5	3	15	7	103.79	38.49	6.20	5.98	0.17
28	5 x 4	5	4	20	7	139.66	40.53	6.37	4.56	0.10
29	8 x 1	8	1	8	23	54.74	18.31	4.28	7.82	0.29
30	8 x 8	8	8	64	2	437.96	779.59	27.92	6.38	0.19

H = Number of Rows, V = Number of Columns, X_i = Plot size (experimental units), G.L = Degrees of Freedom, S^2 = Variance between plots of X size, CV = Variation coefficient, *Vx* = Variance of production per unitary area (S^2/X^2).

Table 2. Plot size (m^2) calculated for different number of treatments and replicates to be detected (expressed as percentage of the mean) b = 0.66, C1= 15.78, α = 5%, P = 80%.

Difference to be detected (%)	**Repetitions**					
	3	**4**	**5**	**6**	**7**	**8**
5	1838.44	1188.12	846.84	642.17	508.23	415.01
10	224.31	144.97	103.33	78.35	62.01	50.64
12	128.99	83.36	59.42	45.06	35.66	29.12
14	80.79	52.21	37.22	28.22	22.33	18.24
16	53.87	34.82	24.82	18.82	14.89	12.16
18	37.68	24.35	17.36	13.16	10.42	8.51
20	27.37	17.69	12.61	9.56	7.57	6.18
25	13.90	8.99	6.40	4.86	3.84	3.14
30	8.00	5.17	3.68	2.79	2.21	1.80
35	5.01	3.24	2.31	1.75	1.38	1.13
40	3.34	2.16	1.54	1.17	0.92	0.75

have the highest number of plots involved (8 x 8), indicates that the analysis where up to some point effective, since a lot of variation was found in the arrangements evaluated.

Data in Table 2 and Figure 4 were obtained with the equation **9**. The yield assay had a plot size of 3.708 m^2 and three replicates. According to Figure 5 with these characteristics is possible to detect 38% of difference related to the mean. After the mean comparison test, the general mean was 1848.5 kg/ha and the DMS value = 546.3 that corresponds to 30% of difference with the average. The value obtained by the Hatheway´s methodology (1961) is similar to the ones of the yield assay, corroborating again the effectiveness of the proposed method to estimate the soil heterogeneity coefficient, number of repetitions and the difference to detect compared to the mean.

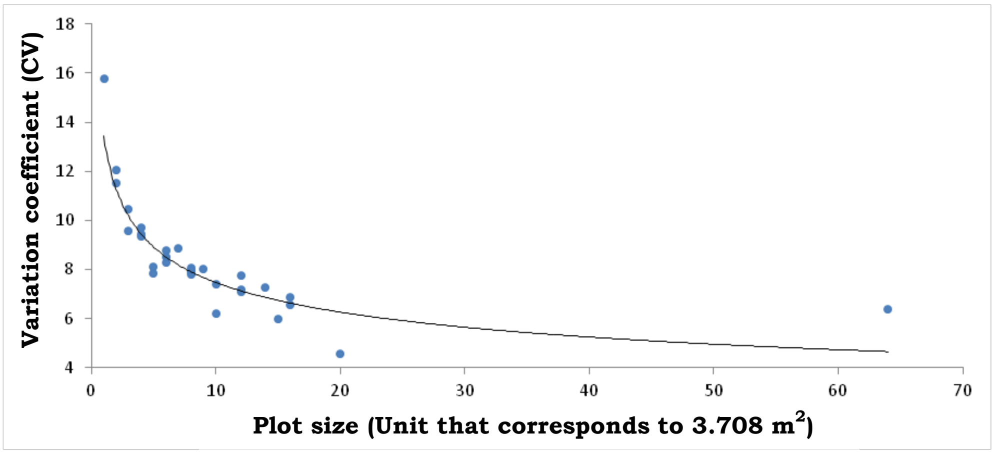

Figure 4. Relation between the yield variation coefficient (CV) and the plot unit.

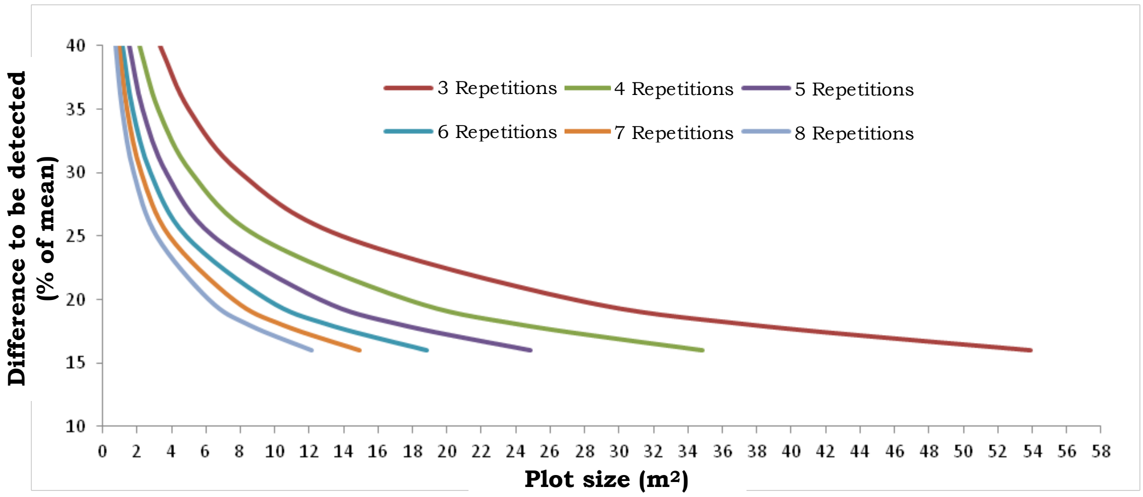

Figure 5. Relation between plot size, number of replicates and true difference to detect in a common bean experiment.

Conclusions

- The area for the yield assay showed neither a gradient nor a defined trend, but it does have a heterogeneity degree.
- To estimate the soil heterogeneity coefficient the Federer´s method (1963) is more reliable because it is a weighted regression coefficient. The area studied had a soil heterogeneity coefficient of 0.66.
- Important information for future experiments in the experimental area can be obtained with the Hatheway´s methodology (1961) since the effects of soil heterogeneity can be reduced by selecting the plot size, number of repetition and adequate experimental designs.
- The methodology proposed in this study is effective, therefore is useful for research center aiming to know their plots conditions without having to perform a uniformity test.
- It is important to state that the data obtained from this research had a huge importance in the breeding of Andean common bean for drought.

References

Baena, D.; Amezquita M. C.; Rodriguez, P. M.; Voysest, O.; and Takegami, F. 1977. Estudio de la heterogeneidad del suelo, del tamaño y forma de parcela y el número de repeticiones optimos en ensayos de uniformidad en frijol. XXIII Reunión Anual del Programa Cooperativo Centroamericano para el Mejoramiento de Cultivos y Animales.

Davis, J. H.; Amézquita, M. C.; and Muñoz, J. E. 1981. Border effects and optimum plot sizes for climbing beans (*Phaseolus vulgaris*) and maize in association and monoculture. Expl. Agric. 17:127 - 135.

Escobar, J. A. 1982. La heterogeneidad del suelo y los ensayos de uniformidad. Centro Internacional de Agricultura Tropical (CIAT). Cali - Colombia.

Escobar G., J. A.; Amézquita, M. C.; Muñoz, J. E.; and García, J. A. 2006. Manual de capacitación en biometría para la experimentacion en fríjol. Universidad Nacional de Colombia Sede Palmira. 133p.

Federer. W. T. 1963. Experimental desing. Nueva York. MacMillan. 544 p.

Gómez, K. A y Gómez, A. A. 1984. Statistical procedures for agricultural research. 2nd Edition. International Rice Research Institute. Filipinas. Pp. 479 - 503.

Hatheway, W. H. 1961. Convenient plot size. Agron. J. 53: 279 - 280.

Rosello, E. and Fernández, G. M. 1993. Guía técnica para ensayos de variedades en campo. Estudio FAO produccion y proteccion vegetal 75.

Smith, H. F. 1938. An empirical low describing heterogeneity in the yield of agriculture crops. J. Agric. Sci. 28: 1 - 23.