Biometrical Letters
Vol. 44 (2007), No. 1, 23-31

On determination of optimum size and shape of plots in field trials

Satyabrata Pal[1], Satyananda Basak[2], Subhabaha Pal[3], Sanpei Kageyama[4]

[1]Bidhan Chandra Krishi Viswavidyalaya, Department of Agricultural Statistics, Mohanpur, Nadia, West Bengal, 741252, India, e-mail: satbrpal@vsnl.net
[2]Uttar Banga Krishi Viswavidyalaya, Department of Agricultural Statistics, Pundibari, Coochbehar, West Bengal, 736165, India, e-mail: satyabasak@rediffmail.com
[3]Post Graduate Student, Calcutta University, India, e-mail: subhabaha@msn.com
[4]Hiroshima University, Department of Mathematics, 739-8524, Japan, e-mail: ksanpei@hiroshima-u.ac.jp

SUMMARY

One of the most important issues confronted by agriculturists/agronomists is to determine the suitable plot size and shape with respect to a particular crop, which will enable to maintain the organization and conduction of the experiment at a high level of precision. The solution to the problem is based on statistical considerations. This paper presents a general method by which the optimum plot size can be determined by a systematic analytic procedure. Though the method is described under isotropic situation, it can be employed to anisotropic situation as well.

Key words: Coefficient of variation, non-random data, radius of curvature, variogram,

1. Introduction

One of the chief difficulties in obtaining reliable results in field trials is the existence of the natural variability in the material with which we are dealing. This variability is reflected through the data obtained on yields from a particular experiment. From detailed studies carried out over the last century it is well accepted that the role of selection of appropriate size and shape of the plots has a substantial impact on the reduction of variability existing in the experimental material. The important references in this area recognize the works of Smith (1938), Modjeska and Rawlings (1983), Webster and Burgess (1984), Sethi

(1985), Bhatti et al. (1991), Zhang et al. (1990, 1994), Fagroud and Meirvenne (2002), etc., in which different procedures, based on some considerations, are spelt out. The fundamental work in this area, however, relates to the proposition of the equation

$$V_x = \frac{V_1}{x_b}, \tag{1.1}$$

where V_x is the variance of yields per unit area for plots of x units of area, V_1 is the variance for the basic plots and b is the heterogeneity coefficient, delineated by Smith (1938). In fact, V_x is calculated from plots of size x units, i.e., the size of the plot here is x times the size of the basic unit. Note that a basic plot is of size unity. If the plots are spatially uncorrelated, then the value of b equals 1 and in absence of heterogeneity the value of b approaches 0. An estimate of b is to be usually found by log-linearising the equation (1.1) using the least squares technique. Smith advocates the construction of a cost function (depending on x) and the optimum plot size is to be determined by minimizing the cost function with respect to x. Under the anisotropic condition of the field, the Smith law is generalised as

$$V_{n,s} = \frac{V_1}{(n_1^{b_1} n_2^{b_2})},$$

where n_1, n_2 are the numbers of basic plots taken along the row and column directions, respectively. Here, $V_{n,s}$ is the variance for plots each of which has $n = n_1 n_2$ basic plots, b_1 and b_2 are the indices which characterize the heterogeneity in the X, Y directions of a 2-dimensional field, respectively. Such anisotropic models are used by Modjeska and Rawlings (1983), Sethi (1985), Zhang et al. (1990, 1994), etc. For an isotropic field, b_1 equals b_2. For a uniform field, $b_1 = b_2 = 0$, and for a field with no spatial correlation, $b_1 = b_2 = 1$.

To obtain the optimum plot size we assume that a cost per unit is known. Let the cost function be of the linear form,

$$k_t = k_1 + k_2 x,$$

where k_t is the total cost for the experimental unit, k_1 is the part of the cost associated with the number of plots only, k_2 is the cost per unit area, and x is the size of plot. Then c being a cost/unit of information is given by

$$c = \frac{k_1 + k_2 x}{\frac{1}{V_x}} = \frac{V_1(k_1 + k_2 x)}{x^b}.$$

Now the value of x which minimizes the cost, c, can be found from the equation $dc/dx = 0$. Solving the equation for x, we obtain $x_{opt} = bk_1/[(1-b)k_2]$, which is the optimum plot size. The value of b, i.e., the index of heterogeneity, is used primarily to derive the optimum plot size. Zhang et al. (1994). obtained the values of n_1 and n_2 by minimizing $c_s=(k_1+k_2n_1n_2)\{V_1/(n_1{}^{b}{}_1n_2{}^{b}{}_2)\}$, where k_1 is a part of the cost associated with the number of plots only and k_2 is a cost per unit area.

Sethi (1985) suggests that Smith's law can be employed to the coefficient of variation, y, of a plot of size x as $y=ax^{-b}$. He suggested maximisation of the value of the radius of curvature,

$$\rho = \left(1 + y_1^2\right)^{3/2} \Big/ y_2,$$

where $y_1=dy/dx$ and $y_2=d^2y/dx^2$, and determined the optimum plot size x as

$$x_{opt} = \left\{a^2b^2(1+2b)/(2+b)\right\}^{1/[2(1+b)]}.$$

Modjeska and Rawlings (1983) have assumed the existence of a spatial correlation in the observations and worked on a model which assumes that the observations lying at equal lag distance (row and column) possess the identical correlation structure. They have determined the optimum plot size by using Smith's cost concept (see Smith, 1938). Fagroud et al. (2002) have exploited the variogram models and proposed the criteria, Nugget/Sill ratio (NSR) and NSR/Range (see also Cressie, 1993). They also advocated that for a plot size to be optimum, the model produces a large value for the NSR and a small value for the NSR/Range. However, the largest/smallest value of the criteria, NSR/ (NSR/Range), respectively, are found by identifying the maximum/minimum value from the set of different values of the corresponding coefficients (NSR or NSR/Range, etc) empirically and not by employing any analytical procedure.

This paper proposes a new method for finding out the optimum plot size from data collected from field experiments in case (i) when the data are random and (ii) when the data are correlated. This method has been used on two real life data sets.

2. Material and method

Sethi (1985) proposes a method of maximum curvature for the determination of the optimum plot size without mentioning the nature of data on which his method should work most appropriately. The present paper advocates to use a method of maximum curvature on values of the coefficient of variation (for different plot sizes) when the data feature is random. Furthermore, we add another stipulation that in order to find out the optimum plot size the curve which fits best through the observed points (with abscissa as plot size and with coefficient of variation as ordinate) is to be taken for the determination of the optimum plot size. In real life situation it is not a usual phenomenon that the observed data are random. This paper is devoted to the development of a distinct method for obtaining the optimum plot size, using the well-known variogram technique, which is used to discover spatial heterogeneity structure in a set of data. The fundamental basis of the method lies in studying the nature of the curvature of the variogram curve (using models which give close fits to the variogram values) for each possible plot size and shape. The theory of minimization of the absolute value of the radius of curvature is then invoked on the variogram curves. With respect to a particular model giving close fits ($R^2 > 0.7$) to the variogram data (developed on the observed variogram values), the plot size for which the absolute value of the radius of curvature is minimum, is selected as a candidate for the optimum plot size. Other models satisfying the above criteria are also to be examined. Final determination of the optimum plot size and shape is based on a study of all the selected candidates after application of the above criteria.

As an illustration, the proposed methodology has been applied on two data sets, one data set is obtained from a uniformity trial with JRO 524 (Naveen, a variety of jute) conducted at the Barokodali State Government Farm, Cooch Behar District, West Bengal, India. The seeds were sown in continuous lines, the distance between line to line being kept at 20cm leaving a border on each side. Uniform management practices were undertaken throughout the field. The field was harvested in continuous units of 448 basic units. There was in total 32 rows, each of which was along N – S direction and 14 columns, each of which was along E – W direction. Yields of dried fiber were calculated from each basic unit of size, 1m x 1m. The units (plots) were combined by taking 1 to 9 units along N$\rightarrow$S with 1 to 9 units across E$\rightarrow$W to form plots of different sizes and shapes. The other data set is obtained from a uniformity trial with MW10 (a variety of rice, Aus paddy) conducted at the Regional Research Station, Terai zone, Cooch Behar district, West Bengal, India. The seedlings were transplanted in lines with a hill to hill spacing of 20cm. The distance between line to line was kept at 20cm leaving a border on each side. Uniform management practices were undertaken throughout the field. There was in all 22 rows, each of which was along N $\rightarrow$ S direction, and 18 columns each of which was along E $\rightarrow$W direction. The field was harvested in units (plots) of 1m x 1m, the size of the sown area was 22m x 18m (with 396 plots). The yields of these plots were separately dried and weighed correct to the nearest gram. In case of the data obtained under designed structure (say, randomised block design), the variogram values are to be calculated on the basis of the residuals obtained from the different plots. Suppose that $\text{var}(Z(s_1)-Z(s_2)) = 2\gamma(s_1 - s_2)$ for all s_1, s_2, where s_1, s_2 are two locations and $Z(s_i)$ denotes the random function at location s_i and the location may represent one- or two-dimensional point in the context of this paper. The quantity $2\gamma(.)$, which is the function of the increment s_1-s_2, has been called a variogram and $\gamma(.)$ is called a semivariogram (see, Cressie, 1993).

Now, $\gamma(h;\theta)$ represents a semivariogram model consisting of parameters h and θ, where h is a lag distance between observations (locations), $\|h\|$ denotes the Euclidean distance, and θ is the vector of model parameters to be estimated. Here a one-dimensional location plot is considered for h being equal to s_1-s_2, s_1 and s_2 being two locations. If we designate the plot in terms of two dimensions, then Z is designated by two co-ordinates, h and k, h and k denoting the corresponding row and column positions.

The classical estimator of the variogram in case of one-dimensional plot as proposed by Matheron (1963) is

$$2\hat{\gamma}(h) \equiv \frac{1}{|N(h)|}\sum_{N(h)}(Z(s_i)-Z(s_j))^2,$$

where the sum extends over $N(h) \equiv \{(i,j): s_i - s_j = h\}$ and $|N(h)|$ is the number of distinct elements of $N(h)$. Cressie and Hawkins (1980) present a more robust approach to the estimation of variogram:

$$2\bar{\gamma}(h) \equiv \left\{\frac{1}{|N(h)|}\sum_{N(h)}|Z(s_i)-Z(s_j)|^{1/2}\right\}^4 \Bigg/ \left(0.457 + \frac{0.494}{|N(h)|}\right).$$

Here the word "robust" is used to describe inference procedures that are stable when model assumptions depart from those of a central model. We have used the robust approach (formula) in this paper.

Different numbers of units along row and column directions are used to construct larger plots with different plot shapes. For each size and shape the semivariogram values are calculated and subsequently the following semivariogram models are fitted:

(i) Spherical model:

$$\gamma(h,\theta)=\begin{cases}0, & \text{if } h=0,\\ c_0+c_1\{(3/2)(\|h\|/a)-(1/2)(\|h\|/a)^3\}, & \text{if } 0<\|h\|\le a,\\ c_0+c_1, & \text{if } \|h\|\ge a,\end{cases}$$

where $\theta=(c_o,c_s,a_s)'$, $c_o\ge 0, c_s\ge 0$, $a_s\ge 0$. Here a is the range, c_0+c_1 is the sill and c_0 is the nugget variance.

(ii) Exponential model:

$$\gamma(h;\theta)=\begin{cases}0, & \text{if } h=0\\ c_o+c_e\{1-\exp(-\|h\|/a_e)\}, & \text{if } h\ne 0\end{cases}$$

where $\theta=(c_o,c_e,a_e)'$, $c_o\ge 0, c_e\ge 0$, $a_e\ge 0$. Here c_0 is the nugget variance, c_0+c_e is the sill and a_e is the distance parameter controlling the spatial extent of the function.

(iii) Gaussian model:

$$\gamma(h;\theta)=\begin{cases}0, & \text{if } h=0\\ c_o+c_g\{1-\exp[-(\|h\|/a_g)^2]\}, & \text{if } h\ne 0\end{cases}$$

where $\theta=(c_o,c_g,a_g)'$, $c_o\ge 0, c_g\ge 0$, $a_g\ge 0$. Here c_0 is the nugget variance, c_0+c_g is the sill and a_g is the distance parameter controlling the spatial extent of the function.

(iv) Michaelis Menton Model:

$$\gamma(h;\theta)=\begin{cases}0, & \text{if } h=0\\ c_o+c_m(\|h\|/a_m)/(1+\|h\|/a_m), & \text{if } h\ne 0\end{cases}$$

where $\theta=(c_o,c_m,a_m)'$, $c_o\ge 0, c_m\ge 0$, $a_m\ge 0$. Here c_0 is the nugget variance, c_0+c_m is the sill and a_m is the distance parameter controlling the spatial extent of the function.

(v) Von Bertalenffy (VB) model:

$$\gamma(h,\theta)=L\infty(1-\exp(-k\,(h-t_o)))$$

where $\theta = (L_\infty, k, t_o)$. Here the parameters, t_0, k, and L_∞ are the base, shape, and limiting parameters, respectively.
These parameters can also be interpreted in terms of sill, nugget, etc. as in the above expressions.
The solutions to h_{opt} and the formulae for the corresponding radius of curvature ρ for different models (Spherical, Exponential, Michaelis-Menton and VB models) are given below:

Model	h_{opt}	Radius of Curvature, ρ
Spherical: $y=a+b(1.5(x/c)-0.5(x/c)^3)$	$c\sqrt{\dfrac{6b+\sqrt{81b^2+20c^2}}{15b}}$	$\dfrac{\left(1+\dfrac{9b^2}{4c^2}\left(1-\dfrac{x^2}{c^2}\right)^2\right)^{1.5}}{-\dfrac{3b}{c^3}x}$
Exponential: $y = a - b\exp(-x/c)$	$c\log_e\left(\dfrac{\sqrt{2}b}{c}\right)$	$\dfrac{\left(1+\dfrac{b^2}{c^2}e^{\frac{-2x}{c}}\right)^{1.5}}{-\dfrac{b}{c^2}e^{-\frac{x}{c}}}$
Michaelis-Menton: $y = a - b\exp[-(x/c)^2]$	$-c+\sqrt{bc}$	$\dfrac{\left\{1+\dfrac{b^2}{c^2}\left(1-\dfrac{x}{c}\right)^{-4}\right\}^{1.5}}{-2\dfrac{b}{c^2}\left(1-\dfrac{x}{c}\right)^{-3}}$
VB: $y= L_\infty[1-exp\{-k(t-t_0)\}]$	$t=t_0+\dfrac{1}{k}\log_e(\sqrt{2}L_\infty k)$	$\dfrac{\left\{1+L_\infty^2k^2e^{-2k(t-t_0)}\right\}^{1.5}}{-L_\infty^2k^2e^{-k(t-t_0)}}$

We have obtained the value of radius of curvature based on h_{opt} value. Here we search for the plot size which minimises the absolute value of the minimum value of radius of curvature at h_{opt}.

This should be noted that the proposed criterion on the determination of the optimum plot-size is to be applied on either uniformity trial data or on data on residuals of the yield observations, recorded from an experiment laid out in randomized blocks, latin squares, etc. The above-mentioned residuals (in case of RBD/LSD) are free from treatment effects and as such the proposed criterion can be employed on such residuals. The essence here is to give importance to the fact that even with Fisherian blocking, the correlation among the residuals is not eliminated, as such residuals remain correlated. Based on the considerations mentioned above the proposed criterion will be a useful tool in the hands of applied statisticians in case of non-random residuals obtained from designed experiments (RBD/LSD).

3. Results and discussion

At the outset, it is mentioned that GENSTAT package was used for non-linear fitting of the different variogram models on the two uniformity trial data sets. For jute data and rice data the R^2 values vary from 75% to 99% and from 74% to 96% respectively in case of the above-mentioned models (excepting Spherical, by application of which the plot sizes are not within acceptable limits under the consideration).

Based on the selection criterion of minimum value of the absolute value of radius of curvature with respect to each model, optimum plot sizes are shown in the following table:

Table 1. Optimum plot sizes for four different models for rice data

Model	ρ_{min}	R^2	Plot size
Michaelis-Menton	-29.51	0.73	2 x 2
VB	-0.012	0.82	4 x 3
Exponential	-0.08	0.96	2 x 5
Gaussian	-0.09	0.81	5 x 3

Table 2 . Optimum plot sizes for four different models for jute data

Model	ρ_{min}	R^2	Plot size
Michaelis-Menton	-9.38	0.81	3 x 2
VB	-0.16	0.99	9 x 1
Exponential	-0.17	0.99	9 x 1
Gaussian	-0.126	0.99	9 x 1

From Tables 1 and 2 the optimum plot sizes and shapes are found to be 2x5 (Exponential model) for rice data and 9x1 (any of the models, VB, Exponential Gaussian) for jute data.

REFERENCES

Bhatti A.U., Mulla D.J., Koehler F.E., Gurmani A.H. (1991). "Identifying and removing spatial correlation from yield experiments. Soil Sci. Soc. Am. J. 55, 1523-1528.

Cressie N.A.C. (1993). Statistics for Spatial Data. John Wiley, New York.

Cressie N., Hawkins D. M. (1980). Robust estimation of the variogram, I. Jour. of International Association for Mathematical Geology 12, 115-125

Fagroud M., Meirvenne M. van (2002). Accounting for soil spatial autocorrelation in the design of experimental trials. Soil Sci. Soc. Am. J. 66, 1134-1142.

Matheron G. (1963). Principles of geostatistics. Economic Geology 58, 1246-1266.

Modjeska J.S., Rawlings J.O. (1983). Spatial correlation analysis of uniformity data. Biometrics 39, 373-384.

Sethi A.S. (1985). A modified approach to determine the optimum size and shape of plots in field experiments on maize grown on terraced land. Indian Jour. of Agril. Sci. 55(1), 48-51

Smith H.F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. Jour. of Agricultural Sciences, Cambridge 28,1-29.

Webster R., Burgess T.M. (1984). Sampling and bulking strategies for estimating soil properties in small regions. J. Soil Sci. 35, 127-140.

Zhang R, Warrick A.W., Myers D.E. (1990). Variance as a function of sample support size. Math. Geol. 22, no.1, 107-121.

Zhang R, Warrick A.W., Myers D.E. (1994). Heterogeneity, plot shape effect and optimum plot size. Geoderma 62, 183-197.