Method for the determination and interpretation of multivariate response functions

Gy. Biczók Ministry of Environmental Protection and Regional Development, Budapest, Hungary
L. Tolner, University of Agricultural Sciences, Gödöllô, Hungary
Gy. Simán, University of Agricultural Sciences, Uppsala, Sweden

ABSTRACT

Soil fertility is influenced by the joint effect of a number of factors. The vastly increasing number of experimental treatments required as the number of variables increases causes problems when multivariate response functions are estimated, as does the interpretation of the large number of parameters in the correlation. Interpretation is complicated by the fact that the more parameters the correlation contains, the greater the problem caused by the interdependence of the estimated parameters.

It is possible to estimate the parameters independently and to reduce the number of experiments to a realistic level if the experiments are set up according to an orthogonal factor design, though these designs are not suitable for calculating the parameters of Baule-Mitscherlich function containing combinations of exponential terms.

The multivariate extension of the parabolic response function recommended here leads to a quadratic polynomial form, the parameters of which can be reasonably estimated by applying an orthogonal factor design. The interpretation of the model parameters was investigated considering the optimum values of the individual factors and the possible characteristics of the available nutrient forms in the soil.

INTRODUCTION

An ever-green problem encountered in soil fertility research is the agronomic interpretation of the often non-linear effects and interactions of the many factors. The first attempts at solving the problem of requirements that are difficult to meet simultaneously involved concentrating on the effect of one factor and ignoring that of the other factors.

A univariate linear correlation was suggested by Liebig (1862) to describe the effect of factors at a relative minimum level. It was found that the linear correlation was only valid over a certain limited range. Mitscherlich (1909) recommended a saturation equation as a response function. The basic idea was that the crop yield increment (dy/dx) is proportional to the distance between the actual yield (y) and the maximum yield (A) (proportionality factor: k) due to the effect of that factor which is at the level of relative minimum (x).

$$\frac{dy}{dx} = k \cdot (A - y)$$

The relationship between the yield (y) and the given factor (x) is obtained by integrating the above differential equation.

$$y = A \cdot (1 - e^{-k \cdot x})$$

In case of extremely high values the effect of any factor may turn to negative, and the plants can even die. Mitscherlich (1928) characterised the effect according to the previous equation modified by the introduction of a damaging factor.

$$y = A \cdot (1 - e^{-k \cdot x}) \cdot e^{-c \cdot x}$$

Liebscher (1887) proposed a quadratic parabolic equation to describe the relationship between the active factor and the yield.

$$y = a + b \cdot x + c \cdot x^2$$

Despite of its simplicity, this model is able to describe both the positive effects and negative ones.

The agrochemical interpretation of the parabolic response function was attempted by Di Gléria (1959), based on the assumption that the crop yield increase or decrease *(dy/dx)* is proportional to the difference between the necessary nutrient quantity *M* for the maximal yield and the just available one *(x)*

$$\frac{dy}{dx} = K \cdot (M - x)$$

where K is the proportionality factor. Integrating this differential equation gives a quadratic parabola as the model, the mathematical description of the response in question:

$$y = A(\frac{2 \cdot x}{M} - \frac{x^2}{M^2})$$

where A is maximum crop yield at a value x = M. When expressed by the terms of K and M, it has a value of : $A = \frac{K \cdot M^2}{2}$.

From the point of view of plant nutrition, the response function representing the correlation between the applied nutrients and the yield is of outstanding importance. By transforming the above model, Di Gléria (1959) made the parabolic function suitable for the description of this correlation.

If the available quantity of the nutrient in the soil is b and the applied (therefore known) quantity is z, in that case the available nutrient quantity for the plants is x=b+z. If maximum yield (A) can be achieved by the application of the nutrient quantity B in that case the total quantity of nutrients required by forming A includes the quantity b which was present in the soil before the fertilization i.e.: M = b + B. If these are substituted into the above equation, we obtain:

$$y = \frac{2 \cdot A \cdot B \cdot b}{(b+B)^2} + \frac{2 \cdot A \cdot B}{(b+B)^2} \cdot \mathbf{z} - \frac{A}{(b+B)^2} \cdot \mathbf{z^2}$$

The parameters of this parabolic response function can be estimated by regression calculation using the data of the various nutrient treatments and the corresponding yields. The parameters obtained can be interpreted in agrochemical sense: A is the maximum yield, B the nutrient quantity distributed to achieve A yield, and b the available nutrient quantity in the soil (Di Gléria 1959).

Despite of the fact that the greater part of the nutrient effect is really caused by the factor functioning at the level of relative minimum, the effects of the other factors are negligible only if the supplying of the other nutrients as factors exists on a satisfactory level and the factor of relative minimum in question causes an extreme nutrient deficiency. The more demand is to supply the balanced plant nutrition, the more importance is to consider the joint effects of all the factors.

The multifactorial models can be derived by the combination of unifactorial model. In an effort to expand the Mitscherlich model (1909), Baule (1953) suggested to multiply the uniform function-terms of the various single factors.

$$y = A \cdot (1 - e^{-k_1 \cdot x_1}) \cdot (1 - e^{-k_2 \cdot x_2}) \cdot (1 - e^{-k_3 \cdot x_3}) \cdot \ldots$$

In spite of its relative complexity, this model is not suitable for the description of the damageous effects and interactions, respectively.

The model of Di Gléria (1959) can be transformed into a multifactorial function in a similar way. The result of the transformation is a multivariate polynome. In the case of normalised variables the terms of higher exponents can be ignored and the multivariate quadratic polynome proves a good approximation (Biczók 1988, Tolner and Biczók 1989). The parameters of the model can be estimated by setting up and evaluating experimental treatments prescribed by quadratic orthogonal factor design (Box and Wilson 1951, Nalimov and Chernova 1965, Kafarov et al. 1976, Adler et al. 1977). Further on it is an relevant information that the treatments are varied on five levels of the individual factors. If the number of the regulated factors in question is *n* and all the individual factors are varied on five different levels the traditional, frequently used fertilization experiments require the adjustment of 5^n treatments. It would mean a cumbersome, superfluous work in comparison with the prescription of quadratic orthogonal factor design that can support not only much smaller number of treatments (2^n+2n+1) but the independent estimate of the model parameters, too.

In order to compare the traditional and quadratic orthogonal design - depending on the number of the experimental factors - the Table 1 presents the number of treatments adjusting 5 levels of factor. According to the figures in Table 1 it must be obvious why the multifactorial model was not widely used in the earlier period of agricultural experimentation.

This model, together with the so-called SITOBI software supporting its application, is gaining ground under the name DISITOBI in agrochemical and soil microbiological research in Hungary. The designation is derived from the names of the scientists involved in its elaboration (**Di** Gléria, **Si**mán, **To**lner, **Bi**czók). Though the model is being applied ever more widely (Loch et al. 1986, Kis et al. 1986, Loch et al. 1987, Németh et al. 1993, Abd El Galil et al. 1993, Szili Kovács et al. 1993), no attempt has yet been made to give an agrochemical interpretation of the model parameters. In the present paper there is an intention to introduce the designing and parameter estimating method for agrochemical and soil microbiological experimentators and to interpret the model parameters.

MATERIAL AND METHODS

The advantage of orthogonal factor designs is especially obvious when more than two factors must be jointly considered. In the course of plant nutrition experiments it is advisable to study the joint effect of as many adequate variables as possible. For this reason orthogonal factor designs are not reasonably used for bivariate experiments. However, the method can be best demonstrated introducing the simplest bivariate case as an example of the fictive experiment on fertilizer effect. Let us take wheat as the test plant and fertilisers containing N and P active agents as the factors to be studied. The N treatment range from 0-120 kg/ha and the P treatments from 0-180 kg/ha active agents are adjusted according to a quadratic orthogonal factor design. The experimental treatments and the corresponding yields are demonstrated in Table 2.

Before carrying out the calculations the values of the variables should be normalised.

$$x_i = \frac{(value.of.\text{var.}) - (mean.of.\text{var.})}{(upper.value.of.\text{var.}) - (mean.of.\text{var.})}$$

In the case of the N treatments this means that

$$x_N = \frac{N - 60.2}{110 - 60.2}.$$

The response function described by the normalized factors is the following bivariate parabolic function:

$$y = b_1 + b_2 x_N + b_3 x_P + b_4 x_N x_P + b_5 (x_N^2 - d) + b_6 (x_P^2 - d)$$

or using matrix designations:

$$\underline{Y} = \underline{B} \cdot \underline{\underline{X}}$$

The experimental design (Nalimov and Chernova 1965) can be written in matrix form ($\underline{\underline{X}}$) and the yields in vector form ($\underline{Y}$). The design matrix contains the normalized numeric values of the treatments related to the constant in the first column, the linear terms (x_N, x_P) in the second and the third columns, the interaction term ($x_N x_P$) in the fourth column and the quadratic terms ($x_N^2 - d$, $x_P^2 - d$) in the fifth and the sixth ones.

$$\underline{\underline{X}} = \begin{vmatrix} +1 & +1 & +1 & +1 & 1-d & 1-d \\ +1 & -1 & +1 & -1 & 1-d & 1-d \\ +1 & +1 & -1 & -1 & 1-d & 1-d \\ +1 & -1 & -1 & +1 & 1-d & 1-d \\ +1 & +\alpha & 0 & 0 & \alpha^2 - d & -d \\ +1 & -\alpha & 0 & 0 & \alpha^2 - d & -d \\ +1 & 0 & +\alpha & 0 & -d & \alpha^2 - d \\ +1 & 0 & -\alpha & 0 & -d & \alpha^2 - d \\ +1 & 0 & 0 & 0 & -d & -d \\ +1 & 0 & 0 & 0 & -d & -d \\ +1 & 0 & 0 & 0 & -d & -d \\ +1 & 0 & 0 & 0 & -d & -d \end{vmatrix} \qquad \underline{Y}_e = \begin{vmatrix} 6.5 \\ 5.0 \\ 5.5 \\ 3.5 \\ 6.0 \\ 4.0 \\ 6.0 \\ 4.5 \\ 5.5 \\ 5.4 \\ 5.5 \\ 5.6 \end{vmatrix}$$

If the values of a (1.21) and d (0.58) are reasonably chosen, this design is orthogonal, which means that the product of any two column vectors in the design matrix ($\underline{\underline{X}}$) is equal to zero. The advantage of the orthogonal design matrix becomes clear in the course of linear regression calculations. The sum of residual square (SQ) can be calculated by the difference between the experimentally measured yield values ($\underline{Y}_e$) and the calculated ones ($\underline{Y}$)

$$SQ = (\underline{Y}_e - \underline{Y})^T (\underline{Y}_e - \underline{Y})$$

approaching the best fitting, the values of the regression coefficients (b_i), i.e. vector $\underline{B}$, can be obtained using a formula derived from the equation expressing the minimum sum of residual squares:

$$\underline{B} = (\underline{\underline{X}}^T \underline{\underline{X}})^{-1} (\underline{\underline{X}}^T \underline{Y}_e)$$

After calculating the information matrix, it is clear why the orthogonality of the design matrix is so important

$$\underline{\underline{X}}^T \underline{\underline{X}} = \begin{vmatrix} 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6.9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.3 \end{vmatrix}$$

In the course of the calculations the scalar products obtained by multiplying the column vectors of the design matrix ($\underline{\underline{X}}$) by themselves form the main diagonal, while all the other components of the information matrix, obtained by multiplying different column vectors by each other, have a value of zero. This enables us to determine the b_i model parameters (i.e. the regression coefficients) independently of each other (Kafarov 1976).

DISCUSSION

The values of the parameters (b_i) are obtained by carrying out the regression calculation. Using normalised variables, the response function can be written, as follows:

$$y = \underset{}{5.47} + \underset{***}{1.00}x_N + \underset{***}{0.48}x_P + \underset{*}{0.13}x_N x_P - \underset{***}{0.59}x_N^2 - \underset{+}{0.08}x_P^2.$$

If the Student probe of the b_i was proved significant on the level of 0.1 % probability it was marked by *** , on the level of 5 % probability the mark of * was applied and on the level of 10 % probability the mark of + was used (below the regression coefficients). The variance of the yield values was estimated by the help of the yield values related to the 4 central treatments (9-12) - as it is shown in the Table 2.

Substituting the normalized variables by their defining formulas

$$x_N = \frac{N - 60.2}{49.8} \qquad\qquad x_P = \frac{P - 86.5}{71.5}$$

into the bivariate response function (described by the normalized variables) the quadratic N-P response function is obtained as follows:

$$y = 2.9 + 0.046N + 0.0073P + 0.000035NP - 0.00024N^2 - 0.000016P^2$$

The calculated response function is illustrated as a graph of trajectories (Fig. 1).

By formal analysis of the response function, remarkable answers can be obtained for important questions concerned with plant nutrition:

1. What is the maximum yield which can be obtained by optimising treatments of a given type?

2. What is the fertilizer equivalent effect of the soil nutrient quantity, i.e. in other words: how much of the nutrient quantity in the soil manifests its own "availability"?

3. What yields can be expected according to the various treatment combination values?

The adequate fertilizer dose belongs to the maximal yield value can be calculated as a derivate of the response function. When the general coefficients are used in the equation, the multivariate quadratic response function has the next form:

$$y = a_0 + a_N N + a_P P + a_{NP} N P - a_{N^2} N^2 - a_{P^2} P^2$$

At the point giving maximum yield ($N=M_N$, $P=M_P$), both the increase and the decrease of either the N or the P doses prove the least effect on the yield. The limiting value of the yield change is zero as it is shown by the next partial differentials:

$$\frac{\delta y}{\delta N} = a_N + a_{NP} M_P + 2 \cdot a_{N^2} M_N = 0 \qquad \frac{\delta y}{\delta P} = a_P + a_{NP} M_N + 2 \cdot a_{P^2} M_P = 0$$

By solving this bivariate equation system and by substituting the relevant coefficient values (a_i), the values of N and P treatments required for maximum yield (M_N, M_P) can be calculated.

$$M_N = \frac{a_P a_{NP} - 2 \cdot a_N a_{P^2}}{4 \cdot a_{N^2} a_{P^2} - a_{NP}^2} = \frac{0.0073 \cdot 0.000035 + 2 \cdot 0.046 \cdot 0.000016}{4 \cdot 0.00024 \cdot 0.000016 - 0.000035^2} = 122.2 \quad \text{(kg/ha)}$$

$$M_P = \frac{a_N a_{NP} - 2 \cdot a_P a_{N^2}}{4 \cdot a_{N^2} a_{P^2} - a_{NP}^2} = \frac{0.046 \cdot 0.000035 + 2 \cdot 0.0073 \cdot 0.00024}{4 \cdot 0.00024 \cdot 0.000016 - 0.000035^2} = 361.8 \quad \text{(kg/ha)}$$

By substituting the values of M_N and M_P into the response function, the value of the maximum yield (A) can be calculated:

$$A = a_0 + a_N M_N + a_P M_P + a_{NP} M_N M_P - a_{N^2} M_N^2 - a_{P^2} M_P^2$$

The calculation showed the maximum yield (A) to be 7 t/ha.

If the response function is examined when one of the variables is constant, we obtain the univariate parabolic function derived by Di Gléria (1959). If the constant chosen is P = 0

$$y = a_0 + a_N N - a_{N^2} N^2 .$$

This is a univariate section of the bivariate response function (Fig. 2). It can be seen from the figure that this parabola crosses the N axis at two points, where y=0. In the case of the treatments characterised by these points, there is no crop yield expected according to the prediction of the model.

The extrapolation of function beyond the real range of measurement must be carefully and informally handled, since going on to the remote points of the treatment scale the interval

of confidence is getting wider and wider depending on the distance of the central point of factor design. In addition to this, the model must lose its validity due to the well-known and unknown plant physiological switches of regulative threshold-values, if the symptoms of toxicity occur at the extreme level of any factor.

The quantity of available soil-N is designated Q_N (Fig. 2) which - assuming a positive sign - is the N quantity required to give a yield of y_0 in the N=0 treatment. Thus, the value of Q_N is the N-fertilizer equivalent of the soil-N, i.e. it is the value of available nutrient estimated by the response function. At point L_N (Fig. 2) the yield is as well as zero, i.e. this treatment has poisoned the plant evenmore occasionaly it may be a lethal dose.

These two values can also be determined by calculations if the univariate parabolic function is solved for y=0.

$$Q_N = \frac{a_N - \sqrt{a_N^2 + 4 \cdot a_0 a_{N^2}}}{2 \cdot a_{N^2}} \qquad L_N = \frac{a_N + \sqrt{a_N^2 + 4 \cdot a_0 a_{N^2}}}{2 \cdot a_{N^2}}$$

The results of the calculation are: Q_N = -49.8, L_N = 241.4, Q_P = -254.6 and L_P = 717.9 (kg/ha). Thus, the available N calculated from the response function is 49.8 kg/ha and the available P is 254.6 kg/ha.

Similarly to the example presented here, the experiments planned to study the effects of more than two factors can also be designed and evaluated.

REFERENCES

ABD EL GALIL A., RADIMSZKY, L. BACZÓ, GY. and NÉMETH T. 1993. Study of the AL-soluble phosphorus content in incubation experiments. Agrokémia és Talajtan 42: 179-182.

ADLER, Y. P., MARKOVA, E. V. and GRANOVSKI, Y. V. 1977. Design of experiments for the determination of optimum conditions. Műszaki Könyvkiadó Bp., Mir Publisher, Moscow.

BAULE, B. 1953. Über die Weiterentwicklung der Ertragsgesetze von Liebig und Mitscherlich. Z. Acker- und Pflanzenbau 96: 173-186.

BICZÓK,GY.(1988) Biocybernetic strategy for engineering of soil fertility, in Model construction use in soil fertility control (Seminarieserie) Sveriges Lantbruksuniversitet, Uppsala, Sverige, April.

BOX, G. E. P. and WILSON, K. B. 1951. On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society, B, 13/1, 1.

DI GLERIA, J. 1959. Agricultural chemistry. Akadémia Kiadó Budapest.

KAFAROV, V. V. 1976. Cybernetic Methods in Chemistry and Chemical Engineering. Mir Publisher, Moscow.

KIS, SZ., BICZÓK, GY. and TOLNER, L. 1986. Effect of N fertilisation at varying P, K and water supply levels. Chemical Products in Agriculture 86 Conference, Keszthely 203-207.

LIEBIG, J. 1862. Einleitung in die Naturgesetze des Feldbaues, 143., Braunschweig.

LIEBSCHER, G. 1887. Der Verlauf der Nährstoffaufnahmen und seine Bedeutung für die Dungerlahre. Journ. für Landwirtschaft 35: 333-518.

LOCH, J., BICZÓK, GY. and TOLNER, L. 1986. Mg fertilisation and K-Ca-Mg antagonism. Chemical Products in Agriculture '86 conference Keszthely 62-67.

LOCH, J., KIS, SZ., VÁGÓ, I., BICZÓK, GY. and TOLNER, L. 1987. Joint effect of N, P, K, Ca, Mg and water supplies on chernozem and acidic sand. I. Yield data. Chemical Products in Agriculture '87 conference Keszthely 53-58.

MITSCHERLICH, E. A. 19O9. Das Gesetz des Minimums und das Gesetz des abnehmenden Bodenertrages. Landw. Jb. 38: 537-552.

MITSCHERLICH, E. A. 1928. Die zweite Annaherung des Wirkungsgesetzes der Wachstumsfaktoren. Z. Pflanzenernahr. Düng. Bodenkunde, A 12: 273-282.

NALIMOV, V. and CHERNOVA, N. 1965. Statisticheszkie metody planirovania ekstremalnyh eksperimentov. Izdatelstvo Nauka, Moscow.

NÉMETH, T., ABD EL GALIL A., BACZÓ, GY. and RADIMSZKY, L. 1993. Study of the ammonium-N and nitrate-N contents of different soils during incubation. Agrokémia és Talajtan 42: 173-178.

SVÁB, J. 1973. Biometric methods in research. 2nd revised, enlarged edition. Mezőgazdasági Kiadó, Budapest.

SZILI KOVÁCS, T., RADIMSZKY, L., ANDÓ, J. and BICZÓK, GY. 1993. CO_2 evolution from soils formed on various parent materials in the eastern Cserhát mountains (Hungary) during laboratory incubation. Agrokémia és Talajtan 42: 140-146.

TOLNER, L. and BICZÓK, GY. 1989. Design and evaluation of agrochemical experiments. A Magyar Biometriai Társaság X. Konferenciája, Martonvásár.