T1992AGR29

ANALYSIS OF SERIES OF VARIETIES TRIALS : A NEW APPROACH

P. LETERME, A. MESSÉAN and B. RIBOUD

CETIOM - 174, Ave Victor Hugo 75116 PARIS

SUMMARY

Each year, a network of about one hundred trials is designed in order to evaluate the performance of sunflower varieties in France. Classical mean performance evaluation is not sufficient and even unreliable in most cases due to the incomplete varieties x trials tables and the high interaction observed between varieties and location sites. The paper presents a method based on the Mandel multiplicative interaction model which characterizes each variety and each location site by two parameters : the mean estimate and a stability parameter. Both methodological and practical aspects are discussed and the results for two series of experiments (1990 and 1991) are presented throughout a two-dimensional map for varieties and location sites.

1. INTRODUCTION

Every year more than 100 field experiments are carried out in France in order to evaluate the performance of Sunflower varieties. After performing a statistical analysis for each individual trial, the trials are pooled together and a regional and/or a national classification of the varieties is obtained. Due to the interaction between varieties and the environmental conditions, the mean rank of each variety is not sufficient, as a great variability in the varietal ranking has been observed when changing the environment.

In the complete case (each variety checked in each location site), the most common analysis of variance model used is the multiplicative model of Mandel where the interaction term is defined as the product of two parameters. The properties of such a model have been studied (Denis, 1983 ; Chadoeuf & Denis, 1988) and it has been implemented in different packages like for example, Stat/ITCF (Tranchefort *et al.*, 1991).

But in most cases, due to the great number of varieties and the great variability in maturity, each variety is not present in each location site and the design is incomplete. The results obtained with the Mandel model in the complete case must thus be extended and we present here their application to the national interpretation of Sunflower varieties'results collected by CETIOM in 1990 and 1991.

2. THE STATISTICAL MODEL

To analyze a serie of varieties trials, we use the following multiplicative interaction model :

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_i.\theta.\delta_j + \varepsilon_{ij}$$

where i=1,2,...,I indicates the levels of the environment factor, j=1,2,...,J the levels of the variety factor. y_{ij} is the mean yield obtained in the i^{th} location site by the j^{th}

variety. The parameters α_i and β_j denote the i^{th} environmental treatment effect and the j^{th} varietal treatment effect, respectively. The parameters γ_i and δ_j are the interaction parameters and denote through the term $\gamma_i.\theta.\delta_j$ the additional response (positive or negative) expected when the j^{th} variety is grown at the i^{th} location. Higher is the absolute value of δ_j, higher is the capability of the j^{th} variety to interact with the specific location site : a variety with a small absolute value $|\delta_j|$ will have quite the same behaviour whereas it is grown. θ denotes the global interaction value and μ the mean yield value. It is assumed that the errors ε_{ij} are independently and normally distributed with mean zero and variance σ^2 but the individual residual standard deviation of each individual trial can also be introduced without lack of consistency of the results.

The standard restrictions

$$\Sigma_i\alpha_i=\Sigma_j\beta_j=\Sigma_i\gamma_i=\Sigma_j\delta_j=0$$
$$\Sigma_i\gamma_i^2=\Sigma_j\delta_j^2=1$$

are used to ensure a complete definition of model.

2.1. Parameter estimation

The usual least squares estimates of the parameters are used. Due to the non-linearity, no analytical solution for the parameter estimates exists and an iterative numerical process must be performed. To estimate these values in the incomplete case, we use the Alternating Least Squares method, which was first described by Kroonenberg (1983) and specifically designed for Linear and Bilinear two-way classification models by Denis (1991). This method consists to the decomposition of the parameters in not necessary disjoint subsets, and to estimate each subset of parameters conditioning the others respectively. As a matter of fact, when one of the multiplicative parameters γ_i or δ_j is considered as a constant, the model becomes linear with respect to the other parameters and classical algorithms for linear models can then be used. Figure 1 presents a schematic description of the fitting algorithm.

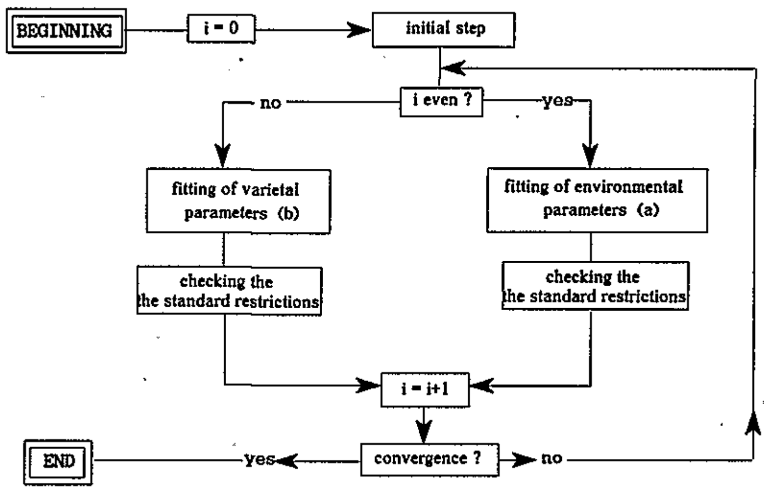
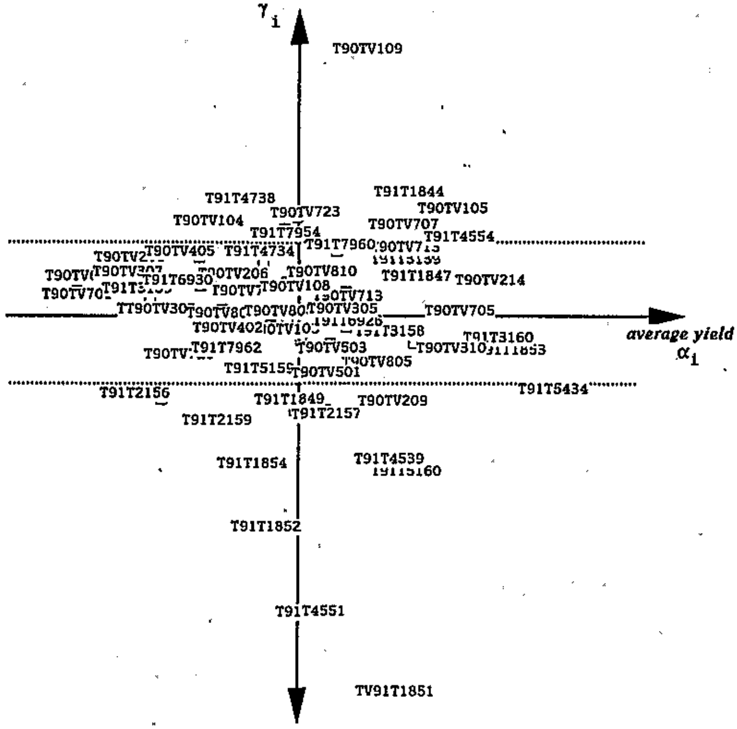


Figure 1. Scheme of the fitting algorithm

Figure 3 represents the two-dimensional environmental map and displays the stabilty environmental parameter γ_i versus the average trial yield $\mu + \alpha_i$. This is thus the symmetric representation of figure 2.



<u>Figure 3.</u> Representation of the trials by αi and γi parameters.

In these representations, αi and βj are interpreted like the meaning behaviour expected for the trial i and the variety j, respectively. And γi (resp. δj) represents the capacity for the i^{th} trial (resp. j^{th} variety) to interact with varieties (resp the environment). Thus δj can be interpreted like the risk for the j^{th} variety to over-step or back-step its expected performance.

In those trials where $|\gamma_i|$ is small, the varieties present the same ranking and no interaction is expressed.

In the trials where $\gamma_i>0$, the varieties which are characterized by $\delta_j>0$, like Vidoc and Oscar, will step over their meaning yield expected, whereas the varieties with $\delta_j<0$ like Frankasol, Eurosol or Albena will step back it, and the varieties with a δ_j near of 0, as Viki, Suntop or Briosol, have the same behaviour as expected through their average yield.

On the other side, in those trials where γi is negative, the varieties with a positive δj present a poorer behaviour while Frankasol and Eurosol are favoured and Viki, Suntop or Briosol still remain with their mean behaviour.

So in the T91VCE4738 trial, Vidoc is the best variety and its yield is 6.5 q/ha higher than Albena and 7.5 q/ha higher than Frankasol, while in the T91VCE1851 trial, the best variety is Frankasol, which is better than Oscar by 15 q/ha and Vidoc by 2.5 q/ha.

Although Albena, Eurosol, Suntop, Oscar and Vidoc have quite the same mean yield value (not significantly different), we thus notice that their behaviour can be very different when changing the location site. In fact, we can see on the figure 2 that the precision of the parameter estimates is sufficient to build three groups of varieties by taking into account the δ_j parameter value :

- Viki, Suntop which have quite the same behaviour whatever are the environmental conditions,
- Frankasol, Albena and Eurosol which interact with the environmental conditions the same way,
- Vidoc and Oscar which highly interact with the environment and in an opposite way with respect to Frankasol, Albena and Eurosol.

Of course, this kind of information is only descriptive and based on a statistical meaningful. Our objective remains to be able to unterstand and to explain why we observe such a behaviour for the variety. Such an analysis requires a large number of observations is collected from the trial, e.g. the soil characteristics, the climate, diseases and pests observations. In our first analysis, we identified the Sclerotinia disease as the main factor explaining the great variability of Vidoc and Oscar. Nevertheless, even if we could not explain the behaviour, a simple descriptive map like figure 2 already gives very useful information for the farmer : for example, if one usually get good results with Albena in a specific environment, one could probably predict that it will be also true for Eurosol.

Our example here show the mean characteristics of the varieties during two years of experimentation. Although we obtained a good precision, one can wonder how variable is the variety map when changing the year conditions. Figure 4 shows the maps obtained by separate analysis performed for 1990 and 1991. We notice that the relative location of the varieties is similar and that the displacement of the varieties on the map is mainly due to the poorer results obtained by Vidoc and Oscar in 1991, a year which was characterized by a high level of Sclerotinia infestation. Further years are required in order to check the reliability of the method to provide a stable map for prediction.

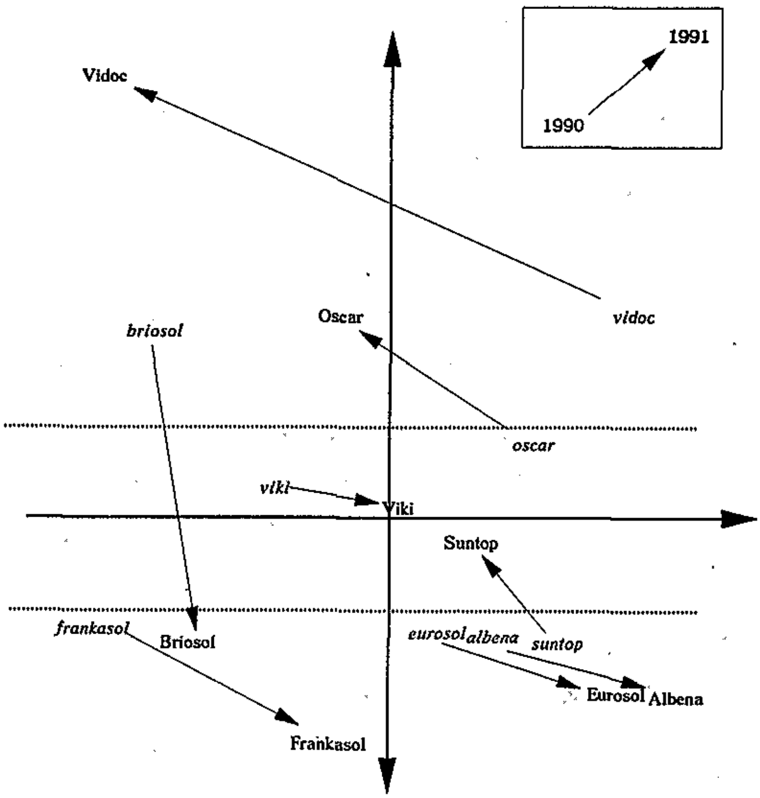


Figure 4. Variety map obtained by separate analysis for 1990 and 1991.

4. CONCLUSION

We have shown through the example that statistical tools are now available for analyzing the interaction between the genotype and the environmental conditions and that such a method is reliable and provides the end-user, i.e. the farmer, a more complete information. The two-dimensional map of figure 2 is being introduced in the dissemination booklets edited by CETIOM towards the farmers. We think it is a simple and useful tool which can help the farmer in choosing his variety : choose a less productive variety but ensure its behaviour will be "stable", choose the variety which is the closest to the variety which already gives him very good results in his particular environmental conditions or choose a more speculative variety. From the experimenter 's point of view, such a method allows us to reduce the number of trials through an optimal choice of locations sites and by the value-added provided by the statistical analysis.

REFERENCES

[1] CHADOEUF, J. & DENIS, J.B. (1988). Asymptotic variances for the multiplicative interaction model. *Journal of Applied Statistics*, 18, 331-353.

[2] DENIS, J.B. (1983). *Interaction entre deux facteurs*. Thèse de Docteur Ingénieur en Sciences Agronomiques, INA-PG, 59-100.

[2] DENIS, J.B. (1991). Ajustements de modèles linéaires et bilinéaires sous contraintes linéaires avec données manquantes. *Revue de Statistique Appliquée*, 29, 5-24.

[3] DECOUX G. & DENIS J.-B. (1991). *INTERA : a package for analysis of the interaction in the incomplete case*. INRA, Département de Biométrie - Versailles

[4] EFRON B. (1979). Bootstrap methods : another look to jackknife. *Annals of Statistics*, 7, 1-26.

[5] GOODMAN, L.A. & HABERMAN, S.J. (1990). The analysis of nonadditivity in two-way analysis of variance. *Journal of the American Statistical Association*. 85, 139-145.

[6] HUET,S., RIBOUD, B. & DENIS J.B. (1992) Parametric Bootstrap applied to multiplicative interactive model. Submitted for publication to *Biometrika*.

[7] KROONENBERG, P.M. (1983). *Three-Mode Principal Component Analysis*. DSWO Press, Leiden, 398pp.

[8] KRUSKAL, J.B. (1977). Some Least-Squares theorems for matrices and N-way Arrays. Communication to the *first International Meeting of Data Analysis and Informatics* INRIA Versailles - France.

[9] RIBOUD, B. (1991). *INTERACT : a Unix package for analysis of interaction*. Rapport Technique du CETIOM. Paris.

[10] TRANCHEFORT J. et al. (1991). *Stat/ITCF : manuel d'utilisation*. ITCF - Boignieville