

REVUE DE STATISTIQUE APPLIQUÉE

C. DEBOUCHE

J. STEINIER

À propos de deux méthodes d'ajustement de modèles mathématiques non linéaires

Revue de statistique appliquée, tome 22, n° 3 (1974), p. 5-22

http://www.numdam.org/item?id=RSA_1974__22_3_5_0

© Société française de statistique, 1974, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

A PROPOS DE DEUX MÉTHODES D'AJUSTEMENT DE MODÈLES MATHÉMATIQUES NON LINÉAIRES (1)

C. DEBOUCHE (2) et J. STEINIER (3)

1. Introduction
2. Le problème général de l'ajustement par la méthode des moindres carrés
3. Le problème linéaire
4. Le problème non linéaire : résolution par la méthode de Gauss
5. Le problème non linéaire : résolution par la construction d'un simplexe
6. Critique de l'ajustement
7. Utilisation pratique des programmes écrits pour l'ordinateur IBM 1130
8. Discussion et conclusion

Le problème général de l'ajustement d'une fonction (ou d'un modèle mathématique) à un ensemble de valeurs observées est présenté dans ces pages.

Deux méthodes itératives sont appliquées pour effectuer cet ajustement. La première est basée sur le calcul des dérivées partielles du modèle mathématique linéarisé en un point donné. La seconde remplace le calcul des dérivées partielles par le calcul des variations finies de la fonction aux environs du point considéré. Ces deux processus sont appliqués à des modèles mathématiques de natures très variées pouvant contenir jusqu'à une dizaine de paramètres. Ils se complètent utilement dans la recherche du meilleur ajustement. Enfin, quelques exemples rencontrés dans des problèmes de spectrométrie, courbes de croissances, ventilation et conserverie sont également présentés. Ils ont été traités avec succès par ces deux méthodes.

(1) Travail subventionné par l'Institut pour l'Encouragement de la Recherche Scientifique dans l'Industrie et l'Agriculture (I.R.S.I.A.) ; remis le 10/9/73

(2) Faculté des Sciences agronomiques de Gembloux. Chaire de Statistique. [BELGIQUE].

(3) Laboratoire central du C.A.M.I.R.A. (I.R.S.I.A.)

1. INTRODUCTION

L'ajustement, au sens des moindres carrés, de fonctions arbitraires à des ensembles de données est un problème fréquent dans de nombreux domaines.

L'observation d'un phénomène fournit en effet souvent à l'expérimentateur des informations qu'il voudrait expliquer par une fonction ou un modèle mathématique. Le choix de ce dernier peut être orienté par des considérations théoriques relatives au phénomène étudié ou, plus empiriquement, par l'allure générale des observations recueillies. Une attention particulière doit toutefois être accordée aux représentations analytiques qui permettent une réelle interprétation, plutôt qu'un ajustement idéal.

Le choix du modèle mathématique étant fait, il convient de l'adapter au mieux aux observations, en calculant les constantes ou paramètres qui assurent la meilleure cohérence entre le modèle théorique et les valeurs observées.

Outre l'obtention de valeurs "idéales" des paramètres, il faut aussi pouvoir chiffrer l'erreur inhérente aux données et donc aux estimations des paramètres.

La solution de ces problèmes d'ajustement, quoique conceptuellement simple et connue de longue date, se heurte cependant à des calculs souvent ardues et, seul l'emploi d'ordinateurs a permis d'en rendre l'utilisation plus systématique.

Une première méthode d'ajustement, dite "méthode de Gauss" nécessite la connaissance de l'expression analytique des dérivées partielles du premier ordre du modèle mathématique. A partir d'approximations de départ suffisamment bonnes pour les paramètres, elle peut conduire rapidement aux "meilleures" valeurs pour les paramètres, ainsi qu'à des estimations de leur précision. Des difficultés de convergence appréciables apparaissent cependant lorsque les valeurs initiales des paramètres sont trop éloignées de la réalité.

Une deuxième méthode, basée sur l'examen des variations finies de la fonction aux environs du point considéré, est employée pour remédier aux difficultés rencontrées lors de l'emploi de la méthode de Gauss. Parmi l'éventail des algorithmes utilisant les variations finies, le choix s'est porté sur le calcul de la fonction aux extrémités d'un simplexe (Nelder et Mead, 1965). Cette méthode semble bien adaptée, d'une façon générale, à l'ensemble des problèmes rencontrés et nécessite un nombre relativement limité de calcul de la fonction. Ce choix n'exclut évidemment pas la possibilité de rencontrer des algorithmes plus adéquats pour certains types de modèles (Vignes, 1969). Une méthode qui s'avèrerait optimale pour tous les problèmes non linéaires semble de toute manière difficile à trouver.

Un ensemble de programmes et sous-programmes ont été écrits en Fortran pour utiliser ces méthodes d'ajustement sur un ordinateur IBM1130 (16 K mots de 16 bits) ou HP 2114 A (8 K mots de 16 bits).

2. LE PROBLEME GENERAL DE L'AJUSTEMENT PAR LA METHODE DES MOINDRES CARRES

Soit une fonction :

$$f(x_1, x_2, \dots, x_m ; t_1, t_2, \dots, t_p), \quad (1)$$

et soit :

$$T = (t_1, t_2, \dots, t_p) \text{ le vecteur des paramètres.}$$

Soit d'autre part l'ensemble des n observations :

$$(y_i, x_{1i}, x_{2i}, \dots, x_{mi}), \quad i = 1, 2, \dots, n, \quad n > p,$$

qui s'écrit encore sous forme vectorielle :

$$(Y, X)$$

où

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad \text{et} \quad X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix}$$

Le problème consiste à déterminer les estimations des p paramètres t_k ($k = 1, 2, \dots, p$) de la fonction (1) qui s'ajuste le mieux à cet ensemble de n observations. Soit \hat{T} cet estimateur du vecteur T . Le critère du meilleur ajustement retenu étant celui des moindres carrés, il s'agit de minimiser la somme S des carrés des écarts entre les valeurs observées y_i ($i = 1, 2, \dots, n$) et les valeurs correspondantes calculées au moyen de (1), soit :

$$S = \sum_{i=1}^n w_i (y_i - f_i)^2, \quad (2)$$

où on a posé :

$$f_i = f(x_{1i}, x_{2i}, \dots, x_{mi} ; t_1, t_2, \dots, t_p).$$

Sous forme vectorielle, l'expression (2) devient :

$$S = (Y - F)' W (Y - F), \quad (3)$$

où F est le vecteur des f_i ($i = 1, 2, \dots, n$) et W la matrice diagonale (de dimensions $n \times n$) des poids w_i associés aux différents y_i . Une condition nécessaire de minimisation de cette somme de carrés S relativement aux t_k est donnée par :

$$\frac{\delta S}{\delta T} = 0.$$

Différentiant, on a :

$$\frac{\delta F'}{\delta T} W (Y - F) = 0, \quad (4)$$

soit un ensemble de p équations à résoudre en les p composantes du vecteur T , ensemble qui constitue les "équations normales" du problème. La forme (4) est, en général, un système de p équations non linéaires ; de plus il n'est pas certain qu'une solution de ce système existe ou que, si elle existe, elle soit unique. Il faut également noter que, bien que pouvant être quelconques, il est possible d'assigner aux w_i une signification particulière.

Supposons chacun des y_i issu d'une certaine distribution de moyenne :

$$\mu(y_i) = f(x_{1i}, \dots, x_{mi} ; \alpha_1, \dots, \alpha_p), \quad (5)$$

et de variance proportionnelle à une fonction donnée de (x_1, \dots, x_m) , c'est-à-dire :

$$\sigma^2(y_i) = \sigma^2 h^2(x_{1i}, \dots, x_{mi}). \quad (6)$$

Ce faisant, un paramètre supplémentaire, σ^2 , est apparu qui doit être estimé. Posant :

$$w_i = \frac{1}{h^2(x_{1i}, \dots, x_{mi}) \sigma^2(y_i)}, \quad (7)$$

et supposant les y_i issus chacun d'une distribution de moyenne et de variance respectivement données par (5) et (6), nous obtenons comme estimateur non biaisé de σ^2 :

$$s^2 = \frac{\sum_{i=1}^n w_i [y_i - f(x_{1i}, \dots, x_{mi} ; \hat{t}_1, \dots, \hat{t}_p)]^2}{n - p}, \quad (8)$$

où l'on note que le numérateur du second membre est la somme minimisée des carrés des écarts obtenus en remplaçant t_k par \hat{t}_k dans (2). Lorsque, pour toutes les valeurs des x_i ,

$$h^2(x_1, \dots, x_m) = c,$$

où c est une constante, la matrice W se réduit à :

$$W = c I,$$

I étant la matrice unitaire de dimension $(n \times n)$. La variance en chaque point est alors estimée en multipliant s^2 par c . Il s'en suit évidemment que, lorsque $c = 1$,

$$W = I,$$

et s^2 est automatiquement une estimation de la variance en chaque point. La quantité s^2 est alors appelée "variance de l'ajustement".

3. LE PROBLEME LINEAIRE

Parmi l'ensemble des fonctions pour lesquelles des solutions du système (4) existent nécessairement, considérons l'ensemble de celles dans lesquelles les paramètres t_k apparaissent comme coefficients linéaires des variables indépendantes. Dans ce cas, le vecteur F prend la forme :

$$F = X T, \text{ (avec } m = p \text{)} \quad (9)$$

et (4) devient :

$$X' W (Y - X T) = 0,$$

d'où l'on tire immédiatement comme estimateur \hat{T} du vecteur T des paramètres:

$$\hat{T} = (X' W X)^{-1} X' W Y. \quad (10)$$

Il n'y a bien sûr aucune solution si la matrice $(X' W X)$ (de dimensions $p \times p$) est singulière ; en fait on peut montrer que ce cas ne se produit que si deux ou plusieurs des variables indépendantes sont hautement corrélées ; dans ce cas, il suffit de sélectionner un sous-ensemble non corrélé de variables indépendantes.

On montre d'autre part que la matrice de dispersion de \hat{T} , soit $V(\hat{T})$ est de la forme :

$$V(\hat{T}) = \sigma^2 (X' W X)^{-1} ; \quad (11)$$

autrement dit, les éléments diagonaux de la matrice $V(\hat{T})$ fournissent les variances $\hat{\sigma}_{\hat{t}_k}^2$ des estimateurs des paramètres t_k , tandis que les éléments non diagonaux de $V(\hat{T})$ fournissent les covariances $\text{cov}(\hat{t}_j, \hat{t}_k)$ des estimateurs des paramètres t_j et t_k ($j = k$).

Dans le cas où les y_i sont normalement distribués, la substitution de l'estimateur s^2 de σ^2 , fourni par (8), permet l'estimation de ces quantités. Enfin, des intervalles de confiance relatifs à chacun des t_k peuvent alors être calculés par :

$$\hat{t}_k - \hat{\sigma}_{\hat{t}_k} \cdot t_{(1-\alpha/2)} < t_k < \hat{t}_k + \hat{\sigma}_{\hat{t}_k} \cdot t_{(1-\alpha/2)},$$

où $t_{(1-\alpha/2)}$ est la valeur de la variable t à $n-p$ degrés de liberté à laquelle correspond une fonction de répartition égale à $1-\alpha/2$. La valeur $1-\alpha$ est le coefficient de confiance ou de sécurité. Remarquons cependant que, lorsque les variables indépendantes X_j sont fortement corrélées, les intervalles de confiance considérés séparément n'ont guère de sens (Dagnelie, 1966).

4. LE PROBLEME NON LINEAIRE : RESOLUTION PAR LA METHODE DE GAUSS

Lorsque les paramètres t_k n'apparaissent pas comme coefficients linéaires des variables indépendantes dans (1), les développements et conclusions du paragraphe ci-dessus ne sont plus applicables. L'estimateur \hat{T} du vecteur des paramètres t_k ne peut plus se calculer au moyen d'une expression telle que (10). Une approche du problème consiste alors à essayer des combinaisons des valeurs des paramètres dans (2), à examiner les sommes S ainsi obtenues et à sélectionner la combinaison fournissant la plus petite valeur de S . Cette technique n'assure pas que le minimum vrai de S ait été atteint. De plus cette méthode n'exclut pas la possibilité d'existence d'un autre ensemble de valeurs des paramètres conduisant à un même minimum de S . Enfin, la masse de calcul requise peut rapidement devenir prohibitive.

La méthode de Gauss permet une meilleure approche du problème. Son principe consiste essentiellement à rendre linéaire en chacun des paramètres la fonction sélectionnée au moyen d'un développement tronqué en série de Taylor. A partir d'estimations initiales des paramètres permettant d'évaluer les coefficients du développement, on déduit de nouvelles estimations qui fournissent elles-mêmes une nouvelle évaluation des coefficients ; le processus est ainsi répété jusqu'à satisfaction d'un critère de convergence donné.

Si l'on note par $\hat{T}_0 = (\hat{t}_{1,0}, \hat{t}_{2,0}, \dots, \hat{t}_{p,0})$ le vecteur des estimations initiales des paramètres de la fonction (1), le vecteur F des valeurs que cette fonction prend en chacune des n observations peut s'écrire, au moyen d'un développement en série de Taylor limité aux termes du 1er ordre :

$$F(X, T) = F(X, \hat{T}_0) + \left[\frac{\delta F}{\delta T} \right]_0 \Delta, \quad (12)$$

où $\left[\frac{\delta F}{\delta T} \right]_0$ désigne la matrice de dimensions (n, p) et de terme général :

$$\left[\begin{array}{c} \delta f_i \\ \delta t_k \end{array} \right]_{t_k} = \hat{t}_{k,0}, \quad (i = 1, 2, \dots, n ; k = 1, 2, \dots, p),$$

et Δ le vecteur de dimensions $(p \times 1)$ de terme général :

$$\delta_k = (t_k - \hat{t}_{k,0}), (k = 1, 2, \dots, p). \quad (13)$$

Il s'agit de nouveau de minimiser la somme S des carrés des écarts :

$$S = (Y - F)' W(Y - F),$$

où le vecteur $(Y - F)$ s'écrit maintenant, vu la relation (12) :

$$(Y - F) = (Y - F_0) - \left[\frac{\delta F}{\delta T} \right]_0 \Delta \quad (14)$$

Posant $R = Y - F_0$ (de dimensions $n \times 1$)

et

$$D = \left[\frac{\delta F}{\delta T} \right]_0 \quad (\text{de dimensions } n \times p),$$

la relation (14) se réécrit :

$$(Y - F) = R - D\Delta.$$

De nouveau, une condition nécessaire de minimisation de la somme S relative aux t_k s'écrit :

$$\frac{\delta S}{\delta T} = 0 = 2 \left[D \frac{\delta \Delta}{\delta T} \right]' W(R - D\Delta). \quad (15)$$

Vu la définition de Δ , la matrice $\frac{\delta \Delta}{\delta T}$ est unitaire de dimensions $p \times p$. Dès lors on a :

$$\left[D \frac{\delta \Delta}{\delta T} \right]' = D'.$$

et la relation (15) se réduit à :

$$D' W(R - D\Delta) = 0,$$

dont on tire facilement l'estimateur $\hat{\Delta}$ de Δ , soit :

$$\hat{\Delta} = (D' W D)^{-1} D' W R. \quad (16)$$

Le vecteur $\hat{\Delta}$ ainsi obtenu constitue donc l'approximation de premier ordre de la correction à apporter au vecteur \hat{T}_0 pour obtenir le vecteur T cherché. Si l'une au moins des composantes δ_k du vecteur $\hat{\Delta}$ est telle que :

$$|\hat{\delta}_k| > \epsilon \quad (17)$$

où ϵ est une constante préalablement définie, on effectue la transformation :

$$\hat{T}_1 = \hat{T}_0 + \hat{\Delta},$$

redéfinissant ainsi un vecteur d'approximations initiales des paramètres. On en déduit les nouvelles valeurs du vecteur R et de la matrice D qui, par (16), conduisent à un nouvel estimateur $\hat{\Delta}$.

Le processus est ainsi répété jusqu'à ce que, après q itérations, toutes les composantes de $\hat{\Delta}$ soient inférieures à ϵ . Le vecteur \hat{T}_q constitue alors l'estimation au sens des moindres carrés du vecteur T des paramètres.

Il convient de remarquer ici que la méthode de Gauss permet en particulier de résoudre le problème linéaire. Il suffit dans ce cas de choisir comme approximation initiale le vecteur $\hat{T}_0 = 0$. Dans ces conditions, en effet, et puisque F est de la forme $F = X T$, il vient

$$F_0 = 0, D = X, R = Y \text{ et } \Delta = T, \quad (19)$$

tandis que (16) se réduit à (10).

D'autre part, la matrice de dispersion de \hat{T} , soit $V(\hat{T})$, est alors de la forme :

$$V(\hat{T}) = \sigma^2 (D' W D)^{-1}, \quad (20)$$

expression dans laquelle l'estimateur s^2 fourni par (8) peut encore être substitué à σ^2 si les résidus sont indépendants entre eux. On peut donc déduire, comme dans le cas linéaire, des intervalles de confiance pour les différents paramètres.

5. LE PROBLEME NON LINEAIRE : RESOLUTION PAR LA CONSTRUCTION D'UN SIMPLEXE

Le problème posé est identique à celui qui a été discuté dans les lignes précédentes : il faut trouver la valeur du vecteur T des paramètres qui minimise la quantité S définie en (3). Rappelons tout d'abord que la méthode présentée dans ce paragraphe (Nelder et Mead, 1965) ne se distingue de la précédente que par le processus d'optimisation de la somme des carrés des écarts entre les valeurs observées et les valeurs calculées au moyen du modèle. Le critère d'ajustement reste le critère des moindres carrés.

La méthode de Gauss étudie les variations à donner aux paramètres à partir d'un point connu, pour minimiser S en calculant les dérivées partielles de la fonction S par rapport à chaque paramètre.

La méthode présentée ici calcule la valeur de S en les différents points d'un simplexe dont l'origine se trouve en \hat{T}_0 (il s'agit du vecteur contenant les estimations initiales des p paramètres). Il convient de préciser que la méthode proposée par Nelder et Mead est différente de l'algorithme dit "du simplexe" employé pour l'optimisation d'une fonction linéaire dans un espace restreint par des contraintes également linéaires. En effet, l'optimisation peut se dérouler dans un espace libre de toutes contraintes et dans le cas de fonction non linéaire.

On fixera donc autour de l'origine p points qui sont les extrémités du simplexe construit dans un espace à p dimensions. Les valeurs prises par la fonction en ces points sont notées S_i et les vecteurs qui les définissent s'écrivent :

$$T_i = (t_1, t_2, \dots, t_i + \Delta t_i, \dots, t_p),$$

si l'origine est située en :

$$T_0 = (t_1, t_2, \dots, t_i, \dots, t_p).$$

La fonction S est calculée en ces $p+1$ points. On cherche la valeur maximale S_h et la valeur minimale S_b de ces $p+1$ valeurs. Un point "moyen" du simplexe \bar{T} est ensuite calculé en ignorant la valeur T_h (correspondant à S_h). C'est à partir de T_h que se fait l'exploration du sens de variation de la fonction et c'est également ce point qui sera modifié à chaque itération par trois opérations possibles ; la réflexion, la contraction ou l'expansion.

La première phase, ou réflexion, se fait en calculant un point T_* , situé sur l'axe $T_h \bar{T}$, par la relation :

$$T_* = (1 + n) \bar{T} - \eta T_h, \quad (21)$$

où η est le coefficient de réflexion, nécessairement positif.

Si S_* est inférieur à S_b , on se situe sur un axe qui peut conduire vers des valeurs plus faibles de S et on cherche donc à progresser dans cette direction par une expansion qui se calcule par l'expression :

$$T_{**} = \gamma T_* + (1 - \gamma) \bar{T}, \quad (22)$$

où γ est le coefficient d'expansion, nécessairement supérieur à 1, ce qui revient à prolonger le chemin qui a déjà été fait au cours de la réflexion. La valeur de la fonction en ce point T_{**} est notée S_{**} .

Si elle est toujours inférieure à la valeur minimale déjà trouvée dans le simplexe initial, le point T_{**} remplace le point T_h parmi les points de départ et le processus recommence. Si la valeur S_{**} est au contraire supérieure au minimum de départ, l'expansion a été trop forte et on remplace T_h par T_* , auquel correspond une valeur S_* plus faible.

Ayant envisagé le cas où, au cours de la réflexion, on a pris un chemin conduisant vers un minimum, il faut maintenant étudier le cas où S_* est supérieur à S_b . Si la valeur de la fonction calculée au point S_* reste proche du minimum et en tout cas inférieur aux autres S_i , on conserve le point T_* comme valeur de T_h pour recommencer le processus.

Dans le cas où S_* est supérieur aux S_i , on calcule un T_{**} qui va se situer entre T_h et \bar{T} , par l'opération dite de contraction, qui s'écrit :

$$T_{**} = \beta T_h + (1 - \beta) \bar{T}, \quad (23)$$

où β est le coefficient de contraction, compris entre 0 et 1. Si on a obtenu une valeur de S_* inférieure à S_h , l'opération de contraction se fait évidemment en remplaçant T_h par T_* .

La valeur S_{**} calculée en ce nouveau point peut être supérieure à la valeur S_h . Dans ce cas on va réduire le simplexe autour du point minimum en remplaçant tous les T_i par les valeurs : $(T_i + T_b)/2$, T_b caractérisant le point où la fonction vaut S_b . La valeur S_{**} peut être inférieure à la valeur maximum S_h , ce qui conduit à remplacer T_h par T_{**} , et à reprendre le processus. Le raisonnement qui vient d'être présenté est repris sous forme d'organigramme à la figure 1.

Le critère qui provoque la fin des calculs doit tenir compte du fait que les valeurs des paramètres en les différents points du simplexe sont pratiquement égales. On est alors arrivé à un point dit "stable" de la fonction. Concrètement, le critère peut être défini comme étant une différence relative minimale entre deux valeurs successives du même paramètre. On peut également calculer la variance des valeurs S_i de la fonction calculée aux différents points du simplexe et décider d'une valeur minimale de cette variance.

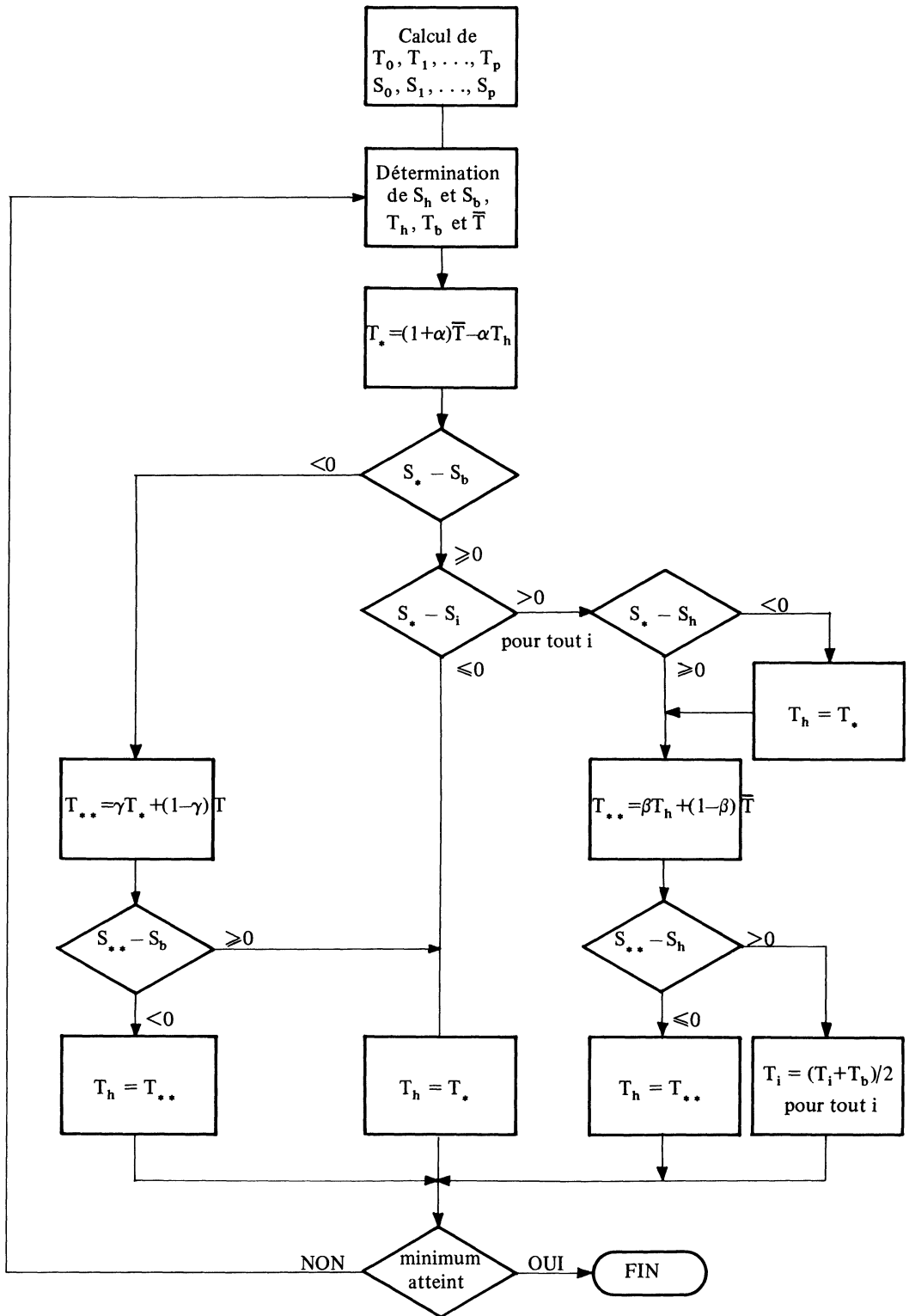


Figure 1 – Organigramme du processus d'optimisation

Pour que l'algorithme ici présenté soit complet, il faut encore attribuer les valeurs aux coefficients η , β et γ .

Il semble que, pour le coefficient de réflexion η , la valeur 1 ait conduit aux optimisations les plus rapides. Cela revient à calculer un point T_* sur l'axe défini par les deux points T_h et \bar{T} , tel que \bar{T} soit strictement à mi-chemin entre T_h et T_* .

La valeur 0,5 a été employée pour le paramètre de contraction, ce qui situe T_{**} à mi-chemin entre T_h et \bar{T} . Le paramètre d'extension est pris égal à 2, ce qui place le point T_{**} sur l'axe du minimum à une distance de \bar{T} égale à deux fois la distance (\bar{T}, T_h) .

L'obtention d'un vecteur de paramètres t correspondant à une variance résiduelle minimale ne résout pas complètement le problème posé. Il reste en effet à calculer la matrice de dispersion. Par la méthode de Gauss on peut obtenir la matrice des dérivées partielles du second ordre de la fonction, calculées par rapport aux paramètres t_i . Cette matrice appelée matrice d'information dans les problèmes à caractère statistique peut donner, après inversion et à une constante près, la matrice des variances et covariances ou matrice de dispersion (équation 20).

Ayant résolu le problème par la méthode du simplexe, on ne dispose pas de cette matrice de dérivées partielles, ces dernières n'étant pas calculées dans ce processus. On fait alors l'hypothèse que la fonction peut, en un point donné, être représentée par une expression quadratique de la forme :

$$u = a_0 + 2 \sum_i a_i z_i + \sum_{ij} b_{ij} z_i z_j, \quad (24)$$

ou, en notation vectorielle :

$$U = a_0 + 2 A' Z + Z' B Z, \quad (25)$$

où A est un vecteur-colonne de paramètres de dimensions $p \times 1$,

Z est le vecteur-colonne des variables de dimensions $p \times 1$,

B est une matrice de paramètres de dimensions $p \times p$.

Les valeurs notées z_i ou z_j sont les sommets d'un simplexe unitaire construit autour du point choisi comme minimum.

Le passage de ces valeurs z_i aux valeurs t_i calculées dans le dernier simplexe se fait par l'opération :

$$T = T_0 + Q Z, \quad (26)$$

où Q est une matrice de dimensions $p \times p$ dont les colonnes sont les vecteurs $T_i - T_0$.

Ces hypothèses étant acquises, on peut calculer la matrice d'information, c'est-à-dire la dérivée seconde de u au point choisi comme minimum. Il convient bien sûr d'exprimer ce résultat non pas en termes de Z mais en termes de T , qui représente les données de notre problème. Cette dérivée seconde s'écrit :

$$(Q^{-1})' B Q^{-1} \quad (27)$$

et la matrice des variances et covariances qui en découle :

$$Q B^{-1} Q'. \quad (28)$$

Il faut cependant rester prudent dans l'interprétation des résultats fournis par cette dernière expression, compte tenu du fait qu'elle suppose possible la représentation de la fonction sous forme quadratique à l'intérieur du dernier simplexe calculé. Il se peut très bien que les dimensions de celui-ci soient telles que l'assimilation à une forme quadratique entraîne un biais important. Le simplexe peut également être tel que les différences entre les valeurs de y soient inférieures ou du même ordre de grandeur que les erreurs d'arrondi. Dans ce dernier cas, il est souhaitable d'agrandir le simplexe final, par exemple en doublant ses dimensions et ce de façon tout à fait arbitraire.

6. CRITIQUE DE L'AJUSTEMENT

Après avoir calculé les meilleures valeurs des paramètres dans le cadre du modèle mathématique choisi, on peut examiner la correspondance existant entre le phénomène observé et la représentation théorique. Il se peut en effet que le modèle mathématique ne corresponde pas à la réalité ou que le processus d'ajustement n'ait pas conduit aux valeurs les plus adéquates des paramètres.

Un premier examen des résultats peut se faire très rapidement au vu du graphique représentant les points observés et la courbe théorique.

Une deuxième critique de l'ajustement peut se faire en examinant la distribution des résidus de régression. L'espérance mathématique de ces résidus étant nulle, leur moyenne doit normalement être proche de 0. Le calcul de leur distribution de fréquences et sa représentation en histogramme permettent de contrôler la disposition des valeurs résiduelles autour de la valeur nulle et éventuellement le caractère normal de leur distribution.

Les résidus supportent la partie aléatoire du phénomène étudié. On peut donc également contrôler ce caractère aléatoire par un test non paramétrique tel que le test des suites (run test) (Siegel, 1956).

La construction d'un test de signification de l'ajustement, basé sur une distribution théorique de probabilité, est moins immédiate que pour les régressions linéaires. Dans ce cas, en effet, pour autant que les résidus puissent être considérés comme indépendants entre eux et distribués normalement, la variance résiduelle est une variable aléatoire de distribution χ^2 , moyennant la transformation :

$$\chi^2 = \frac{(n - k) S^2}{\sigma^2}$$

où S^2 est la variance résiduelle observée,

n est la taille de l'échantillon,

σ^2 est la variance du processus,

k est le nombre de paramètres estimés.

Cette transformation donnera une variable χ^2 à $n-k$ degrés de liberté, étant donné l'existence des k relations linéaires et indépendantes qui ont permis l'estimation des k paramètres (Dagnelie, 1970).

Toujours dans le cas linéaire, une autre variable χ^2 peut être construite à partir de la part de variance expliquée par les k paramètres. On démontre qu'elle est indépendante de la variable χ^2 définie ci-dessus, ce qui permet la construction d'une variable F de Snedecor.

Dans le cas de modèles non linéaires, les relations liant les k paramètres ne sont pas nécessairement linéaires, ni linéairement indépendantes, ce qui rend plus difficile le calcul du nombre de degrés de liberté de la variable χ^2 .

L'erreur commise peut cependant être négligeable si la taille de l'échantillon est suffisante. Pour les mêmes raisons, l'emploi d'une variable F de Snedecor n'est pas mieux adapté au cas non linéaire. L'erreur commise en employant cette variable F ne serait cependant pas trop importante et porterait uniquement sur le niveau de probabilité, tout au moins pour les modèles dont la non-linéarité reste modérée (Beale, 1960 ; Draper et Smith, 1966).

Dans le cas où la non-linéarité du modèle est trop prononcée, il est possible de calculer une région de confiance pour des paramètres transformés qui linéarisent la fonction (Box et Coutie, 1956 ; Hartley, 1964).

7. UTILISATION PRATIQUE DES PROGRAMMES ECRITS POUR L'ORDINATEUR IBM 1130

Les programmes ont été écrits dans une optique modulaire, afin de réduire au maximum l'effort de programmation réclamé aux éventuels utilisateurs. Il leur suffit, en effet, d'identifier leur problème en écrivant un minimum d'instructions Fortran intégrées dans un sous-programme.

L'emploi de la méthode de Gauss nécessite l'écriture d'un sous-programme (sous-routine) explicitant le modèle mathématique choisi, de même que les expressions analytiques des dérivées partielles de la fonction par rapport à chacun des paramètres. L'emploi de la méthode utilisant un simplexe nécessite l'écriture d'un sous-programme (fonction) explicitant uniquement le modèle mathématique. Une vingtaine d'instructions arithmétiques en Fortran suffisent donc à intégrer n'importe quel modèle dans l'ensemble des programmes écrits.

Il convient également d'attirer l'attention des utilisateurs sur la nécessité d'avoir une connaissance approximative de la valeur des paramètres ou du moins de leur ordre de grandeur.

La qualité de ces valeurs initiales va, en effet, conditionner la vitesse de convergence vers la solution finale. De plus, si la somme des carrés des écarts résiduelle, qui est la quantité à minimiser, présente de nombreux minima locaux ou des points de stationarité, un mauvais départ risque de conduire plus souvent à des solutions inadéquates.

A titre d'exemples, les modèles suivants ont été traités avec succès :

1) la somme de plusieurs courbes de Gauss dont la base a été déplacée linéairement (type de problème rencontré en spectrométrie γ ou x) :

$$y = \sum_{i=1}^k \frac{A_i}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} + ax + b,$$

où y est un nombre d'impulsions,

x est une énergie en Kev,

k est le nombre de courbes considérées,

A est la surface de chacune des courbes,

σ_i est l'écart-type de chacune des distributions,

μ_i est la moyenne de chacune des distributions,

a et b sont les paramètres de la déformation linéaire ;

2) une combinaison de courbes de Gauss et de Cauchy (type de problème rencontré en spectrométrie infra-rouge) :

$$y = y_0 \{ f G + (1-f) L \} + c,$$

avec

$$L = \frac{1}{1 + \left[\frac{2(x-x_0)}{\Delta x_{1/2}} \right]^2}$$

$$G = \exp - \ln 2 \left\{ 1 + \frac{2 b(x-x_0)}{\Delta x_{1/2}} \right\} \frac{1}{b}$$

où y est un pourcentage de transmission,

x est une longueur d'onde,

y_0 est la hauteur du pic,

x_0 est la position du pic,

$\Delta x_{1/2}$ est la largeur du pic à mi-hauteur,

b est un paramètre qui rend la courbe de Gauss dissymétrique,

f est la proportion de la courbe de Gauss dans le profil général

c est le fond constant,

la figure 2 présente le résultat d'un tel ajustement ;

3) un modèle de croissance biologique selon une courbe logistique non symétrique :

$$y = \frac{M}{[1 + e^{-M(bX+a)/\theta}]^\theta}$$

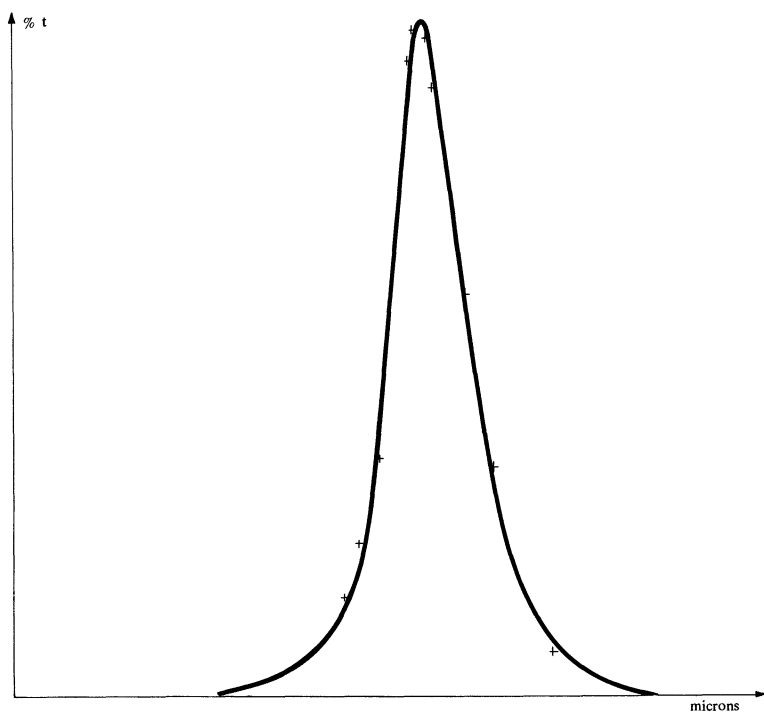


Figure 2 – Evolution du pourcentage de transmission en fonction de la longueur d'onde.

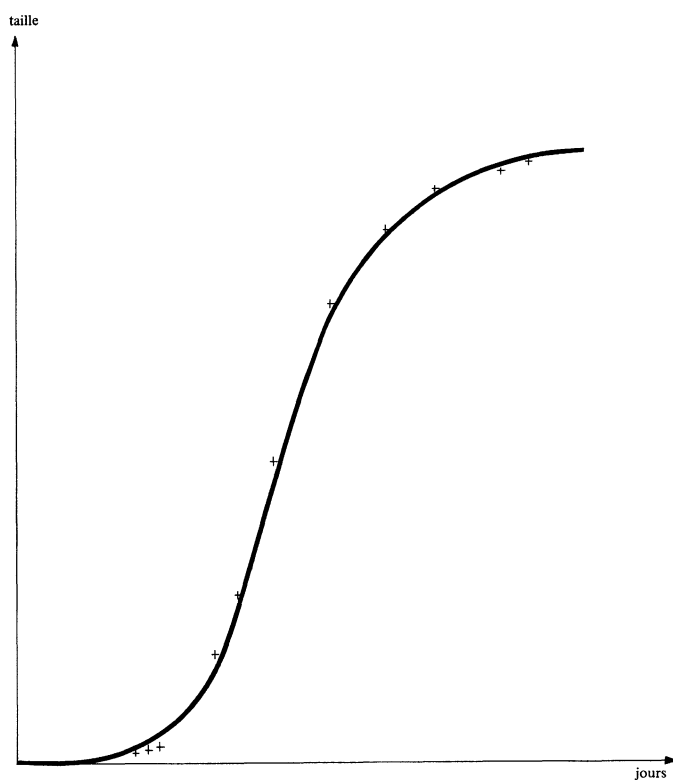


Figure 3 – Croissance de pommiers en fonction du temps.

où y est la taille de l'individu,

X est le temps,

M est la taille maximale,

b est un facteur caractérisant la vitesse de croissance,

a est un paramètre situant le début de la croissance,

θ est un paramètre mesurant la dissymétrie du phénomène,

la figure 3 présente cet ajustement ;

4) un modèle de renouvellement d'air dans une enceinte ventilée :

$$y = B \left\{ \frac{k}{\theta_1^{n_1} \Gamma(n_1)} x^{(n_1-1)} e^{-\frac{x}{\theta_1}} + g(x_0) \frac{1-k}{\theta_2^{n_2} \Gamma(n_2)} (x-x_0)^{n_2-1} e^{-\frac{(x-x_0)}{\theta_2}} \right\}$$

où y est la concentration en CO_2 ,

x est le temps,

B est la quantité totale de CO_2 injectée,

θ_1 et θ_2 situent la position des deux maxima,

n_1 et n_2 définissent la dispersion des deux composantes,

$g(x_0) = 0,5 \left[1 + \tanh \frac{x-x_0}{0,001} \right]$. cette expression étant destinés à annuler la

deuxième composante pour toutes les valeurs de x inférieures à x_0 ;

5) une série numérique, solution de l'équation différentielle de la conduction de la chaleur dans un cylindre fini (type de problème rencontré en conserverie) :

$$u(t) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{(2m-1) \pi j_{0,n} j_1(j_{0,n})} e^{-R_{n,m} \frac{at}{l^2}}$$

avec

$$R_{n,m} = \left(\frac{h}{d}\right)^2 j_{0,n}^2 + (2m-1)^2 \frac{\pi^2}{4}$$

où $h = 2l$, soit la hauteur du cylindre,

d est le diamètre du cylindre,

l est la longueur considérée,

j_1 est la fonction de Bessel d'ordre un,

$j_{0,n}$ est la nième racine de la fonction de Bessel d'ordre 0.

Les "listings" de tous les programmes employés sont disponibles à la Chaire de Statistique de la Faculté des Sciences agronomiques de l'Etat à Gembloux (Belgique).

8. DISCUSSION ET CONCLUSION

Il convient maintenant de discuter des difficultés d'utilisation, avantages et inconvénients des deux méthodes présentées.

En ce qui concerne la méthode de Gauss, il est indispensable de s'assurer que la fonction à ajuster et ses dérivées partielles restent bornées en tous les points durant le processus de minimisation. D'autre part, on doit être sûr qu'aucune des dérivées partielles n'est nulle en *chacun* des n points considérés. Dans le cas contraire, en effet, une ligne et une colonne de la matrice $D' W D$ seraient identiquement nulles et celle-ci serait singulière.

La méthode peut également présenter un défaut de convergence, suite à un mauvais choix des approximations de départ. Malheureusement aucune règle n'existe ici : dans certains cas, le processus peut converger alors que l'approximation initiale s'écarte de la meilleure estimation dans un rapport pouvant aller jusqu'à 10^5 ; dans d'autres cas, un écart de quelques dizaines de % peut suffire à mettre le processus en défaut.

Pour ce qui est de la méthode utilisant un simplexe, il n'y a pas de restriction quant au caractère différentiable de la fonction à optimiser, ni quant à la continuité et à l'indépendance des dérivées partielles. Cette méthode présente aussi l'avantage de ne pas exiger le calcul des dérivées partielles du modèle mathématique choisi.

Il semblerait également que cette méthode puisse converger à partir de valeurs initiales des paramètres plus éloignées des valeurs réelles que dans le cas de la méthode de Gauss, mais, de nouveau, aucune règle n'existe pour apprécier a priori le succès de la convergence.

Au passif de cette seconde méthode, il faut toutefois signaler la lenteur du processus d'ajustement. De plus, l'estimation de la matrice des variances et covariances des paramètres obtenus est imprécise et sujette à un biais important.

Il ressort des différents essais d'ajustements que nous avons été amenés à réaliser que la solution la plus efficace consiste à scinder l'ajustement en deux étapes.

La première étape, destinée à améliorer les estimations initiales des paramètres consiste en l'application de la méthode du simplexe. La seconde étape continue l'ajustement au moyen de la méthode de Gauss afin de trouver plus rapidement la valeur minimale de la variance résiduelle et une estimation plus exacte de la matrice des variances et covariances des paramètres.

Il ne semble pas exister de critère objectif situant le passage le plus opportun de la première étape à la seconde. On peut le décider arbitrairement lorsque la variance expliquée a atteint un certain pourcentage de la variance totale. Si ce passage a eu lieu trop tôt, les valeurs des paramètres obtenus à ce moment

peuvent encore provoquer la divergence de l'ajustement par la méthode de Gauss et il faut revenir à la première méthode.

Cette combinaison des deux méthodes met en évidence l'avantage de pouvoir disposer d'un système de gestion intégrée des variables observées, tel que celui qui a été mis au point à la Chaire de statistique de la Faculté des Sciences agronomiques de Gembloux (Cartelli et al., 1973). Ce système permet de ne pas réintroduire les données lors du passage d'une étape à l'autre, même si les programmes employés sont différents pour chaque méthode et ont été écrits indépendamment l'un de l'autre.

BIBLIOGRAPHIE

- BEALE E.M.L. Confidence regions in non-linear estimation. *J. Roy. Statist. Soc., Ser. B*, 22, 41–88, 1960.
- BOX G.E.P. et COUTIE G.A. Application of digital computers in the exploration of functional relationships. *Proc. I.E.E.* 103. *Part B*, Supp. 1, 100-107, 1956.
- CARLETTI G., CLAUSTRIAUX J.J., DAGNELIE P., DEBOUCHE C., IN K., OGER R. et ROUSSEAU G. Organisation d'une bibliothèque de programmes statistiques pour ordinateur. *Revue belge de Statistique, d'informatique et de Recherche opérationnelle* 12, 4, 2-17, 1973.
- DAGNELIE P. La régression multiple. *Biom. Praxim.* 7, 193-238, 1966.
- DAGNELIE P. *Théorie et méthodes statistiques : applications agronomiques* (2 vol.). Presses Agron., Gembloux, 378 + 451 p., 1969-1970.
- DRAPER N.R. et SMITH H. *Applied regression analysis*. Wiley, New-York, 407 p., 1967.
- HALD A. *Statistical tables and formulas*. Wiley, New-York, 44–46, 1952 .
- HARTLEY H.O. Exact confidence regions for the parameters in non-linear regression laws. *Biometrika* 51, 347, 1964.
- KENDALL M.G. et STUART A. *The advanced theory of statistics*. Vol. 2, Griffin, London, 676 p., 1961.
- NELDER J.A. et MEAD R. A simplex method for function minimization. *Computer J.* 7, 308–313, 1965.
- SIEGEL S. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New-York, 312 p., 1956.
- VIGNE J. *Etude et mise en oeuvre d'algorithmes de recherche d'un extremum d'une fonction de plusieurs variables* (Thèse). Faculté des Sciences de l'Université de Paris, 240 p., 1969.