

Un menu « avec suppléments » qui ne fera pas chuter la courbe de vos revenus.



Ajustements et tracés de courbes sur ordinateur

« personnel »

Modélisateur et analyste rencontrent des problèmes d'ajustement de courbes, ou « fittage », dans des domaines aussi variés que la physique, l'économie, la physiologie, la démographie, etc. Ces spécialistes demanderont ensuite, tout naturellement, à visualiser le résultat de leurs calculs et à en conserver la trace.

Les procédures réunies dans le logiciel donné ici répondent à leurs besoins (fig. 1).

Introduction

Rappelons tout d'abord le problème posé par un ajustement. Il s'agit, étant donné un ensemble de points dans R^n , de déterminer m paramètres d'une fonction de n variables tels que l'écart entre les p points et la fonction soit minimal. Cette fonction peut être imposée par une loi théorique, issue d'une hypothèse, ou bien encore choisie *a priori*.

Différents algorithmes permettent d'obtenir ce résultat. Celui proposé par *Nelder et Mead* (1965) offre des avantages développés dans le second chapitre. Le lecteur trouvera dans le troisième chapitre comment nous avons résolu les problèmes particuliers rencontrés tant au niveau du « hard », comme les standards RS 232, qu'au niveau du « soft ». Nous terminons enfin en donnant un exemple d'application. Mais commençons tout de suite par une présentation d'ensemble du logiciel.

Présentation de l'ensemble des menus

Le menu principal est structuré comme suit.

| GRAPHIX : | |
|-----------------------------|-----|
| Pointé d'un fichier | → 1 |
| Réalisation d'un ajustement | → 2 |
| Tracé d'une fonction | → 3 |
| Fin du travail | → 4 |

La première option de ce menu, « Pointé d'un fichier » permet de charger un fichier de points et de les tracer. Ces fichiers de points peuvent être sous 2 formats différents :

— fichiers standards : ce sont des fi-

chiers de données ne contenant que des coordonnées de points ;

— fichiers « *Nelder-Mead* » : ce sont les fichiers de sortie du programme d'ajustement. On peut les utiliser pour le pointé, bien qu'ils contiennent un grand nombre d'informations inutiles. (Valeur des paramètres, moyenne des carrés des résidus, etc.).

La seconde option permet de réaliser un ajustement selon l'algorithme « *Nelder-Mead* ». Cette option utilise un fichier de points expérimentaux, Il faudra ensuite saisir la fonction de départ ainsi que les valeurs initiales des paramètres.

La troisième option de ce menu est consacrée au tracé : elle permet de saisir une fonction, et éventuellement de charger d'un fichier sur disque les paramètres calculés lors de l'ajustement.

La dernière option de ce menu principal permet de quitter le programme, éventuellement en retournant au système.

Principe

de la méthode d'ajustement

Rappelons tout d'abord les nombreux avantages qu'offre l'algorithme de *Nelder et Mead* par rapport aux autres algorithmes disponibles.

Tout d'abord, cette méthode d'ajustement ne nécessite pas la connaissance des dérivées partielles de la fonction, par conséquent elle évite à l'utilisateur des calculs fastidieux que d'autres algorithmes imposent avant chaque ajustement.

La fonction cherche son minimum absolu, assurant la convergence de la fonction vers un extremum.

De plus, la qualité des paramètres calculés par cette méthode est excellente, et supporte parfaitement la comparaison avec les logiciels fonctionnant sur de très gros systèmes, comme S.A.S., le standard en la matière.

Enfin, si on peut reprocher à *Nelder-Mead* sa relative lenteur, elle est compensée par la sécurité qu'offre la bonne qualité de ses calculs.

Le problème est de trouver les valeurs de m paramètres, d'un modèle $F_p(x)$ tel que la courbe représentant le modèle soit la plus proche possible de la courbe expérimentale $Y_j = F(x_j)$. Le critère de « proximité » entre les 2 courbes peut être apprécié de plusieurs façons. Celle utilisée est la valeur d'une fonction du type :

$$F_p = \sum_{j=1}^p (f_p(x_j) - y_j)^2$$

avec :

- n = nombre de variables,
- m = nombre de paramètres,
- p = nombre de points.

Elle sera progressivement minimisée et ses variations sont suivies au fur et à mesure que la fonction approche de son minimum (fig. 2).

Plusieurs étapes sont nécessaires :

– Etape « 0 » : initialisation et création de la matrice de paramètres.

– Etape « 1 » : la fonction F est considérée comme une fonction à m variables indépendantes (les variables sont les paramètres). Elle est évaluée à partir de p points, $m + 1$ fois dans l'espace des variables. Cet ensemble s'appelle le **simplex**, et les $m + 1$ paramètres sont les sommets d'un polygone.

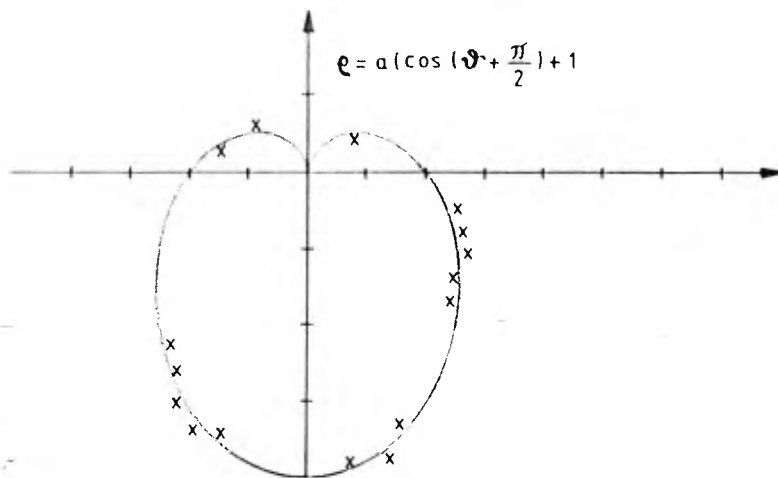


Fig. 1

– Etape « 2 » : les valeurs de F étant évaluées à tous les sommets du polygone, le sommet le plus haut : Ph , où F accuse la valeur maximale : Fh , est réfléchi à travers le centroïde P (point moyen) des autres sommets afin d'obtenir un nouveau point d'essai « P^* » dont les coordonnées sont définies par :

$$P^* = \bar{P} + \alpha(\bar{P} - Ph)$$

où le point moyen est défini par :

$$\bar{P} = \frac{\sum_{i=1}^{m+1} (P_i) - Ph}{m}$$

– Etape « 3 » : si la valeur de F^* en P^* correspond à un nouveau minimum, il est possible de passer directement à l'étape 4.

Si ce n'est pas le cas, et si la comparaison de F^* avec toutes les autres valeurs de F (sauf Fh) montre que F^* ne correspond pas à la plus grande valeur de F , Ph est remplacé par P^* et Fh par F^* . Il est alors possible de passer directement à l'étape 6.

Si la comparaison donne F^* égal à Fh , il faut alors remplacer Ph par P^* et se rendre à l'étape 5.

Si $F^* > Fh$, se rendre à l'étape 5.

– Etape « 4 » : un nouveau point d'essai P^{**} est alors défini de la façon suivante :

$$P^{**} = \bar{P} + \gamma(P^* - \bar{P})$$

Si à F^{**} correspond un nouveau minimum, Ph est remplacé par P^{**} et Fh par F^{**} et les calculs se continuent dans l'étape 6. Le cas échéant, Ph est remplacé par P^* avant de passer à la phase suivante

– Etape « 5 » : un autre point d'essai P^{**} est défini par la contraction du simplex :

$$P^{**} = \bar{P} - \beta(\bar{P} - Ph)$$

Si F^{**} est plus petit que Fh , Ph est remplacé par P^{**} et Fh par F^{**} .

Si F^{**} est plus grand que Fh , la contraction du simplex n'est pas bonne. Dans ce cas, tous les points sont alors ramenés vers le point Pe auquel correspond la valeur minimale de F par la relation :

$$Pe = \frac{P_i + P_b}{2}$$

où P_b est le sommet et où F est la valeur minimale.

– Etape « 6 » : il faut vérifier que le minimum de la fonction F soit atteint. Pour cela, le critère utilisé est la comparaison de l'écart-type des valeurs F à une valeur prédéterminée.

L'écart type est défini par la relation suivante :

$$SD^2 = \frac{\sum_{i=1}^{n+1} (F_i - \bar{F})^2}{n + 1}$$

où \bar{F} = moyenne des valeurs de F :

$$\bar{F} = \frac{\sum_{i=1}^{n+1} F_i}{n + 1}$$

Si SD^2 est supérieur à la valeur prédéterminée, les calculs sont à recommencer depuis l'étape 2.

Sinon la convergence est atteinte et le programme passe à la dernière étape d'impression des résultats.

Ce logiciel est prévu pour fonctionner sur un IBM PC, ou compatible, mais a été développé sur IBM PC-AT dont la vitesse d'exécution permet une utilisation plus souple.

Par conséquent, il est implanté sous le PC-DOS version 3.10, et est écrit sous Basica version 3.10. Cependant, il peut passer sans problème sur n'importe quelle machine de la série PC.

Le traceur, HP-7475A, est un plotter très classique, à 6 couleurs de feutres, et qui accepte les formats de papier A4 et A3. Il dispose d'un jeu d'instructions important qui permettent de simplifier la plupart des tracés. Les instructions sont de la forme : mnémorique, paramètre, terminateur.

Choix du langage utilisé

Parmi tous les langages à notre disposition, nous n'avons retenu que Pascal et Basic pour les facilités de mise au point qu'ils offraient.

Cependant, le Pascal (Microsoft ou Turbo) ne s'est pas avéré satisfaisant pour l'interfaçage avec le traceur, et nous avons dû nous tourner vers Basic.

Malgré tous les défauts connus du Basic, comme sa lenteur ou son manque de structure, il nous est apparu comme étant le seul langage apte à piloter facilement un périphérique.

La syntaxe est simple : OPEN COM1: et arguments. Pour une émission vers le traceur : PRINT £3, instruction traceur. Pour la réception de données : INPUT A\$, n, £3.

Problèmes rencontrés

En premier lieu, la liaison avec le traceur s'est avérée épineuse dans la mesure où le port RS 232 du PC-AT ne répond pas exactement à la norme. Emission et réception se déroulent à peu près normalement, mais il est impossible de se fier au « handshaking » matériel car le brochage sur le PC-AT n'est pas standard.

Par conséquent, les messages émis par le traceur, comme « Données mal reçues » ou « Buffer plein », ne sont pas reçus correctement, et il s'ensuit une perte de données.

Cet inconvénient est particulièrement grave, car le buffer du traceur ne contient que 1024 octets, alors que l'ordinateur le remplit en moins d'une seconde. Bien entendu, le traceur ne peut dessiner à l'allure où il

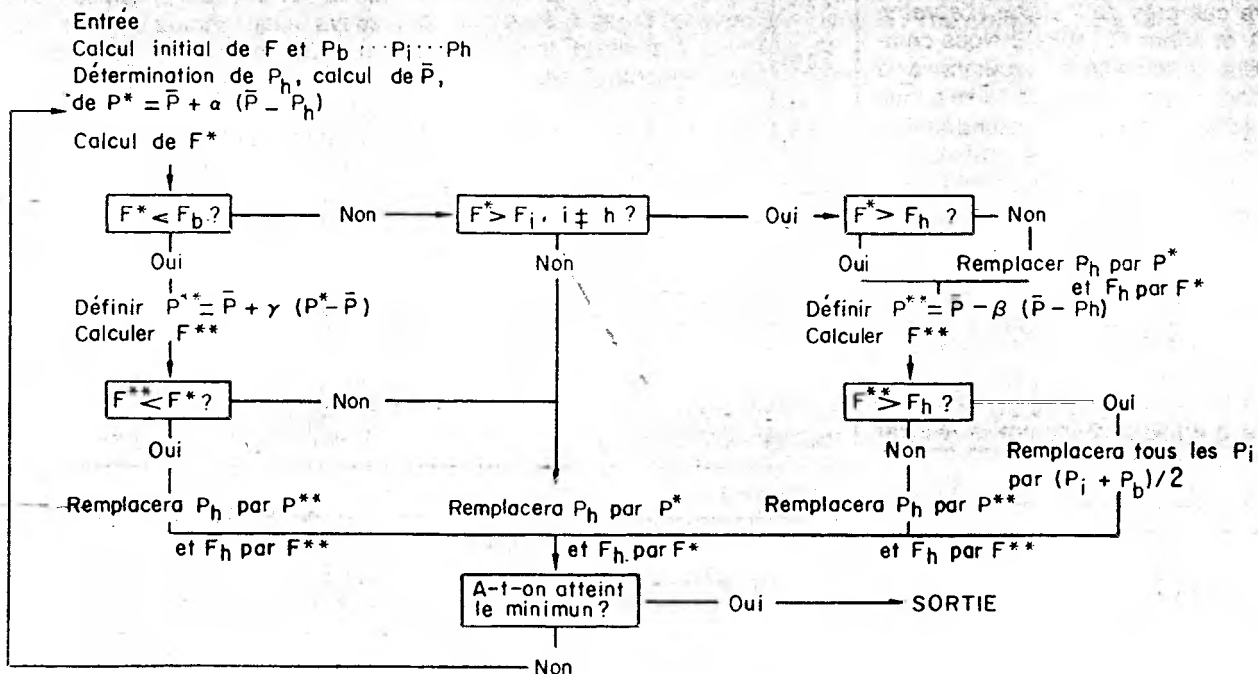
reçoit les données, et il faut programmer un driver artificiel de temporisation, qui teste l'état de remplissage du buffer avant toute émission. Cette méthode permet de remplacer les signaux de ligne, qui seraient sensés remplir le même office, mais que le PC-AT ne sait gérer correctement.

Un second problème délicat est celui du format des données. Le PC-AT émet bien entendu les chiffres sous le format réel, avec mantisse et exposant, alors que le traceur ne reconnaît que le format décimal, avec 4 chiffres significatifs.

Il faut donc utiliser les « Print Using » avec un format identique à celui du traceur. Prévoir également un agrandissement dans le cas où les données traitées nécessitent une précision supérieure à 4 chiffres significatifs, ou une diminution lorsque les données dépassent les limites du traceur (32767).

Exemple d'application

L'exemple d'application suivant fait partie d'une série d'expériences menées, dans un verger près de Montpellier, par G. Selles sous la direction de A. Berger, directeur de recherche au CNRS. Il correspond plus particu-



Organigramme du programme Nelder-Mead (1965)

lièrement à la récupération hydrique d'un plan de pêcher, cultivé en pot, après irrigation. L'instant $t = 0$ correspond à l'instant d'irrigation. La durée de l'irrigation est d'une minute seulement.

Il s'agit d'introduire, dans un modèle du fonctionnement hydrique du pêcher, la cinétique de récupération du diamètre du tronc $\Phi = f(t)$ ainsi que la vitesse de variation $d\Phi/dt$ de celui-ci. Un capteur de déplacement disposé sur le tronc de l'arbre fruitier effectue cette mesure et une centrale d'acquisition de données enregistre une centaine de valeurs par jour. Nous nous trouvons donc en présence d'un ensemble de couples (Φ expérimental, t), représentés en coordonnées cartésiennes sur le **graphique 1** ; le temps t étant exprimé en heures décimales et Φ en microns.

Nous voyons apparaître la forme d'une sigmoïde. Cette forme en « sigma » est assez fréquente dans les phénomènes de croissance. Elle traduit une croissance lente au début, suivie d'une croissance rapide (phase réversible liée à la récupération hydrique), pour terminer à nouveau par une phase de croissance lente. Après quelques essais d'ajustements de sigmoïdes : modèle logistique et modèle de *Gompertz* en particulier, c'est la logistique généralisée passant par l'origine qui a donné le meilleur résultat, le point d'inflexion se situant dans notre cas près de l'origine (*Lebreton J.-D. et Millier C. (1982)*). Nous complétons la fonction en ajoutant, à la logistique généralisée, de la cinétique de récupération, une composante « croissance » de forme linéaire. On obtient ainsi la fonction définitive suivante :

$$\Phi = f(t) = \left(\frac{P1}{P2}\right)^{1/(1-P3)} \cdot (1 - \text{Exp}((P3 - 1) P2 \cdot t))^{1/(1-P3)} + P4 \cdot t \quad (1)$$

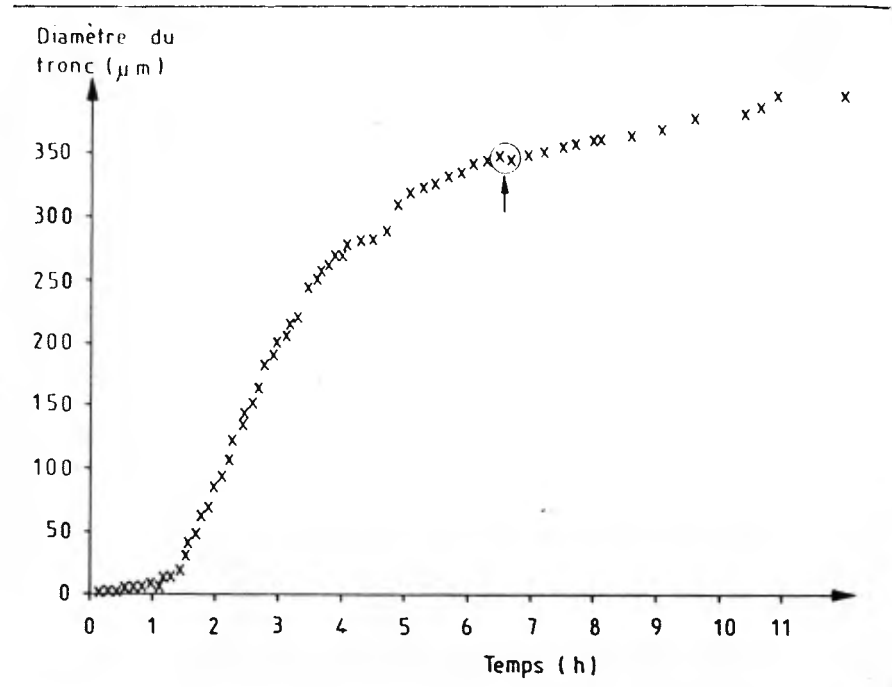
Le **graphique 2** montre le résultat de cet ajustement.

La vitesse de croissance, dérivée de cette fonction, a pour expression :

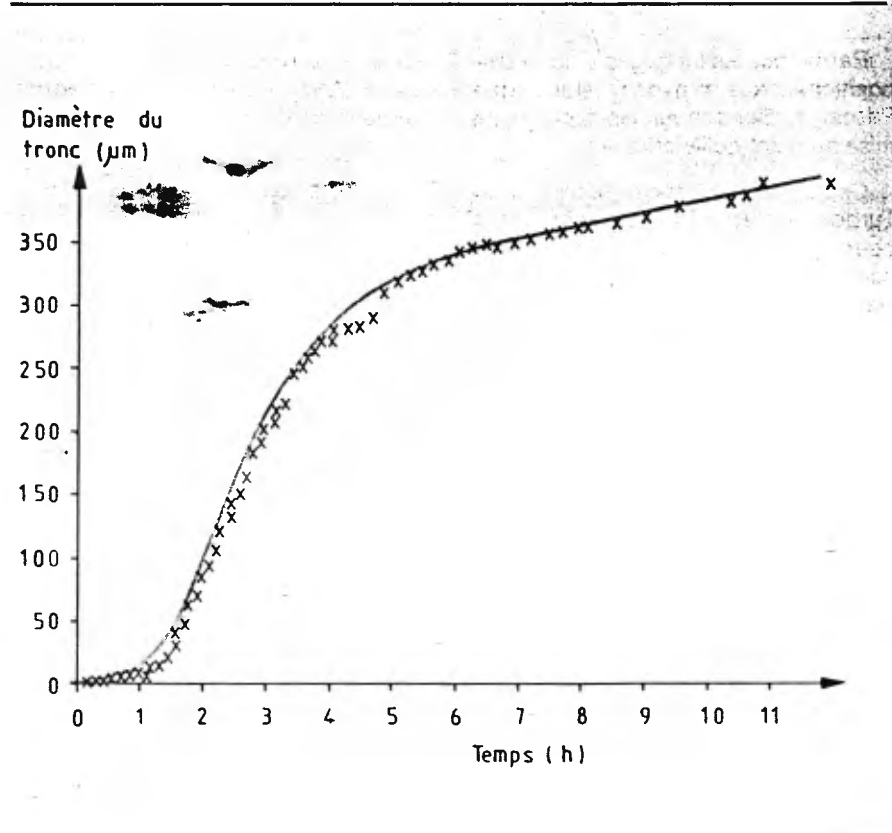
$$\frac{d\Phi}{dt} = P2 \left(\frac{P1}{P2}\right)^{1/(1-P3)} \cdot (1 - \text{Exp}((P3 - 1) P2 \cdot t))^{P3/(1-P3)} \cdot \text{Exp}((P3 - 1) P2 \cdot t) + P4 \quad (2)$$

Remarquons que le calcul de cette dérivée à partir des accroissements finis

$$\Delta\Phi_{\text{expérimentaux}}/\Delta t$$



Graphique 1



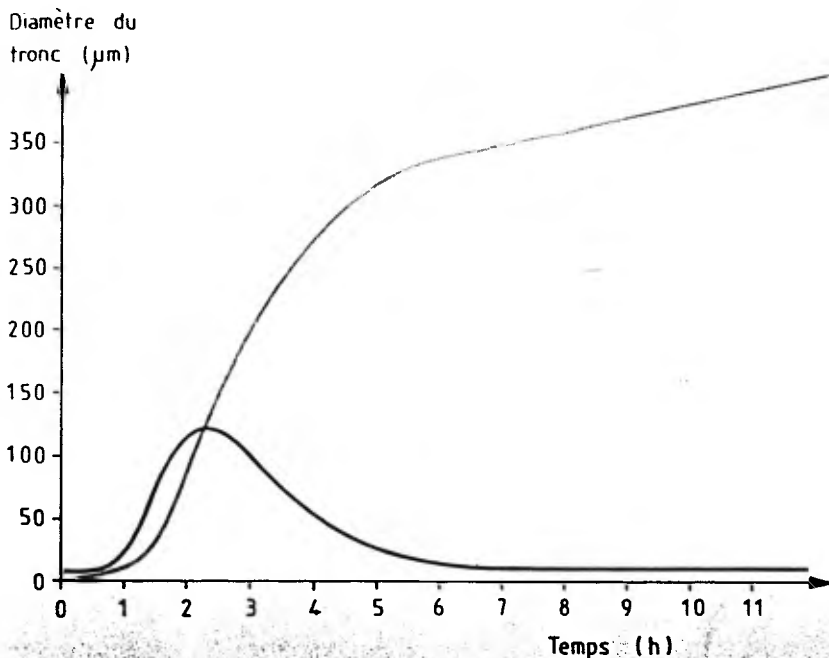
Graphique 2

(sur le **graphique 1**) amplifierait considérablement les erreurs. On trouverait alors une très grande variation entre deux valeurs successives de la dérivée, variation pouvant même aller jusqu'à une inversion de signe dans quelques cas, par exemple pour les deux points cerclés. Il faudrait donc, si ce calcul était retenu, faire appel à d'autres procédés mathématiques de type

moyenne mobile. La méthode que nous avons utilisée : ajuster la fonction, puis calculer sa dérivée, supprime cet inconvénient.

Le **graphique 3** montre la fonction Φ exprimée en μm et sa Φ' en μmh^{-1} .

L'algorithme proposé par *Nelder et Mead* demande une estimation des paramètres au départ. Dans le cas de



Graphique 3

la logistique généralisée, P3 est fonction de la position du point d'inflexion, et

$$\left(\frac{P1}{P2}\right)^{1/(1-P3)}$$

est l'asymptote horizontale ; P4, pente de la croissance, est le plus facile à estimer. Pour ajuster cette fonction aux 76 points, nous avons commencé avec l'estimation du quadruplet (P1, P2, P3, P4) suivante : (20 ; 2 ; 0,7 ; 6). La convergence a été obtenue au bout de 230 itérations en 15 mn en Basic interprété et a donné le résultat suivant (18,44 ; 10,87 ; 0,906 ; 9,1). Le temps d'exécution du tracé du graphique 3 a été de 2,5 mn avec un pas de 0,3 mm.

Sans cet exemple d'application, aucune preuve particulière ne nous permet d'affirmer que la loi de variation du diamètre du tronc au cours d'une journée suit la fonction (1). Mais

nous remarquons que cette fonction décrit très bien le phénomène. Le modélisateur pourra donc avantageusement introduire dans son modèle mathématique cette fonction continue de lissage à la place de ses points expérimentaux. Ceci ne diminuera pas la précision des résultats, puisque les écarts entre les valeurs observées et celles données par la fonction sont le plus souvent inférieurs aux erreurs de mesure. Les avantages sont par contre importants : facilité de mise en équation et de programmation, continuité dans le calcul qui donne une plus grande souplesse dans le fonctionnement du modèle et diminue les risques de « pompage ».

Nous pourrions ainsi donner de nombreux exemples d'applications. Le modèle a servi tout récemment pour des recherches conduisant à une thèse d'Etat, par Gérard Ferrière (1986), qui a nécessité une version

mise au point par *Patricio Soto*, version fonctionnant sur *Apple II*, et comportant une sortie sur imprimante graphique.

J.-B. Puel, P. Soto et J.-L. Salager
CNRS, Centre Louis-Emberger
Montpellier

Remerciements

Les auteurs tiennent à exprimer ici leurs remerciements à M. *Gérard Banco* pour sa contribution à la résolution des problèmes « hardware », à MM. *A. Berger* et *J. Roy* pour avoir bien voulu lire les parties de cet article les concernant et avoir apporté leurs corrections, et à M. *F. Sorrentino* pour avoir fourni une première version (TRS 80) de l'algorithme *Nelder-Mead*.

Bibliographie

- [1] *Ferrière G.*, 1986 : Mouvements naturels des éléments dans une prairie : quantification des échanges d'azote entre lombriciens, sol et plantes. Thèse d'Etat, *Université Claude Bernard*, Lyon I.
- [2] *Lebreton J.D., Millier C.* 1982 : Modèles dynamiques déterministes en biologie.
- [3] *Nelder J.A., Mead R.* 1965 : A simplex method for function minimisation. In : *The Computer Journal* 7.
- [4] *Selles G.*, 1985 : Réponse de quelques paramètres physiologiques à la contrainte hydrique. DAA - *ENSAM Montpellier*.

Une copie du listing complet du programme mis au point par les auteurs est disponible auprès de la rédaction d'*Electronique Applications*, moyennant l'envoi de 5 timbres à 2,20 F.

ELECTRONIQUE APPLICATIONS

est une publication de la Société Parisienne d'Édition.

Société Anonyme au capital de 1 950 000,00 F

Siège social

43, rue de Dunkerque, 75480 Paris Cedex 10

Création : 1909

Durée : 140 ans

Président-Directeur Général

Directeur de la Publication :

J.P. VENTILLARD

Rédacteur en Chef :

Jean-Claude ROUSSEZ

Actionnaires :

Publications Radioélectriques et Scientifiques

Monsieur J.-P. Ventillard

Madame Paule Ventillard

Tirage moyen 1984 :

53 793

Diffusion moyenne 1984 :

29 404

Chiffre d'Affaires 1984 de la Société Parisienne d'Édition :

92 863 848,00 F