

Ajustement d'une hyperbole équilatère

D. LEY

Ingénieur Agronome ENSA (Rennes)
Diplômé d'Agriculture Tropicale de l'ESAT

RÉSUMÉ

Après examen des difficultés liées à l'utilisation de la méthode des moindres carrés dans le cas de l'ajustement d'une hyperbole équilatère, la méthode générale de GAUSS-NEWTON et la modification de HARTLEY sont exposées. Une amélioration par élimination des paramètres linéaires, due à LAWTON et SYLVESTRE, est ensuite examinée.

La combinaison des deux méthodes appliquée à l'estimation des paramètres de l'hyperbole conduit enfin à l'établissement des formules dans ce cas particulier. L'organisation du calcul est précisée en annexe.

ABSTRACT

After examining the difficulties linked to the use of the method of least squares in the adjustment of an equilateral hyperbola, the general method of GAUSS-NEWTON and the modification of HARLY are set forth. A study of an amelioration by linear parameters elimination due to LAWTON and SYLVESTRE follows.

The combination of the two methods applied to the estimation of the hyperbola parameters leads finally to the calculation of formulas relating to this particular case. The calculation is given in detail in the supplement.

Le présent article présente une méthode utilisée pour l'ajustement d'une fonction de régression non-linéaire, en l'occurrence une hyperbole. C'est le résultat d'un travail effectué à l'ORSTOM pour P. FRANQUIN¹ sous la direction de J. DÉJARDIN².

1. INTRODUCTION

1.1. EXPRESSION DE LA FONCTION DE RÉGRESSION

La relation existant entre la photopériode et la température suggère, dans le déterminisme de la flo-

¹ Section d'Agronomie — ORSTOM — 70-74, route d'Aulnay — 93140 Bondy.

² Service de Biométrie — ORSTOM — 70-74, route d'Aulnay — 93140 Bondy.

raison des plantes, une fonction de régression hyperbolique.

D'une manière générale, l'équation d'une hyperbole est :

$$y = (Ax + B) / (Cx + D) \quad (1)$$

Cette équation a quatre paramètres; mais elle peut se ramener à une équation à trois paramètres. En mettant A/C en facteur dans (1), il vient :

$$y = A/C [(x + B/A) / (x + D/C)] \quad (2)$$

Posons alors $B/A = D/C + K$. (2) s'écrit alors :

$$y = A/C \{[(x + D/C) + K] / (x + D/C)\} \quad (3)$$

Posons encore :

$$\begin{aligned} A/C &= a \\ D/C &= c \\ K.A/C &= b \end{aligned}$$

Avec ces notations, (3) s'écrit enfin :

$$y = a + b/(x + c) \quad (4)$$

Cette expression de l'hyperbole ne possède plus que trois paramètres; nous verrons plus loin que cette formulation est beaucoup plus intéressante, et c'est sur ce modèle que nous ferons notre ajustement.

1.2. MODÈLE D'AJUSTEMENT

Soient :

— x , une variable non aléatoire, dite indépendante
— y , une variable aléatoire, dite dépendante, dont la valeur moyenne est fonction de x .

Soit $y = f(x; \theta)$ cette fonction où nous notons θ l'ensemble des w paramètres θ_u ($u = 1, 2, \dots, w$) dont nous nous proposons de déterminer la valeur.

Pour $x = x_i$, $i = 1, 2, \dots, n$, une valeur observée y_i diffère d'une certaine quantité ε_i de la valeur théorique $\bar{y}_i = f(x_i; \theta)$. On peut donc écrire :

$$y_i = \bar{y}_i + \varepsilon_i = f(x_i; \theta) + \varepsilon_i \quad (5)$$

ε_i est une fluctuation aléatoire.

D'une manière générale, pour déterminer la courbe (C) d'équation $y = f(x; \theta)$ passant au plus près des n points (x_i, y_i) observés, il faut choisir les valeurs des w paramètres θ_u , $u = 1, 2, \dots, w$ de $f(x; \theta)$ tels que l'ensemble des écarts $y_i - \bar{y}_i = \varepsilon_i$ soit minimum.

En raison des difficultés de calcul sur les valeurs absolues, on préfère minimiser la somme des carrés

des écarts $Q = \sum_i (y_i - \hat{y}_i)^2$ plutôt que de minimiser

la somme des valeurs absolues des écarts :

$$\Omega = \sum_i |\varepsilon_i| = \sum_i |y_i - \hat{y}_i|$$

Par ailleurs, le critère des moindres carrés présente un autre intérêt.

En effet, si les écarts $\varepsilon_i = y_i - \hat{y}_i$ sont gaussiens et de même écart-type σ pour tout x_i , la vraisemblance d'un ε_i est proportionnelle à $\exp(-\varepsilon_i^2/2\sigma^2)$. Donc, la vraisemblance de l'ensemble des n écarts ε_i peut s'écrire, à condition que les ε_i soient indépendants en probabilité :

$$\prod_i \exp(-\varepsilon_i^2/2\sigma^2) = \exp[-\sum_i \varepsilon_i^2/2\sigma^2], i = 1, 2, \dots, n$$

Cette vraisemblance est maximale quand

$$\sum_i \varepsilon_i^2 = \sum_i (y_i - \hat{y}_i)^2$$

est minimal.

Dans ces conditions, c'est-à-dire, si les écarts ε_i :

a) sont normalement distribués :

$$\mathcal{L}(\varepsilon_i) = \mathcal{N}(0, \sigma^2)$$

b) ne sont pas corrélés :

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i, \forall j, i \neq j$$

la méthode des moindres carrés s'identifie à la méthode du maximum de vraisemblance.

Ces hypothèses sont nécessaires si l'on fait des tests. Dans le cas contraire, elles ne sont pas indispensables : alors les deux méthodes ne s'identifient plus, mais la méthode des moindres carrés n'en constitue pas moins en soi un critère d'ajustement.

L'hypothèse d'homoscédasticité est néanmoins nécessaire aux régressions non pondérées auxquelles nous nous limiterons ici.

1.3. APPLICATION DE LA MÉTHODE DES MOINDRES CARRÉS A LA RÉGRESSION HYPERBOLIQUE

Les estimateurs des moindres carrés sont ceux minimisant la somme des carrés des écarts :

$$Q = \sum_i \varepsilon_i^2 = \sum_i [y_i - a - b/(x_i + c)]^2$$

x_i et y_i étant des observations, donc fixés, Q est fonction uniquement des paramètres a, b , et c .

Soit \hat{a}, \hat{b} , et \hat{c} les estimations des moindres carrés de a, b , et c . Ces estimations seront les solutions des équations normales, obtenues en annulant les dérivées partielles de $Q(a, b, c)$ par rapport à a, b et c .

Ces équations normales sont :

$$\partial Q / \partial a = 0 \Leftrightarrow -2 \sum_i [y_i - \hat{a} - \hat{b}/(x_i + \hat{c})] = 0$$

$$\partial Q / \partial b = 0 \Leftrightarrow -2 \sum_i [y_i - \hat{a} - \hat{b}/(x_i + \hat{c})] / (x_i + \hat{c}) = 0 \quad (7)$$

$$\partial Q / \partial c = 0 \Leftrightarrow 2 \sum_i [y_i - \hat{a} - \hat{b}/(x_i + \hat{c})] / (x_i + \hat{c})^2 = 0$$

avec $i = 1, 2, \dots, n$.

Résoudre le système d'équation (7) n'est pas simple. La difficulté d'application de la méthode des moindres carrés au cas non linéaire provient du fait que les dérivées ne sont pas indépendantes des paramètres comme dans le cas linéaire, et que, par conséquent, le système d'équations normales n'est pas un système linéaire. Une autre méthode doit être utilisée.

Dans la première partie de cet article, nous allons exposer une méthode générale d'ajustement applicable aux fonctions de régression non linéaires. Puis, dans la deuxième partie, nous verrons une méthode complémentaire, applicable dans le cas où la fonction de régression possède des paramètres linéaires, et qui permet d'accélérer les calculs. Enfin, nous exposerons la méthode utilisée ici pour ajuster une fonction de régression hyperbolique, et son application aux calculateurs de bureau à programme.

2. MÉTHODE GÉNÉRALE D'AJUSTEMENT DE FONCTIONS NON-LINÉAIRES

La méthode que nous exposons ici est générale. Elle dérive de la méthode de GAUSS-NEWTON, modifiée par H.O. HARTLEY (1). C'est pourquoi nous adopterons une formulation compacte.

2.1. PRÉSENTATION DU MODÈLE

Nous utiliserons la notation matricielle.

Soit le modèle suivant :

$$y = f(x_1, x_2, \dots, x_j, \dots, x_p; \theta_1, \theta_2, \dots, \theta_u, \dots, \theta_w) + \varepsilon \quad (8)$$

où y est la variable expliquée

$x_j, j = 1, 2, \dots, p$ sont les p variables explicatives

$\theta_u, u = 1, 2, \dots, w$ sont les w paramètres de la fonction

Nous supposons que nous disposons des n ($p + 1$) plets $y_i, x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{ip}$ entre lesquels nous avons la relation suivante :

$$y_i = f(x_{i1}; x_{i2}; \dots, x_{ij}, \dots, x_{ip}; \theta_1, \theta_2, \dots, \theta_u, \dots, \theta_w) + \varepsilon_i \quad (9)$$

Notons :

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & & & \cdot \\ \cdot & & x_{ij} & \cdot \\ \cdot & & & \cdot \\ x_{n1} & & & x_{np} \end{pmatrix}$$

$$\theta = [\theta_1 \quad \theta_2 \quad \dots \quad \theta_u \quad \dots \quad \theta_w]'$$

$$\varepsilon = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_i \quad \dots \quad \varepsilon_n]'$$

$$Y = [y_1 \quad y_2 \quad \dots \quad y_i \quad \dots \quad y_n]'$$

¹ Par convention A' est la transposée de A .

Compte tenu de ces notations, le modèle (8) s'écrit :

$$Y = f(X; \theta) + \varepsilon \quad (10)$$

Dans le cas où l'on fait des tests, le vecteur ε doit suivre une loi normale $(0, I\sigma^2)$ où 0 est le vecteur nul et I la matrice unité, tous deux de taille appropriée.

2.2. HYPOTHÈSES

Aux hypothèses faites dans l'introduction, il faut ajouter les hypothèses suivantes sur la fonction $f(X; \theta)$:

a) $f(X; \theta)$ est supposée continue et dérivable par rapport à θ ; $Q(X; \theta)$ est donc dérivable par rapport à θ . Soit $f_u(X; \theta)$ la dérivée de $f(X; \theta)$ par rapport à θ_u : $\partial f / \partial \theta_u = f_u(X; \theta)$.

b) Nous supposons que pour une suite de ξ_u , $u = 1, 2, \dots, w$, avec

$$\sum_u \xi_u^2 > 0 : \sum_{i=1}^n \left[\sum_{u=1}^w \xi_u f_u(x_i; \theta) \right]^2 > 0$$

pour les valeurs observées X et tous les θ d'une suite convergente bornée S de l'espace des paramètres $\theta_1, \theta_2, \dots, \theta_u, \dots, \theta_w$. En pratique, cette hypothèse est fréquemment satisfaite.

c) Notons par $Q = \lim_{\bar{S}} \inf Q(X; \theta)$, où \bar{S} est le complément de S . On suppose qu'il est possible de trouver un vecteur θ_0 à l'intérieur de S tel que $Q(X; \theta_0) < Q$.

2.3. PROCÉDURE ITÉRATIVE DE LA MÉTHODE DE GAUSS-NEWTON MODIFIÉE

2.3.1. Principe

La fonction de régression non linéaire $f(X; \theta)$ est développée d'après la formule de Taylor, au voisinage de θ_0 , défini à l'hypothèse c, θ_0 proche de $\hat{\theta}$, estimation des moindres carrés de θ . On se limite aux termes du premier degré dans le développement.

Le modèle obtenu, linéaire en termes correctifs, est alors ajusté par les méthodes de la régression linéaire simple ou multiple suivant le cas. Cet ajustement génère un vecteur θ_1 , plus proche de $\hat{\theta}$ que θ_0 , si les hypothèses ci-dessus sont vérifiées.

On recherche alors une nouvelle correction pour ce vecteur θ_1 , en reprenant intégralement la procédure ci-dessus, avec θ_1 à la place de θ_0 . Cette procédure est répétée le nombre de fois nécessaire pour atteindre $\hat{\theta}$. Sous les hypothèses du paragraphe 12, H.O. HARTLEY montre que la procédure itérative décrite converge¹.

2.3.2. Développement limité de Taylor de la fonction de régression

Si la fonction $f(X; \theta)$ est continue et dérivable en θ , ce que nous avons supposé, et si l'ensemble des $\hat{\theta}_u - \theta_{u0}$, $u = 1, 2, \dots, w$ sont petits, la formule de

Taylor nous permet d'écrire, en nous limitant au terme du premier degré dans le développement :

$$f(X; \theta) = f(X; \theta_0) + \sum_{i=1}^n (\hat{\theta}_u - \theta_{u0}) \left[\frac{\partial f(X; \theta)}{\partial \theta_u} \right]_{\theta_u = \theta_{u0}} + R^2(\theta) \quad (11)$$

$R^2(\theta)$ est un résidu, du deuxième ordre au moins que nous négligerons.

Posons :

$$\begin{aligned} f(X; \theta_0) &= f^0 \\ \beta_u^0 &= \hat{\theta}_u - \theta_{u0} \\ Z_u^0 &= \left[\frac{\partial f(X; \theta)}{\partial \theta_u} \right]_{\theta_u = \theta_{u0}} = \theta_{u0} \end{aligned}$$

Avec ces notations, (11) s'écrit, toujours en négligeant $R^2(\theta)$:

$$f(X; \theta) = f^0 + \sum_u \beta_u z_u^0 + \varepsilon, u = 1, 2, \dots, w \quad (12)$$

et il est donc relativement aisé d'obtenir une estimation b_u des paramètres β_u .

Cette estimation sera obtenue d'après les résultats de la régression linéaire, multiple dans le cas général.

Exprimons le modèle (12) sous forme matricielle, et notons :

$$Z_0 = \begin{bmatrix} z_{11}^0 & z_{12}^0 & \dots & z_{1w}^0 \\ \dots & \dots & \dots & \dots \\ z_{21}^0 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & z_{iw}^0 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ z_{n1}^0 & \dots & \dots & z_{nw}^0 \end{bmatrix}$$

$$\begin{aligned} Y_0 &= [y_1 - f_{01} \quad y_2 - f_{02} \quad \dots \quad y_i - f_{0i} \quad \dots \quad y_n - f_{0n}] \\ \beta_0 &= [\beta_1^0 \quad \beta_2^0 \quad \dots \quad \beta_u^0 \quad \dots \quad \beta_w^0] \end{aligned}$$

(12) s'écrit alors :

$$Y_0 = Z_0 \beta_0 + \varepsilon \quad (13)$$

et les estimations b_0 du vecteur des paramètres β_0 sont données par

$$b_0 = (Z_0' Z_0)^{-1} Z_0' Y_0 \quad (14)$$

De cette estimation b_0 du vecteur des corrections β_0 , nous pouvons obtenir un vecteur $\theta_1 = \theta_0 + b_0$ plus proche de $\hat{\theta}$ que θ_0 : c'est la méthode de GAUSS-NEWTON simple.

2.3.3. La procédure itérative

Considérons maintenant la fonction de h :

$$Q(h) = Q(x; \theta_0 + hb_0) \quad \text{avec } h \in [0, 1]$$

Soit h' la valeur de h pour laquelle $Q(h)$ est minimum et soit $\sim \theta_1 = \theta_0 + h' b_0$; $\sim \theta_1$ constituera une nouvelle estimation du vrai vecteur $\hat{\theta}$, pour laquelle on recherchera à nouveau une correction. Cette procédure est recommencée le nombre de fois nécessaire pour que les estimations obtenues soient aussi proches que possible des vraies estimations.

En pratique, on arrête les itérations dès que la quantité $Q(\hat{\theta})$ décroît moins qu'une quantité δ , fixée à l'avance, entre la l ème et la $(l+1)$ ème valeur de θ_1 :

¹ Voir (1) pour les démonstrations, non reprises ici.

$$| Q_{(1+p)}(\theta) - Q_1(\theta) | < \delta$$

On peut aussi décider d'arrêter les itérations dès que l'ensemble des quantités $|\theta_{u(1+p)} - \theta_{u1}|$ est inférieur à une valeur δ' fixée à l'avance.

Mais c'est seulement en utilisant le premier critère que H.O. HARTLEY montre que la procédure itérative indiquée converge.

Notons l'ingéniosité du procédé : au lieu de chercher w facteurs de corrections h , on n'en cherche qu'un. Notons encore que si l'on prend systématiquement $h = 1$ (méthode de GAUSS-NEWTON classique), la convergence, souvent beaucoup plus lente, n'est pas assurée.

Néanmoins, une autre difficulté peut aussi surgir. La surface $Q(X; \theta)$ peut aussi avoir de nombreux minima ou maxima locaux, des points stationnaires. Dans de telles conditions, il est nécessaire de rechercher la solution dans un espace large afin de localiser plus sûrement la région S^+ , région convexe de l'espace des paramètres contenant le minimum minimorum de $Q(X; \theta)$. Si un vecteur de départ θ_0 peut être localisé dans la région S^+ , cette méthode peut être utilisée pour converger vers $\hat{\theta}$.

Si ce minimum minimorum n'est pas unique, le principe des moindres carrés n'est pas une méthode appropriée d'estimation, puisqu'il est impossible de faire une distinction entre deux solutions.

2.3.4. Application de la méthode de Gauss-Newton au cas de la régression hyperbolique

Soit a_0, b_0, c_0 , trois valeurs proches respectivement de $\hat{a}, \hat{b}, \hat{c}$. Le développement selon la formule de TAYLOR s'écrira, en négligeant le résidu du développement :

$$a + b/(x+c) = a_0 + b_0/(x+c_0) + (a-a_0) + (b-b_0)/(x+c_0) - (c-c_0)b_0/(x+c_0)^2$$

Notons :

$$1/(x+c_0) = X_1$$

$$1/(x+c_0)^2 = X_2$$

$$c - c_0 = r$$

$Y = f(X; \theta) + \varepsilon$ s'écrira alors :

$$y = a + b X_1 + r X_2 + \varepsilon$$

Ce modèle peut être ajusté relativement simplement d'après les résultats de la régression linéaire multiple avec terme constant.

Nous allons voir maintenant une méthode, applicable dans les cas semblables au présent, qui simplifie les calculs et accélère la convergence de la procédure itérative.

3. MÉTHODE D'ÉLIMINATION DES PARAMÈTRES LINÉAIRES

Dans les méthodes itératives, le nombre d'itérations nécessaires pour atteindre approximativement les estimations des moindres carrés dépend :

- de la méthode utilisée
- de la qualité de l'estimation du vecteur initial
- et bien entendu du modèle choisi

Nous allons voir maintenant une méthode qui permet d'accélérer la procédure itérative en diminuant le nombre de paramètres difficiles à estimer de la fonction. Cette méthode est due à W.H. LAWTON et E.A. SYLVESTRE (2).

3.1. EXPOSÉ PRÉLIMINAIRE DE LA MÉTHODE SUR UN CAS SIMPLE

Considérons le modèle suivant à deux paramètres :

$$f(x; \theta_1, \theta_2) = \theta_1 \exp(\theta_2 x) \quad (15)$$

Les estimations des moindres carrés $\hat{\theta}_1$ et $\hat{\theta}_2$ de θ_1 et θ_2 minimisent :

$$Q(\theta_1, \theta_2) = \sum_i [y_i - \theta_1 \exp(\theta_2 x_i)]^2 \quad i = 1, 2, \dots, n$$

Pour obtenir $\hat{\theta}_1$ et $\hat{\theta}_2$, nous pourrions appliquer telle quelle la méthode précédente, après avoir recherché des estimations initiales θ_{10}, θ_{20} de θ_1 et θ_2 .

Le modèle (15) est linéaire en θ_1 . Pour un θ_2 fixé, soit $\hat{\theta}_1(\theta_2)$ la valeur de θ_1 qui minimise $Q(\theta_1, \theta_2)$; $\hat{\theta}_1(\theta_2)$ a pour expression, d'après les résultats de la régression linéaire simple :

$$\hat{\theta}_1(\theta_2) = \frac{[\sum_i y_i \exp(\theta_2 x_i)]}{\sum_i \exp(\theta_2 x_i)^2} \quad (16)$$

Il est clair que pour θ_2 fixé, à tout autre θ_1 différent de $\hat{\theta}_1(\theta_2)$ correspondra une somme des carrés supérieure.

Le modèle (15) peut alors s'écrire en tenant compte de (16) :

$$f(x; \theta_1, \theta_2) = \left\{ \frac{[\sum_i y_i \exp(\theta_2 x_i)]}{\sum_i \exp(\theta_2 x_i)^2} \right\} \exp(\theta_1 x) \quad (17)$$

Ce modèle (17) ne possède plus qu'un seul paramètre. Son ajustement pourra se faire d'après la méthode de GAUSS-NEWTON modifiée. L'utilisation du modèle (17) est plus intéressante car :

- d'une part, une seule estimation initiale est nécessaire au lieu de deux,
- d'autre part, le problème à résoudre étant de dimension moindre, la convergence est plus rapide avec le modèle (17).

3.2. CAS GÉNÉRAL

Il s'agit d'une simple généralisation. On remplace tous les paramètres linéaires par leurs estimations linéaires des moindres carrés, fonction des paramètres non linéaires du modèle. On obtient ainsi un modèle ne contenant que des paramètres non linéaires, dont l'estimation se fera par la méthode de GAUSS-NEWTON modifiée.

Le modèle général (10) $Y = f(X; \theta) + \varepsilon$ peut se mettre sous la forme suivante :

$$y = \sum_{t=1}^q \beta_t g_t(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m; X) + \varepsilon_t \quad (18)$$

où β_t ; $t = 1, 2, \dots, q$ sont les paramètres linéaires du modèle

α_k ; $k = 1, 2, \dots, m$ les paramètres non linéaires

$g_t(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m; X)$ sont des fonctions seulement des paramètres non linéaires x , et des variables explicatives indépendantes x .

Le modèle a donc en tout $q + m = p$ paramètres.

A une suite initiale de paramètres non linéaires x correspond une suite de paramètres β , dont les valeurs minimisent la somme des carrés des écarts pour une suite de α données.

Soit $\hat{\beta}(\alpha) = [\hat{\beta}_1(\alpha), \hat{\beta}_2(\alpha) \dots \hat{\beta}_q(\alpha)]'$ le vecteur des estimations des moindres carrés de β associés à une suite de $[\alpha_1, \alpha_2, \dots, \alpha_k, \alpha_m]'$.

Soit G_α , la matrice (n, q) dont les éléments sont :

$$g_t(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m; x_i) \quad \begin{array}{l} t = 1, 2, \dots, q \\ i = 1, 2, \dots, n \end{array}$$

Il vient, d'après les résultats de la régression linéaire multiple et avec les notations adoptées :

$$\hat{B}(\alpha) = [G'_\alpha \quad G_\alpha]^{-1} G'_\alpha Y \quad (19)$$

Le modèle réduit a pour expression :

$$y = \sum_{t=1}^q \hat{\beta}_t(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m) g_t(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m; x) + \varepsilon \quad (20)$$

Les $\hat{\beta}_t(\alpha)$ étant fonctions des α , le modèle (20) est un modèle de régression non linéaire avec seulement m paramètres au lieu de $q + m = p$.

Nous avons vu dans l'exemple précédent (par. 2.1.) qu'il était relativement aisé d'obtenir une fonction de forme simple pour $\hat{\beta}(\alpha)$, en l'occurrence (16).

Mais dans le cas général, $\hat{\beta}(\alpha)$ est une fonction complexe difficilement maniable, et il est généralement impossible d'obtenir les dérivées analytiques de (20) par rapport aux α .

Or ces dérivées sont nécessaires dans la plupart des méthodes itératives d'ajustement de fonctions non linéaires. Pratiquement, on résout ce problème en remplaçant les dérivées par des différences finies. Ces dernières sont obtenues en évaluant (20) à $(\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m)$, $(\alpha_1 + \delta\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_m)$, ... etc. Par cette méthode (19) peut être évalué pour chaque choix d'un des α par une opération matricielle. L'estimation des moindres carrés des α est obtenue par la méthode de GAUSS-NEWTON modifiée.

Les avantages du modèle réduit par rapport au modèle initial sont :

— tout d'abord, de ne nécessiter que m valeurs de départ au lieu de $m + q = p$

— de réduire le nombre de dimensions de l'espace dans lequel la procédure itérative doit converger; cet espace est de dimension m au lieu de $m + q$.

En général, la convergence est plus rapide, et elle est assurée pour un ensemble plus large de valeurs initiales.

4. MÉTHODE PRATIQUE D'AJUSTEMENT D'UNE FONCTION DE RÉGRESSION HYPERBOLIQUE

Nous allons maintenant appliquer les deux méthodes décrites ci-dessus au cas de la régression hyperbolique.

4.1. ELIMINATION DES PARAMÈTRES LINÉAIRES DANS L'EXPRESSION DE L'HYPERBOLE

Nous avons vu, dans l'introduction, que l'expression générale d'une hyperbole pouvait être :

$$\bar{y} = a + b/(x+c) \quad (21)$$

Ce modèle (21) est linéaire en a et b ; le seul paramètre non linéaire est c .

Posons $1/(x+c) = X$. Le modèle (21) s'écrit :

$$\bar{y} = a + bX \quad (22)$$

Nous avons donc à ajuster :

$$y = a + bX + \varepsilon \quad (23)$$

Ce dernier modèle (23) va nous permettre d'obtenir les estimations linéaires des moindres carrés fonction des paramètres non linéaires du modèle.

Pour c fixé, et donc pour X fixé, les expressions analytiques $\hat{b}(c)$ et $\hat{a}(c)$ sont obtenues simplement d'après les résultats de la régression linéaire simple :

$$\hat{b}(c) = \frac{(\sum X_i y_i - \sum X_i \sum y_i/n) / [\sum X_i^2 - (\sum X_i)^2/n]}{}$$

$$\hat{a}(c) = \frac{\sum y_i/n - \hat{b}(c) \sum X_i/n}{}$$

En notant par $\bar{y} = \sum y_i/n$, le modèle à ajuster est

donc :

$$y = \bar{y} + \{(\sum X_i y_i - \bar{y} \sum X_i) / [\sum X_i^2 - (\sum X_i)^2/n]\} (X_i - \sum X_i/n) + \varepsilon \quad (24)$$

(24) est un modèle non linéaire à un seul paramètre, dont l'ajustement se fera par la méthode de GAUSS-NEWTON modifiée.

4.2. DÉVELOPPEMENT LIMITÉ AU VOISINAGE DE c_0 DE LA FONCTION DE RÉGRESSION

Notons :

$$X_0 = 1/(x+c_0)$$

$[\partial \bar{y} / \partial c]_{c=c_0}$ la dérivée de \bar{y} par rapport à c , pour $c=c_0$

$$\bar{y}_0 = \hat{a}(c_0) + \hat{b}(c_0) X_0$$

Le développement de (21) au voisinage de c_0 s'écrit, selon la formule de TAYLOR, et en se limitant au terme du premier degré dans le développement :

$$\bar{y} = \bar{y}_0 + [\partial \bar{y} / \partial c]_{c=c_0} (c - c_0) + R^2(c) \quad (25)$$

$R^2(c)$ est un résidu d'ordre deux au moins.

Par rapport à c , y est de la forme $uv + Cte$. La dérivée $\partial \bar{y} / \partial c$ sera de la forme $uv' + vu'$.

Notons :

$$U = \frac{(\sum X_i y_i - \bar{y} \sum X_i) / [\sum X_i^2 - (\sum X_i)^2 / n]}{1}$$

\tilde{y} s'écrit alors :

$$\tilde{y} = \bar{y} + U (X_i - \sum X_i / n)$$

En notant de plus $\partial U / \partial c$ par V,

$$\frac{\partial U}{\partial c} = \frac{2(\sum X_i y_i - \bar{y} \sum X_i)(\sum X_i^3 - \sum X_i \sum X_i^2 / n) - [\sum X_i^2 - (\sum X_i)^2 / n] (\sum X_i y_i - \bar{y} \sum X_i)}{[\sum X_i^2 - (\sum X_i)^2 / n]^2}$$

$\partial \tilde{y} / \partial c$ que nous noterons plus simplement par z aura pour expression :

$$\partial \tilde{y} / \partial c = z = U (\sum X_i^2 / n - X_i^2) + V (X_i - \sum X_i / n)$$

En notant encore par :

— U_0 et V_0 la valeur de U et V pour $X = X_0$, ($c = c_0$)

— z_0 la valeur de z pour $c = c_0$

— $\tilde{y}_{res} = \tilde{y} - \tilde{y}_0 = \tilde{y} - \bar{y} U_0 (X_{oi} - \sum X_{oi} / n)$

— $\delta c = c - c_0$

(25) s'écrit alors, en négligeant $R^2(c)$:

$$\tilde{y}_{res} = z_0 \delta c \tag{26}$$

(26) est un modèle linéaire sans terme constant, qu'il nous faut maintenant ajuster.

4.3. La méthode itérative

Nous allons choisir le critère d'estimation de δc . Le modèle à ajuster est

$$y_{res,i} = z_{oi} \delta c + \varepsilon_i \tag{27}$$

ce qui signifie que $y_{res,i}$ est proportionnel à z_{oi} , aux erreurs d'échantillonnage près. Pour évaluer δc , c'est-à-dire pour évaluer globalement le rapport $y_{res,i} / z_{oi}$, trois critères sont utilisables.

Dans le cas d'un modèle de la forme $Y = \beta X + \varepsilon$, SNEDECOR et COCHRAN préconisent d'utiliser (3) comme estimation de β :

- $\sum XY / \sum X^2$ si la variance des ε est constante
- $\sum Y / \sum X$ si la variance des ε croît moins que proportionnellement ou proportionnellement à X
- $\sum (Y/X) / n$ si la variance des ε croît proportionnellement à X^2

Plusieurs séries de données concernant la relation entre la photopériode et la somme des températures de plusieurs espèces cultivées ont été étudiées, et il semble que le critère à adopter soit celui correspondant à une variance constante :

$$\delta c = (\sum y_{res,i} z_{oi}) / (\sum z_{oi}^2)$$

Pour trouver le coefficient de correction h optimal, nous allons évaluer la somme des carrés des écarts $Q(h)$ correspondant à $c_1 = c_0 + h \delta c$, à $h = 0$, $h = 1/2$, $h = 1$, et nous déterminons (fig. 1) l'abscisse du minimum de la parabole passant par $Q(0)$, $Q(1/2)$ et $Q(1)$.

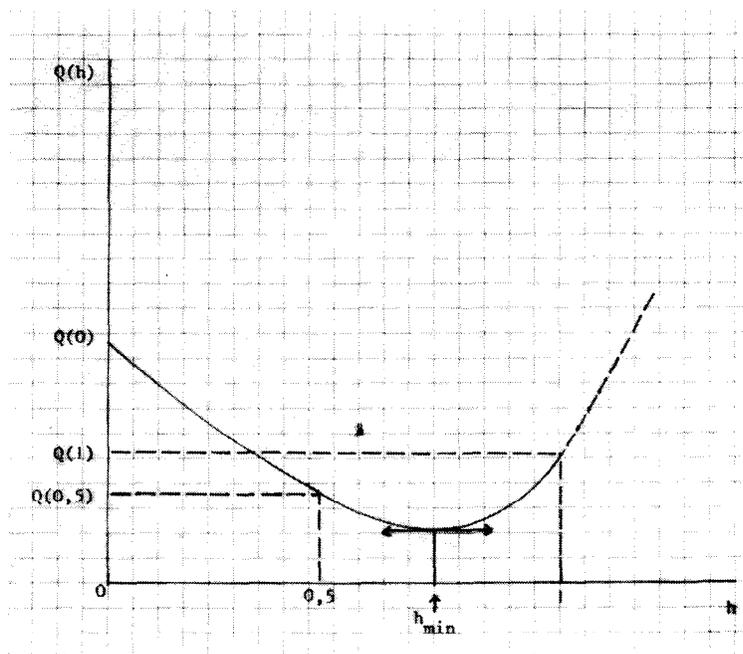


Fig. 1

Nous noterons h_{\min} ce minimum. Il a pour expression :

$$h_{\min} = 1/2 + 1/4[Q(0) - Q(1)] / [Q(1) - 2Q(1/2) + Q(0)]$$

On évalue alors $Q(h_{\min})$, la somme des carrés des écarts correspondant à $c_2 = c_0 + h_{\min} \delta c$. Nous la noterons simplement SCE :

$$SCE = \sum_i (y_{\text{res},i}^2) \quad c = c_2$$

La décroissance observée de SCE est alors comparée à la décroissance minimale désirée de la somme des carrés des écarts : Δ .

Si Δ est supérieur à $[Q(0) - SCE]$, où $Q(0)$ est la somme des carrés correspondant à c_0 , on reprend intégralement la procédure ci-dessus avec c_2 à la place de c_0 .

Dans le cas contraire on calcule :

$$\hat{b} = \frac{(\sum_i X_i y_i - \bar{y} \sum_i X_i) / [\sum_i X_i^2 - (\sum_i X_i)^2 / n]}$$

$$\hat{a} = \bar{y} - \hat{b} \sum_i X_i / n$$

où X_i désigne maintenant $1/(x_i + \hat{c}) = 1/(x_i + c_2)$

On calcule aussi la somme des carrés des écarts correspondant à \hat{a} , \hat{b} et \hat{c} , ainsi que

$$i^2_{y/x} = \frac{[\sum_i (\hat{y}_i - \bar{y})^2] / [\sum_i (y_i - \bar{y})^2]}$$

indice de corrélation curvilinéaire.

4.4. ESTIMATION INITIALE c_0 DE c

L'hyperbole (H) d'équation $y = a + b/(x + c)$ admet pour asymptotes les droites D_1 et D_2 , respectivement d'équation $y = a$ et $x = -c$.

Il faut d'abord estimer graphiquement c et a , et ensuite calculer b . Mais dans la méthode utilisée ici, seule l'estimation de c est nécessaire.

Cette estimation graphique (fig. 2) semble suffisamment précise pour la méthode d'ajustement présentée.

PROGRAMME D'AJUSTEMENT D'UNE FONCTION DE RÉGRESSION HYPERBOLIQUE

L'organigramme du calcul est donné en annexe.

Le texte du programme n'a pas été inséré dans cet article car il a été écrit en langage machine pour un ordinateur programmable WANG 700 C. Il est disponible au service de Biométrie de l'Office de la Recherche Scientifique et Technique Outre-Mer.

Le WANG 70 C est un ordinateur possédant une mémoire de 120 registres, chaque paire de registres pouvant contenir deux nombres réels ou seize instructions programme. Les programmes peuvent être stockés sur une cassette et être chargés en mémoire automatiquement au moment voulu. Mais il n'y a qu'un sens de déroulement de la bande. Le réenbobinage doit être manuel.

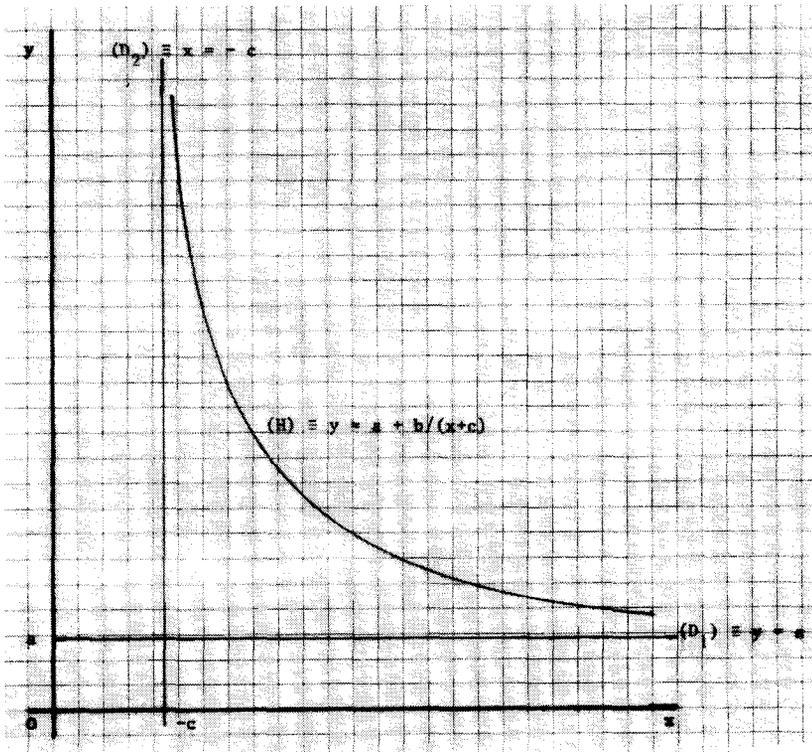


Fig. 2

Le programme d'ajustement d'une fonction de régression hyperbolique a été divisé en dix blocs, le premier d'impression, préparation des sorties de calcul et contenant un sous-programme, les neuf autres exécutant les calculs décrits plus haut. Le premier bloc n'étant exécuté qu'une fois, seuls les neuf blocs suivants ont été reproduits sur toute la longueur de la bande de la cassette.

Ce programme est utilisable pour l'ajustement d'une hyperbole, lorsque le nombre de couples de données est inférieur ou égal à 85. Prévu pour des données de la forme $X = xxx.xxx$ et $Y = yyyy.yy$, il doit être modifié si les données ne sont plus de cette forme. Si les estimations des moindres carrés peuvent être atteintes en dix cycles au plus, aucune intervention manuelle n'est nécessaire.

Nous donnons ci-dessous les séquences de calculs effectués par les différents blocs.

Bloc 1	Sous-programme. Impression des titres et préparation de la machine à écrire pour les sorties de calcul.
Blocs 2-3-4	Entrée des n couples de données, et calcul de δc .
Blocs 5-6-7-8	Calcul et impression des 3 $Q(h)$, $h = 0, 1/2, 1$ Calcul et impression de h_{\min} Calcul et impression de $Q(h_{\min}) = SCE$, et test de comparaison de Δ à ($Q(0) - SCE$)

Blocs 9-10	Si $\Delta > (Q(0) - SCE)$ Calcul et impression de $\hat{a}, \hat{b}, \hat{c}$, Calcul et impression de $x_i, y_i, \tilde{y}_i = \hat{a} + \hat{b}/(x_i + \hat{c})$ Calcul et impression de $e_i = y_i - \tilde{y}_i$ Calcul et impression de $i_{y/x}^2 = \frac{[\sum_i (y_i - \bar{y})^2]}{[\sum_i (y_i - \bar{y})^2]}$ Sinon, le calculateur saute ces deux blocs.
------------	--

BIBLIOGRAPHIE

- (1) H.O. HARTLEY — 1961 — The modified Gauss-Newton Method for the fitting of non-linear Regression functions by least Squares — *Technometrics* — vol. 3, no 2, pp. 269-280 (May 1961).
- (2) W.H. LAWTON and E.A. SYLVESTRE — 1971 — Elimination of linear Parameters in non linear regression — *Technometrics* — vol. 13, no 3, pp. 461-481 (August 1971).
- (3) W.G. SNEDECOR and G.N. COCHRAN — 1968 — *Statistical Methods* — The Iowa State University Press-Ames — USA. (1968).

ANNEXE

ORGANIGRAMME : AJUSTEMENT D'UNE HYPERBOLE D'EQUATION
 $y = a + b/(x + c)$

