

REVUE DE STATISTIQUE APPLIQUÉE

P. CAZÈS

P. Y. TURPIN

Régression sous contraintes : application à l'estimation de la courbe granulométrique d'un aérosol

Revue de statistique appliquée, tome 19, n° 4 (1971), p. 23-44

http://www.numdam.org/item?id=RSA_1971__19_4_23_0

© Société française de statistique, 1971, 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/>

RÉGRESSION SOUS CONTRAINTES APPLICATION A L'ESTIMATION DE LA COURBE GRANULOMÉTRIQUE D'UN AÉROSOL

P. CAZES ⁽¹⁾ et P. Y. TURPIN ⁽²⁾

	Pages
I - INTRODUCTION	25
II - FORMALISATION MATHÉMATIQUE DU PROBLÈME. INTER- PRÉTATION GÉOMÉTRIQUE	25
1 - Formalisation	25
2 - Etude géométrique.....	26
2.1 - Interprétation géométrique.....	26
2.2 - Application	26
2.3 - Cas particulier d'une seule contrainte : les coef- ficients de régression doivent être positifs Coefficient de corrélation multiple sous contrainte	27
2.4 - Cas général	28
III - ETUDE ALGORITHMIQUE	29
1 - Méthode de Wolfe	30
2 - Méthode de d'Esopo	31
3 - Cas où la somme des coefficients de régression vaut un	31
IV - EXEMPLES D'UTILISATION DE LA RÉGRESSION SOUS CON- TRAINTES	32
1 - Problèmes de régression où la variable à expliquer est positive	32
2 - Problèmes de régression entre variables positives	33

(1) Laboratoire de Statistique mathématique, ISUP, Université de Paris 6

(2) Laboratoire de Physique des Aérosols, Université de Paris 6

	Pages
V - ESTIMATION DE LA COURBE GRANULOMETRIQUE D'UN AEROSOL	33
1 - Montage expérimental	34
2 - Détection.....	34
3 - Interprétation des mesures.....	35
4 - Résultats.....	36
VI - CONCLUSION	44

I - INTRODUCTION

Dans un certain nombre de problèmes économiques, physiques, géologiques, statistiques, etc..., on ne peut toujours appliquer directement la méthode des moindres carrés pour expliquer une variable y en fonction d'autres variables x_1, x_2, \dots, x_p . C'est le cas si les coefficients de régression b_1, b_2, \dots, b_p , -ou les valeurs approchées des observations de y - doivent être comprises entre des bornes données pour avoir un sens, compte tenu du problème traité. On a alors affaire à des problèmes de régression sous contraintes, c'est-à-dire des problèmes de programmation quadratique.

Après la formulation mathématique du problème que l'on vient d'exposer, nous en donnons une interprétation géométrique, avant de rappeler les techniques classiques de programmation quadratique, de donner un algorithme dans le cas où les coefficients de régression doivent être positifs, et dans le cas où on estime une loi de probabilité.

Nous citerons quelques exemples où ces techniques de régression sous contraintes s'appliquent, avant d'exposer plus en détail les résultats obtenus lors de l'estimation de la courbe granulométrique d'un aérosol.

II - FORMALISATION MATHÉMATIQUE DU PROBLÈME ET INTERPRÉTATION GÉOMÉTRIQUE

1 - Formalisation

Soient y, x_1, x_2, \dots, x_p , $p + 1$ variables (aléatoires ou non) dont on possède un n -échantillon : $\{y^j, x_1^j, \dots, x_p^j \mid j = 1, n\}$: on se propose "d'expliquer" y en fonction des x ; au sens des moindres carrés, i. e. on recherche la combinaison linéaire $\Sigma \{b^i x_i \mid i = 1, p\}$ telle que

$$\Sigma \{(y^j - \Sigma \{b^i x_i^j \mid i = 1, p\})^2 \mid j = 1, n\} \quad (1)$$

soit minimum, sous les contraintes supposées compatibles :

$$\left. \begin{aligned} \forall i, 1 \leq i \leq p : \alpha_i \leq b^i \leq \beta_i & \quad (a) \\ \forall j, 1 \leq j \leq n : \alpha \leq \Sigma \{b^i x_i^j \mid i = 1, p\} \leq \beta & \quad (b) \end{aligned} \right\} \quad (2)$$

Ces contraintes peuvent ne pas être indépendantes, c'est en particulier le cas si l'on a affaire à des variables positives, et si l'on impose aux coefficients de régression d'être positifs ($\alpha_i = \alpha = 0$; $\beta_i = \beta = +\infty$; $x_i^j \geq 0$), les contraintes (2) (b) étant alors satisfaites, dès que le sont les contraintes (2) (a).

De façon plus générale, la donnée d'un n échantillon de $p + 1$ variables y, x_1, x_2, \dots, x_p conduit à considérer $p + 1$ vecteurs de \mathbb{R}^n $\vec{y}, \vec{x}_1, \vec{x}_2, \dots, \vec{x}_p$ (\vec{y} ayant pour composantes : $\{y^j \mid j = 1, n\}$, et $\vec{x}_i : \{x_i^j \mid j = 1, n\}$). Ayant muni \mathbb{R}^n de la métrique⁽¹⁾ N , et défini un vecteur \vec{z} tel que

$$\vec{z} = \Sigma \{b^i x_i \mid i = 1, p\}$$

(1) i. e. une forme bilinéaire symétrique définie positive.

on désire minimiser la norme de $\vec{y} - \vec{z}$, ou son carré

$$N(\vec{y} - \vec{z}, \vec{y} - \vec{z}), \text{ sous les contraintes (2).}$$

2 - Etude géométrique

2.1 - Interprétation géométrique

D'après II 1, \vec{z} est donc la projection (au sens de N) de \vec{y} sur le convexe fermé C défini par :

$$C = \left\{ \vec{u} \mid \vec{u} = \sum \{ b^i \vec{x}_i \mid i = 1, p \}, \forall i, 1 \leq i \leq p : \alpha_i \leq b^i \leq \beta_i, \right. \\ \left. \forall j, 1 \leq j \leq n : \alpha \leq u^j < \beta \right\} \quad (3)$$

Cette projection est unique et existe toujours, puisque les contraintes (2) n'étant pas incompatibles, C n'est pas vide.

Si les vecteurs $\{\vec{x}_i \mid i = 1, p\}$ sont linéairement indépendants, ce que nous supposons, ils forment une base du sous espace vectoriel H qu'ils engendrent : \vec{z} , qui appartient à H puisque C est contenu dans H, a pour coordonnées par rapport à cette base les "Coefficients de régression" b^i qui sont donc uniques.

Dans certains problèmes, en particulier dans l'estimation d'une mesure de probabilité discrète, les coefficients b^i ne sont pas indépendants, et sont soumis à la contrainte :

$$\sum \{ b^i \mid i = 1, p \} = 1 \quad (4)$$

ce qui revient à projeter \vec{y} sur le convexe C_1 (supposé non vide) contenu dans C et tel que (4) soit satisfaite. Nous désignerons par \vec{z}_1 cette projection.

2.2 - Application

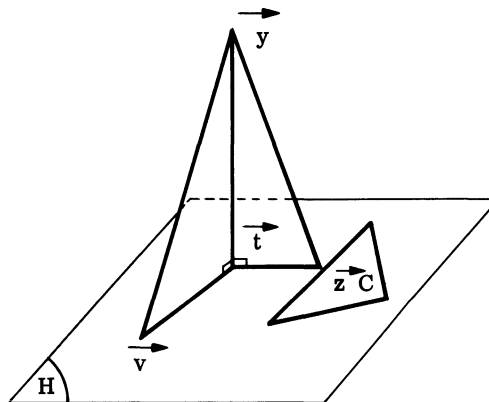


Figure 1

Soit \vec{t} la projection de \vec{y} sur le sous espace vectoriel H engendré par les $\{\vec{x}_i \mid i = 1, p\}$: si $\|v\|$ désigne la norme d'un vecteur de R^n

($\|\vec{v}\|^2 = N(\vec{v}, \vec{v})$), on a :

$$\forall \vec{v} \in H : \|\vec{y} - \vec{v}\|^2 = \|\vec{y} - \vec{t}\|^2 + \|\vec{t} - \vec{v}\|^2 \quad (5)$$

Rendre minimum $\|\vec{y} - \vec{v}\|$ pour \vec{v} parcourant C, revient donc à rendre minimum $\|\vec{t} - \vec{v}\|$ pour \vec{v} parcourant C ; ce minimum qui est atteint pour $\vec{v} = \vec{z}$ est tel que :

$$\|\vec{y} - \vec{z}\|^2 = \|\vec{y} - \vec{t}\|^2 + \|\vec{t} - \vec{z}\|^2 \geq \|\vec{y} - \vec{t}\|^2$$

l'égalité n'ayant lieu que si $\vec{t} = \vec{z}$, i. e. si \vec{t} appartient à C.

Par ailleurs, si \vec{t} n'appartient pas à l'intérieur $\overset{\circ}{C}$ de C, \vec{z} appartient à la frontière δC de C, car si ce n'était pas le cas, \vec{z} appartiendrait à C, et la droite joignant \vec{t} à \vec{z} couperait δC en un point \vec{z}' plus proche de \vec{t} (donc de \vec{y} , d'après (5)) que \vec{z} .

Les résultats précédents peuvent se traduire par le théorème suivant:

- a) La projection \vec{z} de \vec{y} sur C n'est autre que la projection de \vec{t} sur C
- b) \vec{t} est un point plus proche de C que \vec{y}
- c) Si $\vec{t} \in C$, $\vec{z} = \vec{t}$ et réciproquement
- d) Si $\vec{t} \notin \overset{\circ}{C}$, alors $\vec{z} \in \delta C$, et réciproquement.

On a bien sûr des résultats analogues si on projette sur C_1 .

2.3 - Cas particulier d'une seule contrainte : les coefficients de régression doivent être positifs. Coefficient de corrélation multiple sous contrainte

Dans ce cas, le convexe C se réduit au cône positif de H. Désignons par H' le sous espace vectoriel engendré par les vecteurs \vec{x}_1 dont les coefficients de régression sous contrainte sont strictement positifs, et par C' le cône positif associé : \vec{z} qui appartient à $\overset{\circ}{C}$, et qui est le point le plus proche de \vec{y} dans C' (car si ce n'était pas le cas, il existerait un point de C', donc de C plus proche de \vec{y} que \vec{z}) est donc d'après II 2.2 confondu avec la projection \vec{t}' de \vec{y} sur H', d'où l'on déduit :

$$\left. \begin{aligned} \|\vec{y}\|^2 &= \|\vec{y} - \vec{z}\|^2 + \|\vec{z}\|^2 \\ \|\vec{t}\|^2 &= \|\vec{t} - \vec{z}\|^2 + \|\vec{z}\|^2 \end{aligned} \right\} \quad (6)$$

Le théorème de Pythagore est donc encore valable.

On posera :

$$\left. \begin{aligned} R^2 &= \|\vec{z}\|^2 / \|\vec{y}\|^2 : \text{qualité de l'approximation} \\ &\quad \text{de } \vec{y} \text{ par les } \vec{x}_1 \\ S^2 &= \|\vec{y} - \vec{z}\|^2 / \|\vec{y}\|^2 : \text{résidu} \end{aligned} \right\} \quad (7)$$

On a donc :

$$R^2 + S^2 = 1 \quad (8)$$

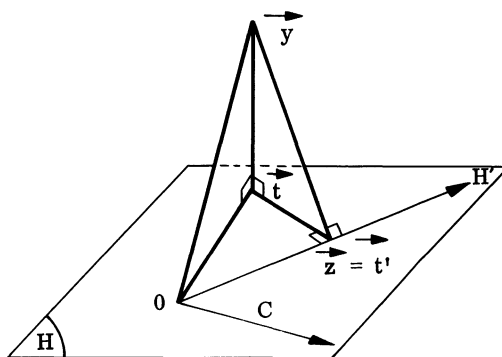


Figure 2

Si les moyennes empiriques de y et des x_i sont nulles, ou si l'on a introduit parmi les variables x_i , la variable certaine et égale à un, alors R^2 n'est autre que le carré du coefficient de corrélation multiple empirique classique⁽¹⁾ de la variable y par rapport à celles des variables x_i auxquelles sont associés des "coefficients de régression sous contrainte" b^i strictement positifs.

On peut dire que R est le coefficient de corrélation multiple sous contraintes de y par rapport aux variables $\{x_i \mid i = 1, p\}$, et il est bien sûr inférieur au coefficient de corrélation multiple classique R_o ⁽²⁾ ($R_o = \|\vec{t}\| / \|\vec{y}\|$) de y par rapport aux $\{x_i \mid i = 1, p\}$, puisque

$$\begin{aligned} R^2 &= \|\vec{z}\|^2 / \|\vec{y}\|^2 = (\|\vec{z}\|^2 / \|\vec{t}\|^2) \times (\|\vec{t}\|^2 / \|\vec{y}\|^2) \\ &= (\|\vec{z}\|^2 / \|\vec{t}\|^2) R_o^2 \leq R_o^2, \text{ d'après (6)} \end{aligned}$$

2.4 - Cas général

Dans le cas où le convexe C est quelconque, on définit toujours R^2 et S^2 (R_1^2 et S_1^2 , si on projette sur C_1) par les relations (7), mais les relations (6) et (8) ne sont pas en général vérifiées : R^2 et même S^2 , si C ne contient pas l'origine, peuvent être supérieurs à 1.

En pratique, si le modèle correspond à la réalité physique, S^2 (ou S_1^2) est toujours inférieur à 1 ; R^2 (ou R_1^2) peut être supérieur à 1 du fait des fluctuations d'échantillonnage. Cela arrive quand on a une bonne approximation de y comme combinaison linéaire (à coefficients positifs) des x_i .

(1) En prenant bien sûr pour norme N la norme classique.

(2) Dans le cas de variables non centrées, ou d'une norme quelconque,

$$R_o^2 = \|\vec{t}\|^2 / \|\vec{y}\|^2 = 1 - S_o^2 \quad (\text{avec } S_o^2 = \|\vec{y} - \vec{t}\|^2 / \|\vec{y}\|^2)$$

sera appelé "qualité de l'approximation" de \vec{y} par les vecteurs \vec{x}_i pour la régression sans contrainte.

III - ETUDE ALGORITHMIQUE

La résolution du problème formulé au II 1, se ramène à celle d'un programme dont la fonction de coût à minimiser est quadratique et convexe. Plusieurs types de procédures sont possibles pour trouver la solution.

La première résulte de l'étude géométrique et revient à projeter le vecteur \vec{y} en \vec{t} sur l'espace H engendré par les vecteurs $\{\vec{x}_i \mid i = 1, p\}$, puis si \vec{t} n'appartient pas au convexe C (ou C_1 suivant le cas), à projeter \vec{y} (ou \vec{t}) suivant une procédure itérative (cf. réf. 4) sur les différentes faces de C, (puisque \vec{z} projection de \vec{y} sur C appartient à la frontière de C).

Un autre algorithme est exposé dans (5) dans le cas où on exige uniquement que les valeurs reconstituées soient positives / i. e.

$$\alpha_1 = -\infty, \alpha = 0, \beta_1 = \beta = \infty$$

Nous allons rappeler brièvement les procédures itératives classiques, en particulier celle du gradient réduit de Wolfe et ses variantes, lorsqu'il n'y a pas de contraintes sur les valeurs reconstituées (contraintes (2) (b)), ou s'il y en a lorsque ces contraintes sont vérifiées, dès que le sont les contraintes (2) (a)).

Dans un premier temps, nous supposons les coefficients b^i indépendants, c'est-à-dire non soumis à la contrainte (4); nous verrons comment modifier les algorithmes dans le cas contraire.

On cherche donc le vecteur \vec{b} de composantes $\{b^i \mid i = 1, p\}$, minimisant la fonction quadratique convexe de \vec{b} :

$$Q(\vec{b}) = N(\vec{y} - \sum \{b^i \vec{x}_i \mid i = 1, p\}, \vec{y} - \sum \{b^i \vec{x}_i \mid i = 1, p\}) \quad (9)$$

sous les contraintes

$$\alpha_1 \leq b^i \leq \beta_1 \quad (10)$$

Les vecteurs \vec{x} étant supposés linéairement indépendants, il y a comme on l'a vu une solution unique.

La procédure itérative s'écrit à l'itération $t + 1$:

$$\vec{b}_{t+1} = \vec{b}_t + \theta d \vec{b} \quad (11)$$

où

- $d\vec{b}$ est le vecteur de composantes $\{db^i \mid i = 1, p\}$, caractérisant la direction de déplacement, et dont l'expression dépend de la méthode utilisée.
- θ est déterminé de telle sorte que dans la direction considérée il minimise Q tout en respectant les contraintes (10) pour \vec{b}_{t+1} ; le point de départ \vec{b}_1 est choisi de façon à ce que les contraintes soient vérifiées.

Posons :

$$\left. \begin{aligned} \varphi(\theta) &= Q(\vec{b}_{t+1}) = Q(\vec{b}_t + \theta \vec{db}) \\ \vec{z}_t &= \Sigma\{b_t^i \vec{x}_i \mid i = 1, p\} \\ \vec{dz} &= \Sigma\{db^i \vec{x}_i \mid i = 1, p\} \end{aligned} \right\} \quad (12)$$

$\varphi(\theta)$ qui se met sous la forme :

$$\left. \begin{aligned} N(\vec{y} - \vec{z}_t - \theta \vec{dz}, \vec{y} - \vec{z}_t - \theta \vec{dz}) &= N(\vec{dz}, \vec{dz}) \theta^2 - 2N(\vec{y} - \vec{z}_t, \vec{dz})\theta \\ &+ N(\vec{y} - \vec{z}_t, \vec{y} - \vec{z}_t) \end{aligned} \right\} \quad (13)$$

est un trinôme du second degré ;

le coefficient de θ^2 , $A = N(\vec{dz}, \vec{dz}) = ||\vec{dz}||^2$, est strictement positif :

$\varphi(\theta)$ passe donc par un minimum pour la valeur

$$\theta_0 = N(\vec{y} - \vec{z}_t, \vec{dz}) / N(\vec{dz}, \vec{dz}) \quad (14)$$

Par ailleurs, en écrivant que \vec{b}_{t+1} satisfait les contraintes (10), on obtient pour θ un système d'inégalités linéaires toujours compatibles, puisqu'il est vérifié pour $\theta = 0$ (\vec{b}_t satisfaisait les contraintes (10)); ce système définit pour θ un intervalle (θ_1, θ_2) avec $\theta_1 \leq 0 \leq \theta_2$.

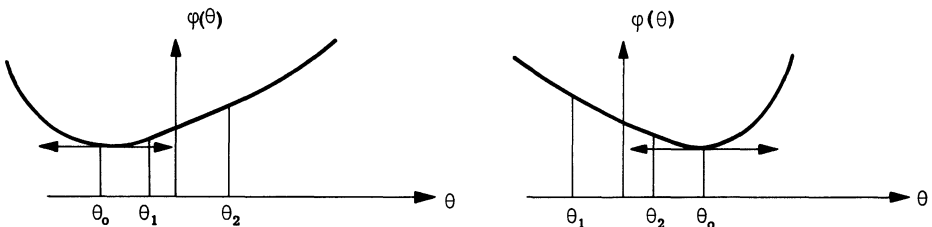


Figure 3

Si donc θ_0 est à l'intérieur de cet intervalle, on choisira $\theta = \theta_0$, sinon, si θ_0 est plus petit que θ_1 , $\varphi(\theta)$ étant croissante dans l'intervalle (θ_0, θ_2) , on prendra $\theta = \theta_1$, et si θ_0 est plus grand que θ_2 , $\varphi(\theta)$ étant décroissante dans l'intervalle (θ_1, θ_0) , on prendra $\theta = \theta_2$ (cf. fig. 3).

1 - Méthode de Wolfe

Dans la méthode de Wolfe, on choisit comme direction de déplacement :

$$\left. \begin{aligned} db^i &= 0 \quad \text{si} \quad b_t^i = \alpha_i \quad \text{et} \quad (\partial Q / \partial b^i)_{\vec{b}=\vec{b}_t} > 0 \\ \text{ou} \\ b_t^i &= \beta_i \quad \text{et} \quad (\partial Q / \partial b^i)_{\vec{b}=\vec{b}_t} < 0 \\ \text{sinon} \\ db^i &= - (\partial Q / \partial b^i)_{\vec{b}=\vec{b}_t} \end{aligned} \right\} \quad \begin{array}{l} (a) \\ (15) \\ (b) \end{array}$$

On a alors :

$$\begin{aligned} \varphi(\theta) &= Q(\vec{b}_t + \theta \vec{db}) = Q(\vec{b}_t) + \theta \sum \left\{ \left(\frac{\partial Q}{\partial b^i} \right)_{\vec{b}=\vec{b}_t} db^i \mid i = 1, p \right\} + A \theta^2 \\ &= Q(\vec{b}_t) - \theta \sum \{ (db^i)^2 \mid i = 1, p \} + A \theta^2 \end{aligned} \quad (16)$$

où A est un coefficient positif.

Au voisinage de $\theta = 0$, $\varphi(\theta)$ est décroissante, θ_0 est donc positif, et supérieur à θ_1 ; on a donc :

$$\theta = \min. (\theta_0, \theta_2) \quad (17)$$

Par ailleurs, si $\theta_0 = 0$, i.e si \vec{db} (ou \vec{dz}) est nul, on a atteint le point optimum, puisque si l'on se déplace dans H sur n'importe quelle direction parallèle aux \vec{x}_1 , on ne peut pas diminuer Q, sauf peut-être en sortant de C.

La suite $Q(b_t)$ étant strictement décroissante, et bornée par 0, converge ; l'algorithme est donc convergent.

2 - Méthode de d'Esopo

Dans cette méthode, la direction de déplacement \vec{dz} est prise successivement suivant chaque vecteur \vec{x}_1 ; on prendra donc si on pose :

$$\left. \begin{aligned} \vec{z}_{t+1} &= \vec{z}_t + \theta \vec{dz}_t \\ \vec{dz}_t &= \vec{x}_t, \text{ avec } 1 \leq t \leq p \end{aligned} \right\} \quad (18)$$

ce qui revient à prendre $b_{t+1}^t = b_t^t + \theta$, et si $i \neq t$, $b_{t+1}^i = b_t^i$; on a alors :

$$\varphi(\theta) = Q(\vec{b}_{t+1}) = Q(\vec{b}_t) + \theta \left(\frac{\partial Q}{\partial b^t} \right)_{\vec{b}=\vec{b}_t} + A \theta^2, \quad (19)$$

l'intervalle (θ_1, θ_2) n'étant autre que l'intervalle $(\alpha_t - b^t, \beta_t - b^t)$.

Quand on a fait p itérations à partir de \vec{z}_1 (une boucle) si \vec{z}_{p+1} est différent de \vec{z}_1 , on recommence le processus, avec comme nouveau point de départ $\vec{z}_1 = \vec{z}_{p+1}$. Si $\vec{z}_{p+1} = \vec{z}_1$ (i.e. $\vec{b}_{p+1} = \vec{b}_1$), il est facile de voir d'après (18) et l'expression de l'intervalle (θ_1, θ_2) , que l'on a atteint l'optimum, i.e. que l'on se trouve dans le cas où $\forall i, 1 \leq i \leq p$, les conditions (15) (a) de la méthode de Wolfe sont vérifiées.

Si l'on désigne par Q_1 la valeur de Q à la suite de la $i^{\text{ème}}$ boucle, la suite Q_1 étant strictement décroissante et bornée par 0 est convergente ; l'algorithme est donc convergent.

3 - Etude du cas où la somme des b^i vaut 1⁽¹⁾

Si on impose la contrainte $\sum b^i = 1$, les coefficients b^i ne sont plus "indépendants". On les divisera en deux parties :

(1) On fait bien sûr l'hypothèse que les contraintes (10) sont compatibles avec la contrainte $\sum \{ b^i \mid i = 1, p \} = 1$.

b^1, b^2, \dots, b^{p-1} , d'une part, et b^p d'autre part, b^p étant donné par :

$$b^p = 1 - \sum \{b^i \mid i = 1, p-1\}$$

b^p étant la variable basique (ou dépendante), b^1, \dots, b^{p-1} les variables hors base (ou indépendantes) ; on supposera toujours que b^p est différent de α_p ou β_p ; si ce n'était pas le cas, on intervertirait b^p avec une variable non basique b^i , telle que b^i soit différent de α_i ou β_i .

On raisonne alors comme précédemment sur les variables non basiques, et les composantes de la direction de déplacement \vec{db} sont calculées pour $1 \leq i \leq p-1$, à partir des formules (12), (15) ou (18) (où Q a été exprimé uniquement en fonction des variables hors base) suivant que l'on emploie la méthode de Wolfe ou de d'Esopo ; on a alors :

$$db^p = - \sum \{db^i \mid i = 1, p-1\}$$

et θ est déterminé à partir de la relation (14) de telle sorte que les contraintes (10) soient vérifiées (i.e. $\theta_1 \leq \theta \leq \theta_2$). Si b^p atteint une de ses bornes, on intervertit b^p avec un coefficient b^i appartenant à l'intervalle ouvert $]\alpha_i, \beta_i[$ (pivotage des variables i et p). Le point de départ \vec{b}_1 de la procédure doit être tel que

$$\sum \{b_1^i \mid i = 1, p\} = 1; \quad b_1^p \in]\alpha_p, \beta_p[; \quad b_1^i \in [\alpha_i, \beta_i]$$

Les algorithmes précédents ont été programmés dans les cas suivants :

- a) Méthode de Wolfe, avec les contraintes $b^i \geq 0$
- b) " de d'Esopo " " "
- c) " " " " " et $\sum \{b^i \mid i = 1, p\} = 1$

Quoiqu'exigeant un nombre d'itérations plus grand, la méthode de d'Esopo, dont les calculs sont plus simples, s'est révélée plus rapide que la méthode de Wolfe.

C'est pour cette raison que nous avons programmé cette méthode dans le cas où

$$\sum \{b^i \mid i = 1, p\} = 1$$

IV - EXEMPLES D'UTILISATION DE LA REGRESSION SOUS CONTRAINTES

1 - Problèmes de régression où la variable à expliquer est positive

On a alors des contraintes du type :

$$\forall j, 1 \leq j \leq n : \sum \{b^i x_1^j \mid i = 1, p\} \geq 0$$

C'est le cas rencontré en particulier en hydrologie (cf. réf. 5), dans les problèmes de corrélations hydro pluviométriques quand on veut expliquer l'écoulement en fonction de p variables aléatoires telles que la précipitation, le nombre de séquences pluvieuses durant l'année, etc...

2 - Problèmes de régression entre variables positives

Il est assez fréquent que les variables y, x_1, \dots, x_p soient positives ; si les coefficients de régression doivent être positifs, pour avoir un sens alors les valeurs reconstituées sont automatiquement positives (i.e. les contraintes (2) (b) où $\alpha = 0, \beta = +\infty$:

$$\Sigma \{b^i x_i^j \mid i = 1, p\} \geq 0 \text{ sont satisfaites).}$$

L'intérêt de ce modèle explicatif réside dans l'additivité des effets permettant d'approcher la variable à expliquer y par des contributions toutes de même sens, ce qui protège la régression contre les erreurs pouvant survenir quand, ayant des contributions de sens opposés, de fortes valeurs se soustraient entre elles pour fournir une explication de y qui peut être fortuite (cf. réf. 2).

Citons trois exemples où ces méthodes ont été appliquées :

- le premier est relatif à la géologie (cf. réf. 3) ;

On veut expliquer le taux d'un oligo-élément dans divers échantillons de roches par les apports de différents minéraux majeurs.

- le deuxième est relatif à une enquête alimentaire où l'on a relevé le poids d'un produit (de la viande par exemple) consommé dans n familles, et où l'on désire "expliquer" ce poids en fonction des effectifs des familles répartis en différentes classes d'âges, ce qui revient à estimer la ration moyenne de ce produit consommé par un individu d'une classe d'âge donnée.

- le troisième est relatif à certains processus physiques régis par une équation intégrale de la forme :

$$f(x) = \int N(x, y) g(y) dy$$

où N, f et g sont essentiellement positifs, N étant connu, et f déterminé expérimentalement. L'estimation de g se ramène à un problème du type précédent, comme on le verra au paragraphe suivant où est traité en détail le cas où f et g sont des densités de probabilité.

V - ESTIMATION DE LA COURBE GRANULOMETRIQUE D'UN AEROSOL (cf. réf. 6)

On se propose de rechercher la courbe granulométrique, $g(x) = \frac{dn}{dx}$ d'un nuage de particules en suspension dans un gaz quelconque, à l'aide d'une méthode photoélectrique. Celle-ci consiste à mesurer le flux lumineux diffusé par chaque particule lors de son passage dans le champ uniformément éclairé d'un photomètre (fig. 4) : l'information ainsi recueillie est fiable, par application des lois de la diffusion, à la dimension de chacune des particules.

1 - Montage expérimental

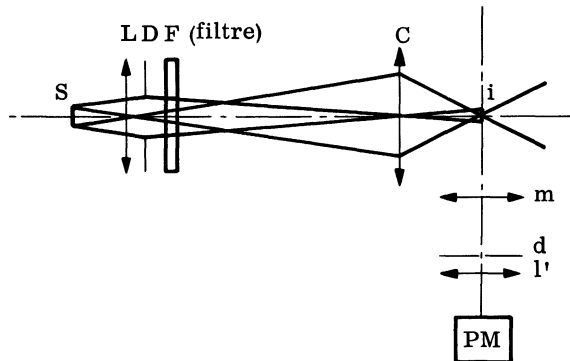


Figure 4

Un condenseur C fait l'image i d'un diaphragme D uniformément éclairé (par la source S) dans le champ d'un objectif m de microscope dont l'axe est à 90° de celui du condenseur. Le champ de mesure est également défini par l'image à travers m d'un diaphragme d, et une lentille l' conjugue la pupille d'entrée du photomètre avec un photomultiplicateur PM. Les particules à observer qui sont en suspension dans l'air, passent en i perpendiculairement au plan de la figure. Ce dispositif permet d'une part d'éviter les pertes de flux, d'autre part d'éclairer de la même façon le photodétecteur, quelle que soit la position de la particule dans le volume de mesure.

On voit que le champ a la forme d'un parallélépipède éclairé uniformément : pour que la granulométrie ait un sens, son volume est choisi de telle sorte que la probabilité d'y trouver plus d'une particule soit très inférieure à celle d'en avoir une seule. Cette probabilité est donnée par la loi :

$$p(n) = \frac{(NV)^n e^{-NV}}{n !}$$

où N est la concentration des particules en nombre par unité de volume et V le volume du champ de mesure.

Si on prend $NV = 0,1$; on aura :

$$\{\sum p(i) \mid i = 2, \infty\} / p(1) = 5,1 \%$$

2 - Détection

Cette méthode de mesure granulométrique des aérosols, maintenant bien connue, est employée couramment pour classer des particules de diamètre supérieur à $0,5 \mu$. Dans ce cas, le flux diffusé par chacune d'elles et capté dans l'ouverture du photomètre donne lieu, à l'anode du photodétecteur, à une impulsion de tension dont la durée est égale au temps de passage de la particule dans le champ, et dont l'amplitude est approximativement proportionnelle au carré de son diamètre. Actuellement, la sélection d'amplitude ne pose plus aucun problème.

Par contre dans le domaine submicronique ($d < 0,3 \mu$) le flux capté par le photomètre est suffisamment faible pour que chaque photon diffusé puisse être détecté individuellement (à condition bien sûr d'employer une électronique suffisamment rapide).

Le nombre de photons diffusé par unité de temps par une particule est aléatoire et suit une loi de Poisson dont le paramètre est fonction du diamètre de la particule. Le rendement quantique très faible (quelques %) du photodétecteur conserve la distribution temporelle poissonnienne des intervalles de temps lors de la conversion photon-électron. L'information associée à chaque particule lors de son passage dans le champ de mesure est alors constituée par une suite discontinue d'impulsions de tension, chacune étant obtenue à partir d'un seul photoélectron au départ de la photocathode du détecteur : nous sommes dans le domaine de "la spectrométrie à photoélectron unique". Toutes les méthodes de sélection d'amplitude sont inopérantes : elles doivent être remplacées par des méthodes digitales consistant à compter des nombres d'impulsions par unité de temps.

Dans l'appareil décrit précédemment, le temps de passage d'une particule est compris entre 70 et 120 μ s. Dès qu'une particule pénètre dans le champ, elle est détectée et un dispositif de comptage donne le nombre de photons émis pendant 50 μ s, valeur prise comme unité de temps. Le dispositif est ensuite bloqué pendant 70 μ s pour éviter de déclencher à nouveau une mesure sur la même particule. On obtient ainsi l'histogramme du nombre de photons émis par unité de temps par l'ensemble des particules d'un nuage.

3 - Interprétation des mesures

L'histogramme précédent est l'estimation de la loi de probabilité du nombre de photons détectés pour un échantillon de particules du nuage ; nous désignerons par $f(m)$ cette loi de probabilité, i.e. la probabilité d'obtenir m photons à partir de cet échantillon. Comme on l'a vu, le nombre de photons émis par une particule de diamètre x est aléatoire, et suit une loi de Poisson de paramètre $\alpha(x)$. Les calculs effectués à partir des lois de la diffusion et vérifiés par ailleurs expérimentalement permettent d'obtenir la courbe d'étalonnage donnant $\alpha(x)$ en fonction de x .

Si $g(x) dx$ désigne la proportion de particules du nuage ayant un diamètre compris entre x et $x + dx$, la probabilité d'obtenir m photons à partir d'une particule de cette classe est,

$$\frac{e^{-\alpha(x)} [\alpha(x)]^m}{m !} g(x) dx$$

la probabilité d'obtenir m photons à partir du nuage étant :

$$f(m) = \int \frac{e^{-\alpha(x)} [\alpha(x)]^m}{m !} g(x) dx \quad (20)$$

On divise l'histogramme précédent en n classes :

$$I_1 = (m_0, m_1), I_2 = (m_1 + 1, m_2), \dots I_n = (m_{n-1} + 1, m_n).$$

La probabilité de se trouver dans la $j^{\text{ème}}$ classe est donc :

$$f^j = \sum \{f(m) \mid m \in I_j\} = \int e^{-\alpha(x)} g(x) \sum \left\{ \frac{(\alpha(x))^m}{m!} \mid m \in I_j \right\} dx \quad (21)$$

Par ailleurs, si on découpe l'intervalle d'intégration en p classes

$(x_0, x_1), (x_1, x_2), \dots, (x_{p-1}, x_p)$, on a :

$$f^j = \sum \left\{ \int_{x_{i-1}}^{x_i} e^{-\alpha(x)} g(x) \sum \left\{ \frac{[\alpha(x)]^m}{m!} \mid m \in I_j \right\} dx \mid i = 1, p \right\} \quad (22)$$

ce qui s'écrit, en appliquant la formule de la moyenne et en désignant par b^i la proportion des particules du nuage ayant un diamètre compris entre x_{i-1} et x_i :

$$\left. \begin{aligned} f^j &= \sum \{b^i x_i^j \mid i = 1, p\} \\ \text{avec : } b^i &= \int_{x_{i-1}}^{x_i} g(x) dx \\ x_i^j &= e^{-\alpha(x_{i_j}^*)} \sum \left\{ \frac{(\alpha(x_{i_j}^*))^m}{m!} \mid m \in I_j \right\} \end{aligned} \right\} \quad (23)$$

$x_{i_j}^*$ appartenant à l'intervalle (x_{i-1}, x_i) .

f^j est estimée par k_j/k , où k_j est le nombre de particules appartenant à la $j^{\text{ème}}$ classe de l'histogramme, et k le nombre total de particules.

On a donc le modèle suivant, si on pose $y^j = k_j/k$:

$$\left. \begin{aligned} \forall j, 1 \leq j \leq n : y^j &= \sum \{b^i x_i^j \mid i = 1, p\} + e^j \\ \forall i, 1 \leq i \leq p : b^i &\geq 0 \\ \sum \{b^i \mid i = 1, p\} &= 1 \end{aligned} \right\} \quad (24)$$

Ceci implique que les b^i soient inférieurs à 1, et que les valeurs reconstituées soient positives. On désire, après avoir choisi une norme dans R^n , estimer les b^i de telle sorte que le vecteur \vec{e} de composante $\{e^j \mid j = 1, n\}$ ait une norme minimale, ce qui revient à projeter sur le convexe C_1 contenu dans le cône positif C de l'espace vectoriel H , défini par les équations (3) (où $\alpha_1 = \alpha = 0$; $\beta_1 = \beta = +\infty$) et (4) (cf. II 2).

4 - Résultats

Nous exposons ici à titre d'exemple le traitement des résultats observés lors d'une expérience particulière. Le nuage de particules est obtenu par pulvérisation pneumatique d'une solution de sphérules de latex monodispersées, de diamètre $0,23 \mu$, diluée dans de l'eau bidistillée. La dilution est calculée de telle sorte que la probabilité de trouver des agglomérats de deux ou plusieurs particules soit très faible. Dans les conditions de nos expériences, on obtient, outre des particules de latex, différents résidus d'évapora-

tion de la solution provenant en particulier de l'émulsifiant de la solution mère des sphérules en suspension : lors de la pulvérisation, l'émulsifiant forme des noyaux hygroscopiques fixant l'eau bidistillée et donnant ainsi des gouttelettes assez stables ne s'évaporant pas complètement malgré toutes les précautions expérimentales.

Nous avons découpé l'intervalle de variation du diamètre des particules (de $0,08 \mu$ à $0,40 \mu$) en huit classes, et ceci de quatre manières différentes, de façon à pouvoir vérifier, en considérant les sommes cumulées des pourcentages, que les résultats restent cohérents d'un découpage à l'autre.

$$\left. \begin{array}{l} \text{1er découpage : } 0,08 - 0,12; 0,12 - 0,16; \dots; 0,28 - 0,32; 0,32 - 0,40 \\ \text{2ème découpage : } 0,09 - 0,13; 0,13 - 0,17; \dots; 0,29 - 0,33; 0,33 - 0,40 \\ \text{3ème découpage : } 0,10 - 0,14; 0,14 - 0,18; \dots; 0,30 - 0,34; 0,34 - 0,40 \\ \text{4ème découpage : } 0,11 - 0,15; 0,15 - 0,19; \dots; 0,31 - 0,35; 0,35 - 0,40 \end{array} \right\} \quad (25)$$

En d'autres termes, la juxtaposition des quatre courbes cumulées (croissantes par construction) relatives aux différents découpages, qui donne la courbe des pourcentages cumulés estimée pour les diamètres $0,08, 0,09, 0,1, \dots, 0,34, 0,35, 0,40 \mu$ doit également être croissante.

On dispose par ailleurs de la courbe d'étalonnage donnant $\alpha(x)$ en fonction de x (fig. 5).

Enfin, pour calculer x_{1j}^* dans la la formule (23), on a approximé $\alpha(x_{1j}^*)$ par

$$[\alpha(x_{1j-1}) + \alpha(x_{1j})] / 2$$

x_{1j}^* étant le diamètre associé à l'intervalle (x_{1j-1}, x_{1j}) .

On a utilisé deux métriques :

- la première est la métrique usuelle
- la seconde tient compte de la dispersion des résultats, en introduisant des coefficients de pondération inversement proportionnels à

$$y^j(1 - y^j)$$

dans la formule de distance :

$$|| \vec{e} ||^2 = \sum \{(e^j)^2 / (y^j(1 - y^j)) \mid j = 1, n\} \quad (26)$$

Le tableau 1 montre l'histogramme du nombre d'impulsions mesuré par unité de temps : il est divisé en seize classes. Dans chacune, k_j représente le nombre de particules "tombant" dans la $j^{\text{ème}}$ classe.

- k_1 est le nombre de particules ayant donné lieu au comptage de deux impulsions.

- k_j , pour $j \neq 1$, est le nombre de particules ayant donné lieu à $2j - 1$ ou $2j$ impulsions.

Les cinq dernières classes étant d'effectifs faibles, on les a regroupées, obtenant ainsi un second histogramme à douze classes, les onze premières étant identiques à celles de l'histogramme précédent.

Sur ces deux histogrammes, on a effectué quatre régressions correspondant successivement aux quatre découpages et avec chacune des deux normes définies précédemment.

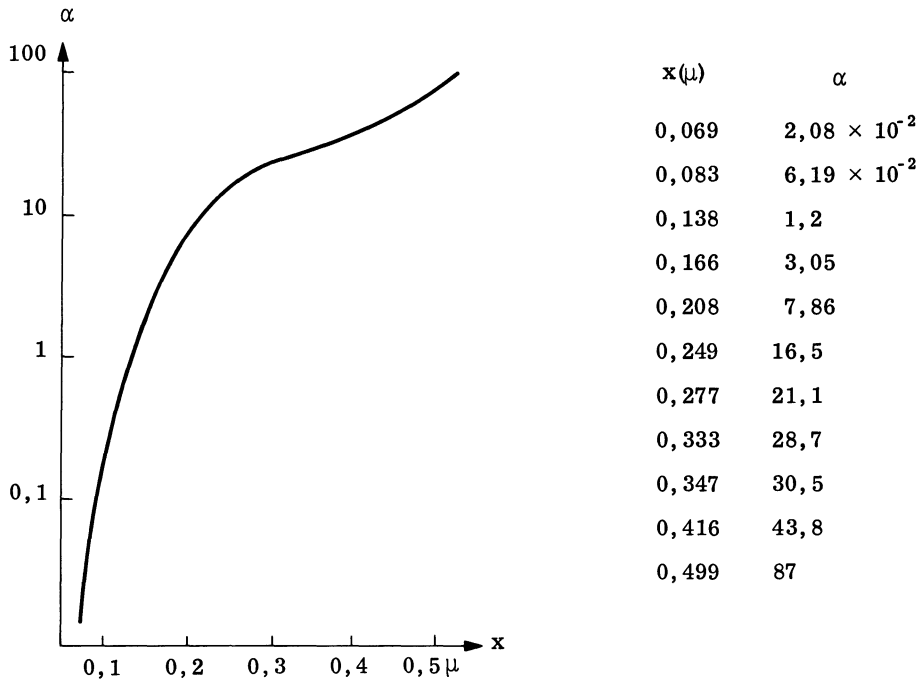


Figure 5 - Courbe d'étalonnage

Tableau 1

numéro de la classe : j	1	2	3	4	5	6	7	8
classe j	2	3-4	5-6	7-8	9-10	11-12	13-14	15-16
k_j	1228	676	455	349	554	495	624	676
numéro de la classe : j	9	10	11	12	13	14	15	16
classe j	17-18	19-20	21-22	23-24	25-26	27-28	29-30	≥ 31
k_j	588	396	197	54	29	41	30	18
						172		

Dans tous les cas, on a effectué la régression classique (projection sur l'espace vectoriel H engendré par les vecteurs \vec{x}_i (cf. II 2.1), puis en imposant des coefficients de régression positifs $\{b^i \mid i = 1, p\}$ (projection sur le cône positif C de H), et enfin en imposant de plus, que la somme des b^i soit égale à 1 (projection sur C_1).

La régression classique a toujours donné des résultats aberrants, certains b^i étant négatifs, ou ayant des valeurs absolues supérieures à un et pouvant atteindre quelques centaines et même quelques milliers. La qualité

de l'approximation R_0^2 de \vec{y} par les vecteurs \vec{x}_1 (cf. II 2.3) était très bonne car toujours supérieure à 0,97 : \vec{y} est donc pratiquement dans l'espace vectoriel H.

Le seul fait d'imposer des coefficients de régression positifs a également eu pour résultat que tous ces coefficients sont devenus inférieurs à un, leur somme étant voisine de un (comprise entre 0,93 et 1,05 dans le cas de la norme usuelle, et entre 0,75 et 0,91 dans le cas de la seconde norme) ; la qualité de l'approximation R^2 de \vec{y} par les vecteurs \vec{x}_1 était bien sûr plus faible que précédemment, mais restait bonne, supérieure dans tous les cas à 0,73.

On retrouve expérimentalement le fait déjà évoqué (cf. IV.2) qu'en imposant aux contributions d'être additives, on "protège" la régression alors que l'on obtient des résultats aberrants sans cette protection, et ceci sans améliorer pratiquement l'explication (R^2 voisin de R_0^2).

En outre, la somme des b^1 étant voisine de 1, la projection sur le convexe C est voisine de celle sur C_1 .

Les résultats obtenus en imposant en outre la contrainte (4) (projection sur C_1) sont résumés sur les tableaux 2 à 5.

Les tableaux 2 et 3 sont relatifs au premier histogramme (16 classes), On y a fait figurer :

- les estimations des pourcentages $F(z)$ de particules de diamètre inférieur ou égal à z (avec $z = x, x + 0,01, x + 0,02, x + 0,03$ selon le mode de partition), i.e. la somme cumulée des coefficients de régression pour chaque découpage.

- $R_0^2 = 1 - S_0^2$: qualité de l'approximation dans le cas de la régression classique (projection sur H).

- $R^2 = 1 - S^2$: qualité de l'approximation dans le cas où l'on impose aux coefficients de régression d'être positifs (projection sur C).

- $\sum b^1$: somme des coefficients de régression quand on projette sur C.

- R_1^2 : qualité de l'approximation quand on projette sur C_1 (coefficients de régression positifs et de somme 1).

- S_1^2 : résidu, quand on projette sur C_1 .

Le tableau 2 est obtenu en adoptant la norme usuelle, le tableau 3 en adoptant la norme définie par (26).

Les tableaux 4 et 5 sont respectivement les homologues des tableaux 2 et 3, mais sont relatifs au second histogramme (12 classes).

Pour chacun de ces tableaux, à part de faibles anomalies, intervenant sur les tableaux 2, 4, 5 pour les diamètres de 0,18 et 0,19 μ , les résultats obtenus entre les quatre découpages sont cohérents.

En ce qui concerne R_1^2 et S_1^2 , R_1^2 pouvant prendre des valeurs supérieures à 1, c'est S_1^2 qui permet de juger de l'efficacité de l'ajustement, qui sera d'autant meilleur que S_1^2 sera faible, R_1^2 étant donné à titre indicatif, pour le comparer à R^2 et R_0^2 .

On peut remarquer :

- que S_1^2 est voisin de S^2 , tout en lui étant bien sûr supérieur, et reste toujours inférieur à 0,36.

- que R_1^2 qui varie suivant les cas entre 0,75 et 1,31 est supérieur à R^2 , chaque fois que la somme des b^i (quand on projette sur C) est inférieure à 1, et qu'il lui est inférieur, si cette somme est plus grande que 1, ces résultats s'expliquant aisément, comme le montre la figure n° 6 où l'on a supposé la somme des b^i plus petite que 1.

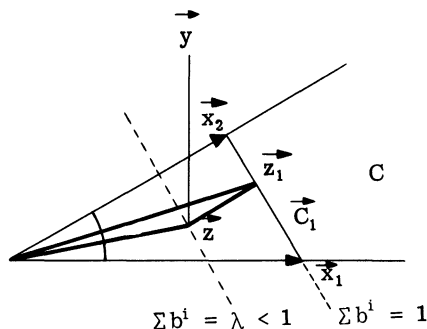


Figure 6

$$R_1^2 = \frac{\|\vec{z}_1\|^2}{\|\vec{y}\|^2} > R^2 = \frac{\|\vec{z}\|^2}{\|\vec{y}\|^2}$$

- que lorsque S_1^2 est pratiquement égal à S^2 (cas de la norme usuelle où $\Sigma b^i = 1$), il n'en est pas de même de R_1^2 et de R^2 , ce qui traduit le fait qu'au voisinage de son minimum (sur C), la somme des moindres carrés varie peu.

On pouvait s'attendre aux résultats précédents qui sont bons : ils traduisent simplement, comme on l'a signalé plus haut que la projection sur le cône positif C est très voisine de la projection sur C_1 , les résultats obtenus en projetant sur C étant très bons.

Pour comparer les différentes estimations de la courbe granulométrique de l'aérosol, nous avons reporté sur la figure 7 les valeurs indiquées dans les tableaux 2 et 4 (norme usuelle) : les deux courbes obtenues sont bien concordantes. On remarquera cependant sur les tableaux 3 et 5 que dans le cas de la norme (26) les résultats sont un peu moins satisfaisants.

Les résultats obtenus pourraient être encore améliorés par l'emploi d'un appareillage (sélecteur multi canaux) permettant d'analyser des densités d'information beaucoup plus grandes, et de mieux définir les classes extrêmes des histogrammes, dans la mesure où l'on pourra alors compter plusieurs centaines d'impulsions élémentaires pour chaque particule.

Tableau 2

Norme usuelle, 1er histogramme

x	8	12	16	20	24	28	32	40	R ₀ ²	R ²	R ₁ ²	S ₁ ²	Σ b ¹
1er découpage F(x)	0	0	17,4	36,15	61,65	100	100	100	0,994	0,80	0,75	0,20	1,04
2ème découpage F(x + 1)	0	0	31,26	38,30	78,04	100	100	100	0,992	0,87	0,81	0,13	1,05
3ème découpage F(x + 2)	0	0	33,93	41,12	93,07	100	100	100	0,993	0,88	0,91	0,12	0,98
4ème découpage F(x + 3)	0	4,53	33,84	48,51	100	100	100	100	0,995	0,82	0,84	0,18	0,99

- F(z) : % de particules de diamètre inférieur ou égal à z.

- x est en centième de microns

Tableau 3

Norme (26), 1er histogramme

x	8	12	16	20	24	28	32	40	R ₀ ²	R ²	R ₁ ²	S ₁ ²	Σ b ¹
1er découpage F(x)	0	0	0,34	29,55	65,37	100	100	100	0,98	0,74	1,28	0,33	0,76
2ème découpage F(x + 1)	0	0	4,06	30,25	83	100	100	100	0,976	0,73	1,31	0,36	0,75
3ème découpage F(x + 2)	0	0	16	35,66	97,30	100	100	100	0,984	0,81	1,19	0,23	0,82
4ème découpage F(x + 3)	0	0	26,72	47,71	100	100	100	100	0,992	0,83	1,16	0,20	0,85

- F(z) : % de particules de diamètre inférieur ou égal à z.

- x est en centième de microns.

Tableau 4

Norme usuelle, 2^{on}d histogramme

x	8	12	16	20	24	28	32	40	R ₀ ²	R ₁ ²	R ₁ ²	S ₁ ²	Σ b ¹
1er découpage F(x)	0	0	15,12	36,56	67,44	100	100	100	0,997	0,76	0,86	0,25	0,94
2ème découpage F(x + 1)	0	0	34,13	41,30	91,15	100	100	100	0,996	0,85	0,97	0,16	0,93
3ème découpage F(x + 2)	0	0	35,54	42,62	99,80	100	100	100	0,995	0,89	0,98	0,12	0,95
4ème découpage F(x + 3)	0	2,82	33,14	48,80	100	100	100	100	0,996	0,82	0,88	0,19	0,96

- F(z) : % de particules de diamètre inférieur ou égal à z.

- x est en centième de microns.

Tableau 5

Norme (26), 2^{on}d histogramme

x	8	12	16	20	24	28	32	40	R ₀ ²	R ₁ ²	R ₁ ²	S ₁ ²	Σ b ¹
1er découpage F(x)	0	0	0	28,71	62,36	100	100	100	0,995	0,76	1,24	0,29	0,79
2ème découpage F(x + 1)	0	0	23,72	38,42	93,62	100	100	100	0,995	0,80	1,22	0,25	0,81
3ème découpage F(x + 2)	0	0	30,88	41,27	100	100	100	100	0,994	0,90	1,08	0,105	0,91
4ème découpage F(x + 3)	0	0	28,59	47,36	100	100	100	100	0,994	0,86	1,12	0,16	0,88

- F(z) : % de particules de diamètre inférieur ou égal à z.

- x est en centième de microns.

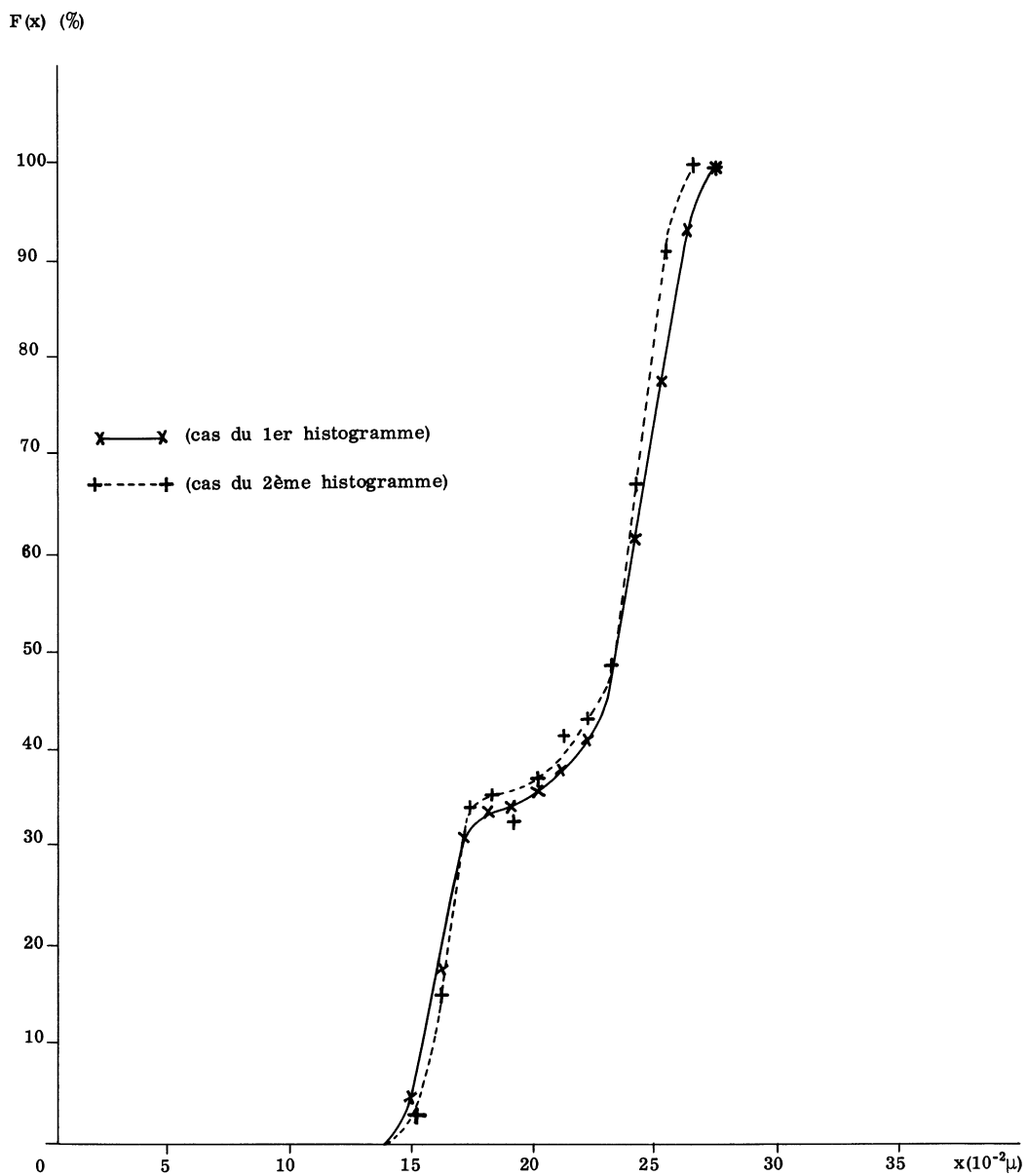


Figure 7 - Courbe granulométrique des pourcentages cumulés estimés de l'aérosol (norme usuelle).

VI - CONCLUSION

L'emploi de la régression sous contraintes permet d'obtenir des modèles d'explication satisfaisants dans de nombreux problèmes où la régression classique s'applique mal ; dans certains cas, il permet de protéger la régression contre une explication fortuite de la variable à expliquer (contributions toutes de même sens).

Ces problèmes de régression sous contraintes rentrent dans le cadre de la programmation quadratique, et leur compréhension est facilitée par une étude géométrique, où au lieu de projeter sur un espace vectoriel, comme dans la régression classique, on projette sur un convexe contenu dans cet espace.

Les résultats obtenus lors de l'estimation de la courbe granulométrique d'un aérosol, et qui ont pu être contrôlés par la méthode même de calcul, sont très bons et pourront être encore améliorés par l'emploi d'un dispositif de mesure plus perfectionné.

BIBLIOGRAPHIE

- [1] J. ABADIE - Mathématique des programmes économiques. Cours photocopié, ISUP.
- [2] J. P. BENZECRI - La régression. Publication multigraphiée du laboratoire de Statistique Mathématique (Faculté des Sciences de Paris).
- [3] P. CAZES - Application de l'analyse des données au traitement de problèmes géologiques. Thèse 3ème cycle, Faculté des Sciences de Paris, 1970.
- [4] S. KHAN - La programmation mathématique en théorie des sondages multivariates. Thèse de doctorat, 1968, Paris.
- [5] M. ROSENBERG - Problèmes de corrélations multiples avec contraintes en hydrologie. Bulletin of the International Association of Scientific Hydrology, XV, 3 9/1970.
- [6] P. Y. TURPIN - Thèse de Docteur Ingénieur, Revue d'Optique théorique et instrumentale, Juin 1967, Tome 46, n° 6.