



HAL
open science

Régression linéaire avec erreur sur les variables : application à l'étalonnage d'un gammadensimètre à transmission et d'un humidimètre à neutrons

Patrick Bertuzzi, Laurent Bruckler, Christian Gros

► To cite this version:

Patrick Bertuzzi, Laurent Bruckler, Christian Gros. Régression linéaire avec erreur sur les variables : application à l'étalonnage d'un gammadensimètre à transmission et d'un humidimètre à neutrons. *Agronomie*, 1987, 7 (7), pp.507-515. <hal-00885020>

HAL Id: hal-00885020

<https://hal.science/hal-00885020v1>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Régression linéaire avec erreur sur les variables : application à l'étalonnage d'un gammadensimètre à transmission et d'un humidimètre à neutrons

Patrick BERTUZZI, Laurent BRUCKLER, Christian GROS (*)

I.N.R.A., Station de Science du Sol, Centre de Recherches d'Avignon, Domaine Saint Paul, B.P. 91, F 84140 Montfavet

() Département de Mathématiques, Université d'Avignon, 33, rue Louis-Pasteur, F 84000 Avignon*

RÉSUMÉ

On se propose de calculer des estimateurs non biaisés de la régression linéaire dans le cas où les variables (x) et (y) de la régression, sont entachées d'erreurs. Dans une première partie, on présente les bases théoriques de la réduction du biais des estimateurs. Ces bases permettent d'aboutir à la construction d'estimateurs sans biais à l'ordre 1. De plus, on décrit la méthodologie permettant de quantifier rigoureusement l'ensemble des sources d'erreur à prendre en compte successivement pour les variables (x) et (y).

Dans une seconde partie, ces résultats théoriques sont appliqués à deux instruments de mesure nécessitant un étalonnage préliminaire, un gammadensimètre à transmission (étalonné en laboratoire) et une sonde à neutrons (étalonnée *in situ*).

Les résultats numériques montrent que les erreurs de mesure sont négligeables pour les variables (x) (taux de comptage) et (y) (masse volumique sèche) dans le cas du gammadensimètre à transmission ($\leq 0,4$ p. 100). Dans le cas de la sonde à neutrons, il en est de même pour la variable (x) (taux de comptage), mais les erreurs deviennent non négligeables (9 à 16 p. 100) pour la variable (y) (teneur en eau volumique).

Lorsque les erreurs de mesure restent faibles (cas du gammadensimètre à transmission), il y a peu de différences entre les estimateurs de la droite d'étalonnage avec ou sans prise en compte des erreurs de mesure. Par contre, les différences deviennent plus sensibles lorsque les erreurs deviennent plus importantes (cas de l'humidimètre à neutrons), notamment en ce qui concerne la précision de l'estimation de valeurs prédites à partir de l'étalonnage.

Mots clés additionnels : *Modèle Linéaire, erreurs de mesure.*

SUMMARY

Linear regression analysis with error in variables : application to the calibration of a gamma-ray attenuation probe and a neutron probe.

This paper proposes unbiased estimators of linear regression when the variables (x, y) present measurement errors. In the first part, theoretical calculations are described to obtain first-order unbiased estimators, and an error analysis is developed to take into account all the error components involved in each variable (x or y). In the second part, these theoretical results were applied to the calibration of a gamma-ray attenuation probe (laboratory calibration) and a neutron probe (field calibration). Numerical results showed that the measurement errors were negligible in the case of the gamma-ray attenuation equipment (≤ 0.4 %), but were greater (9 to 16 %) for the neutron probe, particularly for variable (y). When the measurement errors are small (gamma-ray attenuation probe), there is no difference between the parameters of the calibration estimated according to the classical approach or the proposed model. But when the measurement errors are greater (neutron probe), the differences between these two calibrations become higher especially for the precision of new predicted values.

Additional key words : *Linear model, measurement errors.*

I. INTRODUCTION

En physique du sol, la mesure « in situ » de grandeurs importantes, telles que la teneur en eau volumique ou la masse volumique sèche, fait appel à l'utilisation d'appareils de mesure nécessitant un étalonnage préliminaire. L'étalonnage d'un appareil de mesure suppose généralement l'existence d'une relation linéaire entre deux variables x et y , à un résidu aléatoire « e » près :

$$y = \alpha + \beta x + e.$$

Étalonner c'est estimer les coefficients α et β , représentant respectivement l'ordonnée à l'origine et la pente de la droite d'étalonnage. La théorie classique de la régression linéaire (TOMASSONE *et al.*, 1983) résout ce problème lorsque les mesures expérimentales du régresseur x sont connues avec exactitude. Dans la majorité des procédures d'étalonnage, cette hypothèse fondamentale n'est pas vérifiée. Les observations x_i , aussi bien que les observations y_i , sont entachées d'erreur de mesure et dans ces conditions, les estimateurs classiques introduisent un biais. Cela signifie par exemple que si l'on répétait un très grand nombre de fois l'étalonnage d'un même appareil, en un même lieu, la moyenne des estimations de la pente serait systématiquement différente de sa valeur réelle.

Comme de plus ce biais ne disparaît pas lorsque la taille de l'échantillon devient très grande, la plupart des auteurs se sont attachés à chercher des estimateurs consistants ; c'est-à-dire des estimateurs $\hat{\lambda}_n$ tels que si λ désigne le paramètre à estimer, la probabilité pour que $|\hat{\lambda}_n - \lambda|$ soit inférieure à un nombre ε donné à l'avance, tende vers 1 lorsque la taille n de l'échantillon devient infinie (KARNI & WEISSMAN, 1974 ; SCHNEEWEISS, 1976 ; WARE, 1972).

En Science du sol, HAVERKAMP *et al.* (1984) ont montré combien il était important d'analyser toutes les sources d'erreur de mesure et par conséquent d'éviter tout biais à l'étalonnage. Ces auteurs ont proposé des estimateurs qui réduisent partiellement le biais introduit par la régression linéaire classique. Nous avons montré pour notre part (GROS *et al.*, 1985) comment le biais des estimateurs de la régression linéaire en présence d'erreurs sur les variables pouvait être calculé de manière approchée au moyen d'un développement en série. Cela nous a permis de construire des estimateurs n'ayant plus qu'un biais résiduel d'ordre 2 relativement au rapport — supposé assez petit — de la variance des erreurs commises lors de la mesure du régresseur x sur la variance de ce même régresseur x .

La mise en œuvre de ces estimateurs nécessite une connaissance précise des sources d'erreur. C'est l'objet de ce travail de décrire une méthodologie d'analyse des erreurs expérimentales et de montrer ensuite, une fois l'étalonnage effectué, l'incidence de celles-ci sur l'erreur de prédiction, c'est-à-dire finalement sur l'amélioration de la précision de l'appareil de mesure considéré. Notre démarche est illustrée par deux exemples concrets : l'étalonnage d'un gammadensimètre à transmission et celui d'un humidimètre à neutrons.

II. MATÉRIELS ET MÉTHODES

A. Le gammadensimètre à transmission

Le gammadensimètre à transmission est destiné à la mesure de la masse volumique sèche du sol (STENGEL *et al.*, 1986). Le principe de la mesure est basé sur la loi d'atténuation par la matière d'un rayonnement gamma. Il consiste à introduire simultanément dans le sol, une source radioactive (^{137}Cs) émettrice de photons gamma et un détecteur, dans deux trous parallèles verticaux préalablement forés et séparés par une épaisseur de sol connue. La mesure de la masse volumique du sol revient à un comptage de photons gamma transmis à travers le sol et atteignant le détecteur.

Les expressions algébriques respectives des variables x , y , utilisées dans la procédure d'étalonnage sont reportées au tableau 1 (relations a et b). L'étalonnage est réalisé au laboratoire. Il consiste à utiliser les comptages (n_i) obtenus sur 4 blocs étalons dont la masse volumique (ρ_i) est connue (tabl. 1, relation b). Celles-ci sont normalisées par référence au coefficient d'absorption massique du calcaire (μ_0), cette correction étant nécessaire pour tenir compte de l'influence éventuelle de la nature chimique du matériau scruté.

B. L'humidimètre à neutrons

L'humidimètre à neutrons est destiné à la mesure de l'humidité volumique du sol (COUCHAT, 1977). Le principe de la mesure consiste à introduire dans le sol, une source radioactive ($^{241}\text{AmBe}$) émettrice de neutrons « rapides ». Ils entrent en collision avec les atomes d'hydrogène principalement localisés dans les molécules d'eau contenues dans le sol. Un flux de neutrons « lents », considérablement ralenti, est ainsi généré. La mesure de l'humidité volumique revient à un comptage de neutrons « lents ».

Les expressions algébriques respectives des variables x , y , utilisées dans la procédure d'étalonnage sont reportées au tableau 1 (relations c et d). La réponse de l'humidimètre à neutrons est influencée par la composition chimique et la masse volumique du sol. Ces deux facteurs rendent obligatoire la réalisation d'un étalonnage spécifique sur chaque site étudié.

En plusieurs endroits du site, convenablement choisis, une série de forages est réalisée. Sur les carottes de sol extrait, on prélève des échantillons de sol, par tranches successives de 20 cm d'épaisseur, pour la mesure du profil d'humidité pondérale (w_i) du sol. Des tubes en aluminium sont implantés dans les forages. Ceux-ci sont utilisés à la fois pour la mesure des profils :

- de masse volumique humide (ρ_i),
- des comptages neutroniques (n_i).

Les mesures sont centrées sur la tranche de sol prélevée lors du forage. Ensuite, les humidités volumiques (θ_i) sont calculées par la relation d (tabl. 1). Elles sont mises en relation avec les comptages neutroni-

TABLEAU 1
Expressions algébriques des variables (x, y).
Algebraic formulae for the variables (x, y).

Appareil	Variable	
Gammadensimètre à Transmission	x_i (a)	$LN \left(\frac{n_i}{t_i} \right)$ LN logarithme népérien n_i comptage total mesuré sur le bloc étalon i pendant le temps t_i t_i temps de comptage sur le bloc étalon i (sec)
	y_i (b)	$\rho_i = \frac{M_i}{V_i} \cdot \frac{\mu_i}{\mu_0}$ ρ_i masse volumique du bloc étalon i M_i masse du bloc étalon i (g) V_i volume du bloc étalon i (cm ³) μ_i coefficient d'absorption massique du bloc étalon i (cm ² g ⁻¹) μ_0 coefficient d'absorption massique du calcaire (cm ² g ⁻¹)
Humidimètre à Neutrons	x_i (c)	$r_i = \frac{n_i}{n_r} \cdot \frac{t_r}{t_i}$ r_i ratio de comptage n_i comptage total i mesuré pendant le temps t_i n_r comptage total milieu de référence pendant le temps t_r t_i temps de comptage mesure n_i (sec) t_r temps de comptage mesure n_r (sec)
	y_i (d)	$\theta_i = \frac{\rho_i}{\rho_w} \cdot \frac{w_i}{1 + w_i}$ θ_i humidité volumique ρ_i masse volumique humide de sol, mesure i (g cm ⁻³) ρ_w masse volumique de l'eau (g cm ⁻³) w_i humidité pondérale

ques (n_i) correspondants. Ceux-ci sont rapportés à des comptages mesurés dans un milieu standard constitué par de l'eau, de manière à définir un ratio de comptage (r_i), (tabl. 1, relation c).

sont connues exactement (ou estimées de manière indépendante).

Si l'on néglige les erreurs de mesure sur les variables ($\hat{x} = x$ et $\hat{y} = y$) et si l'on pose :

$$\hat{\beta}_0 = \frac{s(\hat{x} \cdot \hat{y})}{s^2(\hat{x})} \tag{3}$$

III. THÉORIE

A. La régression linéaire avec erreur sur les variables

Etant donné deux variables x et y liées par une relation de la forme :

$$y = \alpha + \beta x + e \tag{1}$$

où α et β désignent des coefficients constants et e un terme d'erreur aléatoire, on cherche à estimer les coefficients α et β à partir de m observations indépendantes des 2 variables x et y :

$$(\hat{x}_1, \hat{y}_1), \dots, (\hat{x}_i, \hat{y}_i), \dots, (\hat{x}_m, \hat{y}_m).$$

Chaque mesure \hat{x}_i ou \hat{y}_i comporte respectivement une erreur u_i ou v_i que l'on suppose centrée et indépendante — ainsi que sa variance — de la variable mesurée. On peut donc écrire :

$$\hat{x} = x + u \quad \text{et} \quad \hat{y} = y + v \tag{2}$$

avec $E(u) = E(v) = 0$.

On suppose en outre que les variances $V(u)$ et $V(v)$

avec les notations habituelles :

$$\bar{x} = \frac{1}{m} \sum_{i=1}^m \hat{x}_i \quad \text{et} \quad \bar{y} = \frac{1}{m} \sum_{i=1}^m \hat{y}_i \tag{4}$$

$$s(\hat{x}, \hat{y}) = \frac{1}{m-1} \sum_{i=1}^m (\hat{x}_i - \bar{x})(\hat{y}_i - \bar{y}) \tag{5}$$

$$s^2(\hat{x}) = \frac{1}{m-1} \sum_{i=1}^m (\hat{x}_i - \bar{x})^2 \tag{6}$$

alors la méthode des moindres carrés permet de montrer que $\hat{\beta}_0$ est un estimateur non biaisé de la pente de la droite d'ajustement.

En présence d'erreurs, on sait que les méthodes classiques — telle celle du maximum de vraisemblance — de recherche d'estimateur non biaisé de β , ne s'appliquent pas (JOHNSTON, 1963).

HAVERKAMP *et al.* (1984) ont proposé dans ces conditions d'estimer la pente β au moyen de l'estimateur :

$$\hat{\beta}_1 = \frac{s(\hat{x}, \hat{y})}{s^2(\hat{x}) - \frac{m-2}{m-1} V(u)} \tag{7}$$

En réalité, la vraie valeur du coefficient à estimer est :

$$\beta = \frac{s(x, y)}{s^2(x)}. \quad (8)$$

Pour étudier le biais de $\hat{\beta}_1$, calculons l'espérance mathématique de :

$$\hat{\beta} = \frac{s(\hat{x}, \hat{y})}{s^2(\hat{x}) - \lambda \cdot V(u)} \quad \text{où } \lambda \text{ désigne un paramètre.} \quad (9)$$

On remarque que $\hat{\beta}$ est un rapport de la forme $\frac{X}{Y}$ que l'on peut développer en série entière au voisinage de $(E(X), E(Y))$:

$$\begin{aligned} \hat{\beta} = \frac{X}{Y} &= \frac{E(X)}{E(Y)} + \sum_{i=1}^{\infty} \frac{(-1)^i}{E(Y)^{i+1}} \\ &\times [(X - E(X))(Y - E(Y))^i \\ &- \frac{E(X)}{E(Y)}(Y - E(Y))^{i+1}] \end{aligned} \quad (10)$$

On peut alors calculer l'espérance mathématique $\frac{X}{Y}$ et on obtient sans difficulté la formule :

$$E\left(\frac{X}{Y}\right) = \frac{E(X)}{E(Y)} \left[1 + \sum_{i=1}^{\infty} (-1)^i (v_{1i} - v_{0i+1}) \right] \quad (11)$$

où v_{rs} désigne $E\left[\left(\frac{X - E(X)}{E(X)}\right)^r \cdot \left(\frac{Y - E(Y)}{E(Y)}\right)^s\right]$.

Nous nous limiterons au terme du premier ordre de ce développement qui s'écrit :

$$E\left(\frac{X}{Y}\right) \approx \frac{E(X)}{E(Y)} \left[1 - \frac{\text{cov}(X, Y)}{E(X) \cdot E(Y)} + \frac{V(Y)}{E(Y)^2} \right]. \quad (12)$$

Appliquons cette formule à :

$$X = s(\hat{x}, \hat{y}) \quad \text{et} \quad Y = s^2(\hat{x}) - \lambda V(u).$$

Il vient successivement :

$$E(X) = \beta s^2(x) \quad (13)$$

$$E(Y) = s^2(x) + (1 - \lambda) V(u) \quad (14)$$

$$\text{cov}(X, Y) = \frac{2\beta}{m-1} s^2(x) \cdot V(u) \quad (15)$$

$$V(Y) = \frac{2V(u)}{m-1} [2s^2(x) + V(u)]. \quad (16)$$

On en déduit :

$$E(\hat{\beta}) = E\left(\frac{X}{Y}\right) = \beta \left[1 + \left(\lambda - \frac{m-3}{m-1} \right) r + 0(r) \right] \quad (17)$$

avec $r = \frac{V(u)}{s^2(x)}$ supposé petit.

Pour réduire le biais de l'estimateur de la pente β , nous proposons de choisir :

$$\lambda = \frac{m-3}{m-1}. \quad (18)$$

Ainsi $\hat{\beta}_2 = \frac{s(\hat{x}, \hat{y})}{s^2(\hat{x}) - \frac{m-3}{m-1} V(u)}$ est un estimateur

de β sans biais à l'ordre 1 relativement à r :

$$E(\hat{\beta}_2) = \beta [1 + 0(r)]. \quad (19)$$

Cela constitue une précision suffisante pour les applications.

La formule (17) permet également d'apprécier la réduction de biais ainsi obtenue par rapport à l'estimateur $\hat{\beta}_0$ (correspondant à $\lambda = 1$) obtenu par la méthode des moindres carrés lorsqu'on néglige les erreurs de mesure :

$$E(\hat{\beta}_0) = \beta \left[1 + \frac{2}{m-1} r + 0(r) \right] \quad (20)$$

et par rapport à l'estimateur $\hat{\beta}_1$ proposé par HAVERKAMP *et al.* (1984) :

$$E(\hat{\beta}_1) = \beta \left[1 + \frac{1}{m-1} r + 0(r) \right]. \quad (21)$$

Cette méthode de construction d'estimateurs sans biais à l'ordre 1 a été appliquée aux autres estimateurs de la régression linéaire.

On aboutit ainsi (tabl. 2 et 3) à une formulation nouvelle des estimateurs de la régression linéaire. En posant $V(u) = 0$, on retrouve les estimateurs classiques.

Les paramètres de l'étalonnage étant déterminés, à partir d'une valeur mesurée $\hat{x}_0 = x_0 + u_0$ du régresseur, on peut prédire la valeur :

$$\tilde{y}_0 = \hat{\alpha} + \hat{\beta} \cdot \hat{x}_0 \quad (22)$$

de la variable expliquée dont la valeur réelle inconnue est y_0 .

L'erreur ainsi commise est $\tilde{y}_0 - y_0$.

On vérifie sans difficulté que cette erreur est sans biais à l'ordre 1 et que sa variance peut être estimée par l'estimateur également sans biais à l'ordre 1 relativement à $r = \frac{V(u)}{s^2(x)}$:

$$\begin{aligned} s^2(\tilde{y}_0 - y_0) &= \hat{x}_0^2 \cdot s^2(\hat{\beta}) + 2 \hat{x}_0 \cdot s(\hat{\alpha}, \hat{\beta}) + s^2(\hat{\alpha}) + \\ &+ s^2(e) + V(u) [\hat{\beta}^2 - s^2(\hat{\beta})]. \end{aligned} \quad (23)$$

B. Analyse des erreurs expérimentales

L'application de la théorie décrite précédemment suppose la connaissance préalable des termes $V(u)$, $V(v)$. Il faut donc procéder à une analyse précise de toutes les sources d'erreur intervenant dans les m observations (\hat{x}_i, \hat{y}_i) utilisées pour l'étalonnage.

On peut alors estimer séparément les variances $V(u)$ et $V(v)$ des erreurs sur x et y par les relations (HAVERKAMP *et al.*, 1984) :

$$V(u) \approx \frac{1}{m-1} \cdot \sum_{i=1}^m s^2(\hat{x}_i) \quad (24)$$

$$V(v) \approx \frac{1}{m-1} \cdot \sum_{i=1}^m s^2(\hat{y}_i). \quad (25)$$

TABLEAU 2

Expressions algébriques des estimateurs « biaisés » de la régression linéaire.
Algebraic formulae of the "biased" estimators of the linear regression.

Estimateurs classiques de la régression linéaire	(Cas : $\hat{x} = x$, $\hat{y} = y$, $V(u) = V(v) = 0$)		
Pente	$\hat{\beta} = \frac{s(\hat{x}, \hat{y})}{s^2(\hat{x})}$		
Ordonnée à l'origine	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$	$\bar{x} = \frac{1}{m} \sum_{i=1}^m \hat{x}_i$	$\bar{y} = \frac{1}{m} \sum_{i=1}^m \hat{y}_i$
Variance résiduelle	$s^2(e) = \frac{m-1}{m-2} [s^2(\hat{y}) - \hat{\beta}^2 s^2(\hat{x})]$		
Variance de la pente	$s^2(\hat{\beta}) = \frac{1}{m-2} \left[\frac{s^2(\hat{y})}{s^2(\hat{x})} - \hat{\beta}^2 \right]$		
Variance de l'ordonnée à l'origine	$s^2(\hat{\alpha}) = s^2(e) \left[\frac{1}{m} + \frac{\bar{x}^2}{(m-1)s^2(\hat{x})} \right]$		
Covariance pente-ordonnée	$s(\hat{\alpha}, \hat{\beta}) = -\bar{x} s^2(\hat{\beta})$		
Variance de l'erreur de prédiction	$s^2(\hat{y}_0 - y_0) = \hat{x}_0^2 s^2(\hat{\beta}) + 2 \hat{x}_0 s(\hat{\alpha}, \hat{\beta}) + s^2(\hat{\alpha}) + s^2(e)$		

TABLEAU 3

Expressions algébriques des estimateurs « non biaisés » de la régression linéaire.
Algebraic formulae of the "unbiased" estimators of the linear regression.

Estimateurs non biaisés à l'ordre 1	(Cas ; $\hat{x} = x + u$, $\hat{y} = y + v$, avec $V(u)$ et $V(v)$ supposées connues)		
Pente	$\hat{\beta} = \frac{s(\hat{x}, \hat{y})}{s^2(\hat{x}) - \frac{m-3}{m-1} V(u)}$		
Ordonnée à l'origine	$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$	$\bar{x} = \frac{1}{m} \sum_{i=1}^m \hat{x}_i$,	$\bar{y} = \frac{1}{m} \sum_{i=1}^m \hat{y}_i$
Variance résiduelle	$s^2(e) = \frac{m-1}{m-2} [s^2(\hat{y}) - \hat{\beta}^2 s^2(\hat{x})] + \frac{V(u)}{m-2} \left[\frac{s^2(\hat{y})}{s^2(\hat{x})} + (m-5)\hat{\beta}^2 \right] - V(v)$		
Variance de la pente	$s^2(\hat{\beta}) = \frac{1}{m-2} \left[\frac{s^2(\hat{y})}{s^2(\hat{x}) - \frac{m-5}{m-1} V(u)} - \hat{\beta}^2 \right] + \frac{1}{(m-2)^2} \cdot \frac{V(u)}{s^2(\hat{x})} \cdot \left[(m-6) \frac{s^2(\hat{y})}{s^2(\hat{x})} + 4\hat{\beta}^2 \right]$		
Variance de l'ordonnée à l'origine	$s^2(\hat{\alpha}) = \frac{m-1}{m(m-2)} [s^2(\hat{y}) - \hat{\beta}^2 s^2(\hat{x})] + \frac{\bar{x}^2}{(m-2)} \left[\frac{s^2(\hat{y})}{s^2(\hat{x}) - \frac{m-5}{m-1} V(u)} - \hat{\beta}^2 \right] + \frac{V(u)}{m-2} \left[\frac{2(m-3)}{m} \hat{\beta}^2 + \frac{1}{m-2} \frac{\bar{x}^2}{s^2(\hat{x})} \left(4\hat{\beta}^2 + (m-6) \frac{s^2(\hat{y})}{s^2(\hat{x})} \right) \right]$		
Covariance pente-ordonnée	$s(\hat{\alpha}, \hat{\beta}) = -\bar{x} s^2(\hat{\beta})$		
Variance de l'erreur de prédiction	$s^2(\hat{y}_0 - y_0) = \hat{x}_0^2 s^2(\hat{\beta}) + 2 \hat{x}_0 s(\hat{\alpha}, \hat{\beta}) + s^2(\hat{\alpha}) + [\hat{\beta}^2 - s^2(\hat{\beta})] s^2(u) + s^2(e)$		

Nous supposons que les valeurs ainsi estimées sont les vraies valeurs de $V(u)$ et $V(v)$.

Le calcul des variances des estimateurs $s^2(\hat{x}_i)$, $s^2(\hat{y}_i)$

est réalisé de la manière suivante : si Z désigne une variable aléatoire quelconque, elle-même fonction de « p » variables aléatoires indépendantes (a, b, \dots, d),

la variance de Z peut être estimée par la relation suivante :

$$s^2(Z) \approx \left(\frac{\partial Z}{\partial a} \right)^2 \cdot s^2(a) + \dots + \left(\frac{\partial Z}{\partial d} \right)^2 \cdot s^2(d), \quad (26)$$

à condition que les écarts-types $s(a)$, $s(b)$, ..., $s(d)$ soient petits respectivement par rapport aux termes a , b , ..., d .

Dans notre cas, la variable aléatoire Z correspondra successivement aux variables \hat{x}_i , \hat{y}_i avec $i = 1, \dots, m$. En utilisant la relation générale (26), les expressions algébriques des variances $s^2(\hat{x}_i)$, $s^2(\hat{y}_i)$ sont reportées dans le tableau 4.

Cette méthodologie permet finalement de dégager les propriétés structurales de l'erreur. Le calcul des contributions, en pourcentage, de chaque terme à la variance totale de Z permet de mettre en évidence les termes prépondérants de l'erreur.

IV. RÉSULTATS ET DISCUSSION

A. Analyse des erreurs de mesure

Le calcul des erreurs de mesure suppose que les termes $s^2(a)$, $s^2(b)$, ..., $s^2(d)$ soient connus, ou du moins estimés avec un ordre de grandeur compatible avec la réalité expérimentale. Les valeurs retenues pour chaque appareil de mesure sont données par le tableau 5.

En utilisant ces valeurs numériques (tabl. 5) et les résultats algébriques obtenus précédemment (tabl. 4), les résultats du calcul final des erreurs de mesure estimées pour chaque appareil sont donnés par le tableau 6. Pour ce calcul, les critères retenus pour quantifier et comparer les résultats sont les suivants :

- l'écart-type estimé $s(Z)$,
- l'incertitude relative estimée

$$IR (\%) = \frac{2 \cdot s(Z)}{Z} \times 100,$$

TABLEAU 4

Expressions algébriques des erreurs de mesures sur les variables (x, y).
Algebraic formulae for the measurement errors of the variables (x, y).

Appareil	Variable	Expression algébrique des estimateurs des variances
Gammadensimètre à Transmission	x_i (1), (2)	$s^2(x_i) = \frac{1}{n_i}$
	y_i	$s^2(y_i) = \left(\frac{\mu_i}{\mu_0} \right)^2 \cdot \left(\frac{M_i}{V_i} \right)^2 \cdot \left[\frac{1}{M_i^2} s^2(M_i) + \frac{1}{V_i^2} s^2(V_i) + \frac{1}{\mu_i^2} s^2(\mu_i) + \frac{1}{\mu_0^2} s^2(\mu_0) \right]$
Humidimètre à Neutrons	x_i (1), (2)	$s^2(x_i) = \left(\frac{n_i \cdot t_r}{n_r \cdot t_i} \right)^2 \left(\frac{1}{n_i} + \frac{1}{n_r} \right)$
	y_i (3)	$s^2(y_i) = \frac{1}{(1 + w_i)^2} \left[\frac{1}{\rho_w^2} \cdot w_i^2 \cdot s^2(\rho_i) + \left(\frac{\rho_i}{\rho_w} \right)^2 \cdot \frac{1}{(1 + w_i)^2} \cdot s^2(w_i) \right]$

(1) Les comptages suivent une loi de Poisson pour laquelle on a $s^2(n_i) = n_i$.

(2) Les termes (t_i) ou (t_r) sont considérés comme une constante dans le calcul.

(3) Le terme (ρ_w) est considéré comme une constante dans le calcul.

TABLEAU 5

Variances estimées des erreurs de mesure pour les variables (x, y).
Estimates of associated error sources for the variables (x, y).

Appareil	Variable	Source d'erreur	Variance de l'erreur	Ecart-type de l'erreur
Gammadensimètre à transmission	x_i	n_i (impulsions)	300 000-500 000	(2) 548-707
		M_i (g)	100	(1) 10
	y_i	V_i (cm ³)	25	(1) 5
		μ_i (cm ² g ⁻¹)	10 ⁻⁸	(1) 10 ⁻⁴
		μ_0 (cm ² g ⁻¹)	10 ⁻⁸	(1) 10 ⁻⁴
Humidimètre à neutrons	x_i	n_i (impulsions)	10 500-17 500	(2) 102-132
		n_r (impulsions)	71 000-72 000	(2) 266-268
	y_i	ρ_i (g cm ⁻³)	0,46-1,35 · 10 ⁻⁴	(2) 0,7-1,2 · 10 ⁻²
		w_i (g · g ⁻¹)	10 ⁻⁴	(1) 10 ⁻²

(1) Variance indépendante de la valeur observée.

(2) Variance dépendante de la valeur observée.

au seuil de confiance de 5 p. 100 et en supposant que l'on est proche de la normalité.

Dans le cas des appareils utilisés, les valeurs mesurées \hat{x}_i proviennent du comptage d'un grand nombre d'impulsions ($n_i > 10\ 000$). Les incertitudes relatives sur les comptages observés sont petites, généralement inférieures à 2 p. 100. Les erreurs expérimentales de mesure sont négligeables sur les observations \hat{x}_i (tabl. 6).

Les erreurs de mesure sur \hat{y}_i ne sont pas du même ordre de grandeur pour les deux appareils étudiés :

Pour le gammadensimètre à transmission, l'étalonnage est réalisé au laboratoire. Il implique l'utilisation de blocs étalons dont les caractéristiques géométriques et massiques sont connues avec précision (tabl. 5). Les incertitudes relatives sur leur masse volumique sont finalement inférieures à 0,5 p. 100 (tabl. 6).

Dans le cas de l'humidimètre à neutrons, la procédure d'étalonnage est plus complexe. Réalisée sur le terrain, elle est la source d'erreurs non négligeables sur les valeurs des humidités volumiques calculées. Les incertitudes relatives obtenues sont comprises entre 9 et 16 p. 100 selon la valeur de l'humidité volumique (tabl. 6). A titre indicatif, en supposant que l'on est proche de la normalité, l'intervalle de confiance de la mesure est de l'ordre de $\pm 0,032$ à $\pm 0,036$ (au seuil de confiance de 5 p. 100), ordre de grandeur que nous considérons comme non négligeable, compte tenu de la précision souhaitée des mesures dans une procédure d'étalonnage.

B. Relations d'étalonnage

Pour chaque appareil de mesure, nous avons construit deux relations d'étalonnage :

— relation « biaisée » d'étalonnage construite à partir de l'utilisation des estimateurs classiques de la régression linéaire (tabl. 2) ;

— relation « non biaisée » d'étalonnage construite à partir de l'utilisation des estimateurs « non biaisés » de la régression linéaire avec erreur (tabl. 3).

1. Cas du gammadensimètre à transmission

L'analyse des erreurs expérimentales sur les observations a montré que celles-ci sont négligeables dans les conditions expérimentales de l'étalonnage. C'est pourquoi, il n'y a pas de différence sensible dans les estimations « biaisées » ou « non biaisées » des paramètres de l'étalonnage. Le choix, *a posteriori*, des estimateurs de la régression linéaire avec erreur ne se justifie pas pratiquement pour cet appareil de mesure (tabl. 7).

2. Cas de l'humidimètre à neutrons

Dans les conditions expérimentales de l'étalonnage, les erreurs de mesure sur la variable x sont petites. Compte tenu de la gamme étendue d'observation des variables (x, y) le terme V(u) influence peu le calcul de l'estimation « non biaisée » de la pente $\hat{\beta}$ et par conséquent, de l'ordonnée à l'origine $\hat{\alpha}$: il n'y a donc pas de différence sensible dans les valeurs numériques de ces deux paramètres, selon le mode d'estimation (tabl. 8, fig. 1).

Sur la figure 2 sont reportés les écarts-types $s(\hat{\theta}_0)$ de la prédiction d'une gamme croissante de valeurs d'humidité volumique ($\hat{\theta}_0$) prédites par la relation d'étalonnage. Les écarts-types obtenus par les deux modes d'estimation sont du même ordre de grandeur, mais ils sont néanmoins systématiquement plus petits dans le cadre de l'utilisation des estimations « non biaisées » des paramètres de la régression linéaire.

Il n'est pas indifférent, *a posteriori*, de retenir le mode d'estimation qui garantit la meilleure précision de la mesure, dans l'hypothèse où, par exemple, les mesures d'humidités volumiques sont utilisées pour le calcul du bilan hydrique d'un sol (HAVERKAMP *et al.*, 1984).

En comparant, pour un même type d'appareil, nos résultats avec ceux de HAVERKAMP *et al.* (1984), il apparaît que les conclusions sont similaires, tant sur un plan qualitatif que quantitatif. Il n'en demeure pas moins cependant, que les estimateurs proposés ici sont

TABLEAU 6
Analyse des erreurs de mesure pour les variables (x, y).
Error analysis of the variables (x, y).

Appareil	Variable	Contribution source d'erreur		Erreur expérimentale sur la mesure		
		Source d'erreur	Contribution %	Variance totale	Ecart-type total	IR (%)
Gammadensimètre à transmission	x_i	n_i	100	$1,9-3,5 \cdot 10^{-6}$	$1,3-1,9 \cdot 10^{-3}$	0,02-0,04
		M_i	0,0-8,0			
	y_i	V_i	0,0-3,0	$0,3-2,5 \cdot 10^{-5}$	$1,7-5,0 \cdot 10^{-3}$	0,35-0,40
		μ_i μ_0	40,0-48,0 45,0-50,0			
Humidimètre à neutrons	x_i	n_i	61-79	$3,3-6,6 \cdot 10^{-5}$	$0,6-0,8 \cdot 10^{-2}$	1,9-2,5
		n_r	21-39			
	y_i	ρ_i	18-41	$2,4-3,3 \cdot 10^{-4}$	$1,6-1,8 \cdot 10^{-2}$	9,0-16,0
		w_i	59-82			

TABLEAU 7

Valeurs numériques des paramètres de l'étalonnage. Comparaison des estimations « biaisée » et « non biaisée ».
Cas du gammadensimètre à transmission.

Estimated parameters of the calibration line. Comparison between "biased" and "unbiased" estimates for the gamma-ray attenuation probe.

Estimation		« Biaisée »	« Non Biaisée »
Coefficients	$\hat{\beta}$	- 0,986	- 0,986
	$\hat{\alpha}$	9,487	9,487
	$s^2(\hat{\beta})$	$4,153 \cdot 10^{-4}$	$4,152 \cdot 10^{-4}$
	$s^2(\hat{\alpha})$	$2,583 \cdot 10^{-2}$	$2,582 \cdot 10^{-2}$
	$s(\hat{\alpha}, \hat{\beta})$	$- 3,266 \cdot 10^{-3}$	$- 3,265 \cdot 10^{-3}$
	$s^2(e)$	$5,730 \cdot 10^{-4}$	$4,542 \cdot 10^{-4}$
Corrélation	r	- 0,999	- 0,999

Nombre couples (\hat{x}_i, \hat{y}_i)	Variable	Gamme de variation	Erreur de mesure
4	x_i	$7,129 < x < 8,510$	$V(u) = 3,089 \cdot 10^{-6}$
	y_i	$1,107 < x < 2,486$	$V(v) = 1,156 \cdot 10^{-4}$

TABLEAU 8

Valeurs numériques des paramètres de l'étalonnage. Comparaison des estimations « biaisée » et « non biaisée ».
Cas de l'humidimètre à neutrons.

Estimated parameters of the calibration line. Comparison between "biased" and "unbiased" estimates for the neutron probe.

Estimation		« Biaisée »	« Non Biaisée »
Coefficients	$\hat{\beta}$	0,812	0,824
	$\hat{\alpha}$	- 0,258	- 0,267
	$s^2(\hat{\beta})$	$0,589 \cdot 10^{-2}$	$0,604 \cdot 10^{-2}$
	$s^2(\hat{\alpha})$	$0,297 \cdot 10^{-2}$	$0,305 \cdot 10^{-2}$
	$s(\hat{\alpha}, \hat{\beta})$	$- 0,417 \cdot 10^{-2}$	$- 0,428 \cdot 10^{-2}$
	$s^2(e)$	$0,723 \cdot 10^{-3}$	$0,384 \cdot 10^{-3}$
Corrélation	r	- 0,863	- 0,930

Nombre couples (\hat{x}_i, \hat{y}_i)	Variable	Gamme de variation	Erreur de mesure
39	x_i	$0,528 < x < 0,752$	$V(u) = 0,500 \cdot 10^{-4}$
	y_i	$0,193 < x < 0,402$	$V(v) = 0,306 \cdot 10^{-3}$

différents de ceux proposés antérieurement, et que ces différences peuvent devenir quantitativement importantes pour les faibles valeurs du nombre d'observations « m ». En conséquence, le caractère similaire des résultats numériques obtenus n'a pas nécessairement une portée générale.

V. CONCLUSION

Notre démarche s'inscrit dans la continuité de nombreux travaux statistiques théoriques sur la régression linéaire. HAVERKAMP *et al.* (1984) ont ouvert la voie de leur application à la Science du sol. Nous nous

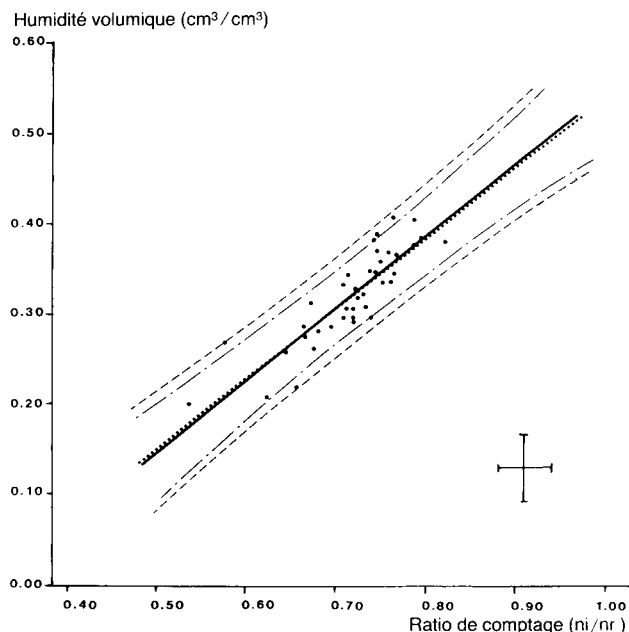


Figure 1
 Comparaison des relations d'étalonnage.
 Cas de l'humidimètre à neutrons.

- *Points expérimentaux.*
 Les barres horizontales et verticales correspondent en moyenne à ± 2 écarts-types ($s(\hat{x}_i)$, $s(\hat{y}_i)$) associés au couple d'observations (\hat{x}_i , \hat{y}_i).
- *Régression « Biaisée ».*
- *Régression « Non Biaisée ».*
- - - *Intervalle de confiance Régression « Biaisée » ($\alpha = 0,05$).*
- — - *Intervalle de confiance Régression « Non Biaisée » ($\alpha = 0,05$).*

Comparison between the calibration curves for the neutron probe.

- *Experimental values.*
 The horizontal and vertical bars correspond to ± 2 mean standard deviation ($s(\hat{x}_i)$, $s(\hat{y}_i)$) associated to each couple of observations (\hat{x}_i , \hat{y}_i).
- *“Biased” regression.*
- *“Unbiased” regression.*
- - - *Confidence limits of “Biased” regression ($\alpha = 0.05$).*
- — - *Confidence limits of “Unbiased” regression ($\alpha = 0.05$).*

sommes attachés, pour notre part, à développer une méthode rigoureuse de recherche approchée d'estimateurs sans biais des différents coefficients de la régression linéaire en présence d'erreurs de mesure. Nous montrons comment des estimateurs, sans biais, à

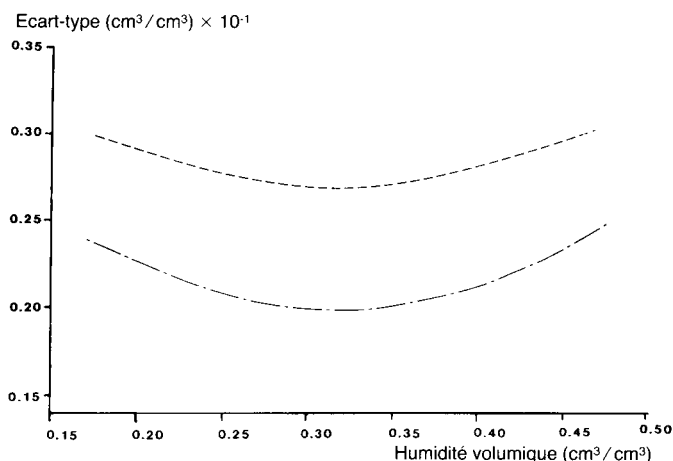


Figure 2
 Comparaison des écarts-types de la prédiction d'une gamme croissante d'humidités volumiques.

- *Ecart-type Régression « Biaisée ».*
 - *Ecart-type Régression « Non Biaisée ».*
- Comparison between standard deviations as function of predicted values of volumetric water content.
- *Standard deviation “Biased” regression.*
 - *Standard deviation “Unbiased” regression.*

l'ordre 1, permettent d'améliorer la qualité de l'étalonnage.

Du point de vue de l'utilisateur, il est certainement intéressant de recourir au modèle plus rigoureux de régression avec erreur, car il permet d'améliorer la précision de l'estimation. Néanmoins, cet avantage ne devient significatif que pour des erreurs de mesure dont les ordres de grandeurs sont déjà notables (cas de l'humidimètre à neutrons). Dans les autres cas, les différences entre les modes d'estimation sont peu sensibles (cas du gammadensimètre). Dans tous les cas, l'analyse exhaustive de la décomposition de la variance totale relative à une variable quelconque apparaît comme une étape indispensable à la connaissance préliminaire de l'ordre de grandeur de l'erreur, et de sa structure : cette étape méthodologique doit ainsi constituer une phase préliminaire à toute analyse de données proprement dite.

Reçu le 7 février 1986.
 Accepté le 30 mars 1987.

RÉFÉRENCES BIBLIOGRAPHIQUES

Couchat Ph., 1977. Aspects méthodologiques et technologiques de la mesure neutronique de l'humidité des sols. *Ann. Agron.*, 28 (5), 77-488.

Gros C., Bertuzzi P., Bruckler L., 1985. Influence des erreurs de mesure dans un modèle linéaire. Application à l'étalonnage d'une sonde à neutrons. *Compte-Rendu des Quatrième Journées Internationales « Analyse des Données et Informatique »*. 9-10-11 octobre 1985. North Holland Publishing Company.

Haverkamp R., Vauclin M., Vachaud C., 1984. Error analysis in estimating soil water content from neutron probe measurements. 1 - Local standpoint. *Soil Sci.*, 137 (2), 78-90. 2 - Spatial standpoint. *Soil Sci.*, 138 (3), 141-148.

Johnston J., 1963. *Econometric Methods*. McGraw-Hill Book Company, New York, 148-176.

Karni E., Weissman I., 1974. A consistent estimator of the slope in a regression model with errors in the variables. *J. Am. statist. Assoc.*, 69 (345), 211-213.

Schneeweiss H., 1976. Consistent estimation of a regression with errors in the variables. *Metrika*, 23 (2) 101-115.

Stengel P., Gabilly Y., Bertuzzi P., Gaudu J. C., Brifault A., 1986. La double sonde LPC-INRA. *Bull. liaison Lab. P. et Ch.*, 141 (janvier-février).

Tomassone R., Lesquoy E., Millier C., 1983. *La Régression. Nouveaux regards sur une ancienne méthode statistique*. Ed. Masson, Paris.

Ware J. H., 1972. The fitting of straight lines when both variables are subject to error and the ranks of the means are known. *J. Am. statist. Assoc.*, 67 (340), 891-897.