JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

ANTOINE DE FALGUEROLLES
DIDIER PINCHON

Une commémoration du bicentenaire de la publication (1805-1806) de la méthode des moindres quarrés par Adrien Marie Legendre

Journal de la société française de statistique, tome 147, n° 2 (2006), p. 81-105

<http://www.numdam.org/item?id=JSFS_2006__147_2_81_0>

© Société française de statistique, 2006, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) 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/

UNE COMMÉMORATION DU BICENTENAIRE DE LA PUBLICATION (1805-1806) DE LA MÉTHODE DES MOINDRES QUARRÉS PAR ADRIEN MARIE LEGENDRE*

Antoine de FALGUEROLLES**, Didier PINCHON***

RÉSUMÉ

La méthode des moindres carrés et ses avatars classiques, moindres carrés généralisés, moindres carrés pondérés itérés, et moindres carrés (pondérés itérés) alternés sous-tendent un grand nombre d'analyses statistiques couramment pratiquées de nos jours. Il semble généralement accepté qu'Adrien Marie Legendre (1752-1833) est bien l'auteur de la première publication sur la méthode des moindres carrés, méthode dont il a forgé le nom (1805). Mais, il semble aussi souvent accepté que Carl Friedrich Gauss (1777-1855) en fut l'inventeur dès 1795 comme il l'a toujours revendiqué. Il n'est pas inopportun de commémorer, deux cents ans après, l'ingéniosité et la maîtrise de Legendre en reprenant l'exemple illustratif qu'il a donné de « sa » méthode. Comme on pourra le constater, Legendre pourrait bien avoir aussi publié le premier exemple de moindres carrés généralisés et l'application qu'il présente fait de lui l'un des premiers géostatisticiens.

Mots clés : Moindres carrés (moindres quarrés), moindres carrés généralisés, moindres carrés alternés, moindres carrés itérés, modèle linéaire, analyse de covariance, modèle bilinéaire ou bi-additif.

ABSTRACT

The least squares method and its standard variations such as generalised least squares, iteratively weighted least squares and alternated (iteratively weighted) least squares underlie a large number of statistical analyses routinely performed nowadays. It seems generally accepted that Adrien Marie Legendre (1752-1833) is the author of the first publication on the method of least squares, of which he coined the name (1805). But it is also often accepted that Carl Friedrich Gauss (1777-1855) had invented it since 1795 as he always claimed. It is not unappropriate, two centuries later, to commemorate the ingenuity and mastery of Legendre by revisiting the illustrative example he gave to "his" method. It turns out that Legendre may have

* Les auteurs remercient vivement Steve Stigler pour ses commentaires encourageants, les Archives départementales du Bas-Rhin pour leur aide et les deux arbitres pour leurs remarques pertinentes.

** Laboratoire de Statistique et Probabilités (LSP), Université de Toulouse III (Paul Sabatier), 118, route de Narbonne, 31062 Toulouse Cedex 9.

*** Laboratoire de Mathématiques pour l'Industrie et la Physique (MIP), Université de Toulouse III (Paul Sabatier), 118, route de Narbonne, 31062 Toulouse Cedex 9.

also published the first paper on generalized least squares and that his application qualifies him as an early geostatistician.

Keywords : Least squares, generalized least squares, alternated least squares, iterated least squares, linear model, covariance analysis, bilinear or bi-additive model.

1. Introduction

« La règle par laquelle on prend le milieu entre les résultats de différentes observations, n'est qu'une conséquence très-simple de notre méthode générale que nous appellerons *Méthode des moindres quarrés.* »

En 1805, Adrien Marie Legendre publiait sous forme d'annexe à un petit livre une méthode permettant de trouver une solution moyenne d'un système incompatible d'équations linéaires (Legendre, 1805). Cette annexe était publiée à nouveau en 1806 et *circa* 1830 (Legendre, 1806, ca 1830). Ces deux dernières sont disponibles sur le réseau international (*internet*) [1].

En cette période de la fin du XVIII^{e} siècle et du début du XIX^{e} siècle, les besoins d'une méthode simple de recherche de solutions acceptables à des systèmes d'équations linéaires, théoriquement compatibles mais numériquement incompatibles, étaient clairement ressentis. De nombreux chercheurs et notamment Rogerius Josephus Boscovich (1711-1787), Johann Tobias Mayer (1723-1762), Pierre-Simon Laplace (1749-1829), et Carl Friedrich Gauss (1777-1855) étudiaient activement cette question (Stigler, 1986 et 1999, Stewart dans sa postface à l'édition de la *Theoria combinationis observationum erroribus minimis obnoxiae* de Gauss,1995, et Farebrother, 1999). Le mérite de Legendre (1752-1833) est d'y avoir apporté une solution générale simple, de l'avoir exprimée en termes clairs, de l'avoir publiée et fait connaître.

C'est l'exemple d'application de la méthode tel qu'il figure dans la présentation de 1806 qui fait l'objet essentiel de cette publication. En effet, dans un langage contemporain, Legendre donne là un intéressant exemple de régression linéaire généralisée (erreurs à modèle de type moyenne mobile d'ordre 1). La démarche, élégante, n'est pas sans rappeler celle utilisée un siècle et demi après dans des situations où les erreurs suivent un modèle autorégressif d'ordre 1 (Durbin et Watson, 1950, 1951). Cet aspect est donc développé dans cet article mais il est reporté en appendice B.

De par sa simplicité et sa flexibilité, la méthode des moindres carrés était promise à un bel avenir. Mais avec quel délai ? Il est toujours difficile de mesurer le laps de temps mis pour qu'une méthode sorte réellement du cercle étroit de ses inventeurs. D'ailleurs, n'avait-elle pas déjà un passé puisque

1. La publication de 1806 est disponible grâce à l'obligeance du Service interétablissements de coopération documentaire de Toulouse (SICDT) et de l'Observatoire Midi-Pyrénées, la seconde grâce à celle de la bibliothèque numérique « gallica ». Elles sont aussi disponibles à l'adresse suivante : http ://www.lsp.ups-tlse.fr/Fp/Falguerolles/FACSIMILE/index.html

Carl Friedrich Gauss publiait en 1809 qu'il en avait eu l'idée en 1795 ! Mais qu'en est-il pour le grand public ? L'espace d'une génération pourrait en être une estimation raisonnable. Un exemple publié en 1834 indique que l'ajustement d'une droite par la méthode des moindres carrés était alors parfaitement maîtrisé par un jeune ingénieur du corps des ponts et chaussées, Georges Muntz [2] (1807-1887). L'exemple dépasse encore une fois la simple application d'une méthode puisque Muntz s'attaque à la résolution d'un problème d'analyse de la covariance en des temps où l'idée de considérer une variable indicatrice comme une variable statistique classique était encore inconnue (Muntz, 1834). Cet exemple est aussi réexaminé dans cet article commémoratif.

Pour séparer les aspects historiques des aspects techniques ou des prolongements contemporains, ces derniers sont présentés sous forme d'appendices à cet article. Ainsi les formules utilisées dans le problème étudié par Legendre sont revues et vérifiées dans l'appendice A. L'explicitation des liens entre deux situations maintenant classiques de dépendance temporelle ou spatiale (AR, MA) fait l'objet de l'appendice B. Enfin, quelques avatars des moindres carrés applicables aux données de Legendre sont décrits dans l'appendice C ; ces procédures justifient pleinement la « fécondité » que Legendre reconnaissait aux « moindres quarrés ».

2. Le méridien de Paris

L'exemple traité par Legendre concerne la mesure du méridien de Paris et donc touche à la définition du mètre. L'histoire de la quête des données nécessaires à sa mesure est magistralement contée dans les ouvrages de Denis Guedj (1986) et de Ken Alder [3] (2002, traduction française en 2005). Les mesures du méridien effectuées (en France et ailleurs) sont liées à des interrogations récurrentes du monde savant au XVII^{e} et XVIII^{e} siècle sur la forme de notre planète Terre. Pour ne citer que les Français, mentionnons Jean Picard (1620-1682), Jean Richer (1630-1696), Pierre Bouguer (1698-1758), Pierre Louis Moreau de Maupertuis (1698-1759), Charles-Marie de La Condamine (1701-1774) ... Pour une description des prolongements récents de cette modélisation on pourra se rapporter très utilement à l'article « commémoratif » de Georges Balmino (2005) [4]. Nous ne rappelons ici (et dans l'appendice A) que la modélisation admise au tournant du XVIII^{e} siècle. Puis, nous reproduisons les données utilisées par Legendre.

2. Ou encore Müntz selon la graphie retenue par la Fédération des Sociétés d'histoire et d'archéologie d'Alsace dans le *Nouveau dictionnaire de biographie alsacienne* (1996) et par Léon Lalanne (1850).

3. Le premier auteur de cet article remercie John Gower de lui avoir fait connaître cet ouvrage et profite de l'occasion pour lui exprimer sa gratitude pour ses encouragements à étudier les modèles que John appelle « bi-additifs » et que l'auteur persiste à appeler modèles « bi-linéaires ». Voir appendice C.

4. Le premier auteur de cet article s'excuse des erreurs typographiques qu'il avait introduites dans le fac-similé qu'il a réalisé de l'appendice de Legendre, erreurs que Balmino a facilement identifiées dans son article (et attribuées au typographe de Legendre).

2.1. Le modèle de la Terre

L'approximation usuelle de la forme de la Terre est celle d'un ellipsoïde de révolution aplati aux pôles et d'axe de révolution la ligne des pôles. Les dimensions de l'ellipsoïde terrestre sont classiquement définies par la longueur a de son grand axe (ou demi-axe équatorial) et par son aplatissement $\dfrac{a-b}{a}$, b désignant le demi-petit axe (ou demi-axe polaire). De nos jours, on admet que $a = 6378140\ m$, $\dfrac{a-b}{a} = \dfrac{1}{298{,}257}$, et donc que $b = 6356755\ m$. Le paramètre d'aplatissement est noté α ; lui est associé le paramètre d'excentricité e défini par

$$e^2 = \frac{a^2 - b^2}{a^2} = 2\alpha - \alpha^2$$

dont la valeur admise de nos jours est 0,08181922.

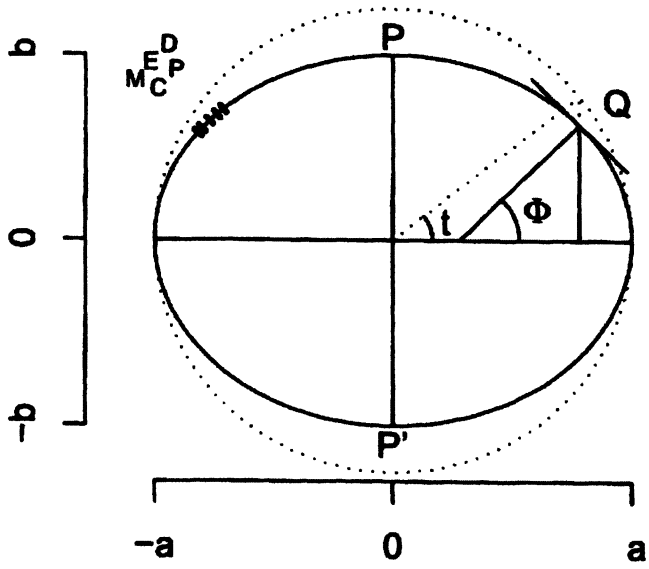

FIG 1. — Représentation schématique de la forme de la Terre : l'excentricité de l'ellipsoïde de révolution est ici égale 0,6 alors que l'excentricité « admise » est approximativement égale à 0,082, (autrement dit, aplatissement égal à $\dfrac{1}{5}$ alors que l'aplatissement « admis » est approximativement égal à $\dfrac{1}{298}$). Côté gauche, les points D, P, E, C, et M figurent Dunkerque, le Panthéon à Paris, Évaux, Carcassonne, et Montjouy sur l'analogue du méridien de Paris ; côté droit, le point Q représente un lieu de 45 degrés de latitude ($\Phi = \dfrac{\pi}{4}$).

L'intersection de l'ellipsoïde terrestre et d'un plan passant par la ligne des pôles définit une ellipse dont la forme polaire est donnée par :

$$\begin{cases} x(t) &= a\cos t \\ y(t) &= b\sin t \end{cases}$$

où t désigne l'anomalie excentrique.

Une autre représentation paramétrique utilise comme paramètre la latitude géodésique Φ (angle que fait l'équateur avec la verticale du lieu) :

$$\begin{cases} X(\Phi) &= a\dfrac{\cos\Phi}{\sqrt{1-e^2\sin^2\Phi}} \\ Y(\Phi) &= a\dfrac{(1-\alpha)^2\sin\Phi}{\sqrt{1-e^2\sin^2\Phi}} \end{cases}$$

les deux paramètres t et Φ étant liés de la façon suivante $t = \Phi$ si $t = \frac{\pi}{2} + k\pi$ avec k entier relatif, et si $t = t_0 + k\pi$ avec $-\frac{\pi}{2} < t_0 < \frac{\pi}{2}$ alors $\Phi = \Phi_0 + k\pi$ et

$$\tan t_0 = (1-\alpha)\tan\Phi_0 \,.$$

L'articulation de a, b, t, et Φ est représentée de façon schématique dans la figure 1.

Le calcul de la longueur algébrique d'un arc d'ellipse compris entre les points de paramètres t_1 et t_2 pour la première représentation paramétrique est donné par [5] :

$$\begin{aligned} s &= \int_{t_1}^{t_2} \sqrt{x'(t)^2 + y'(t)^2}\, dt \\ &= a\int_{t_1}^{t_2} \sqrt{1-e^2\cos^2 t}\, dt. \end{aligned}$$

Une formule analogue est obtenue pour la seconde représentation paramétrique. C'est d'ailleurs cette formule que considère implicitement Legendre avec une définition particulière de l'aplatissement $(\dfrac{a-b}{b})$. Dans les deux cas bien sûr, ces intégrales doivent être calculées par des procédures de calcul numérique reprises dans l'annexe A. On obtient ainsi la formule que Legendre a soumise aux moindres carrés pour en estimer les coefficients inconnus [6] à partir d'observations de terrain.

5. La formule (10) de l'article de Balmino donne une expression erronée de s ce qui conduit l'auteur à des résultats faux dès lors qu'il utilise cette formule dans ses calculs ultérieurs qui eux ne contiennent pas d'erreurs.

6. En fait, Legendre disposait d'estimations de ces coefficients obtenues par d'autres procédés et pouvait ainsi apprécier la validité de son approche.

2.2. Les données d'observation

Il s'agit là d'un jeu de données assez classique de l'histoire de la métrologie. Ces données sont reproduites dans le tableau 1. Rassemblées vers 1795, elles ont été publiées en particulier dans le *Allgemeine Geographische Ephemeriden* en 1799 où elles ont notamment retenu l'attention de Gauss [7] et fait l'objet d'un ajustement. Par une méthode de moindres carrés ? Peut-être. Cette antériorité [8] possible est très soigneusement discutée par Stigler (1999, chapitre 17 et, avec plus de détails, 1981).

TABLEAU 1. — Les données utilisées par Legendre telles que publiées dans son appendice «sur la méthode des moindres quarrés».

Lieu de l'observation	*Sa latitude*	*Arcs compris exprimés en modules*	*L'-L*	*L'+L*
Dunkerque	51° 2'10"50			
		DP 62472,59	2° 11'20"75	99° 53' 0"
Panthéon à Paris	48° 50'49"75			
		PE 76145,74	2° 40' 7"25	95° 1'32"
Evaux	46° 10'42"50			
		EC 84424,55	2° 57'48"10	89° 23'37"
Carcassonne	43° 12'54"40			
		CM 52749,48	1° 51' 9"60	84° 34'39"
Montjouy	41° 21'44"80			

Le méridien de Paris ainsi que les cinq lieux de mesure (Dunkerque, le Panthéon à Paris, Évaux, Carcassonne, Montjouy [9]) sont représentés sur notre schéma de l'ellipsoïde terrestre (voir figure 1). C'est la latitude géodésique de ces cinq lieux qui est utilisée dans cette représentation. Au vu du graphique, on peut se demander s'il est réaliste de penser pouvoir reconstruire l'ellipsoïde terrestre compte tenu du positionnement très voisin de ces points de mesure. Et les résultats numériques ne font que confirmer cette impression pessimiste.

3. Les moindres carrés

Comme on peut le lire dans son article de 1806, Legendre considère deux modèles que l'on obtient, selon lui, par des « formules connues ». Ces modèles expriment les relations approchées entre latitudes et mesure de l'arc compris entre deux latitudes. Leur obtention est détaillée en annexe (voir annexe A).

7. Rappelons que ce jeu de données avait été publié dans *Allgemeine Geographische Ephemeriden* avec une entrée erronée que Gauss avait identifiée et corrigée (voir Stigler, 1999, table 17.1, page 324).

8. Sur la querelle des priorités Gauss-Legendre, il faut aussi se reporter à Stigler (1977).

9. *Montjuïc* à Barcelone.

3.1. Les équations à « faire concourir »

Notant i ($i = 0, \ldots, 4$) les « lieux de l'observation », L_i leurs latitudes, et S_i les mesures des 4 arcs i ($i = 1, \ldots, 4$) entre les lieux de latitudes L_{i-1} et L_i, les deux modèles proposées par Legendre ont pour expression formelle :

1. Variable réponse : S_i ; prédicteur linéaire de la forme :

$$\beta_1(L_i - L_{i-1}) + \beta_2 K \sin(L_i - L_{i-1}) \cos(L_i + L_{i-1})$$

où K est un coefficient connu, et où β_1 et β_2 désignent les coefficients génériques inconnus du modèle.

2. Variable réponse : $L_i - L_{i-1}$; prédicteur linéaire de la forme :

$$\frac{S_i}{K'} + \beta_1 \frac{S_i}{K'} + \beta_2 K'' \sin(L_i - L_{i-1}) \cos(L_i + L_{i-1})$$

où K' et K'' sont des coefficients connus, et où β_1 et β_2 désignent encore les coefficients génériques inconnus du modèle.

On notera que, dans les deux expressions, le prédicteur linéaire ne comporte pas de terme constant (*intercept*). (L'exposé général sur les moindres carrés de Legendre reconnait cependant la possibilité d'un tel terme.) Par ailleurs, on pourrait voir dans le second, sauf à considérer un reparamétrage évident, le premier emploi d'une tare (*offset*) dans un modèle linéaire !

Legendre privilégie le second modèle. Son choix est motivé par de meilleures possibilités d'interprétation physique des coefficients, notamment l'estimation du coefficient α d'aplatissement de la terre [10]. En effet, il s'avère que, si l'aplatissement α est considéré comme inconnu, le premier modèle dépend en fait de trois coefficients inconnus β_1, β_2 et α liés par la relation $\beta_2 = \beta_1 \alpha$; le modèle est donc bilinéaire ou bi-additif en β_1 et α (voir appendice C). Il s'avère aussi que, dans le second modèle, les paramètres β_1 et β_2 sont indépendants (avec $\beta_2 = \alpha$). D'où son choix. Mais il est alors douteux, d'un point de vue de stricte orthodoxie statistique, de retrouver dans le prédicteur linéaire une fonction de la variable réponse.

3.2. Le critère d'ajustement de Legendre

En tout état de cause, Legendre obtient ce que nous appelons de nos jours un modèle de régression linéaire. Rappelons-en les ingrédients en des termes contemporains : un vecteur réponse, $\underline{y}$, et un modèle linéaire pour son prédicteur, $\underline{\mu} = \mathbf{X}\underline{\beta}$. La méthode des moindres carrés consiste alors à estimer les coefficients $\underline{\beta}$ du prédicteur en minimisant $|| \underline{y} - \underline{\mu}(\underline{\beta}) ||^2$ et le génie de Legendre est d'avoir montré que le problème se ramenait à celui de la recherche de la solution d'un système d'équations linéaires, $\mathbf{X}'\mathbf{X}\underline{\beta} = \mathbf{X}'\underline{y}$, admettant,

10. En fait Legendre (voir annexe A) définit l'aplatissement comme le rapport de la différence $(a - b)$ des demi-axes de l'ellipse au demi-petit axe (b).

sous des conditions assez évidentes, une solution unique [11]. On peut penser qu'un des intérêts de la méthode était qu'elle se ramenait techniquement au calcul des coefficients d'un système (de Cramer) de petite taille et à sa résolution par éliminations successives des variables.

Mais, comme on va le voir, Legendre en traitant son exemple va au delà de l'application ordinaire des moindres carrés.

4. L'analyse numérique de Legendre

Les observations sont naturellement ordonnées le long du méridien de Paris et Legendre veut intégrer à son ajustement les effets d'une propagation possible des erreurs de mesure le long de ce méridien. De façon intuitive, l'ajustement ne doit pas minimiser la somme des carrés des erreurs ($u_i = y_i - \mu_i$) mais la somme des carrés d'erreurs latentes (e_i) gouvernant les premières. C'est cette démarche qui est examinée ci-après en la replongeant dans un cadre contemporain et en lui conférant une certaine généralité.

Legendre propose donc un modèle linéaire plausible de dépendance entre les u_i ($i = 1, \ldots, 4$) et les erreurs latentes e_i ($i = 0, 1, \ldots, 4$) : $u_i = e_i - e_{i-1}$. Legendre utilise alors le fait que la somme des erreurs latentes ($\sum e_i$) doit être nulle pour construire un problème augmenté dont les e_i sont les erreurs. Le procédé de décorrélation ne serait pas nouveau et Stigler (Stigler, 1986, p. 60) l'attribue à Pierre Simon Laplace (1749–1827). Mais l'argument invoqué par Legendre en l'absence notamment de terme constant (intercept) est-il recevable ?

Suivant la démarche présentée par Legendre mais en la généralisant quelque peu, supposons que le vecteur des erreurs ($\underline{u}$ de dimension n) soit une transformation linéaire d'un vecteur d'erreurs latentes ($\underline{e}$ de dimension $n+k$) : $\underline{u} = \mathbf{L}\underline{e}$, la matrice $\mathbf{L}$ de dimension $(n, n+k)$ étant supposée de rang n. Soit alors une matrice ℓ de dimension $(k, n+k)$ et de rang k telle que $\ell\mathbf{L}' = \mathbf{0}$.

En posant $\ell\underline{e} = \underline{0}$, l'argument de Legendre, et compte tenu des hypothèses, on a :

$$\begin{bmatrix} \underline{u} \\ \underline{0} \end{bmatrix} = \begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix} \underline{e} \quad \text{soit encore} \quad \underline{e} = \begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix}^{-1} \begin{bmatrix} \underline{u} \\ \underline{0} \end{bmatrix};$$

il est alors facile de vérifier que

$$||\underline{e}||^2 = \underline{e}'\underline{e} = \underline{u}(\mathbf{L}\mathbf{L}')^{-1}\underline{u} = ||\underline{u}'||^2_{(\mathbf{L}\mathbf{L}')^{-1}}.$$

Legendre se ramène ainsi à la résolution d'un problème ordinaire de moindres carrés entre un vecteur réponse augmenté $\underline{y_1} = \begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix}^{-1} \begin{bmatrix} \underline{y} \\ \underline{0} \end{bmatrix}$ et une matrice expérimentale augmentée $\mathbf{X_1} = \begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X} \\ \mathbf{0} \end{bmatrix}$. Autrement dit, par un simple pré-traitement des données, lié à un choix heureux de modèle pour les erreurs

11. Précisons que ces équations ne s'appelaient pas alors les équations normales (voir infra note 12), que l'on ne parlait pas encore de système de Cramer, que le mot de régression ne sera introduit que plus tard et dans un tout autre contexte...

latentes (la matrice $\mathbf{L}$) et un choix *ad hoc* de la matrice ℓ, Legendre résout un problème particulier de moindres carrés généralisés.

Les valeurs numériques des matrices implicitement utilisées par Legendre dans son application figurent ci-dessous :

$$\mathbf{L} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad \mathbf{LL}' = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{L} \\ \ell \end{bmatrix}^{-1} = \frac{1}{5} \begin{bmatrix} -4 & -3 & -2 & -1 & 1 \\ 1 & -3 & -2 & -1 & 1 \\ 1 & 2 & -2 & -1 & 1 \\ 1 & 2 & 3 & -1 & 1 \\ 1 & 2 & 3 & 4 & 1 \end{bmatrix}$$

De nos jours, la méthode utilisée par Legendre, augmentation et inversion, peut faire sourire car il existe des méthodes numériques d'emploi nettement plus simple pour effectuer des moindres carrés généralisés. Connaissant $\mathbf{LL}'$, on peut procéder sans augmentation des données en effectuant des moindres carrés ordinaires avec des données transformées $\mathbf{H}\underline{y}$ et $\mathbf{HX}$. Un vaste choix s'offre pour $\mathbf{H}$: $\mathbf{H} = (\mathbf{LL}')^{-\frac{1}{2}}$, l'inverse $\mathbf{C}^{-1}$ de la décomposition de Cholesky $\mathbf{C}$ de $\mathbf{LL}'$, la décomposition de Cholesky D de l'inverse $(\mathbf{LL}')^{-1}$ de $\mathbf{LL}'$ ...

Toutefois, au terme de son analyse, Legendre est fortement déçu par la valeur estimée de l'aplatissement de la terre qu'il obtient dans cet exemple. En effet, la régression donne une valeur égale à $\frac{1}{149}$ alors que la valeur connue à cette époque était de $\frac{1}{321}$, et celle fixée en 1976 par l'Union Astronomique Internationale de $\frac{1}{298{,}257}$. Mais ces résultats n'altèrent en rien la confiance que Legendre place dans la méthode des moindres carrés puisqu'il n'hésite pas à écrire : « L'application que nous allons faire de cette méthode à la mesure de la méridienne, achèvera de mettre dans tout son jour sa simplicité et sa fécondité. »

Dans l'exemple de Legendre, tout se passe bien, trop bien même. La situation est-elle aussi simple pour d'autres situations de dépendance des erreurs ? Il est assez naturel de replacer l'approche de Legendre dans le cadre des modèles *ARMA(p,q)* introduits par Georges Box et Gwilym Jenkins (1970) mais en se

limitant aux modèles les plus simples *MA(1)* ou *AR(1)*. C'est ce qui est fait dans l'appendice B en considérant toujours le problème étudié par Legendre.

5. Georges Muntz et les chemins de l'analyse de covariance

TABLEAU 2. — Les données de Muntz (p. 91).

Groupe (i)	*N° d'ordre (j)*	*Prix (y)*	*Distance (x)*
1	1	18,8	5,50
1	2	0,8	0,55
1	3	2,2	1,00
1	4	12,3	4,00
1	5	9,9	3,50
1	6	9,0	3,00
1	7	10,2	4,00
1	8	8,4	3,00
1	9	5,0	2,70
1	10	2,2	1,20
1	11	2,7	1,50
1	12	2,1	1,00
1	13	1,6	0,80
1	14	1,5	1,00
1	15	13,3	4,10
2	1	7,6	3,40
2	2	6,7	3,35
2	3	8,9	4,50

Presque trente ans plus tard, en 1834, Georges Muntz publie les résultats d'une étude faite sur l'évolution du prix du transport total du mètre cube de divers matériaux de construction (et notamment de calcaire concassé) en fonction de la distance à parcourir. Il pose pour la moyenne du prix demandé (y, la variable réponse) un modèle de prédicteur linéaire simple classique ($b+ax$). Il fait alors « concourir toutes les équations à la détermination de a et b : en les assujettissant à la condition que la somme de leurs quarrés soit un *minimum* ». Il rappelle alors les formules des équations normales (sans les désigner sous ce nom [12]) et leur solution. L'exposé de la méthode de moindres carrés appliqué au cas simple de l'ajustement d'une droite est d'une grande clarté [13].

12. Aimé Laussedat (1819-1907) utilise cette dénomination dans la traduction qu'il publie en 1860 des travaux de triangulation des colonels Ibañez et Saavedra (voir p. 128 de l'ouvrage de la Commission de la carte d'Espagne cité dans les références bibliographiques). Maurice d'Ocagne, comme d'ailleurs Gauss et Legendre, ne leur donne pas de nom dans son exposé des moindres carrés et des moindres carrés contraints (d'Ocagne, 1910). Emmanuel Carvallo les désigne sous le nom d'« équations résultantes » dans son ouvrage de 1912 ...

13. Léon (Chrétien-)Lalanne (1811–1892) avait repéré la qualité de cet article puisqu'il le cite et en reprend assez fidèlement l'exposé théorique dans la section intitulée « Quelques

De fait, les données sont plus complexes. Muntz dispose dans l'exemple du calcaire concassé de 15 observations relatives à des transports sur des chemins normaux (y_{1j}) et de 3 observations relatives à des transports sur des chemins en terre glaise (y_{2j}). Comment estimer une ordonnée commune (partie fixe du prix) et deux pentes distinctes (une par groupe) ? On peut penser que si Muntz avait su coder sa variable qualitative de groupement il aurait su aussi, comme Legendre l'avait présenté en 1805, faire une régression linéaire multiple.

Toutefois, Muntz résout le problème en profitant du fort déséquilibre des deux groupes ($n_1 > n_2$). Sa procédure comporte deux étapes : estimation de b (ordonnée commune aux deux groupes) et de a_1 (pente spécifique d'un groupe) sur le groupe le plus nombreux (numéroté 1) par une régression linéaire simple classique ; estimation de a_2 sur le groupe le moins nombreux en moyennant les réponses transformées $\dfrac{y_{2j} - b}{x_{2j}}$.

Les deux étapes relèvent bien sûr de l'application des moindres carrés et le lecteur pourra vérifier que Muntz définit ainsi des estimateurs linéaires. Certes, ces estimateurs ne possèdent pas l'optimalité garantie par le théorème de Gauss-Markov dans une estimation globale de b, a_1 et a_2. Mais les estimations obtenues sont ici très proches des estimations optimales.

On peut aussi se demander si une procédure itérative pourrait être tentée : des estimations de a_1 et a_2 étant ainsi disponibles, procéder à la révision de l'estimation de b, puis à celle des estimations de a_1 et a_2 ..., et ce, jusqu'à une éventuelle convergence. Mais il est assez facile de vérifier qu'une formule de type moyenne des $\dfrac{y_{ij} - b}{x_{ij}}$ pour calculer la valeur révisée de a_i ne permet pas de converger vers la solution globale. Ce point de vue algorithmique est repris dans l'appendice C où certaines procédures de moindres carrés itérés sont étudiées.

6. Conclusion

Les deux exemples ainsi évoqués relèvent de la statistique descriptive (ou encore empirique). En effet, ni Legendre ni Muntz n'introduisent de modèle probabiliste pour rendre compte des dissonances entre les relations observées [14]. Pourtant, dans son mémoire publié vers 1830 [15], Legendre reproduit presque mot à mot l'exposé de 1805 (et de 1806) mais introduit la méthode des moindres carrés en rappelant que « M. le comte Laplace ayant trouvé par des considérations fondées sur le calcul des probabilités, que la méthode des moindres quarrés doit être employée préférablement à toute autre, pour

applications du calcul des probabilités» dans un article consacré au *Calcul des probabilités* publié dans une petite encyclopédie populaire (Lalanne, 1850).

14. On notera que l'appendice de 1806 contient 22 occurrences du terme erreur mais en dehors de tout concept explicitement probabiliste.

15. Il s'agit d'un texte qui aurait été lu le 24 septembre 1811.

trouver la valeur moyenne la plus exacte d'un ou de plusieurs éléments inconnus ...» [16]. Mais la manière dont les données sont traitées reste exemplaire pour des statisticiens.

Un regret en passant. Il est un peu surprenant de noter l'absence totale de graphiques statistiques dans ces publications. Et pourtant, Legendre est l'auteur d'un traité de géométrie de base orné de fines illustrations et souvent réédité (par exemple la 14^e édition en 1843). On sait aussi que les ingénieurs des ponts et chaussées recevaient une excellente formation au dessin scientifique. Il n'est donc pas étonnant que la Bibliothèque nationale et universitaire de Strasbourg conserve une carte hydraulique du Haut-Rhin concernant le canal de Vauban et les eaux du Wandelbach [17]. Mais Muntz n'a pas publié de représentation graphique de son problème d'ajustement.

Enfin, cette commémoration des moindres carrés ne saurait être complète sans que soit évoqué l'excellent ouvrage publié par Åke Björck il y a quelques années et consacré aux méthodes numériques pour les moindres carrés (1996). Le statisticien ignore souvent l'obscure algorithmique qui se met en œuvre à l'occasion de régressions notamment et les rend numériquement acceptables. Le livre de Björck permet d'en mesurer les enjeux. L'ouvrage était attendu et il avait fait lors de sa parution l'objet d'une publicité assez amusante qui souligne l'intérêt des commémorations. L'éditeur, la puissante *Society for Industrial and Applied Mathematics* pour ne pas la nommer, n'hésitait pas dans ce prospectus à rappeler *"1795 – Gauss discovers the method of least-squares ... 1995 – Björck writes a monograph that covers the full spectrum of relevant problems and methods in least squares"*. Si la seconde partie du message publicitaire n'est pas exagérée, force est de constater que la première reste un sujet de controverse. C'est Legendre qui a publié, le premier, un exposé précis des moindres carrés et déposé ainsi le nom de ce que Gauss persistait à appeler en 1809 *"meine Methode"* [18].

Références

AITKIN Murray, FRANCIS Brian, and HINDE John (2005). *Statistical modelling in GLIM 4*. Oxford University Press, 2nd edition.

ALDER Ken (2002). *The measure of all things*. Little, Brown, Boston.

ALDER Ken (2005). *Mesurer le monde : l'incroyable histoire de l'invention du mètre*. Flammarion, Paris.

BALMINO Georges (2005). Legendre et la mesure du méridien : 200 ans après. In *Guide des données astronomiques 2006 pour l'observation du ciel, Annuaire du Bureau des Longitudes*, pp. 341-358, EDP Sciences, Paris.

16. On peut regretter que Legendre ne cite pas aussi l'apport théorique de Gauss publié en 1809 chez Perthes et Besser à Hamburg sous le titre de *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.

17. Voir la notice consacrée à Georges Müntz dans le *Nouveau dictionnaire de biographie alsacienne* publié par la Fédération des Sociétés d'histoire et d'archéologie d'Alsace (1996).

18. ou encore *"principium nostrum"*.

BJÖRK Åke (1996). *Numerical methods for least squares problems.* SIAM, Philadelphia.

BOX Georges and JENKINS Gwilym (1970). *Times series analysis : Forecasting and control.* Holden-Day, San Francisco.

BREZINSKI Claude (2005). La méthode de Cholesky. *Revue d'Histoire des Mathématiques*, 11 :205-238.

CARVALLO Emmanuel (1912). *Le calcul des probabilités et ses applications.* Gauthier-Villars, Paris.

DURBIN James and WATSON Geoffrey S. (1950). Testing for serial correlation in least squares regression. *Biometrika*, 37 :409-428.

DURBIN James and WATSON Geoffrey S. (1951). Testing for serial correlation in least squares regression. *Biometrika*, 38 :159-178.

FALGUEROLLES Antoine de and FRANCIS Brian (1992). Algorithmic approaches for fitting bilinear models. In Yadolah Dodge and Joe Whittaker, editors, *COMPSTAT 92, Proceeding in Computational Statistics*, volume 1, pp. 77-82, Heidelberg, Physical-Verlag.

FAREBROTHER Richard William (1998). *Fitting linear relationships, a History of the calculus of observations 1750-1900.* Springer, New York.

Fédération des Sociétés d'histoire et d'archéologie d'Alsace (éditeur) (1996). *Nouveau dictionnaire de biographie alsacienne*, volume 6, page 2777, Strasbourg.

GABRIEL K. Ruben and ZAMIR Shmuel (1979). Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, 21 :489-498.

GAUSS Carl Friedrich (1995). *Theoretica combinationis observationum erroribus minimis obnoxiae (avec une traduction an anglais et une postface de Gilbert W. Stewart).* Society for industrial and Applied Mathematics, Philadelphia.

GUEDJ Denis (1986). *La méridienne (1792-1799).* Seghers, Paris.

(CHRÉTIEN-)LALANNE Léon (1850). Calcul des probabilités. In *Un million de faits, aide-mémoire universel des Sciences, des Arts et des Lettres*, colonnes 255-266, Garniers, Paris (5^{e} édition).

LAUSSEDAT Aimé (traducteur) (1860). *Expériences faites avec l'appareil à mesurer les bases appartenant à la Commission de la Carte d'Espagne.* Librairie Militaire, Paris.

LEGENDRE Adrien Marie (1806). *Nouvelles méthodes pour la détermination des orbites des comètes avec un supplément contenant divers perfectionnements de ces méthodes et leur application aux deux comètes de 1805.* Courcier, Paris.

LEGENDRE Adrien Marie (ca. 1830). *Mémoire sur la méthode des moindres quarrés et sur l'attraction des ellipsoïdes homogènes.* s.n., s.l.

LEGENDRE Adrien Marie (1843). *Éléments de géométrie.* Firmin Didot frères, Paris, 14^{e} édition.

MCCULLAGH Peter and NELDER John A. (1989). *Generalized linear models.* Chapman and Hall, London, 2nd edition.

MUNTZ Georges (1834). Note sur l'évaluation du prix du transport des matériaux de construction dans l'arrondissement du Nord. *Annales des Ponts et Chaussées*, 167 :86-100, premier semestre.

D'OCAGNE Maurice (1910). *Notions élémentaires sur la probabilité des erreurs.* Gauthier-Villars, Paris.

STIGLER Stephen M. (1977). An attack on Gauss, published by Legendre in 1820. *Historia Mathematica*, 4 :31-35.

STIGLER Stephen M. (1981). Gauss and the invention of least squares. *Annals of statistics*, 9 :465-474.

STIGLER Stephen M. (1986). *The history of statistics : The measurement of uncertainly before 1900.* Harvard University Press, Cambridge, Mass.

STIGLER Stephen M. (1999). *Statistics on the Table : The history of statistical concepts and methods.* Harvard University Press, Cambridge, Mass.

WOLD Herman (1946). Nonlinear estimation by iterative least squares procedures. In Florence N. David, editor, *Research paper in statistics, Festschrift for Jerzy Neyman*, volume 1, pp. 391-420, Wiley, New York.

Annexe A

Les « formules connues » d'après Legendre

Rappelons d'abord ce qu'écrit Legendre page 76 de son appendice de 1806. « Supposant que le méridien est une ellipse dont les axes sont dans le rapport de 1 à $1 + \alpha_L$ [19], si on désigne par D la longueur du 45^{eme} degré, et par S celle de l'arc compris entre les deux latitudes L et L', on aura par les formules connues, et en exprimant $L' - L$ en degrés :

$$S = D(L' - L) - \frac{3}{2}\alpha_L D\frac{180}{\pi}\sin(L' - L)\cos(L' + L) ;$$

d'où résulte

$$L' - L = \frac{S}{D} + \frac{3}{2}\alpha_L\frac{180}{\pi}\sin(L' - L)\cos(L' + L) ;$$

Comme le $45^{\text{ème}}$ degré est d'environ 28500 modules, égaux chacun à deux toises, on peut faire $\frac{1}{D} = \frac{1+C}{28500}$, C étant une fraction très petite, et on aura

$$L' - L = \frac{S}{28500} + C\frac{S}{28500} + \alpha_L\frac{270}{\pi}\sin(L' - L)\cos(L' + L), \qquad (a)$$

qui pour chaque arc dont au connoît la longueur avec la latitude de ses extrémités, donnera une relation entre α_L et C. »

Retrouvons donc ces « formules connues ».

---

19. Legendre utilise en fait la notation α. On a introduit ici la notation α_L pour distinguer la définition de l'aplatissement selon Legendre ($\alpha_L = \frac{a-b}{b}$) de celle couramment admise ($\alpha = \frac{a-b}{a}$).

A.1. La mesure d'un arc de méridien

Le calcul de la longueur algébrique d'un arc d'ellipse compris entre les points de paramètres t_1 et t_2 pour la première représentation paramétrique est donné par :

$$\begin{aligned} s &= \int_{t_1}^{t_2} \sqrt{x'(t)^2 + y'(t)^2}\, dt \\ &= a \int_{t_1}^{t_2} \sqrt{1 - e^2 \cos^2 t}\, dt. \end{aligned} \tag{1}$$

Une formule analogue est obtenue pour la seconde représentation paramétrique. La longueur algébrique de l'arc d'ellipse compris entre les points de paramètres Φ_1 et Φ_2 est

$$\begin{aligned} S &= \int_{\Phi_1}^{\Phi_2} \sqrt{X'(\Phi)^2 + Y'(\Phi)^2}\, d\Phi \\ &= a(1-\alpha)^2 \int_{\Phi_1}^{\Phi_2} \left((1-\alpha)^2 - e^2 \cos^2 \Phi\right)^{-\frac{3}{2}}\, d\Phi\,. \end{aligned} \tag{2}$$

Les deux paramètres t et Φ sont liés de la façon suivante $t = \Phi$ si $t = \frac{\pi}{2} + k\pi$ avec k entier relatif, et si $t = t_0 + k\pi$ avec $-\frac{\pi}{2} < t_0 < \frac{\pi}{2}$ alors $\Phi = \Phi_0 + k\pi$ et

$$\tan t_0 = (1-\alpha) \tan \Phi_0\,. \tag{3}$$

Considérant l'aplatissement α comme un petit paramètre, on souhaite obtenir les premiers termes d'un développement limité de S en $\alpha = 0$, dont les coefficients sont des fonctions de Φ_1 et Φ_2.

Deux stratégies de calcul sont possibles :

1. Remplacer l'intégrande de (1) par un développement à l'ordre n au voisinage de $\alpha = 0$, puis intégrer chacun des termes de la série. Les coefficients sont des fonctions de t_1 et t_2. La formule (3) permet de remplacer t_1 et t_2 par leur développement à l'ordre n au voisinage de $\alpha = 0$ dans les coefficients. Une troncature à l'ordre n permet alors d'obtenir le résultat.
2. On remplace directement l'intégrande de (2) par un développement au voisinage de $\alpha = 0$ à l'ordre n et on intègre terme à terme.

La première méthode est utilisée dans l'article de Balmino (ses formules (11) à (16)).

Dans la suite de cette annexe, on présente le calcul au deuxième ordre en utilisant chacune des deux méthodes. L'ensemble des calculs, y compris la vérification des formules de base, a été effectué avec le logiciel de calcul formel Maple.

Première méthode

L'intégrande dans (1) admet le développement

$$\sqrt{1-e^2\cos^2 t} = 1 - \cos^2 t\,\alpha + \frac{1}{2}(\cos^2 t - \cos^4 t)\alpha^2 + o(\alpha^2)\ ,$$

d'où l'on déduit par intégration entre t_1 et t_2

$$\begin{aligned} s = a(t_2 - t_1) \quad &- \quad \frac{a}{4}\,(2(t_2 - t_1) + \sin 2t_2 - \sin 2t_1)\,\alpha \\ &+ \quad \frac{a}{64}\,(4(t_2 - t_1) - \sin 4t_2 + \sin 4t_1)\,\alpha^2 + o(\alpha^2)\ . \end{aligned} \tag{4}$$

De la relation (3), on déduit

$$t = k\pi + \operatorname{atan}((1-\alpha)\tan\Phi)\ ,$$

d'où

$$t = \Phi - \frac{1}{2}\sin 2\Phi\,\alpha + \frac{1}{8}(\sin 4\Phi - 2\sin 2\Phi)\alpha^2 + o(\alpha^2)\ . \tag{5}$$

En substituant (5) dans (4), il vient

$$\begin{aligned} S \;=\; a\Big[\quad & \Phi_2 - \Phi_1 \\ & -\frac{1}{2}\left(\Phi_2 - \Phi_1 + \frac{3}{2}(\sin 2\Phi_2 - \sin 2\Phi_1)\right)\alpha \\ & +\frac{1}{16}\left(\Phi_2 - \Phi_1 + \frac{15}{4}(\sin 4\Phi_2 - \sin 4\Phi_1)\right)\alpha^2\Big] + o(\alpha^2)\ . \end{aligned} \tag{6}$$

Deuxième méthode

Le développement en $\alpha = 0$ à l'ordre 2 de (2) donne directement (6). Au premier ordre, on obtient

$$S = a\left[(1 - \frac{\alpha}{2})(\Phi_2 - \Phi_1) - \frac{3\alpha}{4}(\sin 2\Phi_2 - \sin 2\Phi_1)\right]\ ,$$

ce qui compte tenu de l'égalité

$$\sin 2\Phi_2 - \sin 2\Phi_1 = 2\sin(\Phi_2 - \Phi_1)\cos(\Phi_1 + \Phi_2)\ ,$$

s'écrit encore

$$S = a\left[(1 - \frac{\alpha}{2})(\Phi_2 - \Phi_1) - \frac{3\alpha}{2}\sin(\Phi_2 - \Phi_1)\cos(\Phi_1 + \Phi_2)\right]\ .$$

Le 45^e degré

Legendre introduit la «longueur du 45^e degré», D, en posant dans S, $\Phi_1 = \frac{\pi}{4} - \frac{\pi}{360}$ et $\Phi_2 = \frac{\pi}{4} + \frac{\pi}{360}$, ce qui donne en reportant dans (6)

$$D = a\left[\frac{\pi}{180} - \frac{1}{2}\,\frac{\pi}{180}\alpha + \frac{1}{16}\left(\frac{\pi}{180} + \frac{15}{4}(-2\sin\frac{\pi}{90})\right)\alpha^2\right] + o(\alpha^2)\,. \qquad (7)$$

Avec (6) et (7), on obtient

$$\begin{aligned}\frac{S}{D} &= \frac{180}{\pi}(\Phi_2 - \Phi_1) - \frac{135}{\pi}(\sin 2\Phi_2 - \sin 2\Phi_1)\,\alpha \\ &\quad + \frac{135}{16\pi^2}\Big[(1800\sin\frac{\pi}{90})(\Phi_2 - \Phi_1) - 8\pi(\sin 2\Phi_2 - \sin 2\Phi_1) \\ &\quad + 5\pi(\sin 4\Phi_2 - \sin 4\Phi_1)\Big]\,\alpha^2 + o(\alpha^2)\,.\end{aligned} \qquad (8)$$

Legendre utilise une mesure des angles en degrés au lieu des radians, $L = \frac{180}{\pi}\Phi$ et sa fonction sinus (cosinus) est la fonction sinus appliquée à des angles mesurés en degrés. Notons $\sin_L$ et $\cos_L$ ces fonctions, notées sin et cos dans les écrits de Legendre, qui sont définies par

$$\begin{aligned}\sin\Phi &= \sin_L(\frac{180}{\pi}\Phi) \\ \cos\Phi &= \cos_L(\frac{180}{\pi}\Phi)\,.\end{aligned}$$

En posant $L_1 = \frac{180}{\pi}\Phi_1$ et $L_2 = \frac{180}{\pi}\Phi_2$ et en remplaçant dans (8), il vient

$$\begin{aligned}\frac{S}{D} &= L_2 - L_1 - \frac{135}{\pi}(\sin_L 2L_2 - \sin_L 2L_1)\alpha \\ &\quad + \frac{135}{16\pi}\Big[(10\sin\frac{\pi}{90})(L_2 - L_1) - 8(\sin_L L_2 - \sin_L 2L_1) \\ &\quad + 5(\sin_L 4L_2 - \sin_L 4L_1)\Big]\alpha^2 + o(\alpha^2)\,.\end{aligned} \qquad (9)$$

L'aplatissement selon Legendre

Si au lieu d'utiliser le paramètre α, on utilise le paramètre α_L défini par

$$\alpha = \frac{\alpha_L}{1 + \alpha_L}\,,$$

on obtient le développement au deuxième ordre

$$\begin{aligned}\frac{S}{D} &= L_2 - L_1 - \frac{135}{\pi}(\sin_L 2L_2 - \sin_L 2L_1)\alpha_L \\ &\quad + \frac{135}{16\pi}\Big[(10\sin\frac{\pi}{90})(L_2 - L_1) + 8(\sin_L 2L_2 - \sin_L 2L_1) \\ &\quad + 5(\sin_L 4L_2 - \sin_L 4L_1)\Big]\alpha_L^2 + o(\alpha_L^2)\,.\end{aligned} \qquad (10)$$

A.2 Les prédicteurs linéaires considérés par Legendre

On retrouve donc bien au premier ordre la formule (a) de Legendre :

$$\begin{aligned} \frac{S}{D} &= L_2 - L_1 - \frac{135}{\pi}(\sin_L 2L_2 - \sin_L 2L_1)\alpha_L + o(\alpha_L) \\ &= L_2 - L_1 - \frac{3}{2}\alpha_L \frac{180}{\pi}\sin_L(L_2 - L_1)\cos_L(L_2 + L_1) + o(\alpha_L)\ , \end{aligned} \tag{11}$$

et, par suite,

$$S = D(L_2 - L_1) - \frac{3}{2}\alpha_L D\frac{180}{\pi}\sin_L(L_2 - L_1)\cos_L(L_2 + L_1) + o(\alpha_L)\ . \tag{12}$$

En ce qui concerne l'emploi de la formule (12), la variable réponse est S_i et, au premier ordre, le prédicteur linéaire est bien de la forme :

$$\beta_1(L_i - L_{i-1}) + \beta_2 K\sin_L(L_i - L_{i-1})\cos_L(L_i + L_{i-1})$$

où K est un coefficient connu, et où $\beta_1 = D$ et $\beta_2 = \alpha_L D$ désignent les coefficients génériques inconnus du modèle linéaire.

En ce qui concerne l'emploi de la formule (11), D n'étant qu'imparfaitement connu, Legendre remplace $\dfrac{1}{D}$ par $\dfrac{1+C}{28500}$. Il obtient ainsi

$$L_2 - L_1 = \frac{S}{28500} + C\frac{S}{28500} + \alpha_L\frac{270}{\pi}\sin_L(L_2 - L_1)\cos_L(L_2 + L_1) + o(\alpha_L)\ .$$

La variable réponse retenue par Legendre devient alors $L_i - L_{i-1}$ et le prédicteur linéaire est de la forme :

$$\frac{S_i}{K'} + \beta_1\frac{S_i}{K'} + \beta_2 K''\sin_L(L_i - L_{i-1})\cos_L(L_i + L_{i-1})$$

où K' et K'' sont des coefficients connus, et où $\beta_1 = C$ et $\beta_2 = \alpha_L$ désignent encore les coefficients génériques inconnus du modèle linéaire. C'est ce modèle de régression que Legendre privilégie.

Annexe B

Modèles simples de dépendance des erreurs

Replaçons l'approche de Legendre dans le cadre des modèles *ARMA(p,q)* introduits par Georges Box et Gwilym Jenkins (1970) mais nous limitant aux modèles les plus simples. Envisageons donc les deux schémas suivants de modèle de dépendance pour les erreurs : le schéma de type moyenne mobile d'ordre 1 ($u_i = e_i + \theta e_{i-1}$, $|\theta| \leqslant 1$), autrement dit un modèle *ARMA(0,1)* =

MA(1), et le schéma de type autorégressif d'ordre 1 ($u_i = \theta u_{i-1} + e_i$, $|\theta| < 1$), autrement dit un modèle *ARMA(1,0) = AR(1)*. La forme des matrices de variances associées $\mathbf{W}(\theta)$ des erreurs u_i est alors bien connue :

– pour une moyenne mobile d'ordre 1,

$$\mathbf{W}_L(\sigma^2,\theta) = \sigma^2 \begin{bmatrix} 1+\theta^2 & \theta & 0 & \dots & 0 & 0 & 0 \\ \theta & 1+\theta^2 & \theta & \dots & 0 & 0 & 0 \\ 0 & \theta & 1+\theta^2 & \dots & 0 & 0 & 0 \\ & & & & & & \\ 0 & 0 & 0 & \dots & 1+\theta^2 & \theta & 0 \\ 0 & 0 & 0 & \dots & \theta & 1+\theta^2 & \theta \\ 0 & 0 & 0 & \dots & 0 & \theta & 1+\theta^2 \end{bmatrix};$$

– pour un schéma autorégressif d'ordre 1,

$$\mathbf{W}_D(\sigma^2,\theta) = \sigma^2 \frac{1}{1-\theta^2} \begin{bmatrix} 1 & \theta & \theta^2 & \dots & \theta^{n-3} & \theta^{n-2} & \theta^{n-1} \\ \theta & 1 & \theta & \dots & \theta^{n-4} & \theta^{n-3} & \theta^{n-2} \\ \theta^2 & \theta & 1 & \dots & \theta^{n-5} & \theta^{n-4} & \theta^{n-3} \\ \vdots & & & & & & \vdots \\ \theta^{n-3} & \theta^{n-4} & \theta^{n-5} & \dots & 1 & \theta & \theta^2 \\ \theta^{n-2} & \theta^{n-3} & \theta^{n-4} & \dots & \theta & 1 & \theta \\ \theta^{n-1} & \theta^{n-2} & \theta^{n-3} & \dots & \theta^2 & \theta & 1 \end{bmatrix}.$$

La première de ces structures a pour cas particulier celle utilisée par Legendre dans son analyse : *MA(1)* avec $\theta = -1$. La seconde, *AR(1)* (et métrique $\mathbf{W}_D(\theta)^{-1}$), est bien connue ; c'est celle notoirement étudiée par James Durbin et Geoffrey Watson (1950, 1951). On sait que les calculs sont alors fortement simplifiés. En effet, l'expression de l'inverse de $\mathbf{W}_D(\theta)$, une matrice tridiagonale, est bien connue. (Elle ne diffère de celle de $\mathbf{W}_L(\theta)$ que par le premier et le dernier élément de sa diagonale principale.) Sa décomposition de Cholesky [20] $\mathbf{D}_D(\theta)$ est évidente à calculer. Dans la pratique, on choisit plutôt une variante, triangulaire inférieure, de la décomposition de Cholesky : $\mathbf{H}_D(\theta)$, $\mathbf{H}_D(\theta)'\mathbf{H}_D(\theta) = \mathbf{W}_D(\theta)^{-1}$. C'est une matrice bidiagonale dont les éléments non nuls (diagonaux ou juste au dessous) sont donnés par la récurrence : $c_{1,1} = \sqrt{1-\theta^2}$; pour $i = 2,\dots,n$, $c_{i,i-1} = -\theta$ et $c_{i,i} = 1$.

En ce qui concerne le cas *MA(1)* (et métrique $\mathbf{W}_L(\theta)^{-1}$), il faut utiliser des procédures numériques. On retrouve cependant certaines analogies avec la démarche précédente. La décomposition de Cholesky $\mathbf{C}_L(\theta)$ de $\mathbf{W}_L(\theta)$ reste relativement simple à calculer. Dans la pratique, on choisit en fait la variante, triangulaire inférieure, de cette décomposition : $\mathbf{H}_L(\theta)$ est une matrice bidiagonale dont les éléments non nuls (diagonaux ou juste au dessous) sont donnés par la récurrence : $h_{n,n} = \sqrt{1+\theta^2}$; pour $i = 1,\dots,n-1$, $h_{n+1-i,n-i} = \dfrac{\theta}{h_{n+1-i,n+1-i}}$ et $h_{n-i,n-i} = \sqrt{1+\theta^2-h^2_{n+1-i,n-i}}$. L'inversion

20. Sur Cholesky et la méthode de Cholesky on se reportera à l'article de Claude Brezinski (2005).

de cette matrice bidiagonale ne pose pas de difficulté et factorise ainsi finalement $\mathbf{W}_L(\theta)^{-1}$.

Il n'y a là rien de surprenant. On ne fait que retrouver une version approximative de la propriété bien connue selon laquelle les formes de la fonction d'autocovariance d'un *MA(q)* (respectivement, d'un *AR(q)*) et de la fonction d'autocovariance partielle d'un *AR(q)* (respectivement, d'un *MA(q)*) sont identiques. De ce point de vue, les stratégies de moindres carrés généralisés associées aux schémas considérés sont duales. Dans le cas « moyenne mobile », factoriser $\mathbf{W}_L(\theta)$ puis inverser $\mathbf{C}_L(\theta)$; dans le cas « autorégressif », inverser $\mathbf{W}_D(\theta)$ puis factoriser $\mathbf{W}_D(\theta)^{-1}$. Dans les deux schémas, les moindres carrés généralisés des données initiales se ramènent à des moindres carrés ordinaires sur les données transformées.

On notera aussi que l'approche considérée par Legendre, augmentation et inversion, se généralise aussi sans difficulté à tous les cas *MA(1)*. Il suffit de choisir $\ell = (-\theta, \theta^2, -\theta^3, \ldots)$. L'inversion de la matrice augmentée $\mathbf{L}_1(\theta)$, $\mathbf{L}_1(\theta)' = [\mathbf{L}(\theta)' \; \ell'(\theta)]$, devient alors un peu plus technique.

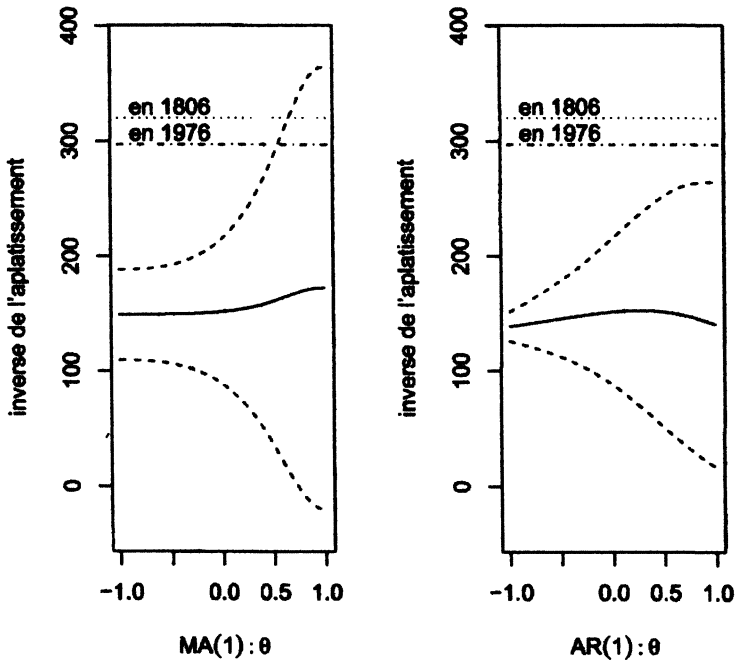

FIG 2. — Profil de variation de l'inverse de l'aplatissement de la terre en fonction de θ pour les modèles de dépendance *MA(1)* (à gauche) et *AR(1)* (à droite), intervalle de confiance de niveau de coefficient de sécurité asymptotique 95 % ; valeurs théoriques de cet inverse en 1806 et 1976.

Les schémas de dépendance *MA(1)* et *AR(1)* requièrent tous deux une valeur de θ qu'il faudrait donc estimer. Cet aspect n'est pas abordé ici. Mais les moyens actuels de calcul permettent d'étudier la variation d'un critère numérique d'intérêt lorsque le paramètre θ parcourt l'intervalle $(-1, 1)$. Legendre aurait-il mieux estimé l'aplatissement de la terre pour certains des schémas de dépendance généralisant son approche? L'examen du profil de variation de l'aplatissement de la terre en fonction de θ pour les modèles de dépendance *MA(1)* et *AR(1)* (voir figure 2) montre qu'il n'en est rien.

Annexe C
Moindres carrés pour certains modèles non linéaires du méridien

En partant de la transcription de la première équation de Legendre avec les notations utilisées dans l'appendice A,

$$S = D\,(L_2 - L_1) - \frac{3}{2}\,\alpha_L\,D\,\frac{180}{\pi}\,\sin_L(L_2 - L_1)\cos_L(L_2 + L_1)$$

et en poussant ce développement au second ordre, il vient comme on l'a vu :

$$\begin{aligned} S \;=\;& D\,(L_2 - L_1) \\ & - D\,\alpha_L\,\frac{135}{\pi}(\sin_L 2L_2 - \sin_L 2L_1) \\ & + D\,\alpha_L^2\,\frac{135}{16\pi}\Big[(10\sin_L\frac{\pi}{90})(L_2 - L_1) + 8(\sin_L L_2 - \sin_L 2L_1) \\ & + 5(\sin_L 4L_2 - \sin_L 4L_1)\Big] \end{aligned}$$

On reconnaîtra facilement dans cette formule, celle d'un modèle non linéaire dont les coefficients inconnus sont D et α_L.

En effet, avec des notations plus classiques pour un statisticien – espérance mathématique de la variable réponse S notée $\mu(\alpha, \beta)$, variables explicatives notées x^1, x^2, x^3, et coefficients inconnus notés α et β au lieu de α_L et D – le modèle admet la forme suivante :

$$\mu_i(\alpha, \beta) = \beta\,x_i^1 + \beta\alpha\,x_i^2 + \beta\alpha^2\,x_i^3$$

Le problème est alors celui de la recherche de a et b tels que :

$$(a, b) = \arg\min_{\alpha,\beta} \sum_i (S_i - \mu_i(\alpha, \beta))^2 \ .$$

Il va sans dire que cette extension des moindres carrés n'est pas considérée par Legendre. Mais une adaptation de sa méthode, les moindres carrés itérés, que les moyens actuels de calcul rendent possible, permet l'estimation des coefficients de ce type de modèle.

C.1 Conditions du premier ordre et l'inverse de la matrice d'information de Fisher

Les conditions du premier ordre du problème de moindres carrés (ou équations normales) sont données par :

$$\begin{cases} -2\sum_i (S_i - \mu_i(\alpha,\beta))\ \beta\ (x_i^2 + 2\alpha x_i^3) & = & 0 \\ -2\sum_i (S_i - \mu_i(\alpha,\beta))\ (x_i^1 + \alpha x_i^2 + \alpha^2 x_i^3) & = & 0 \end{cases}$$

Il est assez facile de vérifier que si la réponse est gaussienne, la matrice d'information de Fisher relative aux coefficients α et β a pour expression :

$$\sum_i \begin{bmatrix} \beta^2\ (x_i^2 + 2\alpha x_i^3)^2 & \beta\ (x_i^2 + 2\alpha x_i^3)(x_i^1 + \alpha x_i^2 + \alpha^2 x_i^3) \\ \beta\ (x_i^2 + 2\alpha x_i^3)(x_i^1 + \alpha x_i^2 + \alpha^2 x_i^3) & (x_i^1 + \alpha x_i^2 + \alpha^2 x_i^3)^2 \end{bmatrix}$$

Sous l'hypothèse gaussienne, le calcul de la matrice de covariance asymptotique des estimateurs de α et β, sous réserve certes de pouvoir en calculer des estimations, est immédiat. Or des adaptions assez simples des moindres carrés permettent dans la plupart des cas de les obtenir au terme de calculs itératifs.

C.2 Moindres carrés itérés

L'idée de base des moindres carrés itérés consiste remplacer à chaque itération $\mu_i(\alpha,\beta)$ par son développement limité d'ordre 1 au voisinage de la solution courante (a,b) :

$$\begin{aligned} \mu_i(\alpha,\beta) &\approx \mu_i(a,b) + [\alpha - a, \beta - b][\nabla \mu_i(\alpha,\beta)]_{\alpha=a,\beta=b} \\ &\approx \mu_i(a,b) + [\alpha - a, \beta - b] \begin{bmatrix} b\ (x_i^2 + 2a x_i^3) \\ (x_i^1 + a x_i^2 + a^2 x_i^3) \end{bmatrix} \end{aligned}$$

À chaque itération, on résout donc un problème classique de moindres carrés, la régression linéaire d'une variable réponse de travail, $S_i^{(a,b)} = S_i + abx_i^2 + 2a^2 b x_i^3$, sur des variables explicatives de travail, $X_i^a = b(x_i^2 + 2ax_i^3)$ et $X_i^b = x_i^1 + ax_i^2 + a^2 x_i^3$, obtenant ainsi :

$$\arg\min_{\alpha,\beta} \sum_i (S_i^{(a,b)} - (\alpha X_i^a + \beta X_i^b))^2$$

Lorsque la procédure converge, il est facile de vérifier que

$$\sum_i (S_i - \mu_i(a,b))^2 = \sum_i (S_i^{(a,b)} - (aX_i^a + bX_i^b))^2 .$$

Les conditions du premier ordre du problème initial sont alors vérifiées : en effet, les deux problèmes admettent les mêmes équations normales. Par

ailleurs, la matrice de Fisher des coefficients est encore un sous-produit de la régression itérée finale :

$$[X^a \quad X^b]'[X^a \quad X^b] \ .$$

La méthode des moindres carrés itérés peut être facilement adaptée à la prise en compte d'une métrique W^{-1}. C'est d'ailleurs ce qui est fait pour l'estimation des modèles linéaires généralisés (Peter McCullagh et John Nelder, 1989 et Murray Aitkin *et al.*, 2005) pour lesquels la métrique diagonale et la variable réponse sont éventuellement révisées à chaque itération. L'algorithme est alors dit des moindres carrés pondérés itérés (*iteratively weighted least squares*).

C.3 Moindres carrés alternés itérés

Compte tenu de la forme bi-linéaire (bi-additive ou encore multiplicative, selon les auteurs) en α et β de la formule du prédicteur, on peut alterner les régressions en itérant cette procédure :

- connaissant a, la régression de S_i sur $X_i^b = x_i^1 + ax_i^2 + a^2x_i^3$ donne l'estimation révisée de β :

$$\arg\min_{\beta} \sum_i (S_i - \beta X_i^b)^2 \ .$$

- Connaissant a et b, la régression de $S_i^a = S_i - bx_i^1 + ba^2x_i^3$ sur $X_i^a = b(x_i^2 + 2ax_i^3)$ donne une estimation révisée de α :

$$\arg\min_{\alpha} \sum_i (S_i^a - \alpha X_i^a)^2 \ .$$

Il est encore facile de vérifier que, lorsque l'algorithme converge, la solution obtenue satisfait aux conditions du premier ordre du problème initial.

Ce type de procédure a été souvent utilisé en statistique. Citons quelques exemples.

- Herman Wold (1946) obtient le premier élément singulier de la décomposition en valeurs singulières d'une matrice de données en alternant les régressions. En effet, la recherche de

$$||Y - d\,\underline{u}\,\underline{v}'||^2 \ ,$$

où $\mathbf{Y} = [y_{ij}]$ désigne une matrice de données, se ramène à la résolution des problèmes alternés suivants :

$$\arg\min_{\underline{u}} \sum_i \sum_j (y_{ij} - u_i v_j)^2 \quad \text{et} \quad \arg\min_{\underline{v}} \sum_i \sum_j (y_{ij} - u_i v_j)^2 \ .$$

Après convergence, et $\underline{u}$ et $\underline{v}$ étant normés, la valeur singulière d est obtenue par régression,

$$\arg\min_{d} \sum_{i} \sum_{j} (y_{ij} - d u_i v_j)^2 \ .$$

Les éléments singuliers successifs sont alors calculés par déflation de la matrice $\mathbf{Y}$. On notera qu'en compliquant un peu les formules des régressions alternées la procédure peut être adaptée au calcul simultané de plusieurs éléments singuliers.

- K. Ruben Gabriel et Shmuel Zamir (1979) introduisent aussi des régressions alternées (ou *criss cross regressions*) pondérées pour calculer une décomposition en valeurs singulières à l'ordre r sous contrainte de poids pour les observations d'une matrice de données :

$$\arg \min_{(d_k, \underline{u_k}, \underline{v_k})} \sum_{i} \sum_{j} w_{ij} (y_{ij} - \sum_{k=1}^{r} d_k u_{k,i} v_{k,j})^2 \ .$$

Dans leur article, les poids w_{ij} sont des scalaires positifs ou nuls connus, les poids nuls permettant de prendre en compte des données (structurellement) manquantes.

- Antoine de Falguerolles et Brian Francis (1982) utilisent des régressions linéaires généralisées [21] alternées pour estimer, par maximum de vraisemblance, certains modèles bilinéaires (ou bi-additifs) non nécessairement gaussiens.

C.4 Application numérique

Les moindres carrés itérés sont donc susceptibles de permettre d'estimer les coefficients α_L et D dans la formule considérée au début de cet appendice. Comme l'avait constaté Legendre pour le premier ordre et Balmino pour le second ordre, les estimations obtenues ne sont pas très satisfaisantes. Le tableau 3 donne les estimations obtenues par moindres carrés itérés (avec comme valeurs initiales des coefficients $\alpha_L = 0$ et $D = 30000$) et permet de les comparer à celles connues au moment de la publication de Legendre ainsi qu'à celles obtenues par Legendre en appliquant ses moindres carrés généralisés.

En introduisant dans ses formules la longueur approchée du 45$^{\text{ème}}$ degré (soit 28500 modules de deux toises), Legendre considérait d'une certaine façon une observation supplémentaire. L'introduction de cette donnée dans l'approximation du second ordre laisse pratiquement inchangées les estimations obtenues. Des mesures effectuées plus près du pôle et plus près de l'équateur permettraient sans doute d'améliorer les résultats obtenus. Souvenons nous des expéditions de Maupertuis en Laponie et de Bouguer et La Condamine au Pérou ; mais ceci est une autre histoire.

---

21. Rappelons qu'une régression linéaire généralisée est souvent effectuée par une procédure de type moindres carrés pondérés itérés.

Legendre, conscient du caractère décevant des résultats de son ajustement « local », reste cependant convaincu de la portée générale de la méthode des moindres carrés et publie son exemple d'application en l'état.

TABLEAU 3. — Valeurs de la longueur D de l'arc de 1 degré centré au 45^e degré (en modules de 2 toises) et de l'aplatissement α_L de la terre. L'aplatissement connu est « déterminé par la longueur du pendule et par quelques phénomènes astronomiques » et la longueur de l'arc 1 degré telle « qu'on la déduit de la comparaison des mesures faites en France avec les mesures faites au Pérou » (Legendre, 1806, p. 78).

	α_L	D
valeurs connues de Legendre	$\frac{1}{320}$	28504,10
valeurs estimées par Legendre (premier ordre et moindres carrés généralisés)	$\frac{1}{148}$	28497,78
valeurs estimées (second ordre et moindre carrés itérés)	$\frac{1}{150}$	28497,21