Ajustement des données pluviométriques à l'aide des splines plaque mince de lissage: Application à la pluviométrie Camerounaise

Madeleine Nyamsi

Ajustement des données pluviométriques à l'aide des splines plaque mince de lissage: Application à la pluviométrie Camerounaise

Madeleine NYAMSI

Department of computer science
University of Buea, Cameroon
Email : nyamsi_madeleine@yahoo.fr

RÉSUMÉ. Dans cet article, nous présentons une méthode pour le tracé de cartes de pluviométrie. Il s'agit pour cela de lisser et interpoler les mesures des précipitations. La technique des splines, de type plaque mince, est utilisée pour modéliser la surface de lissage et d'ajustement des données pluviométriques. La construction de cette spline nécessite la détermination de paramètres d'ajustement, ce qui fait recours à un problème de minimisation non linéaire multidimensionnelle. Nous proposons une approche heuristique pour la détermination des paramètres d'ajustement. La méthode est testée sur des données pluviométriques de la direction nationale de la météorologie du Cameroun. Les résultats que nous présentons dans cet article confirment la robustesse de la méthode.

ABSTRACT. In this paper we present a numerical method for drawing a rainfall map. We use thin plate adjustment spline to fit a smooth surface through meteorological data such as rainfall observations. The determination of the adjustment parameters gives rise to the minimization of a multidimensional nonlinear function, whose evaluation is very computational intensive. We propose an heuristic approach for the calculation of the adjustment parameters which is cost effective. The proposed method is tested on a set of rainfall observations gathered by the department of national meteorology in Cameroon.

MOTS-CLÉS : pluviométrie, spline, plaque mince, lissage, ajustement, minimisation, météorologie, cumul de données, reconstitution

KEYWORDS : rainfall data, adjustment, minimization, spline, thin plate, meteorology, smooth

0. Je voudrais remercier le Pr. Emmanuel Kamgnia (Department of computer science at Madonna University, Elele - Nigeria) et M. Bernard Philippe (Directeur de recherche à IRISA, Campus de Beaulieu Rennes) tous deux à la retraite, pour avoir initié et suivi ce travail dans le cadre de la collaboration Nord-Sud.

0. Je remercie M. Philippe Richard, Chef Service de la Coopération Technique Internationale, Direction de la Météorologie Nationale du Cameroun et toute son équipe pour leur implication à la réussite de ce projet.

1. Introduction

Dans un bon nombre de domaines, les informations ou données observées sont relevées, stockées parfois en archive et presque inexploitées. Aujourd'hui, la Direction de la Météorologie du Cameroun a en sa possession un ensemble de relevés pluviométriques lus à des stations pluviométriques disposées sur son territoire national. Malheureusement, la répartition de ces stations de mesures n'est pas uniforme et, surtout, ne couvre pas tout le territoire national. Il serait donc utile que la Direction de la Météorologie dispose d'un outil, qui servirait à estimer la hauteur de précipitation à des endroits où les stations n'existent pas, ou bien lorsque les mesures manquent ou sont mal relevées dans certaines stations. Le problème abordé dans cet article est celui de la conception d'un modèle numérique de précipitation, à partir des relevés pluviométriques numérisés. les usages principaux d'un tel modèle étant la capacité d'interpoler la hauteur de précipitation en un point ou ensemble de points.

Le problème de la représentation d'une courbe qui passe par des points donnés est classique en analyse numérique. Pendant longtemps, les polynômes ont été la seule approche utilisée ; les polynômes présentent cependant de nombreux inconvénients, surtout lorsque le nombre de points est élevé. On peut se servir des fonctions splines d'interpolation pour faire passer une courbe lisse par des points donnés.

Les fonctions splines ont la caractéristique de minimiser l'énergie de flexion définie en dimension 1 par :

$$\min \int_a^b \frac{(u^{''}(t))^2}{(1+u'(t)^2)^{\frac{5}{2}}} dt.$$

Mais, mathématiquement, le calcul de cette expression est très difficile. On suppose pour cela $u^{'}$ négligeable et on résout le problème :

$$\min \int_a^b (u^{''}(t))^2 dt.$$

Les bonnes propriétés des splines cubiques d'interpolation en dimension 1 ont conduit à utiliser des splines en dimension supérieure, tout en définissant d'autres critères de minimisation [16], et une notion moins contraignante que l'interpolation : l'ajustement.

L'objectif visé dans cet article est de trouver les paramètres de minimisation sur la base des données existantes, de les utiliser dans les fonctions d'ajustement et d'en déduire les moyennes des mesures aux points météorologiques où les données sont manquantes. D'autre part, une analyse de la robustesse des solutions proposées est faite, tant sur la qualité des résultats que sur les temps mis pour le calcul et l'exploitation des paramètres d'ajustement des dites données.

La détermination des paramètres d'ajustement fait recours à un problème de minimisation non linéaire multidimensionnelle, puis, à la solution d'un système linéaire d'équations assez mal conditionné. Nous proposons une approche heuristique pour la détermination de ces paramètres d'ajustement. Dans la section 2 nous présentons l'environnement de l'étude et posons le problème mathématiquement. Dans la section 3, nous présentons la méthode de Duchon pour la construction de spline plaque mince, ainsi que 2 approches pour le calcul des paramètres d'ajustement. Dans la section 4, nous présentons quelques méthodes utilisées pour le calcul des paramètres d'ajustement. La section 5 présente les résultats expérimentaux. La section 6 conclut l'article en proposant quelques remarques.

2. Présentation mathématique du problème de lissage par la spline

Pour cette partie, nous retenons la présentation de [19]. Soient U, V et Z trois espaces de Hilbert ; soient L une application linéaire continue de U sur V et T une application linéaire continue de U sur Z ; pour notre application, L peut représenter une fonction qui mesure la flexion de la surface, et T l'ajustement des données. Soit z un élément de Z.

Le problème de lissage par la spline consiste à rechercher un élément σ de U tel que

$$\|L(\sigma)\|_V^2 + \rho\|T(\sigma) - z\|_Z^2 = \min_{u\in U} \|L(u)\|_V^2 + \rho\|T(u) - z\|_Z^2 \qquad (1)$$

où $\|.\|_V$ et $\|.\|_Z$ désignent les normes associées à V et Z respectivement, et où ρ est un réel strictement positif. σ est appelé fonction spline de lissage relative à L, T, ρ et z.

Le problème (1) admet une solution pour tout $z \in Z$ si et seulement si $\mathcal{N}(L)+\mathcal{N}(T)$ est fermé dans U. Cette solution est unique si en plus $\mathcal{N}(L) \cap \mathcal{N}(T) = \{0\}$.

3. Lissage par la spline d'ajustement

Dans cette section nous définissons la spline plaque mince d'ajustement et présentons une méthode numérique pour sa construction. Nous commençons par une définition de la courbure globale d'une surface.

3.1. Courbure globale d'une surface

Soit $u(x, y)$ une représentation d'une surface de classe C^2. Soient κ_1 et κ_2 les courbures principales en un point (x, y) de cette surface. Les valeurs possibles de la courbure sont :

– la courbure de Gauss : $\Phi = \kappa_1\kappa_2$,

– la courbure moyenne : $\Psi = \frac{\kappa_1+\kappa_2}{2}$.

On montre dans [8] que

$$\phi = \frac{LN - M^2}{TG - F^2}, \quad \Psi = \frac{1}{2}\frac{NT - 2MF + LG}{TG - F^2},$$

où

$$T = 1 + u_x^2, \quad F = u_x u_y, \quad G = 1 + u_y^2, \quad L = \frac{1}{D}u_{xx},$$

$$M = \frac{1}{D}u_{xy}, \quad N = \frac{1}{D}u_{yy}, \quad D = \sqrt{TG - F^2} = \sqrt{1 + u_x^2 + u_y^2}.$$

Comme Φ et Ψ peuvent s'annuler même pour une surface courbée [8], une meilleure mesure de courbure est donnée par

$$\Upsilon = \kappa_1^2 + \kappa_2^2 = 4\Psi^2 - 2\Phi.$$

Une expression de la courbure plus simple à calculer peut être obtenue, en considérant plutôt la courbure associée à la surface $\bar{u}(x,y) = u(x,y) - g(x,y)$, écart entre $u(x,y)$ et son plan tangent $g(x,y)$. En un point (x_0, y_0) l'équation de $\bar{u}(x,y)$ est donnée par

$$\bar{u}(x,y) = u(x,y) - u(x_0,y_0) - u_x(x_0,y_0)(x-x_0) - u_y(x_0,y_0)(y-y_0).$$

On vérifie que $\bar{u}_x(x_0,y_0) = \bar{u}_y(x_0,y_0) = 0$, $\bar{u}_{xx}(x_0,y_0) = u_{xx}(x_0,y_0)$, et $\bar{u}_{yy}(x_0,y_0) = u_{yy}(x_0,y_0)$. Pour $\bar{u}$ au point (x_0,y_0) on a

$$\bar{\Upsilon} = \bar{\kappa}_1^2 + \bar{\kappa}_2^2 = u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2.$$

Comme la tangente g n'a pas de courbure, la quantité $\bar{\Upsilon} = \bar{\kappa}_1^2 + \bar{\kappa}_2^2$ relative à $\bar{u}$, contient toute l'information sur la courbure de u ; on peut donc l'utiliser pour estimer la courbure de u. La courbure globale de la surface u est donc donnée par :

$$E(u) = \iint_{R^2} ((\frac{\partial^2 u}{\partial x^2})^2 + 2(\frac{\partial^2 u}{\partial x \partial y})^2 + (\frac{\partial^2 u}{\partial y^2})^2)\, dx\, dy \tag{2}$$

On montre [8] que la fonction $(\frac{\partial^2 u}{\partial x^2})^2 + 2(\frac{\partial^2 u}{\partial x \partial y})^2 + (\frac{\partial^2 u}{\partial y^2})^2$ est invariante par rotation de repère, et par conséquent, la mesure $E(u)$ l'est aussi.

3.2. Spline plaque mince d'ajustement

Soient U et V deux espaces de Hilbert. Soit L une application linéaire continue de U sur V et n formes linéaires linéairement indépendantes sur U notées f_i, $i = 1, \ldots, n$. Soient $t^i = (t_1^i, t_2^i), i = 1, \ldots, n$, n points du plan non tous alignés et $z_i, i = 1, \ldots, n$ les mesures à ajuster.

On définit les fonctionnelles suivantes :

$$I_0(u) = \|Lu\|_V^2,$$

et

$$I_\rho(u) = \|Lu\|_V^2 + \sum_{i=1}^{n} \frac{1}{\rho_i}(f_i(u) - z_i)^2,$$

où $\|.\|_V$ désigne la norme de V, et le vecteur de coefficients $\rho = (\rho_i) \in \mathbb{R}^n$ définit des poids qu'on associe aux données. Si U et V vérifient

$$U \subset H^2(\mathbb{R}^2) \text{ et } V = (L^2(\mathbb{R}^2))^4,$$

où l'opérateur L est défini par

$$L(u) = (\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial y^2}, \frac{\partial^2 u}{\partial y \partial x}, \frac{\partial^2 u}{\partial x \partial y}),$$

et les formes linéaires f_i sont définies par

$$f_i(u) = u(t^i),\ i = 1, \ldots, n,$$

alors la fonction $\sigma_0(t)$ solution de

$$\min_{u\in U}\left\{I_0(u)\equiv\iint_{R^2}((\frac{\partial^2 u}{\partial x^2})^2+2(\frac{\partial^2 u}{\partial x\partial y})^2+(\frac{\partial^2 u}{\partial y^2})^2)\,dx\,dy,\ u(t^i)=z_i\right\} \quad (3)$$

est appelée spline plaque mince d'interpolation. Pour un vecteur $\rho\in\mathbb{R}^n$ donné, la fonction $\sigma_\rho(t)$ solution de

$$\min_{u\in U}\left\{I_\rho(u)\equiv\iint_{R^2}((\frac{\partial^2 u}{\partial x^2})^2+2(\frac{\partial^2 u}{\partial x\partial y})^2+(\frac{\partial^2 u}{\partial y^2})^2)\,dx\,dy+\sum_{i=1}^{n}\frac{1}{\rho_i}(u(t^i)-z_i)^2\right\} \quad (4)$$

est appelée spline plaque mince d'ajustement. On montre [4, 5, 14] l'existence d'une solution unique du problème (4), qui s'exprime sous la forme

$$\sigma_\rho(t)=a_0+a_1t_1+a_2t_2+\frac{1}{16\pi}\sum_{i=1}^{n}\lambda_i\|t-t^i\|_2^2\log\|t-t^i\|_2 \quad (5)$$

avec

$$\lambda_i=\frac{1}{\rho_i}(\sigma_\rho(t^i)-z_i) \quad (6)$$

et

$$\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}\lambda_it_1^i=\sum_{i=1}^{n}\lambda_it_2^i=0 \quad (7)$$

où les $\rho_i>0$, $i=1,\ldots,n$ sont des poids ou paramètres d'ajustement, qui représentent une pondération de la contrainte $\sigma_\rho(t^i)$ voisin de z_i. L'expression $E(u)$ de l'équation (2) est une approximation de l'énergie de flexion de la plaque mince infinie d'équation $z=\sigma_\rho(t)$. Avant de donner une méthode de calcul de ces paramètres d'ajustement, nous présentons une méthode numérique de construction de la fonction spline d'ajustement de type plaque mince.

3.3. Méthode numérique de construction de la spline plaque mince d'ajustement

Si on substitue l'égalité (6) dans l'équation (5), on obtient que le système linéaire à résoudre pour obtenir la fonction spline plaque mince, correspondant aux coefficients d'ajustement $\rho_i, i=1,\ldots,n$ est :

$$\begin{cases}\sum_{j=1}^{n}(\frac{1}{16\pi}\|t^i-t^j\|_2^2\log\|t^i-t^j\|_2+\rho_i\delta_{ij})\lambda_j+a_0+a_1t_1^i+a_2t_2^i=z_i \quad i=1,\ldots,n\\ \sum_{j=1}^{n}\lambda_j=\sum_{j=1}^{n}\lambda_jt_1^j=\sum_{j=1}^{n}\lambda_jt_2^j=0\end{cases}$$

où δ_{ij} représente le symbole de Kronecker et (t_1^i,t_2^i) sont les coordonnées de t^i ; le système peut aussi s'écrire sous une forme matricielle :

$$\begin{pmatrix} F & E^T \\ E & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ a \end{pmatrix} = \begin{pmatrix} z \\ 0 \end{pmatrix} \tag{8}$$

où

$$a = (a_0, a_1, a_2)^T, \quad \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$$

, F est la matrice carrée d'éléments (f_{ij}) définie par :

$$f_{ij} = \frac{\|t^i - t^j\|_2^2 \log \|t^i - t^j\|_2}{16\pi}, \ i \neq j, \ i, j = 1, \ldots, n$$
$$f_{ii} = \rho_i, \ i = 1, \ldots, n$$

et la matrice rectangulaire E est :

$$E = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1^1 & t_1^2 & \cdots & t_1^n \\ t_2^1 & t_2^2 & \cdots & t_2^n \end{pmatrix}$$

On a ainsi un système linéaire de $n + 3$ équations à $n + 3$ inconnues. Les λ_i sont appelés les paramètres de lissage. La matrice F de ce système est pleine, symétrique et sa diagonale est formée des éléments ρ_i.
On peut, lorsque le système (8) est de grande taille, utiliser une méthode itérative telle que la méthode du gradient conjugué pour le résoudre, même comme la matrice de ce système n'est pas définie positive ; Sinon, pour des tailles modestes, on utilise une méthode comme celle que nous décrivons.

Soit Q une matrice orthogonale de dimension $n \times n$ telle que $QE^T = R$. R est une matrice de dimension $n \times 3$ qu'on peut partitioner en :

$$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix};$$

où R_1 est une matrice triangulaire supérieure d'ordre 3 . Les matrices Q et R sont obtenues par factorisation QR (Housholder ou Givens) [7]. Soit $Q^T = (Q_1, Q_2)$ un partitionnement de Q^T où Q_1 est une matrice de dimension $n \times 3$ et Q_2 une matrice de dimension $n \times (n - 3)$. On a alors : $QE^T = R$ et en exploitant le partitionnement des matrices Q^T et R, on écrit :

$$Q_1^T E^T = R_1$$
$$Q_2^T E^T = 0.$$

Le système $E\lambda = 0$ peut alors s'écrire $EQ^T\nu = 0$ avec $\nu = Q\lambda$. Soit

$$\nu = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}$$

où ν_1 sont les 3 premières composantes de ν et, ν_2 ses $(n-3)$ dernières composantes. Comme $EQ^T\nu = 0$ et $EQ^T = (R_1^T, 0)$, il s'en suit que

$$R_1^T \nu_1 = 0 \Longrightarrow \nu_1 = 0.$$

On a alors

$$Q_2^T \lambda = \nu_2$$

. Le système

$$F\lambda + E^T a = z$$

peut s'écrire maintenant

$$FQ_2\nu_2 + E^T a = z$$

et, en pré-multipliant par Q_2^T, on obtient :

$$Q_2^T FQ_2\nu_2 = Q_2^T z.$$

La matrice $Q_2^T FQ_2$ étant symétrique définie positive [4, 16], on peut déterminer ν_2 par la méthode de Cholesky. On dérive ensuite $\lambda = Q_2\nu_2$. Enfin, on trouve les coefficients a en résolvant le système

$$R_1 a = Q_1(z - F\lambda).$$

Dans cet article, nous avons utilisé une méthode directe de type factorisation LU pour la solution du système d'équations (8). Il a été prouvé dans [4, 5, 14] l'existence d'une solution unique du problème (4) dont la forme matricielle simplifiée est le système d'équations (8). Le choix de la méthode direct se justifie par le fait que la taille des données est inférieure à 50. Lorsque la matrice est plus grande, elle peut être mal conditionnée, il sera par conséquent important d'utiliser une méthode ittérative. La section suivante présente quelques approches pour le calcul des paramètres d'ajustement.

4. Calcul des paramètres d'ajustement

Nous présentons quelques approches de détermination des paramètres d'ajustement ρ_i. L'objectif est de minimiser l'écart entre les mesures faites sur le terrain et les valeurs estimées par la spline $\sigma_{\rho,k}$ solution de

$$\min_u \{\|Lu\|_V^2 + \sum_{i=1, i\neq k}^{n} \frac{1}{\rho_i}(\sigma_{\rho,k}(t^i) - z_i)^2\},\ u \in U \text{ et } t^i \in R^2,$$

$\sigma_{\rho,k}(t^i)$ étant une estimation de z_i.

4.1. Minimisation par l'approche de validation croisée généralisée

Pour déterminer les paramètres ρ_i d'ajustement optimal, on peut utiliser l'approche classique de validation croisée généralisée que l'on rappelle maintenant.

Soit $K \in \mathbb{R}^{m\times n}$, $m \geq n$ une matrice de rang n et soit $y \in \mathbb{R}^m$. Le problème de minimisation au sens des moindres carrés est de trouver un $x_{opt} \in \mathbb{R}^n$ qui minimise par rapport à x l'expression

$$\|Kx - y\|_2^2.$$

Mais lorsque la matrice K est très mal conditionnée, on transforme le problème initial en un problème mieux conditionné

$$\min_x \{\|Kx - y\|_2^2 + m\lambda\|Lx\|_2^2\},$$

où $L \in R^{m \times n}$, $m \geq n$ est une matrice de normalisation utilisée pour améliorer le conditionnement de K et λ un paramètre de régularisation [3, 6, 7]. Pour résoudre ce problème, on commence par déterminer le paramètre de lissage optimal λ_{gcv}. Pour cela, on doit minimiser par rapport à λ la fonction GCV (Generalized Cross Validation) définie dans [6, 17] par :

$$GCV(\lambda) = m \frac{\|(I - K(K^T K + m\lambda L^T L)^{-1} K^T) y\|_2^2}{(trace(I - K(K^T K + m\lambda L^T L)^{-1} K^T))^2}. \quad (9)$$

Plusieurs approches de solution pour la fonction GCV ont été décrites, la difficulté dans toutes ces approches réside dans l'estimation de la trace de l'inverse d'une matrice. Pour K assez grand, Hutchinson a proposé une méthode stochastique pour l'estimation de la trace de l'inverse d'une matrice.

Si L est une matrice carrée inversible, le problème GCV avec semi norme peut être ramené à un problème GCV sans semi norme [1, 6], il suffit de poser $K_L = KL^{-1}$. Alors, on a :

$$GCV(\lambda) = m \frac{\|(I - K_L(K_L^T K_L + m\lambda I)^{-1} K_L^T) y\|_2^2}{(\mathrm{trace}(I - K_L(K_L^T K_L + m\lambda I)^{-1} K_L^T))^2} \quad (10)$$

Proposition 4.1 *Posons $\mu = m\lambda$ et considérons la décomposition aux valeurs singulières de la matrice $K_L = U\Sigma V^T$, où K_L est de rang n,*

$$\Sigma = \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix},$$

$\Sigma_1 = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n), 0 \in \mathbb{R}^{(m-n) \times n}, U \in \mathbb{R}^{m \times m}$ et $V \in \mathbb{R}^{n \times n}$ sont des matrices orthogonales. Alors

$$\mathrm{trace}(I - K_L(K_L^T K_L + m\lambda I)^{-1} K_L^T) = m - \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \mu}.$$

Preuve. Nous commençons par démontrer que

$$I - K_L(K_L^T K_L + \mu I)^{-1} K_L^T = U(I - D)U^T,$$

où

$$D = \begin{pmatrix} D_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

avec $D_{11} = \mathrm{diag}(\frac{\sigma_1^2}{\sigma_1^2+\mu}, \frac{\sigma_2^2}{\sigma_2^2+\mu}, \cdots, \frac{\sigma_n^2}{\sigma_n^2+\mu})$.

A partir de :

$$\begin{aligned} K_L^T K_L &= V\Sigma^T U^T U \Sigma V^T \\ &= V\Sigma^T \Sigma V^T \\ &= V\Sigma_1^2 V^T, \end{aligned}$$

on a :

$$K_L(K_L^T K_L + \mu I)^{-1} K_L^T = U\Sigma V^T (V\Sigma_1^2 V^T + \mu I)^{-1} V\Sigma^T U^T.$$

Comme

$$\begin{aligned}(V\Sigma_1^2V^T+\mu I)^{-1} &= (V(\Sigma_1^2+\mu I)V^T)^{-1}\\ &= V(\Sigma_1^2+\mu I)^{-1}V^T,\end{aligned}$$

il s'en suit que

$$\begin{aligned}K_L(K_L^TK_L+\mu I)^{-1}K_L^T &= U\Sigma V^TV(\Sigma_1^2+\mu I)^{-1}V^TV\Sigma^TU^T\\ &= U\Sigma(\Sigma_1^2+\mu I)^{-1}\Sigma^TU^T\\ &= U\begin{pmatrix} D_{11} & 0\\ 0 & 0\end{pmatrix}U^T\end{aligned}$$

d'où

$$I-K_L(K_L^TK_L+\mu I)^{-1}K_L^T=U(I-D)U^T.$$

On en déduit que les valeurs propres de $I-K_L(K_L^TK_L+\mu I)^{-1}K_L^T$ sont les éléments de la diagonale de $I-D$. Comme la trace d'une matrice est égale à la somme de ses valeurs propres, il s'en suit que :

$$\mathrm{trace}(I-K_L(K_L^TK_L+m\lambda I)^{-1}K_L^T)=m-\sum_{i=1}^{n}\frac{\sigma_i^2}{\sigma_i^2+\mu} \tag{11}$$

□

En utilisant une approche stochastique pour estimer la trace, on montre dans [1, 6] que la fonction GCV peut être approchée par :

$$GCV(\lambda)\approx m\frac{y^Ty-v^T(Q+\mu I)^{-1}v-\mu v^T(Q+\theta I)^{-2}v}{(m-u^T(Q+\mu I)^{-1}u)^2}, \tag{12}$$

où $Q=K_L^TK_L$, $v=K_L^Ty$, $u=K_L^Tr$, $\mu=m\lambda$ et r représente un vecteur dont les composantes sont dérivées de variables aléatoires et sont égales à $\{-1,1\}$ avec la même probabilité 1/2.

Le processus de minimisation en utilisant l'approche GCV est, du point de vue du nombre d'évaluation de la fonction, très coûteux. Or la fonction GCV est écrite en utilisant deux fonctions de base : $v^T(Q+\mu I)^{-1}v$ et $v^T(Q+\mu I)^{-2}v$. Q est une matrice symétrique et parfois (en fonction des applications) définie positive ou semi-definie positive. Golub [6] a montré comment utiliser la méthode quadratique de Gauss pour évaluer la fonction $GCV(\lambda)$ en exploitant l'algorithme de Stieltjes et la bidiagonalisation de Lanczos.

Soient $f_v(\mu)$ et $g_v(\mu)$, deux fonctions scalaires définies par :

$$f_v(\mu)=\frac{v^T}{\|v\|_2}(Q+\mu I)^{-2}\frac{v}{\|v\|_2}, \tag{13}$$

et

$$g_v(\mu)=\frac{v^T}{\|v\|_2}(Q+\mu I)^{-1}\frac{v}{\|v\|_2}. \tag{14}$$

La fonction GCV peut simplement être réécrite comme suit :

$$GCV(\lambda) \approx m\frac{y^T y - \|v\|_2^2(g_v(\mu) + \mu f_v(\mu))}{(m - \|u\|_2^2 g_v(\mu))^2} \tag{15}$$

Pour deux (2) vecteurs de départ u et v, toute évaluation de la fonction GCV fait intervenir les évaluations des fonctions $f_v(\mu)$, $g_v(\mu)$ et $g_u(\mu)$. Sidjé et al. [1, 17] ont développé des algorithmes basés sur le procédé de Lanczos pour l'évaluation rapide de la fonction GCV.

4.2. Approche par validation croisée

L'idée ici est de trouver les valeurs du paramètre d'ajustement ρ qui minimise au sens des moindres carrés l'erreur absolue entre la mesure estimée par les splines $\sigma_{\rho,k}$ aux points t^k, et la mesure réelle z_k pour $k = 1, \ldots, n$. Mathématiquement, on écrit :

$$\min_{\rho}(V(\rho)) = \sum_{i=1}^{n} \frac{1}{\rho_i}(\sigma_{\rho,i}(t^i) - z_i)^2.$$

Ce qui revient à résoudre un problème de minimisation multidimensionnelle au sens des moindres carrés non linéaires.

Considérons l'expression de la spline donnée dans les équations (5), (6) et (7). En remplaçant λ_i par son expression dans l'écriture de $\sigma_\rho(t^i)$, il vient :

$$\sigma_\rho(t^i) = a_0 + a_1 t_1^i + a_2 t_2^i + \frac{1}{16\pi}\sum_{j=1}^{n}(z_j - \sigma_\rho(t^j))\frac{\|t^i - t^j\|_2^2 \log\|t^i - t^j\|_2}{\rho_j}.$$

Posons $A = (\alpha_{i,j})$ une matrice carrée d'ordre n définie par :

$$\alpha_{i,j} = \frac{\|t^i - t^j\|_2^2 \log\|t^i - t^j\|_2}{16\pi\rho_j} \quad i \neq j, \quad \alpha_{i,i} = 0.$$

Alors,

$$\sigma_\rho(t^i) = a_0 + a_1 t_1^i + a_2 t_2^i + \sum_{j=1}^{n}\alpha_{i,j} z_j - \sum_{j=1}^{n}\alpha_{i,j}\sigma_\rho(t^j),$$

qui peut se mettre sous la forme matricielle

$$(I + A)\mu = E^T a + Az$$

où : $\mu = (\sigma_\rho(t^1), \ldots, \sigma_\rho(t^n))^T$ et $a = (a_0, a_1, a_2)^T$. Soit D une matrice diagonale telle que $D(i,i) = \rho_i$; Alors on peut vérifier

$$F = (I + A)D,$$

où F est la matrice du système (8). La matrice F étant inversible il s'en suit que $(I + A)$ est inversible et on écrit : $\mu = (I + A)^{-1}E^T a + (I + A)^{-1}Az$.
Posons $B = (I + A)^{-1}E^T$ et $C = (I + A)^{-1}A$. Alors

$$\mu = Ba + Cz.$$

Considérons le vecteur w formé des z_j, $j = 1, \ldots, n$, dont la i^{eme} composante est remplacée par $\sigma_{\rho,i}(t^i)$. Alors,

$$\sigma_{\rho,i}(t^i) = e_i^T Ba + e_i^T Cw,$$

et après quelques manipulations algébriques on a :

$$(1 - c_{ii})(\sigma_{\rho,i}(t^i) - z_i) = e_i^T Ba + e_i^T Cz - z_i.$$

C'est à dire :

$$\sigma_{\rho,i}(t^i) - z_i = \frac{e_i^T(Ba + Cz - z)}{1 - c_{ii}}.$$

Si la matrice C était circulante, toutes les composantes c_{ii} seraient égales et on aurait :

$$1 - c_{ii} = 1 - \frac{1}{n} trace(C).$$

Et par conséquent

$$V(\rho) = \sum_{i=1}^{n} \frac{(B(\rho)a + (C(\rho) - I)z)_i^2}{\rho_i(1 - c_{ii})^2} = \frac{\|B(\rho)a + (C(\rho) - I)z\|_\rho^2}{(1 - \frac{1}{n} trace(C))^2}.$$

Le résultat reste vrai même si la matrice C n'est pas circulante [4, 6] et, le problème de minimisation de $V(\rho)$ peut donc être résolu en utilisant l'approche GCV avec ou sans semi-norme.

Si on suppose tous les ρ_i égaux, le problème est ramené à celui de la minimisation de la fonction d'une variable réelle

$$V(\rho) = \frac{1}{\rho} \sum_{i=1}^{n} (\sigma_{\rho,i}(t^i) - z_i)^2 \quad (\rho \in \mathbb{R}).$$

Quelle que soit l'approche de minimisation, chaque évaluation de $V(\rho)$ nécessite la construction de n splines $\sigma_{\rho,i}$, $i = 1, \ldots, n$, ce qui rend l'approche très coûteuse.

4.3. Approche heuristique

Soient a et b deux réels strictement positifs tels que $a < b$. On considère l'intervalle $[a, b]$ de $\mathbb{R}$. On se donne un entier positif m et on définit un pas $h = \frac{b-a}{m}$ dans $[a, b]$. Soient $\rho_j = a + jh$, $j = 0, \ldots, m$, $(m + 1)$ valeurs de ρ dans $[a, b]$. Considérons la matrice $M = (\xi_{ij})$, $M \in \mathbb{R}^{n \times (m+1)}$ où $\xi_{ij} = (\sigma_{\rho_j,i}(t^i) - z_i)$.
M est la matrice des erreurs absolues entre la valeur lisse obtenue de la spline $\sigma_{\rho,k}$ et la mesure z_k.
Dans les deux procédés de détermination du ρ optimal on fait usage d'une norme vectorielle. L'un des procédés normalise d'abord la matrice M des erreurs absolues. L'autre par contre, passe l'étape de normalisation et extrait la valeur de ρ optimale, pour laquelle la norme vectorielle appliquée aux colonnes de M est la plus petite.

Procédé de normalisation

Initialement , on se fixe :

1) une marge d'erreur absolue qu'on note θ_a. C'est l'erreur absolue maximale qu'on déduit de l'information sur les mesures initiales.

2) une marge d'erreur relative autorisée notée θ_r. On peut la considérer comme étant la moyenne des erreurs relatives ou alors le minimum des erreurs relatives sur les mesures.

3) une valeur seuil notée η, qui dépend des mesures, et à partir de laquelle on ajuste les données.

Pour une valeur quelconque de ρ_j donnée dans $[a, b]$ on a $\xi_{ij} = \sigma_{\rho_j,i}(t^i) - z_i$, $\forall i = 1, \ldots, n$ et $j = 1, \ldots, m + 1$. Alors,

$$\begin{cases} \text{si} \quad z_i \leq \eta \quad \text{alors} & \tau_{ij} = \frac{\xi_{ij}}{\theta_a} \\ \text{sinon} & \tau_{ij} = \frac{\xi_{ij}}{\theta_a + \theta_r(z_i - \eta)} \end{cases}$$

On construit, à partir de cette définition, une matrice normalisée $P \in R^{n\times(m+1)}$ avec $P = (\tau_{ij})$.
Une fois que les poids à associer aux mesures sont déterminés, on peut déterminer les paramètres λ_i, $i = 1, \ldots, n$ et a_j, $j = 0, 1, 2$ de la spline plaque mince d'ajustement σ_ρ.

5. Résultats Expérimentaux

5.1. Présentation de l'environnement de nos tests

Le Cameroun est un pays d'Afrique centrale très étendu en latitude, situé aux coordonnées polaires $6^\circ N$ et $12^\circ E$. Riverain du bassin du Congo au sud, il atteint au nord les rives sahéliennes du Lac Tchad. Bordé par l'océan Atlantique, le pays est dominé par l'un des massifs montagneux les plus hauts d'Afrique. L'ensemble constitue une très grande variété de domaines bio-géographiques. Le Cameroun est séparé en deux grands domaines climatiques : le domaine équatorial et subéquatorial, au sud, et les domaines tropicaux au nord.

Le climat est caractérisé par des précipitations abondantes et en particulier à Debunsha, second point le plus pluvieux au monde. En fonction de la latitude avec des modulations dues au relief, le climat tropical est de trois types très différents : la pluviométrie s'abaisse, la durée de la saison sèche augmente. Les moyennes de pluies annuelles par station sont présentées dans la figure 2 pour l'année 2012.

5.2. Présentation des données utilisées

Pour tester la robustesse de l'application que nous avons développé, nous avons utilisé les données pluviométriques de la Direction de la Météorologie du Cameroun. Ces mesures ont été relevées dans les 41 stations pluviométriques irrégulièrement réparties sur le territoire camerounais comme l'indique la figure 2. Afin de constituer la base de données en entrée de notre application, nous avons procédé en plusieurs étapes parmi lesquelles :

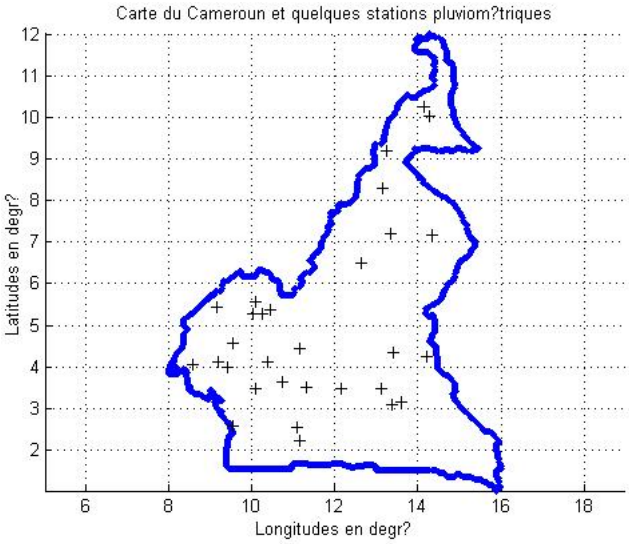

Figure 1. *Stations pluviométriques*

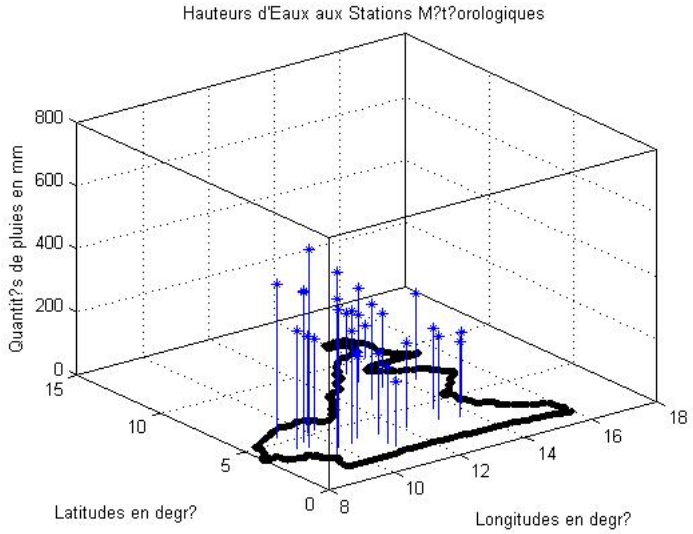

Figure 2. *Hauteurs moyennes des pluies aux stations pluviométriques*

– **Numérisation des données quotidiennes** : les relevés étant généralement sur des supports papiers, nous les avons fait saisir et vérifier à plusieurs niveaux par les agents de la Direction de la Météorologie Nationale. En effet, les relevées se font à une heure précise de la journée sur tout le territoire national ; Les relevés sont gradués en millimètres (mm). Les données saisies sont des quantités de pluies journalières, par station météorologique et sur la majorité des stations qui recouvrent le pays. Cependant, sur certaines plages de dates, les relevés étaient manquants et, lorsque les quantités valent moins de un millimètre ($1\ mm$), on les ramène à zero millimètre ($0\ mm$).

– **Épuration des données** : sur la base des connaissances accumulées et de l'expertise des services de la météorologie, certaines données douteuses ont été retirées et remplacées par le signe '$-$', signifiant données manquantes ; Ces valeurs saisies étaient soit en contradiction avec les saisons, soit en contradiction avec la zone géographique. De même les relevés relatives à la station de Debundsha (longitude 8.59 et latitude 4.06), qui sont trop élevés par rapport aux autres, ont été retirés de la liste pour éviter de biaiser les résultats : il faut rappeler que Debundsha est le second point le plus pluvieux du monde. Le modèle peut être utilisé pour estimer les mesures aux points où les relevés douteux ont été rejetés.

– **Construction des échantillons de données** :pour construire les paramètres λ et a du modèle mathématique présenté dans les sections 3.2, 3.3 et 4, nous avons utilisé des moyennes de pluviométrie, plutôt que des cumuls ; Les moyennes sont soit mensuelles, soit annuelles. L'année 2000 a été choisie pour l'échantillon des moyennes mensuelles à cause de la régularité et de la disponibilité des données à plusieurs stations. Les moyennes annuelles se portent sur les années 1974 à 2000. Nous avons aussi construit un échantillon de données pour valider les résultats de notre simulation.

5.3. Présentation des résultats

Nous rappelons que l'objectif de notre travail était de concevoir un outil que la Direction de la Météorologie du Cameroun pourrait utiliser pour estimer les niveaux de précipitations là où les stations de mesure n'existent pas, ou là où elles ont été mal relevées. Nous avons à cet effet conçu un modèle mathématique de type splines plaques minces en dimension 2, et une méthode heuristique pour le calcul des paramètres d'ajustement de la spline. Les programmes ont été implémentés sous Matlab, sur un ordinateur PC i5. Nous présentons dans un premier temps les splines plaques minces de notre modèle mathématique ; Nous présentons ensuite un tableau comparatif des valeurs obtenues par les modèles de $V(\rho)$ à des points où ces mesures sont connues, ceci pour nous permettre de tester la robustesse de notre modèle.

5.3.1. Splines Plaques Minces

La fonction $V(\rho)$ que nous utilisons pour faire de l'ajustement des mesures est une fonction qui n'est pas monotone ; Elle admet plusieurs minimums locaux, les plus petits se situant vers l'extrémité supérieur de l'intervalle (figure 3). Nous nous sommes pour cela limités à l'intervalle $[10^{-3}, 10^{-1}]$ dans lequel on ne fait ni de l'interpolation ni le lissage maximal.

Nous présentons quatre splines dérivées à partir des mesures relevées sur 41 stations pluviométriques. Les mesures utilisées sont les moyennes mensuelles de l'année 2000 ; Nous avons choisi cette année à cause de la régularité et de la disponibilité des données à plusieurs stations. La spline de la figure 4 est la spline d'interpolation ($\rho = 0$). La spline de la figure 5 est celle obtenue par la méthode heuristique avec normalisation sur les moyennes mensuelles de pluies ; La norme utilisée est la norme 2. Elle est très proche de la spline d'interpolation. La spline de la figure 6 est la spline d'ajustement obtenue par minimisation multidimensionnelle de la fonction $V(\rho)$. On utilise le minimum du vecteur minimal $V(\rho)$ pour le calcul des paramètres d'ajustement de la spline. Elle est très proche des points de l'espace et se rapproche au mieux de la spline de la méthode heuristique avec la norme 2 et par conséquent de la spline d'interpolation. Enfin, la spline de la la figure 7 est la spline d'ajustement obtenue par minimisation multidimensionnelle de la fonction

$V(\rho)$. Contrairement à la figure précédente, on utilise tout le vecteur minimum pour le calcul des paramètres d'ajustement de la spline. Elle est très proche de l'interpolation en quelques points seulement de l'espace. Cette démarche est moins précise que la méthode heuristique utilisée avec la norme 2.

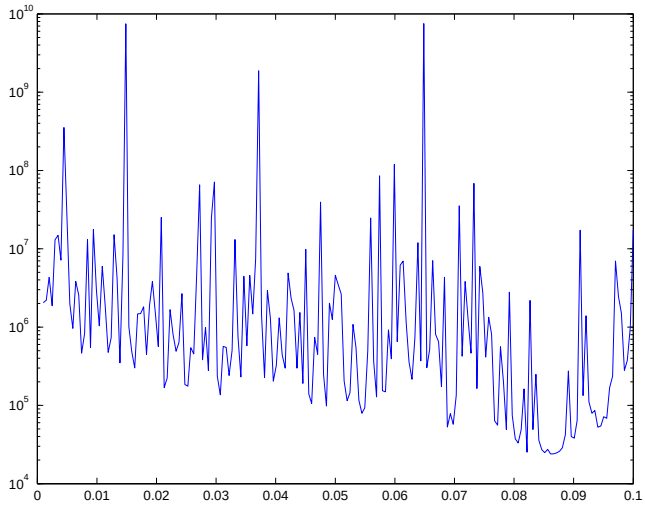

Figure 3. *Courbe de $V(\rho)$ pour la minimisation d'une variable*

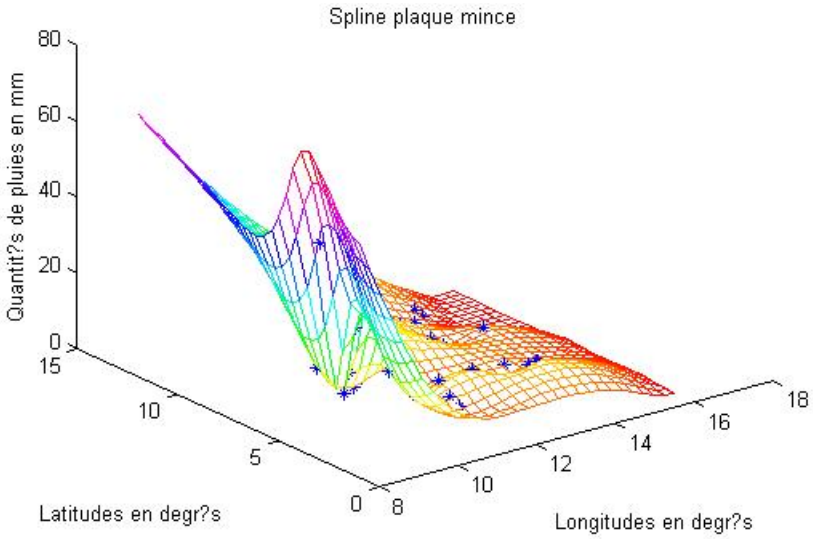

Figure 4. *Spline d'interpolation ($\rho = 0$)*

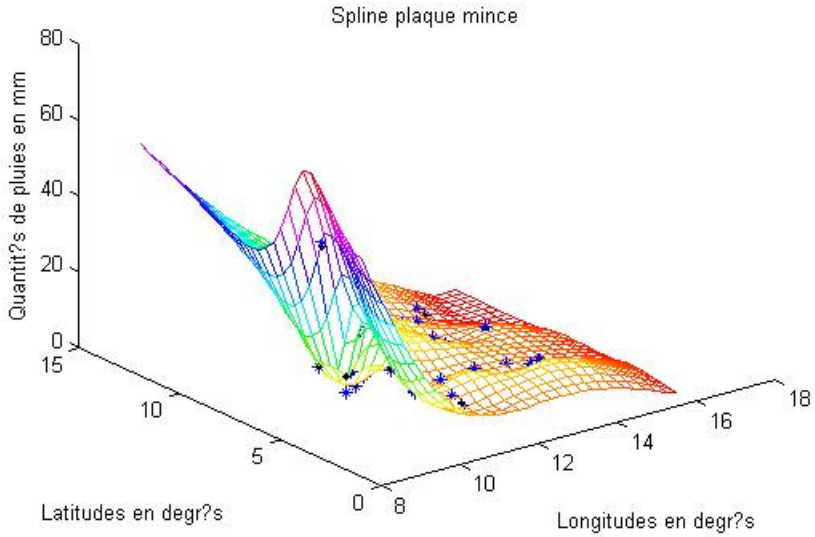

Figure 5. *Spline d'ajustement : Méthode heuristique avec la norme 2*

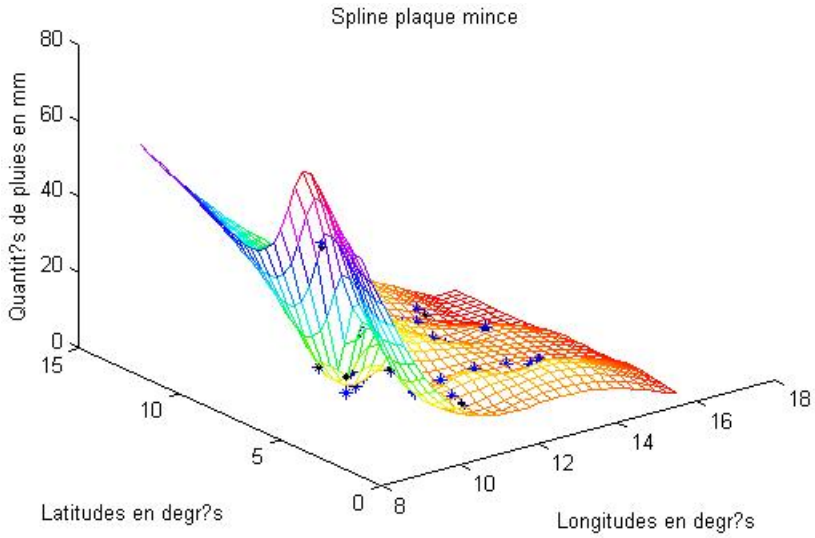

Figure 6. *Spline d'ajustement :Méthode multidimensionnelle avec le minimum du vecteur*

5.3.2. Robustesse du Model

Pour tester la robustesse du model, nous avons utilisé les quatre splines que nous venons de décrire pour estimer les quantités de pluies à certaines stations où les mesures relevées sur le terrain sont disponibles à la Direction de la Météorologie du Cameroun. Nous avons supprimé une mesure lors des simulations, dans le but de la retrouver à partir des différentes méthodes utilisées. Nous présentons dans le tableau 1 les résultats comparatifs

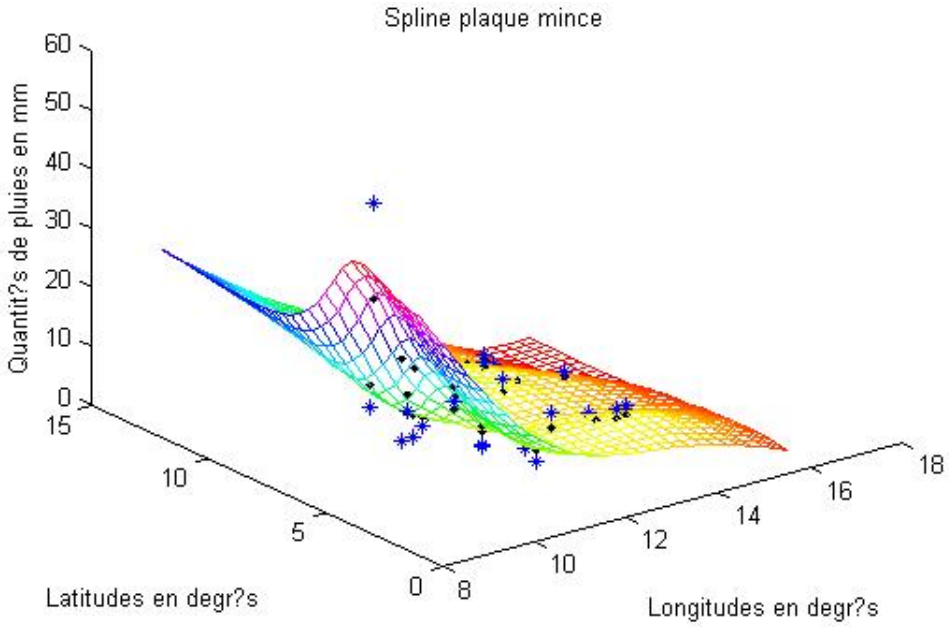

Figure 7. *Spline d'ajustement : Méthode multidimensionnelle utilisant le vecteur minimal* pour le point de coordonnées (13.11, 3.46). Les mesures sont des moyennes mensuelles sur les douze mois de l'année 2000.

(L=13.11, l=3.46)		**Mesures expérimentales**				
Mois	Mesures	Dim. 1	Heur.	Dim 1 ρ min.	Mult. Dim	Mult. Dim ρ min.
1	0.0	0.4	0.4	0.2	0.6	0.2
2	0.9	1.0	1.0	1.0	1.0	0.9
3	2.8	4.7	3.1	4.6	4.8	3.9
4	4.8	4.7	4.7	3.9	5.0	4.7
5	5.6	5.3	5.5	5.2	5.3	5.5
6	5.9	6.1	6.0	5.3	6.1	6.0
7	1.9	2.0	2.0	0.4	2.2	2.1
8	4.1	4.8	4.2	4.5	5.5	4.8
9	5.5	9.3	9.8	7.1	9.3	5.9
10	12.1	11.0	11.3	11.2	11.2	11.4
11	1.5	2.5	1.6	2.6	2.9	1.7
12	1.8	0.8	0.8	0.8	0.8	0.8

Tableau 1. *Tableau comparatif des moyennes mensuelles aux coordonnées (13.11, 3.46)*

Il ressort de la courbe des erreurs relatives (9) que, la spline plaque mince d'ajustement construite par la méthode multidimensionnelle, avec la plus petite valeur du vecteur minimal sur les moyennes de pluies (figure 6), et la spline plaque mince d'ajustement, construite par la méthode heuristique avec la norme 2 sur la moyenne des pluies (figure 5), sont plus robustes. Pour ces deux splines, les erreurs relatives sont de l'ordre de $10^{-}2$, ordre de précision accepté à la Direction de la Météorologie du Cameroun.

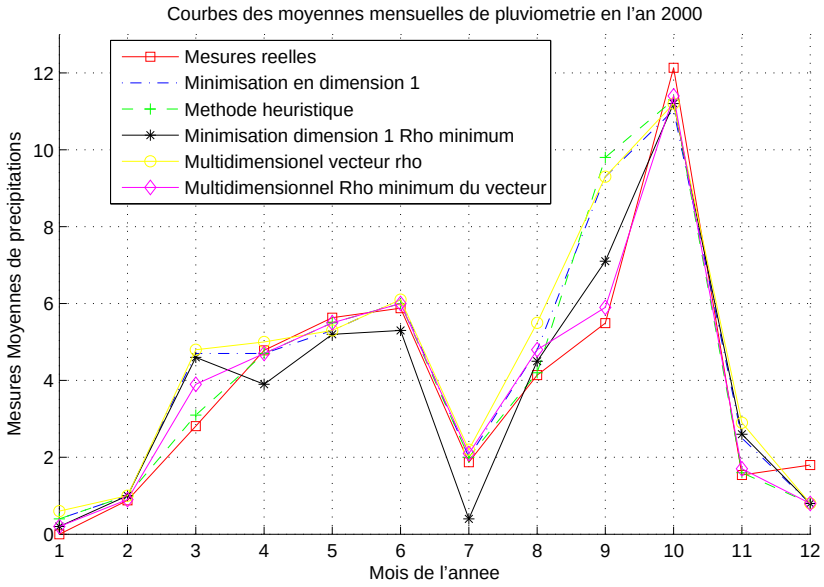

Figure 8. *Courbe comparative des moyennes mensuelles au point (13.11, 3.46)*

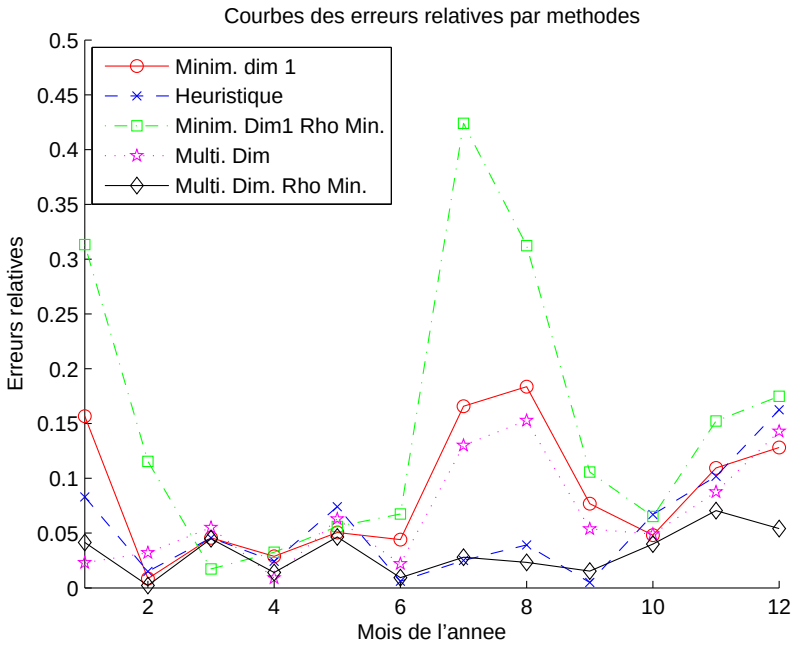

Figure 9. *Courbe comparative des erreurs relatives au point (13.11, 3.46)*

6. Conclusion

Dans cet article nous avons utilisé la fonction spline d'ajustement de type plaque mince pour modéliser une surface de lissage et d'ajustement des données pluviométriques. On en a déduit une méthode numérique pour le tracé de carte de pluviométrie du Cameroun. La quantité assez faible des données ainsi que leur répartition ne nous ont cependant pas permis d'obtenir une carte de pluviométrie suffisamment représentative. La méthode heuristique que nous présentons pour la calcul des paramètres de lissage est assez robuste, la spline qui en découle nous a permis d'estimer des moyennes de pluie à certains points, avec un ordre de précision acceptable à la Direction de la Météorologie du Cameroun. Il faudrait cependent confirmer cette robustesse sur une quantité de données plus grande et régulièrement réparties sur le territoire.

7. Bibliographie

[1] I. ALTAS, K. BURRAGE, M. HEGLAND, S. ROBERTS, R.B. SIDJE, « Extended Fast Generelised Cross Validation for Data Mining Applications », DEPARTMENT OF MATHEMATICS, UNIVERSITY OF QUEENSLAND, AUSTRALIA, 1999.

[2] ARNOLD NEUMAIER, « Solving Ill-Conditioned And Singular Linear Systems : A Tutorial on Regularization », *Siam Review*, vol. 3, n° 40, 636-666, September 1998.

[3] AKE BJÖRK, « Numerical Methods for Least Squares Problems », *SIAM*, 1996.

[4] CLAUDE CARASSO, « Lissage de Données à l'aide de fonctions splines », *Analyse Numérique, Ed. Herman*, 357-415, 1991.

[5] JEAN DUCHON, « Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces », *R.A.I.R.O.*, vol. 10, n° 12, 5-12, December 1976.

[6] G. H. GOLUB, URS VON MATT, « Generalized cross-validation for large scale problems », *Journal of computational and graphical Statistics*, vol. 6, n° 1, 1-34, 1997

[7] G.H. GOLUB, C. F. VAN LOAN « Matrix Computations », THE JOHNS HOPKINS UNIVERSITY PRESS, 2ED 1989

[8] G. R. GONÇALVES, « Qualités requises en modélisation numérique de relief : Etude de l'équivalence entre modèle carthographique et modèle numérique », THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE MARNE-LA-VALLÉE PRÉSENTÉE ET SOUTENUE LE 17 NOVEMBRE 1999 PAR GIL RITO CONÇALVES

[9] M. F. HUTCHINSON, « Interpolating mean rainfall using thin plate smoothing splines », *International Journal of Geographical Information Systems* n° 9, 385-403, 1995.

[10] M. F. HUTCHINSON, « Interpolation of rainfall data with thin plate smoothing splines-Part I : two dimensional smoothing of data with short range correlation » *Journal of Geographic Information and Decision Analysis* vol. 2, n° 2, 139-151,1998a

[11] M. F. HUTCHINSON, « Interpolation of rainfall data with thin plate smoothing splines-Part II : analysis of topographic dependence » *Journal of Geographic Information and Decision Analysis* vol. 2, n° 2, 152-167,1998b

[12] M. F. HUTCHINSON, « A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines » *Communications in Statistics,Simulation and Computation* n° 19, 433-450, 1990.

[13] M. F. HUTCHINSON, P. E. GUESSLER, « Splines-more than just a smooth interpolator » *Geoderma* n° 62, 45-67, 1994.

[14] JEAN MEINGUET, « Multivariate Interpolation at Arbitrary Points Made Simple », *Journal of Applied Mathematics and Physics (ZAMP)*, n° 30, 292-304, 1979.

[15] M. NYAMSI, E. KAMGNIA AND B. PHILIPPE, « Ajustement des données météorologiques : Application à la pluviométrie du Cameroun », *Proceeding of the 6th Africam Conference on Research in Computer Science (CARI02), Yaoundé, 411-418, 2002*.

[16] R. SIBSON AND G. STONE, « Computation of Thin Plate Spline », *SIAM J. Sci. Statist. Comput*, n° 12, 1304-1313, 1991.

[17] R. B. SIDJE AND A. B. WILLIAMS, « Fast Generelised Cross Validation », *Department on Mathematics, University of Queensland, Australia*, 1996.

[18] A. TAIT, R. HENDERSON, R. TURNER AND X. ZHENG « Thin plate smoothing spline interpolation of daily rainfall for New Zealand using climatological rainfall surface » *Int. J. Climatol*, n° 26, 2097-2115, 2006.

[19] M. TAZEROUALTI, « Modélisation de surfaces à l'aide de fonctions splines », THÈSE DE DOCTORAT DE L'UNIVERSITÉ JOSEPH FOURRIER I, PRÉSENTÉE LE 26 FÉVRIER 1993 PAR MOHAMMED TAZEROUALTI

[20] X. ZHENG, R. BASHER, « Thin-plate smoothing spline modelling of spatial climate data and its application to mapping South Pacific rainfalls », *Monthly Weather Review*, n° 123, 3086-3102, 1995.