



HAL
open science

Logiciel de résolution d'un problème inverse à 1 variable.

I: Exposé des méthodes

C. Delannoy, François Lefeuvre

► To cite this version:

C. Delannoy, François Lefeuvre. Logiciel de résolution d'un problème inverse à 1 variable. I: Exposé des méthodes. [Rapport de recherche] Centre de recherches en physique de l'environnement terrestre et planétaire (CRPE). 1981, 44 p. hal-02191520

HAL Id: hal-02191520

<https://hal-lara.archives-ouvertes.fr/hal-02191520>

Submitted on 23 Jul 2019

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

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

RP

182

(44)

**CENTRE NATIONAL D'ETUDES
DES TELECOMMUNICATIONS**

**CENTRE NATIONAL DE LA
RECHERCHE SCIENTIFIQUE**

**CENTRE DE
RECHERCHES
EN PHYSIQUE DE
L'ENVIRONNEMENT
TERRESTRE
ET PLANETAIRE**

CRPE

**NOTE TECHNIQUE
CRPE / 80**

110

**LOGICIEL DE RESOLUTION
D'UN PROBLEME INVERSE
A 1 VARIABLE**

I. Exposé des méthodes

C.N.R.S.
Centre de Documentation
Scientifique et Technique
Bibliothèque

Par

C. DELANNOY & F. LEFEUVRE

C.R.P.E. ORLEANS

16 SEP. 1981

CENTRE DE RECHERCHE EN PHYSIQUE DE
L'ENVIRONNEMENT TERRESTRE ET PLANETAIRE

NOTE TECHNIQUE CRPE/80

LOGICIEL DE RESOLUTION D'UN PROBLEME INVERSE A 1 VARIABLE

I - EXPOSE DES METHODES

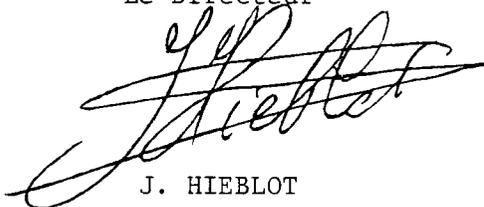
par

C. DELANNOY et F. LEFEUVRE

C.R.P.E./P.C.E.

45045 - ORLEANS CEDEX

Le Directeur



J. HIEBLOT

mars 1981

PLAN

1. INTRODUCTION

2. APPROCHE SPECIFIQUE DU PROBLEME

2.1. Introduction

2.2. Conditionnement des données

2.2.1. Généralités

2.2.2. Méthode proposée

2.2.3. Discussion

2.3. Validation des solutions

2.3.1. Stabilité

2.3.2. Pouvoir de prédiction

2.3.3. Pouvoir de résolution

2.4. Discussion (choix interactif d'un compromis)

3. LES METHODES D'INVERSION

3.1. Généralités

3.1.1. Introduction

3.1.2. Position du problème

3.2. Méthode de Backus et Gilbert

3.2.1. Définition de la méthode

3.2.1.1. Contexte d'application

3.2.1.2. Noyaux filtrants

3.2.1.3. Sélection des noyaux filtrants

3.2.1.4. Noyaux de Dirichlet

3.2.2. Techniques numériques spécifiques

3.2.3. Validation des solutions

3.2.3.1. Pouvoir de résolution

3.2.3.2. Pouvoir de prédiction

3.2.3.3. Stabilité

3.2.4. Discussion

- 3.3. Méthode de discrétisation
 - 3.3.1. Définition de la méthode
 - 3.3.1.1. Cadre général
 - 3.3.1.2. Méthode effectivement employée
 - 3.3.2. Techniques numériques
 - 3.3.3. Validation des solutions
 - 3.3.3.1. Pouvoir de résolution
 - 3.3.3.2. Pouvoir de prédiction
 - 3.3.3.3. Stabilité
 - 3.3.4. Discussion
- 3.4. Méthode du maximum d'entropie
 - 3.4.1. Définition de la méthode
 - 3.4.1.1. Introduction
 - 3.4.1.2. Méthode
 - 3.4.1.3. Recherche pratique de la solution
 - 3.4.1.4. Prise en compte des erreurs
 - 3.4.2. Méthode programmée
 - 3.4.3. Recherche d'une solution initiale
 - 3.4.4. Validation des solutions
 - 3.4.4.1. Pouvoir de résolution
 - 3.4.4.2. Pouvoir de prédiction
 - 3.4.4.3. Stabilité
 - 3.4.5. Discussion

BIBLIOGRAPHIE

1. INTRODUCTION

Cette note est la première d'une série de 3 décrivant un logiciel permettant de restituer une fonction réelle non négative d'une variable réelle, lorsqu'on en connaît un certain nombre de fonctionnelles linéaires. Plus précisément, on cherche à déterminer sur un intervalle réel I , une fonction f , connaissant les n valeurs \hat{s}_i liées à f par les relations

$$(1.1.) \quad \hat{s}_i = \int_I a_i(x) f(x) dx \quad i = 1, n$$

Les n fonctions connues a_i nommées noyaux sont supposées de carrés intégrables. On fait également l'hypothèse que les erreurs sur les données \hat{s}_i sont statistiquement indépendantes.

Le problème ainsi exposé est un "problème inverse". Il est généralement "mal posé" en ce sens qu'il admet une infinité de solutions. Le logiciel proposé a pour principal objectif la recherche de la solution particulière de (1.1.) qui possède l'entropie maximum (Lefeuvre, 1977). Une telle recherche nécessite généralement la connaissance d'une solution approchée que l'on calcule par l'une des 2 méthodes suivantes :

- recherche d'une fonction discrète positive satisfaisant les données (Daval et Al., 1974),
- recherche d'une solution "moyenne" celle des "noyaux de Dirichlet" qui est un cas particulier de la méthode de Backus et Gilbert (Backus et Gilbert, 1967, 1968, 1970 ; Sabatier, 1974).

Chacune de ces 2 méthodes peut constituer une fin en soi ; le logiciel permet, en définitive, de réaliser l'une des quatre opérations suivantes :

1. Recherche d'une solution particulière discrète.
2. Recherche d'une solution moyenne.
3. Recherche de la solution "maximum d'entropie" initialisée à partir de l'une des solutions 1 ou 2.
4. Recherche de la solution "maximum d'entropie" initialisée par l'utilisateur lui-même.

Ce logiciel a été développé et appliqué dans le cadre de l'analyse de champs d'ondes électromagnétiques aléatoires. (Storey, 1971 ; Storey and Lefeuvre, 1974 ; Lefeuvre, 1977 ; Lefeuvre and Delannoy, 1978).

Cependant il a été conçu dans un esprit suffisamment général pour permette de résoudre tout problème inverse du type défini en (1.1.).

Cette première note est consacrée à l'exposé des méthodes. Nous y décrirons dans le chapitre 2 les principaux concepts qui interviennent dans notre approche du problème inverse, et ceci, indépendamment de la méthode choisie pour le résoudre. Notamment, on montrera comment il est possible de conditionner les n données \hat{s}_i et l'incidence que cela peut avoir sur les solutions que l'on jugera à travers certains critères de stabilité, de prédiction et de résolution que nous serons amenés à définir.

Dans le chapitre 3, nous décrirons les 3 méthodes évoquées ci-dessus sous un triple aspect :

- + définition de la méthode,
- + critères de validation des solutions,
- + techniques numériques à mettre en jeu pour aboutir à une solution.

La note numéro II précisera comment utiliser le logiciel proposé. La note III sera essentiellement une description détaillée du logiciel en vue de son éventuelle maintenance.

2. APPROCHE SPECIFIQUE DU PROBLEME

2.1. INTRODUCTION

Pour résoudre le problème inverse décrit par la formule (1.1.), de nombreuses méthodes numériques peuvent être utilisées.

Cependant, dans tous les cas, pour obtenir une solution numérique satisfaisante, on a tout intérêt à utiliser des données convenablement "conditionnées". Dans ce logiciel on a choisi une méthode de conditionnement "intrinsèque", c'est à dire indépendante des données et de leurs erreurs ; nous justifions ce point de vue d'une part par la rapidité de traitement qui en découle, d'autre part par un souci de ne pas optimiser la recherche d'une solution sur la base d'erreurs généralement mal connues.

Par ailleurs, on notera qu'on ne cherche pas à obtenir n'importe quelle solution au problème (1.1.), mais plutôt à déterminer une solution ayant une signification physique. On pourra juger de la qualité de la solution obtenue grâce à 3 critères qui sont :

- l'instabilité de la solution définie comme la susceptibilité de la solution aux erreurs sur les données \hat{s}_i ,
- le pouvoir de prédiction qui traduira l'aptitude de la solution à décrire les données, compte tenu des erreurs.
- le pouvoir de résolution qui exprimera la finesse avec laquelle la solution est apte à décrire les détails de la véritable fonction ayant engendré les données.

Les 2 premiers critères s'exprimeront très simplement par une valeur réelle. Le troisième n'aura pas ce pouvoir de caractérisation globale et il ne pourra pas intervenir aussi aisément que les autres dans la validation des solutions.

Une des caractéristiques de ce logiciel est que, non seulement l'utilisateur disposera, pour chaque solution des valeurs précises concernant stabilité et prédiction, mais qu'en outre il pourra agir sur leur valeur, en introduisant ou en supprimant des informations à l'aide du conditionnement intrinsèque préalablement décrit.

Dans ce chapitre, nous décrirons tout d'abord le conditionnement intrinsèque que nous réalisons, puis nous définirons formellement les critères utilisés pour valider les solutions obtenues.

2.2. CONDITIONNEMENT DES DONNEES

2.2.1. Généralités

Les n données \hat{s}_i apparaissent comme des produits scalaires de la fonction inconnue f par les noyaux a_i . Dans ces conditions, on conçoit que vont apparaître sur les \hat{s}_i des dépendances et "facteurs d'échelle" qui seront inhérents à la nature même des noyaux a_i .

Plus explicitement, si un noyau a_2 est 10 fois plus grand (au sens de la norme relative au produit scalaire mentionné) qu'un noyau a_1 la donnée \hat{s}_2 "aura tendance" à être plus grande que la donnée \hat{s}_1 (on ne peut guère donner de terme plus précis que celui de tendance, du moins en l'absence d'hypothèse sur f ou sur la densité de probabilité à priori des différentes fonctions f susceptibles de se prêter à une mesure.). De même si un noyau a_n est une combinaison linéaire formelle des $n-1$ autres, la donnée \hat{s}_n correspondante n'apportera aucune information aux $n-1$ autres données. On conçoit clairement, dans le cas de telles liaisons, qu'une "réduction" d'information soit possible ; on verra même lors de l'étude de stabilité des solutions qu'elle est nécessaire.

Enfin, en dehors de liaisons formelles, on conçoit que 2 données \hat{s}_i et \hat{s}_j auront "tendance" à être d'autant plus corrélées que les noyaux correspondants a_i et a_j formeront un "angle" d'autant plus petit.

Dans ces conditions, on conçoit que le conditionnement des données peut être réalisé à partir de l'orthogonalisation des noyaux a_i .

N'importe quelle base orthogonale de l'espace vectoriel des noyaux engendré par les a_i permet d'éliminer les dépendances linéaires existant entre les \hat{s}_i , s'il y en a, et de réduire le nombre n d'informations à une valeur p qui est la dimension de l'espace des a_i ($p < n$).

Par contre, l'analyse en vecteurs propres, tout en permettant cette orthogonalisation, a en plus le mérite de traduire l'importance des corrélations existant entre les noyaux a_i . Ceci se révèle très précieux, dans la mesure où les données \hat{S}_i sont toujours bruitées. Dans ces conditions, et dans le cas où l'on ne fait aucune hypothèse spécifique sur les erreurs (ou, ce qui revient au même, si l'on suppose les erreurs sur les \hat{S}_i indépendantes et égales), la base de vecteurs propres peut être "hiérarchisée" à l'aide des valeurs propres correspondantes. Les données, dans cette nouvelle base, sont, en effet, d'autant plus bruitées que les valeurs propres correspondantes sont faibles.

2.2.2. Méthode proposée

Notons Λ l'espace vectoriel engendré par les n noyaux a_i . Nous sommes amenés à chercher une base de Λ qui soit orthogonale vis à vis du produit scalaire usuel, noté $\langle ., . \rangle$ défini par

$$(2.1.) \quad \langle \alpha, \mu \rangle = \int_I \alpha(x) \mu(x) dx$$

L'analyse en vecteurs propres va nous permettre de déterminer une (parmi une infinité possible) base orthonormale. Elle consiste à rechercher les vecteurs et valeurs propres de la matrice A d'éléments a_{ij} définis par

$$(2.2.) \quad a_{ij} = \langle a_i, a_j \rangle = \int_I a_i(x) a_j(x) dx$$

Cette matrice, dont on peut montrer qu'elle est symétrique, définie semi-positive aura des valeurs propres réelles non négatives. Le nombre p de valeurs propres non nulles donne le rang de A ; c'est aussi la dimension de l'espace Λ .

Soient : Λ_p la matrice diagonale des p valeurs propres non nulles λ_i rangées par ordre décroissant.

U_p la matrice (n, p) des p vecteurs propres correspondants, rangés dans le même ordre.

On peut montrer que la matrice

$$(2.3.) \quad T = \Lambda^{-1/2} U_p^*$$

Où U_p^* désigne la matrice transposée de U_p .

transforme les n noyaux a_i en p noyaux orthogonaux Π_i . Plus précisément les p fonctions Π_i définies par

$$(2.4.) \quad \Pi_i(x) = \sum_{j=1}^n t_{ij} a_j(x)$$

(où t_{ij} est l'élément courant de T) sont orthonormales vis à vis du produit scalaire usuel. Autrement dit

$$\int_I \Pi_i^2(x) dx = 1$$

$$(2.5.) \quad \int_I \Pi_i(x) \Pi_j(x) dx = 0 \quad \text{si } i \neq j$$

Le problème initial portant sur n données est ainsi transformé en un problème portant sur p données, à savoir rechercher la fonction f telle que

$$(2.6.) \quad \int_I \Pi_i(x) f(x) dx = \hat{s}'_i \quad i = 1, p$$

où les nouvelles données \hat{s}'_i sont définies par

$$(2.7.) \quad \hat{s}'_i = \sum_{j=1}^n t_{ij} \hat{s}_j$$

Si $\delta\hat{S}_i$ est l'erreur sur \hat{S}_i et $\delta\hat{S}'_i$ l'erreur sur \hat{S}'_i on peut écrire

$$(2.8.) \quad E[\delta\hat{S}'_i \delta\hat{S}'_j] = \sum_k t_{ik} t_{jk} E[\delta\hat{S}_k^2]$$

(E désignant l'espérance mathématique).

Si l'on suppose les erreurs sur les données \hat{S}_i indépendantes et égales, c'est à dire si

$$E[\delta\hat{S}_i^2] = \delta^2 \quad i = 1, n$$

$$E[\delta\hat{S}_i \delta\hat{S}_j] = 0 \quad i \neq j$$

alors

$$E[\delta\hat{S}_i' \delta\hat{S}_j'] = \sum_{k=1}^n \frac{u_{ki} u_{kj}}{\lambda_k} \delta^2$$

et

$$(2.9.) \quad E[\delta\hat{S}_i'^2] = \sum_{k=1}^n \frac{u_{ki}^2}{\lambda_k} \delta^2$$

On voit que l'erreur sur \hat{S}_i' est inversement proportionnelle à la valeur propre correspondante.

En définitive l'analyse en valeurs propres nous permet de définir un changement de base intrinsèque (défini une fois pour toutes par la matrice T) permettant de passer de n données \hat{S}_i à p données \hat{S}_i' . Ces données, rangées dans l'ordre des valeurs propres décroissantes, se trouvent ordonnées suivant leur "susceptibilité" aux erreurs de mesure $\delta\hat{S}_i'$.

Ce conditionnement nous apparaît donc comme réaliste si l'on désire restreindre l'information à traiter en l'absence de connaissance précise des erreurs sur les données \hat{S}_i .

2.2.3. Discussion

En pratique, les erreurs sur les \hat{S}_i ne seront généralement ni égales, ni même indépendantes. Dans ces conditions, on pourrait être tenté de remettre en cause le caractère intrinsèque du conditionnement décrit ci-dessus. Toutefois, il faut être conscient que l'on serait alors amené à réaliser autant de conditionnements que de jeux de données à traiter et qu'en plus, il faudrait être capable d'estimer les erreurs avec une bonne précision. En outre, il faut mentionner que dans les problèmes que nous avons été amenés à traiter, nous avons pu vérifier que cette prise en compte précise des erreurs n'aurait pas amélioré de façon sensible le conditionnement.

En définitive, le conditionnement intrinsèque permet premièrement d'extraire l'information utile : \hat{S}_i' $i = 1, p$ à partir des n données \hat{S}_i en éliminant les dépendances formelles existant entre elles. Deuxièmement, comme on le verra dans le paragraphe suivant, la hiérarchisation des nouvelles données \hat{S}_i' se révélera fort précieuse dans la recherche d'une solution satisfaisant à certains critères de stabilité et de prédiction.

2.3. VALIDATION DES SOLUTIONS

Comme nous l'avons déjà mentionné, nous cherchons à obtenir une solution ayant une signification physique. Nous nous proposons ici de décrire les 3 critères que nous utilisons pour juger de la qualité de la solution obtenue : instabilité, prédiction, résolution.

2.3.1. Stabilité

Intuitivement, on dira qu'une solution est d'autant plus stable qu'une petite perturbation des données entraîne une déformation d'autant plus faible de la solution.

Il est clair qu'une solution instable présente peu d'intérêt physique ; en plus sa recherche sur le plan numérique est toujours hasardeuse.

Pour préciser cette idée intuitive, nous allons définir un indicateur d'instabilité d'une solution.

Nous supposerons que la méthode employée conduit à une solution unique.

Soit $f_{\hat{S}_i}$ la solution relative aux données \hat{S}_i

$f_{\hat{S}_i + \delta\hat{S}_i}$ la solution relative à des données perturbées $\hat{S}_i + \delta\hat{S}_i$

La fluctuation de la solution entraînée par les fluctuations $\delta\hat{S}_i$ des données est

$$\delta f = f_{\hat{S}_i + \delta\hat{S}_i} - f_{\hat{S}_i}$$

Pour obtenir un critère simple d'instabilité, on choisit d'exprimer de manière globale la fluctuation de f , à savoir la norme (usuelle) de l'écart-type de cette fluctuation. On normalise cette quantité par la norme de la solution f . D'où le coefficient d'instabilité Q défini par

$$(2.10.) \quad Q = \frac{\int_I E[\delta f^2(x)] dx}{\int_I f^2(x) dx}$$

Une solution f sera d'autant plus stable que, pour des perturbations $\delta\hat{S}_i$ données, le coefficient Q sera petit.

On verra, par la suite, que le coefficient d'instabilité Q est une fonction croissante du nombre p de nombre de noyaux orthogonaux Π_i utilisés dans l'analyse.

Pour obtenir une solution suffisamment stable, on sera donc amené à ne prendre en compte qu'un nombre limité p' ($p' < p$) de noyaux Π_i et donc également que ce même nombre de données \hat{S}_i' . On ne perdra pas de vue que chacune des données \hat{S}_i' est une combinaison linéaire de n données \hat{S}_i . Autrement dit, le fait de ne prendre que p' valeurs \hat{S}_i' ne doit pas être interprété comme la non utilisation de certaines mesures \hat{S}_i .

Dans ces conditions, il semblerait qu'on doive choisir p' le plus petit possible. Mais alors, deux problèmes se posent : tout d'abord, avec quelle précision, cette solution déterminée à partir de peu de données \hat{S}_i' satisfait-elle à toutes les données d'origine \hat{S}_i ; ensuite, dans quelle mesure cette solution la plus stable est-elle représentative du phénomène physique étudié. On va donner des éléments permettant de répondre à ces questions, en définissant deux autres critères : pouvoir de prédiction et pouvoir de résolution.

2.3.2. Pouvoir de prédiction

Ayant obtenu une solution f à partir d'un nombre p' de données \hat{S}_i' inférieur au nombre n de données effectives \hat{S}_i , on aimerait savoir dans quelle mesure la solution ainsi obtenue satisfait aux erreurs de mesure près, l'ensemble des n données d'origine \hat{S}_i .

Pour cela on définit le coefficient suivant

$$(2.11.) \quad P_r = \sum_{i=1}^n \left(\frac{\hat{S}_i^{\wedge} - \hat{S}_i}{\sigma_i} \right)^2$$

où \hat{S}_i^{\wedge} est la donnée "prédite" relative à la solution $f_{\hat{S}_i}$, c'est à dire $\hat{S}_i^{\wedge} = \int_I a_i(x) f_{\hat{S}_i}(x) dx$ et σ_i la variance de la donnée \hat{S}_i (erreur de mesure).

On cherchera bien sûr à obtenir une solution $f_{\hat{S}_i}$ telle que

$$|\tilde{S}_i - \hat{S}_i| < \sigma_i \quad i = 1, n$$

autrement dit

$$P_r < n$$

Comme on peut s'y attendre, on verra que le pouvoir de prédiction P_r , croît avec le nombre p' de noyaux orthogonaux Π_i pris en compte. On voit donc que lorsque l'on voudra stabiliser la solution en diminuant la valeur de p' , on diminuera le pouvoir de prédiction. Il sera donc nécessaire de trouver un compromis. Lorsque celui-ci sera réalisé, l'utilisateur pourra, le cas échéant, s'intéresser au pouvoir de résolution de sa solution.

2.3.3. Pouvoir de résolution

Ainsi qu'on l'a précisé dans le paragraphe 2.1. on cherche ici à exprimer la finesse avec laquelle on est capable de déterminer une solution.

Le pouvoir de résolution présente deux différences fondamentales par rapport aux 2 critères précédents.

+ Tout d'abord, il n'a pas un caractère global. Théoriquement, on devrait définir le pouvoir de résolution d'une solution en chaque point de I. On verra que cela est possible dans le cas de la méthode des noyaux de Dirichlet. Dans les autres cas, l'utilisateur n'aura d'autres ressources que de générer des données \hat{S}_i à partir de fonctions connues (par exemple des approximations de distributions de Dirac en certains points "critiques" de I) ; et d'examiner les solutions fournies par la méthode à partir de ces données synthétiques.

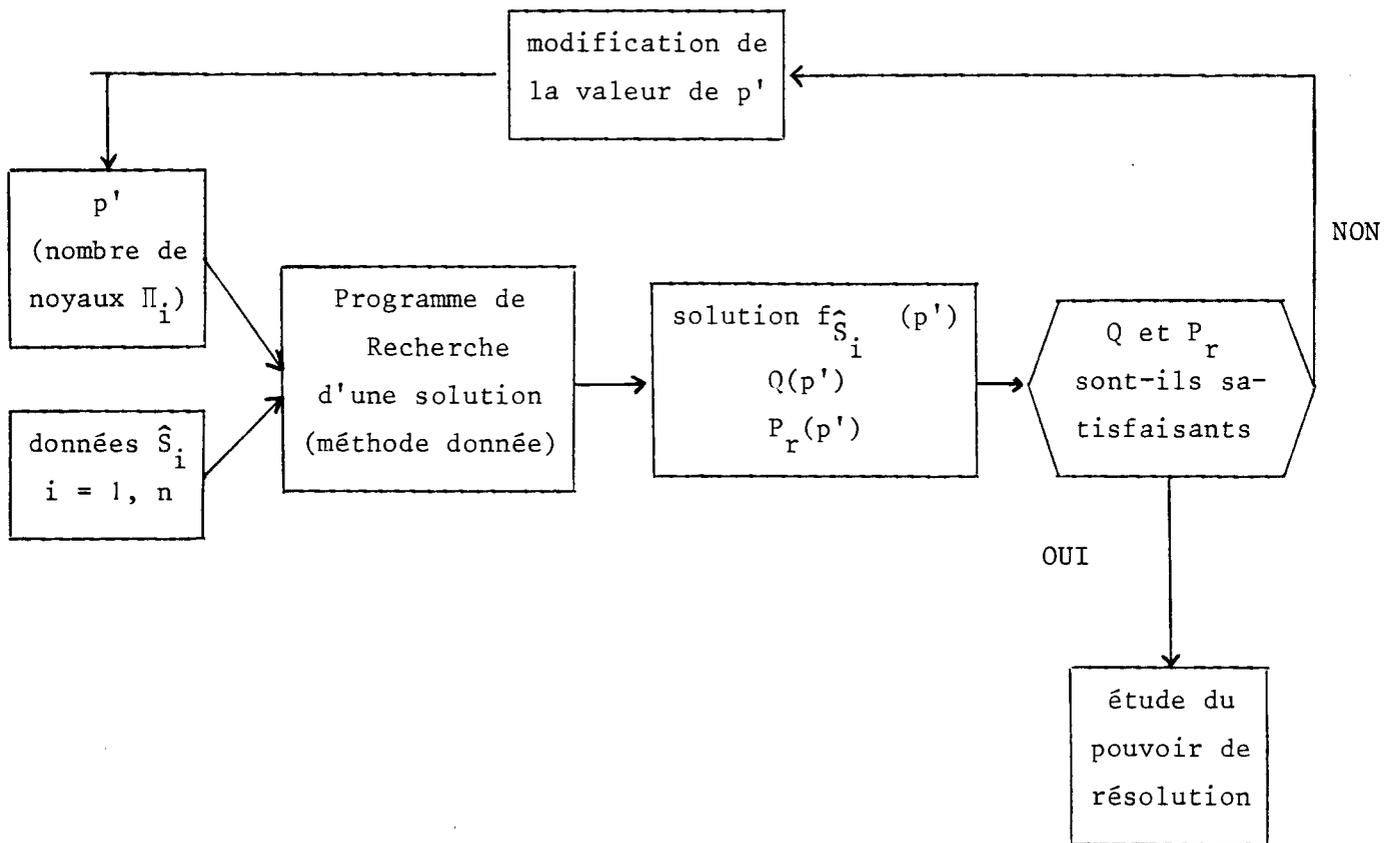
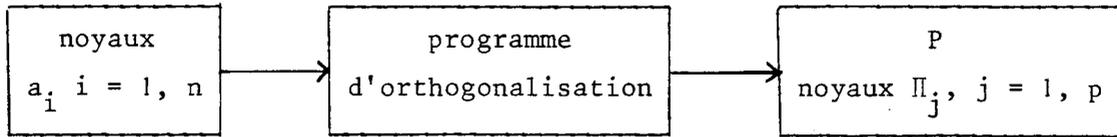
+ D'autre part, le pouvoir de résolution ne dépend pas directement des données \hat{S}_i . Il ne dépend en fait que de la méthode employée et du nombre p' de données \hat{S}_i choisies. Dans ces conditions, l'utilisateur pourrait envisager d'examiner ce critère en premier lieu. En pratique cependant (exception faite de la méthode des noyaux de Dirichlet) l'utilisateur examinera le pouvoir de résolution en quelques points de I qui lui paraissent critiques. Ces points, quant à eux ne seront pas toujours prévisibles à priori.

2.4. DISCUSSION (CHOIX INTERACTIF D'UN COMPROMIS)

On peut maintenant dégager les grandes lignes de la recherche d'une solution par une méthode donnée.

L'utilisateur cherchera, à l'aide du logiciel proposé, une solution relative à une certaine valeur de p' . Au vu des valeurs des coefficients d'instabilité Q et de prédiction P_r , il modifiera en conséquence la valeur de p' jusqu'à obtenir une solution satisfaisante. Il en étudiera alors le pouvoir de résolution. Notons que, exception faite pour la méthode des noyaux de Dirichlet où une étude systématique peut être faite, ce dernier point est à la charge de l'utilisateur, en ce qui concerne la génération de données synthétiques \hat{S}_i à partir de fonctions connues.

On peut schématiser cela ainsi



3. METHODES D'INVERSION

3.1. GENERALITES

3.1.1. Introduction

Dans ce chapitre, nous décrivons, sous leur aspect théorique, les 3 méthodes d'inversion utilisées dans le logiciel proposé.

Dans ce premier paragraphe, nous rappellerons comment s'exprime le problème général exposé en (1.1.), compte tenu du conditionnement intrinsèque des noyaux a_i . Nous tenterons alors de situer les 3 "méthodes" proposées par rapport aux diverses attitudes qu'il est possible d'avoir devant un tel problème.

Les paragraphes 3.2., 3.3. et 3.4. seront alors consacrés à l'exposé théorique des méthodes. Pour chacune d'entre elles, nous adopterons la même présentation. Dans un premier temps, nous définirons la méthode, en la situant éventuellement dans un contexte général. On précisera, s'il y a lieu, quelles sont les techniques numériques nécessaires à la mise en oeuvre de la méthode. Nous montrerons ensuite, comment calculer les différents paramètres de qualité de la solution. Enfin une discussion sera faite.

3.1.2. Position du problème

Le conditionnement intrinsèque décrit dans le paragraphe 2.2. nous a conduit à une base orthonormale de p fonctions Π_i ($p < n$). Nous avons vu au paragraphe 2.4. que nous sommes amenés, pour arriver à un compromis satisfaisant, à choisir les p' premiers de ces noyaux ($p' < p$), autrement dit à n'utiliser que p' données \hat{S}_i' . La valeur de p' sera généralement déterminée de manière interactive, c'est à dire en résolvant plusieurs problèmes, de mêmes données \hat{S}_i , correspondant à diverses valeurs de p' .

Dans tout ce qui suit, nous raisonnerons pour p' fixé. Le problème (1.1.) s'exprime alors comme suit :

Soient p' fonctions réelles Π_i d'une variable réelle, nommées noyaux, définies sur un segment I de R , linéairement indépendants et orthonommés (c'est à dire tels que $\int_I \Pi_i(x) \Pi_j(x) dx = 0$ si $i \neq j$

tels que

$$\int_I \Pi_i(x) \Pi_k(x) dx = 0 \quad \text{si } i \neq j$$

$$\text{et} \quad \int_I \Pi_i^2(x) dx = 1$$

On suppose connues p' fonctionnelles linéaires \hat{S}_i' d'une fonction inconnue f , non négative sur I , définies par

$$(3.0.) \quad \hat{S}_i' = \int_I \Pi_i(x) f(x) dx \quad i = 1, p'$$

Généralement, malgré la contrainte de non-négativité imposée à f , un tel problème admet une infinité de solutions.

Deux attitudes sont alors possibles

+ ajouter de nouvelles contraintes au problème pour le rendre "bien posé", c'est à dire pour qu'il n'admette plus qu'une solution.

+ se refuser à induire une information sous quelque forme que ce soit et tenter d'exprimer ce que représente l'ensemble des solutions possibles.

Dans le cadre de la première attitude, on peut faire choix pour f de un ou plusieurs modèles paramétriques. C'est ce qui se passera dans la méthode de discrétisation. On notera que le logiciel décrit ici ne permet pas la prise en compte d'un modèle paramétrique quelconque.

On peut également se donner des contraintes de lissage, au sens large. C'est ce qui se passe, ici, lorsque l'on cherche l'unique solution d'entropie maximum. On remarquera d'ailleurs dans ce cas qu'en toute rigueur, on a encore à faire à un modèle paramétrique particulier.

Dans le cadre de la deuxième attitude, on peut chercher une solution représentant, d'une certaine manière, la moyenne de toutes les solutions possibles. C'est ce que se propose de faire la méthode des noyaux de Dirichlet. En toute rigueur, il faut noter qu'une information nouvelle est alors introduite qui est la "manière" de faire cette moyenne.

Par contre la méthode de recherches d'extrémales (Sabatier, 1977) non réalisée dans ce logiciel, traduirait complètement la 2ème attitude ; elle permettrait de décrire la multiplicité des solutions possibles.

3.2. METHODE DE BACKUS ET GILBERT

3.2.1. Définition de la méthode

Après avoir montré que notre problème pouvait s'insérer dans les hypothèses requises par la méthode générale de Backus et Gilbert, nous ferons un bref résumé de cette dernière. La méthode des noyaux de Dirichlet (utilisée effectivement dans le logiciel décrit) sera présentée comme un cas particulier.

3.2.1.1. Contexte d'application de la méthode de Backus et Gilbert

Les hypothèses sont les suivantes :

- + Soit M un espace vectoriel d'éléments nommés modèles,
- + Soit p' fonctionnelles linéaires g_i de M dans R , différentiables au sens de Fréchet (supposées linéairement indépendantes).

On cherche à restituer des informations moyennes sur l'ensemble des modèles de M susceptibles d'engendrer p' valeurs données \hat{s}_i des p' fonctionnelles g_i .

Dans notre problème, on peut écrire

$$(3.1.) \quad g_i(f) = \int_I \Pi_i(x) f(x) dx$$

On se ramène bien au contexte mentionné sous réserve de ne pas tenir compte de la contrainte de positivité sur f (afin que M soit bien linéaire).

3.2.1.2. Noyaux filtrants

La méthode de Backus et Gilbert consiste à chercher une fonction qui d'une part satisfait les p' données \hat{S}_i' et qui d'autre part est en chaque point x de I , une "certaine moyenne" des valeurs de toutes les solutions dans un voisinage aussi petit que possible de x_0 . Pour cela, on cherche en chaque point de x_0 de I une application, que l'on notera $\langle \cdot \rangle_{x_0}$ de l'ensemble des modèles M dans R qui traduira "assez bien" la valeur moyenne dans un "voisinage" de r_0 . Plus précisément, si on impose à cette application d'être linéaire on montre que cela revient à chercher en chaque point x_0 une fonction A de 2 variables (nommée noyau filtrant) telle que

$$(3.2.) \quad \langle f \rangle_{x_0} = \int A(x_0, x) f(x) dx$$

On va en outre, imposer à cette "valeur locale moyenne" de ne dépendre que des p données \hat{S}_i' . Ce qui signifie qu'une telle valeur moyenne sera la même pour tout modèle susceptible d'engendrer les données \hat{S}_i' .

On montre qu'alors la fonction A appelée noyau filtrant est nécessairement de la forme

$$(3.3.) \quad A(x_0, x) = \sum_{i=1}^{p'} \alpha_i(x_0) \Pi_i(x)$$

et qu'alors

$$(3.4.) \quad \langle f \rangle_{x_0} = \sum_{i=1}^{p'} \alpha_i(x_0) \hat{S}_i'$$

Autrement dit, en chaque point x_0 , le noyau filtrant est une combinaison linéaire des noyaux Π_i , les coefficients α_i dépendant du point x_0 .

3.2.1.3. Sélection des noyaux filtrants

Les noyaux filtrants sont donc à rechercher parmi les expressions de la forme (3.3.). Pour choisir entre l'infinité de noyaux possibles, Backus et Gilbert proposent d'introduire un critère dit de "diracité" permettant de définir, dans quelle mesure $\int_I A(x_0, x) f(x) dx$ représente la moyenne de f dans un faible voisinage de x_0 . On conçoit que plus les noyaux filtrants seront proches d'une distribution de Dirac, plus le pouvoir de résolution en sera élevé.

Plusieurs définitions sont bien sûr possibles concernant la "diracité". Un critère, proposé par Backus et Gilbert consiste à trouver, en chaque point x_0 la fonction d'une variable $A(x_0, x)$ qui minimise la quantité

$$(3.5.) \quad \Delta = \int_I (x - x_0)^2 A^2(x_0, x) dx$$

Ceci conduira généralement à des noyaux, dont on ne pourra définir une expression analytique qu'en chaque point x_0 .

3.2.1.4. Noyaux de Dirichlet

La méthode proposée dans ce logiciel consiste à choisir comme noyau filtrant $A(x, x_0)$ la projection orthogonale d'une distribution de Dirac δ_{x_0} en x_0 sur l'espace vectoriel engendré par les noyaux Π_i (supposé muni du produit scalaire usuel). On va montrer que les noyaux filtrants sont encore du type 3.3. et qu'on a donc bien à faire à un cas particulier de la méthode de Backus et Gilbert.

Comme la base des Π_i est orthonormale, les composantes de δ_{x_0} sur les Π_i ne sont rien d'autre que les produits scalaires de δ_{x_0} par Π_i ; autrement dit

$$(3.6.) \quad A(x, x_0) = \sum_{i=1}^{p'} \lambda_i \Pi_i(x)$$

avec

$$(3.7.) \quad \lambda_i = \int_I \Pi_i(x) \delta_{x_0}(x) dx = \Pi_i(x_0)$$

D'où

$$(3.8.) \quad A(x, x_0) = \sum_{i=1}^{p'} \Pi_i(x_0) \Pi_i(x)$$

Cette expression est bien de la forme (3.3.) ce qui prouve qu'on a bien à faire à un cas particulier de la méthode de Backus et Gilbert.

L'application $\langle f \rangle_{x_0}$ définie en (3.2.) de l'espace M des modèles dans \mathcal{R} , qui permet de définir la moyenne d'un modèle f en x_0 , n'est autre ici que la projection orthogonale de f sur l'espace vectoriel engendré par les Π_i .

En résumé la méthode des noyaux de Dirichlet consiste à rechercher dans l'espace vectoriel engendré par les Π_i l'unique solution f_0 satisfaisant aux données \hat{s}_i qui s'écrit ; elle a pour expression

$$(3.9.) \quad f_0(x) = \sum_{i=1}^p \hat{s}_i \Pi_i(x)$$

On peut aussi considérer $f_0(x_0)$ comme représentant en chaque x_0 , la moyenne de n'importe quelle fonction f susceptible d'engendrer les données \hat{s}_i ; cette moyenne de f en x_0 étant obtenue par convolution de f par un noyau filtrant $A(x_0, x)$ qui est ici la projection orthogonale d'une distribution de Dirac δ_{x_0} sur la base des Π_i .

3.2.2. Techniques numériques spécifiques

Néant

3.2.3. Validation des solutions

3.2.3.1. Pouvoir de résolution

De par la conception même de la méthode, le pouvoir résolvant en chaque point x_0 de I peut être "visualisé" par les noyaux filtrants

$$A(x_0, x) = \sum_{i=1}^p \Pi_i(x) \Pi_i(x_0)$$

Ces noyaux représentent d'ailleurs la solution que fournirait la méthode si on lui fournissait en entrée les données correspondant au cas où f est une distribution de Dirac en x_0 .

Rappelons que ce pouvoir de résolution est indépendant des données \hat{s}_i .

3.2.3.2. Pouvoir de prédiction

Si on l'exprime sur la base des p noyaux Π_i , il a pour expression

$$P_r = \sum_{k=1}^p \frac{(\hat{s}'_k - \tilde{s}'_k)^2}{\delta S'_K{}^2}$$

où : \tilde{s}'_K désigne la valeur de la donnée engendrée par la solution f_0 de Backus et Gilbert

$$\tilde{s}'_K = \int_I \Pi_K(x) f_0(x) dx$$

$\delta S'_K$ la variance de la donnée s'_k

Il est clair que : $\tilde{s}'_K = \hat{s}'_K$ pour $k = 1, p'$
 $\tilde{s}'_K = 0$ pour $k = p' + 1, p$
 (car par construction f_0 , combinaison linéaire des p' premiers Π_K est orthogonale aux autres noyaux Π_K).

Par suite, le pouvoir de prédiction se réduit à

$$P_r = \sum_{k=p'+1}^p \frac{s'_k{}^2}{\delta S'_K{}^2} \quad \text{si } p' < p$$

(3.10.)

$$P_r = 0 \quad \text{si } p' = p$$

3.2.3.3. Stabilité

Voyons tout d'abord quelle est l'incidence sur la solution f_0 d'erreurs sur les données \hat{s}'_i . Soient $\delta \hat{S}'_i$ les espérances mathématiques de ces erreurs, et σ'_i leur variance. On se propose de calculer l'espérance mathématique Δ de la norme de l'erreur induite sur f_0

$$\Delta = E \left[\int_I \delta f_0^2(x) dx \right] = \int_I E(\delta f_0^2(x)) dx$$

Or

$$\hat{S}'_i = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^n u_{ij} \hat{S}_i$$

(si λ_i sont les valeurs propres de A et u_{ij} l'élément courant de la matrice U_p , des vecteurs propres correspondants).

D'où

$$\delta \hat{S}'_i = \frac{1}{\sqrt{\lambda_i}} \sum_{j=1}^n u_{ij} \delta \hat{S}_i$$

l'hypothèse d'indépendance des erreurs sur les données S_i se traduit par

$$E[\delta \hat{S}_i \delta S_j] = 0 \quad \text{si } i \neq j$$

Par suite

$$(3.11.) \quad \Delta = \sum_{i=1}^p \sum_{j=1}^n \frac{u_{ij}^2}{\lambda_i} \sigma_i^2$$

où σ_i représente l'écart type de l'erreur sur \hat{S}_i .

Pour interpréter cette relation, supposons toutes les erreurs sur les \hat{S}_i égales à σ^2 ; alors, compte tenu de ce que la matrice U_p des vecteurs propres est unitaire ($U_p^* U_p = I$)

$$\Delta = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}$$

Il apparaît clairement que l'instabilité de la solution f_0 est gouvernée par les valeurs propres les plus faibles. Ceci justifie le conditionnement intrinsèque réalisé puisque les noyaux Π_i sont classés en fonction de l'instabilité qu'ils introduisent dans la solution f_0 .

Le coefficient d'instabilité Q défini en 2.3.1. a pour expression

$$Q = \frac{E(\int \delta f_o^2(x) dx)}{\int \delta f_o^2(x) dx} = \frac{p' \sum_{i=1}^{p'} \delta S_i^2}{\sum_{i=1}^{p'} \delta S_i'^2}$$

Si l'on souhaite optimiser la sélection des données à prendre en compte, on serait amené pour chaque valeur de p' à rechercher la combinaison optimale de p' données \hat{S}'_i assurant la meilleure stabilité Q. Dans le souci de rapidité de calcul et afin de rendre plus aisée la comparaison de diverses solutions relatives à des données différentes, nous avons préféré que les données sélectionnées le soient toujours dans le même ordre. La valeur de p' reste ajustable par l'utilisateur, au vu du pouvoir de résolution et de la valeur de Q.

En résumé donc, l'utilisateur a à faire choix, de manière interactive, de la valeur de p' lui assurant le meilleur compromis entre stabilité, résolution et prédiction.

3.2.4. Discussion

Notons tout d'abord que la méthode des noyaux de Dirichlet présente par rapport à la méthode générale de Backus et Gilbert l'avantage d'être beaucoup plus rapide au niveau du calcul numérique. Notamment on peut obtenir une expression analytique des noyaux filtrants $A(x_o, x)$ ainsi que de la solution.

Par contre, aucune de ces méthodes ne permet d'assurer la positivité de f , ni celle des noyaux filtrants $A(x_o, x)$. Ce dernier point est fondamental et montre l'intérêt qu'aura l'utilisateur à visualiser les noyaux filtrants $A(x_o, x)$ en suffisamment de points x_o de I. Notamment, il ne faudra peut être pas se contenter de considérer la demi-largeur d'un noyau filtrant $A(x, x_o)$ comme représentant le pouvoir de résolution en x_o , si les parties négatives de ce noyau filtrant sont importantes. La notion de valeur moyenne serait en effet alors sérieusement compromise.

Notons que l'expérience montre que les noyaux filtrants de Dirichlet sont généralement plus "piqués" que ceux obtenus par Backus et Gilbert avec un critère quadratique de "diracité". En revanche leurs parties négatives sont souvent plus importantes.

Enfin il faut noter que le logiciel proposé permet une bonne maîtrise de la méthode. L'utilisateur n'a en effet qu'un paramètre p' à ajuster d'une manière très limitée puisqu'il ne peut prendre que p valeurs distinctes. Pour ce faire, il a à sa disposition

- les noyaux filtrants $A(x_0, x)$ aux points x_0 qui l'intéressent pour une valeur de p' donnée et ceci de manière indépendante des données
- pour chaque solution f_0 , le pouvoir de prédiction P_r et le coefficient de stabilité Q

3.3. METHODE DE DISCRETISATION

Cette méthode consiste à déterminer une solution f_1 de manière discrète.

Nous allons présenter cette méthode, en l'inscrivant dans un cadre plus général de développement de la solution sur une base de fonctions quelconques.

3.3.1. Définition de la méthode

3.3.1.1. Cadre général

Soient ψ_i , $i = 1, m$, m fonctions réelles non négatives, d'une variable réelle.

On se propose de chercher une solution non négative f_1 de la forme

$$(3.12.) \quad f_1(x) = \sum_{i=1}^m M_i \psi_i(x) \quad M_i > 0$$

Si l'on note M le vecteur colonne des coefficients M_i , il est clair que les M_i sont solution du système

$$(3.13.) \quad B \cdot M = S'$$

où S' est le vecteur colonne des $p' \hat{s}_j'$

B est la matrice (p, m) d'élément courant

$$(3.14.) \quad b_{ij} = \int_I \Pi_i(x) \psi_j(x) dx$$

Généralement, tant que $m < p$, ce système n'a aucune solution ; pour $m = p$, la contrainte de positivité fait qu'on n'a pas toujours une solution. En augmentant suffisamment la valeur de m on peut aboutir à un système admettant des solutions. Mais généralement on passe d'un système impossible à un système indéterminé (infinité de solutions). Il est alors nécessaire d'imposer une contrainte supplémentaire. On choisit ici de minimiser la "norme" de la solution f_1 . Il existe quantités de normes ; nous supposons que celle-ci est une forme quadratique définie, positive des coefficients M_i . Soit H la matrice correspondante.

Notre problème se ramène alors à

$$(3.15.) \quad \begin{array}{ll} \text{minimiser} & M^* H M \\ \text{sous les contraintes} & B M = S' \\ & M \in (\mathbb{R}^+)^m \end{array}$$

On démontre (Daval, 1977) que ce problème est équivalent à chercher la limite, quand $\epsilon_a \rightarrow 0$ de la solution du problème suivant

$$\begin{array}{ll} \text{minimiser} & (B M - S')^* (B M - S') + \epsilon_a M^* H M \\ \text{sous la contrainte} & M \in (\mathbb{R}^+)^m \end{array}$$

ce qui par dérivation, conduit à

$$(3.16.) \quad \begin{array}{ll} (B^* B + \epsilon_a H) M = B^* S \\ \text{avec} & M \in (\mathbb{R}^+)^n \end{array}$$

On voit donc que la solution $f_1(x) = \sum_{i=1}^m M_i \psi_i(x)$ apparaît en fait comme la limite des solutions de systèmes linéaires, sous contrainte de positivité.

3.3.1.2. Méthode effectivement employée

On s'est limité, dans ce logiciel, au cas où les fonctions ψ_i étaient des "portes", contigües, de même largeur Δh , réparties dans l'intervalle I, de hauteur unité. Si l'on note x_1, x_u les bornes de l'intervalle I, on a alors

$$b_{ij} = \int_{x_j - \frac{\Delta h}{2}}^{x_j + \frac{\Delta h}{2}} \Pi_i(x) dx$$

(3.17.)

avec $x_j = x_1 + (j - \frac{1}{2}) \frac{x_u - x_1}{m}$

qu'on approxime, en remplaçant $\Pi_i(x)$ par $\Pi_i(x_j)$ sur tout l'intervalle $[x_j - \frac{\Delta h}{2}, x_j + \frac{\Delta h}{2}]$; soit

$$b_{ij} \simeq \Delta h \Pi_i(x_j)$$

Les coefficients μ_i de la formule 3.12. sont alors assimilables à la valeur de la fonction inconnue f_1 au point x_i . On adoptera une nouvelle notation, en remplaçant μ_i par F_i et M par F.

On choisit comme "norme" pour f_1 , une quantité $\|f_1\|$ définie par

$$\|f_1\|^2 = \int_I f_1^2(x) dx = \sum_{i=1}^m F_i^2 \int_I \psi_i^2(x) dx$$

(3.18.)

qu'on approxime, comme précédemment par

$$\|f_1\|^2 = \Delta h^2 \sum_{i=1}^m F_i^2$$

(3.19.)

La matrice H définie en 3.3.1.1. se réduit donc ici à la matrice identité. Le système (3.16.) s'écrit donc

$$\Delta h (B B^* + \varepsilon_a I) F = B^* S'$$

(3.20.)

avec $F \in (R^+)^m$

En résumé, on obtient théoriquement une solution discrète en un nombre donné m de points, en cherchant la limite de la solution de systèmes linéaires soumis à des contraintes de positivité.

En fait, le logiciel proposé cherche la solution pour une valeur ϵ_a donnée. Ce n'est là qu'une solution approchée que l'utilisateur pourra valider grâce à la valeur du coefficient de prédiction.

3.3.2. Techniques numériques

On montre que l'algorithme classique de Gauss-Seidel modifié pour permettre la prise en compte de la contrainte de positivité converge vers la solution du système (3.20.). Pour cela, il suffit, de remplacer, à chaque fin d'itération de la méthode classique de Gauss-Seidel, toutes les composantes négatives de la solution par des valeurs nulles (cf Daval, 1977).

Nous allons décrire cet algorithme en posant

$$(3.21.) \quad A = B^* B + \epsilon_a I$$

$$Y = B^* S$$

$$\begin{aligned} \text{on a} \quad A &= (a_{ij}) \quad i = 1, m; j = 1, m \\ Y &= (y_i) \quad i = 1, m \\ F &= (f_i) \quad i = 1, m \end{aligned}$$

La résolution de $AF = Y$ avec $F \in (\mathbb{R}^+)^m$ se fait par un processus itératif de relaxation. Notons F_i^k la valeur de la i ème composante de F obtenue après la k ème itération.

On choisit comme valeurs initiales

$$(3.22.) \quad F_i^0 = \max \left(0, \frac{y_i}{a_{ii}} \right) \quad i = 1, m$$

L'itération de rang k se compose de m sous-itérations nommées "relaxations" relatives chacune à une composante F_i^k . Chaque relaxation de rang i est réalisée de manière à annuler "l'erreur" sur la ième équation du système $AF = Y$, en prenant comme valeurs des inconnues autres que la ième, celles connues au moment de la relaxation ; c'est à dire qu'on détermine F_i^k à partir de F_j^k pour $j = 1, 2, \dots, i - 1$

$$F_j^{k-1} \text{ pour } j = i + 1, \dots, m$$

Cela revient à poser, pour chaque itération de rang k

$$(3.23.) \quad F_i^k = \frac{1}{a_{ii}} \left(y_i - \sum_{j=1}^{i-1} a_{ij} F_j^k - \sum_{j=i+1}^m a_{ij} F_j^{k-1} \right)$$

La solution (pour ε_a donné) est considérée comme "satisfaisante" quand, à la fin d'une itération, le vecteur résidu : $Y - Y_e^K$ a toutes ses composantes inférieures à une valeur donnée ε_c .

Y_e^K est défini par

$$(3.24.) \quad Y_e^K = A F^K$$

avec $F^K = (F_i^K) \quad K = 1, m$

3.3.3. Validation des solutions

3.3.3.1. Pouvoir de résolution

Le pouvoir de résolution n'est pas fourni explicitement par la méthode. Comme nous l'avons mentionné en 2.3.3., l'utilisateur peut, s'il le souhaite, générer des données \hat{S}_1 synthétiques correspondant à une fonction connue et comparer cette dernière à la solution fournie par la méthode.

3.3.3.2. Pouvoir de prédiction

On est amené à calculer les valeurs des données \hat{S}_1 générées par la solution f_1 formée de m portes contigües sur I (cf 3.3.1.2.).

On a

$$\hat{S}_i = \int_I a_i(x) f_j(x) dx$$

En procédant comme précédemment, assimilant $a_i(x)$ à $a_i(x_j)$ sur l'intervalle $[x_j - \frac{\Delta h}{2}, x_j + \frac{\Delta h}{2}]$, on trouve

$$(3.25.) \quad \tilde{S}_i = \Delta h \sum_{i=1}^m a_i(x_j) F_i$$

F_i désignant la valeur finale des F_i^k , après arrêt des itérations (3.23.). F_i représente aussi la valeur de f_j au point x_i .

D'où l'on peut déduire l'erreur de prédiction

$$(3.26.) \quad P_r = \sum_{i=1}^m \left(\frac{\tilde{S}_i - \hat{S}_i}{\sigma_i} \right)^2$$

3.3.3.3. Stabilité

La stabilité de cette méthode n'a pas été étudiée en soi. Toutefois, on sait que dans un système linéaire du type $AF = Y$, on définit ce que l'on appelle conditionnement par

$$(3.27.) \quad C = \|A\| \|A^{-1}\|$$

où $\| \cdot \|$ désigne une norme matricielle.

La norme euclidienne, pour une matrice symétrique, comme c'est le cas dans notre problème, conduit à

$$(3.28.) \quad C = \frac{|\lambda_M|}{|\lambda_m|}$$

où λ_M et λ_m désignent respectivement la plus grande et le plus petite valeur propre de A .

D'autre part, on démontre que la matrice A d'élément courant $\alpha_{ij} = \sum_{k=1}^m \Pi_i(x_k) \Pi_j(x_k)$ tend vers la matrice A d'élément

$$a_{ij} = \int_I \Pi_i(x) \Pi_j(x) dx$$

qui est celle définie au paragraphe 2.2. lors du conditionnement intrinsèque.

Il semble donc raisonnable d'admettre que le conditionnement intrinsèque est utilisable pour cette méthode. Cependant, contrairement à ce qui se passe pour les 2 autres méthodes, le logiciel ne fournit ici aucune valeur de coefficient de stabilité.

3.3.4. Discussion

Par rapport à la méthode de Backus et Gilbert, la méthode décrite permet d'imposer la positivité de la solution.

Cependant de nombreux arbitraires sont présents ici. On y trouve le choix de la norme de la solution, le choix du nombre points qui doit être assez grand pour qu'il existe au moins une solution et qu'elle présente suffisamment de détails. En même temps m doit rester suffisamment petit pour éviter des problèmes d'instabilité numérique, et pour que le temps de calcul ne soit pas prohibitif. Le choix de ε_a reste assez délicat.

En définitive, on ne perdra pas de vue qu'il s'agit, ici, plus encore que pour la méthode de Backus et Gilbert, d'un moyen d'obtention d'une solution initiale nécessaire à la recherche de la solution d'entropie maximum. Les arbitraires présents font qu'il est conseillé de ne recourir à cette méthode discrète que lorsque la méthode de Backus et Gilbert n'aura pas permis la détermination de solution initiale satisfaisante.

Dans le cas où l'on désirerait faire de cette méthode une fin en soi, il semble nécessaire d'en faire une étude théorique plus détaillée, notamment en ce qui concerne la prise en compte des erreurs sur les données. Le logiciel demanderait alors d'être quelque peu aménagé.

3.4. LA METHODE DU MAXIMUM D'ENTROPIE

3.4.1. Définition de la méthode

3.4.1.1. Introduction

Cette méthode consiste à rechercher, parmi toutes les solutions possibles (satisfaisant les p données \hat{S}_i), celle qui possède l'entropie maximum.

L'application du concept d'entropie à des fonction conti-
nues et quelque peu discutable (Roubine, 1970). Cependant, cette méthode
apparaît aussi comme l'application d'un critère simple de lissage (Lefeuvre,
1977) qui a le mérite de conduire à une solution non-négative.

3.4.1.2. La méthode

On est amené à chercher une fonction f_2 (dont on montre
que si elle existe, elle est unique) vérifiant les p' données \hat{s}_i'

$$(3.31.) \quad \int_I \Pi_i f_2(x) = \hat{s}_i' \quad i = 1, p'$$

et maximisant

$$(3.32.) \quad \int_I f_2(x) \text{Log} f_2(x) dx$$

Cela revient à rendre stationnaire

$$(3.33.) \quad \int_I f_2(x) \text{Log} f_2(x) + \sum_{i=1}^{p'} \lambda_i \int_I \Pi_i f_2(x) dx$$

où les λ_i sont des multiplicateurs de Lagrange. Le calcul des variations
montre que la solution est définie par

$$(3.34.) \quad f_2(x) = \exp \left\{ -1 + \sum_{i=1}^{p'} \lambda_i \Pi_i(x) \right\}$$

les λ_i étant déterminés par les p' equations

$$(3.35.) \quad \hat{s}_j' = \int_I \Pi_j(x) f_2(x) dx$$

3.4.1.3. Recherche de la solution

En pratique, pour estimer les λ_i , on utilise la méthode
des moindres carrés et on cherche à minimiser

$$(3.36.) \quad \sum_{i=1}^{p'} \left[\hat{s}_i' - \int_I \Pi_j(x) \exp \left(-1 + \sum_{j=1}^{p'} \lambda_j \Pi_j(x) \right) dx \right]^2$$

En l'absence de toute erreur sur les données, cette méthode conduit à la
solution unique de la forme (3.34.) satisfaisant exactement les données.

3.4.1.4. Prise en compte des erreurs

Dès que les données sont bruitées, il est raisonnable, dans cette estimation par moindres carrés, de pondérer les quantités à minimiser en fonction des erreurs.

En toute rigueur, les données \hat{S}_i' n'ayant pas leurs erreurs indépendantes, on est amené à minimiser la quantité

$$(3.36.) \quad \sum_{i=1}^{p'} \sum_{j=1}^{p'} \frac{1}{E[\delta\hat{S}_i' \delta\hat{S}_j']} (\hat{S}_i' - \int_I \Pi_i(x) f_2(x) dx) (\hat{S}_j' - \int_I \Pi_j(x) f_2(x) dx)$$

Cependant, il est possible de choisir dans l'espace de dimension p' des noyaux Π_i , une nouvelle base μ_i telle que les nouvelles données \hat{S}_i'' correspondantes voient leurs erreurs indépendantes (autrement dit que la matrice de variance covariance des erreurs sur \hat{S}_i'' soit diagonale). Dans ces conditions si σ_i'' représente la variance de l'erreur sur \hat{S}_i'' , on est amené à minimiser

$$(3.37.) \quad \Phi = \sum_{i=1}^{p'} \frac{1}{\sigma_i''^2} (\hat{S}_i'' - \int_I \mu_i(x) f_2(x) dx)^2$$

Cette deuxième formule (3.37.) ne fait apparaître que p' termes contre $\frac{p'(p'+1)}{2}$ dans la première (3.36.). D'où le choix effectué pour la recherche numérique.

3.4.2. Méthode programmée

A partir des variances σ_i des erreurs (supposées indépendantes) sur les données \hat{S}_i , on calcule la matrice de covariance Σ' des erreurs sur les données \hat{S}_i' . Elle a pour terme général

$$\sigma_{ij}' = E[\delta\hat{S}_i' \delta\hat{S}_j']$$

avec

$$\delta\hat{S}_i' = \sum_{k=1}^n t_{ik} \delta\hat{S}_k$$

d'où, compte tenu de ce que $E[\delta S_i \delta S_j] = 0$ si $i \neq j$

$$(3.38.) \quad \sigma_{ij}' = \sum_{k=1}^n t_{ik} t_{jk} \sigma_k^2$$

On recherche alors vecteurs propres et valeurs propres de Σ' . Si V' désigne la matrice des vecteurs propres et Λ' la matrice diagonale des valeurs propres λ_i' , on peut écrire

$$\Lambda' = U'^* \Sigma' U'$$

Autrement dit U représente la matrice de changement de base dans $\mathbb{R}^{p'}$ rendant diagonale la matrice de covariance des erreurs. La quantité à minimiser (3.37.) s'écrit

$$(3.39.) \quad \Phi = \sum_{i=1}^{p'} \frac{1}{\lambda_i'} \left[\hat{S}_i'' - \int_I \mu_i(x) \exp\{-1 + \sum_{j=1}^{p'} \lambda_j \Pi_j(x)\} dx \right]^2$$

avec

$$\hat{S}_i'' = \sum_{j=1}^{p'} u'_{ij} \hat{S}_j'$$

$$\mu_i(x) = \sum_{j=1}^{p'} u'_{ij} \Pi_j(x)$$

u'_{ij} désignant le terme général de la matrice U' .

La recherche des p' coefficients λ_i maximisant (3.39.) est faite à l'aide du sous-programme VAO7AD de la bibliothèque d'Harwell. Celui-ci est particulièrement adapté au cas où l'on recherche à minimiser une somme de carrés, dans le cas où l'on sait en calculer numériquement les dérivées de Φ par rapport aux variables λ_i à estimer. Il utilise une méthode de Marquardt modifiée par Fletcher, dont une particularité est le modifier la "métrique" de l'espace des coefficients, au fur et à mesure des itérations.

L'arrêt des itérations se fait lorsque chaque "résidu" sur chaque donnée \hat{S}_i' est inférieur à une valeur fournie par l'utilisateur. Comme on le verra dans la documentation d'utilisation, on pourra choisir entre 2 options d'arrêt

$$+ \text{résidus de rang } i \quad [\hat{S}_i'' - \tilde{S}_i'']^2 < 0,5 \sigma_i''$$

+ résidus inférieurs à des valeurs fournies par l'utilisateur

Par ailleurs l'algorithme requiert, pour initialiser la suite des itérations, une valeur initiale λ_i^0 pour chacun des paramètres à estimer. Ces valeurs initiales doivent être choisies avec d'autant plus de soin que le problème d'optimisation rencontré ici n'est généralement pas convexe. Ce qui signifie que la recherche numérique d'un optimum peut aboutir à un optimum local qui ne soit pas l'optimum absolu. On voit alors tout l'intérêt de disposer de solutions approchées telles que celles fournies par la méthode de Backus et Gilbert ou par la méthode de discrétisation. En effet, comme nous allons le montrer, il est possible, à partir de telles solutions, de fournir à l'algorithme d'optimisation, des valeurs initiales λ_i^0 par une technique d'identification de modèles.

3.4.3. Recherche d'une solution initiale

Nous avons vu que la méthode de Backus et Gilbert nous conduisait à une solution f_0 de la forme

$$f_0(x) = \sum_{i=1}^{p'} \hat{S}_i \Pi_i(x)$$

tandis que la méthode de discrétisation nous conduirait à une solution f_1 définie par sa valeur en m points x_i équirepartis sur I

$$f_1(x_i) = y_i$$

Le problème consiste à estimer les paramètres λ_i^0 ($i = 1, p'$) de telle façon que la fonction f_2 définie par

$$(3.40.) \quad f_2^0(x) = \exp\left\{-1 + \sum_{i=1}^{p'} \lambda_i^0 \Pi_i(x)\right\}$$

soit "suffisamment proche" de f_0 ou de f_1 .

Le premier cas où f_0 est connue sous forme analytique se ramène aisément au second (connaissance des valeurs y_i de la fonction en m points x_i). Cependant, le problème ainsi posé, n'est pas linéaire par rapport aux paramètres λ_i^0 à estimer. On peut le transformer en un problème linéaire par passage au logarithme, sous réserve de ne pas avoir de valeurs négatives ou nulles de la fonction à estimer.

En définitive, on se fixe une valeur ε_m positive et l'on définit m valeurs z_i de la façon suivante

$$(3.41.) \quad \begin{aligned} z_i &= \text{Log } \varepsilon_m & \text{si } & y_i \leq \varepsilon_m \\ z_i &= \text{Log } y_i & \text{si } & y_i > \varepsilon_m \end{aligned}$$

On cherche alors à identifier cette nouvelle fonction avec un modèle entropie en recherchant les paramètres λ_i^0 tels que

$$(3.42.) \quad \sum_{j=1}^{p'} \lambda_j^0 \Pi_j(x_i) = y_i + \varepsilon_i$$

avec $|\varepsilon_i|$ aussi petit que possible.

Plus précisément, on choisit ici comme critère d'estimation celui des moindres carrés, ce qui nous conduit à minimiser

$$(3.43.) \quad \sum_{i=1}^m \left(\sum_{j=1}^{p'} \lambda_j \Pi_j(x_i) - z_i \right)^2$$

Remarquons que lorsque l'on applique cette méthode à f_0 , on peut être amenés, le cas échéant, à remplacer par ε_m des valeurs négatives ; ceci peut se révéler gênant dans le processus d'identification. Aussi a-t-on prévu, dans ce cas, la possibilité de ne sélectionner pour l'identification que les m' ($m' < m$) points où f_0 est supérieur à ε_m .

Remarquons que pour que la solution soit unique, il est nécessaire que le nombre de points sélectionnés soit supérieur ou égal au nombre p' de paramètres à estimer.

REMARQUES :

On notera que cette méthode d'identification introduit plusieurs éléments arbitraires.

+ l'élimination des parties négatives de la fonction à identifier,

+ le passage au logarithme : son seul intérêt est de rendre le problème linéaire et donc beaucoup moins gourmand en temps de calcul. Il est certain que l'estimation par moindres carrés faite sur la fonction elle-même conduirait à des résultats différents,

+ la présence d'un seuil ε_m . Sa valeur est indispensable pour limiter le poids qu'on risquerait de donner, après passage au logarithme, aux faibles valeurs de la fonction à identifier (puisqu'à une petite erreur sur l'estimation de la fonction correspondrait une grande erreur sur l'estimation du logarithme).

Toutes ces remarques doivent rester présentes à l'esprit de l'utilisateur pour lui permettre de conserver un sens critique vis à vis des résultats obtenus. En particulier, s'il s'avérait que la solution d'entropie maximum obtenue avec une telle solution initiale semble insatisfaisante, quel que soit le choix de m et de ε_m , on pourrait être amenés à affiner la technique d'identification.

3.4.4. Validation des solutions

3.4.4.1. Pouvoir de résolution

Il n'est pas fourni explicitement par la méthode. Comme nous l'avons mentionné en 2.3.3., l'utilisateur peut, s'il le souhaite, générer des données \hat{S}_i synthétiques correspondant à une fonction connue et comparer cette dernière à la solution fournie par la méthode.

3.4.4.2. Pouvoir de prédiction

On l'évalue, à postériori. Si f_2 est la solution d'entropie maximum, le pouvoir de prédiction est

$$P_r = \sum_{k=1}^p \frac{(\hat{S}_k - \int_I a_k(x) f_2(x) dx)^2}{\delta S_K^2}$$

3.4.4.3. Stabilité

Contrairement à ce qui se passait dans la méthode des noyaux de Dirichlet, on ne peut mettre en évidence, de manière explicite, l'influence de p' sur la stabilité de la solution.

On peut cependant concevoir que les causes d'instabilité sont qualitativement les mêmes (Lefevre, 1977). On est ainsi amené à penser que l'instabilité sera approximativement une fonction croissante de p' . On peut à postériori, après obtention d'une solution pour p' donné, calculer le coefficient de stabilité Q défini en 2.3.1. Il a ici pour expression approchée

$$Q \simeq 2H^{-1}$$

où H est la matrice Hessienne d'élément

$$H_{kl} = \frac{\partial^2 \phi}{(\partial \lambda_k \partial \lambda_l)} \Big|_{\lambda_k^0 \lambda_l^0}$$

où ϕ est la quantité à minimiser définie en 3.37.

3.4.5. Discussion

La recherche de la solution maximum d'entropie nécessite de la part de l'utilisateur

+ la connaissance d'une solution approchée. Le degré de ressemblance entre cette solution approchée et le solution dépendant essentiellement du problème de l'utilisateur, c'est à dire de la nature des noyaux. La "qualification" de cette solution initiale ne pourra se faire qu'à postériori.

Notons qu'il nous est arrivé qu'une solution initiale constante ($\lambda_i = 0 \ i = 1, p'$) permette d'arriver à une solution identique à celle obtenue à partir de valeurs initiales plus précises.

+ Les paramètres d'arrêt des itérations (3.4.2.). Ils ne posent pas de problèmes lorsque les erreurs sur les données sont bien connues.

+ Le choix du nombre p' de noyaux orthogonaux. C'est l'élément le plus difficile à évaluer à priori. Cependant la tâche est rendue aisée à l'utilisateur puisque

- d'une part p' ne peut prendre qu'un nombre limité de valeurs,

- d'autre part il dispose pour chaque solution de 3 critères d'évaluation qui permettent de bien cerner la validité des solutions obtenues.

BIBLIOGRAPHIE

- BACKUS G. and GILBERT F., "Numerical application of a formalism for geophysical inverse problem", Geophys. J.R. Astr. Soc., 13, 247-276, 1967.
- BACKUS G. and GILBERT F., "The resolving power of gross earth data", Geophys. J.R. Astr. Soc., 16, 2, 169-205, 1968.
- BACKUS G. and GILBERT F., "Uniqueness in the inversion of inaccurate gross earth data", Phil. Trans. Roy. Soc., London, 266, A 1173, 123-192, 1970.
- COURTILLOT V.E., DUCRUIX J. et LE MOUËL J.L., "Le prolongement d'un champ de potentiel d'un contour quelconque sur un contour horizontal : une application de la méthode de Backus et Gilbert", Ann. Geophys., 29, 361-366, 1973.
- DAVAL E., NAVET M. and SURY E., "Déconvolution par optimisation", Note Technique GRI/NTP/108, 1974.
- JOBERT G., "Problèmes inverses en géophysique", Traité de géophysique interne, ch 20, J. Coulomb et G. Jobert, Masson, Paris, 1973.
- LEFEUVRE F., "Analyse de champs d'ondes aléatoires observés dans la magnétosphère, à partir de la mesure simultanée de leurs six composantes", Thèse d'Etat, 19-12, 1977.
- LEFEUVRE F. and DELANNOY C., "Analysis of random electromagnetic wave field by a maximum entropy method", Annales des Télécommunications, Tome 35 n° 3-4, Avril 1979.
- ROUBINE E., "Introduction à la théorie de la communication, Tome III, Théorie de l'information", Masson & Cie, Paris, 1970.
- SABATIER P.C., "Remarks on approximate methods in geophysical inverse problems", Proc. R. Soc. London, 1, 337, 49-71, 1974.

- SABATIER P.C., "On geophysical inverse problem and constraints". Invited lecture at eleventh international symposium on mathematical geophysics, Seeheim, 1976.
- SABATIER P.C., "Positivity constraints in linear inverse problems-I", Geophys. J.R. Astr. Soc. 48, 415-441, 1977-a.
- SABATIER P.C., "Positivity constraints in linear inverse problems II", Geophys. J.R. Astr. Soc. 48, 443-459, 1977-b.
- STOREY L.R.O., "Electric field experiments : alternating fields", in "the ESRO Geostationary Magnetospheric Satellite", European Space Research Organization, Neuilly sur Seine, Report n° SP-60, 267-279, 1971.
- STOREY L.R.O. and LEFEUVRE F., "Theory for the interpretation of measurements of a random electromagnetic wave field in space", in "Space Research XIV" (Eds : M.J. Rycroft and R.D. Reasenberg), Akademie Verlag, Berlin, 381-386, 1974.
- STOREY L.R.O. and LEFEUVRE F., "The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma. 1-The direct problem", Note Technique CRPE/12, 1976.
- TWOMEY S., "The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements", J. Franklin Inst. 279, 95-109, 1965.
- TWOMEY S., "Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of a particle size distributions", J. Comp. Physics, 18, 188-200, 1975.

CRPE ORLEANS Cairo
Debrie
Décréau Pierrette
Feix
Hamelin
Parrot
Storey
Hayakawa

CRPE ISSY Cornilleau
Etcheto
Gendrin
Robert

LGE Pottelette
Malingre
Lavergnat
Berthelier

CEPHAG Glangeaud
Lacoume

STANFORD Carpenter
T. Miller

CORNELL R. Pfaff

COPENHAGUE T. Neubert

CANADA G. James

DELANNOY 20 exemplaires
LEFEUVRE 20 exemplaires
Bibliothèque 20 exemplaires

CRPE
*Centre de Recherches
en Physique de l'Environnement
terrestre et planétaire*

*Avenue de la Recherche scientifique
45045 ORLEANS CEDEX*

Département PCE
*Physique et Chimie
de l'Environnement*

*Avenue de la Recherche scientifique
45045 ORLEANS CEDEX*

Département ETE
*Etudes par Télédétection
de l'Environnement*

*CNET - 38-40 rue du général Leclerc
92131 ISSY-LES-MOULINEAUX*