HAL
open science

Etude méthodologique de l'analyse en composantes principales de processus bidimensionnels : effets des approximations numériques et de l'échantillonnage et utilisation pour la simulation de champs aléatoires : application au traitement des températures mensuelles de surface de la mer sur l'Atlantique intertropical

Isabelle Braud

To cite this version:

Isabelle Braud. Etude méthodologique de l'analyse en composantes principales de processus bidimensionnels : effets des approximations numériques et de l'échantillonnage et utilisation pour la simulation de champs aléatoires : application au traitement des températures mensuelles de surface de la mer sur l'Atlantique intertropical. Hydrologie. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1990. Français. NNT : . tel-00704988

HAL Id: tel-00704988

https://theses.hal.science/tel-00704988

Submitted on 6 Jun 2012

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.

THESE

présentée par

Isabelle BRAUD
Ingénieur E.N.S.H.M.G.
(I.N.P.G. Grenoble)

Pour obtenir le titre de DOCTEUR

de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(Arrêté ministériel du 23 novembre 1988)

Spécialité : Mécanique

Etude méthodologique de l'analyse en composantes principales de processus bidimensionnels. Effets des approximations numériques et de l'échantillonnage et utilisation pour la simulation de champs aléatoires Application au traitement des températures mensuelles de surface de la mer sur l'Atlantique intertropical.

Date de soutenance : 10 avril 1990

Composition du jury :

M. G. DE MARSILY		Président
M. M. DEQUE		Rapporteurs
M. PHAM DINH TUAN		
M. D. CADET		Examinateurs
M. Ch. OBLED		

Thèse préparée au sein de l'Institut de Mécanique de Grenoble
Laboratoire de l'I.N.P.G., U.J.F., associé au C.N.R.S. (UMR 101)

A mes parents.

"A peine exprimons-nous quelque chose qu'étrangement nous le dévaluons. Nous pensons avoir plongé au plus profond des abîmes, et quand nous revenons à la surface, la goutte d'eau ramenée à la pointe pâle de nos doigts ne ressemble plus à la mer dont elle provient. Nous nous figurons avoir découvert une mine de trésors inestimables, et la lumière du jour ne nous montre plus que des pierres fausses et des tessons de verre; et le trésor, inaltéré, n'en continue pas moins à briller dans l'obscur."

Maeterlinck.

AVANT-PROPOS

Le travail de recherche présenté dans ce mémoire a été réalisé dans l'équipe "Hydrométéorologie" du groupe Hydrologie de l'Institut de Mécanique de Grenoble, sous la direction de Mr. Ch. OBLED, Professeur à l'Ecole Nationale Supérieure d'Hydraulique et de Mécanique de Grenoble.

Je le remercie vivement d'avoir su me transmettre les fruits de sa longue expérience en Traitement de données et son enthousiasme pour la technique d'Analyse en Composantes Principales (ACP) de processus et pour ses perpétuels encouragements à aller plus loin dans l'approfondissement des mécanismes de cette méthode. De plus, son souci permanent de diffuser les résultats de cette recherche m'ont permis, par l'intermédiaire de séminaires ou de congrès, d'élargir le champ de mes connaissances, en évitant le piège du chercheur enfermé dans son domaine très pointu.

Qu'il me soit maintenant permis de remercier les autres membres du Jury:

J'exprime toute ma reconnaissance à Mr G. De MARSILY, Professeur à l'Université Pierre et Marie Curie (Paris VI), qui a bien voulu assurer la Présidence du Jury et apporter toute ses compétences de Géostatisticien aux débats.

Que Mr M. DEQUE, Ingénieur de la Météorologie Nationale, soit remercié d'avoir accepté d'être Rapporteur de cette thèse. Il a su s'acquitter de cette tâche en y apportant tout son savoir-faire de Météorologue et de praticien de l'ACP. C'est aussi grâce à lui que j'ai pu avoir connaissance des fichiers de données existant, susceptibles de servir d'illustration à mon travail.

Mr PHAM DINH TUAN, Directeur de recherche au CNRS, m'a amené toute sa rigueur de Mathématicien, et son aide m'a été précieuse pour mener à terme toute la partie théorique du chapitre 2. Je l'en remercie vivement, ainsi que d'avoir accepté d'être rapporteur de ce mémoire.

Enfin, j'exprime toute ma gratitude à Mr D. CADET, Directeur de recherche au CNRS et Chargé de Mission à l'INSU, qui a su replacer mes travaux dans le contexte plus large de la compréhension de l'évolution des climats.

Que MM. M. DEQUE et J. SERVAIN soient remerciés pour la fourniture grâcieuse des données qui m'ont permis d'illustrer, sur un cas réel, les approches méthodologiques et théoriques développées.

Un grand merci à mon collègue Y. RODRIGUEZ qui, en acceptant de jouer le rôle du "candide", m'a aidé à rendre les développements théoriques et la présentation des résultats plus accessibles aux non-spécialistes.

Je n'oublierai pas non plus mes collègues C. BARANCOURT, Ph. BOIS, J.D. CREUTIN, G. DELRIEU, K. KLADOU, J.M. NEYRET, F. SABBAGH, D. SEMPERE et tous les autres membres du groupe Hydrologie pour leur sympathie, la chaleur de leur accueil et leur disponibilité pour entamer des discussions amicales et scientifiques toujours enrichissantes.

Je terminerai en exprimant ma reconnaissance à MM. LAGARDE et RABATEL qui ont su répondre avec patience et gentillesse à mes questions concernant l'Informatique ainsi qu'à tout le personnel technique ayant participé à la réalisation matérielle de ce mémoire.

Enfin, je voudrais signaler que, sans le soutien financier du CNRS, ce travail n'aurait pas été possible.

SOMMAIRE

	Page
AVANT-PROPOS	3
SOMMAIRE	7
INTRODUCTION.	13
CHAPITRE 1:	
ANALYSE EN COMPOSANTES PRINCIPALES DE PROCESSUS. FORMULATION INTEGRALE, APPROXIMATION NUMERIQUE ET LIEN AVEC D'AUTRES TECHNIQUES D'ANALYSE.	17
1-1 INTRODUCTION	17
1-2 FORMULATION DE L'ACPP	17
1-2-1 Quelques rappels sur l'ACP classique.	17
1-2-2 Formulation de l'ACPP.	18
1-3 APPROXIMATION NUMERIQUE.	20
1-3-1 Approche générale.	20
1-3-2 Quelques algorithmes numériques.	23
1-3-3 Choix des fonctions de base $e_i(x)$.	24
1-4 LIENS ENTRE EOFs ET MODES NORMAUX DE SYSTEMES DYNAMIQUES	25
1-5 LIENS ENTRE ACPP ET METHODES D'INTERPOLATION OPTIMALE.	28
1-5-1 Rappels sur le krigeage.	28
1-5-1-1 *Le krigeage simple.*	28
1-5-1-2 *Formulation duale du krigeage simple.*	31
1-5-1-3 *Krigeage simple de données entachées d'erreurs de mesure.*	31
1-5-2 Liens entre l'analyse EOF et le krigeage.	33
1-5-2-1 *Estimateur proposé.*	33
1-5-2-2 *Analogie entre l'ACPP et le krigeage de données entachées d'erreur.*	34
1-5-2-3 *Analogie avec le krigeage dual*	37
1-6 CONCLUSION	38

CHAPITRE 2 :

COMPARAISON ENTRE FONCTIONS PROPRES D'UN NOYAU DE COVARIANCE ET LEURS APPROXIMATIONS NUMERIQUES. 39

2-1 POURQUOI CETTE PREOCCUPATION? 39

2-2 OBTENTION DES FONCTIONS PROPRES ANALYTIQUES D'UN NOYAU DE COVARIANCE A DEUX DIMENSIONS. 41

2-2-1 Cadre de la solution analytique. 41

2-2-2 Méthode générale de résolution. 44

2-2-3 Calcul complet des fonctions propres analytiques du modèle de covariance proposé par Fortus (1975). 48

2-2-4 Extension à d'autres modèles de covariance. 51

2-3 INFLUENCE DE LA DISCRETISATION SPATIALE DU PROBLEME SUR LA PRECISON DES FONCTIONS PROPRES. 53

2-3-1 Méthode de comparaison entre solution analytique et solution numérique. 53

2-3-2 Influence de la densité du réseau. 55

2-3-2-1 *Précision du calcul des valeurs propres.* 55

2-3-2-2 *Précision du calcul des fonctions propres.* 59

2-3-3 Influence de la régularité du réseau. 61

2-3-4 Choix des fonctions de base $e_i(x)$. 65

2-3-5 Influence de la portée du modèle de covariance. 68

2-4 INFLUENCE DE L'ECHANTILLONNAGE STATISTIQUE SUR LA PRECISION DES VALEURS PROPRES ET FONCTIONS PROPRES. 70

2-4-1 Estimation des valeurs propres. 71

2-4-2 Estimation des fonctions propres. 74

2-5 CONCLUSIONS 78

CHAPITRE 3 :

UTILISATION DE L'ACPP POUR LA SIMULATION DE CHAMPS BIDIMENSIONNELS. 81

3-1 INTRODUCTION 81

3-2 TECHNIQUES DE SIMULATION. 82

3-2-1 Méthodes spectrales. 83

3-2-2 Méthodes géométriques. 84

3-2-3 La méthode des bandes tournantes. 85

3-2-4 Les méthodes matricielles. 88

3-3 UTILISATION DE L'ACPP POUR LA SIMULATION DE CHAMPS BIDIMENSIONNELS. 89

3-4 SENSIBILITE DE LA SIMULATION AU CHOIX DES PARAMETRES. 93

3-4-1 Choix du nombre K de fonctions propres retenues dans la décomposition. 93

3-4-2 Influence de la discrétisation spatiale sur la covariance simulée. 98

3-4-3 Quelques remarques sur la distribution des $Z_k(\xi)$. 101

3-5 COMPARAISON DES PERFORMANCES DE LA SIMULATION PAR ACPP AVEC CELLES DES METHODES LU ET DES BANDES TOURNANTES. 108

3-6 EXEMPLE SYNTHETIQUE DE SIMULATION DE CHAMPS NON STATIONNAIRES PAR ACPP. 113

3-7 CONCLUSIONS 118

CHAPITRE 4:

APPLICATION DE L'ACPP A LA TEMPERATURE MENSUELLE DE SURFACE DE LA MER (SST) SUR L'ATLANTIQUE INTERTROPICAL DE 1964 A 1987. 121

4-1 INTRODUCTION 121

4-2 PRESENTATION DES DONNEES 123

4-2-1 Obtention du fichier final. 123

4-2-2 Statistiques simples sur les données. 125

4-3 ANALYSE DE LA STRUCTURE SPATIALE DE LA SST. 129

4-3-1 Méthodologie utilisée. 129

4-3-2 Analyse de la structure spatiale de la SST brute. 133

4-3-2-1 *Comportement au voisinage de l'origine.* 133

4-3-2-2 *Comportement pour les grandes distances.* 133

4-3-2-3 *Conclusions.* 137

4-3-3 Analyse de la structure spatiale de la SST désaisonnalisée. 139

4-3-3-1 *Comportement au voisinage de l'origine.* 139

4-3-3-2 *Comportement au voisinage de l'infini.* 139

4-4 APPLICATION DE L'ACPP AUX DONNEES DE SST ET COMPARAISON AVEC L'ACP CLASSIQUE. 143

4-4-1 Résultats déjà obtenus par ACP classique des données. 143

4-4-1-1 *ACP du cycle moyen.* 143

4-4-1-2 *ACP des données brutes.* 145

4-4-1-3 *ACP classique de données désaisonnalisées.* 145

4-4-2 Application pratique de l'ACPP aux données de SST. 149

4-4-2-1 *Les problèmes posés.* 149

4-4-2-2 *Mise en oeuvre pratique de l'ACPP lorsqu'on dispose de plus de points de mesure que de réalisations.* 152

4-4-3 Stabilité des F.P. calculées par ACP ou par ACPP lorsqu'on dégrade aléatoirement le réseau. 155

4-4-4 Comparaison entre F.P. calculées par ACP classique et par ACPP. 160

4-5 COMPARAISON DE DIFFERENTES TECHNIQUES DE RECONSTITUTION BASEES SUR L'ACP OU L'ACPP POUR LA RECONSTITUTION DES CHAMPS DE 1986 ET 1987. 164

4-5-1 Introduction. 164

4-5-2 Les différentes techniques de reconstitution testées. 164

4-5-3 Les critères de comparaison. 166

4-5-4 Analyse des résultats globaux. 167

4-5-5 Analyse des résultats champ par champ. 171

4-6 CONCLUSIONS 175

CONCLUSIONS GENERALES 177

REFERENCES BIBLIOGRAPHIQUES 181

ANNEXES: 191

ANNEXE A : OBTENTION DES FONCTIONS PROPRES ET DES COMPOSANTES PRINCIPALES PAR MINIMISATION D'UNE FONCTIONNELLE 191

ANNEXE B: QUELQUES PROPRIETES DES FONCTIONS DE COVARIANCE 197

ANNEXE B1 : La fonction de covariance $C^*(\underline{x},\underline{x}')$ du processus interpolé est de type positif. 197

ANNEXE B2 : La fonction de covariance $C_s(\underline{x},\underline{x}')$ du processus simulé est de type positif. 198

ANNEXE C : CALCUL PRATIQUE DE LA MATRICE E DES PRODUITS SCALAIRES DANS LE CAS D'UTILISATION DE FONCTIONS DE BASE LINEAIRES PAR FACETTE. 199

ANNEXE C1 : Cas d'une surface plane. 199

ANNEXE C2: Cas d'une surface sphérique. 201

ANNEXE D: OBTENTION DES FONCTIONS PROPRES ANALYTIQUES DU NOYAU DE COVARIANCE PROPOSE PAR FORTUS (1975) 203

ANNEXE D1: Rappels sur les fonctions de Bessel. 203

ANNEXE D2: Solution à l'intérieur du cercle: 203

ANNEXE D3: Solution à l'extérieur du cercle. 205

ANNEXE D4: Calcul des constantes C_1^a et C_2^a. 205

ANNEXE D5: Calcul de la constante C_1^a. 207

ANNEXE E: OBTENTION DES FONCTIONS PROPRES ANALYTIQUES DU NOYAU DE COVARIANCE $C(r) = \gamma r \,.\, K_1(\gamma r)$. 211

ANNEXE F: NOTION D'ECHELLE INTEGRALE. CALCUL POUR QUELQUES MODELES DE COVARIANCE USUELS. 213

INTRODUCTION

Dans de nombreuses disciplines, (Océanographie ou Météorologie par exemple), il est nécessaire de disposer de techniques efficaces permettant, à partir de tableaux de données souvent volumineux, d'extraire de précieux renseignements sur la structure ou la dynamique des phénomènes étudiés et, si possible, de condenser cette information. Ce dernier besoin est renforcé, depuis quelques années, par l'apparition, au côté de réseaux de mesure "ponctuels" traditionnels (stations météorolologiques, pluviographes, bouées marines, etc..), de moyens de télédétection (radars météorologiques, satellites..) qui fournissent, avec des fréquences assez élevées, des images échantillonnées sur des grilles très denses.

L'Analyse en Composantes Principales (ACP) ou analyse en Fonctions Orthogonales Empiriques (EOFs) pour les Anglo-saxons, appelée encore parfois décomposition de Karhunen-Loeve (Loeve,1963) permet de répondre à ce type de préoccupations en effectuant une transformation des variables initiales, qui, tout en extrayant le maximum de variance des données, permet d'obtenir de nouvelles variables statistiquement orthogonales.

Depuis ses débuts, la technique a fait l'objet de nombreux développements théoriques (voir Preisendorfer (1988) pour un bref historique de la méthode), mais surtout de nombreuses applications dans des domaines très variés. En effet, à la suite de Lorentz (1956), l'ACP a souvent été utilisée pour l'étude de la variabilité des champs géophysiques tels que la température de surface de la mer mesurée par des navires marchands (Servain et Legler, 1986) ou, plus récemment par satellite (Kelly, 1985; Lagerloeff et Bernstein, 1988); la tension de vent sur les océans (Barnier,1986; Mac Veigh et al., 1987); ou encore la hauteur du géopotentiel 500 mb (Rinne et Karhila,1975,1979; Dequé,1986; Molteni, 1987; Molteni et al., 1988). En Hydrométéorologie, Dyer (1975) puis Tabony (1981) ont utilisé l'analyse EOFs pour la description de la pluviométrie sur l'Europe occidentale, Obled et Creutin (1986) pour l'étude de champs de précipitation intense, Galle (1987), pour celle de la durée d'insolation quotidienne en France mesurée par un réseau sol d'héliographes ou par satellite. Son intérêt a aussi été montré pour l'étude de la stabilité de structures métalliques soumises à des sollicitations aléatoires (Kettani-Wagner, 1986). Nous avons fait ici le choix de citer des travaux récents, afin de souligner toute l'actualité de la méthode; mais on pourra trouver chez Obled (1979) des références de travaux plus anciens.

Grâce à l'analyse EOFs, de nombreux auteurs ont par ailleurs pu mettre en évidence l'existence de régions particulières du globe, jouant un rôle spécifique dans la climatologie mondiale et donnant lieu à ce qu'on appelle les oscillations Atlantique Nord ou

Pacifique Nord par exemple (Wallace et Gutzler, 1981; Horel, 1981; Simmons et al., 1983). L'ACP a aussi permis d'expliciter les liens existant entre plusieurs types de champs: Dequé et Servain (1989) pour les téléconnexions entre la température de surface de la mer sur l'Atlantique intertropical et la circulation atmosphérique aux latitudes tempérées de l'hémisphère Nord; Morlière et Rebert (1986) pour la relation entre l'oscillation Sud du courant El Niño et les précipitations en Nouvelle-Calédonie par exemple.

L'analyse EOF s'est aussi avérée utile pour l'étude et la description de la turbulence En effet, dès 1967, Lumley l'a utilisée pour l'étude des fluctuations de vitesse d'un écoulement turbulent inhomogène. Le tenseur des covariances associées au vecteur vitesse est, en effet, décomposé en une somme de fonctions propres orthogonales pondérées par des coefficients aléatoires (Lumley, 1970). Les tourbillons les plus "énergétiques" sont alors associés aux plus grandes valeurs propres. Ce type d'étude se poursuit toujours à l'heure actuelle, puisque Moin et Moser (1989) utilisent des décompositions de Karhunen-Loeve à deux ou trois dimensions pour décrire la turbulence dans un canal. Ils obtiennent par exemple que, près des parois, les tourbillons semblent mieux organisés et qu'il faut moins d'EOFs pour les caractériser que ceux du centre du canal.

Par ailleurs, certains auteurs ont souligné les analogies entre les EOFs, qui s'attachent plutôt à traduire les propriétés statistiques des phénomènes étudiés, et l'analyse spectrale (Fortus, 1973; Obled, 1979) ou les modes normaux des systèmes dynamiques (Buell, 1975; North, 1984) qui décrivent quant à eux la physique de ces phénomènes. Néanmoins, dans le cas de systèmes plus complexes (en particulier non linéaires), la relation entre EOFs et modes normaux n'est plus aussi simple (Smith, 1987); et inférer les processus physiques régissant les phénomènes étudiés à partir des EOFs ne paraît plus aussi simple.

Notons que toutes ces applications font intervenir des variables présentant une <u>continuité spatiale et/ou temporelle</u>. Mais, l'approche classique décrite en détail par Preisendorfer (1988), et largement développée en France dans les années 60 (Lebart et Fénelon, 1973, par exemple) ne prend pas en compte cette continuité, puisqu'on se contente de rechercher les valeurs propres et vecteurs propres d'une matrice de covariance de variables qui peuvent même être de nature différente. C'est pourquoi Escouffier et al. (1968), Lobert et Dauty (1969), Buell (1971), et plus récemment North et al. (1982) préconisent l'utilisation d'une <u>formulation intégrale</u>, qui a l'intérêt de tenir compte explicitement de la continuité des phénomènes étudiés. On est en effet conduit à rechercher les fonctions propres d'un noyau continu de covariance et la formulation fait aussi intervenir le domaine d'étude, ce qui n'est pas le cas avec la formulation classique. C'est cette approche continue qui a été à la base des travaux présentés dans ce mémoire.

Dans le premier chapitre, nous essaierons de mettre en évidence sa spécificité par rapport à l'ACP classique, puis de montrer les étapes permettant de passer d'une formulation intégrale à un problème aisément soluble sur ordinateur, avec un rappel des techniques numériques utilisées (Obled et Creutin, 1986). Nous présenterons ensuite une petite synthèse des travaux essayant de relier les EOFs aux modes normaux de systèmes dynamiques. Nous nous attacherons enfin, pour clore ce premier chapitre, à montrer comment, dans sa version continue, (notée par la suite ACPP, comme ACP de Processus), l'ACP peut se rattacher aux techniques d'interpolation optimale ou à l'approche géostatistique (Obled et Braud, 1989).

Dans une première partie de notre travail, nous avons essayé de quantifier les erreurs commises sur l'estimation des fonctions propres lors de la discrétisation numérique du problème. En effet, à notre connaissance, ce type d'étude n'avait pas été mené à deux dimensions, dans le cadre de l'approche continue; Bouhaddou (1984) ayant quant à lui traité le cas monodimensionnel. Buell (1975) avait aussi proposé une étude de l'influence du domaine d'étude sur les fonctions propres ou en 1978 quelques règles sur le nombre de fonctions propres significatives dans une ACPP, thème repris et approfondi par North et al. (1982). Nous avons jugé qu'une étude plus systématique pouvait être utile pour guider les utilisateurs de l'ACPP lors de la discrétisation de leur problème. C'est pourquoi, dans la lignée des travaux de Fortus (1973,1975) ou de Bouhaddou et al. (1987) , nous avons recherché, dans un cas simple, une solution analytique du problème intégral. Les différentes étapes de ce calcul sont résumées dans la première partie du chapitre 2. Dans la seconde, nous présentons les résultats de la comparaison entre solutions analytiques et numériques, qui nous a permis de dégager certains critères de choix des paramètres entrant en jeu dans la discrétisation du problème (Braud et Obled, 1989).

Après cette étude métholologique, nous nous sommes intéressés à la possibilité d'utiliser l'ACPP pour faire de la simulation de champs spatiaux. Ceci est un besoin souvent rencontré en pratique et l'ACPP peut fournir une alternative intéressante pour y répondre, comme nous le verrons au chapitre 3.

Enfin, dans le chapitre 4, nous essaierons d'illustrer toutes ces recherches sur un exemple pratique d'application. Il s'agit de données de température mensuelle de surface de la mer (SST) sur l'Atlantique intertropical qui avaient déjà fait l'objet d'un traitement par une ACP classique (Servain et Legler, 1986).

CHAPITRE 1: ANALYSE EN COMPOSANTES PRINCIPALES DE PROCESSUS. FORMULATION INTEGRALE , APPROXIMATION NUMERIQUE ET LIEN AVEC D'AUTRES TECHNIQUES D'ANALYSE.

1-1 INTRODUCTION

Ce premier chapitre est destiné à introduire la technique d'Analyse en Composantes Principales de Processus (ACPP), dont nous essaierons de montrer l'originalité par rapport à l'Analyse en Composantes Principales classique (ACP). Cette originalité réside essentiellement dans la prise en compte de la continuité spatiale et/ou temporelle des processus étudiés. Mais, comme en pratique ces derniers, même continus, sont échantillonnés, nous verrons comment, à l'aide d'algorithmes simples, on peut résoudre numériquement le problème. Ensuite, nous essaierons de voir comment l'ACPP peut être reliée à d'autres techniques d'analyse en évoquant tout d'abord les liens entre EOFs et modes normaux de systèmes dynamiques puis, dans la dernière partie de ce chapitre, en replaçant l'ACPP dans le cadre plus général des techniques d'interpolation optimale. Nous verrons en effet que, formellement, la méthode peut être considérée comme un cas particulier de krigeage.

1-2 FORMULATION DE L'ACPP

1-2-1 Quelques rappels sur l'ACP classique.

Considérons un tableau **X** de données x_{ij}, i=1..N, j=1..P, de P variables X_j, supposées centrées, dont on possède N observations x_{ij}. Les variables X_j peuvent être de nature différente, par exemple plusieurs variables météorologiques.

L'ACP de **X** consiste à rechercher de nouvelles variables Z_k, décorrélées et se déduisant des variables initiales par simple transformation linéaire:

$$Z_k = \sum_{j=1}^{P} X_j \cdot f_{jk}$$

On montre alors que les vecteurs F_k (dont les f_{jk} forment les coefficients) sont vecteurs propres de la matrice de variance-covariance empirique $\mathbf{C} = 1/N \,.\mathbf{X}^T.\,\mathbf{X}$ ($\mathbf{X}^T$ étant la

transposée de **X**) associée au tableau **X**. Ces vecteurs propres sont orthogonaux (et on peut les normer). Les valeurs propres associées λ_k sont alors les variances des nouvelles variables Z_k. Par ailleurs, si on classe les valeurs propres par ordre décroissant, la conservation des K<P premières composantes principales Z_k maximise la variance retenue.

Pour une description plus détaillée de cette méthode on pourra de référer à Lebart et Fénelon (1973) ou à Preisendorfer (1988).

1-2-2 Formulation de l'ACPP.

Nous avons souligné que l'ACP classique pouvait s'appliquer à des variables de nature différente. Mais, lorsqu'il s'agit de traiter des variables continues dans l'espace et/ ou le temps, (par exemple une même grandeur physique mesurée en plusieurs stations comme en Météorologie ou en Océanographie), une approche de type processus aléatoire est mieux adaptée à la prise en compte de cette continuité (Buell, 1971). On utilise alors une formulation intégrale, que l'on rencontre déjà chez Obukhov (1960) ou Holmström (1963). Elle est le pendant, pour des champs continus, de l'ACP classique décrite au § précédent, même si pratiquement , on est obligé de discrétiser le problème.

Cette approche présente l'avantage de prendre explicitement en compte:

i) la forme du domaine d'étude

ii) la typologie des réseaux de mesure, par l'intermédiaire de pondérations différentes si les stations sont irrégulièrement réparties.

Considérons donc un processus aléatoire $X(\xi,\underline{x})$, connu sur un domaine borné D de R ($\underline{x}$ est alors par exemple le temps t) ou de R^2 ($\underline{x}=(x,y)$ est alors le vecteur des coordonnées de l'espace).

ξ désigne ici un événement du processus aléatoire (ça pourrait aussi être l'état à un temps t pour un processus spatio-temporel).

Nous supposerons par la suite, sans perte de généralité, que $X(\xi,\underline{x})$ est centré, c'est-à-dire qu'en espérance, $E[\ X(\xi,\underline{x})\ ] = 0$. Nous ferons aussi l'hypothèse que les différentes réalisations aléatoires ξ sont indépendantes.

Par ailleurs, nous n'envisagerons que des processus aléatoires possédant une variance $E[\ X^2(\xi,\underline{x})\ ]$ finie, de façon à assurer l'existence de la fonction de covariance $C(\underline{x}, \underline{x}') = E[\ X(\xi,\underline{x})\ .\ X(\xi,\underline{x}')\ ]$. Cette dernière hypothèse est en général vérifiée par les processus physiques traités à l'aide de la technique d'ACP.

Plutôt qu'une décomposition en séries de Fourier, usuelle dans le domaine temporel (Davenport et Root, 1958; Papoulis, 1965 ou Brillinger, 1975), l'ACPP de $X(\xi,\underline{x})$ consiste à rechercher une décompositon de la forme:

$$X(\xi,\underline{x}) = \sum_{k=1}^{\infty} Z_k(\xi) \cdot F_k(\underline{x}) \qquad (1\text{-}1)$$

qui permet d'isoler des composantes spatiales $F_k(\underline{x})$ "déterministes" et des coefficients aléatoires, les $Z_k(\xi)$, dépendant uniquement de la réalisation aléatoire ξ considérée.

Le critère d'optimisation retenu pour déterminer à la fois les $Z_k(\xi)$ et les $F_k(\underline{x})$ est de type moindres carrés classiques puisqu'on cherche à minimiser, en espérance, l'erreur quadratique, moyennée sur l'ensemble du domaine D, entre $X(\xi,\underline{x})$ et $X_K(\xi,\underline{x})$ correspondant à la série tronquée à K termes.

$$E\left[\int_D \{ X(\xi,\underline{x}) - X_K(\xi,\underline{x}) \}^2 . d\underline{x} \right] \qquad \text{minimum} \qquad (1\text{-}2)$$

avec $$X_K(\xi,\underline{x}) = \sum_{k=1}^{K} Z_k(\xi).F_k(\underline{x})$$

L'optimisation est faite pas à pas pour K=1, puis K=2, jusqu'à K=∞. On ajoute par ailleurs, pour assurer l'unicité de la solution, une condition de normalisation à 1 des fonctions $F_k(\underline{x})$.

$$\int_D F_k^2(\underline{x}).d\underline{x} = 1 \qquad (1\text{-}3)$$

Plusieurs auteurs ont proposé des développements des calculs d'optimisation, mais le plus souvent sous une forme discrète (Davis, 1976; Von Storch et Hannoschöck, 1985). Dans l'Annexe A, nous proposons une démonstration, conservant jusqu'au bout les notations intégrales, et s'appuyant sur une approche variationnelle proposée par Holmström (1963, 1977).

On montre alors que:

i) les $F_k(\underline{x})$ sont des fonctions <u>orthonormales sur D</u>:

$$\int_D F_k(\underline{x}) \cdot F_m(\underline{x}) \cdot d\underline{x} = \delta_{km} \qquad (1\text{-}4)$$

où δ_{km} est le symbole de Kronecker.

Ce sont par ailleurs les fonctions propres (notées F.P. par la suite) du noyau de covariance $C(\underline{x}, \underline{x}') = E[X(\xi,\underline{x}) \cdot X(\xi,\underline{x}')]$ associées aux valeurs propres λ_k:

$$\int_D C(\underline{x},\underline{x}') \cdot F_k(\underline{x}') \cdot d\underline{x}' = \lambda_k \cdot F_k(\underline{x}) \qquad (1\text{-}5)$$

ii) les $Z_k(\xi)$ sont quant à elles des variables aléatoires, appelées composantes principales, <u>décorrélées, de variance λ_k décroissantes</u>:

$$E[Z_k(\xi) \cdot Z_m(\xi)] = \lambda_k \cdot \delta_{km} \qquad (1\text{-}6)$$

obtenues par "projection de $X(\xi,\underline{x})$ sur la kième fonction propre:

$$Z_k(\xi) = \int_D X(\xi,\underline{x}) \cdot F_k(\underline{x}) \cdot d\underline{x} \qquad (1\text{-}7)$$

1-3 APPROXIMATION NUMERIQUE.

Pour calculer explicitement les $Z_k(\xi)$ à l'aide de (1-7) ou estimer les fonctions propres en résolvant (1-5), l'emploi d'une formulation intégrale nécessite la connaissance du processus X en tout point du domaine D. Or, en pratique, X est spatialement discrétisé, c'est-à-dire connu en un nombre fini P de points ou stations de mesure $\underline{x}_i$, i=1..P. On ne pourra alors calculer que P fonctions propres et il sera nécessaire d'avoir recours à l'interpolation, donc d'approximer.

1-3-1 <u>Approche générale</u>.

La méthode utilisée, proposée par Deville (1974) à une dimension, a été généralisée à deux dimensions par Obled et Creutin (1986). En voici les caractéristiques principales.

On choisit, à priori, une base de fonctions ($e_i(\underline{x})$, i=1..P) ayant la structure d'espace vectoriel sur D, sur laquelle le processus $X(\xi,\underline{x})$ est interpolé. Si, de plus, les $\underline{x}_i$, i=1..P sont les P points de mesure, on peut choisir les $e_i(\underline{x})$ de façon canonique, c'est-à-dire telles que $e_i(\underline{x}_j) = \delta_{ij}$.

$X(\xi,\underline{x})$ s'écrit alors:

$$X^*(\xi,\underline{x}) = \sum_{i=1}^{P} X(\xi,\underline{x}_i) \cdot e_i(\underline{x})$$

où les $X(\xi,\underline{x}_i)$ sont les valeurs de X mesurées aux P stations.

Le processus $X(\xi,\underline{x})$ étant supposé centré, la fonction de covariance $C^*(\underline{x},\underline{x}')$ du processus interpolé $X^*(\xi,\underline{x})$, est alors donnée par une double interpolation sur la base des $e_i(\underline{x})$:

$$C^*(\underline{x},\underline{x}') = \sum_{i=1}^{P} \sum_{j=1}^{P} C(\underline{x}_i,\underline{x}_j).e_i(\underline{x}).e_j(\underline{x}')$$

où les $C(\underline{x}_i,\underline{x}_j)$ sont les covariances "vraies" entre deux points quelconques du réseau. En pratique, ces covariances sont estimées sur un échantillon fini de N réalisations, ce qui introduit des erreurs supplémentaires dues à l'échantillonnage statistique du processus. Dans ce paragraphe, seul le problème posé par la discrétisation spatiale est traité.

Il est par ailleurs assez aisé de montrer que $C^*(\underline{x},\underline{x}')$, fonction de covariance du processus interpolé, est bien de type positif, c'est-à-dire un modèle de covariance admissible au sens de Matheron (1972-a) ou Christakos (1984). (Annexe B1).

Les F.P. $F_k(\underline{x})$, solutions de (1-5), sont aussi approximées sur la base des $e_i(\underline{x})$ par l'intermédiaire de coefficients f_{ik}, i=1..P, à déterminer. (f_{ik} représente la valeur de la fonction $F_k(\underline{x})$ au point $\underline{x}_i$).

$$F_k^*(\underline{x}) = \sum_{i=1}^{P} f_{ik} \cdot e_i(\underline{x}) \qquad (1\text{-}8)$$

Ensuite, dans (1-5), on remplace la covariance vraie $C(\underline{x},\underline{x}')$ et les F.P. vraies par leurs valeurs approchées $C^*(\underline{x},\underline{x}')$ et $F_k^*(\underline{x})$, ce qui donne:

$$\int_D \{ \sum_{i=1}^{P} \sum_{j=1}^{P} C(\underline{x}_i,\underline{x}_j).e_i(\underline{x}).e_j(\underline{x}') \}.\{\sum_{m=1}^{P} f_{mk} \cdot e_m(\underline{x}')\}.d\underline{x}' = \lambda_k. \sum_{i=1}^{P} f_{ik} \cdot e_i(\underline{x})$$

soit encore:

$$\sum_{i=1}^{P} \{ \sum_{j=1}^{P} \sum_{m=1}^{P} C(\underline{x}_i,\underline{x}_j). \int_D e_j(\underline{x}').e_m(\underline{x}').d\underline{x}' \,.\, f_{mk} - \lambda_k.\ f_{ik} \} \,.\, e_i(\underline{x}) = 0$$

Les $e_i(\underline{x})$ forment une base du plan. Tous les termes entre accolades dont donc nuls, et, si on pose $E_{jm} = \int_D e_j(\underline{x}').e_m(\underline{x}').d\underline{x}'$, on en déduit:

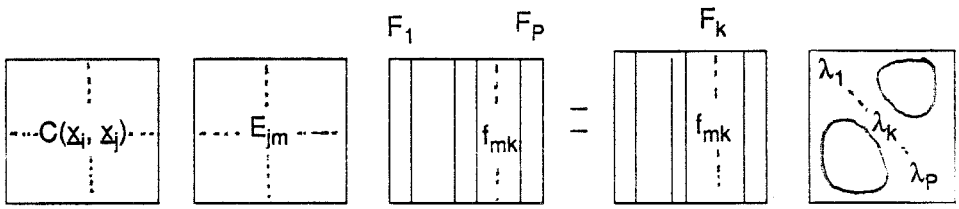
$$\sum_{j=1}^{P} \sum_{m=1}^{P} C(\underline{x}_i,\underline{x}_j).\ E_{jm} \,.\, f_{mk} = \lambda_k.\ f_{ik} \qquad i=1..P$$

ou encore sous forme matricielle:

$$\mathbf{C} \,.\, \mathbf{E} \,.\, \mathbf{F} = \mathbf{F} \,.\, \Lambda \qquad (1\text{-}9)$$

avec

$\mathbf{C} = [\ C(\underline{x}_i,\underline{x}_j)\ \}\ i=1..p,\ j=1..P$	matrice carrée des covariances entre deux points du réseau.
$\mathbf{E} = [\ E_{jm}\ ]\ j=1..P,\ m=1..P$	matrice carrée des produits scalaires construits à partir des $e_i(\underline{x})$.
$\mathbf{F} = [\ f_{mk}\ ]\ m=1..P,\ k=1..P$	matrice des valeurs des F.P. aux points de mesure.
Λ	matrice diagonale des valeurs propres λ_k, k=1..P.



On est donc ramené à la résolution d'un problème matriciel de valeurs propres et de vecteurs propres, la seule différence avec l'ACP classique résidant dans <u>l'introduction de la matrice **E** des produits scalaires</u> qui prend explicitement en compte le domaine d'étude et la typologie du réseau des points de mesure.

La seconde étape du calcul consiste à estimer les $Z_k(\xi)$. Comme précédemment, en utilisant la base des $e_i(\underline{x})$, (1-7) est transformée de la façon suivante:

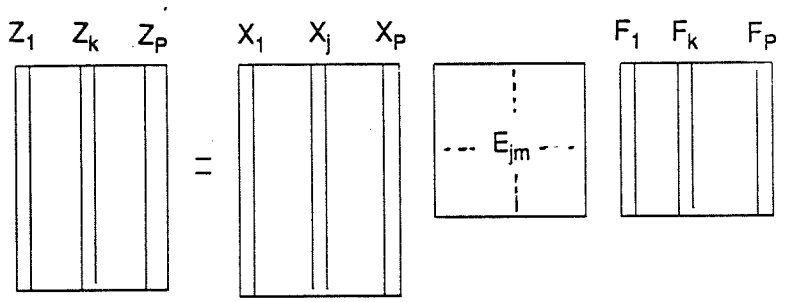
$$Z_k(\xi) = \int_D \sum_{i=1}^{P} X(\xi,\underline{x}_i) \cdot e_i(\underline{x}) \cdot \sum_{j=1}^{P} f_{jk} \cdot e_j(\underline{x}) \cdot d\underline{x}$$

$$Z_k(\xi) = \sum_{i=1}^{P} \sum_{j=1}^{P} X(\xi,\underline{x}_i) \cdot \underbrace{\int_D e_i(\underline{x}) \cdot e_j(\underline{x}) \cdot d\underline{x}}_{E_{ij}} \cdot f_{jk}$$

$$Z_k(\xi) = \sum_{i=1}^{P} \sum_{j=1}^{P} X(\xi,\underline{x}_i) \cdot E_{ij} \cdot f_{jk} \qquad (1\text{-}10)$$

Soit sous forme matricielle:

$$\mathbf{Z = X \cdot E \cdot F} \qquad (1\text{-}11)$$



si **X** = [$X(\xi_i,\underline{x}_j)$] i=1..N, j=1..P est la matrice des N réalisations (ou champs) du processus $X(\xi,\underline{x})$ et **Z** = [$Z_k(\xi_i)$] i=1..N, k=1...P, la matrice des composantes principales correspondantes.

1-3-2 Quelques algorithmes numériques.

Nous ne détaillerons pas ici les algorithmes de calcul des valeurs propres et vecteurs propres dont on pourra trouver des descriptions détaillées chez Faye (1982). Nous nous bornerons à souligner que les algorithmes sont beaucoup plus performants lorsque la matrice à diagonaliser est symétrique, ce qui n'est pas, en général, le cas pour la matrice **C.E** de (1-9), même si **C** et **E**, séparément, le sont. Nous allons donc voir comment on peut se ramener simplement à la diagonalisation d'une matrice symétrique.

E étant symétrique, elle peut se décomposer sous la forme $\mathbf{E} = \mathbf{Q} \,.\, \mathbf{Q}^T$. Obled et Creutin (1986) proposaient d'obtenir **Q** par diagonalisation de **E**. Pour des questions de gain en temps de calcul, nous avons retenu une décomposition de Choleski où **Q** est, cette fois, une matrice triangulaire inférieure.

En prémultipliant, à gauche, les deux membres de (1-9) par $\mathbf{Q}^T$, on obtient:

$$(\mathbf{Q}^T \,.\, \mathbf{C} \,.\, \mathbf{Q}) \,.\, \mathbf{V} = \mathbf{V} \,.\Lambda$$

en ayant posé $\mathbf{V} = \mathbf{Q}^T \,.\, \mathbf{F}$

$(\mathbf{Q}^T \,.\, \mathbf{C} \,.\, \mathbf{Q})$ est, cette fois, symétrique. On calcule donc ses vecteurs propres **V**, à l'aide d'algorithmes classiques, puis on en déduit **F** par $\mathbf{F} = (\mathbf{Q}^T)^{-1} \,.\, \mathbf{V}$. Comme $\mathbf{Q}^T$ est triangulaire supérieure, son inversion est rapide.

Enfin, on peut montrer simplement (Obled et Creutin ,1986) que la condition d'orthogonalité (1-4) des F.P. se traduit matriciellement par:

$$\mathbf{F}^T \,.\, \mathbf{E} \,.\, \mathbf{F} = \mathbf{E} \,.\, \mathbf{F}^T \,.\, \mathbf{F} = \mathbf{F} \,.\, \mathbf{F}^T \,.\, \mathbf{E} = \mathbf{I}_P \qquad (1\text{-}12)$$

où $\mathbf{I}_P$ est la matrice identité de rang P.

1-3-3 Choix des fonctions de base $e_i(\underline{x})$.

Nous avons souligné que la principale différence entre l'ACP classique et l'ACPP résidait, au niveau de la solution numérique, dans l'introduction de la matrice **E** des produits scalaires. Encore faut-il définir comment elle sera remplie. En pratique, à deux dimensions, deux types de fonctions $e_i(\underline{x})$ ont été utilisés.

i) *des fonctions constantes par morceau (CPM)*
où $e_i(\underline{x})$ vaut un sur le polygone contenant tous les points les plus proches de $\underline{x}_i$ (polygone de Thiessen) et zéro partout ailleurs (Fig. 1-1-a). Dans ce cas, **E** est diagonale et les termes diagonaux sont égaux aux surfaces des polygones.

ii) *des fonctions linéaires par facette (LPF)*
$e_i(\underline{x})$ vaut un au point $\underline{x}_i$ et zéro aux autres points, avec une décroissance linéaire sur les facettes reliant les différents sommets (Fig. 1-1-b). Dans ce cas, il est nécessaire d'établir une triangulation du réseau de mesure. Le calcul des E_{ij}, décrit en détail dans Obled et Creutin (1986) utilise alors les techniques des éléments finis (Norrie et De Vries, 1973). Ce calcul est repris dans l'Annexe C1.

Signalons par ailleurs que la définition des frontières du domaine d'étude peut poser quelques problèmes, notamment lorsque ce dernier n'est pas défini par une frontière physique ou géographique intrinsèque au phénomène (étude des champs de pression sur l'Europe par exemple). De plus, la triangulation du réseau, dans le cas LPF, n'est pas

nécessairement unique et, si une station est très excentrée par rapport aux autres, on peut hésiter à lui donner un poids trop important. Nous reviendrons plus en détail sur ces problèmes liés au domaine d'étude au chapitre 2.

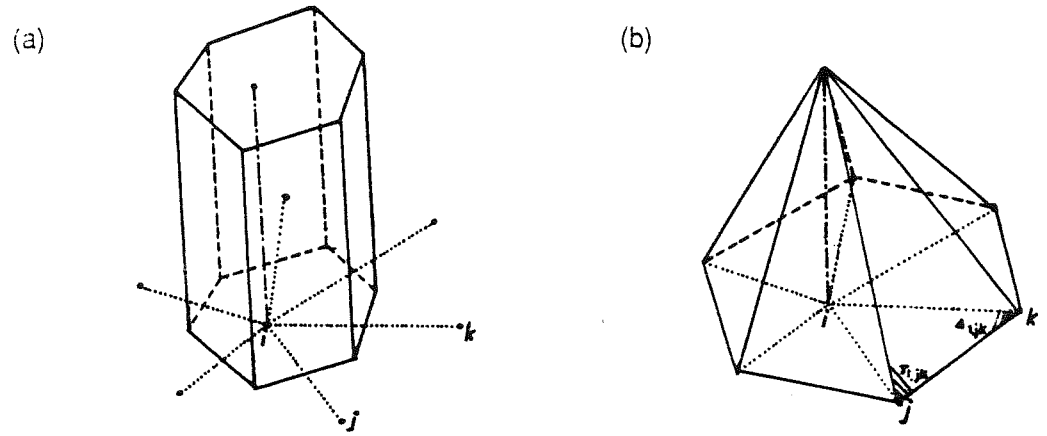


Fig. 1-1: Fonctions de base $e_i(\underline{x})$. (a) Fonctions constantes par morceau (CPM)
(b) Fonctions linéaires par facette (LPF).

Après avoir précisé la formulation de l'ACPP et vu comment, numériquement, on pouvait traiter le problème, nous allons essayer de montrer que cette technique peut être aisément rattachée à d'autres techniques d'analyse, notamment la recherche de modes normaux de systèmes dynamiques ou les techniques géostatistiques d'interpolation optimale; ce qui peut permettre des interprétations plus aisées des résultats.

1-4 LIENS ENTRE EOFs ET MODES NORMAUX DE SYSTEMES DYNAMIQUES

L'étude de séries temporelles se base souvent sur une analyse spectrale du signal, et les travaux de Fortus (1973) sur les EOFs à une dimension, ont montré que, lorsque l'intervalle de temps étudié tend vers l'infini, EOFs et séries de Fourier sont confondues. Obled (1979) propose aussi quelques développements sur ce sujet.

Si maintenant on s'intéresse à l'évolution de variables spatio-temporelles (telles que la pression atmosphérique ou le champ de vent par exemple), obéissant à un système d'équations aux dérivées partielles, il est souvent très instructif de rechercher les modes normaux de ces systèmes dynamiques. Ceux-ci constituent en effet une base de réponses élémentaires du système - ne dépendant en général que de la variable d'espace- lorsque ce dernier n'est soumis à aucune excitation (ou forçage) extérieure. Ils permettent une simplification de l'étude de l'évolution temporelle du système.

Par exemple, considérons le cas simple où $X(t,\underline{x})$ vérifie:

$$(\partial^2 / \partial t^2 - a^2 . \Delta) X(t,\underline{x}) = f(t,\underline{x})$$

Δ étant le symbole du Laplacien, $f(t,\underline{x})$ un forçage quelconque.

Si, en l'absence de forçage, c'est-à-dire en supposant $f(t,\underline{x})=0$, on cherche $X(t,\underline{x})$ sous la forme $X(t,\underline{x}) = e^{i\nu t} . \Psi(\underline{x})$, $\Psi(\underline{x})$ sera un mode normal vérifiant:

$$a^2 . \Delta\Psi(\underline{x}) + \nu^2 . \Psi(\underline{x}) = 0$$

De la même manière, l'analyse EOF peut être vue comme une technique qui permet de séparer les composantes spatiales et temporelles des champs étudiés. La recherche d'une analogie entre EOFs et modes normaux devient alors justifiée, d'autant plus qu'elle est confirmée par certaines observations expérimentales. Ainsi, Buell (1975) remarque que les EOFs obtenues sur un domaine rectangulaire avec une fonction de covariance de type exponentiel ou gaussien ressemblent fort aux modes normaux d'une plaque vibrante. Preisendorfer (1979), est allé plus loin puisque, à partir de l'étude systématique d'un système dynamique de masses et de ressorts, soumis ou pas à un amortissement, il a montré que, lorsque le temps d'observation du système tendait vers l'infini, vecteurs propres de la matrice de covariance construite à partir de l'enregistrement d'un grand nombre de trajectoires et modes propres du système dynamique étaient confondus, alors que les valeurs propres étaient reliées au carré des amplitudes de ces modes. Plus récemment (Preisendorfer, 1988) cette étude a été généralisée à un système de masses et de ressorts soumis à une excitation aléatoire (ou forçage) quelconque. La propriété précédente (que Preisendorfer nomme "propriété asymptotique des composantes principales") reste vérifiée quelles que soient les conditions initiales; avec ou sans frottement, pour un forçage particulier dont les composantes temporelles, sur la base des modes normaux, sont décorrélées. Si le forçage est quelconque mais stationnaire, modes normaux et EOFs peuvent se déduire les uns des autres par simple rotation. Ces conclusions restent vraies si on étudie un champ aléatoire de la variable d'espace à deux ou trois dimensions, régi par une équation aux dérivées partielles d'ordre deux (comme celles des grandes oscillations du vent sur les océans ou de la pression atmosphérique au niveau de la mer, après linéarisation), (Preisendorfer, 1988).

Il semble même que dès 1965, les Russes Monin et Obukhov (cités par Fortus (1980) et North (1984)) étaient arrivés à ces résultats avec une équation aux dérivées partielles plus simple. North (1984) a permis de franchir un pas supplémentaire en montrant que, pour un opérateur différentiel normal (*) et pour un forçage stationnaire dans le temps et blanc dans l'espace, EOFs et modes normaux étaient confondus.

(*) Soit Φ et Ψ, deux fonctions quelconques et (,), le produit scalaire défini par

$(\Phi , \Psi) = \int_D \Phi(\underline{x}) . \Psi(\underline{x}) . d\underline{x}$

Un opérateur différentiel A est <u>normal</u> s'il commute avec son adjoint A^+, défini par $(\Phi, A\Psi) = (A^+\Phi, \Psi)$ pour toutes fonctions Φ et Ψ.

Il est <u>hermitien</u> s'il est confondu avec son adjoint, c'est-à-dire si $A = A^+$.

En revanche, pour des opérateurs non normaux, notamment lorsqu'il y a des non linéarités ou du frottement, cette propriété n'est plus vérifiée. Smith (1987) a essayé d'aller plus loin en montrant que pour des opérateurs hermitiens (*) soumis à un forçage quelconque, EOFs et modes normaux se déduisent les uns des autres par simple rotation. Il a aussi abordé le cas d'opérateurs non hermitiens (les plus intéressants en pratique car tout système physique est amorti), en montrant qu'on pouvait relier EOFs et modes normaux assez facilement. En effet, le signal initial peut être décomposé sur la base des modes normaux. Les séries temporelles obtenues sont appelées amplitudes modales. On peut en calculer les transformées de Fourier et construire la matrice spectrale associée. Cette matrice est hermitienne et il est aisé d'en calculer les vecteurs propres. Les modes normaux formant une base complète, on peut aussi décomposer les EOFs sur cette base. On constate alors que les coefficients qui interviennent dans cette décomposition sont les vecteurs propres de la matrice spectrale introduite précédemment. On peut donc déduire les EOFs des modes normaux et de la matrice spectrale et réciproquement.

Néanmoins, tous ces développements restent purement formels et à notre connaissance n'ont pas donné lieu à des applications pratiques. Il semblerait pourtant séduisant, comme certains auteurs l'ont déjà formulé (ECMWF, 1977; Preisendorfer, 1988) d'utiliser les EOFs pour tenter d'inférer la structure des modèles dynamiques ou prendre en compte certains phénomènes de manière statistique tant qu'on ne sait pas les modéliser de façon déterministe.

Ces espoirs ne font pourtant pas l'unanimité puisque Von Storch et Hannoschöck (1985) pensent que rechercher des liens entre EOFs et modes normaux sera rarement couronné de succès dans la mesure où les EOFs sont un objet mathématique soumis à des contraintes d'orthogonalité et de maximisation de la variance retenue, alors que les modes normaux ne sont orthogonaux que dans des cas très particuliers, en général éloignés des cas réels. Ils préconisent donc de limiter l'utilisation de l'ACPP à la condensation de l'information sans vouloir chercher à les interpréter. Cette interprétation s'avère d'ailleurs d'autant plus délicate que les EOFs dont on dispose sont estimées à partir d'un échantillon fini et donc entachées de biais et d'erreurs, comme ils l'explicitent dans la suite de leur article, et comme nous le verrons au chapitre 2.

On voit donc que l'interprétation des EOFs en terme de dynamique des systèmes étudiés, pour séduisante qu'elle soit, doit être menée avec précaution et on ne doit pas oublier que les EOFs ne décrivent que des statistiques sur les réponses de ces systèmes. Elles peuvent en revanche être employées efficacement comme bases de projection des processus et être introduites directement dans les modèles (Rinne et Karhila, 1975; Schubert, 1985).

Les liens entre analyse EOF et méthodes d'interpolation optimale que nous allons étudier maintenant, vont aussi permettre de faire entrer l'ACPP dans la grande famille des méthodes géostatistiques.

1-5 LIENS ENTRE ACPP ET METHODES D'INTERPOLATION OPTIMALE.

Comme nous l'avons souligné lors de la présentation de l'appoximation numérique de l'analyse EOF, les processus étudiés sont, en général, spatialement discrétisés en un nombre fini P de points. Or, en pratique, (Météorologie, Sciences de la terre, etc..) il est souvent nécessaire de les connaître en d'autres points, non mesurés (cartographie) ou sur une surface ou un volume donné (estimation de lames d'eau moyennes sur un bassin, ou de la puissance de gisements miniers). Il faut donc interpoler. De nombreuses techniques existent: interpolation polynomiale (Appert et Louveau, 1973), fonctions splines de type plaque mince (Duchon, 1976), krigeage (Matheron, 1965) auxquelles il convient d'ajouter l'analyse EOF. En effet, Creutin (1979), Creutin et Obled (1982) ont testé et validé l'utilisation de l'ACPP comme méthode d'estimation de surfaces. Obled et Creutin (1986), après application à des champs pluviométriques intenses des Cévennes, ont obtenu des résultats comparables pour l'ACPP et le krigeage, et légèrement supérieurs aux autres techniques d'interpolation.

Dans ce paragraphe, nous ne reprendrons pas ces travaux mais, nous nous bornerons, après un rappel des principes du krigeage, à montrer les analogies formelles qui existent entre ACPP et krigeage simple de données entachées d'erreurs.

1-5-1 Rappels sur le krigeage.

On pourra trouver une présentation complète de la méthode chez Matheron (1965) qui, le premier, en a explicité les fondements mathématiques. Nous nous contenterons ici de rappeler le formalisme du krigeage simple en covariance (Gandin, 1965) sous la forme classique en pondérateur et sous sa forme duale, puis celui du krigeage de données entachées d'erreurs de mesure.

1-5-1-1 *Le krigeage simple.*

Soit un processus $X(\xi,\underline{x})$, centré ($E[X(\xi,\underline{x})] = 0$), échantillonné en P points $\underline{x}_j$, $j=1..P$ d'un domaine borné D de R^2 (Fig. 1-2). Nous nous plaçons par ailleurs dans un contexte de multiréalisations notées $\xi=i$, $i=1..N$, de façon à pouvoir estimer les espérances et les covariances.

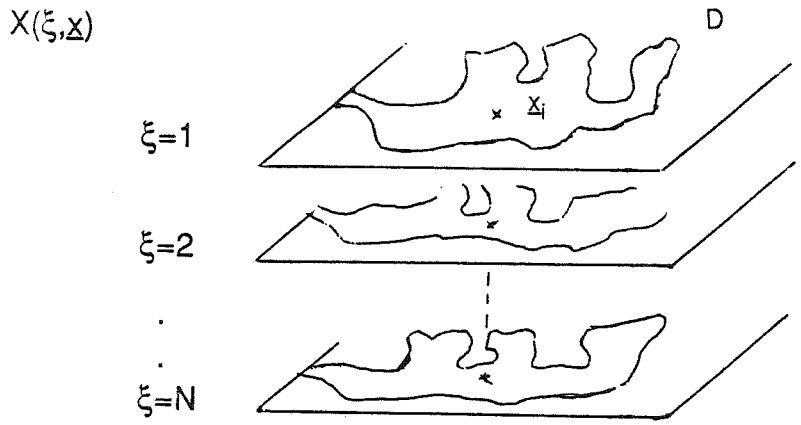


Fig. 1-2: Echantillonnage spatial du processus $X(\xi,\underline{x})$ sur un domaine borné D, dans un contexte de multiréalisations.

Le problème que l'on se pose est le suivant: connaissant $X(\xi,\underline{x})$ aux P points $\underline{x}_j$ de mesure, comment peut-on estimer la valeur en $\underline{x}_o$ qui n'a pas été mesurée?

Le krigeage simple propose un estimateur linéaire $X^*(\xi,\underline{x}_o)$:

$$X^*(\xi,\underline{x}_o) = \sum_{i=1}^{P} \mu_i \cdot X(\xi,\underline{x}_i) \qquad (1\text{-}13)$$

qui apparaît comme une pondération des valeurs mesurées.

Les critères qui permettent de calculer les μ_i sont doubles:

i) l'estimateur doit être sans biais, i.e. $E\,[\,X(\xi,\underline{x}_o) - X^*(\xi,\underline{x}_o)\,] = 0$. Cette condition est automatiquement vérifiée si le processus est centré.

ii) l'estimateur doit être optimal dans le sens où il doit minimiser la variance de l'erreur de reconstitution:

$$\sigma^2(\underline{x}_o) = E\left[\,\{\,X(\xi,\underline{x}_o) - X^*(\xi,\underline{x}_o)\,\}^2\,\right] \quad \text{minimale} \qquad (1\text{-}14)$$

L'optimisation par rapport aux paramètres μ_i conduit à la résolution du système linéaire suivant:

$$\sum_{i=1}^{P} \mu_i \cdot C(\underline{x}_i, \underline{x}_j) = C(\underline{x}_j, \underline{x}_o) \qquad j = 1,..P \qquad (1\text{-}15)$$

(voir par exemple Obled (1987) pour le détail des calculs).

Sous forme matricielle, on écrira:

$$\mathbf{C} \cdot \boldsymbol{\mu} = \mathbf{C_o} \qquad (1\text{-}16)$$

$$\begin{bmatrix} & \vdots & \\ \cdots & C(\underline{x}_i, \underline{x}_j) & \cdots \\ & \vdots & \end{bmatrix} \begin{bmatrix} \mu_1 \\ \cdot \\ \mu_j \\ \cdot \\ \mu_P \end{bmatrix} = \begin{bmatrix} C_o(\underline{x}_1, \underline{x}_o) \\ \cdot\cdot \\ C_o(\underline{x}_j, \underline{x}_o) \\ \cdot\cdot \\ C_o(\underline{x}_P, \underline{x}_o) \end{bmatrix}$$

avec $\mathbf{C} = [\, C(\underline{x}_i, \underline{x}_j) \,]$ i=1..P, j=1..P — matrice des covariances entre deux points quelconques du réseau

$\boldsymbol{\mu} = [\, \mu_1, \mu_2, \ldots, \mu_P \,]^T$ — vecteur des pondérations inconnues

$\mathbf{C_o} = [\, C_o(\underline{x}_1, \underline{x}_o), \ldots\ldots, C_o(\underline{x}_P, \underline{x}_o) \,]^T$ vecteur des covariances entre $\underline{x}_o$ et un point quelconque du réseau.

En pratique, puisqu'on se place dans un contexte de multiréalisations, on a accès aux $C(\underline{x}_i, \underline{x}_j)$ que l'on peut estimer sur le jeu de données. En revanche, les $C(\underline{x}_j, \underline{x}_o)$ sont inconnues, puisqu'on n'a pas de mesure en $\underline{x}_o$. Pour lever ce problème, l'approche classique consiste à supposer le processus stationnaire d'ordre deux (c'est-à-dire que ses moments d'ordre un et deux sont invariants par translation dans l'espace), donc $C(\underline{x}, \underline{x}') = C(\underline{x} - \underline{x}')$. On ajuste alors un modèle sur le nuage des $C(\underline{x}_i, \underline{x}_j)$ qui sera utilisé pour remplir le vecteur $\mathbf{C_o}$.

Le krigeage nous donnera, outre une estimation de $X(\xi, \underline{x}_o)$, la variance de l'erreur commise:

$$\sigma^2(\underline{x}_o) = C(\underline{x}_o, \underline{x}_o) - \sum_{i=1}^{P} \mu_i \cdot C(\underline{x}_i, \underline{x}_o) \qquad (1\text{-}17)$$

Notons par ailleurs que le krigeage est une méthode d'<u>interpolation</u>, donc la surface estimée passe par les points de mesure, i.e.: $X^*(\xi, \underline{x}_i) = X(\xi, \underline{x}_i)$

<u>Remarque</u>: Si la moyenne du processus est inconnue, la vérification de la condition de non-biais introduit des inconnues supplémentaires sous la forme de multiplicateurs de Lagrange qui viennent s'ajouter aux μ_i.

1-5-1-2 *Formulation duale du krigeage simple.*

Le krigeage classique (formule (1-13)) propose un estimateur sous forme de pondération des valeurs mesurées. Il existe une autre formulation équivalente à (1-13), appelée formulation duale où l'estimateur $X^*(\xi,\underline{x}_o)$ est obtenu par interpolation sur une base de fonctions que nous expliciterons plus loin (Matheron, 1970; Dubrule, 1982; Royer et Viera, 1984; Obled, 1987).

En effet, si on écrit (1-13) sous forme matricielle:

$$X^*(\xi,\underline{x}_o) = [\,\mathbf{X}(\xi)\,]^T \,.\, \boldsymbol{\mu}$$

avec $$[\,\mathbf{X}(\xi)\,]^T = [\,X(\xi,\underline{x}_1), \ldots, X(\xi,\underline{x}_P)\,]^T$$

et, si de (1-16), on tire $\boldsymbol{\mu} = \mathbf{C}^{-1} \,.\, \mathbf{C}_o$, on obtient:

$$X^*(\xi,\underline{x}_o) = [\,\mathbf{X}(\xi)\,]^T \,.\, \mathbf{C}^{-1} \,.\, \mathbf{C}_o = \mathbf{B}^T \,.\, \mathbf{C}_o$$

si on pose $$\mathbf{B}^T = [\,\mathbf{X}(\xi)\,]^T \,.\, \mathbf{C}^{-1} = [b_1, \ldots, b_P]^T \qquad (1\text{-}18)$$

L'interpolation de $X(\xi,\underline{x}_o)$ se fait donc sur une base de fonctions formée des covariances $C(\underline{x}_i, \underline{x}_o)$, i=1,...P, entre un point du réseau et le point $\underline{x}_o$, les coefficients d'interpolation étant quant à eux les b_i, qui se déduisent des μ_i par (1-18).

$$X^*(\xi,\underline{x}_o) = \sum_{i=1}^{P} b_i \,.\, C(\underline{x}_i, \underline{x}_o) \qquad (1\text{-}19)$$

1-5-1-3 *Krigeage simple de données entachées d'erreurs de mesure.*

Dans la pratique, les données mesurées sont souvent entachées d'erreur. Soit $Y(\xi,\underline{x})$, la variable mesurée et $X(\xi,\underline{x})$, la "vraie" valeur associée.

Supposons que $Y(\xi,\underline{x})$ s'écrive:

$$Y(\xi,\underline{x}) = X(\xi,\underline{x}) + er(\xi,\underline{x})$$

où $er(\xi,\underline{x})$ est un terme d'erreur que l'on supposera de moyenne nulle, de variance $\sigma^2(\underline{x})$, blanc dans l'espace et décorrelé de X, i.e.:

$$E[\, er(\xi,\underline{x}) \,] = 0$$

$$E[\, er(\xi,\underline{x}) \,.\, er(\xi,\underline{x}') \,] = \delta(\underline{x} - \underline{x}') \,.\, \sigma^2(\underline{x})$$

$$E[\, er(\xi,\underline{x}) \,.\, X(\xi,\underline{x}') \,] = 0$$

On cherche toujours une estimation de X au point $\underline{x}_o$, mais, cette fois-ci, on ne dispose que des valeurs mesurées $Y(\xi,\underline{x}_i)$, i= 1, ..P, entachées d'erreurs. L'estimateur linéaire proposé est toujours une combinaison linéaire de ces valeurs mesurées:

$$X^*(\xi,\underline{x}_o) = \sum_{i=1}^{P} \nu_i \,.\, Y(\xi,\underline{x}_i) \qquad (1\text{-}20)$$

Les critères permettant le calcul des ν_i sont les mêmes qu'au § 1-5-1-1. La condition de non-biais est automatiquement vérifiée puisqu'on suppose toujours les variables de moyenne nulle. La condition d'optimalité (1-14) conduit à la résolution du nouveau système linéaire:

$$\sum_{i=1}^{P} \nu_i \,.\, C_{YY}(\underline{x}_i, \underline{x}_j) = C_{XX}(\underline{x}_j, \underline{x}_o) \qquad j = 1,..p \qquad (1\text{-}21)$$

où, si on compare à (1-15), ce sont maintenant les covariances du processus Y, entaché d'erreur, qui interviennent à gauche, et les covariances -inconnues- du processus vrai X à droite.

On peut montrer facilement (Obled, 1987), que C_{XX} et C_{YY} sont reliées par:

$$C_{YY}(\underline{x}, \underline{x}') = C_{XX}(\underline{x}, \underline{x}') + \delta(\underline{x} - \underline{x}') \,.\, \sigma^2(\underline{x}) \qquad (1\text{-}22)$$

A partir des mesures $Y(\xi,\underline{x}_i)$, on peut donc inférer les covariances $C_{YY}(\underline{x}_i, \underline{x}_j)$, et, après modélisation, remplir le membre de gauche de (1-21) à l'aide de (1-22).

La remarque la plus intéressante dans ce cas de figure est qu'il n'y a <u>plus interpolation mais approximation</u> de la surface puisque $X^*(\xi,\underline{x}_i) \neq Y(\xi,\underline{x}_i)$. En effet, si on suppose la mesure entachée d'erreur, il n'est plus nécessaire d'imposer que la surface d'interpolation passe exactement par les points de mesure. C'est une propriété que l'on retrouvera avec l'ACPP au paragraphe suivant.

1-5-2 Liens entre l'analyse EOF et le krigeage.

Pour développer cette analogie, nous nous placerons dans le cas d'un nombre fini P de points de mesure, c'est-à-dire dans le cadre de l'approximation numérique développée au § 1-3.

1-5-2-1 *Estimateur proposé.*

Au point $\underline{x}_o$ d'intérêt, l'estimateur fourni par l'analyse EOF, noté $X_K^*(\xi,\underline{x}_o)$, est décomposé sur une base de fonctions formée des fonctions propres du noyau de covariance:

$$X_K^*(\xi,\underline{x}_o) = \sum_{k=1}^{K} Z_k(\xi) \cdot F_k^*(\underline{x}_o) \qquad (1\text{-}23)$$

avec rappelons-le:

$F_k^*(\underline{x}_o) = \sum_{i=1}^{P} f_{ik} \cdot e_i(\underline{x}_o)$ les f_{ik} ayant été obtenus comme solution de $\mathbf{C} \cdot \mathbf{E} \cdot \mathbf{F} = \mathbf{F} \cdot \Lambda$

$Z_k(\xi)$ élément de la kième colonne de la matrice $\mathbf{Z} = \mathbf{X} \cdot \mathbf{E} \cdot \mathbf{F}$ correspondant à l'événement ξ.

Il est par ailleurs important de noter que la série des P fonctions propres disponibles est tronquée à K<P éléments. Ceci implique que l'estimation n'est pas exacte, i.e., aux points de mesure $\underline{x}_i$, i= 1..P, $X_K^*(\xi,\underline{x}_i) \neq X(\xi,\underline{x}_i)$. La surface estimée passe par les points de mesure uniquement si K=P, mais, dans ces conditions, on se ramène à une interpolation purement déterministe sur la base des $e_i(\underline{x})$. (voir § suivant). Cette propriété nous conduira donc à faire une analogie avec le krigeage de données entachées d'erreurs (cf § 1-5-1-3), puis, compte-tenu de la formulation proposée, avec sa forme duale.

1-5-2-2 *Analogie entre l'ACPP et le krigeage de données entachées d'erreur.*

Dans un premier temps, nous allons voir que l'on peut passer très simplement de l'expression (1-23) de l'estimateur sur la base des fonctions propres, à une expression sous forme de pondération des valeurs mesurées donnée par:

$$X_K^*(\xi,\underline{x}_o) = \sum_{i=1}^{P} \mu_i^K . X(\xi,\underline{x}_i) \qquad (1\text{-}24)$$

ou encore sous forme matricielle:

$$X_K^*(\xi,\underline{x}_o) = [\, \mathbf{X}(\xi)\,]^T . \boldsymbol{\mu}^K$$

En effet, en utilisant les relations (1-8) et (1-10):

$$F_k^*(\underline{x}) = \sum_{i=1}^{P} f_{ik} \cdot e_i(\underline{x}) \qquad (1\text{-}8)$$

$$Z_k(\xi) = \sum_{i=1}^{P} \sum_{j=1}^{P} X(\xi,\underline{x}_i) \cdot E_{ij} \cdot f_{jk} \qquad (1\text{-}10)$$

$X_K^*(\xi,\underline{x}_o)$ s'écrit: (voir aussi Obled et Braud 1989)

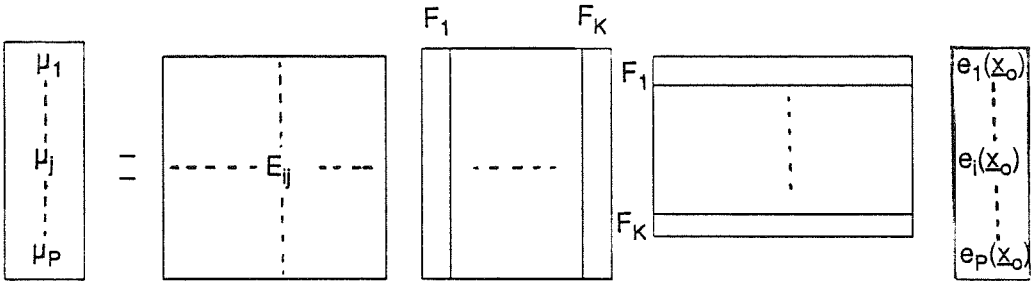
$$X_K^*(\xi,\underline{x}_o) = \sum_{k=1}^{K} \{ \sum_{i=1}^{P} \sum_{j=1}^{P} X(\xi,\underline{x}_i) .E_{ij} .f_{jk} \}.\{ \sum_{m=1}^{P} f_{mk} \cdot e_m(\underline{x}_o) \}$$

$$= \sum_{i=1}^{P} \{ \sum_{k=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{P} E_{ij} \cdot f_{jk} . f_{mk} \cdot e_m(\underline{x}_o) \} . X(\xi,\underline{x}_i)$$

$$= \sum_{i=1}^{P} \mu_i^K . X(\xi,\underline{x}_i)$$

avec $\mu_i^K = \sum_{k=1}^{K} \sum_{j=1}^{P} \sum_{m=1}^{P} E_{ij} \cdot f_{jk} . f_{mk} \cdot e_m(\underline{x}_o)$

ou encore $\mu^K = E \,.\, F_K \,.\, F_K^T \,.\, [\, e(\underline{x}_o)\,]^T$ (1-25)



si on pose $[\, e(\underline{x}_o)\,]^T = [\, e_1(\underline{x}_o), \ldots, e_P(\underline{x}_o)\,]^T$ et si on convient d'appeler F_K, la matrice (P,K) des K premiers vecteurs propres F_k, lorsque les valeurs propres correspondantes λ_k ont été classées par ordre décroissant.

<u>Remarque</u>: Si K=P, $F_P = F$ et $\mu^P = E \,.\, F \,.\, F^T \,.\, [\, e(\underline{x}_o)\,]^T$

Or, d'après (1-12), $E \,.\, F \,.\, F^T = I_P$.

Donc, si K=P, c'est-à-dire si on conserve toutes les F.P., $\mu^P = [\, e(\underline{x}_o)\,]^T$ et l'estimation devient purement déterministe. Par ailleurs, on passe alors exactement par les points de mesure puisque les $e_i(\underline{x})$ vérifient:
$e_i(\underline{x}_j) = \delta_{ij}$, i,j =1...P.

Pour avancer dans notre analogie, il nous faut maintenant montrer que le vecteur μ^K des pondérations est solution d'un système analogue à (1-16)

En effet, la covariance entre $\underline{x}_o$ et un point quelconque du réseau s'écrit, pour l'estimateur proposé par l'ACPP:

$$C_o^K(\underline{x}_i, \underline{x}_o) = E\,[\, X_K^*(\xi,\underline{x}_i) \,.\, X_K^*(\xi,\underline{x}_o)\,]$$

$$= \sum_{k=1}^{K} \lambda_k \,.\, F_k^*(\underline{x}_i) \,.\, F_k^*(\underline{x}_o)$$

$$= \sum_{k=1}^{K} \lambda_k \,.\, f_{ik} \,.\, \{ \sum_{m=1}^{P} f_{mk} \,.\, e_m(\underline{x}_o) \}$$

Donc, sous forme matricielle, on aura, en posant:

$$\mathbf{C_o}^K = [\ C_o^K(\underline{x}_1, \underline{x}_o), \ldots\ldots, C_o^K(\underline{x}_P, \underline{x}_o)\ ]^T$$

Λ_K matrice carrée de dimension K, des K premières valeurs propres λ_k

$$\mathbf{C_o}^K = \mathbf{F_K} \cdot \Lambda_K \cdot \mathbf{F_K}^T \cdot [\ e(\underline{x}_o)\ ]^T \qquad (1\text{-}26)$$

Par ailleurs, on peut écrire:

$$\mathbf{C} \cdot \mu^K = \mathbf{C} \cdot \mathbf{E} \cdot \mathbf{F_K} \cdot \mathbf{F_K}^T \cdot [\ e(\underline{x}_o)\ ]^T$$

avec $\mathbf{C} \cdot \mathbf{E} \cdot \mathbf{F_K} = \mathbf{F_K} \cdot \Lambda_K$ d'après (1-9) restreinte à $\mathbf{F_K}$, et où, rappelons-le, **C** est la matrice des covariances, utilisée pour le calcul des f_{ik}.

En utilisant (1-26), on obtient alors:

$$\mathbf{C} \cdot \mu^K = \mathbf{F_K} \cdot \Lambda_K \cdot \mathbf{F_K}^T \cdot [\ e(\underline{x}_o)\ ]^T = \mathbf{C_o}^K \qquad (1\text{-}27)$$

Comment peut-on interpréter cette relation en terme de krigeage? L'analogie formelle avec (1-16) est immédiate. Si on se rappelle maintenant que la surface estimée par ACPP ne passe pas par les valeurs mesurées, on peut rapprocher, par analogie avec le krigeage de données entachées d'erreurs, les $C_{YY}(\underline{x}_i, \underline{x}_j)$ de (1-21) et les $C(\underline{x}_i, \underline{x}_j)$ de **C** dans (1-27) d'une part, et les $C_{XX}(\underline{x}_i, \underline{x}_o)$ de (1-21) avec les $C_o^K(\underline{x}_i, \underline{x}_o)$ de (1-27) d'autre part.

En d'autres termes, les $C(\underline{x}_i, \underline{x}_j)$ sont estimées à partir des données mesurées entachées d'erreur (i.e. les $Y(\xi,\underline{x}_i)$ du § 1-5-1-3), alors que les $C_o^K(\underline{x}_i, \underline{x}_o)$ nous fournissent un modèle particulier de covariance, permettant de remplir le membre de droite du krigeage. Par ailleurs, nous associons $C_o^K(\underline{x}, \underline{x}')$ au processus vrai, puisque, ayant tronqué la série des F.P. à K termes, on peut estimer avoir filtré les erreurs de mesure. L'ACPP, utilisée en interpolation apparaît donc comme un cas particulier de krigeage, fournissant un modèle spécifique de covariance, s'appuyant sur les F.P. Ce modèle s'écrit:

$$C_o^K(\underline{x}, \underline{x}') = \sum_{k=1}^{K} \lambda_k \cdot F_k^*(\underline{x}) \cdot F_k^*(\underline{x}') \qquad (1\text{-}28)$$

Il est donc assez différent des modèles sphériques, exponentiels ou linéaires utilisés classiquement. Par ailleurs, il est intéressant de noter que la résolution du problème n'a pas

nécessité d'hypothèse de stationnarité d'ordre deux, ce qui apparaît clairement si on observe le modèle de covariance (1-28) proposé.

1-5-2-3 *Analogie avec le krigeage dual*

Nous avons vu au § 1-5-1-2, que l'on pouvait aisément passer de la formulation en pondérateur du krigeage à sa forme duale. Nous allons voir maintenant que ce passage est encore possible avec l'ACPP. Il s'agit en fait de calculer le vecteur **B** de (1-18), donc d'expliciter l'expression de la matrice $\mathbf{C}^{-1}$.

Nous savons que **C** vérifie (1-9), d'où:

$$\mathbf{C} . \mathbf{E} . \mathbf{F} = \mathbf{F} . \Lambda \quad \rightarrow \quad \mathbf{C}^{-1} . \mathbf{C} . \mathbf{E} . \mathbf{F} = \mathbf{C}^{-1} . \mathbf{F} . \Lambda$$

$$\rightarrow \quad \mathbf{E} . \mathbf{F} = \mathbf{C}^{-1} . \mathbf{F} . \Lambda$$

$$\rightarrow \quad \mathbf{E} . \mathbf{F} . \Lambda^{-1} = \mathbf{C}^{-1} . \mathbf{F}$$

$$\rightarrow \quad \mathbf{E} . \mathbf{F} . \Lambda^{-1} . \mathbf{F}^{T} . \mathbf{E} = \mathbf{C}^{-1} . \mathbf{F} . \mathbf{F}^{T} . \mathbf{E}$$

Soit $\mathbf{C}^{-1} = \mathbf{E} . \mathbf{F} . \Lambda^{-1} . \mathbf{F}^{T} . \mathbf{E}$ car $\mathbf{F} . \mathbf{F}^{T} . \mathbf{E} = \mathbf{I}_p$ d'après (1-12)

Nous pouvons donc écrire la forme duale de (1-24):

$$X_K^*(\xi, \underline{x}_o) = \mathbf{B}^T . \mathbf{C}_o^K$$

avec $$\mathbf{B}^T = [\, \mathbf{X}(\xi) \,]^T . \mathbf{C}^{-1} = [\, \mathbf{X}(\xi) \,]^T . \mathbf{E} . \mathbf{F} . \Lambda^{-1} . \mathbf{F}^T . \mathbf{E} = [b_1, \ldots, b_p]^T \qquad (1\text{-}29)$$

ou encore en développant:

$$X_K^*(\xi, \underline{x}_o) = \sum_{i=1}^{p} b_i . C_o^K(\underline{x}_i, \underline{x}_o) \qquad (1\text{-}30)$$

Notons que dans cette dernière expression les coefficients b_i ne dépendent pas de la troncature de la série des F.P., c'est-à-dire de K, contrairement aux μ_i^K du paragraphe précédent. Ils sont uniquement déterminés par la réalisation considérée et, cette dernière approche, une fois le modèle de covariance identifié, peut permettre une interpolation plus rapide puisque le calcul des b_i se fait simplement par un produit matriciel (cf (1-29)) où il faut introduire les valeurs mesurées pour cette réalisation.

1-6 CONCLUSION

Dans ce premier chapitre, nous avons tenté de faire une présentation aussi exhaustive que possible de l'analyse EOF, en en détaillant, bien sûr, les fondements mathématiques et les étapes permettant de passer à une implémentation sur ordinateur. Dans un deuxième temps, nous nous sommes attachés à montrer que l'ACPP n'était pas une technique "isolée", mais qu'on pouvait la rapprocher de disciplines aussi diverses que la recherche de modes normaux de systèmes dynamiques et les techniques géostatistiques d'interpolation optimale.

Néanmoins, jusqu'à présent, la précision de l'approximation numérique présentée au § 1-3, n'a pas été testée de façon systématique, pour le moins à deux dimensions, Bouhaddou (1984), ayant quant à lui traité le cas monodimensionnel. Or, si on veut utiliser l'ACPP pour approximer des surfaces ou interpréter les EOFs en termes de physique, il semble naturel de connaître quelles sont les erreurs qui les affectent. Nous essaierons de répondre à cette question au chapitre 2, en comparant des solutions analytiques du problème intégral à leurs estimations numériques.

CHAPITRE 2 : COMPARAISON ENTRE FONCTIONS PROPRES D'UN NOYAU DE COVARIANCE ET LEURS APPROXIMATIONS NUMERIQUES.

2-1 POURQUOI CETTE PREOCCUPATION?

Nous avons déjà souligné que l'interprétation des EOFs en liaison avec la physique sous-jacente à la variable étudiée était pratique courante. C'est ainsi que dans l'ACP de champs climatologiques, on cherche à retrouver, dans la forme des EOFs, des circulations météorologiques dominantes, voire des "types de temps". Assez fréquemment, on arrive à reconnaître, dans le plan des premières EOFs, la carte géographique du domaine étudié, mais affectée de distorsions que l'on cherche à interpréter. Certes, les avis sont partagés sur l'opportunité d'une telle interprétation (Von Storch et Hannoschöck, 1985, par exemple). Cette méfiance à l'égard des résultats bruts de l'analyse EOF est d'ailleurs confirmée par la somme d'énergie dépensée pour mettre au point des techniques de rotations des composantes principales, afin de les rendre.... interprétables en Climatologie. (On trouvera chez Richman (1986) une revue de toutes ces méthodes).

Mais, quelle confiance accorder à tous ces résultats si on n'a aucune idée de la précision des calculs, sachant que des approximations sont faites lors du passage de la formulation intégrale à la solution numérique (cf § 1-3)? La forme des EOFs n'est-elle pas uniquement une conséquence des contraintes d'orthogonalité imposées?

Pour ce faire, deux outils privilégiés sont disponibles. Le premier est l'analyse mathématique, rigoureuse mais complexe, surtout à plus d'une dimension, et dans laquelle il est difficile de prendre en compte à la fois les aspects fonctionnels et les aspects statistiques, en particulier l'échantillonnage. Le second est la simulation stochastique, qui consiste à examiner expérimentalement le rôle des diverses conditions (choix du domaine, du noyau de covariance, de l'échantillonnage, etc...) au risque de ne pouvoir en donner une vue synthétique.

Une autre technique, assez ancienne et beaucoup plus simple à mettre en oeuvre, consiste à séparer l'échantillon disponible en plusieurs sous-échantillons et à analyser la stabilité des résultats obtenus. On peut ainsi mettre en évidence à partir de quand une EOF estimée est trop entachée d'erreurs d'échantillonnage pour être significative.

Plus récemment, Kueny (1977), partant de données sur grille régulière (champs de pression atmosphérique) a ajusté à la fonction de corrélation empirique des sommes de cosinus, puis a cherché, analytiquement, les solutions matricielles exactes sur la grille

utilisée. On peut considérer, qu'à notre connaissance, c'était la première tentative en France d'aborder la recherche de solutions exactes. Malheureusement, le fait d'inclure à la fois l'aspect analytique (choix d'un noyau de corrélation, qui n'était d'ailleurs pas simple) et l'aspect numérique (recherche d'une solution discrète sur une grille choisie a priori) rend les solutions extrêmement complexes, bien que très réalistes graphiquement. Cependant le rôle du pas de la grille n'est pas perçu directement. Une approche voisine à une dimension a été proposée par Brillinger (1975), mais à pas de temps discret, fixé là aussi, a priori.

Abordant le problème par la simulation, Obled (1979) montre, à une dimension d'abord, le rôle modeste du modèle de covariance choisi, mais le rôle déterminant de la taille du domaine par rapport à la longueur de corrélation, et aussi celui du pas de grille et de sa régularité. Il esquisse aussi le lien avec l'analyse harmonique classique. Quelques essais à deux dimensions confirment ces résultats, mais l'auteur se consacre ensuite aux analogies de l'ACP avec l'interpolation optimale. Parallèlement, Buell (1975, 1978, 1979), aux Etats-Unis, fait des essais similaires, à deux dimensions, surtout axés sur l'approximation numérique.

Ce n'est que tardivement que l'on prend connaissance, sous la forme d' articles très condensés, des travaux menés en URSS par Fortus (1973, 1975). Il propose, mais sans décrire entièrement les moyens de l'obtenir, des solutions analytiques au problème intégral (1-4), respectivement sur un segment de R et un cercle de R^2. Ces recherches, à une dimension, sont reprises par Bouhaddou (1984), Bouhaddou et al. (1987). Elles mettent en évidence le rôle des frontières, puisque le problème intégral est transformé en système d'équations différentielles avec conditions aux limites. La solution analytique à une dimension a été exploitée par Bouhaddou (1984) pour évaluer la précision du calcul des F.P., mais plutôt dans l'optique d'une utilisation en interpolation.

Pour notre part, nous avons repris la même démarche, mais à deux dimensions cette fois, pour aboutir à une solution analytique qui puisse être comparée avec la solution numérique déduite du § 1-3. Le noyau de covariance utilisé est le même que celui étudié par Fortus (1975). Les résultats de cette comparaison sont décrits dans la seconde partie de ce chapitre, en faisant bien la distinction entre les erreurs imputables à la discrétisation spatiale décrite en 1-3 et à l'échantillonnage statistique qui influence l'estimation de la matrice de covariance. La première partie est, quant à elle, consacrée à l'obtention de la solution analytique proprement dite.

A ce niveau, soulignons que, de son côté, le problème de l'échantillonnage statistique a déjà fait l'objet de plusieurs publications récentes. Par exemple, North et al. (1982) proposent des intervalles de confiance sur les valeurs propres calculées à l'aide d'un N-échantillon et montrent que la précision des fonctions propres est d'autant moins grande que les valeurs propres sont proches. Von Storch et Hannoschöck (1985) utilisent la

simulation de Monte-Carlo et les techniques de Jacknife pour déduire un estimateur non biaisé des valeurs propres. Néanmoins, toutes ces études ont un point commun: elles soulignent combien les incertitudes sur les EOFs sont grandes, ce que ne viendront pas démentir les résultats que nous présenterons par la suite.

2-2 OBTENTION DES FONCTIONS PROPRES ANALYTIQUES D'UN NOYAU DE COVARIANCE A DEUX DIMENSIONS.

2-2-1 Cadre de la solution analytique.

Nous considérons donc un processus aléatoire $X(\xi,\underline{x})$, $\underline{x}$ étant le vecteur des coordonnées (x,y) de R^2, défini sur un domaine borné D de R^2. Nous nous donnons la fonction de covariance $C(\underline{x},\underline{x}')$ et nous voulons obtenir les fonctions $F(\underline{x})$ solutions de (1-4) sur D. Ces fonctions seront supposées nulles hors de D, d'où le nouveau système:

$$\begin{cases} \int_D C(\underline{x},\underline{x}').F(\underline{x}').d\underline{x}' = \lambda \,.\, F(\underline{x}) & \underline{x} \in D \qquad (2\text{-}1) \\ F(\underline{x}) = 0 & \underline{x} \notin D \qquad (2\text{-}2) \end{cases}$$

Il est, bien sûr, impossible de résoudre ce problème dans le cas général. En revanche (Fortus, 1975), si on considère une fonction de covariance:

i) stationnaire : $C(\underline{x},\underline{x}') = C(\underline{x} - \underline{x}') = C(\underline{h})$
ii) isotrope,: $C(\underline{h}) = C(r)$ où $r = ||\underline{h}||$

iii) à densité spectrale $S(\underline{w})$ rationnelle (i.e. $S(\underline{w})$, transformée de Fourier inverse de $C(\underline{x},\underline{x}')$, peut s'écrire comme le rapport de deux polynômes)

iv) le domaine D étant un cercle de R^2

on peut en obtenir une solution analytique.

C'est dans ce cadre que nous nous placerons et, après avoir présenté la méthode générale de calcul, nous l'appliquerons, sur un domaine circulaire, au modèle de covariance traité par Fortus (1975).

$$C(r) = 1/Ln(\alpha/\beta).\,[\,K_0(\beta.r) - K_0(\alpha.r)\,] \qquad \alpha>\beta \qquad (2\text{-}3)$$

où $K_0(x)$ est une fonction de Bessel modifiée d'ordre zéro.

La Fig 2-1 présente ce modèle, utilisé par Fortus pour des champs de géopotentiels, pour plusieurs valeurs du couple (α, β) des paramètres.

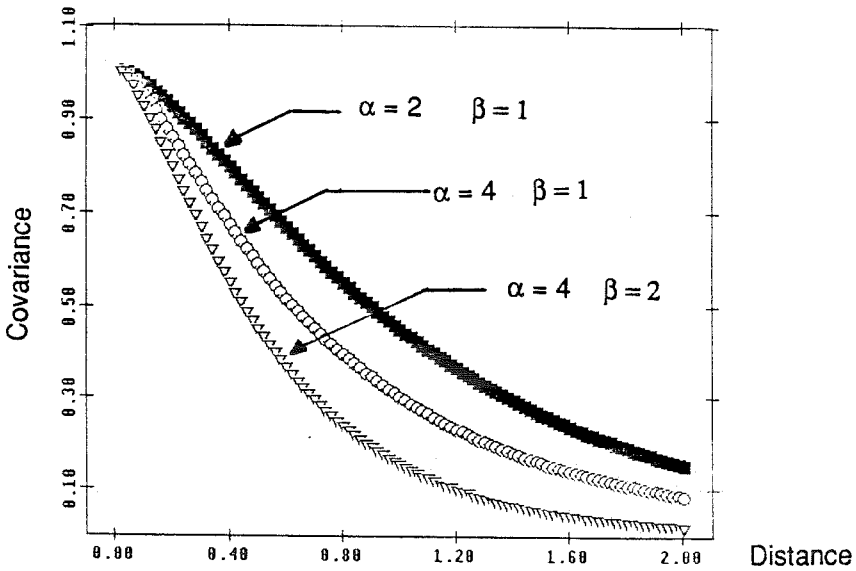


Fig 2-1: Modèle de covariance proposé par Fortus (1975):
$C(r) = 1/Ln(\alpha/\beta).\ [\ K_0(\ \beta.r\ ) - K_0(\ \alpha.r)\ ]$

a) α=2, β=1, b) α=4, β=1, c) α=4, β=2.

Avant d'entrer dans "le vif du sujet", nous allons préciser la définition de la densité spectrale d'une fonction de covariance, les notations utilisées pour les transformées de Fourier directe et inverse, ainsi que deux propriétés qui nous seront utiles par la suite.

Si une fonction de covariance stationnaire $C(\underline{h})$ est continue, absolument intégrable, et à décroissance rapide, le théorème de Wiener-Kichnine permet de l'exprimer comme la transformée de Fourier d'une fonction $S(\underline{w})$, appelée sa densité spectrale (Papoulis, 1962).

i) *Transformée de Fourier directe* **T**

$$C(\underline{h}) = T[\ S(\underline{w})](\underline{h}) = \int_{R2} S(\underline{w})\ .\ \exp(i\underline{h}.\underline{w})\ .\ d\underline{w} \qquad (2\text{-}4)$$

Si on pose, à deux dimensions, $\underline{h}=(h_1,h_2)$, $\underline{w}=(w_1,w_2)$, le produit scalaire des deux vecteurs s'écrit $\underline{h}\ .\ \underline{w} = h_1.w_1 + h_2\ .w_2$ et $C(\underline{h})$ s'écrit complètement:

$$C(\underline{h}) = C(h_1,h_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(w_1,w_2)\ .\ \exp(\ i\ (h_1.w_1 + h_2\ .w_2))\ .\ dw_1\ .\ dw_2$$

ii) *Transformée de Fourier inverse* T^{-1}

Réciproquement, $S(\underline{w})$ peut s'exprimer à l'aide de $C(\underline{h})$ par l'intermédiaire de la transformée de Fourier inverse:

$$S(\underline{w}) = T^{-1}[\, C(\underline{h})\,](\underline{w}) = 1/(2\Pi)^2 \int_{R2} C(\underline{h}) \,.\, \exp(-i\underline{h}.\underline{w}) \,.\, d\underline{h} \qquad (2\text{-}5)$$

Si, de plus $C(\underline{h})$ est <u>isotrope</u>, les relations (2-4) et (2-5) peuvent se simplifier (Mantoglou et Wilson, 1981):

$$C(r) = 2\Pi. \int_0^{\infty} J_0(wr) \,.\, S(w) \,.\, dw \qquad (2\text{-}6)$$

$$S(w) = 1/2\Pi. \int_0^{\infty} J_0(wr) \,.\, C(r) \,.\, dr \qquad (2\text{-}7)$$

où $r = (h_1^2 + h_2^2)^{1/2}$ et $w = (w_1^2 + w_2^2)^{1/2}$ sont les modules des vecteurs $\underline{h}$ et $\underline{w}$ et J_0 une fonction de Bessel de première espèce d'ordre 0.

iii) *Propriété 1*

Soit $f(\underline{x})$ une fonction; $g(\underline{w}) = T[\, f(\underline{x})\,](\underline{w})$, sa transformée de Fourier directe.

Si $f(\underline{x})$ est de carré sommable (c'est-à-dire de carré intégrable), $g(\underline{w})$ l'est aussi. Par ailleurs, si $\underline{x}^k \,.\, f(\underline{x})$ est sommable, $g(\underline{w})$ sera k fois dérivable. (Matheron, 1965).

iv) *Propriété 2* (Matheron, 1965, p. 275)

Si on pose $\underline{x} = (x_1, x_2)$, $\underline{w} = (w_1, w_2)$ et si m1, m2 , m sont trois entiers tels que m1+m2 = m

$$T^{-1}\left[\frac{\partial^m f(\underline{x})}{\partial x_1^{m1}\ \partial x_2^{m2}} \right](\underline{w}) = i^m \,.\, w_1^{m1} \,.\, w_2^{m2} \,.\, T^{-1}[\, f(\underline{x})](\underline{w}) \qquad (2\text{-}8)$$

En particulier, pour $w^{2k} = (w_1^2 + w_2^2)^k$, nous aurons, compte-tenu de la linéarité des transformées de Fourier directe et inverse:

$$A = (w_1^2 + w_2^2)^k \,.\, T^{-1}[\, f(\underline{x})\,](\underline{w}) = \sum_{p=0}^{k} C_p^k \,.\, w_1^{2p} \,.\, w_2^{2(k-p)} \,.\, T^{-1}[\, f(\underline{x})\,](\underline{w})$$

en utilisant la formule du binôme de Newton. Puis, grâce à (2-8), on peut écrire:

$$A = T^{-1}\left[(-1)^k \sum_{p=0}^{k} C_p^k \frac{\partial^{2k} f(\underline{x})}{\partial x_1^{2p}\ \partial x_2^{2(k-p)}} \right](\underline{w}) \qquad (2\text{-}9)$$

Si nous introduisons maintenant l'opérateur Laplacien Δ, défini par:

$$\Delta f(\underline{x}) = \frac{\partial^2 f(\underline{x})}{\partial x_1^2} + \frac{\partial^2 f(\underline{x})}{\partial x_2^2}$$

l'opérateur Δ^k, qui est l'opérateur Laplacien appliqué k fois, peut s'écrire:

$$\Delta^k f(\underline{x}) = \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} \right)^{(k)} f(\underline{x})$$

$$= \sum_{p=0}^{k} C_p^k \frac{\partial^{2k} f(\underline{x})}{\partial x_1^{2p}\ \partial x_2^{2(k-p)}}$$

Nous en déduisons, pour (2-9):

$$w^{2k} \,.\, T^{-1}[\, f(\underline{x})](\underline{w}) = T^{-1}[\, (-\Delta)^{(k)} f(\underline{x})](\underline{w}) \qquad (2\text{-}10)$$

où encore en prenant la transformée de Fourier de (2-10):

$$T\left[\, w^{2k} \,.\, T^{-1}[\, f(\underline{x})](\underline{w}) \,\right](\underline{x}) = (-\Delta)^{(k)} f(\underline{x}) \qquad (2\text{-}11)$$

2-2-2 <u>Méthode générale de résolution</u>.

i) On transforme tout d'abord le membre de gauche de (2-1) en utilisant la <u>stationnarité</u> de $C(\underline{x},\underline{x}')$, i.e. $C(\underline{x},\underline{x}') = C(\underline{x} - \underline{x}')$ puis <u>sa définition en fonction de $S(\underline{w})$</u>,i.e. (2-4), ce qui donne:

$$\int_D C(\underline{x} - \underline{x}').F(\underline{x}').d\underline{x}' = (2\Pi)^2 \,.\, T\left[\, S(\underline{w}) \,.\, T^{-1}[\, F(\underline{x}) \,](\underline{w}) \,\right](\underline{x}) \qquad (2\text{-}12)$$

Démonstration:

$$\Psi(\underline{x}) = \int_D C(\underline{x} - \underline{x}').F(\underline{x}').d\underline{x}' = \int_D [\int_{R2} S(\underline{w}) . \exp(i(\underline{x}-\underline{x}').\underline{w}) . d\underline{w}]. F(\underline{x}') . d\underline{x}'$$

(utilisation de la formule (2-4))

$$\Psi(\underline{x}) = \int_{R2} S(\underline{w}) . \exp(i\underline{x}.\underline{w}) . \{ \int_D \exp(- i\underline{x}'.\underline{w}) . F(\underline{x}') . d\underline{x}' \} . d\underline{w}$$

Nous admettons ici qu'on peut échanger les deux signes $\int_D$ et $\int_{R2}$. Par ailleurs, puisque $F(\underline{x}') = 0$ en dehors du domaine D (formule (2-2)), on peut étendre la seconde intégrale à R^2 . On reconnaît alors, à $(2\Pi)^2$ près, compte-tenu de la définition (2-5), la transformée de Fourier inverse de $F(\underline{x})$: $T^{-1}[F(\underline{x})](\underline{w})$. (Nous supposons ici que $F(\underline{x})$ est de carré intégrable pour que (1-4) soit définie et que $T[F(\underline{x})](\underline{w})$ existe.) Donc:

$$\Psi(\underline{x}) = (2\Pi)^2 . \int_{R2} S(\underline{w}) . T^{-1}[F(\underline{x})](\underline{w}) . \exp(i\underline{x}.\underline{w}) . d\underline{w}$$

Etant donnée la définition (2-4) de la transformée de Fourier, on reconnaît, cette fois,la transformée de Fourier directe du produit $S(\underline{w}) . T^{-1}[F(\underline{x})](\underline{w})$. D'où finalement:

$$\Psi(\underline{x}) = (2\Pi)^2 . T \left[S(\underline{w}) . T^{-1}[F(\underline{x})](\underline{w}) \right](\underline{x})$$

ce qui est bien le résultat (2-12) annoncé.

ii) Pour progresser, nous allons maintenant utiliser <u>l'isotropie</u> du modèle de covariance qui induit une densité spectrale $S(\underline{w})$ ne dépendant que du module w de $\underline{w}$ (formules (2-6) et (2-7))

Nous supposons de plus que $S(\underline{w})$ est <u>rationnelle</u>, c'est-à-dire qu'elle peut s'écrire comme le rapport de deux polynômes en w (Bouhaddou, 1984)

$$S(\underline{w}) = P(w^2) / Q(w^2) \qquad \text{avec } w = || \underline{w} ||$$

où P et Q sont, respectivement, des polynômes de degrés 2p et 2q en w (avec p<q).

L'existence d'une telle décomposition nous permet alors un changement de fonction inconnue. Nous poserons en effet:

$$\phi(\underline{x}) = T \left[T^{-1}[F(\underline{x})](\underline{w}) / Q(w^2) \right](\underline{x}) \qquad (2\text{-}13)$$

L'intérêt d'une telle transformation est le suivant: Q est un polynôme de degré 2q en w, $T^{-1}[\, F(\underline{x})\,](\underline{w})$ est sommable (donc bornée) car $F(\underline{x})$ est de carré sommable, donc

$$\forall\, k,\ k<2q,\quad w^k \,.\, T^{-1}[\, F(\underline{x})\,](\underline{w}) \,/\, Q(w^2) \text{ est aussi sommable.}$$

(elle sera en $1/w^{2q-k}$ au voisinage de l'infini, ce qui assure la convergence des intégrales)

La propriété 1 du § 2-1, nous permet alors d'en déduire que $\phi(\underline{x})$ est <u>2q-1 dérivable</u>.

En prenant la transformée de Fourier directe de (2-13), on obtient:

$$T^{-1}[\, F(\underline{x})\,](\underline{w}) \,/\, Q(w^2) = T^{-1}\,[\, \phi(\underline{x})\,](\underline{w})$$

En utilisant cette expression, (2-12) s'écrit:

$$\begin{cases} \int_D C(\underline{x}-\underline{x}').F(\underline{x}').d\underline{x}' = (2\Pi)^2 \,.\, T\,\big[\, P(w^2) \,.\, T^{-1}\,[\, \phi(\underline{x})\,](\underline{w})\big](\underline{x}) & (2\text{-}14) \\ Q(w^2) \,.\, T^{-1}\,[\, \phi(\underline{x})\,](\underline{w}) = T^{-1}[\, F(\underline{x})\,](\underline{w}) & (2\text{-}15) \end{cases}$$

avec

Nous allons maintenant transformer ce système grâce à la propriété 2 rappelée plus haut. Ecrivons le polynôme au numérateur de la densité spectrale sous la forme:

$$P(w^2) = \sum_{k=0}^{p} a_k \,.\, w^{2k}$$

En utilisant la linéarité de la transformée de Fourier et la propriété (2-11) pour chaque monôme w^{2k}, le membre de droite de (2-14) peut s'écrire $(2\Pi)^2 \,.\, P(-\Delta)\,\phi(\underline{x})$ où la notation $P(-\Delta)\,\phi(\underline{x})$ se traduit par:

$$P(-\Delta)\,\phi(\underline{x}) = \sum_{k=0}^{p} a_k \,.\, (-1)^k \,.\, \Delta^k\phi(\underline{x})$$

Avec les mêmes notations et la formule (2-10), le membre de gauche de (2-15) est transformé en $T^{-1}\,[\, Q(-\Delta)\,\phi(\underline{x})\,](\underline{w})$. Donc le système (2-14), (2-15) devient:

$$\begin{cases} \int_D C(\underline{x}-\underline{x}').F(\underline{x}').d\underline{x}' = (2\Pi)^2 \,.\, P(-\Delta)\,\phi(\underline{x}) & (2\text{-}16) \\ T^{-1}\,[\, Q(-\Delta)\,\phi(\underline{x})\,](\underline{w}) = T^{-1}\,[\, F(\underline{x})\,](\underline{w}) & (2\text{-}17) \end{cases}$$

En prenant ensuite la transformée de Fourier de (2-17), il s'écrira:

$$\begin{cases} \int_D C(\underline{x} - \underline{x}').F(\underline{x}').d\underline{x}' = (2\Pi)^2 \,.\, P(-\Delta)\,\phi(\underline{x}) & (2\text{-}18) \\ Q(-\Delta)\,\phi(\underline{x}) = F(\underline{x}) & (2\text{-}19) \end{cases}$$

Revenons maintenant à notre problème initial constitué des équations (2-1) et (2-2). Appelons $\phi^a(\underline{x})$ et $\phi^b(\underline{x})$ les solutions, respectivement à l'intérieur et à l'extérieur de D et $\phi(\underline{x})$ la fonction égale à $\phi^a(\underline{x})$ si $\underline{x} \in D$ et à $\phi^b(\underline{x})$ si $\underline{x} \notin D$. En utilisant (2-18) et (2-19), ce système s'écrit:

$$\begin{cases} (2\Pi)^2 \,.\, P(-\Delta)\,\phi^a(\underline{x}) = \lambda \,.\, Q(-\Delta)\,\phi^a(\underline{x}) & \underline{x} \in D & (2\text{-}20) \\ Q(-\Delta)\,\phi^b(\underline{x}) = 0 & \underline{x} \notin D & (2\text{-}21) \\ F(\underline{x}) = Q(-\Delta)\,\phi(\underline{x}) & & (2\text{-}22) \end{cases}$$

<u>Nous avons donc transformé le problème intégral initial en un système d'équations aux dérivées partielles d'ordre 2q avec conditions aux limites</u>.

Ces dernières sont en effet déduites des propriétés de dérivabilité de $\phi(\underline{x})$ et, pour tout point $\underline{x}_F$ de la frontière Γ_D, on devra avoir, pour toutes les dérivées jusqu'à l'ordre 2q-1:

$$\phi^{a(k)}(\underline{x}_F) = \phi^{b(k)}(\underline{x}_F) \qquad \underline{x}_F \in \Gamma_D,\ k=0, \ldots\ldots, 2q-1 \qquad (2\text{-}23)$$

La démarche suivie pour l'obtention de la solution complète sera alors la suivante:

i) Résolution de (2-21) à l'extérieur de D. (On demandera à cette solution $\phi^b(\underline{x})$ d'être de carré intégrable et donc bornée à l'infini.)

ii) Recherche de la forme de la solution $\phi^a(\underline{x})$ à l'intérieur de D ((2-20)). Comme q>p, nous aurons alors 2q constantes à déterminer.

iii) Obtention des constantes par application des conditions aux limites (2-23)

iv) Calcul des fonctions propres $F(\underline{x})$ par (2-22), puis normalisation de façon à vérifier (1-4):

$$\int_D F^2(\underline{x}) \,.\, d\underline{x} = 1$$

Voyons maintenant ce que cette démarche donne concrètement dans le cas du modèle de covariance (2-3) proposé par Fortus (1975), lorsque le domaine est un cercle de rayon R.

2-2-3 Calcul complet des fonctions propres analytiques du modèle de covariance proposé par Fortus (1975).

Si $C(r) = A \cdot [K_0(\beta.r) - K_0(\alpha.r)]$ $\alpha>\beta$,

avec $A = 1/Ln(\alpha/\beta)$ pour que $C(0) = 1$, sa densité spectrale est donnée par : (Fortus, 1975)

$$S(w) = \frac{(\alpha^2 - \beta^2) \cdot A}{2\Pi} \cdot \frac{1}{(\alpha^2 + w^2) \cdot (\beta^2 + w^2)}$$

ce qui conduit à: $\begin{cases} P(w^2) = ((\alpha^2 - \beta^2) \cdot A)/2\Pi = cte \\ Q(w^2) = (\alpha^2 + w^2) \cdot (\beta^2 + w^2) \end{cases}$

Les équations (2-20) et (2-21) s'écrivent alors:

$$\begin{cases} 2\Pi. A \cdot (\alpha^2 - \beta^2)\, \phi^a(\underline{x}) = \lambda \cdot (\Delta - \alpha^2) \cdot (\Delta - \beta^2)\, \phi^a(\underline{x}) & \underline{x} \in D \quad (2\text{-}24) \\ (\Delta - \alpha^2) \cdot (\Delta - \beta^2)\, \phi^b(\underline{x}) = 0 & \underline{x} \notin D \quad (2\text{-}25) \end{cases}$$

A ce stade, le problème reste encore très complexe. Mais, en utilisant l'isotropie du domaine d'étude D (qui rappelons-le est un cercle de rayon R), on peut passer en coordonnées polaires $\underline{x} = (r, \theta)$ et rechercher des solutions de la forme:

$$\phi^l(\underline{x}) = \phi^l(r, \theta) = \phi_m{}^l(r) \cdot \exp(im\theta) \qquad l=a,b,\ m \in N \qquad (2\text{-}26)$$

On peut alors montrer, en admettant la régularité de ϕ^a en zéro et que ϕ^b est de carré intégrable, que la solution à l'intérieur (resp. à l'extérieur) du cercle est de la forme: (Annexes D1,D2,D3)

$$\begin{cases} \phi_m{}^a(r, \theta) = [\, C_1{}^a \cdot J_m(\chi r) + C_2{}^a \cdot I_m(\kappa r)\,] \cdot \exp(im\theta) & (2\text{-}27) \\ \phi_m{}^b(r, \theta) = [\, C_1{}^b \cdot K_m(\alpha r) + C_2{}^b \cdot K_m(\beta r)\,] \cdot \exp(im\theta) & (2\text{-}28) \end{cases}$$

J_m, I_m, K_m sont des fonctions de Bessel ou de Bessel modifiées d'ordre m, m appartenant à $[0, +\infty[$. χ et κ sont deux paramètres, dépendants de m et de R qu'il faudra déterminer, sachant qu'ils vérifient la relation :

$$(\alpha^2 + \chi^2) . (\beta^2 + \chi^2) = (\alpha^2 - \kappa^2) . (\beta^2 - \kappa^2) \qquad (2\text{-}29)$$

qui s'écrit encore: (Annexe D)

$$S(\chi) = S(i\kappa) \qquad (2\text{-}30)$$

Cette relation, unique, n'est évidemment pas suffisante pour calculer les deux inconnues. Nous verrons comment le calcul des constantes C_i^j, i=1,2, j=a,b, nous permet d'en obtenir une seconde. Notons enfin (Annexe D) que (2-30) nous donne aussi les valeurs propres puisque:

$$\lambda = (2\Pi)^2 . S(\chi) = (2\Pi)^2 . S(i\kappa) \qquad (2\text{-}31)$$

<u>Remarque</u>: Fortus (1975) a montré que $\lambda < (2\Pi)^2 . S(w_o)$ où w_o correspond au maximum absolu de S(w), de sorte que χ et κ existent toujours. (voir aussi l'Annexe D).

Ayant la forme de la solution, il nous faut maintenant calculer les constantes C_i^j, i=1,2, j=a,b. Pour cela, on pourrait, comme ϕ est trois fois dérivable (cf § 2-2-2), écrire les relations (2-23) à la frontière, ce qui nous donnerait un système linéaire de <u>quatre équations à quatre inconnues</u> C_1^a, C_2^a, C_1^b, C_2^b. Mais, en pratique, seule la solution ϕ^a à l'intérieur du cercle nous intéresse (puisqu'on sait que $F(\underline{x}) = Q(-\Delta)\phi(\underline{x})$ est nulle en dehors du cercle) et donc, seul le calcul de C_1^a et C_2^a est nécessaire. Nous allons donc rechercher une solution "moins lourde" en essayant de trouver deux opérateurs différentiels D_1 et D_2, de degrés inférieurs à 2q-1=3, tels que:

$$D_1 \phi^b = D_2 \phi^b = 0 \qquad \text{en dehors du cercle.}$$

Compte tenu de la régularité de ϕ, à la frontière, c'est-à-dire pour r =R (voir (2-23)):

$$D_1 \phi^a(R, \theta) = D_2 \phi^a(R, \theta) = 0$$

Ceci nous fournira un système de <u>deux équations à deux inconnues</u> C_1^a et C_2^a beaucoup plus facile à résoudre que le système à quatre inconnues initial.

On montre (cf Annexe D4) que les opérateurs différentiels suivants conviennent:

$$\begin{cases} D_1 = \{ \beta \,.\, K_{m-1}(\beta r) + K_m(\beta r) \,[\, d/dr + m/r \,] \,\} \,(B_m - \alpha^2) & (2\text{-}32) \\ D_2 = \{ \alpha \,.\, K_{m-1}(\alpha r) + K_m(\alpha r) \,[\, d/dr + m/r \,] \,\} \,(B_m - \beta^2) & (2\text{-}33) \end{cases}$$

où B_m est l'opérateur différentiel :

$$B_m = d^2/dr^2 + 1/r \,.\, d/dr - m^2/r^2$$

Le système linéaire en $C_1{}^a$ et $C_2{}^a$ est alors donné par (Annexe D4)

$$\begin{cases} d_1(\alpha, \chi) \,.\, C_1{}^a + d_2(\alpha, \kappa) \,.\, C_2{}^a = 0 & (2\text{-}34) \\ d_1(\beta, \chi) \,.\, C_1{}^a + d_2(\beta, \kappa) \,.\, C_2{}^a = 0 & (2\text{-}35) \end{cases}$$

avec

$$\begin{cases} d_1(\gamma, \chi) = [\, \gamma \,.J_m(\chi R) \,.\, K_{m-1}(\gamma R) + \chi \,.\, J_{m-1}(\chi R) \,.\, K_m(\gamma R) \,] \,/\, (\chi^2 + \gamma^2) & (2\text{-}36) \\ d_2(\gamma, \kappa) = [\, \gamma \,.\, I_m(\kappa R) \,.\, K_{m-1}(\gamma R) + \kappa \,.\, I_{m-1}(\kappa R) \,.\, K_m(\gamma R) \,] \,/\, (\kappa^2 - \gamma^2) & (2\text{-}37) \end{cases}$$

Pour que le système formé par les équations (2-34) et (2-35) n'ait pas qu'une solution identiquement nulle, il faut et il suffit que son déterminant soit nul, c'est-à-dire que:

$$d_1(\alpha, \chi) \,.\, d_2(\beta, \kappa) - d_1(\beta, \chi) \,.\, d_2(\alpha, \kappa) = 0 \qquad (2\text{-}38)$$

Cette relation, ajoutée à (2-30): $S(\chi) = S(i\,\kappa)$, nous permet de calculer χ et κ puisque nous avons deux relations et deux inconnues. En fait, pour une valeur de m donnée, ce système admet une <u>infinité de solutions</u>. Chaque couple (χ , κ) peut alors être associé à une valeur propre donnée par (2-31) et à une fonction $\phi^a(r, \theta)$ donnée par (2-27). La fonction propre correspondante $F(r, \theta)$ est calculée par (2-22). Mais, si on remarque que, dans notre cas particulier, $P(w^2)$ est une constante, on voit que $F(r, \theta)$ se déduit de $\phi^a(r, \theta)$ par simple multiplication par une constante:

$$F(r,\theta) = 2\Pi.\, A \,.\, (\alpha^2 - \beta^2) \,/\, \lambda \,.\, \phi^a(r, \theta) = (\, \alpha^2 + \chi^2) \,.\, (\, \beta^2 + \chi^2 \,) \,.\, \phi^a(r, \theta)$$

en utilisant (2-30)

Il reste encore à calculer les constantes C_1^a et C_2^a de (2-27). La relation (2-34) nous permet d'exprimer C_2^a en fonction de C_1^a:

$$C_2^a = d_1(\alpha, \chi) / d_2(\alpha, \kappa) \,.\, C_1^a$$

ce qui donne pour $F(r,\theta)$:

$$F(r,\theta) = C_1^a.(\alpha^2 + \chi^2).(\beta^2 + \chi^2).[\, J_m(\chi r) + d_1(\alpha, \chi) / d_2(\alpha, \kappa) \,.\, I_m(\kappa r)\,] \,.\, \exp(im\theta) \qquad (2\text{-}39)$$

C_1^a est ensuite évaluée en utilisant la condition de normalisation:

$$\int_0^{2\Pi} \int_0^R F^2(r,\theta) \,.\, r \,.\, dr \,.\, d\theta = 1 \qquad \text{(relation (1-4))}$$

Ce calcul est décrit complètement dans l'Annexe D5.

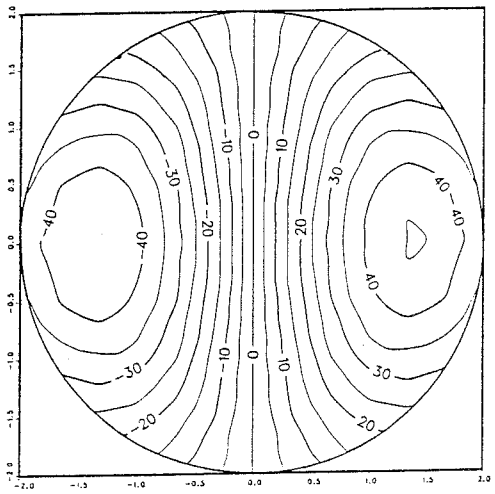
Revenons maintenant un peu sur les valeurs propres. Nous avons dit plus haut que, pour chaque valeur de m, on obtenait une infinité de valeurs propres. Pour se rattacher aux pratiques habituelles, on doit classer toutes les solutions obtenues par ordre décroissant des valeurs propres. Or, sauf pour m=0, où la valeur propre est simple, <u>toutes les valeurs propres sont doubles</u> (elles correspondent respectivement à $\cos(m\theta)$ et $\sin(m\theta)$ dans (2-39) et les fonctions propres correspondantes se déduiront l'une de l'autre par simple rotation d'angle $\pi/2$). A cause du classement effectué, on verra donc s'intercaler quelques valeurs propres simples, issues de m=0, entre des valeurs propres doubles. La Fig. 2-2 l'illustre pour la F.P. 15 et présente la projection, dans le plan (r,θ), de quelques fonctions propres lorsque $R=2$, $\alpha=2$, $\beta=1$.

2-2-4 <u>Extension à d'autres modèles de covariance</u>.

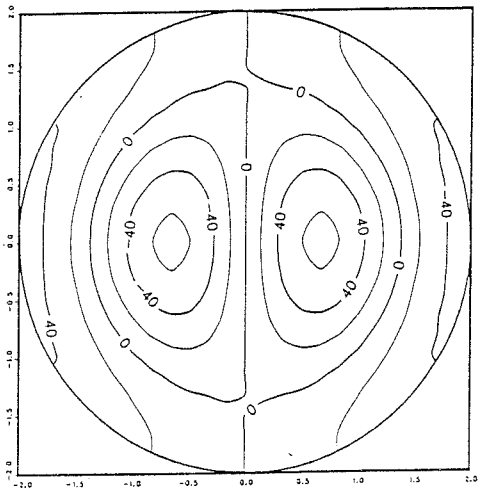
Signalons aussi que, la méthode de résolution présentée dans ce chapitre (§ 2-2-2) est suffisamment générale pour permettre de traiter d'autres modèles de covariance que celui proposé par Fortus.

En particulier, nous avons pu calculer les solutions analytiques, toujours sur un cercle de rayon R, correspondant au modèle:

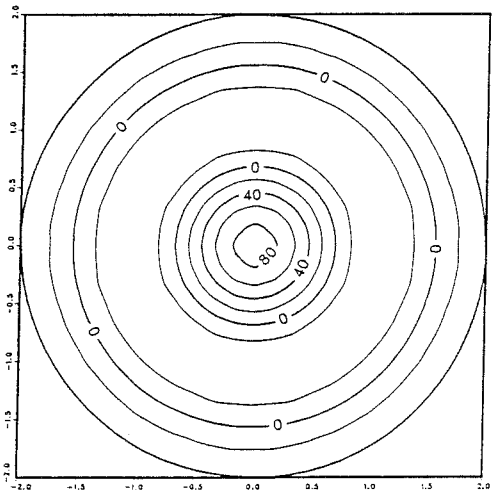
$$C(r) = \gamma r \,.\, K_1(\gamma r) \qquad (2\text{-}40)$$



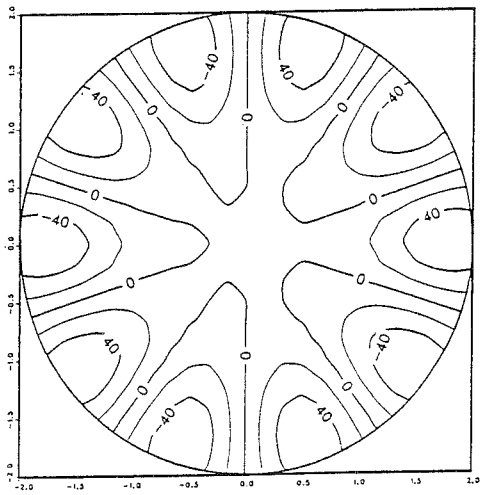
F.P. 2 m=1 λ= 1.57



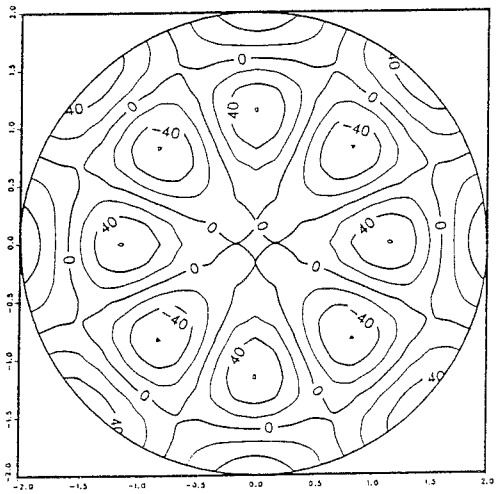
F.P. 9 m=1 λ= 0.27



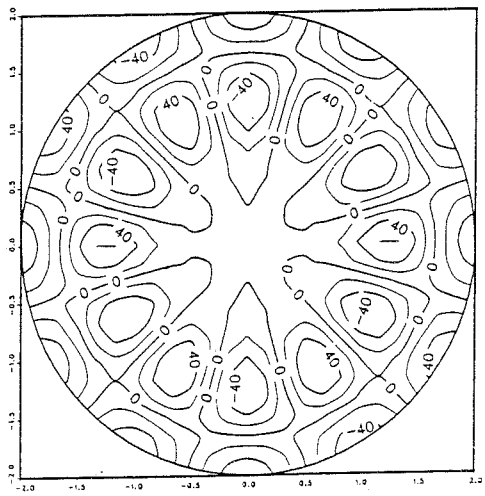
F.P. 15 m=0 λ= 0.13



F.P. 16 m=5 λ= 0.11



F.P. 24 m=4 λ= 0.0044



F.P. 42 m=6 λ= 0.0019

Fig. 2-2: Quelques fonctions propres analytiques pour R=2, α=2, β=1.

où $K_1(x)$ est une fonction de Bessel modifiée d'ordre 1, plus communément utilisé en pratique. Il s'avère que les F.P. ont toujours la même forme analytique à l'intérieur du cercle (équation (2-27)), mais la solution à l'extérieur du cercle est modifiée (car les densités spectrales sont différentes). Ceci change alors les conditions aux limites et donc les valeurs de χ et κ intervenant dans (2-27). L'annexe E présente rapidement la démarche suivie avec ce modèle de covariance.

Après ce paragraphe "mathématique", nous allons passer à l'exploitation concrète des résultats obtenus, en comparant cette solution analytique aux fonctions propres numériques calculées par la méthode décrite au § 1-3. Cette comparaison comporte deux étapes distinctes visant à séparer les erreurs introduites par la discrétisation spatiale du problème d'une part (c'est-à-dire lorsqu'on travaille avec un nombre fini P de points et que l'on fait l'approximation dans une base de fonctions $e_i(\underline{x})$, cf § 1-3) et l'échantillonnage statistique d'autre part (c'est-à-dire lorsqu'on ne dispose que d'un nombre fini N de réalisations pour évaluer la fonction de covariance). Pour cela, dans le premier cas, la matrice de covariance sera remplie en utilisant le modèle (2-3). Aucune erreur d'estimation des covariances ne sera donc introduite dans cette matrice, contrairement au second cas où les covariances seront estimées sur un échantillon simulé de taille finie. La méthode des bandes tournantes (Matheron, 1973), décrite plus en détail au chapitre suivant (§ 3-2-3) nous fournira un tel échantillon.

2-3 INFLUENCE DE LA DISCRETISATION SPATIALE DU PROBLEME SUR LA PRECISON DES FONCTIONS PROPRES.

2-3-1 Méthode de comparaison entre solution analytique et solution numérique.

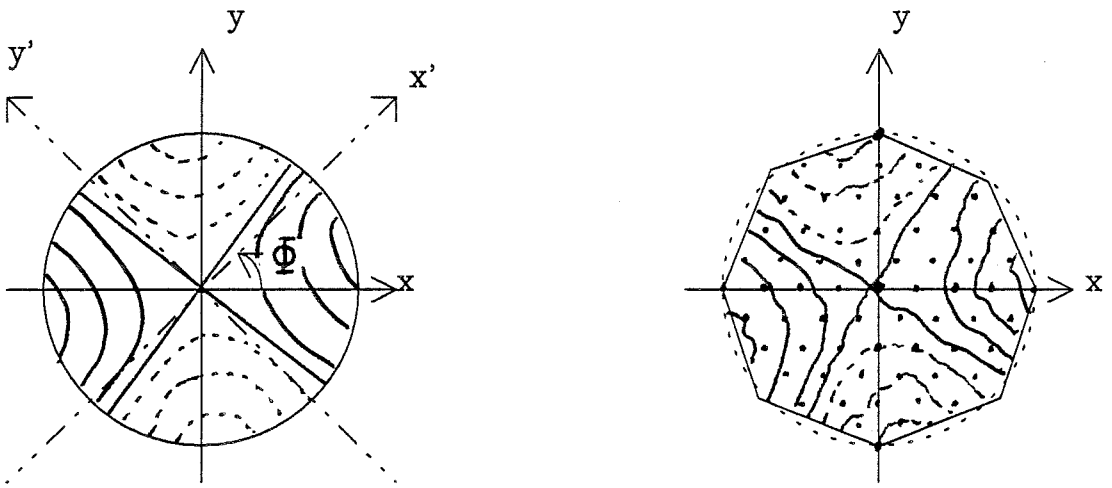
La comparaison peut paraître, a priori, assez difficile dans la mesure où la solution analytique est connue en tout point du domaine D, qui, rappelons-le, est un cercle de rayon R, alors que, par définition, la solution numérique ne l'est qu'en un nombre fini P de points. Nous avons, dans ces conditions, retenu comme critère de qualité le coefficient de détermination, c'est-à-dire le coefficient de corrélation au carré, entre les valeurs aux P points de grille et les valeurs de la solution analytique en ces mêmes P points. Néanmoins, ceci ne suffit pas. En effet, la solution analytique sur le cercle a été déterminée avec un choix arbitraire de l'origine de la coordonnée angulaire θ. Le calcul direct du coefficient de détermination pourrait alors donner des résultats catastrophiques même si, par exemple, les deux solutions étaient identiques à une rotation près. C'est pourquoi, pour chaque fonction

étudiée, nous avons recherché le "meilleur" coefficient de détermination ρ^2 entre la solution numérique et la solution analytique après rotation d'angle Φ. (Fig. 2-3)

$$\rho^2 = \underset{\Phi}{\text{Max}}\ \rho_\Phi^2 \qquad (2\text{-}41)$$

où ρ_Φ^2 est le coefficient de détermination entre les P valeurs de la solution numérique et les P valeurs de la solution analytique aux même points après rotation d'angle Φ.

En pratique, les ρ_Φ^2 ont été calculés tous les 5°, à partir de $\Phi = 0$.



Solution analytique

Solution numérique

Fig. 2-3: Comparaison entre solution analytique et numérique.

Le coefficient de détermination ρ^2 constitue néanmoins un critère global qui pourrait cacher certaines anomalies systématiques. C'est pourquoi une cartographie des solutions numériques a été effectuée afin d'éventuellement les détecter.

Par ailleurs, nous avons calculé les coefficients de détermination entre F.P. analytiques et numériques de même rang, mais aussi entre F.P. numériques et analytiques de rangs différents, afin de voir si des inversions dans l'ordre des F.P. pouvaient se produire, ou si des F.P. numériques pouvaient se "mélanger", notamment lorsque les valeurs propres (notées V.P. par la suite) sont proches.

Dans le cadre de l'analyse des résultats, nous nous intéresserons d'abord à l'estimation des V.P. puis à celle des F.P., en nous concentrant plus particulièrement sur ce dernier point.

2-3-2 Influence de la densité du réseau.

Cette étude a été réalisée en discrétisant le cercle de rayon R=2 à partir d'un maillage carré, (Fig. 2-6-a), les fonctions de base étant LPF. Nous verrons plus loin que le choix d'un maillage carré n'était pas le plus judicieux, mais le temps nous a manqué pour refaire cette étude.

Les valeurs des pas Dx utilisés et le nombre P de points correspondant sont données dans le tableau 2-1.

Les coefficients de détermination ρ^2 ont été calculés sur les P=113 points communs à ces différents réseaux, donc correspondant à Dx=R/6 (sauf pour Dx=R/9 où ils ne coïncidaient pas), ce, afin de pouvoir effectivement les comparer.

Dx	R/6	R/9	R/12	R/18
P	113	269	441	1025

Tableau 2-1: Pas de discrétisation utilisés et nombres de points correspondants.

2-3-2-1 *Précision du calcul des valeurs propres.*

L'analyse suivante est basée sur les graphes de la Fig 2-4 qui donnent les V.P. en fonction de leur rang K, selon le pas de discrétisation choisi. Attention néanmoins, aux différences d'échelle entre les premières et les dernières V.P.

La première chose qui frappe est que, lorsqu'on augmente la densité du réseau, on n'améliore pas systématiquement la précision des premières V.P., contrairement à ce qui se passe plus loin pour K>12. Par ailleurs, on a presque toujours sous-estimation des V.P. Mais ceci peut s'expliquer puisque, à cause de la discrétisation (voir Fig 2-6-a), la surface du domaine "numérique" est pratiquement toujours inférieure à la surface "théorique" ΠR^2 du cercle.

Or, si on définit la variance totale du processus centré par:

$$var_{tot} = E\left[\int_D X^2(\xi, \underline{x}) \cdot d\underline{x} \right]$$

en utilisant (1-1), var_{tot} s'écrit:

$$var_{tot} = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \int_D E[Z_k(\xi) \cdot Z_m(\xi)] \cdot F_k(\underline{x}) \cdot F_m(\underline{x}) \cdot d\underline{x}$$

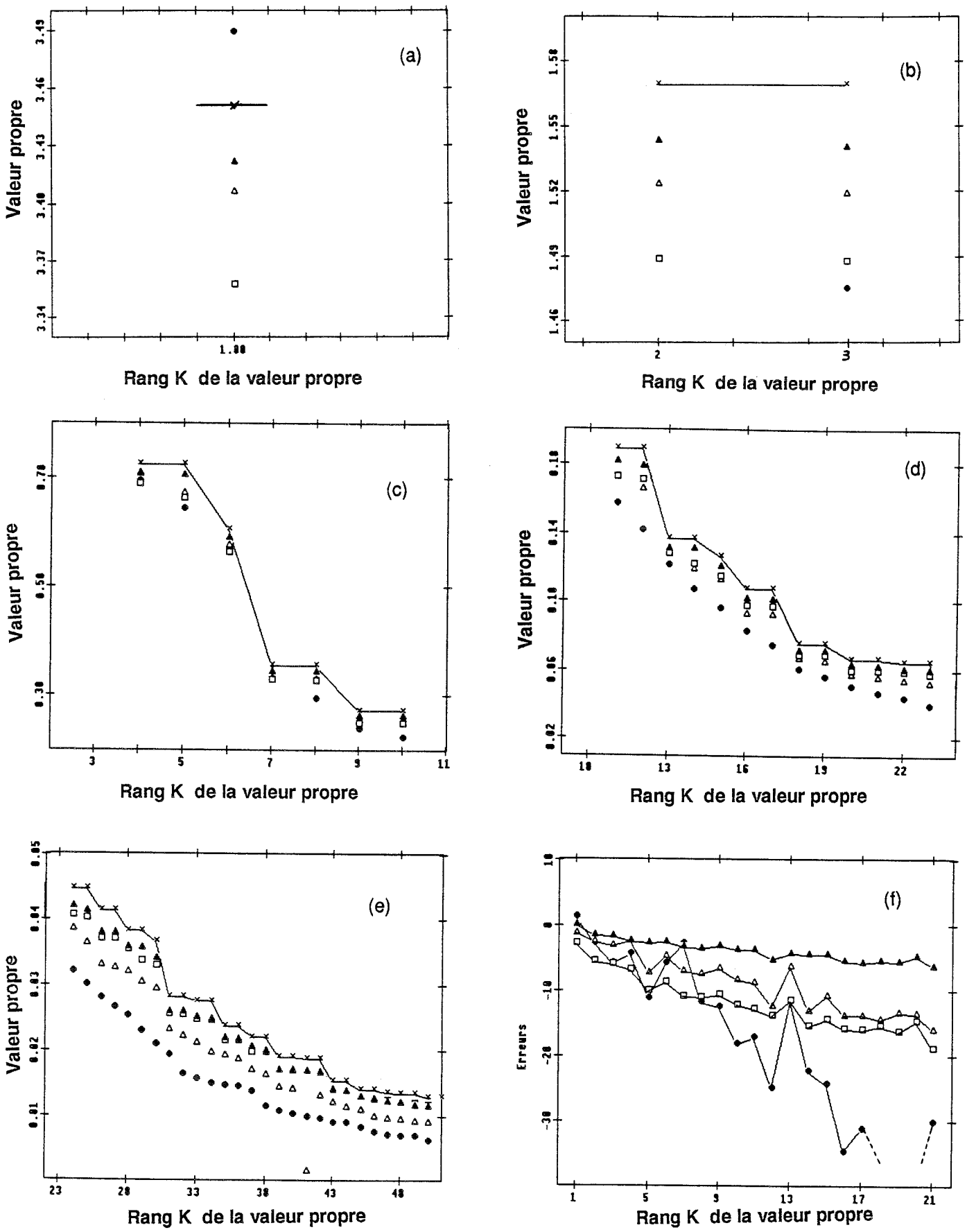


Fig. 2-4: Valeurs propres en fonction de leur rang K (R=2, α=2, β=1) selon la densité P du réseau. (Attention : les échelles des diverses figures sont différentes).

—×— valeurs propres théoriques, ● P=113, △ P=269, □ P=441, ▲ P=1025.

(a) V.P. 1, (b) V.P. 2 et 3, (c) V.P. 4 à 10, (d) V.P. 11 à 23, (e) V.P. 24 à 50, (f) Erreurs relatives

$$var_{tot} = \sum_{k=1}^{\infty} \lambda_k . \int_D F_k^2(\underline{x}) .d\underline{x}$$

puisque les $Z_k(\xi)$ sont décorrélées (formule (1-6)).

La normalisation à un des F.P. (formule (1-3)) nous permet de conclure que:

$$var_{tot} = \sum_{k=1}^{\infty} \lambda_k \qquad (2\text{-}42)$$

Or (Fortus, 1975), la somme des valeurs propres est égale à $C(0).\Pi R^2 = \Pi R^2$ puisque $C(0) = 1$. C'est donc la surface du domaine d'étude qui fixe la variance totale et on peut voir (tableau 2-2) que la discrétisation est responsable, en partie, d'une perte de variance.

P	113	269	441	1025	théorique
Surface réelle	11.55	12,34	11,78	12,37	12,57
Somme des V.P. ou Variance totale	11.12	11.6	11.32	12.01	12.57

Tableau 2-2: Variance totale selon la densité du réseau.

Néanmoins, même si, en valeur absolue, les premières V.P. semblent plus éloignées des valeurs théoriques que les suivantes, le graphique des erreurs relatives (Fig. 2-4-f) montre, qu'en fait, la dégradation de l'estimation des V.P. augmente avec K, rang de la V.P.

Ces résultats ont été confirmés par un autre essai avec R=5, α=1.2, β=1.

Remarque: En augmentant légèrement la taille du rayon R (R=2.06 au lieu de R=2) pour que la surface du domaine numérique coïncide avec la surface théorique, on augmente les premières valeurs propres, mais on diminue les dernières par rapport au cas précédent. Le problème de la perte de variance à cause de la discrétisation n'est alors pas résolu de façon satisfaisante.

Un autre point intéressant à regarder est l'aptitude de la méthode à séparer les valeurs propres doubles. (Toutes sont doubles sauf λ_1, λ_6, λ_{15}, λ_{30}, λ_{46}. Elles correspondent, rappelons-le, au cas m=0 de la solution analytique) Manifestement, (Fig. 2-4), avec P=113 points, le maillage est trop lâche pour y parvenir (ce que confirmera l'étude des F.P.), alors que pour les autres maillages, c'est satisfaisant, (même si on obtient toujours un spectre continûment décroissant), du moins si l'ordre des V.P. n'est pas trop élevé. Par ailleurs, plus

n° FP numérique	n° FP analytique	P=113	P=269	P=441	P=1025
1	1	.94	.997	.997	.997
2	2,3	.99	.999	.999	.999
4	4,5	.71	.998	.999	.997
	6	.22			
5	4,5	.98	.999	.998	.996
	6				
6	4,5	.38			
	6	.59	.997	.997	.999
13	13,14	.53	.96	.99	.99
	15	.22			
14	13,14	.86	.99	.91	.99
	15				
15	13,14	.46			
	15	.47	.93	.89	.99
20	18,19	.21			
	20,21	.74	.76	.95	.99
21	20,21	.12	.98	.67	.99
	22,23	.75			
22	20,21	.54			
	22,23	.33	.99	.75	.999
23	20,21				
	22,23	.78	.96	.98	.999
34	33,34		.89	.98	.97
43	43,44		.40	.69	.95
50	50,51			.72	.94

Tableau 2-3: Coefficients de détermination ρ^2 entre les valeurs des F.P. numériques et analytiques sur le réseau de P=113 points selon la densité du réseau de calcul. (Par convention, nous n'avons pas porté les valeurs de ρ^2<.1).

le maillage est fin, plus la bonne identification des V.P. doubles subsiste longtemps lorsque l'ordre K des V.P. croît.

2-3-2-2 *Précision du calcul des fonctions propres.*

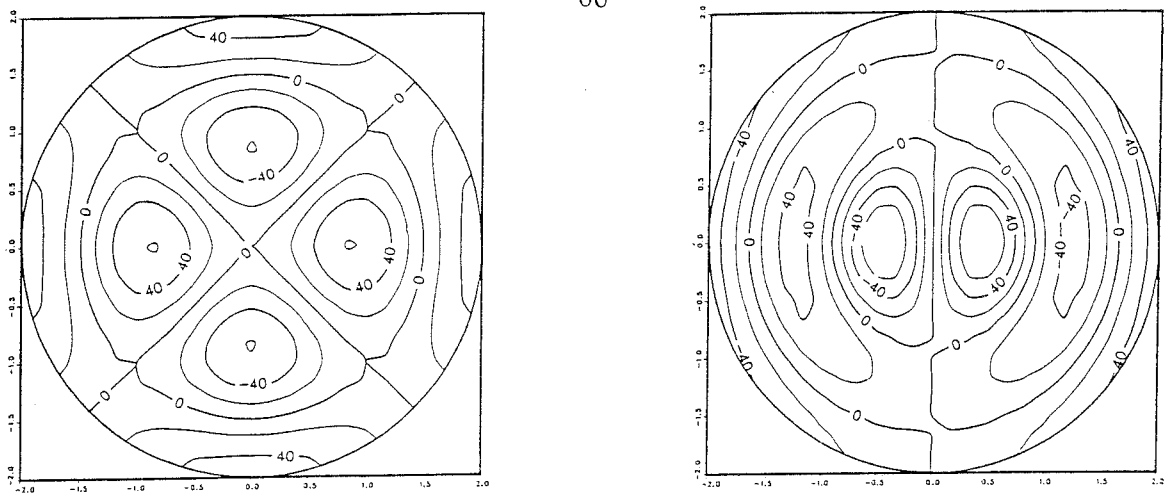
L'analyse est basée sur les coefficients de détermination ρ^2 consignés dans le tableau 2-3. La lecture de ce tableau doit se faire de la façon suivante. La première colonne contient le numéro (ou rang K) de la F.P. numérique étudiée, la seconde celui de la F.P. analytique avec laquelle a été faite la comparaison (et donc la régression). Plusieurs F.P. analytiques ont été testées, afin de voir comment les F.P. numériques pouvaient se mélanger (Une même F.P. numérique est alors comparée avec plusieurs F.P. analytiques, correspondant aux différentes lignes du tableau). Par ailleurs, lorsque une F.P. analytique correspondait à une V.P. double, nous avons mis deux numéros dans la deuxième colonne (par exemple 4,5). Rappelons qu'alors les deux F.P. correspondantes se déduisent l'une de l'autre par rotation d'angle $\pi/2$ et, comme nous recherchons par rotation de la F.P. analytique la meilleure corrélation avec la F.P. numérique (cf § 2-3-1), il est inutile de les distinguer. Les colonnes suivantes contiennent les coefficients de détermination ρ^2, issus de la régression, selon la densité P du réseau. La Fig. 2-5 illustre les résultats pour les F.P.13 et 20.

Globalement, contrairement à ce qui se passait avec les V.P., l'augmentation de la densité du réseau améliore l'estimation des F.P., jusqu'à un ordre d'autant plus grand que P est grand (avec P=1025, ρ^2 = .94 pour la cinquantième F.P.). On remarque encore qu'avec P=113 points, la résolution n'est pas suffisante puisque dès la F.P.13, on ne peut plus isoler chaque F.P.: des "mélanges" se produisent entre plusieurs F.P. dont les V.P. estimées sont proches (F.P.4 et 6 ou 13 et 15 par exemple). Des permutations sont aussi possibles (F.P.21 et 22 par exemple).

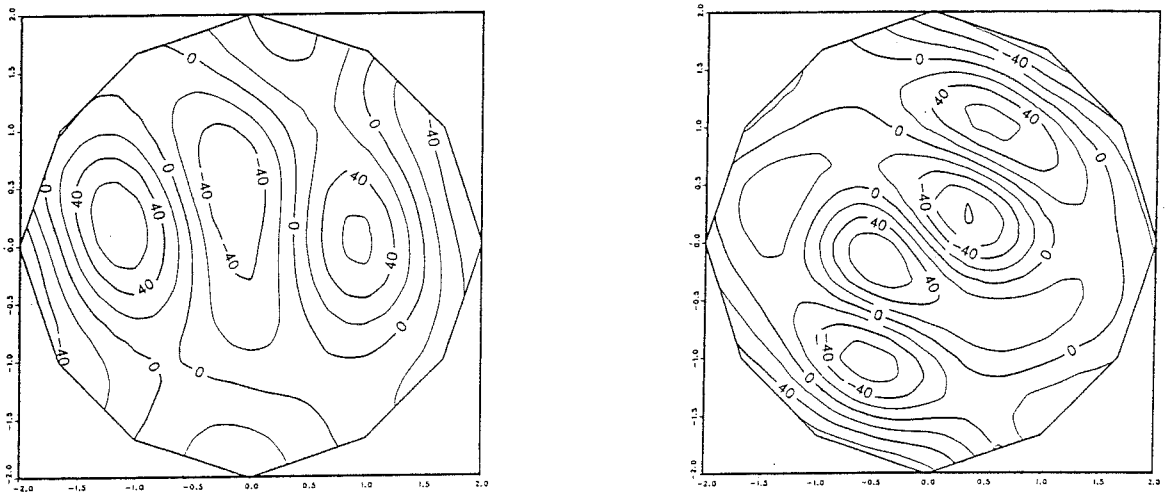
Par ailleurs, pour une densité donnée, il n'y a pas systématiquement dégradation de l'estimation lorsque le rang des F.P. augmente (pour P=113, la F.P.20 est mieux estimée que la F.P.13), même si c'est la tendance générale. Dans ce cas, ça peut être la plus ou moins bonne estimation des V.P. et leur proximité qui peuvent jouer. Et n'oublions pas que nous nous sommes placés d'emblée dans un cas "difficile" quoiqu'assez courant, puisque la majorité des V.P. est double.

F.P. 13 F.P. 20

(a)

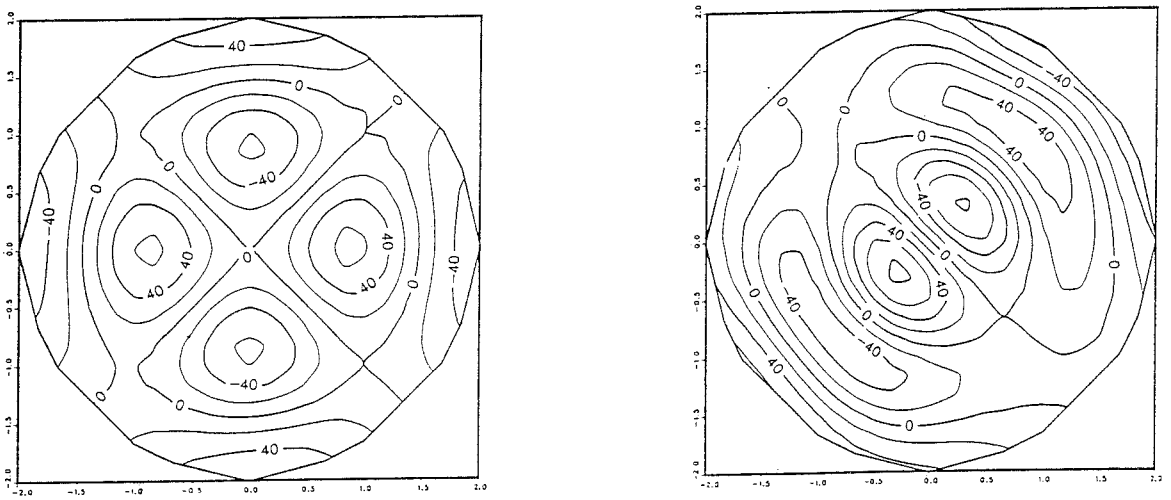


(b)



$\rho^2 = .53$ $\rho^2 = .74$

(c)



$\rho^2 = .99$ $\rho^2 = .95$

Fig. 2-5: Comparaison entre F.P. numériques et analytiques selon la densité P du réseau.
A gauche: F.P. 13, à droite: la F.P. 20.
(a) Solution analytique, (b) P=113, (c) P=441.

Les remarques précédentes restent valables pour le cas R=5, α=1.2, β=1 (rapport taille du domaine/ portée du modèle deux fois plus grand que pour R=2, α=2, β=1), où les densités utilisées pour le réseau ont été les mêmes. Les formes des F.P. sont tout à fait identiques au cas précédent, mais elles sont plus "aplaties", car le domaine est plus grand, et il faut respecter la condition de normalisation à un. Les coefficients de détermination ρ^2 sont aussi strictement analogues au cas R=2, α=2, β=1, quoique légèrement inférieurs. (Par strictement analogues, nous entendons que lorsque ρ^2 était très mauvais, il le reste; lorsqu'il était bon ou moyen, il le demeure encore).

Or, on aurait pu croire, comme le rapport taille du domaine/ portée du modèle est plus grand; que les F.P., avec le même nombre de points auraient été plus mal estimées. Ça n'est manifestement pas vrai sur les premières F.P. , et ça serait alors le rapport taille du domaine/ pas de discrétisation qui gouvernerait la qualité de l'estimation des F.P., plutôt que le rapport taille du domaine/ portée du modèle. Nous verrons au § 2-3-5 comment on peut moduler cette affirmation.

2-3-3 Influence de la régularité du réseau.

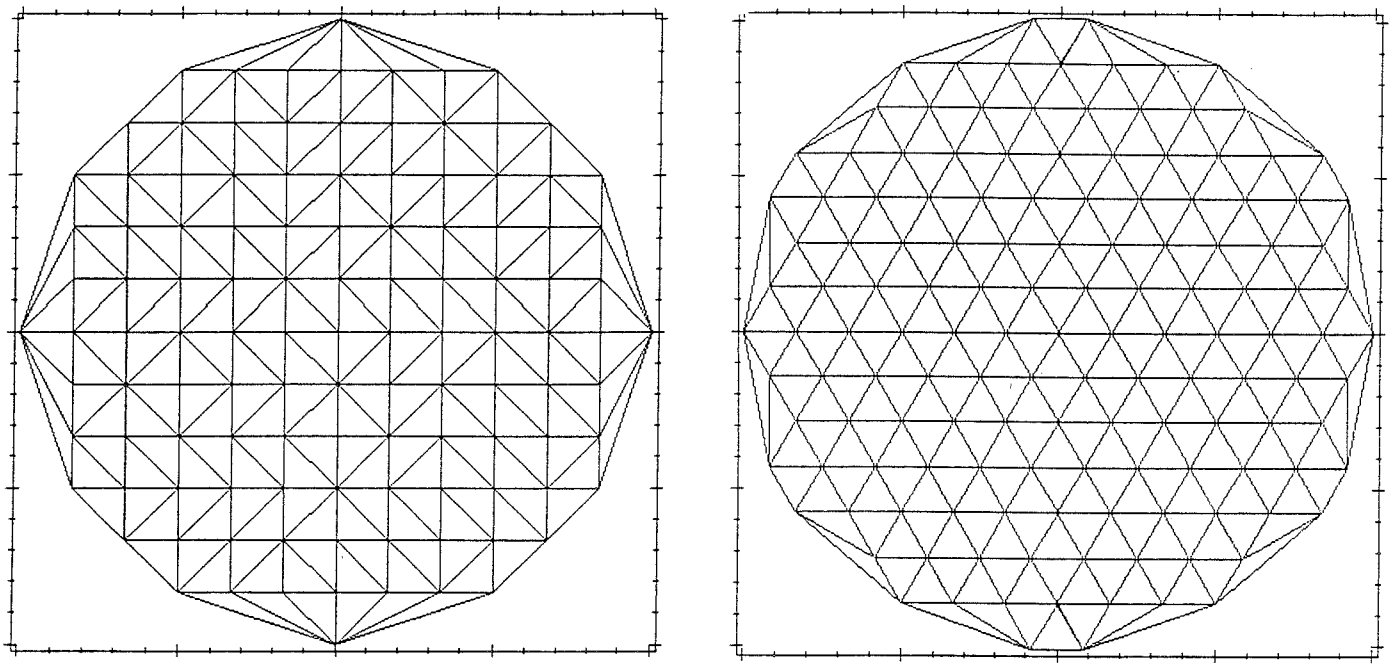
Pour un nombre de points à peu près constant, plusieurs types de maillages, présentés sur la Fig. 2-6 ont été utilisés: carré (P=113), triangulaire (P=139), aléatoire stratifié (P=119) (chaque point du maillage est tiré aléatoirement dans chaque carré défini à partir d'une grille régulière), aléatoire pur (P=110). Les fonctions de base étant linéaires par facettes (LPF), les réseaux ont ensuite été triangulés.

Le domaine D d'étude est le domaine connexe s'appuyant sur le réseau, ce qui implique des surfaces totales différentes et donc des variances totales différentes (cf § 2-3-2-1), comme le montre le tableau 2-4.

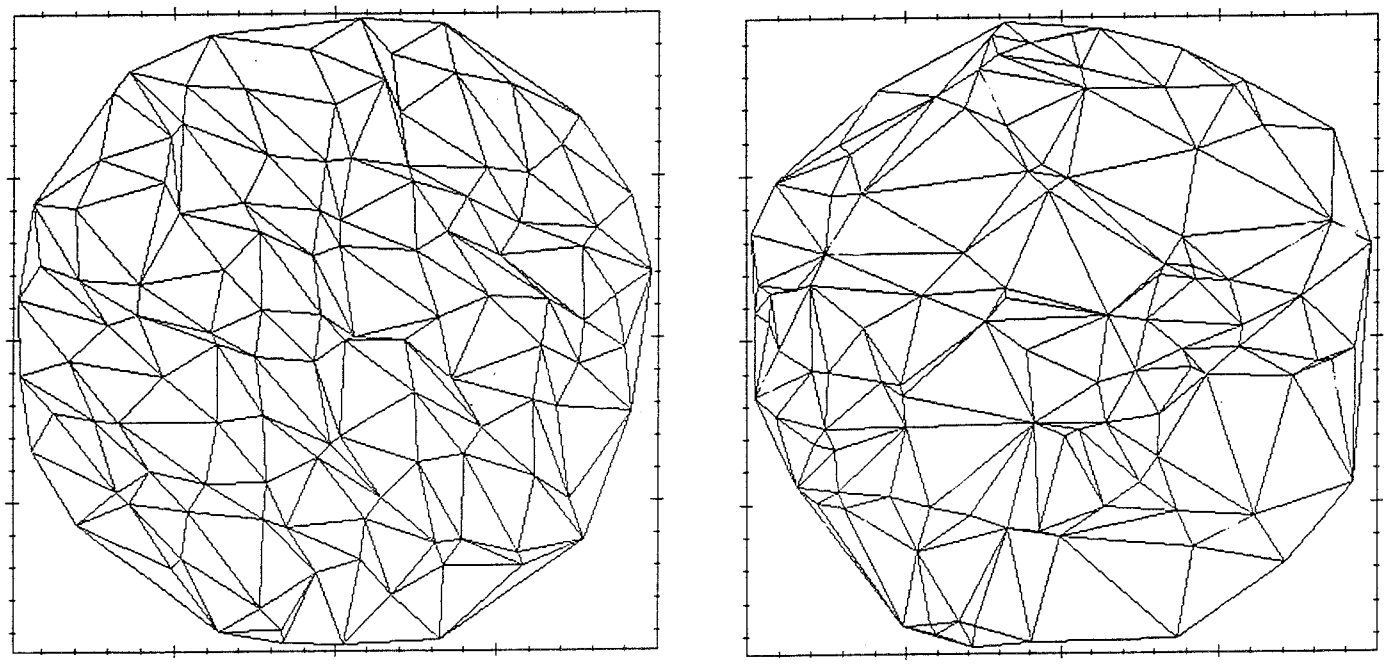
Maillage	Carré	Triangulaire	Aléatoire stratifié	Aléatoire pur	Théorique
Somme des V.P. ou Variance totale	11,12	11,33	10,44	10,11	12,57

Tableau 2-4: Variance totale selon la régularité du réseau.

Ceci a des conséquences directes sur l'estimation des V.P., avec la même sous-estimation générale qu'au § 2-3-2-1, qui est ici d'autant plus importante que la variance totale est faible. En revanche, même si les V.P. elles-mêmes sont différentes, les pourcentages de variance expliqués par les K premières V.P. sont comparables.



(a) Carré (b) Triangulaire

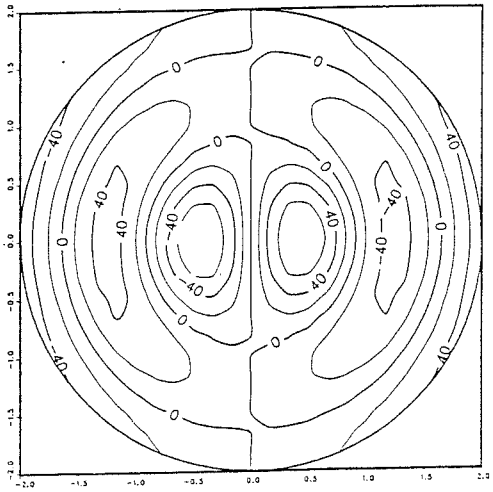


(c) Aléatoire stratifié (d) Aléatoire pur

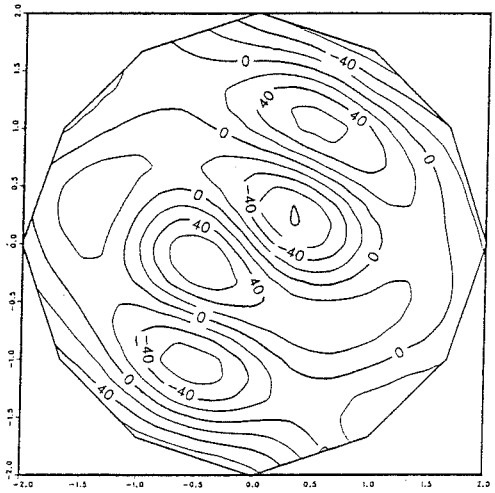
Fig. 2-6: Les différents types du maillages utilisés.

		Maillage			
N° FP numérique	N° FP analytique	Carré	Triangulaire	Aleatoire stratifié	Aléatoire pur
1	1	.95	.999	.997	.997
2	2,3	.99	.999	.999	.999
13	13,14	.50	.99	.81	.94
	15	.19		.20	
14	13,14	.94	.99	.89	.81
	15				.16
15	13,14	.35		.14	
	15	.60	.998	.79	.75
20	20,21	.60	.99	.61	.60
	22,23	.11		.13	.33
21	20,21		.99	.44	.16
	22,23	.63		.47	.61
29	26,27	.21	.97	.69	.28
	28,29	.54			.19
	30	.19			
30	26,27	.11			
	28,29	.39		.49	.17
	30	.42	.99	.29	.18
39	41,42	.61	.98	.39	.15
40	39,40	.46			.29
	41,42		.98		.13
46	43,44	.35	.18	.23	
	45,46	.22	.72	.27	.19
	47			.10	
50	48,49	.16	.84	.11	
	50,51	.57			

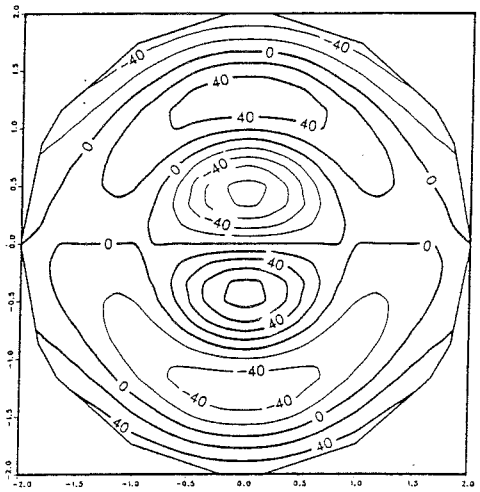
Tableau 2-5: Coefficients de détermination ρ^2 entre valeurs des F.P. numériques et analytiques selon la régularité du réseau.



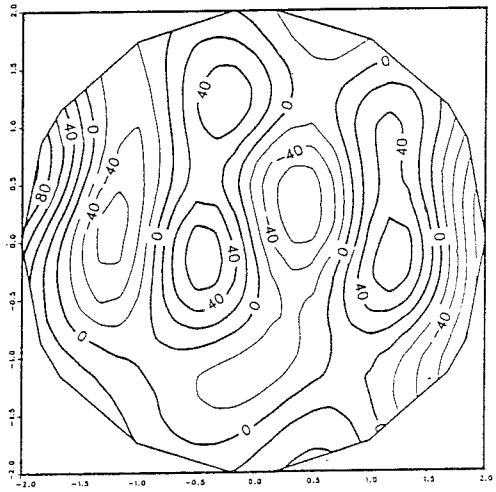
(a) Solution analytique



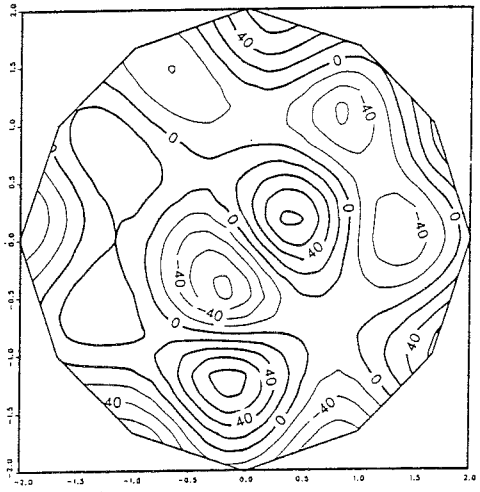
(b) Maillage carré $\rho^2 = .60$



(c) Maillage triangulaire $\rho^2 = .99$



(d) Maillage aléatoire stratifié $\rho^2 = .61$



(e) Maillage aléatoire pur $\rho^2 = .60$

Fig. 2-7: Comparaison, pour la F.P. 20, entre F.P. numériques et analytiques selon le maillage utilisé.

L'analyse des F.P., basée sur le tableau 2-5 (la description des colonnes est la même qu'au § précédent) et la Fig. 2-7 qui illustre les résultats pour la F.P. 20, montre que le maillage triangulaire est le plus performant (ρ^2 de l'ordre de .99 jusqu'à la quarantième F.P.) et surtout qu'il réussit très bien à séparer les différentes F.P. Par ailleurs, on pouvait s'attendre à une précision moindre de la part de maillages irréguliers. Or, jusqu'à la F.P.20, les maillages aléatoires stratifiés et aléatoires purs font mieux que le maillage carré. Ceci tend à montrer que lorsque le domaine est circulaire, une grille carrée n'est pas du tout adaptée, contrairement au cas du maillage triangulaire par exemple. En revanche, on assiste ensuite à une dégradation rapide de ρ^2 pour les maillages aléatoires, notamment aléatoire pur. En effet, lorsque l'ordre des F.P. augmente, un tel maillage ne permet plus de bien séparer les F.P. qui se mélangent (une F.P. numérique est significativement corrélée avec plusieurs F.P. analytiques).

Il semble donc qu'un maillage régulier soit à privilégier, pour peu qu'il soit adapté à la forme du domaine, surtout si l'on veut une bonne estimation des F.P. d'ordre plus élevé. Pour les seules premières F.P., les maillages aléatoires donnent par contre des résultats comparables aux maillages réguliers.

2-3-4 Choix des fonctions de base $e_i(\underline{x})$.

Lorsque les F.P. sont calculées aux noeuds d'une grille, le choix des fonctions de base $e_i(\underline{x})$ a, bien sûr, une influence directe sur les valeurs estimées à ces noeuds.

Mais on peut aussi avoir besoin d'estimer les F.P. en des points différents des points de calcul. Il faudra alors y interpoler les F.P. à l'aide des fonctions $e_i(\underline{x})$ en utilisant l'équation (1-8):

$$F_k^*(\underline{x}) = \sum_{i=1}^{P} f_{ik} \cdot e_i(\underline{x}) \qquad (1\text{-}8)$$

où rappelons-le, les f_{ik} sont les valeurs discrètes des F.P. estimées aux noeuds de la grille de calcul.

Nous devrons donc juger, successivement, les performances des fonctions constantes par morceaux (CPM) ou linéaires par facettes (LPF).

i) *sur le calcul direct des F.P.* (On comparera alors directement les valeurs calculées aux noeuds de la grille à leurs homologues analytiques).

ii) *lorsqu'elles sont utilisées en interpolation.*

Pour cela les F.P ont été calculées aux noeuds d'une grille régulière (carrée ou triangulaire), en remplisssant toujours la matrice **C** à l'aide du modèle (cas R=2, α =2, β =1).

Puis, nous les avons interpolées aux noeuds d'une grille aléatoire stratifiée de P=453 points

N° FP numérique	N° FP analytique	ACPP directe sur rés. aléa. stra. P=453	Maillage carré de P=441 points CPM/CPM	LPF/LPF	LPF/spline	Maillage triangulaire de P=139 points CPM/CPM	LPF/LPF
2	2,3	.997	.996	.999	.999	.99	.999
4	4,5	.97	.994	.999	.999	.97	.999
5	4,5	.99	.99	.999	.999	.97	.999
6	6	.96	.99	.998	.998	.97	.999
13	13,14	.77	.97	.99	.99	.91	.99
14	13,14	.96	.96	.90	.91	.90	.99
15	15	.80	.95	.90	.89	.90	.997
20	20,21	.49	.89	.95	.95	.84	.99
21	20,21	.85	.89	.68	.69	.86	.99
22	22,23	.79	.96	.75	.73	.88	.99
23	22,23	.69	.95	.98	.98	.88	.97
30	28,29	.21		.37	.39		
	30	.60	.87	.52	.52	.80	.98
40	39,40		.87	.83	.84	.73	
	41,42	.85			.10		.94
41	39,40	.36	.90	.63	.63	.76	.93
	41,42	.50		.29	.29		
47	45,46	.51					
	47	.21	.75			.63	.93
50	48,49	.35					.82
	50,51	.43	.86	.74	.74	.63	

Tableau 2-6: Coefficients de détermination ρ^2 entre les F.P. numériques et analytiques interpolées aux noeuds d'une grille aléatoire stratifiée après calcul des F.P. sur une grille régulière en utilisant, soit des fonctions CPM, soit des fonctions LPF. Le sens des notations CPM/CPM, LPF/spline, etc.. est donné dans le texte.

(voir Fig. 2-8) en utilisant des fonctions CPM ou LPF. A titre de comparaison, des fonctions splines ont aussi été testées (pour la phase d'interpolation uniquement). Nous avons aussi calculé directement les F.P. aux noeuds de la grille aléatoire stratifiée, sans nécessité d'interpoler dans ce cas.

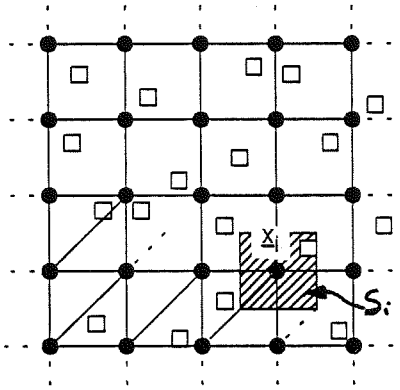


Fig 2-8: Schéma des grilles utilisées, respectivement pour le calcul en CPM (utilisation d'une discrétisation par polygones de Thiessen, ici carrés) ou en LPF (triangulation du réseau) des F.P. (les points de grille sont des cercles), puis leur interpolation sur une grille aléatoire stratifiée (les points de grille sont alors des carrés).

Au niveau des résultats, nous ne nous appesantirons pas sur le point i), puisque, pour un calcul direct des F.P. aux noeuds d'une grille, fonctions CPM et LPF possèdent des performances comparables (et en général très bonnes). Seul le maillage à base carrée de P=113 points donne, comme nous l'avons déjà souligné, de très mauvais résultats en LPF, alors qu'ils sont bons en CPM. Il semblerait donc que l'introduction d'une complexité supplémentaire en LPF ne se traduise pas par de meilleurs résultats.

En revanche, si nous nous intéressons aux résultats en interpolation, les fonctions LPF se montrent les plus performantes. Ces conclusions s'appuient sur les coefficients de détermination ρ^2 (tableau 2-6) entre les valeurs des F.P. interpolées aux points du réseau aléatoire stratifié et les valeurs analytiques correspondantes (R=2, α =2, β =1). La première colonne contient toujours le numéro de la F.P. numérique étudiée, le seconde, celui de la F.P. analytique avec laquelle nous avons effectué la comparaison. La troisième colonne donne quant à elle les ρ^2 lorsqu'on fait un calcul direct des F.P. aux points de la grille aléatoire stratifiée. Pour les trois suivantes, la grille de calcul était carrée avec P=441 points alors que pour les deux dernières, elle était triangulaire (P=139). Pour ces cinq colonnes, LPF/LPF signifie par exemple que les F.P. ont été estimées sur la grille régulière en utilisant des fonctions LPF, puis interpolées à l'aide de ces mêmes fonctions LPF aux points de la grille aléatoire stratifiée.

Pour LPF/spline, l'interpolation a, cette fois, été menée à l'aide de fonctions splines de type plaque mince (Duchon, 1976), les F.P. sur la grille de calcul étant toujours estimées à l'aide de fonctions LPF.

L'analyse de ces coefficients de détermination ρ^2 montre que les différentes techniques sont tout à fait équivalentes pour les premières F.P. En revanche, pour les F.P. d'ordre plus élevé (à partir de la treizième F.P.), le calcul sur une grille régulière (carrée assez dense de P=441 points ou triangulaire peu dense de P=139 points), puis l'interpolation sur le réseau aléatoire stratifié donne de meilleurs résultats que le calcul direct des F.P. sur ce même réseau. On retrouve d'ailleurs ce que nous avons vu au § précédent: sur un réseau aléatoire, les premières F.P. sont bien estimées, mais il y a dégradation et difficulté à séparer les F.P. qui suivent. Par ailleurs, au niveau des techniques d'interpolation, l'approche LPF est plus performante sur un réseau triangulaire. C'est plutôt le contraire pour un réseau carré, mais nous avons déjà souligné la moins bonne adaptation de ce réseau à un domaine circulaire. L'interpolation spline des F.P. calculées par la méthode CPM ou LPF est légèrement meilleure qu'une interpolation respectivement CPM ou LPF. Mais ce calcul n'a été fait qu'à titre indicatif car, en toute logique, il faudrait utiliser une interpolation spline dès le départ....

Ce qui ressort donc de ce paragraphe, c'est que, dans le cas d'un modèle tel que celui que nous avons imposé, le calcul des F.P. sur une grille régulière, puis leur interpolation sur une grille irrégulière semble préférable à un calcul direct sur ce dernier réseau. Par ailleurs, une approche LPF donnera de meilleurs résultats qu'une approche CPM, qui pourra en revanche suffire pour un calcul des F.P. sans interpolation ultérieure. (Notons néanmoins que, pour un réseau irrégulier, l'approche LPF est plus facile à mettre en oeuvre que l'approche CPM où, il faut, pour chaque station, évaluer la surface du polygone de Thiessen associé, calcul qui, le plus souvent, ne peut être que manuel, surtout à la périphérie. Voir par exemple Obled et Creutin (1986) pour un cas pratique d'application.)

2-3-5 Influence de la portée du modèle de covariance.

Pour un domaine fixé (c'est-à-dire une valeur du rayon R=2), une même discrétisation (même réseau de calcul des F.P.) et un même choix des fonctions de base, nous avons fait varier les paramètres α et β du modèle de covariance (2-3), de façon à diminuer progressivement la longueur de corrélation (voir Fig.2-1). Nous avons défini, à l'aide du concept d'échelle intégrale rappelé par Muñoz-Pardo (1987), une portée équivalente à celle d'un modèle sphérique (Annexe F) et les valeurs en sont données dans le tableau 2-7. Rappelons que la portée d'un modèle sphérique est la distance au-delà de laquelle la

covariance est nulle. La valeur équivalente au modèle (2-3) est celle qui permet d'assurer l'égalité des échelles intégrales des deux modèles.

Nous pensions obtenir une dégradation de la qualité des F.P. estimées, puisque nous diminuions le rapport portée/ pas de discrétisation. Or, en fait, nous avons pu constater qu'il n'en était rien et que les formes des F.P. (voire les valeurs des F.P.) étaient tout à fait semblables, du moins pour les premières. Ceci peut en fait se comprendre simplement à l'aide du petit exemple qui va suivre.

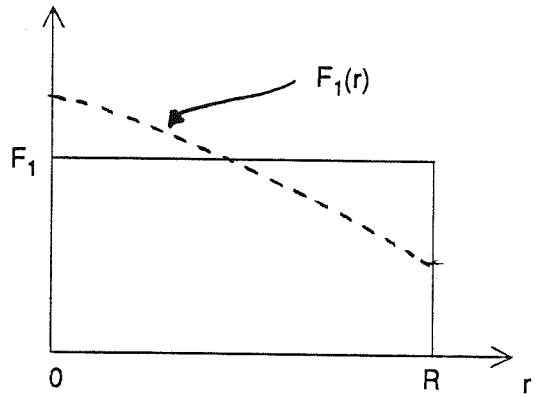


Fig. 2-9: Projection, dans un plan θ = cste, de la fonction $F_1(r,\theta)$.
Schéma de l'approximation utilisée.

Considérons la première F.P. Sa forme est généralement, pour une densité spectrale "classique", assez plate (dôme). En première approximation, nous supposerons qu'elle a une valeur constante sur tout le domaine, i.e. que $F_1(r,\theta) = F_1$. (Fig. 2-9)

Comme nous imposons $\int_D F_1^2(r,\theta) \,.\, r \,.\, dr \,.\, d\theta = 1$ (condition de normalisation), il vient: $\Pi \,.\, R^2 \,.\, F_1^2 = 1$, soit $F_1 = (\Pi \,.\, R^2)^{1/2}$.

Cette valeur approchée est donc totalement indépendante du modèle de covariance et de ses paramètres α et β. C'est uniquement la taille du domaine (et sa forme), qui conditionne la valeur de $F_1(r,\theta)$. Ce type de raisonnement pourrait être étendu aux F.P. d'ordre supérieur. C'est d'ailleurs un résultat déjà obtenu empiriquement par Buell (1975), Kueny (1977) ou Obled (1979) sur quelques simulations: la forme des F.P. est liée au domaine d'étude et le modèle de covariance n'a qu'un rôle secondaire.

Néanmoins, une modification du modèle, si elle affecte peu les F.P., modifie considérablement les valeurs propres et surtout les pourcentages de variance expliqués par les K premières F.P. Rappelons en effet que les V.P. sont données par (2-31):

$$\lambda = (2\Pi)^2 \,.\, S(\chi) = (2\Pi)^2 \,.\, S(i\kappa) \qquad \text{avec}$$

$$S(w) = \frac{(\alpha^2 - \beta^2) \cdot A}{2\Pi} \cdot \frac{1}{(\alpha^2 + w^2) \cdot (\beta^2 + w^2)}$$

On observe alors que, si la longueur de covariance diminue, pour un même rang, les V.P. sont plus faibles. On peut l'interpréter en disant que le processus "blanchit", et donc qu'il reprend de l'énergie dans les hautes fréquences, ce qui augmente les petites V.P. aux dépens des grandes. Pour expliquer le même pourcentage de variance, il faut alors plus de F.P. (voir tableau 2-7 pour 90% de variance expliquée.)

Paramètres du modèle	$\alpha = 2$ $\beta = 1$	$\alpha = 4$ $\beta = 1$	$\alpha = 4$ $\beta = 2$
Portée sphérique asph équivalente	**3.85**	**2.61**	**1.65**
K90%	**23**	**51**	**91**

Tableau 2-7: Paramètres, portées sphériques équivalentes des divers modèles de covariance utilisés. Nombre de F.P. nécessaires pour expliquer 90% de la variance totale.

Pour pouvoir bien estimer ces F.P. d'ordre élevé, il faudra donc diminuer le pas de discrétisation, i.e. augmenter la densité du réseau (cf § 2-3-2-2). On voit donc que, finalement, comme l'intuition nous le suggérait, lorsque la longueur de corrélation du modèle diminue, il faut diminuer le pas de discrétisation, mais uniquement parce qu' on a besoin de plus de F.P. pour expliquer le même pourcentage de variance. Si on n'avait besoin que des premières F.P., ceci ne serait pas nécessaire.

Après l'étude des erreurs introduites par la discrétisation spatiale du problème, étude menée en remplissant la matrice de covariance à l'aide du modèle théorique exact, voyons les erreurs supplémentaires introduites par l'échantillonnage statistique, c'est-à-dire lorsque la matrice de covariance est estimée sur un N-échantillon.

2-4 INFLUENCE DE L' ECHANTILLONNAGE STATISTIQUE SUR LA PRECISION DES VALEURS PROPRES ET FONCTIONS PROPRES.

Pour tester l'influence de l'échantillonnage statistique, il nous fallait tout d'abord pouvoir disposer de champs respectant le modèle de covariance cible donné par (2-3). Nous avons utilisé la méthode des bandes tournantes (décrite au chapitre suivant, § 3-2-3), pour simuler des échantillons de N=100, 300, 1000 champs indépendants, les valeurs du modèle de covariance (2-3) étant fixées à $\alpha = 2$ et $\beta = 1$.

Le réseau utilisé pour l'ACPP est de type aléatoire stratifié avec P=453 points. L'approximation numérique utilise des fonctions linéaires par facettes (LPF).

La méthode des bandes tournantes a été retenue pour générer les réalisations car, d'une part, elle est complètement indépendante de l'ACPP et, d'autre part, les nombreuses applications auxquelles elle a donné lieu en ont montré la robustesse et la fiabilité.

2-4-1 Estimation des valeurs propres.

La Fig. 2-10 présente les V.P. estimées en fonction de leur rang selon la taille N de l'échantillon utilisé (Attention encore une fois aux différences d'échelles entre les premières et les dernières V.P.).

Elle montre que, comme précédemment, les V.P. sont sous-estimées, sauf ici la première. On peut, encore une fois, incriminer une variance totale plus faible à cause de la discrétisation du domaine. Mais, d'après les résultats obtenus par Von Storch et Hannoschöck (1985) grâce à des simulations de Monte-Carlo, lorsqu'une matrice de covariance est estimée sur un N-échantillon, il y a biais systématique impliquant une surestimation des premières valeurs propres et, par suite, une sous-estimation des autres, puisque la trace de la matrice de covariance, et donc la somme des valeurs propres, est non biaisée. Les auteurs proposent d'utiliser des techniques de type "Jacknife" ou des simulations de Monte-Carlo pour déduire, à partir des valeurs propres estimées, des valeurs non biaisées. Mais ceci est assez lourd à mettre en oeuvre.

Pour avoir une idée un peu plus grossière de l'incertitude entachant les V.P. estimées, on peut utiliser une formule au premier ordre donnée par North et al. (1982): si λ est la V.P. théorique, l'incertitude sur son estimation est donnée par:

$$\delta\lambda = \lambda \, . \, (2/N)^{1/2} \qquad (2\text{-}43)$$

La Fig. 2-11 présente ces intervalles pour quelques valeurs propres théoriques. Nous y avons aussi porté leurs estimations empiriques pour N=100, 300, 1000. On note la diminution rapide de la largeur de ces intervalles lorsque N augmente. On voit aussi que, sauf pour λ_1, λ_2, λ_3, les intervalles de confiance entre plusieurs V.P. se chevauchent, ce qui implique qu'elles seront mal séparées et que, par conséquent, les F.P. correspondantes auront tendance à se mélanger. En effet (North et al., 1982), c'est la proximité de deux V.P. qui influe sur l'estimation des F.P. puisque, si λ_k' est la V.P. la plus proche de λ_k:

$$\delta F_k(\underline{x}) = [\, \delta\lambda_k / (\lambda_k - \lambda_k') \,] \, . \, F_k(\underline{x}) \qquad (2\text{-}44)$$

C'est d'ailleurs ce que nous allons voir maintenant en regardant les F.P.

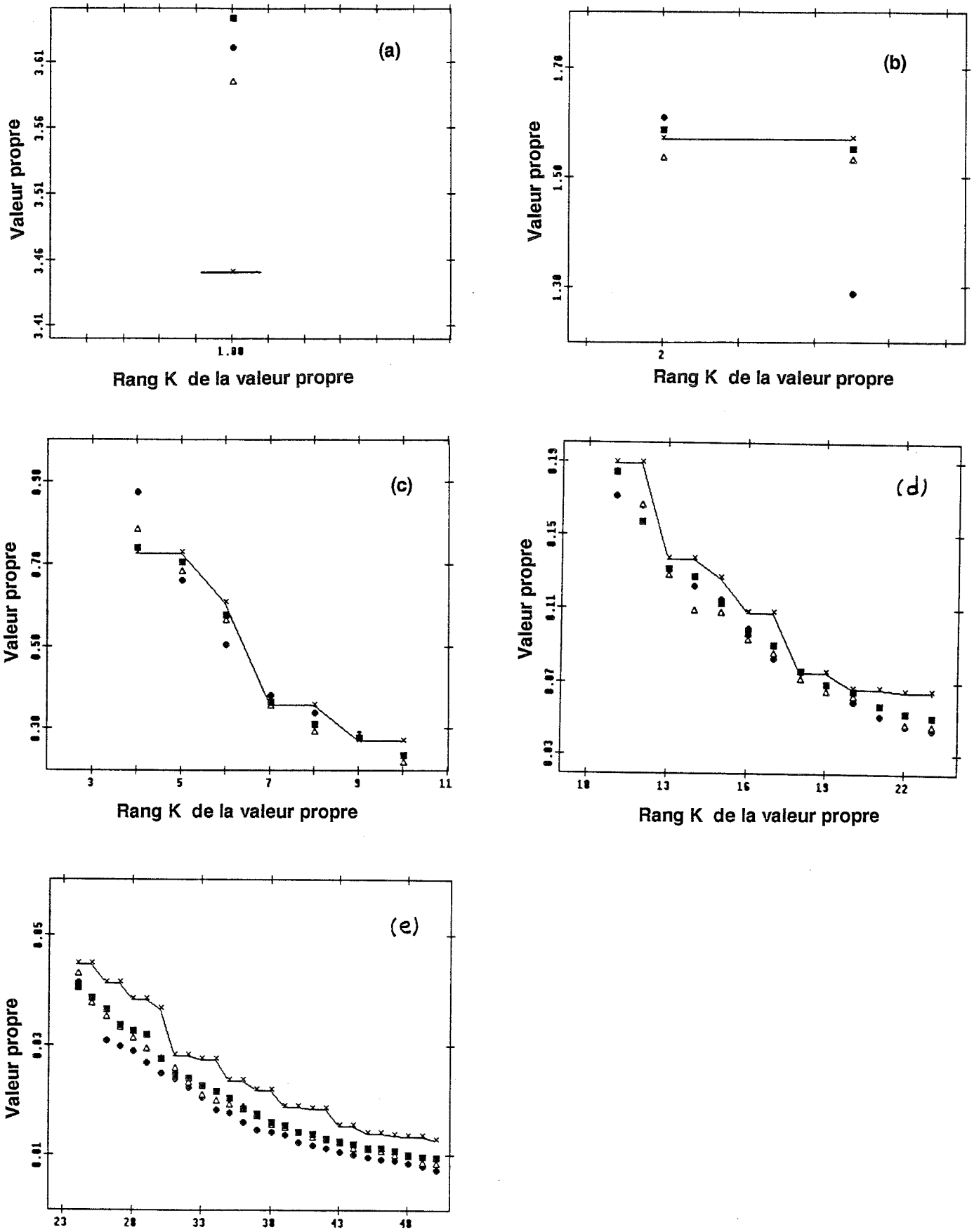


Fig. 2-10:Valeurs propres en fonction de leur rang K (R=2, α=2, β=1) selon la taille N de l'échantillon. (Attention : les échelles des diverses figures sont différentes).

—×— valeurs propres théoriques, ● N=100 , Δ N=300 , ■ N=1000.

(a) V.P. 1, (b) V.P. 2 et 3, (c) V.P. 4 à 10, (d) V.P. 11 à 23, (e) V.P. 24 à 50,

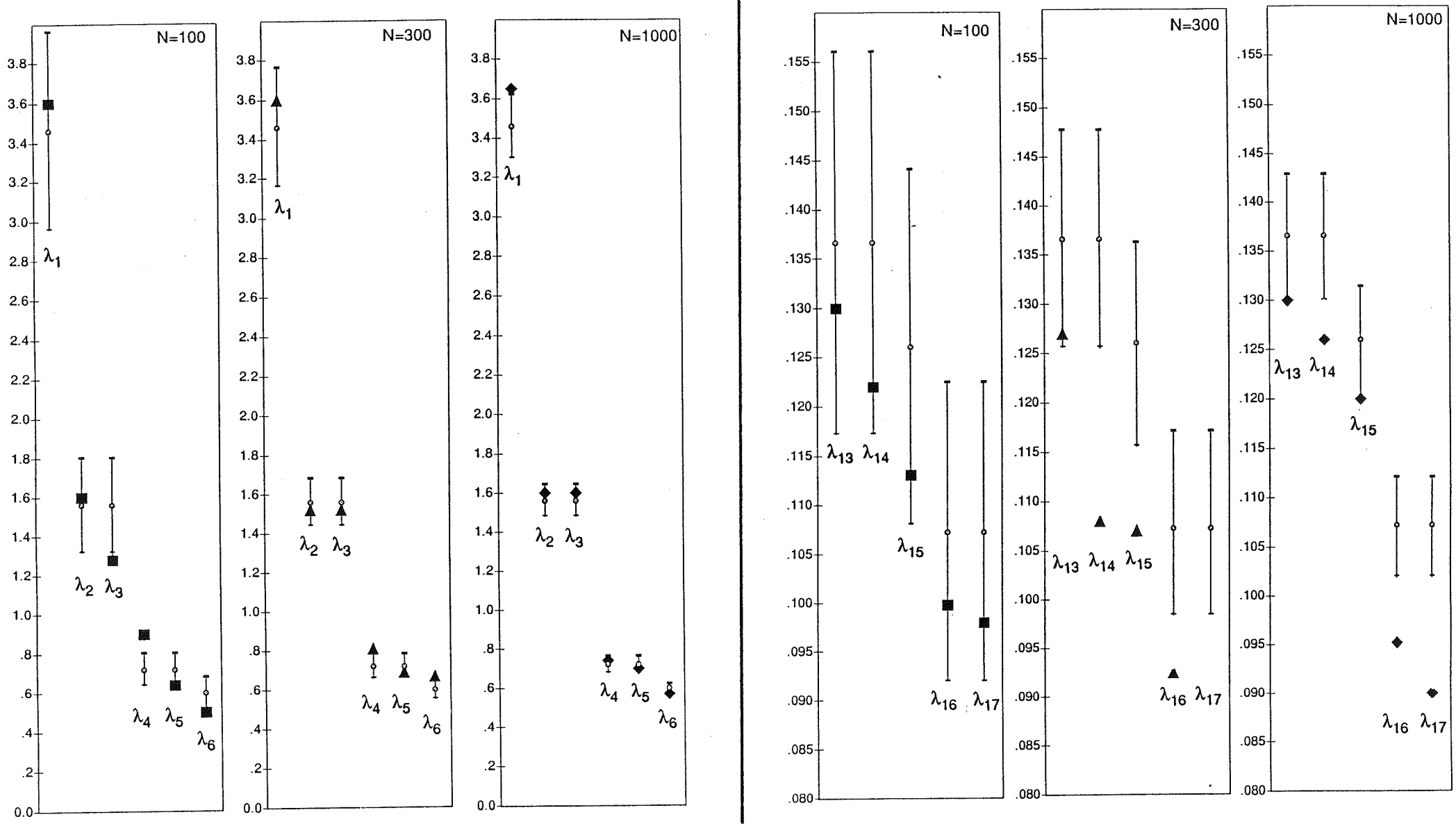


Fig. 2-11:Incertitudes (d'après North et al., 1982) sur les valeurs propres estimées sur un N échantillon.

Valeurs propres théoriques et leur incertitude.

■, ▲, ◆ Valeurs estimées sur un N-échantillon.

2-4-2 Estimation des fonctions propres.

Les coefficients de détermination ρ^2 entre F.P. analytiques et numériques sont donnés dans le tableau 2-8 pour N=100, 300, 1000. La dernière colonne (N=∝) fournit les résultats lorsque la matrice **C** est remplie à l'aide du modèle exact. La Fig. 2-12 les illustre pour les F.P.4 et 20.

Ce qui frappe de prime abord, c'est la dégradation très nette qui apparaît dans l'estimation des F.P., par rapport au § précédent où **C** était remplie à l'aide du modèle et où ρ^2= .99 n'était pas rare. Ici, avec N=100, pour la première F.P., on atteint seulement ρ^2= .42, et dès la quatrième F.P., des mélanges entre F.P. se produisent (qu'on se rappelle les incertitudes sur les V.P. présentées au § précédent). L'augmentation de N permet d'améliorer les résultats, mais, même avec N=1000, on est loin de faire aussi bien que pour N=∝ (cas qui inclut, rappelons-le, les erreurs de discrétisation présentées auparavant). Néanmoins, il ne faudrait pas croire que l'on puisse affirmer que "N=100 observations est insuffisant alors que N=1000 serait un bon chiffre". En effet, les conclusions doivent être relativisées par rapport au nombre de points traités. Obled (1979) a ainsi pu montrer qu'un rapport N/P de l'ordre de dix permettait d'assurer une bonne estimation des V.P., et, dans notre cas, avec P=453 et N=100, 300 ou 1000, nous en sommes assez loins...

Cependant, il apparaît clairement que, dans l'estimation des F.P., c'est vraiment l'échantillonnage statistique qui est le plus pénalisant, beaucoup plus que les erreurs dues à la discrétisation du domaine.

Au vu de ces résultats, il semble judicieux de se poser le problème suivant: lorsqu'on veut traiter un jeu de données, ne vaudrait-il pas mieux ajuster un modèle sur le nuage de corrélation, puis remplir la matrice **C** avec ce modèle? On s'affranchirait alors des bruits introduits par l'estimation des covariances, en gagnant en stabilité sur les F.P. Par ailleurs, nous avons vu qu'une erreur sur la portée du modèle n'avait pas d'incidence notable sur les F.P. estimées. Dans notre cas, si on regarde la Fig 2-13, qui montre les covariances moyennes recalculées sur les échantillons de N=100, 300, 1000 réalisations simulées, on voit que leur modélisation nous donnerait une fonction très proche du modèle théorique. Son utilisation pour remplir la matrice **C** nous permettrait d'obtenir des ρ^2 de l'ordre de ceux de la colonne N=∝, ce qui n'est pas négligeable comme amélioration.

En contrepartie, il est aussi évident que, si les processus sont non homogènes, modéliser le corrélogramme nous fera perdre cette information, qui serait, par contre, implicitement contenue dans la matrice **C** empirique.

Avant d'aborder l'ACPP d'un ensemble de champs, il serait donc très intéressant de faire une analyse du nuage de covariance, de tester la signification des éventuelles

N° FP numérique	N° FP analytique	N=100	N=300	N=1000	N= ∝
1	1	.42	.65	.91	.99
2	2,3	.87	.81	.89	.997
3	2,3	.66	.89	.84	.996
4	4,5	.83	.94	.80	.97
5	4,5	.58	.67	.86	.99
	6	.19			
6	4,5	.12			
	6	.55	.85	.91	.96
13	13,14	.19	.40	.66	.77
	15			.15	.12
14	13,14			.48	.96
	15	.34	.44		
15	13,14	.34	.36		.14
	15			.59	.80
20	20,21	.10		.49	.49
21	20,21	.11	.25	.41	.85
22	20,21				.10
	22,23		.44	.52	.79
23	20,21				.21
	22,23		.41	.62	.69
30	28,29		.24	.19	.21
	30				.60

Tableau 2-8: Coefficient de détermination ρ^2 entre valeurs des F.P. numériques et analytiques selon la taille N de l'échantillon. N=∝ signifie que la matrice **C** a été remplie à l'aide du modèle et donc qu'il n'y a pas d'erreurs introduites par l'échantillonnage statistique.

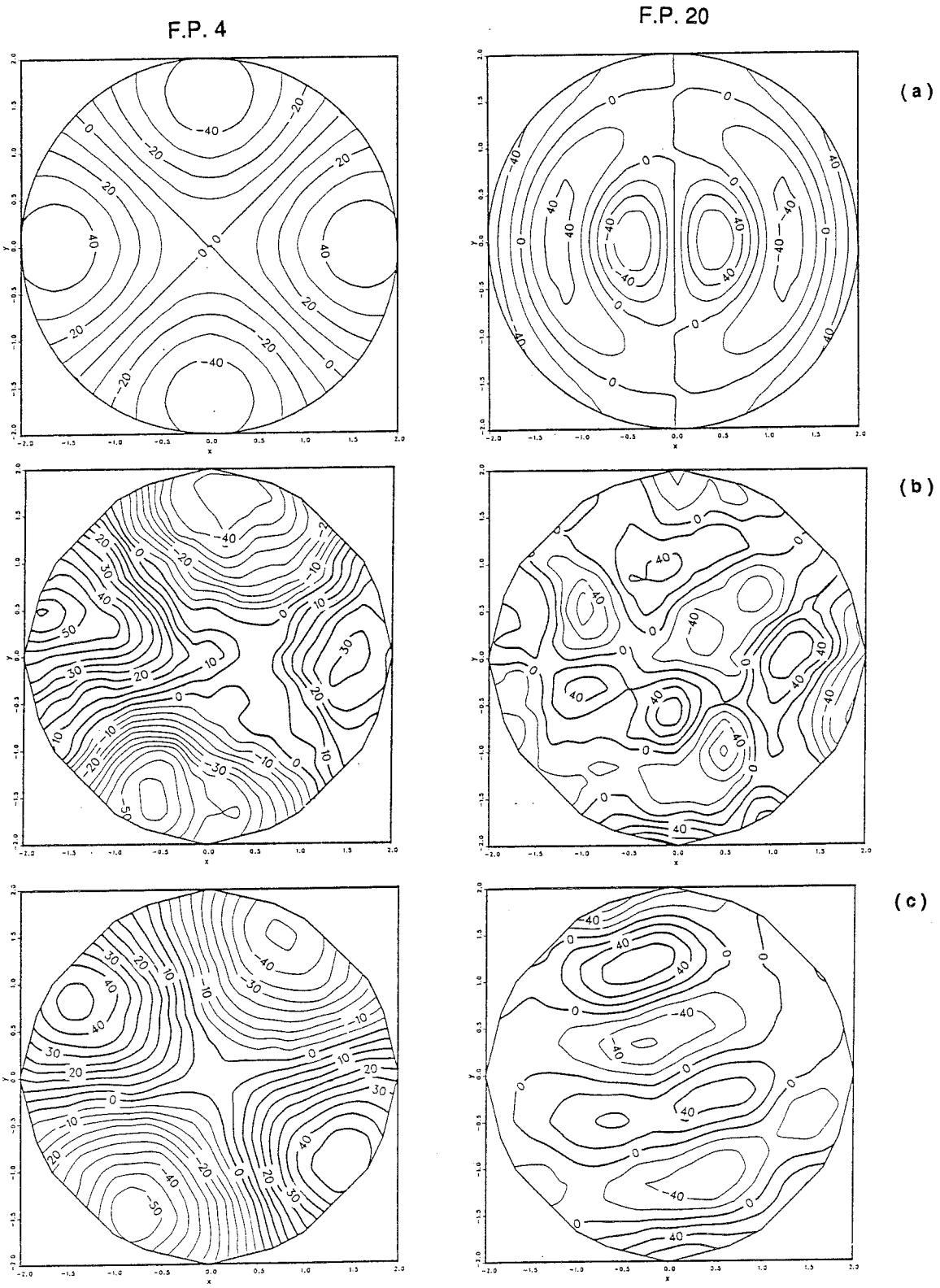
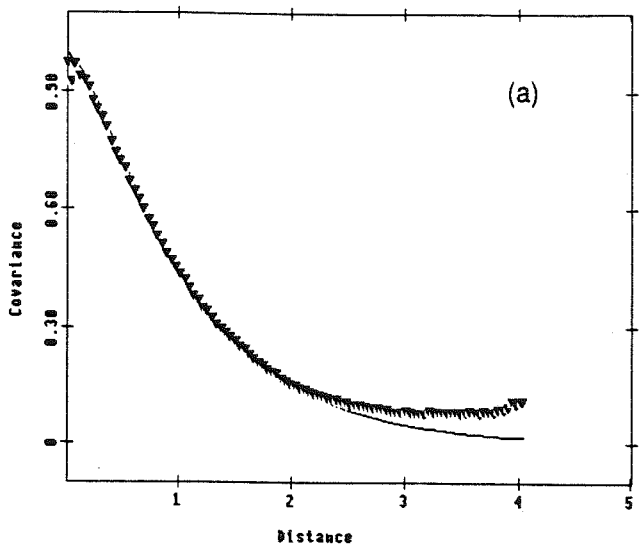
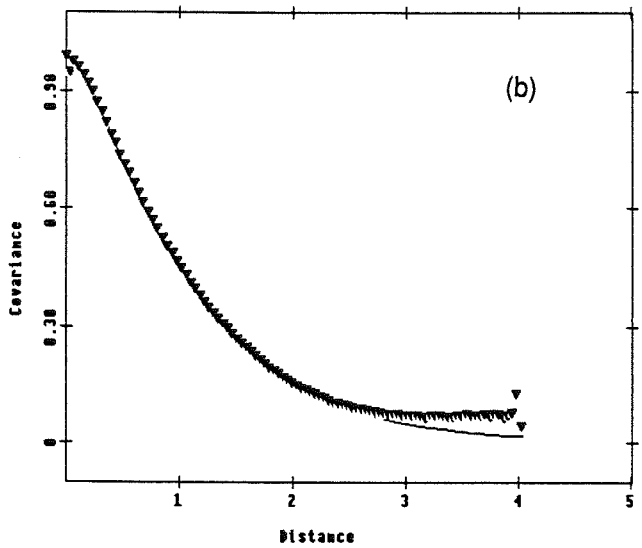


Fig. 2-12:Comparaison entre F.P. numériques et analytiques selon la taille N de l'échantillon.
Gauche: F.P. 4 Droite: F.P. 20
(a) Solution analytique, (b) N=100, (c) N=1000





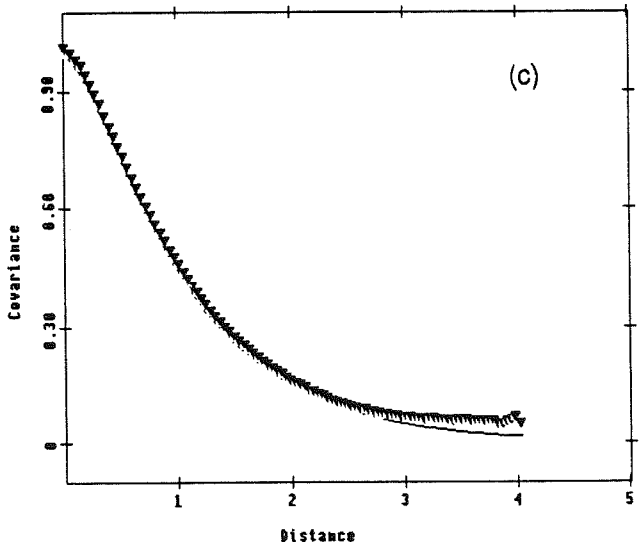


Fig. 2-13:Covariances théoriques ------ et expérimentales ▼ d'un jeu de N réalisations simulées par Bandes tournantes.
(a) N=100, (b) N=300, (c) N=1000.

inhomogénéités détectées. Puis, si ces tests ne sont pas concluants, la modélisation de la covariance et l'utilisation de ce modèle pour remplir la matrice **C**, pourraient améliorer la stabilité et la précision des fonctions propres estimées.

2-5 CONCLUSIONS

Dans ce deuxième chapitre, nous avons donc proposé une méthode de calcul des solutions analytiques de l'équation de Fredholm, pour des modèles de covariance homogènes, isotropes et à densité spectrale rationnelle, les calculs ayant été complètement détaillés pour le modèle proposé par Fortus (1975).

Nous nous sommes ensuite attachés à étudier les erreurs induites par la discrétisation spatiale du problème et l'échantillonnage statistique sur la précision des F.P. Il est clair que, en toute rigueur, les conclusions que nous pouvons proposer sont limitées au modèle et au domaine très particuliers étudiés. Nous espérons néanmoins que les remarques suivantes sont généralisables à toute une classe de modèles de comportement voisin:

i) L'échantillonnage statistique apparaît comme le facteur le plus pénalisant puisque l'estimation des covariances sur un nombre fini de réalisations perturbe fortement les F.P. et notamment réduit considérablement les capacités de la méthode à isoler les différentes valeurs propres. La modélisation préalable du corrélogramme pourrait permettre une plus grande stabilité des résultats, mais probablement en perdant des informations sur d'éventuelles inhomogénéités du processus.

ii) Au niveau de la discrétisation spatiale, si on a le choix, un maillage régulier, adapté à la forme du domaine d'étude semble préférable, avec une densité du réseau vérifiant à peu près: taille du domaine/ pas de discrétisation =20. Dans le cas où on ne maîtrise pas le réseau (stations météorologiques par exemple), et si ce dernier est plutôt irrégulier, seules les premières F.P. pourront éventuellement être interprétées.

Par ailleurs, nous avons vu que la densité du réseau devait aussi être adaptée à la portée du modèle, avec une augmentation lorsque la portée diminue. Nous avons en effet constaté qu'alors, nous avions besoin de plus de F.P. pour expliquer un pourcentage de variance donné.

iii) Le choix des fonctions de base $e_i(\underline{x})$ devient important s'il est nécessaire d'interpoler les F.P. en d'autre points que les noeuds du réseau de calcul. L'utilisation de fonctions linéaires par facettes (LPF) est alors préférable. Nous avons aussi noté que le calcul sur un réseau régulier, puis l'interpolation sur un réseau irrégulier donnait de meilleurs résultats qu'un calcul direct sur ce même réseau irrégulier. On privilégiera alors l'approche LPF.

Après ces quelques remarques, nous voulons souligner le point suivant. Si les recommandations précédentes sont respectées, on peut s'attendre à une bonne estimation des fonctions propres qui pourront alors être interprétées plus facilement. (Gardons cependant en mémoire les réserves émises, tout au long de ce chapitre, quant à une interprétation physique trop poussée....).

D'ailleurs, pour une utilisation ultérieure des EOFs (stockage de données, simulations de champs (voir chapitre suivant), interpolation (cf chapitre 1)), la bonne estimation de chacune des EOFs n'est pas forcément nécessaire. En effet, nous avons vu que certaines F.P. pouvaient se mélanger entre elles et donc être mal corrélées individuellement aux F.P. vraies. Mais, si on conserve toutes les F.P. correspondantes, on ne perdra pas de variance et donc l'utilisation ultérieure de ces F.P. en reconstitution ou en simulation pourra donner satisfaction.

On se retrouve à ce niveau confronté au problème, jamais complètement résolu, du choix du nombre de F.P. significatives. De nombreuses règles, plus ou moins empiriques, ont été proposées. Obled (1979), Preisendorder (1988) en décrivent un certain nombre, et les travaux de North et al. (1982), donnent aussi de bonnes indications. Notre étude, quant à elle, a, nous l'espérons, aidé à comprendre un peu la façon dont se dégradaient les F.P. et quelles étaient les principales sources d'erreur, même si nos conclusions ne peuvent avoir qu'une portée limitée par l'emploi d'un modèle et d'un domaine d'étude bien spécifiques.

CHAPITRE 3 : UTILISATION DE L'ACPP POUR LA SIMULATION DE CHAMPS BIDIMENSIONNELS

3-1 INTRODUCTION

Dans de nombreuses disciplines, les scientifiques sont confrontés à la grande variabilité spatiale et/ou temporelle des phénomènes qu'ils étudient. C'est vrai notamment pour les propriétés des aquifères ou des gisements de pétrole (perméabilité, transmissivité, etc..) (Delhomme, 1979; Mantoglou et Gelhar, 1987), mais aussi en Météorologie où la pluie est un phénomène extrêmement variable (Tourasse, 1981; Barancourt, 1990). Il en va de même en Géologie où les gisements miniers sont constitués de veines plus ou moins riches (Journel et Huijbregts, 1978) . On peut encore citer l'Océanographie où de nombreux auteurs se penchent sur la variabilité des champs de vent (Barnier, 1986) ou de la température de surface de la mer (Servain et al., 1985) et cette liste est loin d'être exhaustive.... (Voir aussi Mantoglou, 1987).

Pour appréhender cette variabilité, on ne dispose, en général, que de points de mesure répartis plus ou moins régulièrement. Or, les grandeurs en question constituent souvent les entrées, ou des paramètres de modèles numériques, reproduisant les équations aux dérivées partielles des écoulements atmosphériques ou océaniques ou encore les cheminements de l'eau ou de polluants dans les sols. Il faut alors affecter une valeur à ces variables partout, ou plus exactement aux noeuds d'une grille, voire une moyenne sur la maille. D'où la nécessité d'interpoler. Mais, les méthodes traditionnelles que nous avons citées au chapitre 1: interpolation polynomiale ou spline, krigeage, etc.. (§1-5) fournissent en général des champs trop lisses et les modèles sont alors incapables de reproduire des phénomènes de grande échelle produits par des variations locales des paramètres (Delhomme, 1979).

Par ailleurs, connaissant les statistiques des entrées de modèles, on peut s'intéresser à l'influence de leur variabilité sur les sorties. On fait alors tourner le modèle un grand nombre de fois avec, en entrée, des champs différents reproduisant cette variabilité : c'est la simulation de Monte-Carlo. Une telle approche a souvent été utilisée dans les différentes disciplines citées plus haut: Meija et Rodriguez-Iturbe (1974) en Météorologie, De Fouquet et al. (1989) dans le domaine minier ou encore Shinuzoka (1971), Kettani-Wagner (1986) pour l'étude de la réponse de structures métalliques à des sollicitations aléatoires de

vent. Muñoz-Pardo (1987) ou Krajewsky et Dufy (1988) se sont quant à eux attachés, grâce à la simulation, à quantifier les erreurs de modélisation des moyennes, des variances ou des fonctions de structure de champs aléatoires.

Pour de telles études, nous avons vu que les champs interpolés ne suffisent pas, mais qu'il faut pouvoir fournir des champs reproduisant bien les petites variations locales, ces champs pouvant parfois être anisotropes ou non homogènes. Pour d'autres applications, on peut aussi avoir besoin de fournir en entrée au modèle, une valeur moyenne du processus sur un domaine donné, par exemple la lame d'eau moyenne sur un bassin, qui sert d'entrée à un modèle pluie-débit.

De nombreuses méthodes de simulation ont été proposées pour répondre à ces besoins. Nous allons essayer d'en donner une vue d'ensemble dans la première partie de ce chapitre avant de voir comment l'analyse EOF peut apporter sa contribution à un tel problème.

3-2 TECHNIQUES DE SIMULATION.

Les méthodes de simulation existantes sont assez nombreuses, mais peuvent être regroupées en quatre grandes familles:

i) les méthodes se basant sur l'analyse spectrale des processus (Shinuzoka, 1971; Shinuzoka et Jan, 1972; Meija et Rodriguez -Iturbe, 1974)

ii) les méthodes " géométriques" (Alfaro, 1979)

iii) la méthodes des bandes tournantes (Matheron, 1973) et ses variantes.

iv) les méthodes matricielles utilisant une décomposition particulière de la matrice de covariance (Alabert, 1987; Davis, 1987-a, 1987-b)

Toutes ces méthodes nécessitent une bonne connaissance, souvent analytique, de la fonction de covariance du processus simulé et peuvent être appliquées à des variables quelconques, mais stationnaires et continues. Pour des types de données particuliers, des modèles spécifiques qui en reproduisent les caractéristiques peuvent bien sûr être construits. La modélisation des champs de précipitation proposée par Amorocho et Wu (1977) ou Bras et Rodriguez-Iturbe (1976) en est un bon exemple. On peut aussi citer les modèles booléens qui sont adaptés à la simulation de champs de variables discontinues (la pluie à faible pas de temps, un gisement minier,etc,...)

Dans ce paragraphe, nous omettrons l'indice ξ des événements, étant entendu que nous parlerons de la simulation d'<u>un</u> champ aléatoire bidimensionnel particulier. Nous nous placerons par ailleurs dans le cas de <u>champs stationnaires</u>, où la covariance peut s'écrire:

$$C(\underline{x},\underline{x}') = C(\underline{x}-\underline{x}') = C(\underline{h})$$

Nous envisagerons cependant, au § 3-6, les champs non-stationnaires.

On définit par ailleurs, comme au § 2-2-1, la densité spectrale $S(\underline{w})$ d'une covariance stationnaire, $\underline{w}$ étant le vecteur des fréquences, comme la transformée de Fourier inverse dans R^2, de la fonction de covariance. Cette fonction peut encore s'interpréter comme la fonction de répartition de l'énergie totale du signal entre toutes les fréquences $\underline{w}$.

$$S(\underline{w}) = T^{-1}[\, C(\underline{h})\,](\underline{w}) = 1/(2\Pi)^2 \,.\, \int_{R2} C(\underline{h}) \,.\, \exp(\, -i\underline{h}.\underline{w}) \,.\, d\underline{h}$$

où $\underline{h} \,.\, \underline{w}$ est le produit scalaire des vecteurs $\underline{h}$ et $\underline{w}$ dans R^2. Réciproquement, on a aussi:

$$C(\underline{h}) = T[\, S(\underline{w})\,](\underline{h}) = \int_{R2} S(\underline{w}) \,.\, \exp(\, i\underline{h}.\underline{w}) \,.\, d\underline{w}$$

Dans toute la suite de ce chapitre, $X_s(\underline{x})$ (ou $X_s(\xi,\underline{x})$) désignera le processus simulé ou une de ses réalisations.

3-2-1 Méthodes spectrales.

Par analogie avec les décompositions en séries de Fourier, le champ simulé s'écrit de façon générique:

$$X_s(\underline{x}) = \sum_{k=1}^{M} a(\underline{w}_k) \,.\, \cos(\, \underline{w}_k \,.\, \underline{x} + \Phi_k) \qquad (3\text{-}1)$$

Dans l'approche proposée par Shinuzoka (1971) ou Meija et Rodriguez-Iturbe (1974), les $\underline{w}_k$ sont des fréquences aléatoires dont la densité de probabilité est donnée par $S(\underline{w}) \,/\, \sigma^2$ où $S(\underline{w})$ est la densité spectrale associée à la covariance cible $C(\underline{h})$ (dont nous avons rappelé la définition au-dessus) et σ^2 la variance du processus. Φ_k est une phase aléatoire tirée dans une loi uniforme sur $[\,0,\, 2\Pi]$. Enfin, les $a(\underline{w}_k)$ sont ici constants et égaux à $\sigma^2 .\, (2/M)^{1/2}$. Notons que, bien que l'amplitude associée à chaque harmonique $a(\underline{w}_k)$ soit constante, le processus simulé respectera la densité spectrale cible $S(\underline{w})$, puisqu'une même fréquence $\underline{w}_k$ sera tirée proportionnellement à son poids dans la répartition de l'énergie totale, c'est-à-dire avec une densité de probabilité $S(\underline{w}_k)/\,\sigma^2$.

On peut alors montrer(Shinuzoka, 1971 ou Meija et Rodriguez-Iturbe, 1974) que le processus simulé possède bien, en espérance, la covariance cible, mais, en revanche, il n'est qu'asymptotiquement ergodique, c'est-à-dire lorsque M tend vers l'infini. Nous rappelons que, pour un processus stationnaire et ergodique, moyenne spatiale dans la réalisation et espérance sur un grand nombre de réalisations sont confondues.

Shinuzoka (1971) montre néanmoins que la convergence est assez lente (en $1/\sqrt{M}$). Aussi, il a proposé une amélioration de cette technique (Shinuzoka et Jan, 1972). Le processus simulé est toujours une série de cosinus, mais, cette fois, les fréquences $\underline{w}_k$ sont choisies à priori et régulièrement espacées. On suppose en effet que le spectre $S(\underline{w})$ est pratiquement nul en dehors d'une plage de fréquences $\underline{w}_k > \underline{\Omega}$ (i.e. si $\underline{w}=(w_1,w_2)$, $\underline{\Omega}=(\Omega_1,\Omega_2)$ pour $w_i>\Omega_i$, i=1,2). L'espace des fréquences est discrétisé avec un pas $\Delta\underline{w}=(\Delta w_1,\Delta w_2)$. La densité spectrale est, elle aussi, discrétisée et supposée constante et égale à $S(\underline{w}_k)$ sur le pavé $(\Delta w_1,\Delta w_2)$. Les amplitudes $a(\underline{w}_k)$ sont alors données par la formule approchée:

$$a(\underline{w}_k) = [\, 4 \,.\, S(\underline{w}_k) \,.\, \Delta w_1 \,.\, \Delta w_2 \,]^{1/2} \qquad (3\text{-}2)$$

Finalement $X_s(\underline{x})$ s'écrit:

$$X_s(\underline{x}) = \sum_{k=1}^{M} [\, 4 \,.\, S(\underline{w}_k) \,.\, \Delta w_1 \,.\, \Delta w_2 \,]^{1/2} \,.\, \cos(\, (\underline{w}_k + \delta\underline{w}) . \underline{x} + \Phi_k) \qquad (3\text{-}3)$$

Les Φ_k sont définies comme précédemment et $\delta\underline{w} = (\delta w_1,\delta w_2)$, tel que $\delta w_i << \Delta w_i$, est un incrément aléatoire ajouté pour éviter une périodicité de $X_s(\underline{x})$.

Encore une fois, $X_s(\underline{x})$ possède, en espérance, la covariance cible. Il est ergodique pour tout M et converge beaucoup plus rapidement qu'avant: en $1/M^2$.

Notons que, dans tous les cas, le processus simulé est gaussien (ce qui se déduit du théorème de la limite centrale).

3-2-2 Méthodes géométriques.

Ces méthodes sont plus spécifiquement adaptées à la simulation de processus stationnaires et isotropes dans R^2.

Pour des modèles de covariance

exponentiels $C(r) = \sigma^2 \,.\, \exp(\, -r/a_e)$ avec $r=\|\, \underline{h} \,\|$

ou *sphériques*: $\begin{cases} C(r) = \sigma^2 \,.\, [\, 1 + 1/2\, (r/a_s)^3 - 3/2\, (r/a_s)\,] & \text{si } r \leq a_s \\ C(r) = 0 & \text{si } r > a_s \end{cases}$

Alfaro (1979) propose la méthode des jetons aléatoires. On implante dans le plan des cercles dont les diamètres sont des variables aléatoires indépendantes de densité de probabilité f(r), reliée à la dérivée seconde de la fonction de covariance cible C(r). Les centres des cercles sont répartis selon un processus de Poisson. On affecte un nombre aléatoire a_i à chaque

cercle, et, en un point $\underline{x}$, le processus simulé est obtenu par sommation des a_i associés à tous les cercles contenant le point $\underline{x}$ (voir Fig. 3-1).

Pour la simulation de la covariance exponentielle dans R^2, on peut aussi utiliser la méthode des polyèdres poissoniens (Alfaro, 1979).

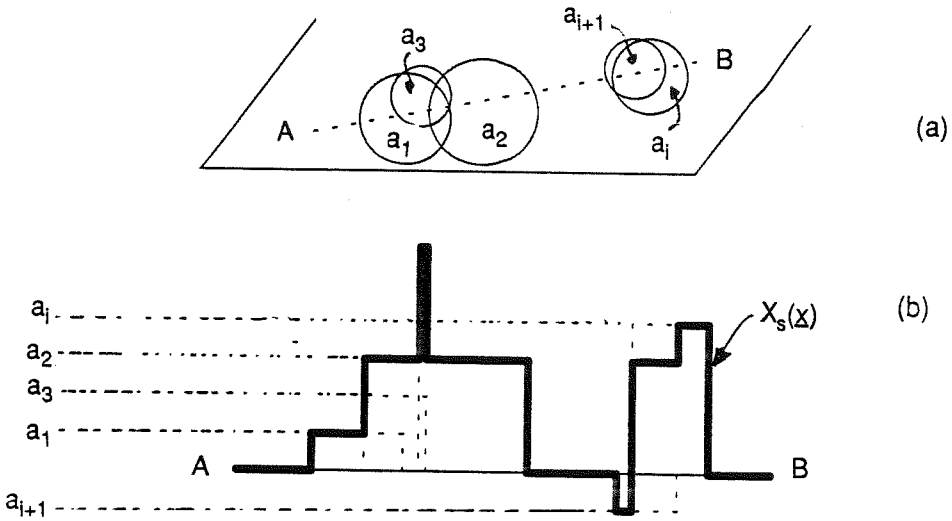


Fig. 3-1: Illustration de la méthode des jetons aléatoires
(a) Implantation des cercles
(b) Calcul des valeurs de $X_s(\underline{x})$ sur la ligne AB.

3-2-3 La méthode des bandes tournantes.

Cette méthode, initialement proposée par Matheron (1973) permet de simuler des fonctions aléatoires stationnaires et isotropes (Elle peut aussi être adaptée pour des champs anisotropes ou non homogènes (Mantoglou et Wilson, 1981; Tompson et al., 1989)).

La simulation dans R^2 ou R^3, est ramenée à une simulation dans R sur des lignes issues d'une même origine arbitraire et régulièrement réparties dans le plan ou l'espace (voir la Fig. 3-2 pour R^2). Christakos (1987) a montré que la méthode des bandes tournantes était en fait un cas particulier de projection de Radon, renforçant ainsi les bases théoriques de cette méthode dont la robustesse avait par ailleurs été prouvée par les maintes utilisations auxquelles elle avait donné lieu.

Pour obtenir la valeur simulée en un point $\underline{x}$, on fait la somme de L processus monodimensionnels $X_s^1(\eta_l)$, L étant le nombre de lignes. $\eta_l = \underline{x} \cdot \underline{u}_l$ est obtenu par projection orthogonale de $\underline{x}$ sur la lième ligne dont $\underline{u}_l$ est un vecteur directeur (voir Fig. 3-2).

$$X_s(\underline{x}) = 1/L \cdot \sum_{l=1}^{L} X_s^1(\underline{x} \cdot \underline{u}_l) \qquad (3\text{-}4)$$

En pratique, la simulation monodimensionnelle est effectuée de façon discrète, avec un pas $\Delta\eta$, si on note η l'abscisse dans R.

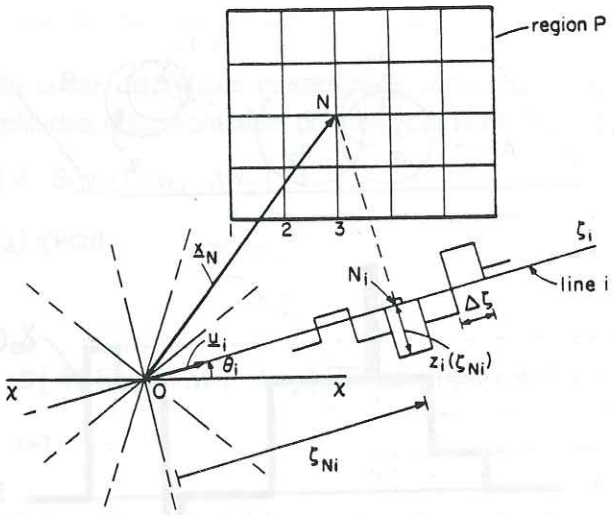


Fig. 3-2: Représentation schématique de la méthode des bandes tournantes. (d'après Mantoglou et Wilson, 1982)

Pour pouvoir simuler dans R, il faut bien sûr préciser la fonction de covariance $C^1(r)$ du processus monodimensionnel qui nous permettra de retrouver la covariance cible C(r) dans R^2 ou R^3. On peut montrer que ces deux fonctions sont liées par: (Matheron, 1973; Mantoglou et Wilson , 1982)

$$C^1(r) = d/dr\,(r \cdot C(r)) \qquad \text{dans } R^3$$

$$C(r) = 2/\Pi \cdot \int^{r} C^1(\eta)/(r^2-\eta^2) \cdot d\eta \qquad \text{dans } R^2$$

On voit immédiatement que, si dans R^3, la relation est simple, dans R^2, on n'a qu'une relation implicite. Mantoglou et Wilson (1981,1982) lèvent cette difficulté en passant dans le domaine des fréquences où, même dans R^2, les densités spectrales des processus mono et bidimensionnels, $S_1(w)$ et $S(\underline{w})$, sont liées par une relation simple. En particulier, pour une fonction de covariance isotrope, nous avons vu au chapitre 2 (formule (2-7)), que $S(\underline{w})$ ne dépendait plus que de $w = \| \underline{w} \| = (w_1^2 + w_2^2)^{1/2}$. Dans ce cas, (Mantoglou et Wilson, 1981, 1982), nous avons $S_1(w) = \Pi.w \cdot S(w)$.

Connaissant la densité spectrale du processus monodimensionnel, on peut alors facilement utiliser les méthodes spectrales décrites au § 3-2-1 pour le générer. Muñoz-Pardo (1987) propose par ailleurs deux autres méthodes pour simuler $X_s^1(\eta)$:

$$X_s^1(\eta) = \sum_{k=1}^{M} a_k^1(w_k) \cdot \cos(w_k \cdot \eta + \Phi_k)$$

avec i) un calcul exact des $a_k^1(w_k)$:

$$a_k^1(w_k) = [\, 4 \int_{w_k - \Delta w/2}^{w_k + \Delta w/2} S_1(w) \cdot dw \,]^{1/2} \qquad (3\text{-}5)$$

qui donne des résultats comparables à ceux obtenus par Shinuzoka et Jan (1972) avec la formule approchée (3-2).

ii) une approche "équiprobable" où on donne aux $a_k^1(w_k)$ une valeur fixée à priori et où on calcule, par itérations successives, le Δw qui vérifiera (3-5). Cette dernière méthode est plus précise (car on a un pas plus fin près de l'origine des fréquences) et plus rapide que les précédentes. (Muñoz-Pardo, 1987)

Lorsque la covariance monodimensionnelle peut se factoriser sous la forme:

$$C^1(\underline{h}) = \int_0^{\infty} g(\underline{u}) \cdot g(\underline{u} + \underline{h}) \cdot d\underline{u}$$

(ce qui est le cas pour les modèles sphériques et exponentiels), on peut utiliser une simulation par moyenne mobile (Moving Average) (Journel et Huijbregts, 1978; Christakos, 1987)

$$X_s^1(\eta) = \sum_{k=-M}^{M} g(k\Delta\eta) \cdot v(\eta + k)$$

où les $v(\eta + k)$ sont des réalisations de variables aléatoires $V(\eta)$, indépendantes, de même distribution, de moyenne nulle et de variance donnée. Cette méthode est assez rapide et peu coûteuse.

On peut encore, pour les covariances tendant vers zéro à l'infini, utiliser des processus autorégressifs (AR) (Christakos, 1987; Boulanger, 1989)

$$X_s^1(\eta) = \sum_{k=1}^{M} \gamma_k \cdot X_s^1(\eta - k) + \omega_\eta$$

où ω_η est un bruit blanc gaussien, de variance à déterminer et les γ_k des coefficients, devant, eux-aussi, être calculés en fonction du modèle de covariance du processus. Cette technique peut être directement utilisée pour une simulation dans R^2 et, d'après Boulanger (1989), elle permet d'éviter le passage de $C(\underline{h})$ à $C^1(\underline{h})$, le non respect du comportement à l'origine des méthodes spectrales ou la discrétisation sur les lignes de la méthode des bandes tournantes. Néanmoins, elle devient vite très lourde si le nombre de points à simuler est grand.

Enfin, Bouhaddou (1984) propose, pour les processus gaussiens $X_s^1(\eta)$ monodimensionnels, à densité spectrale rationnelle d'utiliser la représentation markovienne suivante de ces processus:

$$\begin{cases} X_s^1(\eta) = \mathbf{H} \,.\, Z(\eta) \\ dZ(\eta) = \mathbf{A} \,.\, Z(\eta) \,.\, d\eta + \mathbf{B} \,.\, dW(\eta) \end{cases}$$

où **H, A, B** sont des matrices ligne, carrée et colonne respectivement et $W = \{ W(\eta), \eta \in R \}$, un processus à accroissements orthogonaux avec $E[(W(\eta)-W(\eta'))^2] = |\eta - \eta'|$. Comme $X_s^1(\eta)$ est supposé gaussien, W est un mouvement brownien et Z un processus de Markov dont la probabilité de transition est assez simple (Bouhaddou, 1984).

3-2-4 Les méthodes matricielles.

Ces méthodes utilisent directement le fait que la simulation ne soit réalisée qu'en un nombre fini P de points, ce qui les autorise à travailler avec la matrice de covariance **C** plutôt qu'avec la fonction de covariance elle-même et à chercher des décompositions particulières de cette matrice, sachant qu'elle est symétrique, définie positive.

i) Davis (1987-a) ou Alabert (1987) utilisent la décomposition de Cholesky de **C** sous forme du produit d'une matrice triangulaire inférieure (L=lower) et supérieure (U=upper). On parle de méthode LU de simulation. On considère en effet un vecteur aléatoire $\mathbf{W}^T$, formé de P variables indépendantes w_i tirées dans une loi normale $N(0,1)$. Le vecteur simulé est alors:

$$\mathbf{X_s}^T = \mathbf{L} \,.\, \mathbf{W}$$

où on note $\mathbf{X_s}^T = [\, X_s(\underline{x}_1), \ldots\ldots, X_s(\underline{x}_P) \,]$

$\mathbf{W}^T = [\, w_1, \ldots\ldots, w_P \,]$

On montre, qu'en espérance, la matrice de covariance du processus X_s est bien la matrice **C** cible.

ii) Davis (1987-b) propose de calculer la "racine carrée" **G** de **C**, telle que $\mathbf{C}=\mathbf{G}^2$, et où $\mathbf{G} = \mathbf{H} . \Gamma^{1/2} . \mathbf{H}^T$, **H** étant la matrice orthogonale des vecteurs propres de **C** et Γ, la matrice diagonale des valeurs propres correspondantes. Diagonaliser **C** est très lourd, c'est pourquoi une approximation polynomiale de **G** est proposée. La méthode de simulation est ensuite analogue à la précédente, en remplaçant **L** par **G**.

$$\mathbf{X_s}^T = \mathbf{G} . \mathbf{W}$$

Ce paragraphe nous a permis de voir que de nombreuses méthodes de simulation existent déjà. Nous voudrions y ajouter l'utilisation de l'analyse EOF. Mais, on peut se demander quel est l'intérêt d'une technique supplémentaire. Nous le justifierons par les remarques suivantes. En Océanographie ou en Météorologie, une ACP est en général effectuée pour condenser les tableaux de variables (pression, température de la mer, vent,..), mais surtout comprendre les phénomènes sous-jacents (Barnier, 1986; Déqué, 1986; Servain et Legler, 1986; Mac Veigh et al., 1987). Par ailleurs, pour l'étude de la sensibilité des modèles de circulation générale ou d'évolution des climats aux conditions initiales, il est nécessaire de disposer de champs cohérents, respectant la structure de ces mêmes variables. Il peut alors sembler séduisant d'utiliser directement les résultats de l'ACP, effectuée de toute manière, pour produire ces champs.

Dans la suite de ce chapitre, nous commencerons donc par voir comment la technique d'ACPP peut être adaptée à la génération de champs bidimensionnels. Puis, nous présenterons les sources potentielles d'erreur, susceptibles d'introduire des biais entre covariance cible et covariance simulée, et comment un choix judicieux des paramètres de la simulation peut les réduire. Une comparaison avec la méthode des bandes tournantes et la méthode LU sera ensuite effectuée, avant de terminer par la présentation d'un exemple synthétique d'extension à des champs hétérogènes (voir aussi Braud et Obled, 1990).

3-3 UTILISATION DE L'ACPP POUR LA SIMULATION DE CHAMPS BIDIMENSIONNELS.

Nous cherchons donc à simuler, sur un domaine D donné, un processus $X(\xi,\underline{x})$, supposé de moyenne nulle, dont on se donne la fonction de covariance $C(\underline{x},\underline{x}')$. Le processus simulé, noté $X_s(\xi,\underline{x})$, issu de l'analyse EOF aura la forme suivante:

$$X_s(\xi,\underline{x}) = \sum_{k=1}^{K} Z_k(\xi) \,.\, F_k(\underline{x}) + \varepsilon(\xi,\underline{x}) \qquad (3\text{-}6)$$

Explicitons maintenant les différents termes de cette décomposition.

i) Comme on se donne le modèle de covariance et le domaine d'étude, les $F_k(\underline{x})$ peuvent être calculées directement en résolvant (1-5):

$$\int_D C(\underline{x},\underline{x}') \,.\, F_k(\underline{x}') \,.\, d\underline{x}' = \lambda_k \,.\, F_k(\underline{x}) \qquad (1\text{-}5)$$

Notons que, contrairement à la plupart des méthodes présentées au § 3-2, l'hypothèse d'invariance de la covariance par translation n'est pas nécessaire, puisque, en théorie, (1-5) peut être résolue pour tout modèle $C(\underline{x},\underline{x}')$.

En pratique, bien sûr, les F.P. seront calculées sur une grille, à l'aide de la méthode décrite au § 1-3. Compte-tenu des remarques du chapitre précédent, on choisira un maillage plutôt régulier, de densité compatible avec la taille du domaine et la portée du modèle de covariance.

ii) Les $Z_k(\xi)$ constituent les éléments "aléatoires" de la simulation. Nous en connaissons les moyennes (nulles, comme $X(\xi,\underline{x})$ est centré), et les variances λ_k. Nous savons par ailleurs que les $Z_k(\xi)$ sont décorrélées. Néanmoins, pour pouvoir générer ces variables, il faudrait aussi connaître les distributions marginales de chacune d'elle, voire même leurs distributions conjointes. En pratique, on peut avoir accès aux distributions des $Z_k(\xi)$, ou du moins à une estimation, si on dispose d'un échantillon de réalisations de données réelles.

Pour les essais présentés par la suite, nous avons retenu des lois gaussiennes pour les $Z_k(\xi)$, ce qui conduit à des distributions gaussiennes de $X(\xi,\underline{x})$. Des lois uniformes ont aussi été testées. Nous y reviendrons au paragraphe suivant.

iii) Enfin, $\varepsilon(\xi,\underline{x})$ est un bruit blanc gaussien, de moyenne nulle, décorrélé des $Z_k(\xi)$, dont nous allons maintenant préciser la variance.

La covariance du processus simulé s'écrit:

$$\begin{aligned} C_s(\underline{x},\underline{x}') &= E\,[\, X_s(\xi,\underline{x}) \,.\, X_s(\xi,\underline{x}') \,] \\ &= \sum_{k=1}^{K} \sum_{m=1}^{K} E\,[\, Z_k(\xi) \,.\, Z_m(\xi) \,] \,.\, F_k(\underline{x}) \,.\, F_m(\underline{x}') + E\,[\, \varepsilon(\xi,\underline{x}) \,.\, \varepsilon(\xi,\underline{x}') \,] \end{aligned}$$

Les $Z_k(\xi)$ sont décorrélées (cf (1-6)) et ici, elles le seront par construction, donc:

$$C_s(\underline{x},\underline{x}') = \sum_{k=1}^{K} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') + E[\,\varepsilon(\xi,\underline{x}) \cdot \varepsilon(\xi,\underline{x}')\,]$$

$$= \sum_{k=1}^{\infty} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') - \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') + E[\,\varepsilon(\xi,\underline{x}) \cdot \varepsilon(\xi,\underline{x}')\,]$$

$$= C(\underline{x},\underline{x}') - \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') + E[\,\varepsilon(\xi,\underline{x}) \cdot \varepsilon(\xi,\underline{x}')\,]$$

En effet, la covariance vraie $C(\underline{x},\underline{x}')$ peut être écrite:

$$C(\underline{x},\underline{x}') = E[\,X(\xi,\underline{x}) \cdot X(\xi,\underline{x}')\,] = \sum_{k=1}^{\infty} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') \qquad (3\text{-}7)$$

Pour qu'il y ait égalité stricte entre la covariance du processus simulé et la covariance cible, il suffirait donc d'avoir:

$$E[\,\varepsilon(\xi,\underline{x}) \cdot \varepsilon(\xi,\underline{x}')\,] = \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') \qquad (3\text{-}8)$$

Mais alors, $\varepsilon(\xi,\underline{x})$ ne serait plus, dans le cas général, un bruit blanc dans l'espace et ça reviendrait à retenir toutes les F.P. dans la décomposition. Or, si on cherche à tronquer à K termes la série, c'est par souci de simplification et d'économie, sachant que la plus grande partie de la variance est concentrée sur les premières F.P. Mais, <u>cette simplification se fera au prix d'une approximation</u>. Nous supposerons en effet <u>que (3-8) est voisin de zéro pour $\underline{x} \neq \underline{x}'$</u>, la valeur pour $\underline{x}=\underline{x}'$ nous fournissant quant à elle la variance de $\varepsilon(\xi,\underline{x})$. Finalement, nous écrirons:

$$E[\,\varepsilon(\xi,\underline{x}) \cdot \varepsilon(\xi,\underline{x}')\,] = \delta(\underline{x}-\underline{x}') \cdot \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x})^2 \qquad (3\text{-}9)$$

ce qui donnera pour la covariance de X_s:

$$C_s(\underline{x},\underline{x}') = \sum_{k=1}^{K} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') + \delta(\underline{x}-\underline{x}') \cdot \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x})^2 \qquad (3\text{-}10)$$

On voit par ailleurs immédiatement que:

$$\lim_{K \to \infty} C_s(\underline{x},\underline{x}') = C(\underline{x},\underline{x}')$$

On peut aussi montrer (Annexe B-2) que $C_s(\underline{x},\underline{x}')$ est bien une fonction de type positif au sens de Matheron (1972-a) ou Christakos (1984).

Nous nous intéresserons plus particulièrement aux effets de la troncature sur $C_s(\underline{x},\underline{x}')$ dans la suite de ce chapitre. Mais notons néanmoins que le choix retenu pour var[$\varepsilon(\xi,\underline{x})$] nous permet d'assurer que pour tout K et pour tout $\underline{x}$, les variances simulées et théoriques coïncident, du moins en espérance.

Pour une application pratique, les $F_k(\underline{x})$ intervenant dans les différents termes de la fonction de covariance simulée seront calculées en utilisant l'approximation du § 1-3, c'est-à-dire en introduisant les $e_i(\underline{x})$. Par ailleurs, si nous estimons les F.P. sur un réseau de P points de grille, nous ne pourrons en calculer que P et la sommation jusqu'à l'infini dans (3-9) (qui nous donne var[$\varepsilon(\xi,\underline{x})$]) se limitera à P termes.

Compte-tenu de la formulation proposée en (3-6) pour $X_s(\xi,\underline{x})$, on peut déjà se faire une idée des erreurs qui pourront introduire des biais entre le modèle et la covariance effectivement simulée:

i) Tout d'abord, la troncature de la série des F.P. aux K premiers termes, c'est-à-dire l'hypothèse de bruit blanc sur $\varepsilon(\xi,\underline{x})$ pourra n'être pas justifiée pour certaines valeurs de K ou certaines gammes d'interdistances.

ii) Une deuxième source d'erreur est, bien sûr, liée à la discrétisation spatiale du domaine et au fait que les F.P. seront calculées sur une grille de taille finie (cf chapitre 2).

iii) Le choix des distributions marginales des $Z_k(\xi)$ pourra aussi jouer un rôle "perturbateur" sur la covariance simulée.

C'est pourquoi nous allons maintenant étudier d'un peu plus près les influences respectives de la troncature et de la discrétisation spatiale, en ajoutant un certain nombre

de remarques sur l'influence de la distribution des $Z_k(\xi)$. Nous ne traiterons pas ici les erreurs statistiques susceptibles de s'introduire lors de l'étape, préliminaire à la simulation, qui est la modélisation de la fonction de covariance $C(\underline{x},\underline{x}')$ qui sera ensuite utilisée comme covariance cible.

3-4 SENSIBILITE DE LA SIMULATION AU CHOIX DES PARAMETRES.

Les essais ont été menés avec un modèle de covariance homogène, isotrope, de type exponentiel:

$$C(\underline{x},\underline{x}') = C(\| \underline{x}-\underline{x}' \|) = C(r) = \sigma^2 . \exp(-r/a_e) \qquad (3\text{-}11)$$

Nous avons vu au paragraphe précédent que les hypothèses d'homogénéité et d'isotropie de la covariance n'étaient pas nécessaires à la simulation par ACPP, mais elles permettent une étude de sensibilité plus aisée, car la comparaison modèle / simulation est plus simple et plus robuste dans ce cas. De plus, il n'existe pas de modèles analytiques non stationnaires.

Nous avons par ailleurs cherché à générer des variables gaussiennes, de moyenne nulle et de variance unité: $\sigma^2 = 1$. Le domaine d'étude est un carré de côté Ls. Les fonctions propres ont été calculées sur une grille carrée de pas Dx, formée de P points (Fig 3-3).

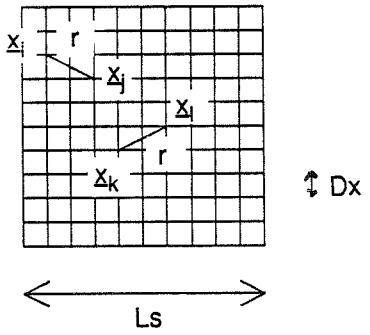


Fig 3-3: Type de grille utilisée pour le calcul des fonctions propres.
Carré de côté Ls, de pas Dx, formée de P points.

3-4-1 Choix du nombre K de fonctions propres retenues dans la décomposition.

Pour cette étude, nous n'avons pas tenu compte du bruit blanc, dont l'introduction aurait seulement ajouté un Dirac à l'origine, pour la fonction de covariance. Nous nous

intéressions en effet plus particulièrement à l'évolution de la forme de cette dernière en fonction du nombre K de termes intervenant dans la décomposition (3-6). Pour l'évaluer, nous avons calculé, pour tout couple de points ($\underline{x}_i$, $\underline{x}_j$), une covariance "empirique" en ne conservant que les K premières F.P.

$$C_s(\underline{x}_i, \underline{x}_j) = \sum_{k=1}^{K} \lambda_k \cdot F_k(\underline{x}_i) \cdot F_k(\underline{x}_j)$$

$$C_s(\underline{x}_i, \underline{x}_j) = \sum_{k=1}^{K} \lambda_k \cdot f_{ik} \cdot f_{jk} \qquad \text{puisque } e_i(\underline{x}_j) = \delta_{ij} \qquad (3\text{-}12)$$

(Nous rappelons que les f_{ik} sont les éléments de la matrice F , c'est-à-dire les valeurs discrètes des F.P. aux points de la grille de calcul).

Puis, pour l'ensemble des couples ($\underline{x}_i$, $\underline{x}_j$), distants de r (Fig 3-3), nous avons évalué la covariance expérimentale:

$$C_s(r) = 1/Nc(r) \cdot \sum_{k=1}^{Nc(r)} C_{sk}(\underline{x}_i, \underline{x}_j) \qquad (3\text{-}13)$$

où Nc(r) est le nombre de couples distants de r.

Ces courbes ont ensuite été comparées au modèle, pour plusieurs valeurs de la portée a_e du modèle.

Le tableau 3-1 donne les pourcentages de variance expliqués par K=20,40,60,80,100 F.P. selon les valeurs de a_e retenues (P=121). Les Fig. 3-4 et 3-5 illustrent l'évolution de $C_s(r)$ avec K pour deux cas étudiés: a_e = Dx/2 et a_e = 3Dx/2 respectivement (Ls/Dx=10).

ae \ K	20	40	60	80	100	121
Dx/2	29	49	65	81	92	100
3.Dx/2	60	74	83	90	96	100
3.Dx	78	86	91	95	98	100
5.Dx	86	92	95	97	99	100

Tableau 3-1: Pourcentages de variance expliqués par K F.P. pour différentes valeurs de la portée a_e du modèle de type exponentiel, Ls/Dx=10.

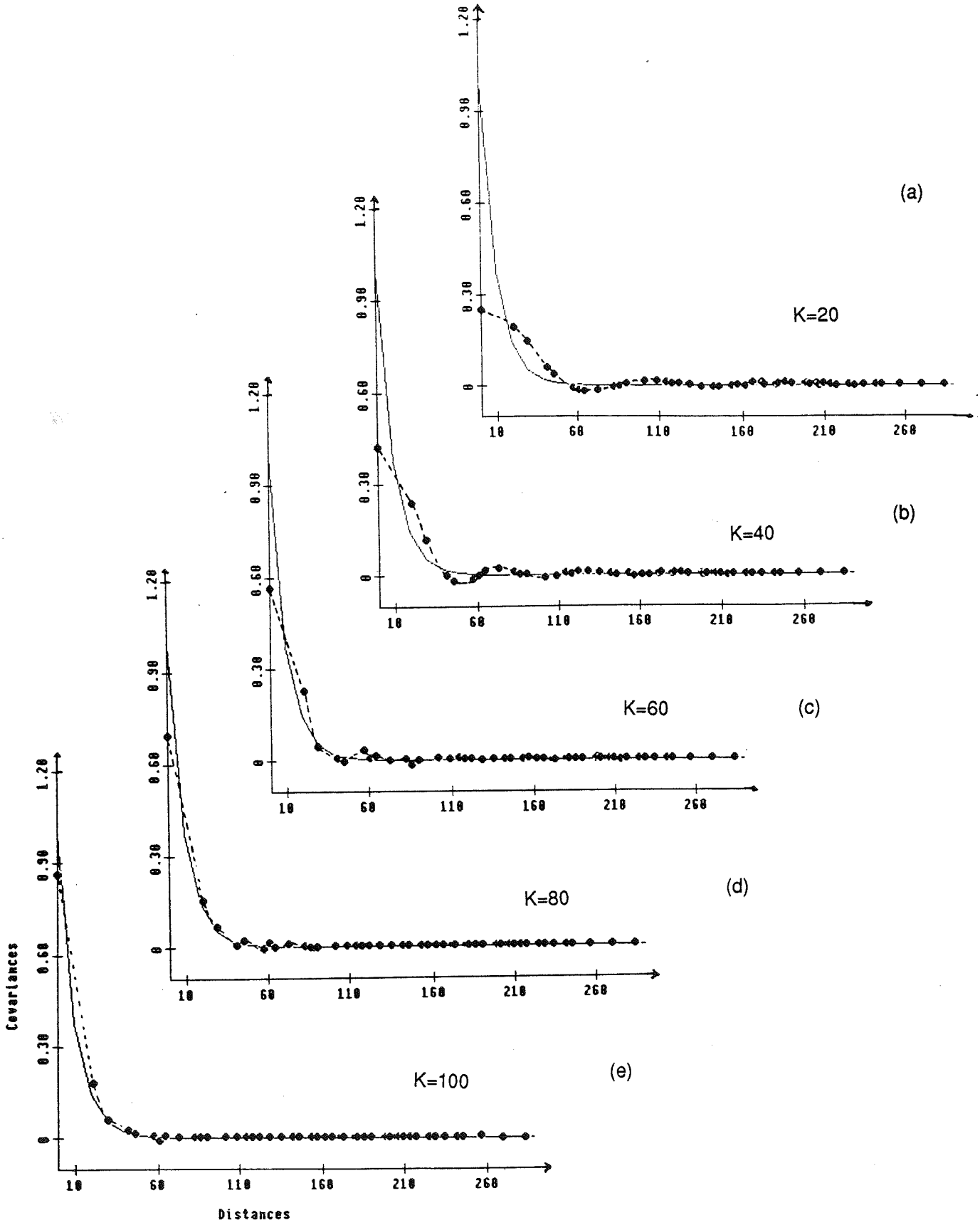


Fig. 3-4: Comparaison des covariances théoriques et expérimentales selon le nombre K de fonctions propres retenues. (Le bruit blanc n'a pas été ajouté).

—— Modèle de covariance C(r) = exp(-2r/Dx)
-◆- Covariance expérimentale avec (a) K=20, (b) K=40, (c) K=60, (d) K=80, (e) K=100

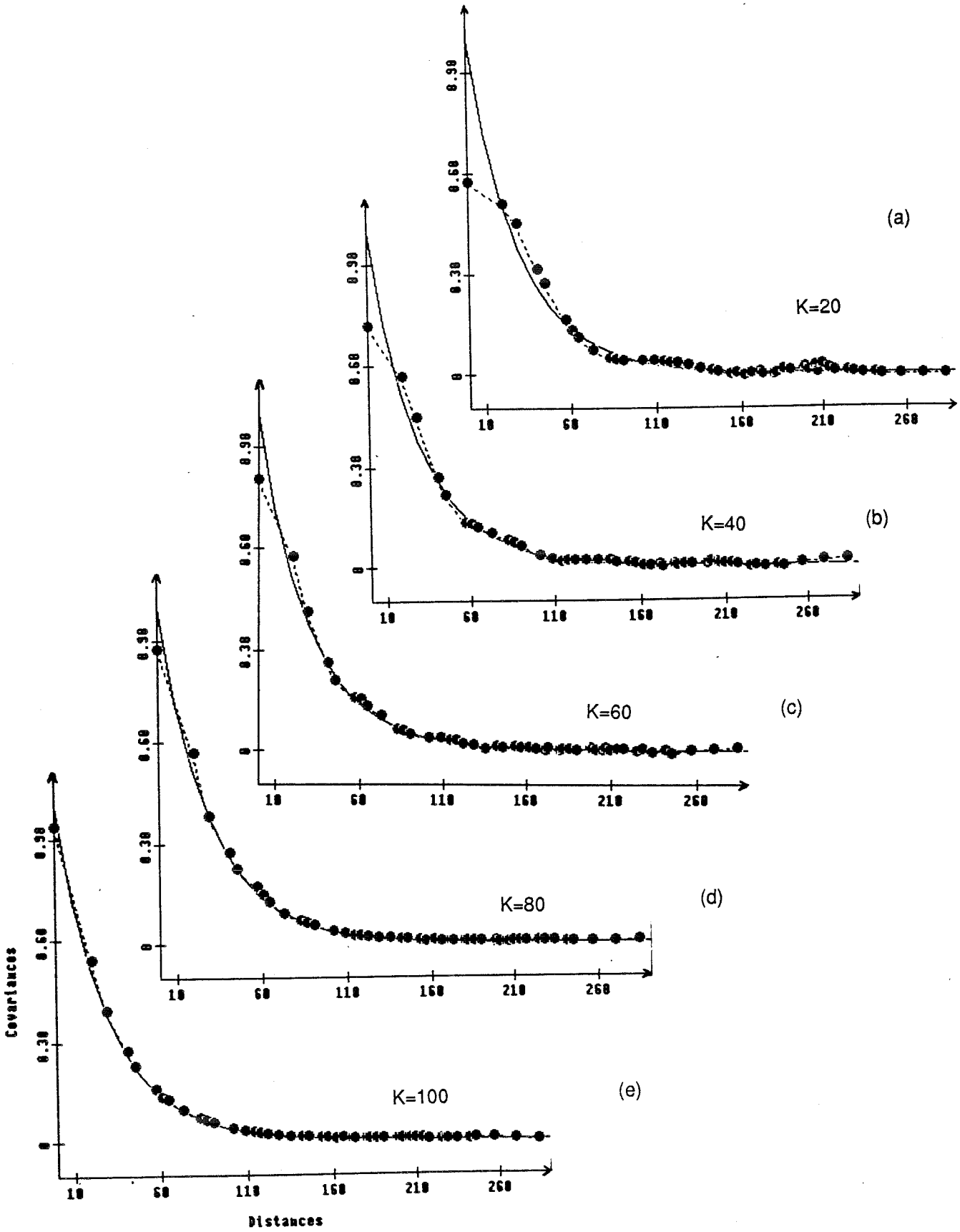


Fig. 3-5: Comparaison des covariances théoriques et expérimentales selon le nombre K de fonctions propres retenues. (Le bruit blanc n'a pas été ajouté).

Modèle de covariance C(r) = exp(-2r/3Dx)
Covariance expérimentale avec (a) K=20, (b) K=40, (c) K=60, (d) K=80, (e) K=100

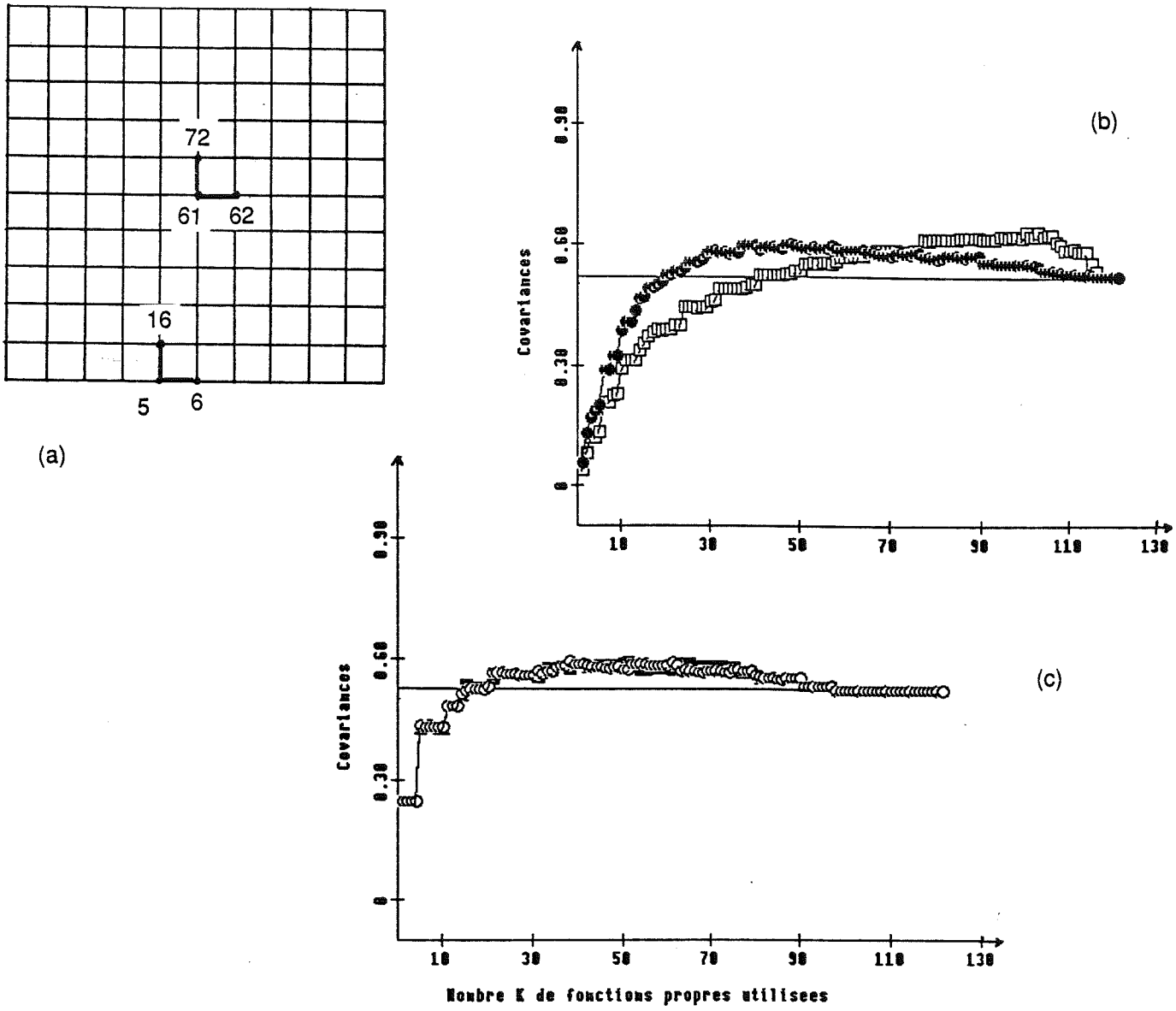


Fig. 3-6: Evolution de la covariance empirique en fonction du nombre K de F.P. utilisées pour 4 couples de points distants de Dx, définis sur la grille (a).

(b) — Valeur théorique C(Dx) = exp(-2/3), stations 5-6, stations 5-16
(c) — Valeur théorique C(Dx) = exp(-2/3), stations 61-62, stations 61-72

Le tableau 3-1 montre clairement que, pour expliquer un même pourcentage de variance, il faudra plus de F.P. lorsque le modèle aura une longueur de corrélation plus courte. La convergence de $C_s(r)$ vers $C(r)$ sera donc d'autant plus rapide que la portée sera longue, ce qu'illustrent les Fig. 3-4 et 3-5. On y voit par ailleurs que la troncature de la série a des effets essentiellement près de l'origine avec des oscillations et des inversions de courbure, lorsque K est petit. Ceci revient à dire que, si on tronque trop brutalement la série, l'hypothèse de bruit blanc sur $\varepsilon(\xi,x)$ n'est plus légitime, du moins pour les faibles interdistances. Avec 90% de variance expliquée, on peut estimer que cette hypothèse est justifiée.

Les Fig. 3-4 et 3-5 présentent néanmoins des résultats moyens sur un grand nombre de couples. Nous avons voulu voir, en choisissant arbitrairement quelques couples de points, si les propriétés d'homogénéité et d'isotropie étaient conservées, malgré la troncature. Les Fig. 3-6-b et c montrent l'évolution de $C_s(\underline{x}_i, \underline{x}_j)$ en fonction de K (cf formule (3-12)) pour quatre couples choisis au hasard et repérés sur la Fig. 3-6-a. Tous ces couples sont distants de r=Dx. La covariance théorique est donc la même pour tous. En revanche, $C_s(\underline{x}_i, \underline{x}_j)$ n'apparaît plus homogène et isotrope, notamment lorsque un ou les deux points appartiennent à la frontière. Pour les couples du centre (Fig. 3-6-c), les deux courbes sont superposables, l'isotropie est donc respectée.

Par ailleurs, la convergence est assez rapide (elle est atteinte pour K=90, sachant qu'on a calculé P=121 F.P.). Mais, pour les couples (5,16) et surtout (5,6) (Fig 3-6-b), il faut prendre quasiment toutes les F.P. pour atteindre la valeur théorique et de plus, pour K=P, on est beaucoup plus éloigné de la valeur théorique que dans le cas précédent. La troncature semble donc induire des effets de bord, les dernières F.P. estimées semblant être plutôt associées aux points de la frontière du domaine. Nous retrouverons ces effets au § 3-5, lors de la comparaison avec d'autres techniques de simulation.

3-4-2 Influence de la discrétisation spatiale sur la covariance simulée.

Nous ne développerons pas outre mesure ce paragraphe puisque nous y retrouverons des résultats déjà énoncés au chapitre 2. En effet, nous avons vu que la discrétisation spatiale influait directement sur le calcul des F.P., qui interviennent elles-mêmes dans le processus simulé (cf formule (3-6)). Pour tous les essais présentés dans ce paragraphe, le nombre K de F.P. retenues correspond à 90% de variance expliquée, afin de minimiser les effets de la troncature.

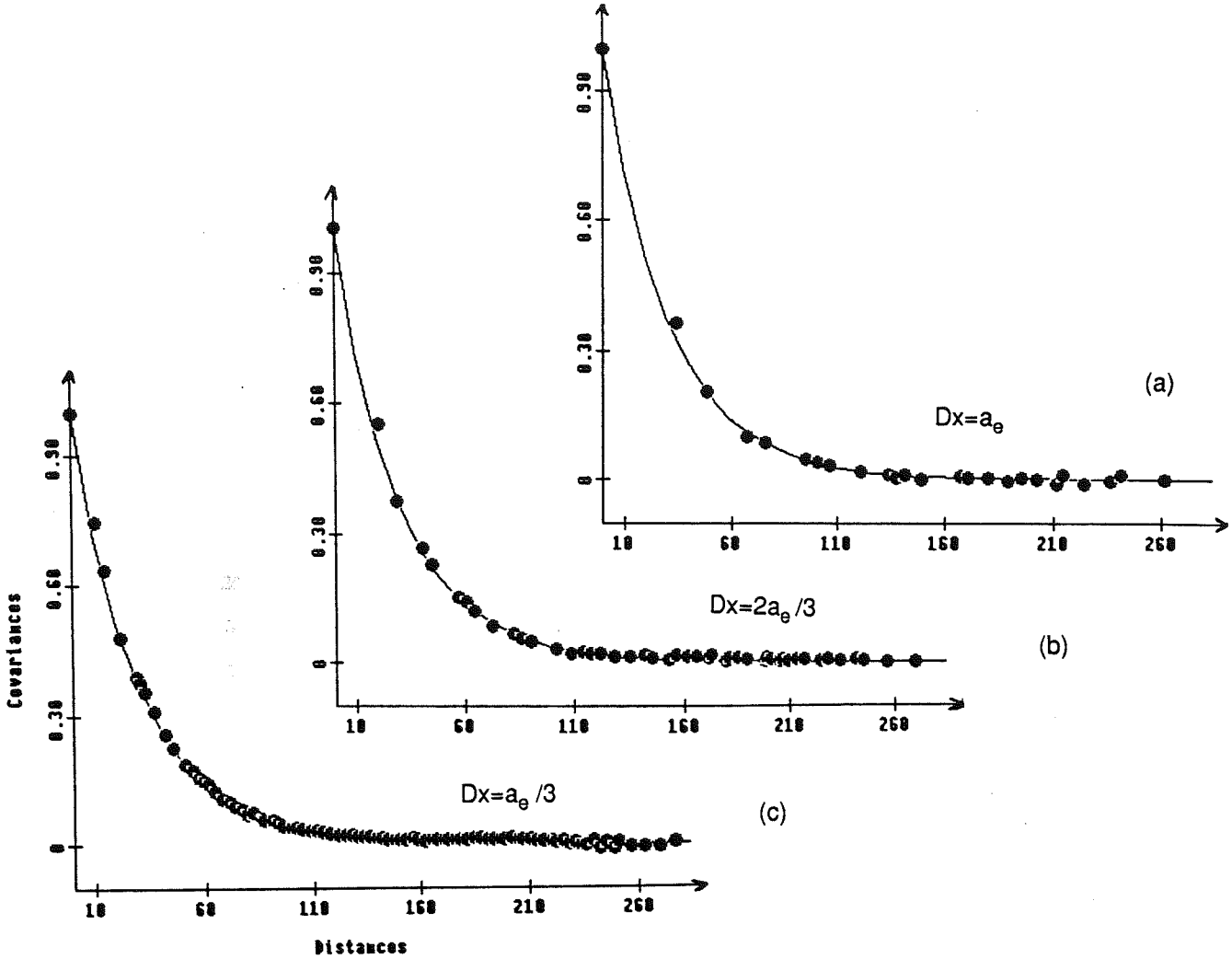


Fig. 3-7: Comparaison des covariances expérimentales et théoriques pour un domaine carré de côté Ls, selon le pas Dx de discrétisation.

—— Modèle de covariance $C(r) = \exp(-r/a_e)$ avec $a_e = 3Ls/20$.

• Covariance expérimentale avec (a) $Dx=a_e$, (b) $Dx=2a_e/3$, (c) $Dx=a_e/3$.

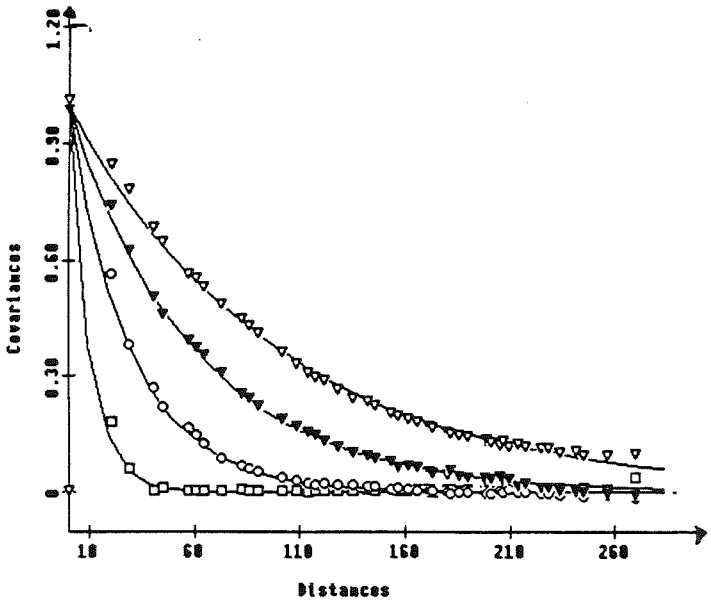


Fig. 3-8: Comparaison des covariances théoriques et expérimentales selon la portée a_e du modèle exponentiel. Ls / Dx =10.

—— Modèle $C(r) = \exp(-r/a_e)$ avec □ $a_e = Dx/2$, o $a_e = 3Dx/2$,

▼ $a_e = 3Dx$, ▽ $a_e = 5Dx$.

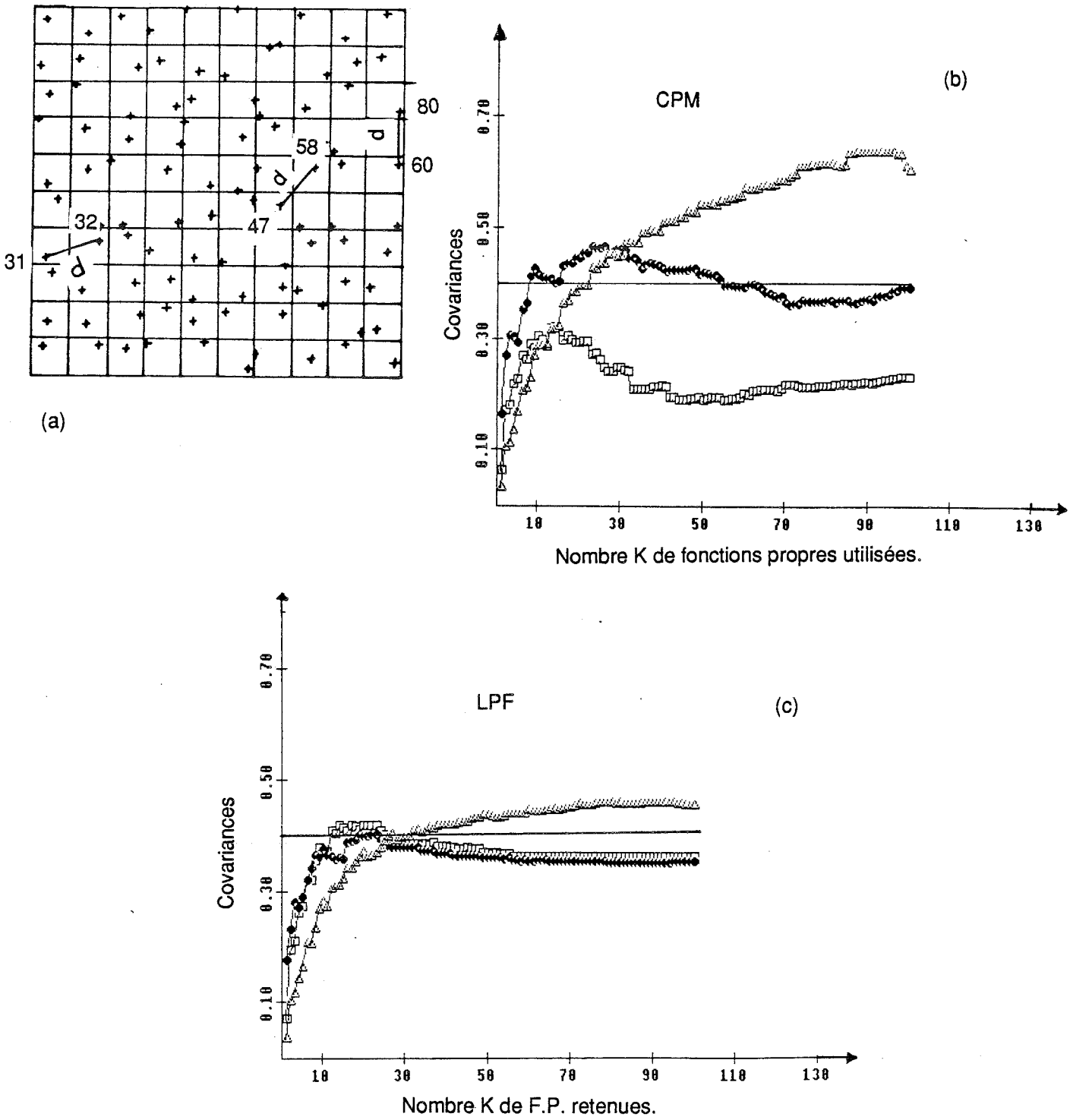


Fig. 3-9: Evolution de la covariance empirique en fonction du nombre K de F.P. utilisées pour 4 couples de points distants de d repérés sur la grille (a).

(a) Grille régulière de calcul des F.P. et réseau aléatoire stratifié + où a été simulé le processus.
(b) Utilisation de fonctions de base CPM. Modèle C(d) = .39 —
—□— stations 31-32, —●— stations 47-58 —△— stations 60-80.
(c) Utilisation de fonctions de base LPF. Modèle C(d) = .39 —
—□— stations 31-32, —●— stations 47-58 —△— stations 60-80.

Pour une taille Ls du domaine et une portée a_e fixées (a_e=3.Ls/20), la diminution progressive du pas Dx, entraîne une meilleure reconstitution de la covariance (Fig. 3-7), ce qui est cohérent avec la meilleure estimation des F.P. observée au chapitre 2. On retrouve aussi, pour avoir une bonne estimation de $C_s(r)$ qu'un rapport Ls/Dx=20 semble suffisant.

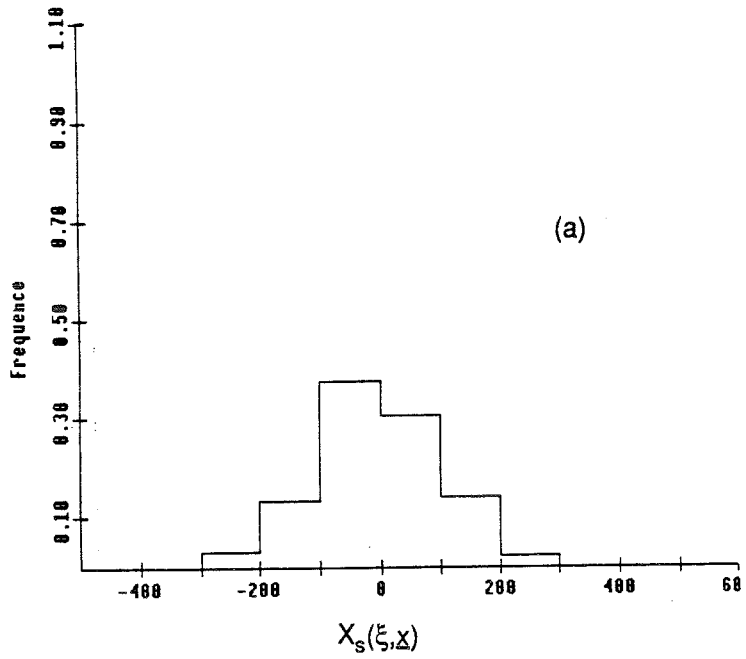
De même, la Fig. 3-8, qui, pour Ls et Dx fixés, présente les covariances simulées pour diverses valeurs de la portée a_e, confirme la moins grande importance de ce dernier paramètre, déjà notée au chapitre 2. C'est donc plus le rapport taille du domaine/ pas qui est important que le rapport portée du modèle /pas. Mais, bien entendu, il reste néanmoins nécessaire de choisir un pas de calcul des F.P. inférieur à la portée. Ceci n'est pas vérifié pour le cas a_e=Dx/2, et on voit que l'essai n'est pas très significatif puisqu'on compare des covariances presque toujours nulles!..

Enfin, pour illustrer l'influence du choix des fonctions de base $e_i(\underline{x})$, l'essai suivant a été réalisé. Les F.P. ont été calculées sur une grille carrée régulière et la simulation a été faite sur un réseau aléatoire stratifié (Fig. 3-9-a). Il était donc nécessaire d'interpoler les F.P. Les Fig. 3-9-b et c permettent de comparer, pour trois couples distants de d, l'évolution de la covariance empirique $C_s(\underline{x}_i, \underline{x}_j)$ avec le nombre K de F.P. retenues en utilisant des fonctions $e_i(\underline{x})$ constantes par morceau (CPM) ou linéaires par facette (LPF). La supériorité de l'utilisation de fonctions LPF est nette: la convergence est plus rapide, même si un léger biais semble apparaître. Avec des fonctions CPM, on peut converger vers des valeurs différentes de la valeur théorique, ce qui pénalise cette approche. Néanmoins, dans les deux cas, des problèmes se posent pour les points frontières (couple 60-80) car on ne converge pas vers la vraie valeur de la covariance, même en LPF, ce qui vient confirmer les effets de bord évoqués au paragraphe précédent.

Comme au chapitre 2, nous pouvons donc recommander de choisir un pas de discrétisation de la grille de calcul des F.P. cohérent avec la taille du domaine et la portée du modèle et de privilégier une approche LPF si la grille de simulation est différente de la grille de calcul des F.P.

3-4-3 Quelques remarques sur la distribution des $Z_k(\xi)$.

Tous les résultats précédents ont été obtenus avec des $Z_k(\xi)$ tirées de lois normales $N(0, \sqrt{\lambda_k})$. Nous avons aussi testé la méthode de simulation en utilisant une distribution uniforme sur $[-\sqrt{3.\lambda_k}, \sqrt{3.\lambda_k}]$. (Les bornes de cet intervalle ont été choisies pour conserver une moyenne nulle aux $Z_k(\xi)$, ainsi que la variance égale à λ_k). Ces tests ont porté essentiellement sur la distribution de $X_s(\xi,\underline{x})$, tant station par station que champ par champ, sur les valeurs et les distributions des moyennes et écarts-types station par station



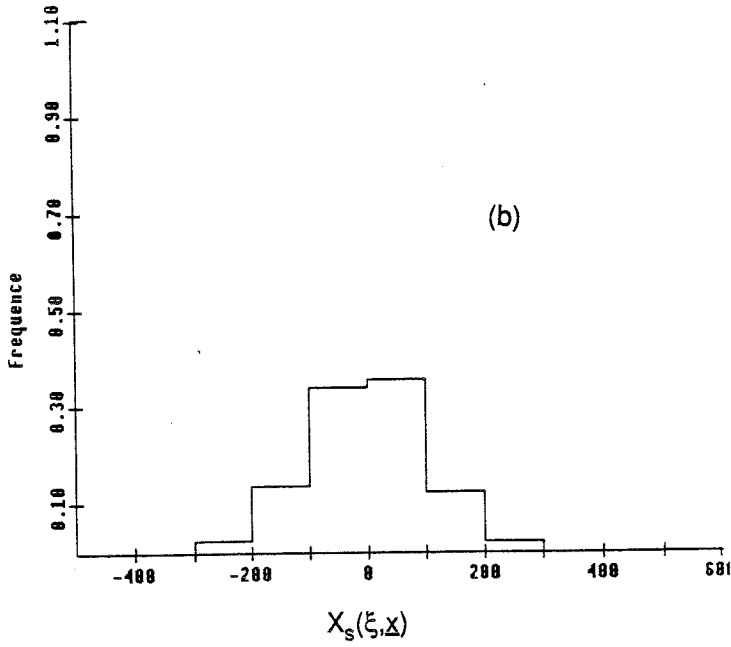
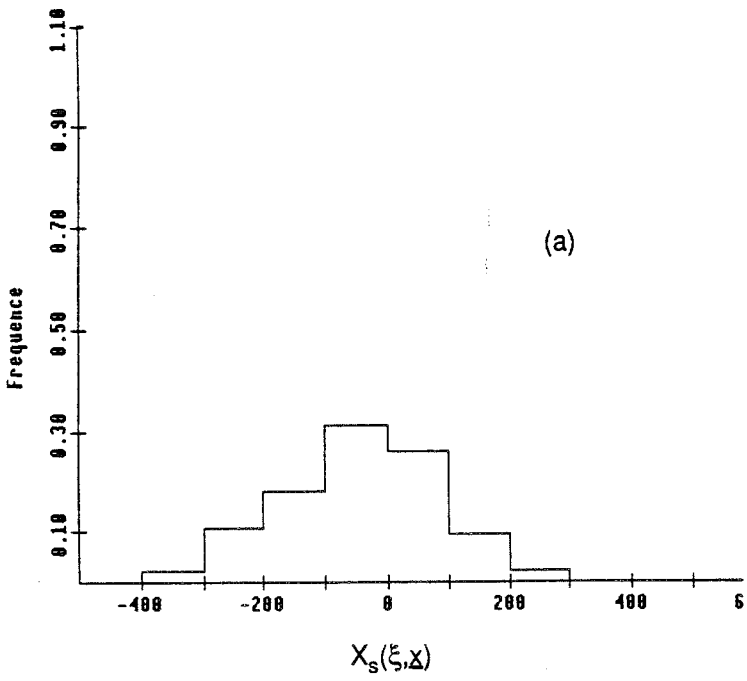


Fig. 3-10:Histogramme de $X_s(\xi,\underline{x})$ pour la station 42 avec des distributions des $Z_k(\xi)$
(a) gaussiennes, (b) uniformes



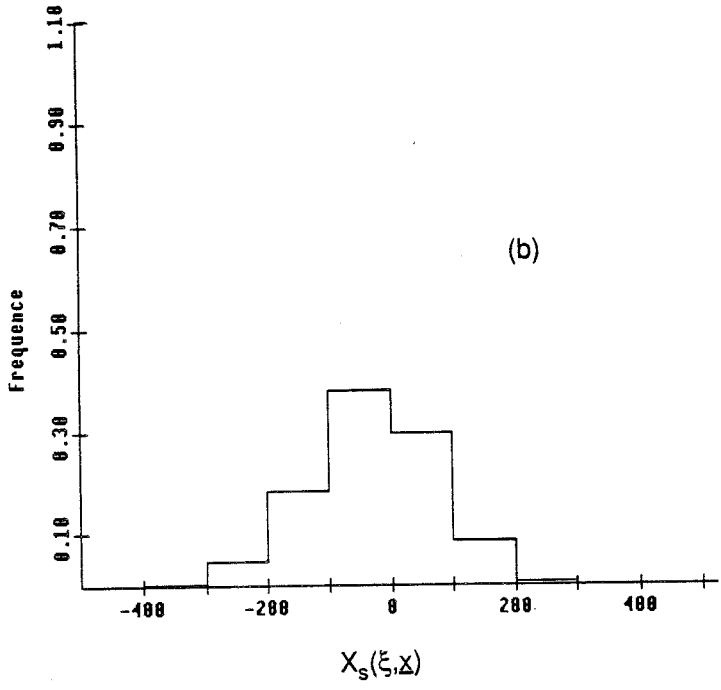


Fig. 3-11:Histogramme de $X_s(\xi,\underline{x})$ pour le champ 114 avec des distributions des $Z_k(\xi)$
(a) gaussiennes, (b) uniformes

(on parlera alors de valeurs climatologiques) ou champ par champ (ce seront alors des valeurs spatiales).

La distribution de X est approximativement gaussienne, que celle des $Z_k(\xi)$ soit uniforme ou gaussienne. Ce résultat est vrai si on trace les histogrammes globaux pour l'ensemble des stations et des champs simulés, mais encore pour les distributions de chaque champ ou de chaque station. Les Fig. 3-10 et 3-11 illustrent ces résultats pour, respectivement, une station et un champ particulier.

Ce résultat peut s'expliquer assez facilement. En effet, nous avons:

$$X_s(\xi,\underline{x}) = \sum_{k=1}^{K} Z_k(\xi) \cdot F_k(\underline{x}) + \varepsilon(\xi,\underline{x}) \qquad (3\text{-}6)$$

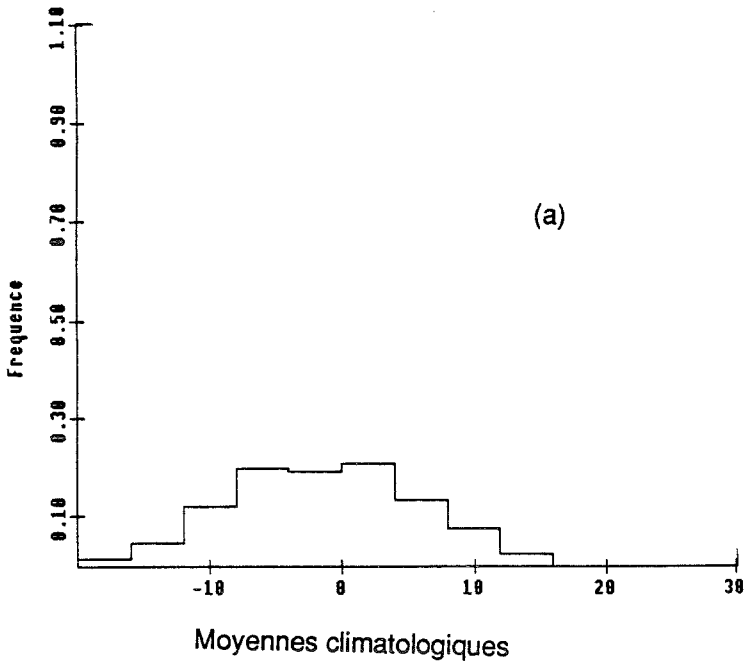
X_s est donc obtenu en sommant un grand nombre de variables aléatoires, les $Z_k(\xi)$. (Ce nombre est effectivement grand car, en un même point, on ne trouvera que peu de valeurs de k telles que $F_k(\underline{x})=0$ simultanément). Or la distribution de la somme de deux variables aléatoires est donnée par la convolution des deux distributions (Bass,1962). Ici, la taille des supports des distributions uniformes des $Z_k(\xi)$ va décroissant avec k. La convolution de deux "rectangles" donnera un trapèze et en réalisant ces opérations pour toutes les valeurs de k, on peut admettre aisément que la distribution finale de $X_s(\xi,\underline{x})$ soit gaussienne.

Si on s'intéresse maintenant aux distributions des moyennes climatologiques (calculées sur N réalisations) ou à leur écart-type, là encore, on retrouve des distributions qui se rapprochent de lois normales (Fig. 3-12 et 3-13). Les tableaux 3-2 et 3-3 résument les statistiques sur ces moyennes et écarts-types climatologiques.

Nous avons utilisé la notation E pour les moyennes climatologiques calculées sur N=300 réalisations, et < >, pour les moyennes spatiales évaluées à l'aide de P=441 stations. x_i^j désigne l'observation i de la station j, X_i, l'observation i et X^j,la station j.

		< E(xi) >	$\sigma_{E(xi)}$	var(E(xi))	Intervalle de Confiance à 80 %	Nombre d'obs Hors intervalle	% d'obs Hors intervalle	Distribution
Z_k gaussiennes	Théorique	0	5,77	33,33	[-7,39 ; 7,39]	145	32,90%	Gaussienne
	Empirique (441 Obs.)	-1,34	6,92	47,96	[-10,2 ; 7,52]	96	21,80%	Gaussienne
Z_k uniformes	Théorique	0	5,77	33,33	[-7,39 ; 7,39]	82	18,50%	Gaussienne
	Empirique (441 Obs.)	0,63	5,87	34,48	[-6,89 ; 8,14]	80	18,10%	Gaussienne

Tableau 3-2: Statistiques sur les moyennes climatologiques.



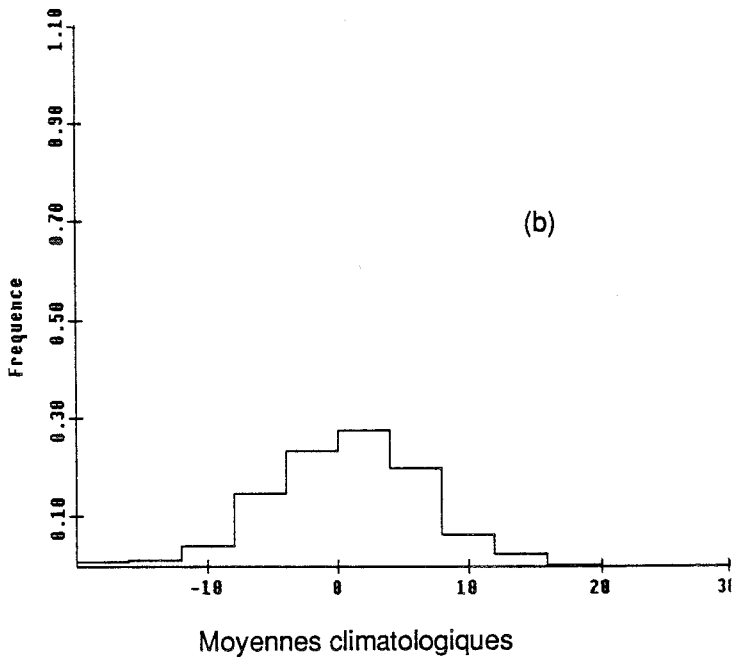
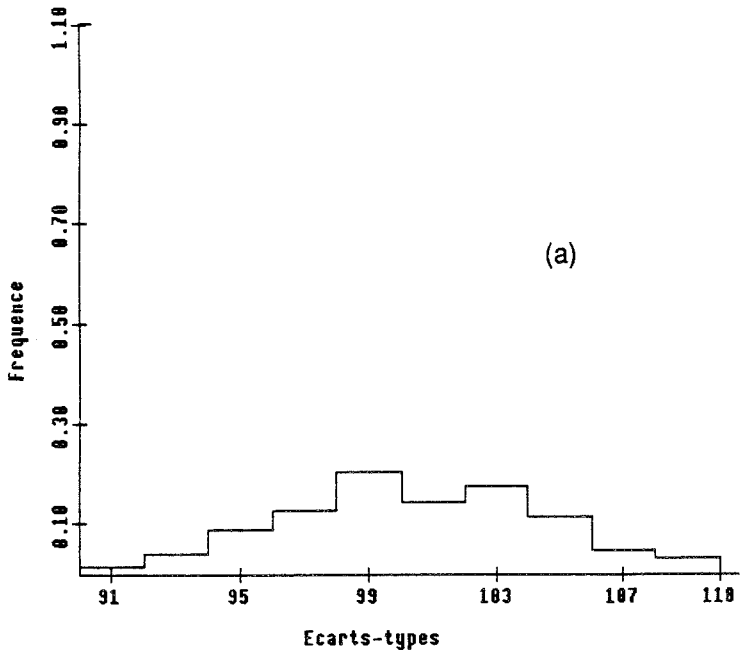


Fig. 3-12:Histogramme des moyennes climatologiques avec des distributions des $Z_k(\xi)$

(a) gaussiennes, (b) uniformes



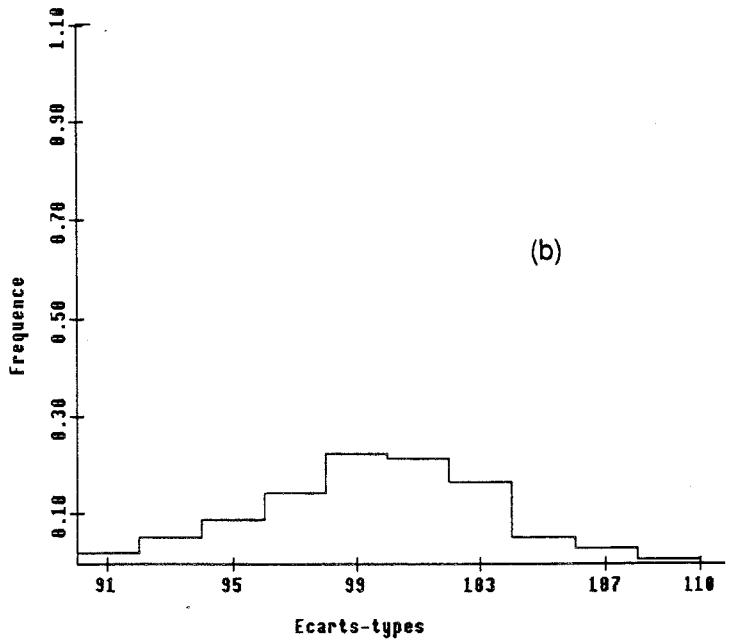


Fig. 3-13:Histogramme des écarts-types climatologiques avec des distributions des $Z_k(\xi)$

(a) gaussiennes, (b) uniformes

Il est néanmoins curieux de noter, qu'avec des lois uniformes pour les $Z_k(\xi)$, la distribution de la moyenne est plus proche d'une gaussienne que si les $Z_k(\xi)$ possèdent aussi cette distribution. C'est encore vrai pour l'écart-type, mais à un degré bien moindre.

		$<\sigma(xi)>$	$\sigma_{\sigma xi}$	var(σxi)	Intervalle de Confiance à 80 %	Nombre d'obs Hors intervalle	% d'obs Hors intervalle	Distribution
Z_k gaussiennes	Théorique	100	4,08	16,66	[94,77 ; 105,22]	97	22,00%	Gaussienne
	Empirique (441 Obs.)	100,52	4,29	18,4	[495,03 ; 106,01]	84	19,00%	Gaussienne
Z_k uniformes	Théorique	100	4,08	16,66	[94,77 ; 105,22]	66	14,50%	Gaussienne
	Empirique (441 Obs.)	99,57	3,62	13,12	[94,93 ; 104,21]	78	17,70%	Gaussienne

Tableau 3-3: Statistiques sur les écarts-types climatologiques.

Remarque: Il faut signaler que les valeurs théoriques ou empiriques des intervalles de confiance ou des écarts-types sont biaisées puisque, lorsqu'on calcule une moyenne spatiale, les P stations ne sont pas indépendantes. On a donc tendance à être trop pessimiste sur les valeurs théoriques, ce qui est en général vérifié, sauf pour l'écart-type climatologique lorsque les $Z_k(\xi)$ sont tirées dans des lois uniformes.

Enfin, l'étude des moyennes spatiales $< X^j >$ permet de mettre en évidence à quel niveau le choix de la distribution des $Z_k(\xi)$ peut introduire des biais. En effet, la Fig. 3-14 montre que, pour des $Z_k(\xi)$ gaussiennes, $< X^j >$ a une distribution plutôt gaussienne, bien qu'assez aplatie, mais qu'en revanche, si on tire les $Z_k(\xi)$ dans des lois uniformes, $< X^j >$ aura aussi un distribution plutôt uniforme. Ce résultat n'est pas conservé pour les écarts-types spatiaux (Fig. 3-15).

		E(<xj>)	$\sigma_{<xj>}$	var(<xj>)	Intervalle de Confiance à 80 %	Nombre d'obs Hors intervalle	% d'obs Hors intervalle	Distribution
Z_k gaussiennes	Théorique	0	4,76	22,65	[-6,09 ; 6,09]	247	82,00%	Gaussienne
	Empirique (300 Obs.)	-1,34	29,76	885,71	[-39,43 ; 36,75]	64	21,30%	Gaussienne
Z_k uniformes	Théorique	0	4,76	22,67	[-6,09 ; 6,09]	259	86,00%	Gaussienne
	Empirique (300 Obs.)	0,63	30,24	914,38	[-48 ; 48] Loi uniforme	35	11,70%	Uniforme sur [-60 ; 60]
					[-34,08 ; 39,33] Gausienne	72	24,00%	

Tableau 3-4: Statistiques sur les moyennes spatiales.

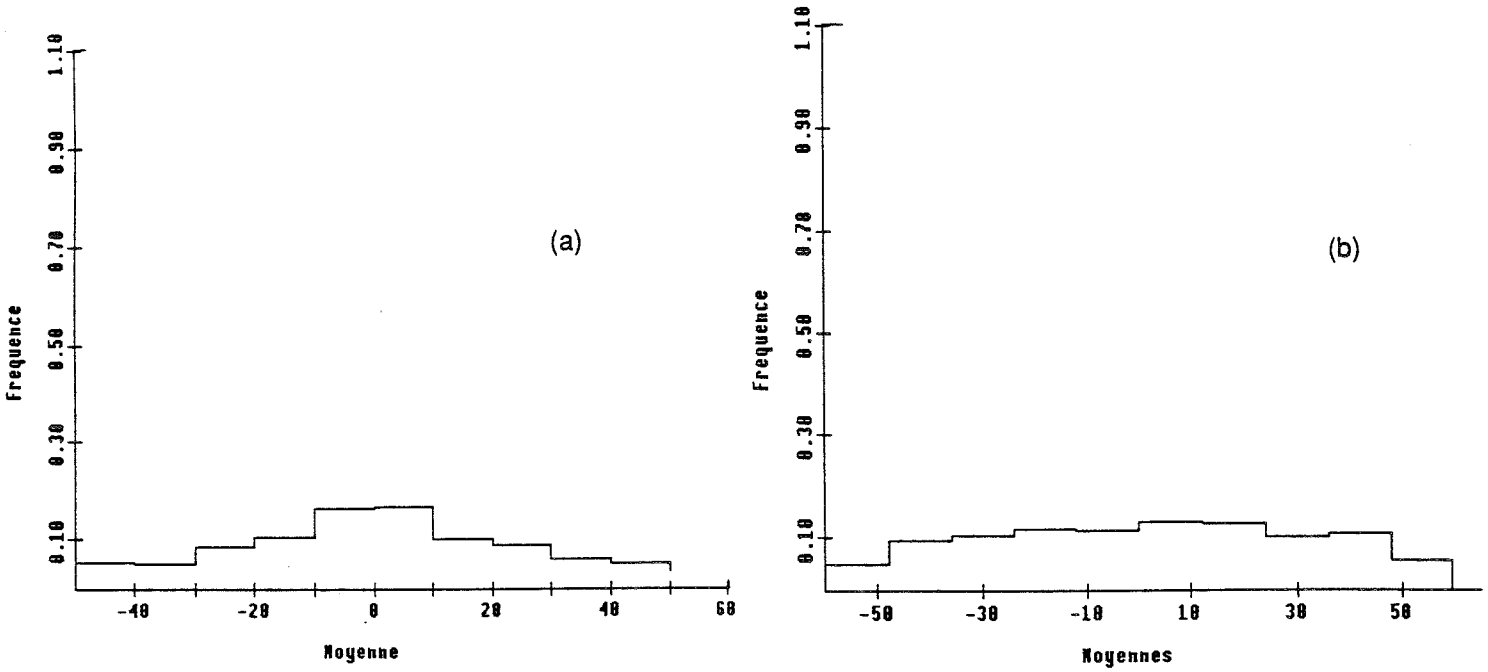


Fig. 3-14:Histogramme des moyennes spatiales avec des distributions des $Z_k(\xi)$

(a) gaussiennes, (b) uniformes

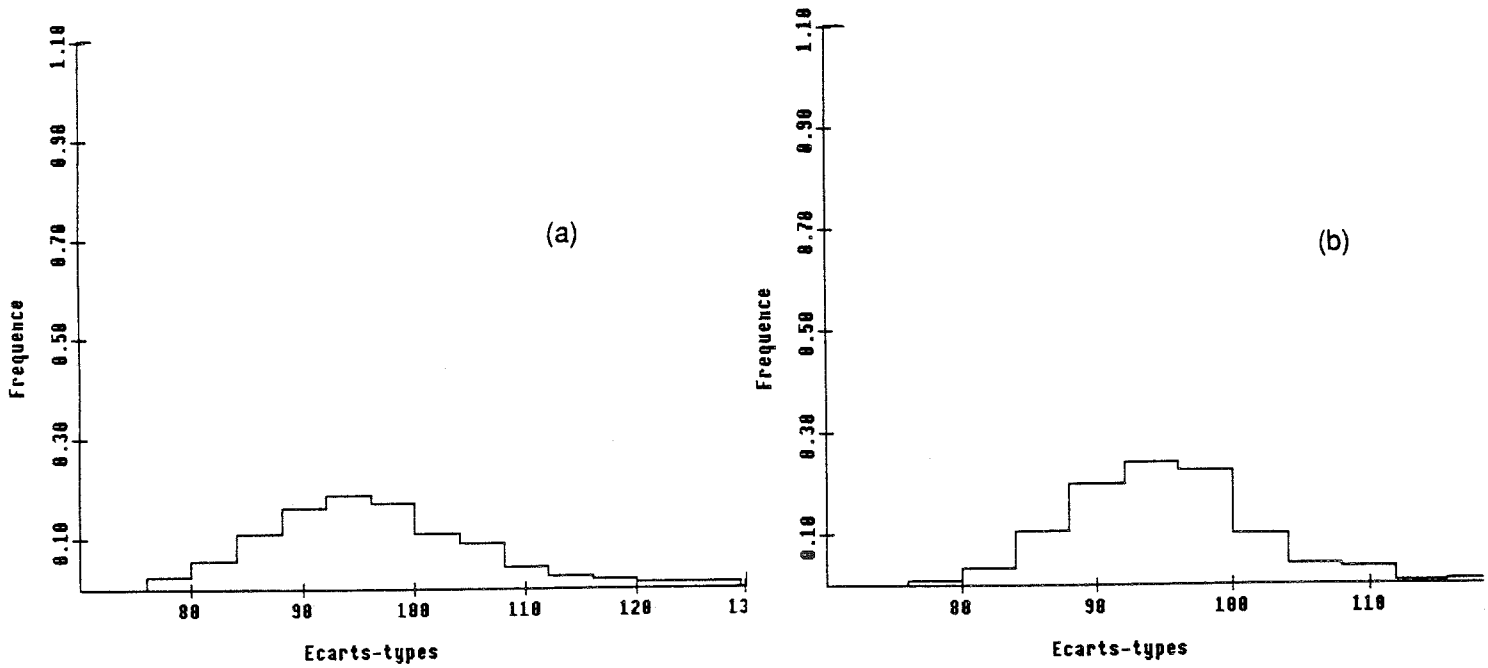


Fig. 3-15:Histogramme des écarts-types spatiaux avec des distributions des $Z_k(\xi)$

(a) gaussiennes, (b) uniformes

		$E(\sigma_{xj})$	$\sigma_{\sigma_{xj}}$	$var(\sigma_{xj})$	Intervalle de Confiance à 80 %	Nombre d'obs Hors intervalle	% d'obs Hors intervalle	Distribution
Z_k gaussiennes	Théorique	100	3,37	11,35	[95,69 ; 104,31]	202	67,00%	Gaussienne
	Empirique (300 Obs.)	95,85	9,16	83,83	[84,14 ; 107,57]	52	17,30%	Gaussienne
Z_k uniformes	Théorique	100	3,37	11,33	[95,69 ; 104,31]	182	61,00%	Gaussienne
	Empirique (300 Obs.)	94,82	6,67	44,49	[86,27 ; 103,35]	59	19,70%	Gaussienne

Tableau 3-5: Statistiques sur les écarts-types spatiaux.

Les tableaux 3-4 et 3-5 complètent ces figures en donnant les statistiques sur ces moyennes et écarts-types champ par champ. Les notations sont identiques aux précédentes. Dans ce cas, les valeurs théoriques estimées pour les intervalles de confiance et les écarts-types sont largement sous-estimées et la majorité des valeurs sort de ces intervalles théoriques. En revanche, avec les intervalles empiriques, les résultats sont nettement plus satisfaisants.

Nous pouvons maintenant essayer d'avancer quelques remarques qui permettraient d'expliquer pourquoi, avec des $Z_k(\xi)$ distribuées uniformément, les moyennes climatologiques sont gaussiennes, alors que les moyennes spatiales sembleraient plutôt uniformes.

Nous avons vu plus haut comment on pouvait justifier la distribution gaussienne de $X_s(\xi,\underline{x})$, même lorsque les $Z_k(\xi)$ sont uniformes. Lorsqu'on calcule une moyenne climatologique $E[\ X_s(\underline{x})\ ]$, en un point $\underline{x}$, on va sommer des variables aléatoires gaussiennes, il semble donc logique que la moyenne climatologique appartienne aussi à une loi normale.

En revanche, lorsqu'on calcule une moyenne spatiale $< X_s(\xi) >$, elle est donnée par:

$$< X_s(\xi) > = \sum_{k=1}^{K} Z_k(\xi) \,.\, < F_k(\underline{x}) > \quad + < \varepsilon(\xi,\underline{x}) >$$

Or, si on définit l'opérateur < > par

$$< F_k(\underline{x}) > = 1/D \int_D F_k(\underline{x}) \,.\, d\underline{x}$$

compte-tenu des changements de signe et des symétries de $F_k(\underline{x})$ sur D (voir chapitre 2), on peut s'attendre à ce que la majorité des $< F_k(\underline{x}) >$ soit nulle (ou proche de zéro) sauf $< F_1(\underline{x}) >$ et éventuellement les valeurs correspondant à des valeurs propres simples au chapitre 2.

$< X_s(\xi,) >$ n'est donc plus la somme d'un grand nombre de variables aléatoires, mais bien de seulement deux ou trois. La distribution de $< X_s(\xi,) >$ sera donc plutôt gouvernée par celle de $Z_1(\xi)$, ou la somme de deux ou trois distributions uniformes et sera donc plus proche d'une loi uniforme que d'une gaussienne.

Compte-tenu de la complexité des calculs, ce type de raisonnement n'a pas été mené pour l'écart-type spatial.

Pour conclure ce paragraphe, nous pouvons donc dire que, pour obtenir des variables gaussiennes, il vaudrait mieux utiliser des distributions normales pour les $Z_k(\xi)$. Le choix d'une autre distribution semble néanmoins n'induire que des biais sur la moyenne spatiale. En revanche, faute de données disponibles, nous n'avons pas essayé de simuler des variables aléatoires avec des distributions non gaussiennes. Néanmoins, compte-tenu des remarques faites, il serait sûrement préférable de faire une anamorphose sur ces données afin de se ramener à des variables de distributions normales qu'on simulerait avant de les retransformer par l'opération inverse.

Notons que, dès 1977, Goussebaïle s'était posé le problème de la simulation par ACP de variables non gaussiennes puisqu'il traitait des pourcentages d'insolation, donc une variable bornée entre 0 et 100, dont la distribution est nettement bimodale. Bien que disposant d'un grand échantillon lui permettant d'estimer avec précision les distributions des $Z_k(\xi)$ et donc de les tirer au hasard dans ces lois, il s'est avéré qu'une fois la sommation (3-6) effectuée avec K F.P., les variables simulées sortaient, parfois de façon très importante, de l'intervalle [0, 100]. Cette expérience assez décevante peut venir conforter la nécessité d'une anamorphose préalable des données.

3-5 COMPARAISON DES PERFORMANCES DE LA SIMULATION PAR ACPP AVEC CELLES DES METHODES LU ET DES BANDES TOURNANTES.

Les deux méthodes retenues pour cette comparaison ont été décrites aux paragraphes 3-2-3 pour les bandes tournantes (BT) et 3-2-4 pour la méthode LU. Cette comparaison a été faite en terme de variogrammes $\gamma(\underline{h}) = 1/2\ E\ [\ (X_s(\underline{x}) - X_s(\underline{x}+\underline{h})^2]$. Comme le modèle de covariance est stationnaire, $\gamma(\underline{h})$ est relié de façon simple à la covariance:

$$\gamma(\underline{h}) = C(0) - C(\underline{h}) \qquad \text{(Obled, 1987 par exemple)}$$

Nous avons en fait calculé un variogramme moyenné sur 100 champs simulés, appelé par la suite variogramme "climatologique", puis nous l'avons modélisé. Le modèle de covariance retenu est toujours de type exponentiel isotrope:

$$C(\underline{h}) = C(||\underline{h}||) = C(r) = \sigma^2 \,.\, \exp(-r/a_e)$$

et les variables ont toujours une moyenne nulle et une variance égale à un.

La modélisation, par une méthode heuristique, du variogramme climatologique nous fournit des valeurs "modélisées" a_e^* et σ^{2*} des paramètres portées et paliers, qui peuvent être comparées aux valeurs théoriques a_e et σ^2. (Ici a_e=60 et σ^2=1).

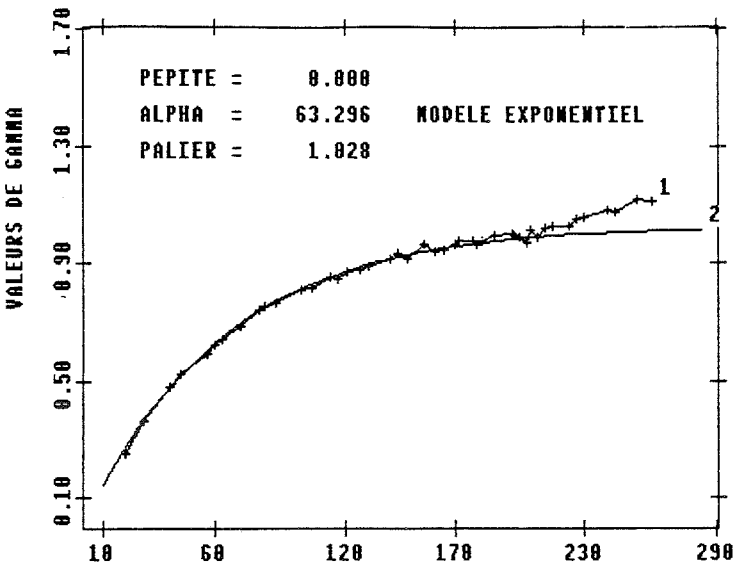
Pour la simulation par ACPP, nous nous sommes placés dans un cas où la restitution de la covariance était bonne, c'est-à-dire en respectant les règles proposées au § 3-4: Ls/Dx=10, a_e=3.Dx, K=62 (90% de variance expliquée). Pour la simulation par BT, nous avons choisi les paramètres en respectant les recommandations de Mantoglou et Wilson (1981).

Les valeurs des portées et paliers modélisés pour les variogrammes climatologiques, tant globaux (Fig. 3-16) que directionnels (Fig 3-17 pour la définition des directions, Fig. 3-18 pour les variogrammes modélisés sur les données simulées par ACPP) sont données dans le tableau 3-6.

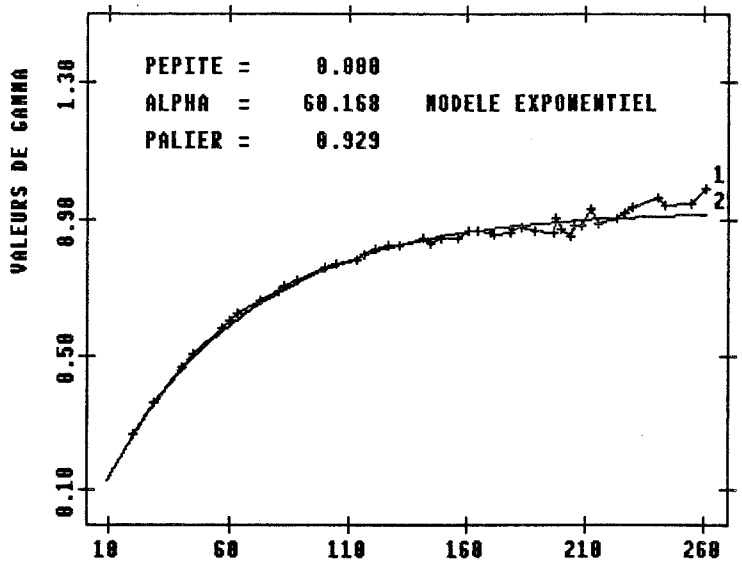
	ACPP		BT		LU	
Direction	ae *	σ^{2*}	ae *	σ^{2*}	ae *	σ^{2*}
Global	63.80	1.028	60.17	0.929	59.92	0.962
1	62.23	1.02	60.32	0.933	61.13	0.983
2	78.72	1.121	60.17	0.973	59.18	0.916
3	63.30	1.011	60.73	0.953	60.53	0.951
4	71.16	1.09	52.74	0.837	60.66	0.997

Tableau 3-6: Portées et paliers des variogrammes globaux et directionnels modélisés pour la simulation par ACPP, par BT ou par la méthode LU. Modèle : a_e=60, σ^2=1.

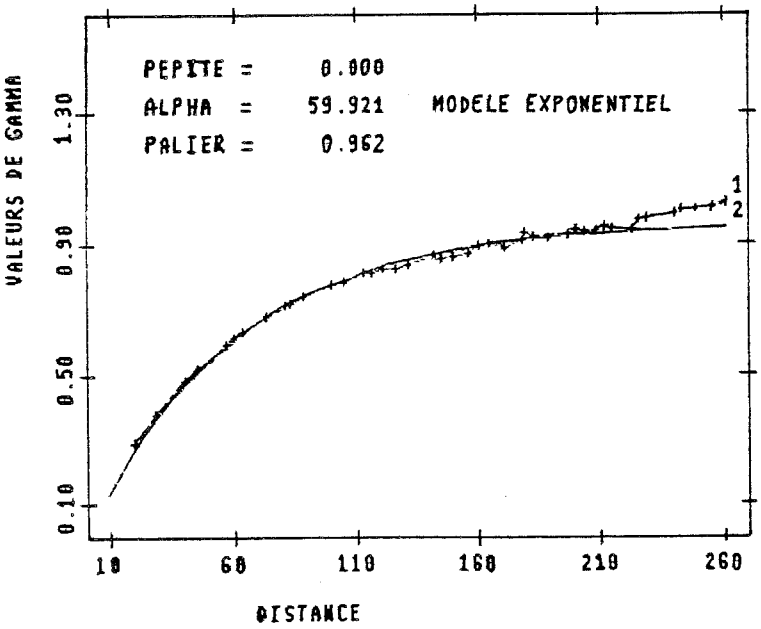
L'analyse de ces valeurs montre que, en ACPP, la valeur a_e^* est plus éloignée de la valeur théorique qu'avec les autres méthodes (surestimation générale) et que, de plus, des anisotropies apparaissent avec une portée plus longue selon les diagonales du carré que selon la direction perpendiculaire. On ne retrouve pas ce problème pour les deux autres techniques (même si, dans la direction 4, les BT semblent avoir quelques problèmes...). Cette anisotropie, compte-tenu des remarques faites au § 3-4-1, peut problablement être imputée à la troncature de la série des F.P. à K termes. Néanmoins, comme nous avons déjà choisi K pour expliquer 90% de variance, ce résultat est assez décevant.



(a)



(b)



(c)

Fig 3-16: Variogrammes climatologiques globaux bruts (+) et modélisés (-) avec une simulation des données (a) par ACPP, (b) par bandes tournantes, (c) par la méthode LU.
Modèle théorique C(r) = exp(-r/6Dx).

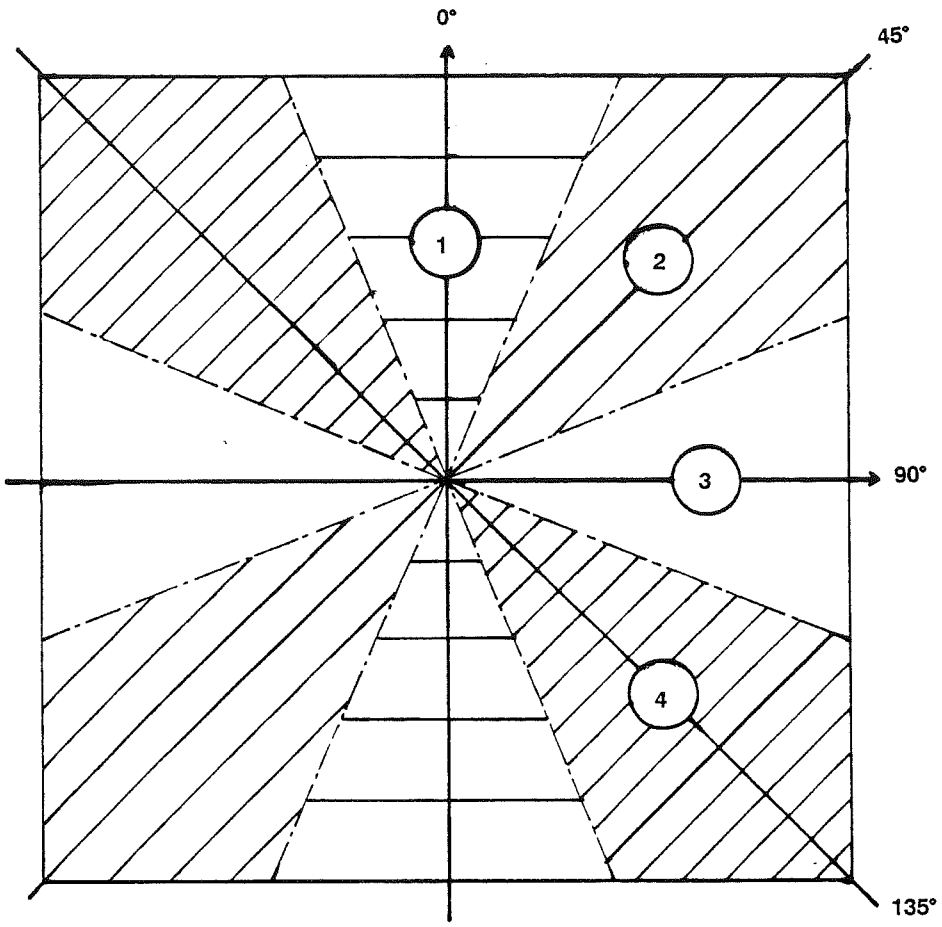


Fig. 3-17:Les différentes directions retenues pour le calcul des variogrammes par azimut.

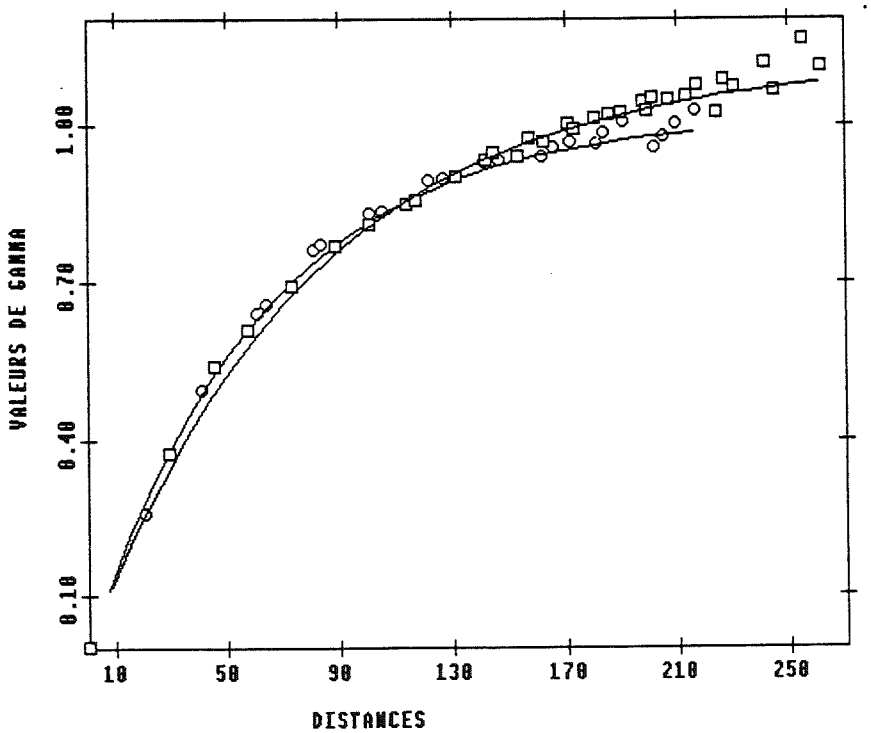


Fig. 3-18:Variogrammes climatologiques globaux bruts et modélisés, les données étant simulées par ACPP.

—○— Direction 1 —□— Direction 2

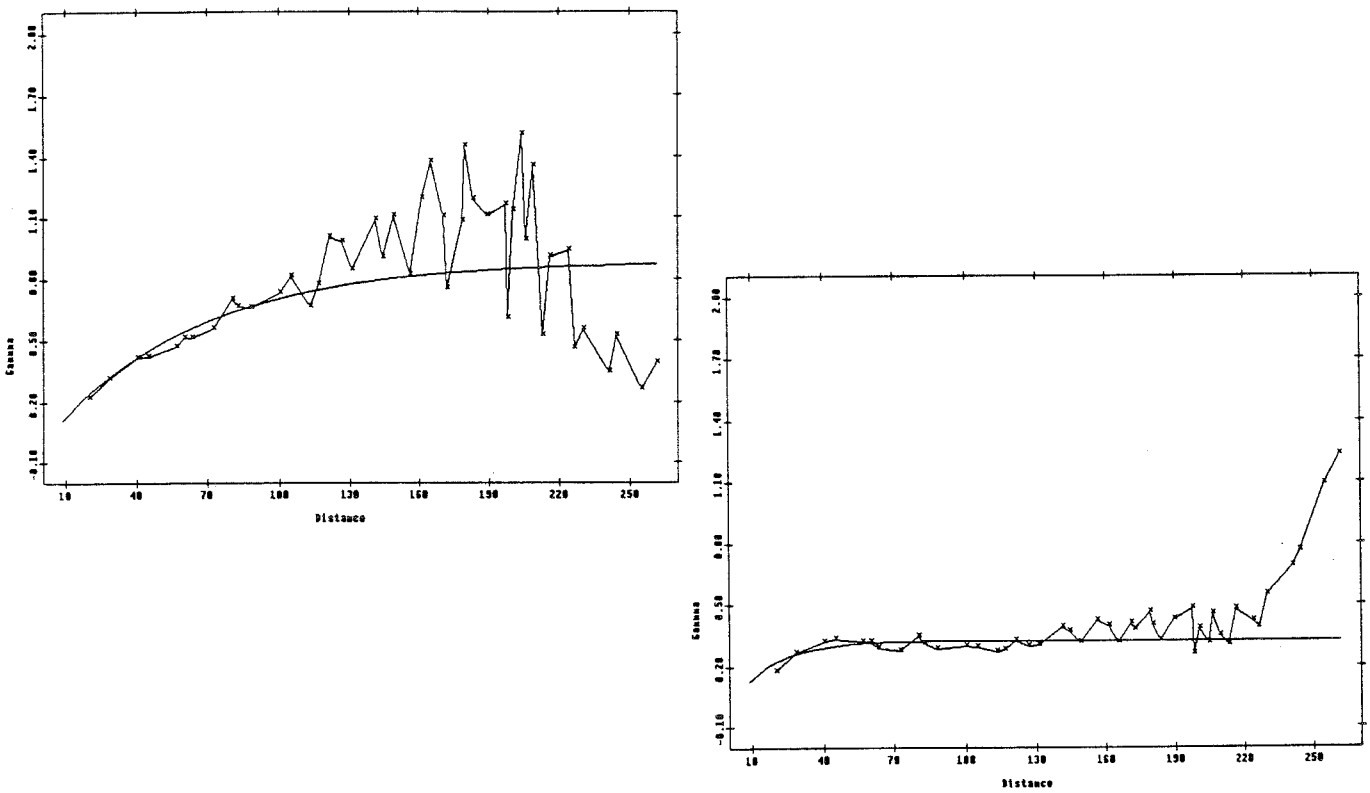


Fig. 3-19:Deux exemples de variogrammes bruts et modélisés pour deux champs particuliers simulés par ACPP.

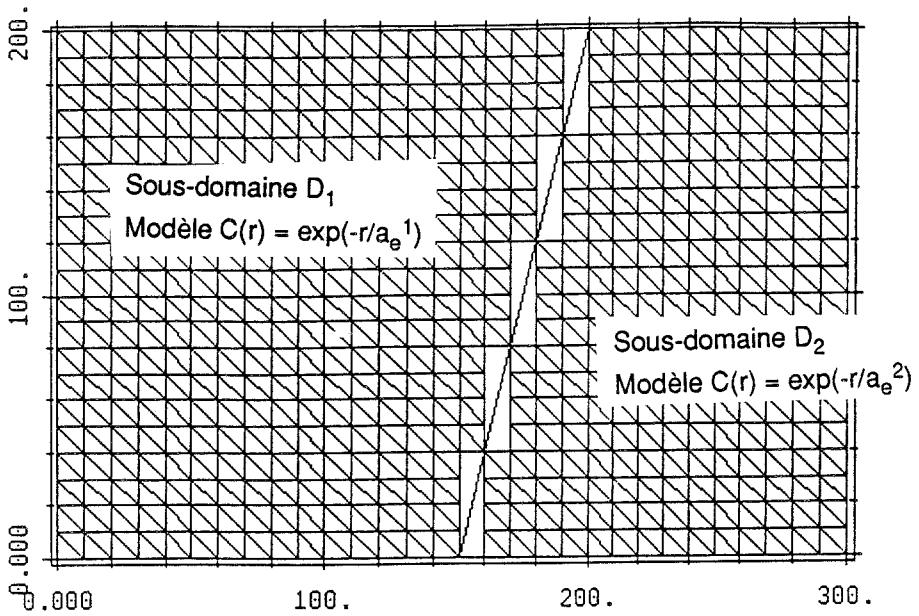


Fig. 3-20:Discrétisation du domaine 300.DX par 200.DX utilisé pour la simulaton de champs non homogènes

En revanche, en ACPP, l'estimation du palier semble meilleure qu'en BT ou qu'avec la méthode LU, mais, il faut aussi se rappeler que nous ajoutons un bruit blanc pour compenser le déficit en variance introduit par la troncature.

Les variogrammes de chaque champ ont aussi été calculés puis modélisés. Quelle que soit la méthode de simulation utilisée, on observe une variabilité très grande des formes (variogrammes en cloche, semblant présenter une dérive, etc...) et des valeurs des paramètres a_e^* et σ^{2*} modélisés. La Fig. 3-19 en donne une illustration pour la simulation par ACPP.

Donc, si globalement, les trois méthodes testées ont des performances comparables, avec une restitution du modèle de covariance acceptable, la simulation par ACPP semble présenter l'inconvénient d'introduire, à cause de la troncature, des anisotropies là où il n'en existe pas, ce qui pénalise la méthode par rapport à ses concurrentes. En revanche, nous allons voir maintenant que, contrairement aux BT, qui génèrent des processus stationnaires, l'ACPP peut permettre facilement de simuler des champs non homogènes, comme nous l'avons déjà signalé au début du § 3-4.

3-6 EXEMPLE SYNTHETIQUE DE SIMULATION DE CHAMPS NON STATIONNAIRES PAR ACPP.

L'exemple qui va être présenté est certes artificiel, mais il permet d'illustrer notre propos de façon assez nette. En effet, nous avons divisé un domaine rectangulaire de 31x21 points en deux sous-domaines <u>décorrélés</u>, dont la discrétisation est donnée par la Fig. 3-20. Pour chacun des sous-domaines, le modèle retenu est homogène, isotrope et de type exponentiel, les portées a_e étant différentes pour chacun des sous-domaines D_1 et D_2. Deux cas ont été envisagés: a_e^1=3.Dx sur D_1, a_e^2=6.Dx sur D_2, puis, a_e^1=6.Dx sur D_1 , a_e^2=3.Dx sur D_2.

Dans le cas particulier considéré (deux sous-domaines décorrélés) , la matrice de covariance globale **C** est formée de trois blocs indépendants:

$$\mathbf{C} = \left[\begin{array}{c|c} \mathbf{C_{11}} & \mathbf{C_{12}} \\ \hline \mathbf{C_{21}} & \mathbf{C_{22}} \end{array}\right]$$

où - $\mathbf{C_{11}}$ (resp. $\mathbf{C_{22}}$) est la matrice de covariance des points de D_1 (resp. de D_2)

- $\mathbf{C_{12}} = \mathbf{C_{21}} = 0$ puisque les deux sous-domaines sont décorrélés.

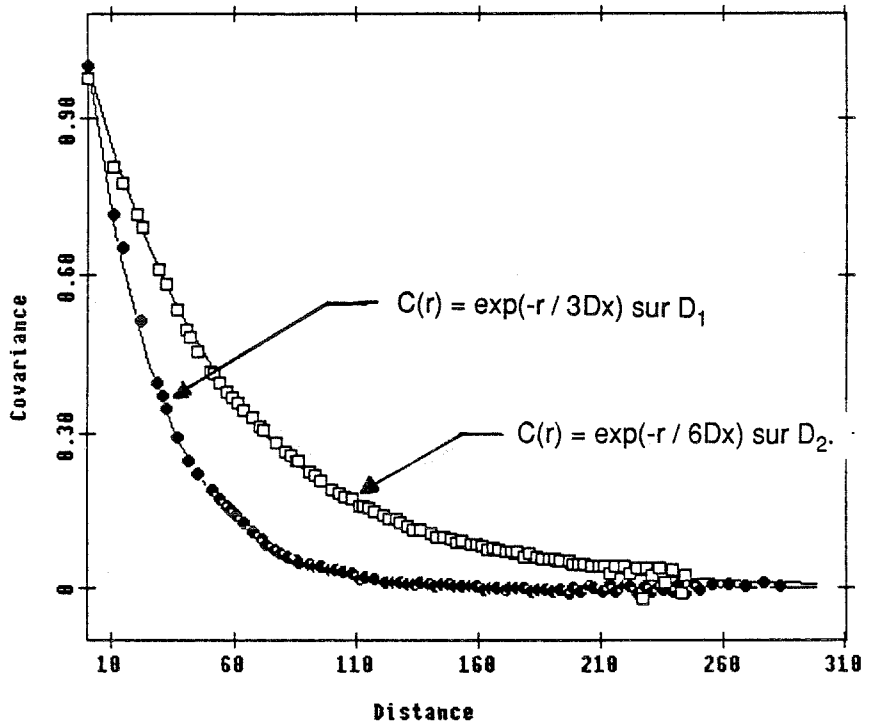


Fig. 3-21:Comparaison entre covariance théorique et covariance expérimentale pour les sous-domaines D_1 et D_2.
——— modèles théoriques; ● covariance expérimentale sur D_1, □ covariance expérimentale sur D_2.

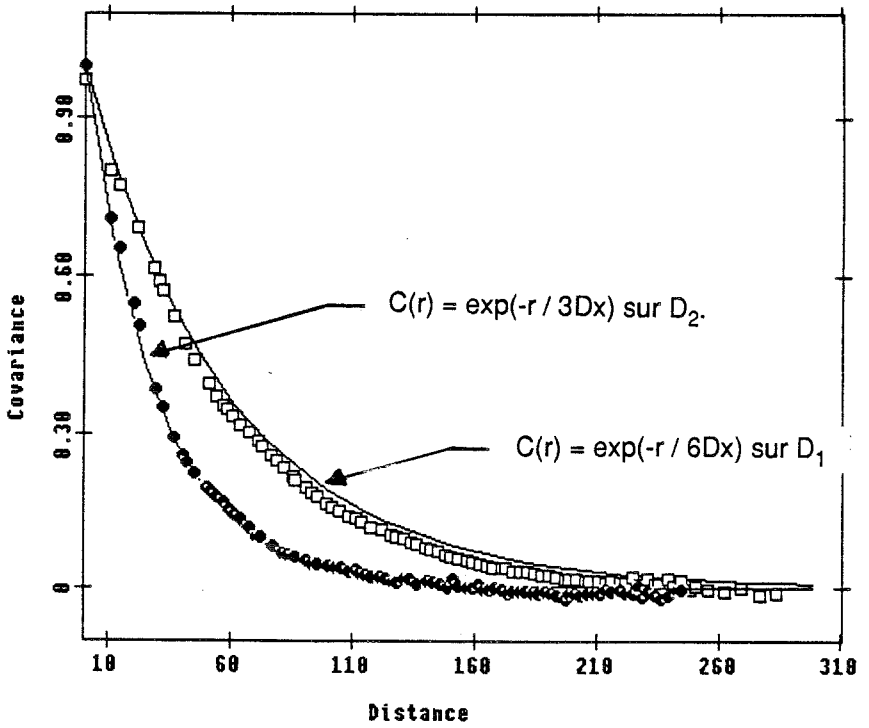


Fig. 3-22:Comparaison entre covariance théorique et covariance expérimentale pour les sous-domaines D_1 et D_2.
———modèles théoriques; □ covariance expérimentale sur D_1, ● covariance expérimentale sur D_2.

On pourrait donc diviser le problème aux valeurs propres en deux problèmes distincts:

$$C_{11} \cdot E_{11} \cdot F_{11} = \Lambda_{11} \cdot F_{11} \qquad \text{sur } D_1$$

$$C_{22} \cdot E_{22} \cdot F_{22} = \Lambda_{22} \cdot F_{22} \qquad \text{sur } D_2$$

Les vecteurs propres sur l'ensemble du domaine d'étude seraient alors obtenus en ajoutant simplement des zéros.

Mais, dans les analyses qui suivent, nous avons fait une ACPP globale sur tout le domaine, en supposant que nous ne connaissions pas l'existence de la "faille".

Voyons maintenant les résultats obtenus pour les deux essais réalisés. Le tableau 3-7 présente le nombre K de F.P. retenues et comment elles se répartissent entre les deux sous-domaines.

	K 90% (D1 + D2)	K1	% Variance expliquée sur D1 par K1	K2	% Variance expliquée sur D2 par K2
ae^1 = 3 Dx ae^2 = 6 Dx	100	67	88	33	92
ae^1 = 6 Dx ae^2 = 3 Dx	88	43	92	45	87

Tableau 3-7: Nombre K de F.P. retenues sur l'ensemble du domaine et répartition entre les deux sous-domaines. K_1(resp. K_2) est le nombre de F.P. associé à D_1(resp. D_2) et on a $K_1+K_2=K$

Dans le second cas (où le modèle à plus longue corrélation est associé au sous-domaine le plus étendu), on constate - et c'est normal compte-tenu de ce que nous avons déjà dit aux chapitres 2 et 3- qu'on a besoin de moins de F.P. pour expliquer le même pourcentage de variance. Par ailleurs, si dans ce cas, l'ensemble des F.P. se répartit équitablement entre les deux sous-domaines (K_1=43, K_2=45), ça n'est pas vrai dans le premier cas (K_1=67, K_2=33), mais on peut le comprendre puisque le plus grand sous-domaine (D_1) possède alors le modèle de covariance de portée la plus faible.

Si on regarde maintenant les courbes des Fig. 3-21 (premier cas: a_e^1=3.Dx , a_e^2=6.Dx) et 3-22 (second cas: a_e^1=6.Dx , a_e^2=3.Dx), on observe que les covariances expérimentales (calculées pour chaque sous-domaine en utilisant (3-13)) coïncident bien avec les covariances théoriques. Un léger problème semble néanmoins apparaître dans le second cas pour la portée 6.Dx, problème d'autant moins explicable qu'on a plus de F.P. que dans le premier cas pour reconstituer le modèle (43 contre 33) ...Par ailleurs, nous avons pu vérifier que sur les N=1000 champs simulés, les deux sous-domaines demeuraient bien décorrélés.

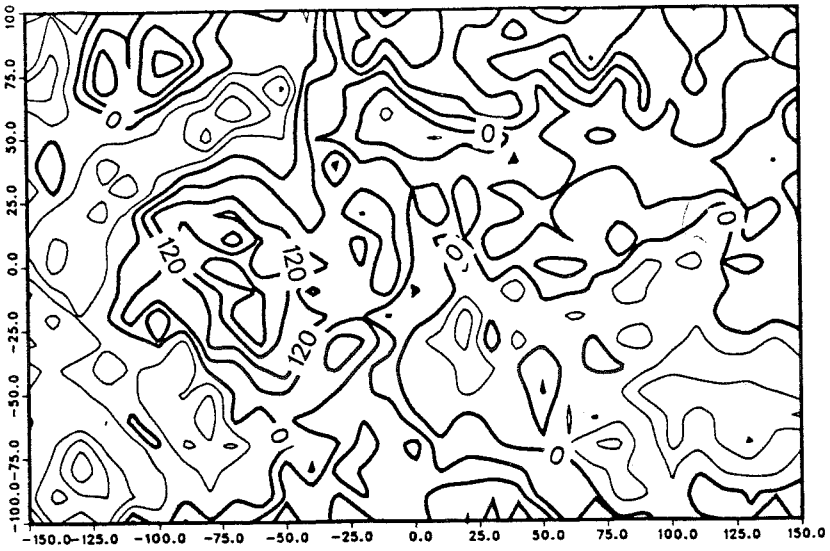
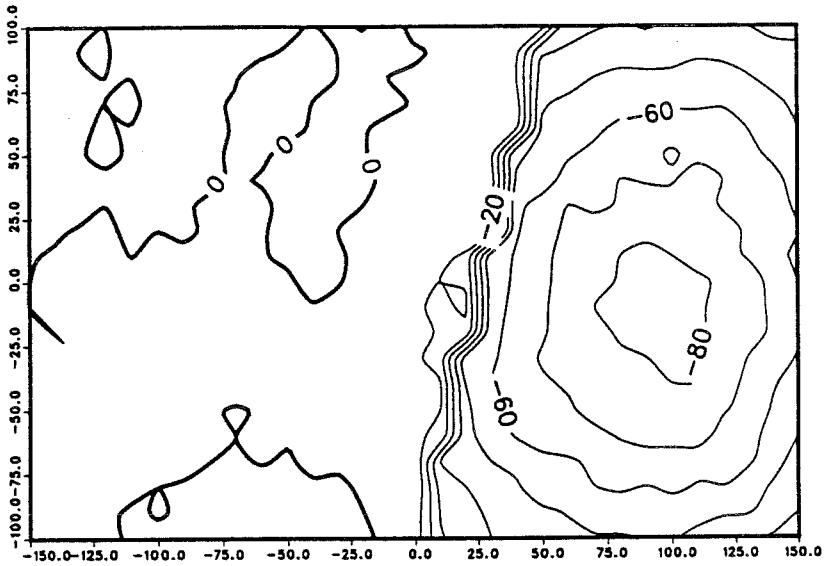


Fig. 3-23:Exemple de champ non homogène simulé sur le domaine de la Fig. 3-20. A gauche C(r) = exp(-r / 3Dx), à droite C(r) = exp(-r / 6Dx).

(a)



(b)

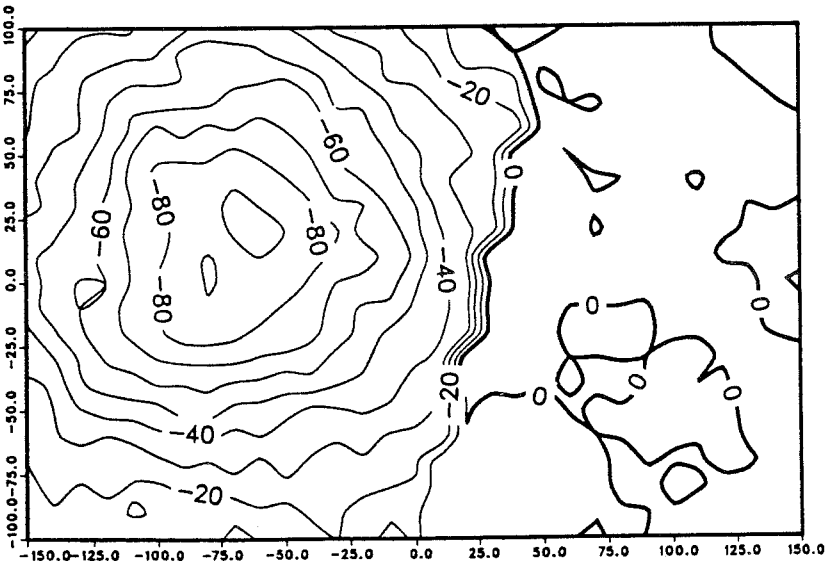


Fig. 3-24:Fonctions propres recalculées à partir de la matrice de covariance empirique de 1000 réalisations de champs analogues à celui de la Fig. 3-2 .
(a) F.P. 1 (b) F.P. 2.

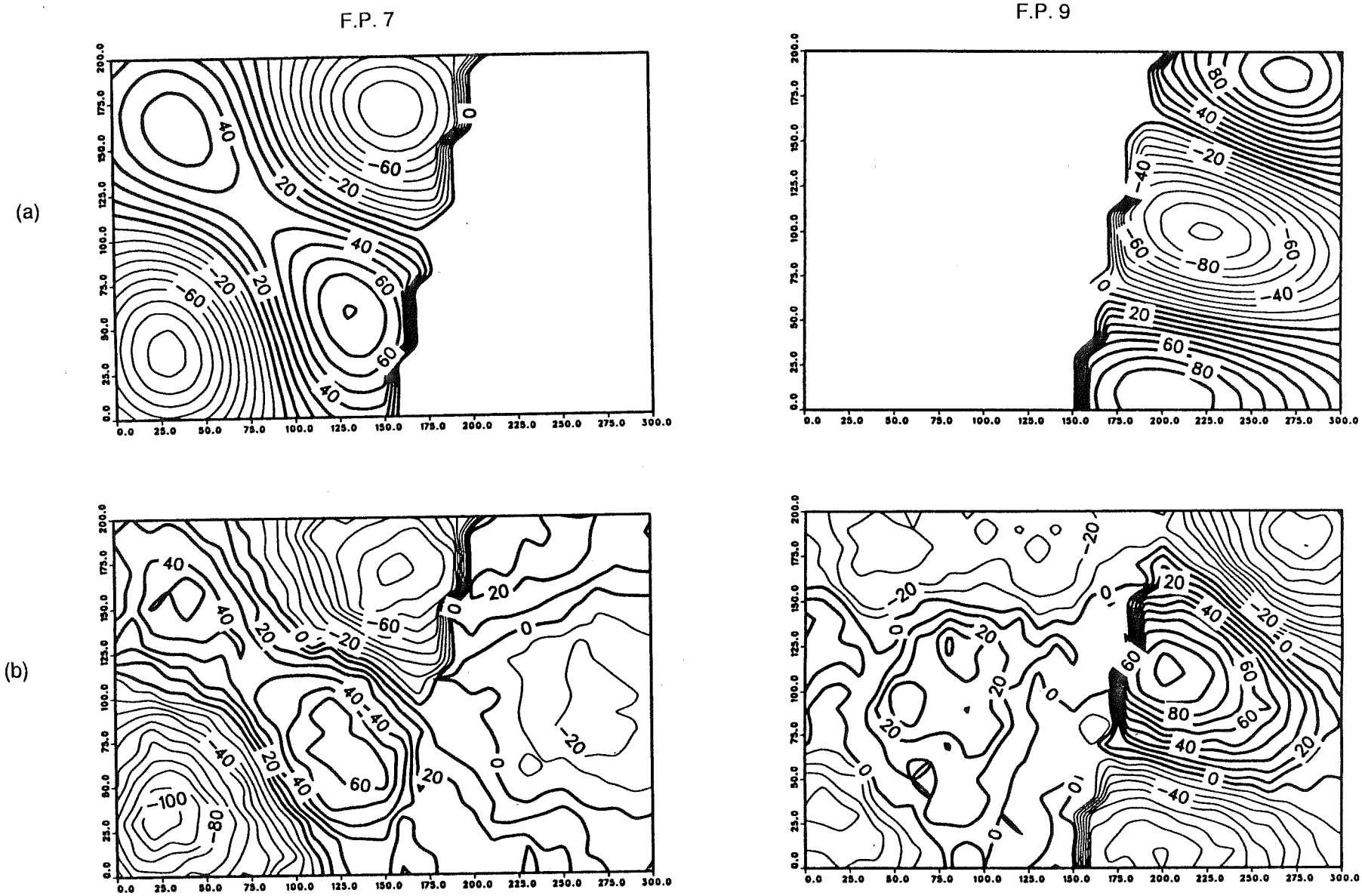


Fig 3-25: Comparaison des F.P. 7 (à gauche) et 9 (à droite) estimées
(a) en remplissant la matrice **C** à l'aide des modèles théoriques
(b) en utilisant la matrice des covariances empiriques de 1000 champs simulés par ACPP.
Sur D_1 $C(r) = \exp(-r / 3Dx)$ et sur D_2 $C(r) = \exp(-r / 6Dx)$

Globalement donc, l'ACPP a permis de répondre de façon satisfaisante au problème posé: la simulation de champs non homogènes. Mais, elle peut nous apporter plus encore en nous donnant, dans un contexte de multiréalisations, les moyens de détecter la présence de la faille. En effet, si on regarde à l'oeil nu un champ simulé (Fig. 3-23), elle n'apparaît pas de façon évidente. En revanche, si on cartographie les premières F.P. obtenues par ACPP de la matrice de covariance empirique du jeu de données simulé, elle est nettement visible (Fig. 3-24). On peut douter qu'une simple analyse variographique sur l'ensemble des données ait pu permettre une localisation aussi claire.

Par ailleurs, le tracé des F.P. recalculées sur le jeu de données simulées, après comparaison avec des F.P. "théoriques" calculées (de façon discrète) en remplissant **C** à l'aide des modèles, nous a permis de mettre en évidence, encore une fois, les incertitudes introduites par l'estimation des covariances sur un échantillon de taille finie. En effet, alors que les F.P. "théoriques" sont uniquement associées à un seul sous-domaine, les F.P. estimées correspondantes peuvent en être des combinaisons linéaires. Une F.P. estimée peut alors posséder des valeurs non nulles sur l'ensemble du domaine (Fig. 3-25). On voit donc encore ici une confirmation que la taille finie de l'échantillon ne permet plus d'assurer une bonne séparabilité des valeurs propres et des fonctions propres.

3-7 CONCLUSIONS

Dans le chapitre 1, nous avons souligné les liens étroits qui existaient entre l'ACPP et le krigeage. Cette dernière technique a conduit au développement d'une méthode de simulation que nous avons décrite au § 3-2-3: la méthode des bandes tournantes (Matheron, 1973). Nous voulions donc voir si l'analogie géostatistique/ ACPP ne pouvait pas être poussée plus loin, et si l'ACPP ne pouvait pas être utilisée pour la simulation de champs bidimensionnels.

Moyennant quelques hypothèses simplificatrices sur la structure du résidu $\varepsilon(\xi,\underline{x})$ ajouté, ceci est effectivement possible. Les champs simulés sont alors gaussiens. Pour des variables non normales, nous avons vu qu'une anamorphose prélable à la simulation était souhaitable. Et, si on choisit judicieusement les paramètres: nombre K de fonctions propres intervenant dans la décomposition (90% de variance expliquée au moins), une discrétisation spatiale qui assure une bonne estimation des F.P. (voir chapitre 2), nous avons pu voir que l'ACPP donnait des résultats comparables à la méthode LU ou aux bandes tournantes, même si quelques problèmes (anisotropie, effets de bord) apparaissent, en liaison avec la troncature de la série des F.P.

Par ailleurs, pour des champs non homogènes, un exemple simple nous a permis de voir que, sans "effort" supplémentaire, l'ACPP permettait de simuler de tels champs, avec un restitution satisfaisante du modèle de covariance. A la limite même, l'ACPP permet de mettre en évidence ces inhomogénéités par la cartographie des F.P. (Attention cependant aux fluctuations d'échantillonnage qui peuvent faire apparaître des non-stationnnarités artificielles...)

Enfin, au niveau des temps de calcul, l'étape la plus longue de la simulation par ACPP est le calcul des F.P. (avec la diagonalisation de **C.E**). Mais, une fois ce calcul effectué, le "coût marginal" d'une simulation supplémentaire est réduit. On voit donc l'intérêt que pourrait présenter son utilisation en Météorologie ou en Océanographie, puisque les EOFs sont en général calculées (pour d'autres objectifs, bien sûr), et, qu'alors, on pourrait simuler aisément un grand nombre de champs, afin de tester, par exemple, la sensibilité d'un modèle de circulation générale à la variabilité des entrées.

CHAPITRE 4: APPLICATION DE L'ACPP A LA TEMPERATURE MENSUELLE DE SURFACE DE LA MER (SST) SUR L'ATLANTIQUE INTERTROPICAL DE 1964 A 1987.

4-1 INTRODUCTION

Après avoir présenté une recherche plutôt méthodologique et surtout basée sur la simulation, il nous a paru souhaitable d'illustrer notre démarche en traitant des données "réelles", c'est-à-dire décrivant une grandeur physique d'intérêt en Océanographie, en Climatologie ou en Météorologie.

Nous aurions pu reprendre des données qui nous sont familières: champs de précipitation (Obled et Creutin, 1986), d'insolation (Goussebaïle, 1977) ou les champs de pression (Berlin et Cendrier, 1986) . Les techniques développées dans cette thèse leur ont déjà été partiellement appliquées. Par contre, dans ces exemples, le domaine d'étude n'est pas naturellement borné, ou imparfaitement, et on a vu l'influence de ce domaine sur les résultats de la méthode. C'est pourquoi nous avons recherché un exemple associé à un domaine bien délimité, par exemple un bassin océanique, ce qui nous a conduit à nous intéresser aux fichiers de températures mensuelles de surface de la mer (Sea Surface Temperature ou SST en Anglais) établis par J. Servain sur l'Atlantique intertropical entre 1964 et 1987. On peut trouver chez Picaut et al. (1985) ou Servain et al. (1987) les détails concernant la constitution de ces fichiers. Nous en redonnerons les principales étapes au § 4-2-1.

Sur ces données, et c'est pour nous l'un de leur intérêt principal, une étude par ACP "classique" a déjà été réalisée (Piedelièvre, 1985; Servain et Legler, 1986). Nous pensions donc avoir tous les éléments pour évaluer les apports éventuels d'une approche de type processus (ACPP) sur ces mêmes données. Il est en effet clair que la SST sur l'Atlantique intertropical possède une structure spatiale forte. S'y ajoute par ailleurs une structure temporelle, puisque les données mensuelles sont aussi autocorrélées. (Soulignons d'ores et déjà qu'à cause de cette autocorrélation temporelle, l'une des hypothèses faites lors des études théoriques précédentes: l'indépendance statistique des différentes réalisations, sera donc violée et nous essaierons d'en mesurer les effets).

Dans ce chapitre, nous présenterons donc, dans un premier temps, le fichier de données et son mode d'obtention ainsi que quelques statistiques de base (analyse des

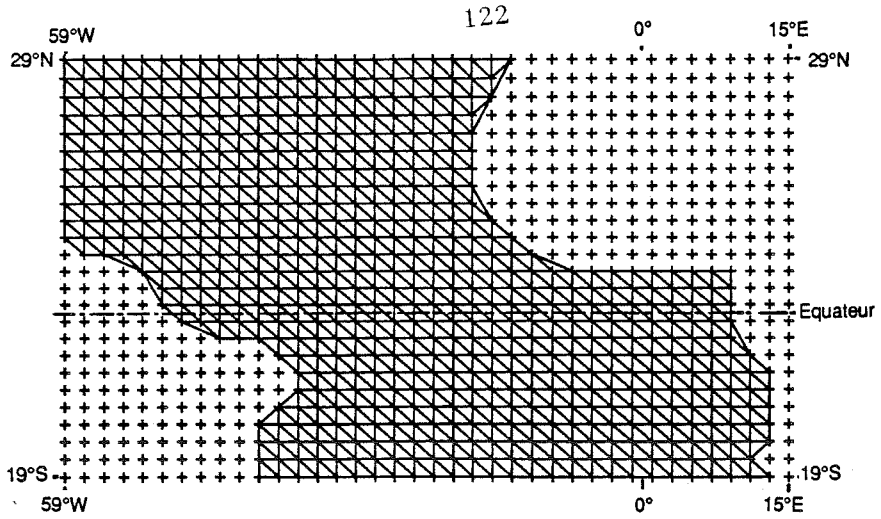
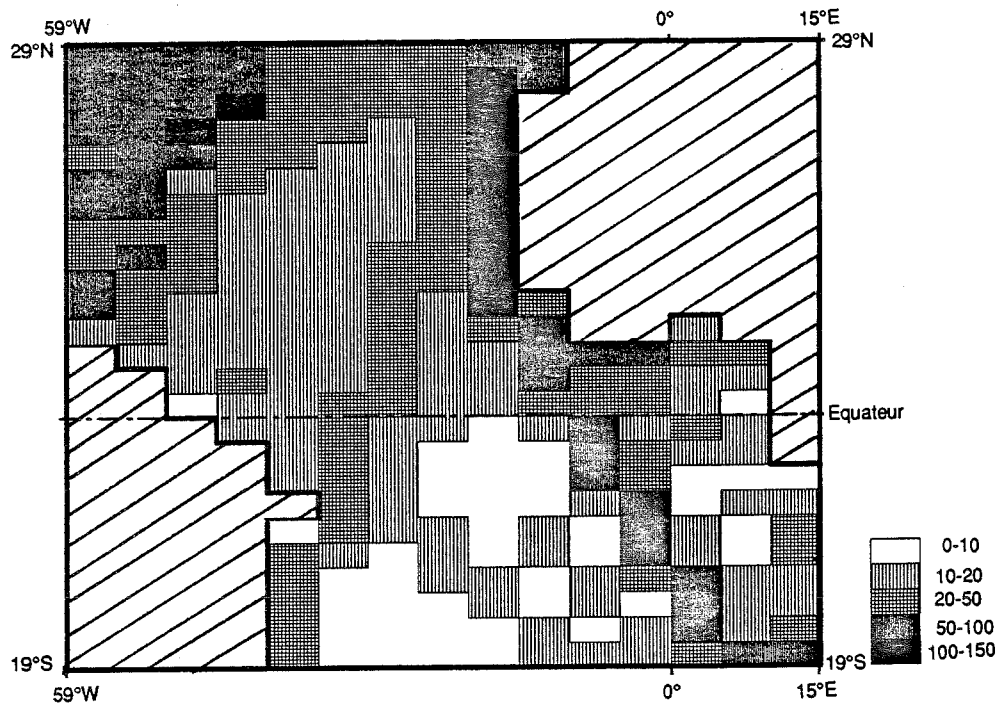


Fig. 4-1: Discrétisation du domaine d'étude (grille de 2° par 2° entre 19°S et 29°N et 59°W et 15°E)

(a)



(b)

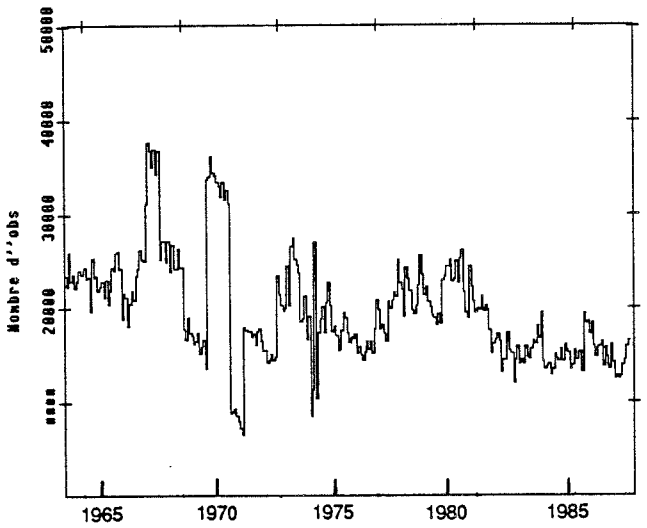


Fig 4-2: (a) Nombre moyen d'observations mensuelles par boîte de 5° par 2°.
(b) Nombre total d'observations par mois.

moyennes, des écarts-types et des distributions). Puis, nous ferons une analyse, désormais classique, de la structure spatiale du phénomène par l'intermédiaire du calcul de variogrammes. Ensuite, nous aborderons l'étude par ACP proprement dite. Tout d'abord, nous présenterons comment on peut mettre en oeuvre sur la sphère, la méthode numérique décrite au chapitre 1 dans le plan, car le domaine considéré est trop étendu pour être assimilé à un plan.

Puis, nous effectuerons la comparaison entre l'approche classique et l'approche de type processus continu. Cette démarche concernera tant le fichier total de 2° de longitude par 2° de latitude, qu'un ficher "dégradé" où nous aurons retiré aléatoirement un certain nombre de points de mesure (étude de la stabilité des résultats). Enfin, nous clôturerons ce chapitre par un essai de validation de plusieurs techniques basées sur l'ACP, en reconstitution (cf chapitre 1) sur les années 1986 et 1987.

4-2 PRESENTATION DES DONNEES

4-2-1 Obtention du fichier final.

Les données ayant servi de base à la constitution du fichier final, une grille régulière de 2° par 2° sur l'Atlantique intertropical entre 20° Sud et 30° Nord (voir Fig. 4-1 pour la discrétisation du domaine), sont des mesures de la SST effectuées par des navires marchands. L'ensemble des données recueillies sur chaque "boîte" de 5° par 2° , a ensuite servi à calculer les moyennes mensuelles correspondantes. Mais, comme le soulignaient déjà Servain et al. (1985), une grande inhomogénéité existe quant au nombre de mesures disponibles pour évaluer chacune de ces moyennes (entre 0 et 150...). En particulier, sur l'Atlantique Sud, on observe des zones pratiquement sans mesures, car c'est une zone très éloignée des routes maritimes où; en revanche, les données sont abondantes (voir la Fig. 4-2-a pour le nombre moyen d'observations mensuelles par boîte). Par ailleurs, si on regarde maintenant l'évolution temporelle du nombre de données recueillies, on observe une diminution assez sensible depuis le début des années 80, avec aussi quelques "trous" en 1971 et 1974 (Fig. 4-2-b).

Néanmoins, des précautions ont été prises et, notamment pour les boîtes contenant moins de 8 observations pour un mois, des tests d'homogénéité temporelle et spatiale (tests de Student) ont été menés avant le calcul des moyennes mensuelles. Les valeurs ne satisfaisant pas ces tests ont été éliminées, de même que celles où on avait moins de deux observations.

Ensuite, les valeurs de ce premier fichier sur une grille de 5° par 2°, ont été complétées et interpolées par la méthode de Cressman (1959), afin d'obtenir la grille finale de

2° par 2°. Cette méthode, encore connue sous le nom d'analyse objective, part d'un champ initial (ici les moyennes mensuelles tirées d'atlas climatologiques). On effectue ensuite une correction de ce champ de départ à l'aide d'un coefficient A, prenant en compte tous les points à l'intérieur d'un cercle de rayon R, centré sur le point étudié. Si on appelle e_i l'erreur entre la moyenne estimée sur les observations et le champ initial, A est donné par (Servain et al. , 1985):

$$A = \sum_{i=1}^{N} \{ [W(r) / d_i(r)] . e_i(r) \} / \sum_{i=1}^{N} W(r) / d_i(r) \qquad (4\text{-}1)$$

où - W(r) est la fonction de Cressman:

$$\begin{cases} W(r) = (R^2 - r^2) / (R^2 + r^2) & \text{si } r \leq R \\ W(r) = 0 & \text{si } r > R \end{cases} \qquad (4\text{-}2)$$

- $d_i(r)$ est la densité des données à l'intérieur d'un cercle de rayon 375 Km entourant chaque point. L'introduction de ce facteur permet de prendre en compte la corrélation spatiale entre points de mesure proches.

- N est le nombre de points intérieurs au cercle de rayon R. Si ce dernier ne contient aucun point, la valeur du champ initial est conservée.

Afin de supprimer des éventuelles instabilités, un filtre spatial (Schuman, 1957) est ensuite appliqué.

Le champ corrigé devient le nouveau champ initial et la procédure est itérée jusqu'à la convergence avec des rayons d'auscultation décroissants. Dans notre cas, R=1400 Km, 1000 Km, puis 600 Km (trois fois) ont été utilisés. Notons que le choix du dernier rayon est très important : il ne doit pas être trop grand pour ne pas déformer les champs près des côtes, ni trop petit car, alors, il n'y a plus assez de points à l'intérieur du cercle.

Cette rapide présentation de la démarche suivie pour l'obtention du fichier final montre clairement que plusieurs aspects se conjuguent pour aboutir à un <u>lissage très important</u> des données: intégration spatiale sur les boîtes de 5° par 2°, interpolation optimale, filtrage de Schuman; alors que l'observation des images infra-rouges de la surface de la mer met en évidence des zones où la variabilité de la SST est grande: bordures côtières, courants type Gulf Stream. Nous retrouverons d'ailleurs l'effet de ce lissage lors de l'analyse variographique des données au § 4-3. Signalons enfin que, plus un champ est lisse, plus il

devient difficile de départager des méthodes, et les conclusions auxquelles nous aboutiront iront dans ce sens. Mais, ce type de données, très lissées, est cependant d'usage courant dans toutes les initialisations ou forçages de modèles océanographiques et le cas étudié en est un exemple assez représentatif.

4-2-2 Statistiques simples sur les données.

Avant d'aborder l'analyse des données par ACPP, nous aimerions présenter quelques statistiques simples sur ce même fichier, notamment les moyennes, les écarts-types et les distributions. Ceci nous fournira une première idée de la stationnarité du phénomène.

La Fig. 4-3-a donne les moyennes à long terme (sur l'ensemble des 12 mois et des 24 années) et la Fig. 4-3-b, les écarts-types correspondants. Une première constatation s'impose: le phénomène n'est pas du tout stationnaire. La moyenne présente un effet zonal marqué, avec des valeurs plus élevées près de l'équateur, mais, néanmoins, les isolignes ne suivent pas complètement les parallèles et se resserrent quand on s'éloigne de l'équateur. Quant à la variabilité de la SST, elle est très forte au large de l'Afrique de l'ouest (îles du Cap Vert). On distingue aussi des zones où l'écart-type est assez élevé au nord du domaine, ainsi que sur l'Atlantique Sud, à l'est du domaine (au large des côtes africaines).

Pour expliquer cette variabilité de la SST, il faut faire appel aux grands courants maritimes de surface, dont Piedelièvre (1985) donne une brève description. Nous en redonnons ici les éléments importants. On distingue ainsi trois courants principaux:

i) *le Courant Sud Equatorial (CSE)*

Il coule vers l'ouest entre 20°S et 4°N et se divise en deux branches au contact de l'Amérique : l'une qui remonte vers le nord le long des côtes brésiliennes, l'autre qui descend vers le Sud.

Ce courant provoque des remontées d'eau froide vers l'équateur entre Avril et Septembre (on retrouve cet effet sur la moyenne à long terme avec une "bosse" (Fig. 4-3-a)) et ceci permet d'expliquer la variabilité plus forte sur l'Atlantique du Sud-Est (Fig 4-3-b).

ii) *le Courant Nord Equatorial (CNE)*

Ce dernier coule aussi vers l'ouest entre 4°N et 25°N. Ses mouvements saisonniers peuvent expliquer la plus forte variabilité notée sur le nord du domaine.

ii) *le Contre Courant Sud Equatorial (CCSE)*

Il vient "s'intercaler" entre les précédents (entre 5°N et 10°N) mais coule vers l'est. Son origine est thermodynamique (rétablissement de l'équilibre de la surface océanique modifiée par les vents d'est) et il se manifeste essentiellement de Juin à Septembre.

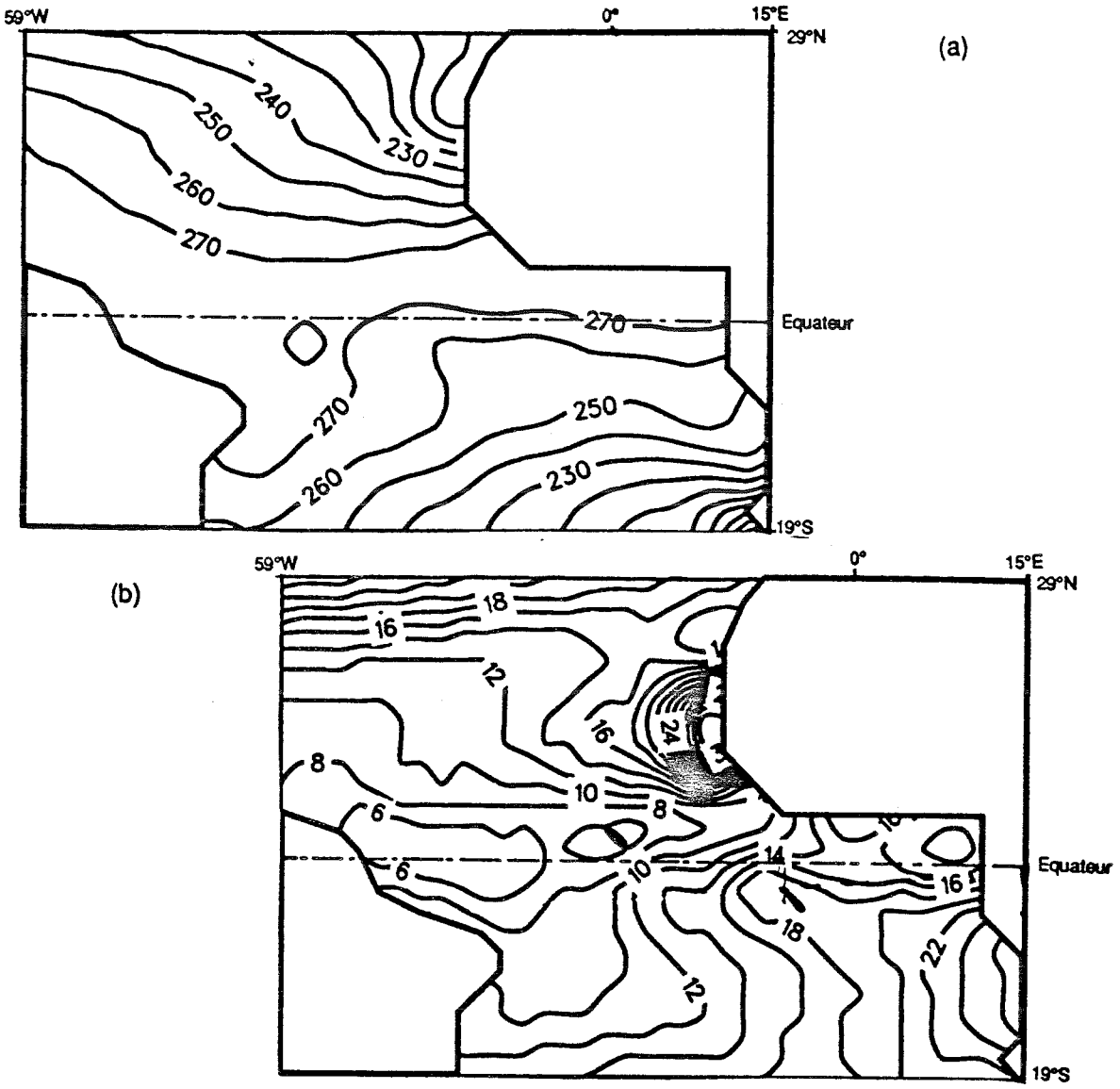


Fig. 4-3: Moyenne (a) et écart-type (b) climatologiques (tous mois et toutes années confondus) de la SST brute sur la période 1964-1985. (SST en 1/10 °C)

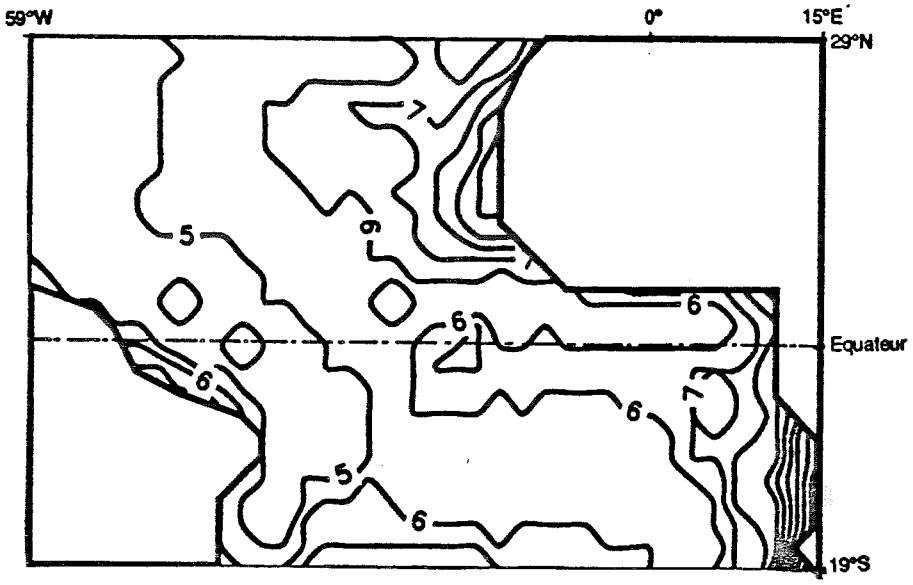


Fig. 4-4: Ecart-type climatologique (tous mois et toutes années confondus) des SST désaisonnalisées sur la période 1964-1987.(SST en 1/10 °C)

A ces grands courants marins viennent s'ajouter des courants plus locaux, liés aux remontées d'eaux froides sub-superficielles: les upwellings.

i) *le courant des Canaries*

Ce courant descend avec la direction NE-SW, le long des côtes africaines et se charge en eaux froides de Mai à Aout-Septembre. Les mouvements provoqués par ces upwellings côtiers induisent la forte variabilité observée autour des îles du Cap Vert (Fig. 4-3-b)

ii) *le golfe de Guinée*

C'est, comme la précédente, une zone d'upwellings côtiers avec des remontées d'eaux froides de Juin à Septembre, l'eau étant plus chaude en hiver. Ceci explique, encore une fois, la variabilité assez forte de la SST sur cette région.

Si on cherche à s'affranchir de la non-stationnarité du phénomène mise en évidence par les cartes de la Fig 4-3, on peut chercher d'abord à "désaisonnaliser" les données. En effet, une forte composante saisonnière qui "écrase" de son poids relatif élevé la variabilité totale, peut alors occulter d'autres effets, tout aussi importants pour la compréhension des mécanismes.

On retranchera alors à la variable SST brute que nous noterons $X(t_i, \xi_l, \underline{x}_j)$ où $\underline{x}_j$, j=1..P est la variable d'espace (pratiquement, nous avons pris les centres des boîtes de 2° par 2°), t_i, i=1..12 désigne le mois et ξ_l, l=1..24 représente l'année considérée, une moyenne, dite saisonnière (moyenne en un lieu $\underline{x}_j$ pour le mois t_i), notée $X_s(t_i, \underline{x}_j)$, j=1..P, i=1..12 qui sera estimée par:

$$X_s(t_i, \underline{x}_j) = 1/24 \sum_{l=1}^{24} X(t_i, \xi_l, \underline{x}_j) \qquad (4\text{-}3)$$

La nouvelle variable "désaisonnalisée" sera alors:

$$X_d(t_i, \xi_l, \underline{x}_j) = X(t_i, \xi_l, \underline{x}_j) - X_s(t_i, \underline{x}_j) \qquad (4\text{-}4)$$

X_d possède une moyenne nulle, mais, si on regarde la carte de l'écart-type (Fig. 4-4), on observe que X_d n'est pas stationnaire pour la variance et que des zones de forte variabilité subsistent, notamment aux abords des îles du Cap Vert et de la côte Angolaise. Néanmoins, les gradients sont moins forts que sur les données brutes.

Par ailleurs, ce centrage a un effet très positif sur la distribution de la SST. En effet, si X possède une distribution très dissymétrique, (Fig. 4-5-a), X_d peut être considérée comme une variable gaussienne (Fig. 4-5-b), souvent préférée dans les traitements statistiques.

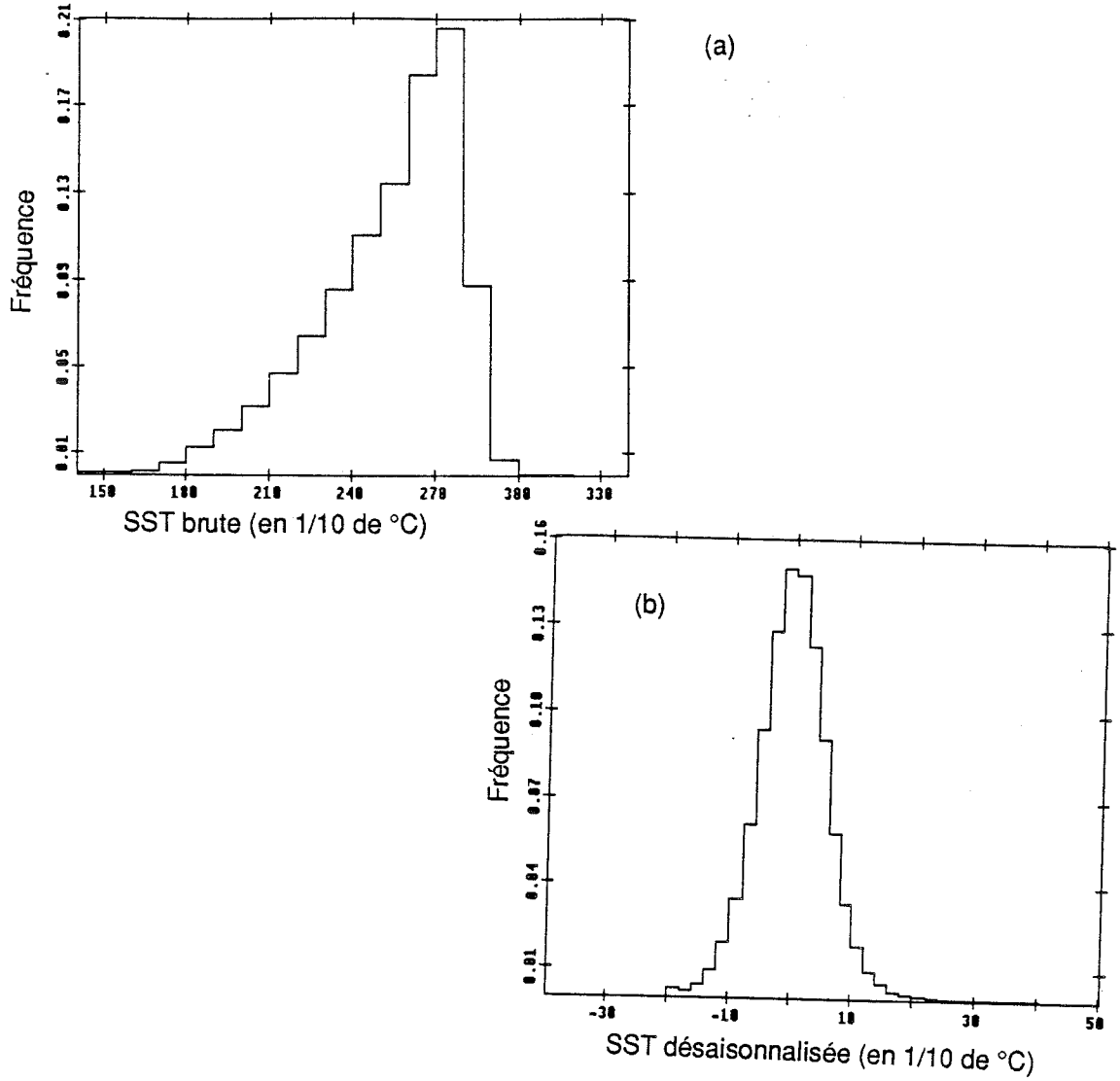


Fig. 4-5: Histogrammes globaux (toutes stations, tous mois et toutes années confondus) des SST (a) brutes
(b) désaisonnalisées.

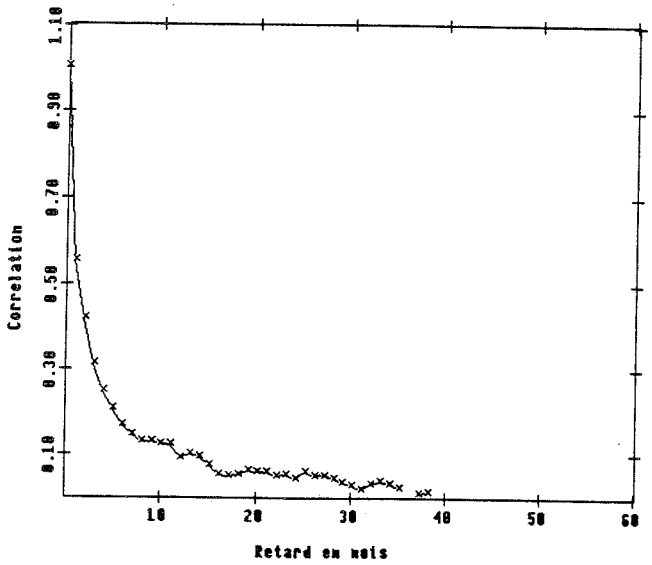


Fig. 4-6: Corrélogramme de l'ensemble du fichier des SST désaisonnalisées.

Pour en terminer avec ces généralités sur les données, nous évoquerons le problème de l'autocorrélation temporelle des séries. Sur l'ensemble des données (toutes stations, tous mois et toutes années confondus), le coefficient de corrélation entre deux mois successifs est de .55, ce qui est loin d'être négligeable (La Fig. 4-6 donne le corrélogramme complet pour les données désaisonnalisées). En particulier, l'incidence sur l'estimation de la matrice de covariance sera important car le nombre d'observations réellement indépendantes ne sera plus N=288, mais Ne= $(\mathrm{Tr}\mathbf{C})^2 / \mathrm{Tr}\mathbf{C}^2$ (Der Megreditchian , 1988, Tome 1, p. 169) où **C** est la matrice de variance-covariance et Tr**C** sa trace. Ceci augmentera encore l'effet des erreurs d'échantillonnage sur la qualité des F.P. calculées par ACP, qui est déjà sûrement très grand car on possède beaucoup plus de "stations" (en réalité le centre des boîtes) que de réalisations (P=628 sur la grille de 2° par 2°, N=12*24 = 288)

4-3 ANALYSE DE LA STRUCTURE SPATIALE DE LA SST.

4-3-1 Méthodologie utilisée.

Nous nous intéresserons successivement aux données brutes $X(t, \xi, \underline{x})$, puis aux données désaisonnalisées $X_d(t, \xi, \underline{x})$ définies par (4-4). Nous ne donnerons les formules que pour X, sachant qu'elles seront analogues pour X_d.

L'étude des moyennes et écarts-types climatologiques nous a montré, au § 4-2-2, que la SST, notamment la SST brute, était loin d'être stationnaire. Néanmoins, nous avons quand même calculé les variogrammes climatologiques:

$$\gamma(\underline{h}) = 1/2 \,.\, E\,[\,\{\, X(t, \xi, \underline{x}) - X(t, \xi, \underline{x} + \underline{h})\,\}^2\,] \qquad (4\text{-}5)$$

en espérant qu'ils présenteraient un palier ($\gamma(\underline{h})$ devient alors indépendant du vecteur $\underline{h}$).

D'après Matheron (1965), le processus étudié peut alors être considéré comme stationnaire d'ordre deux, ce qui suppose:

i) la stationnarité de la moyenne

ii) la stationnarité des moments d'ordre 2. On peut alors admettre que X possède une variance finie σ^2 (donnée par la valeur du palier) et $\gamma(\underline{h})$ est relié à la covariance par:

$$\gamma(\underline{h}) = \sigma^2 - C(\underline{h}) \qquad (4\text{-}6)$$

Si, en revanche, le variogramme ne se stabilise pas, on peut alors espérer que les écarts soient stationnaires d'ordre 2. C'est l'hypothèse intrinsèque:

$$E[\, X(t, \xi, \underline{x}) - X(t, \xi, \underline{x} + \underline{h})\,] = 0$$

$$E[\, \{ X(t, \xi, \underline{x}) - X(t, \xi, \underline{x} + \underline{h}) \}^2 \,] = 2.\, \gamma(\underline{h})$$

c'est-à-dire que le variogramme ne dépend que de l'écart entre deux points, mais pas du point d'appui $\underline{x}$ lui-même.

Physiquement, si le variogramme présentait un palier, la distance à partir de laquelle il serait atteint (la portée) pourrait s'interpréter comme la taille des régions où se développe la variabilité, ce qui, dans notre cas, pourrait définir l'échelle des zones de refroidissement ou de réchauffement de la SST, provoqués par les mouvements des grands courants ou les upwellings. La valeur du palier nous donnerait quant à elle l'ordre de grandeur de ces variations. Mais, comme on travaille sur des données intégrées par "boîtes", on réduit la variabilité et on tend à augmenter un peu la portée réelle.

D'un point de vue pratique, les variogrammes ont été estimés à l'aide des N=288 réalisations disponibles, globalement (c'est-à-dire en prenant en compte toutes les directions possibles pour $\underline{h}$) ou par azimut (en privilégiant une direction α de $\underline{h}$). L'estimateur retenu a alors la forme suivante:

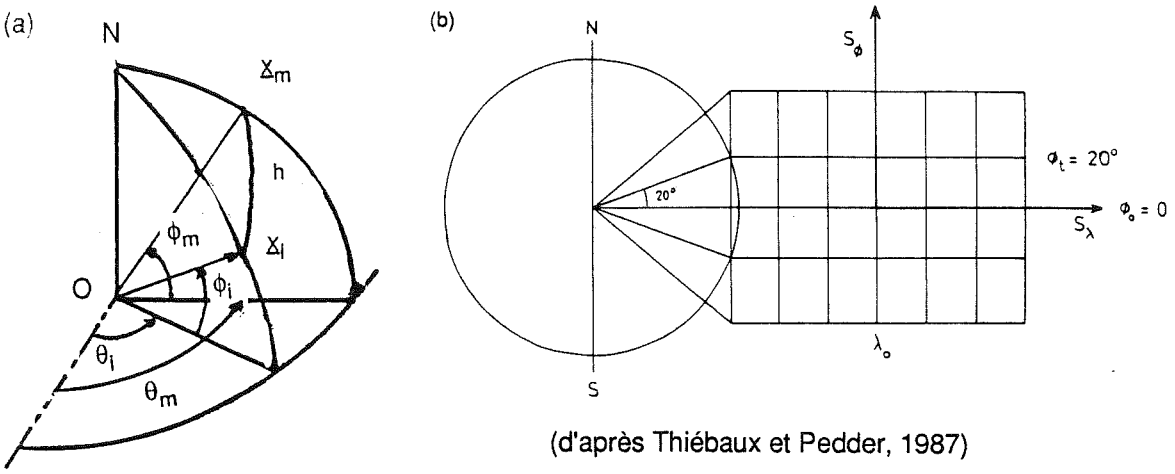
$$\gamma(h,\alpha) = 1/[2.Nc(h,\alpha)] \,.\, \sum_{j=1}^{Nc(h,\alpha)} \sum_{i=1}^{12} \sum_{l=1}^{24} \{ X(t_i, \xi_l, \underline{x}_j) - X(t_i, \xi_l, \underline{x}_j + \underline{h}) \}^2 \qquad (4\text{-}7)$$

où $Nc(h,\alpha)$ est le nombre de couples $(\underline{x}_j, \underline{x}_j + \underline{h})$ distants de h, dans la direction α. En pratique, α est définie par un cône de largeur 2δ, et on retient tous les couples dont l'angle (en coordonnées Mercator, voir Fig. 4-7-b) appartient à $[\, \alpha - \delta, \alpha + \delta \,]$. La Fig. 4-8 présente les quatre directions d'investigation retenues dans cette étude.

Par ailleurs, la distance h entre deux points $\underline{x}_i$ et $\underline{x}_j$, définis par leur latitude et leur longitude respectives (θ,ϕ) (Fig. 4-7-a), calculée sur la sphère est (Thiébaux et Pedder, 1987):

$$h = R_T \,.\, \mathrm{Arcos}\,[\, \cos\phi_i \,.\, \cos\phi_j \,.\, \cos(\theta_i - \theta_j) + \sin\phi_i \,.\, \sin\phi_j \,] \qquad (4\text{-}8)$$

où R_T = 6370 Km est le rayon terrestre moyen.



(d'après Thiébaux et Pedder, 1987)

Fig. 4-7: (a) Définition de la longitude et de la latitude (θ,ϕ) ainsi que de la distance sur la sphère
(b) Projection Mercator $S_\theta = R_T \,.\, (\theta - \theta_o)$; $S_\phi = R_T \,.\, \mathrm{Ln}\,[\,\mathrm{tg}(\Pi/4 + \phi/2)\,]$

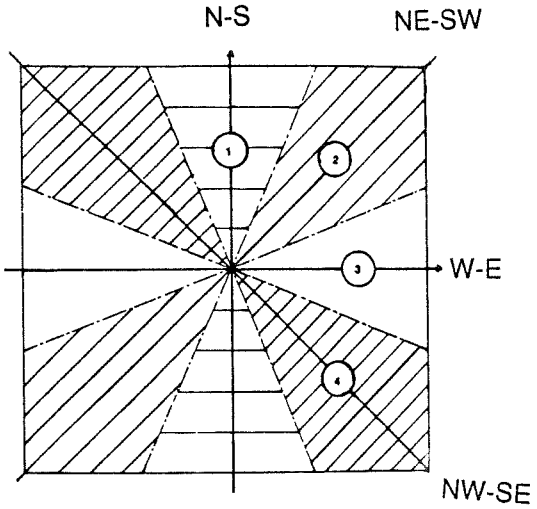


Fig. 4-8: Les différentes directions retenues pour le calcul des variogrammes par azimut.

Après le calcul du variogramme brut $\gamma(\underline{h})$, vient la phase de modélisation de $\gamma(\underline{h})$ à l'aide d'une fonction analytique choisie parmi les modèles:

$$\textit{sphérique:}\ \begin{cases} \gamma^*(\underline{h}) = \sigma^{2*} \,.\, [\,1/2.\,(h/a_{sph}^*)^3 - 3/2.\,(h/a_{sph}^*)\,] & \text{si } h \leq a_{sph}^* \\ \phantom{\gamma^*(\underline{h})} = 0 & \text{sinon} \end{cases} \quad (4\text{-}9)$$

$$\textit{exponentiel:}\ \gamma^*(\underline{h}) = \sigma^{2*} \,.\, [\,1 - \exp(-h/a_e^*)\,] \quad (4\text{-}10)$$

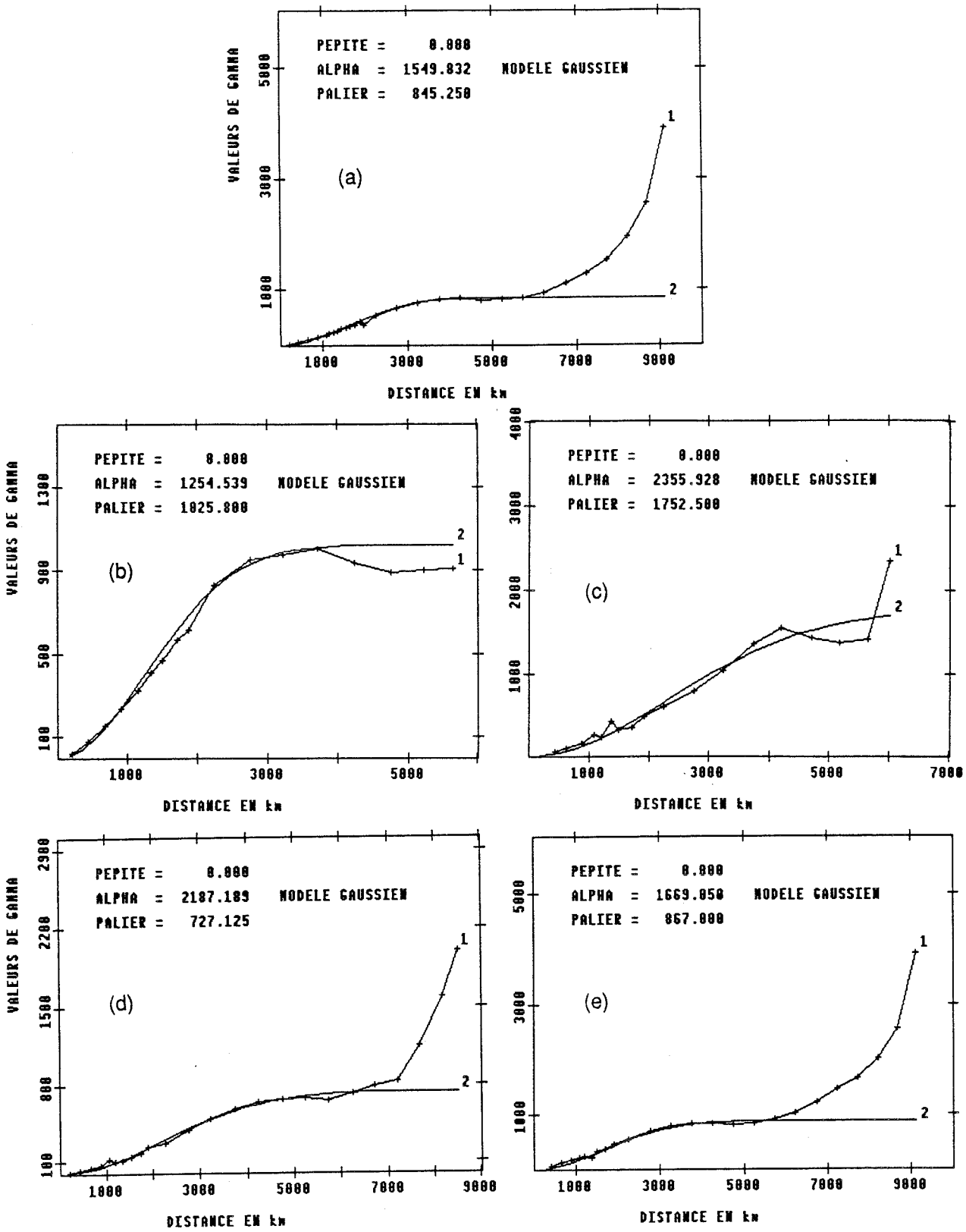


Fig. 4-9: Variogrammes climatologiques de la SST brute calculés sur l'ensemble du domaine.
(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

gaussien: $\gamma^*(\underline{h}) = \sigma^{2*} . [1 - \exp(-h^2 / a_g^{*2})]$ (4-11)

Les paramètres σ^{2*} et a_{sph}^* (resp. a_e^* et a_g^*) nous donnent alors les valeurs des paliers et portées modélisés (Rappelons qu'on peut trouver en Annexe F la définition des portées sphériques équivalentes à a_e^* et a_g^*). Ces modèles peuvent ensuite être utilisés pour faire, par exemple, de la simulation (chapitre 3) ou remplir une matrice de covariance servant à faire l'ACPP du processus.... C'est ce que nous ferons au § 4-5.

4-3-2 Analyse de la structure spatiale de la SST brute.

Les variogrammes climatologiques bruts ont été calculés puis modélisés sur l'ensemble du domaine (Fig. 4-9), puis sur chaque hémisphère Nord et Sud (Fig. 4-10 et 4-11 resp.), afin de voir si une symétrie des résultats se dégageait. Le tableau 4-1 regroupe quant à lui les valeurs des différents paramètres σ^{2*} et a_g^* modélisés (Par rapport aux Fig 4-9 à 4-11, les valeurs de σ^{2*} ont été divisées par 100 pour se ramener à des °C, les calculs ayant été faits en 1/10 de °C).

4-3-2-1 *Comportement au voisinage de l'origine.*

Notons que, pour ces données brutes, seul le modèle gaussien a été utilisé. Cette fonction est continue et dérivable au voisinage de l'origine, ce qui traduit la grande régularité de la variable (Matheron, 1972-b), qui présente alors une bonne continuité spatiale. Néanmoins, on est en droit de se demander si cette régularité est "naturelle" ou si elle ne serait pas due au lissage introduit par la méthode d'interpolation (analyse objective) rappelée au § 4-2-1. En effet, une analyse grossière, restreinte aux faibles distances indique que le choix de la fonction de Cressman (formule (4-2)) induirait, au voisinage de l'origine, un variogramme puissance h^β, avec $\beta>2$, ce qui correspond bien à l'allure expérimentale observée.

Cette incertitude qui pèse sur la source de ce comportement régulier aux faibles distances limite donc l'utilisation des variogrammes dans cette gamme de distances.

4-3-2-2 *Comportement pour les grandes distances.*

i) Les variogrammes sur l'ensemble du domaine (Fig. 4-9) présentent, en général, un palier stable entre 3000 et 7000 Km, la valeur de la portée sphérique équivalente variant entre 2800 et 5200 Km (Tableau 4-1). On pourrait donc être satisfait: le processus serait bien stationnaire d'ordre deux. Mais, hélas, à partir de 7000 Km (dans les directions où cette interdistance est atteinte), le variogramme se met à croître de façon parabolique. En langage

de géostatistique, on dit qu'on observe une dérive : les écarts ne sont plus stationnaires. D'après Creutin (1979), un comportement parabolique à l'infini peut être dû à une dérive linéaire, i.e.:

$$X(t, \xi, \underline{x}) - X(t, \xi, \underline{x} + \underline{h}) = a \cdot \underline{h} + b$$

Comment pourrait-on interpréter cette dérive? Vu les interdistances concernées, les couples mettent en jeu des points situés de part et d'autre de l'équateur. Donc, plus on s'éloigne de l'équateur, plus le gradient sur la SST est important. On le retrouve sur la Fig. 4-3-a de la moyenne climatologique où on observe un resserrement des isolignes lorsqu'on s'écarte de part et d'autre de l'équateur.

Si l'on s'attache maintenant à comparer les résultats par azimut, on observe que les directions W-E et NW-SE sont très proches du variogramme climatologique global, même si la portée est plus grande dans la direction W-E (4890 Km contre 3700 Km dans la direction NW-SE) (3500 Km globalement). Les variabilités sont quant à elles comparables: 2.7 à 2.9°C d'écart-type. En revanche, la direction N-S possède une portée plus faible (2800 Km) pour un écart-type de 3.2°C. Mais, c'est la direction NE-SW qui présente à la fois la plus grande portée (5300 Km) et la plus forte variabilité (4.2°C).

		Global	Direction N-S	Direction NE-SW	Direction W-E	Direction NW-SE
Ensemble du domaine	ag*	1550. (3466.)	1254. (2804.)	2356. (5268.)	2187. (4890.)	1669. (3732.)
	σ^2*	8.45	10.25	17.52	7.27	8.67
	σ*	2.9	3.2	4.2	2.7	2.9
Hémisphère Nord	ag*	1614. (3669.)	1661. (3715.)	2872. (6422.)	2017. (4511.)	1937. (4330.)
	σ^2*	9.33	17.04	39.67	6.32	10.49
	σ*	3.0	4.1	6.3	2.5	3.2
Hémisphère Sud	ag*	1611. (3603.)	841. (1882.)	706. (1579.)	2240. (5009.)	1829. (4091.)
	σ^2*	7.78	6.77	2.39	8.49	16.55
	σ*	2.8	2.6	1.5	2.9	4.1

Tableau 4-1: Paramètres σ^{2*} et $a_g{}^*$ des variogrammes climatologiques modélisés de la SST brute. Pour $a_g{}^*$, la valeur entre parenthèses est la portée sphérique équivalente . Les valeurs de σ^{2*} ont été divisées par 100 par rapport aux graphiques pour avoir des valeurs en °C.

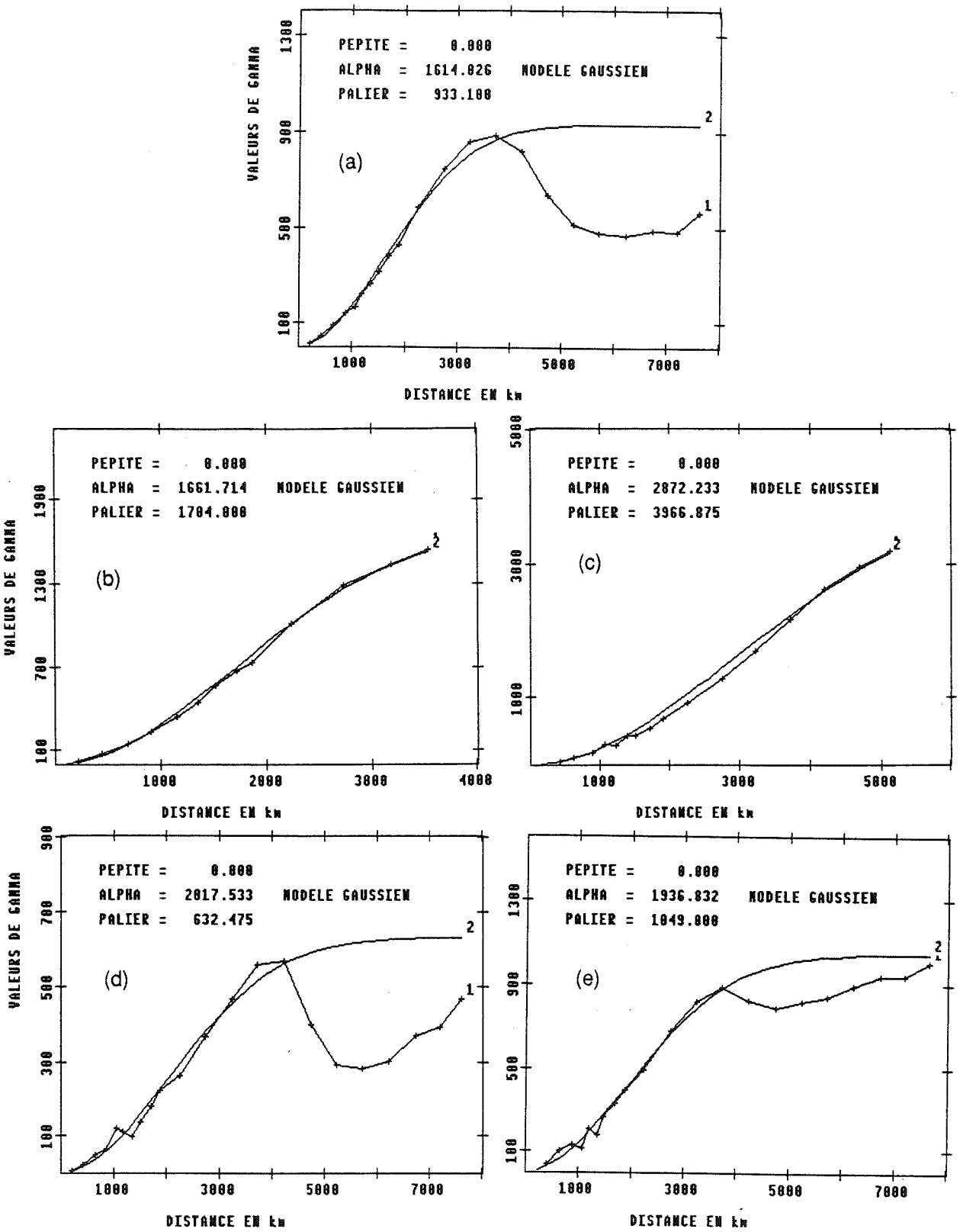


Fig. 4-10:Variogrammes climatologiques de la SST brute calculés sur l'hémisphère Nord.
(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

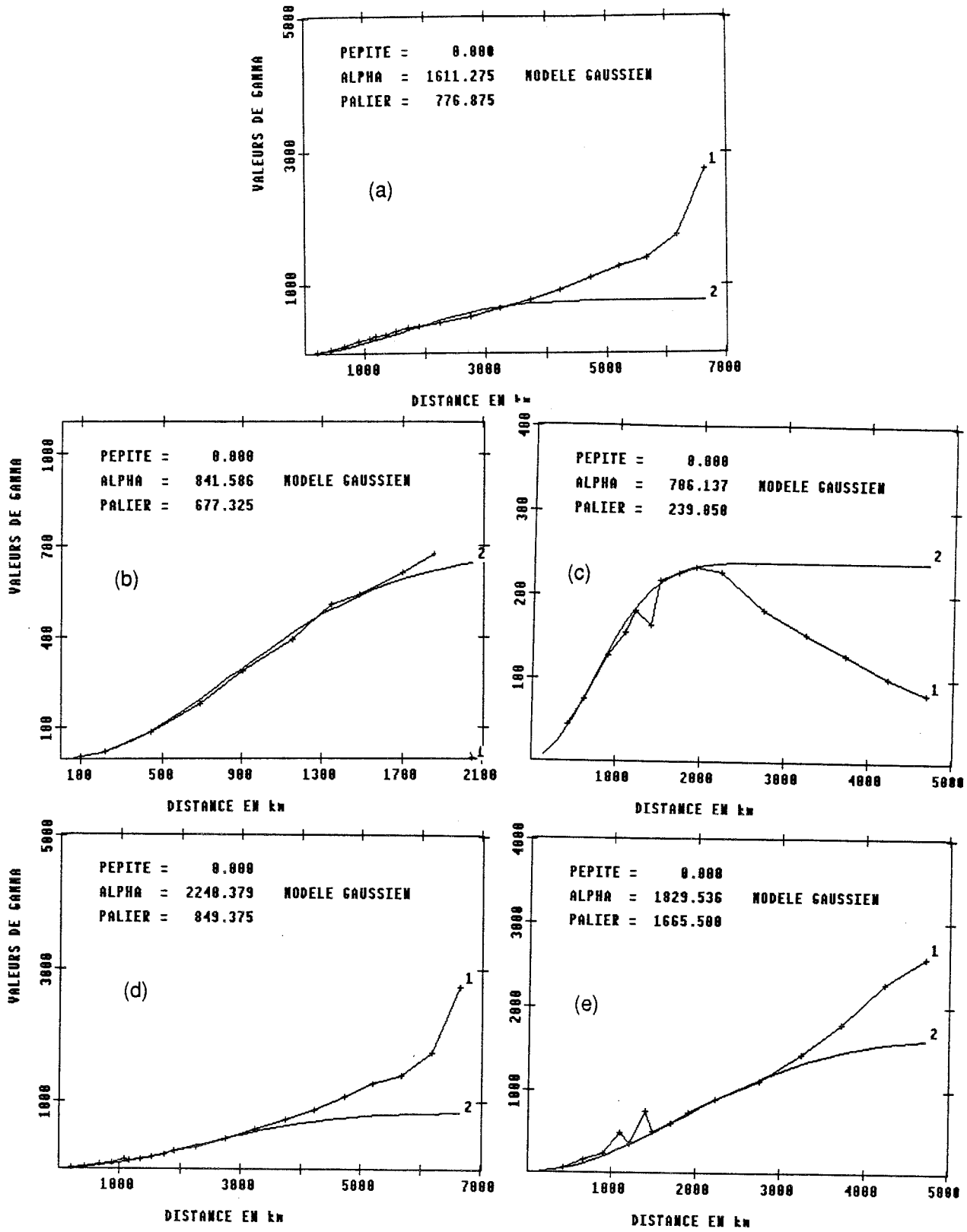


Fig. 4-11:Variogrammes climatologiques de la SST brute calculés sur l'hémisphère Sud.
(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

ii) Mais, afin de donner une interprétation possible de ces différences, il convient d'examiner les résultats selon les deux hémisphères (Fig 4-10 et 4-11). Ils sont cohérents sur les variogrammes globaux avec une variabilité plus grande au Nord. Cependant, l'observation des résultats par azimut vient encore moduler toutes ces affirmations générales.

Dans la direction N-S, la variabilité est beaucoup plus forte au Nord qu'au Sud (4.1°C contre 2.7°C) et est bien plus étendue spatialement (portée de 3700 Km contre 1800 Km). Peut-être pourrait-on relier la forte variabilité au Nord aux mouvements du Courant Nord Equatorial?....

C'est dans la direction NE-SW qu'on observe la plus grande variabilité (6.1°C) au Nord, avec une zone d'extension très grande: 6400 Km. Cette direction correspond aux mouvements de l'upwelling des Canaries dont nous avons vu qu'il provoquait aussi les plus forts gradients d'écart-type (Fig. 4-3-b). En revanche, dans cette direction la SST sur l'hémisphère Sud est très stable: variabilité et portée minimales: 1.5°C et 1500 Km. On conçoit, au vu de ces différences que le variogramme dans cette direction, sur l'ensemble du domaine, soit irrégulier et peu facile à modéliser (Fig. 4-9-c).

Dans les directions W-E et NW-SE, les variogrammes, entre 3000 et 7000 Km, ont des comportements tout à fait opposés: au Nord, ils s'affaissent, alors qu'au Sud, ils ont tendance à croître indéfiniment. La combinaison des deux donne un magnifique palier pour l'ensemble du domaine (Fig. 4-9-d et e)... La variabilité est, cette fois, plus forte au Sud qu'au Nord (2.9°C contre 2.5°C dans la direction W-E et 4.1°C contre 3.2°C dans la direction NW-SE). Les portées varient entre 4000 et 5000 Km et, on pourrait peut-être relier les résultats dans l'hémisphère Sud aux remontées du CSE vers l'équateur. Les phénomènes d'upwellings côtiers dans le golfe de Guinée ou au large de l'Angola se passent à des échelles plus faibles et sont peut-être occultés.

4-3-2-3 *Conclusions.*

Pour conclure sur cette analyse, elle nous a montré, encore une fois, la non-stationnarité de la SST et surtout, les grandes inhomogénéités et anisotropies entre les deux hémisphères et les différentes directions d'étude, qui peuvent être renforcées par la distribution non gaussienne des données, mais aussi par la distribution hétérogène de leur disponibilité (des moyennes sur 10 ou 100 valeurs peuvent induire des variabilités différentes....). On peut néanmoins mettre en évidence des zones de variabilité dont la portée varie de 3000 à 6000 Km, avec des valeurs moyennes de 3°C, pouvant parfois atteindre 4 voire 6°C.

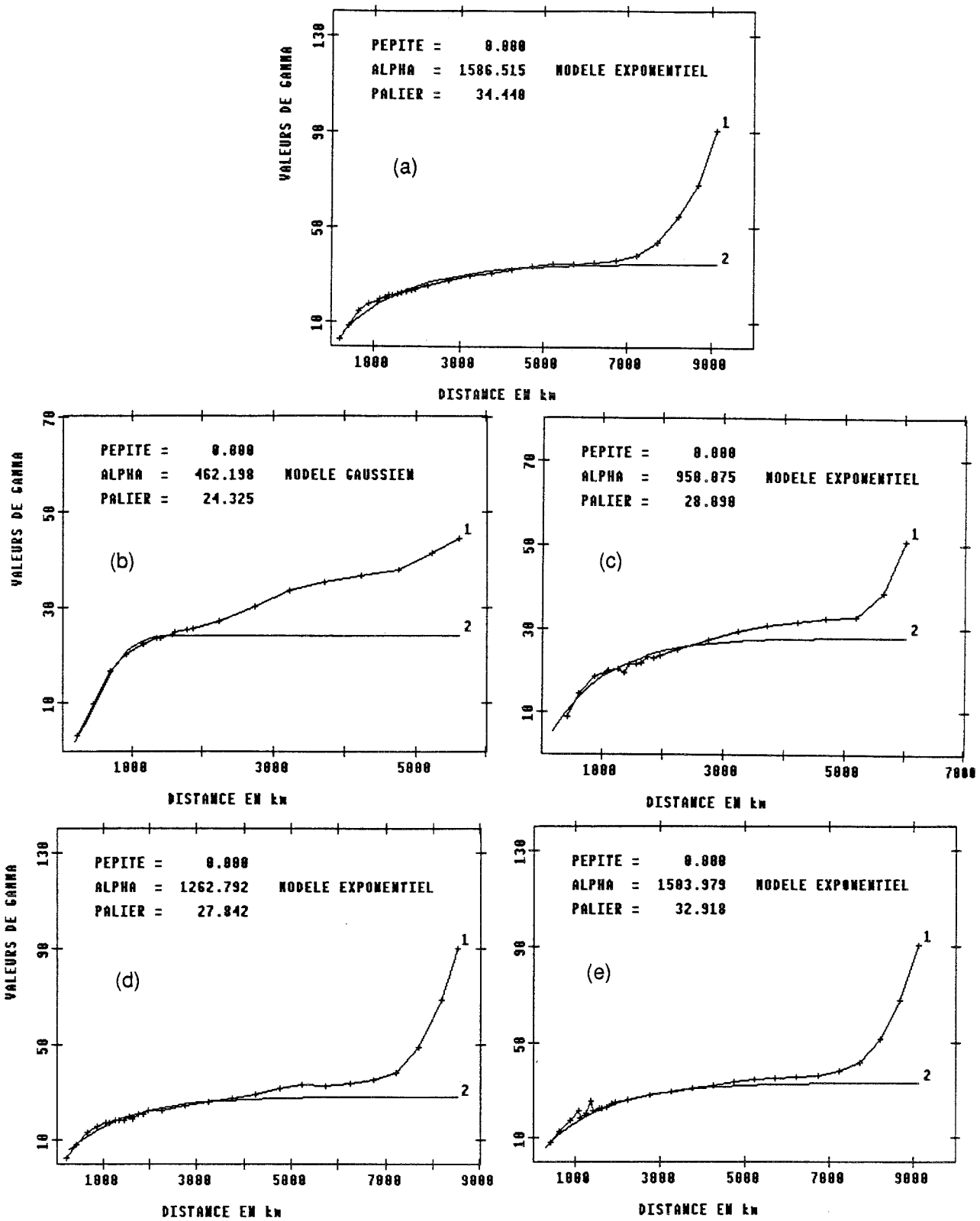


Fig. 4-12: Variogrammes climatologiques de la SST désaisonnalisée calculés sur l'ensemble du domaine.
(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

En ce qui concerne les variogrammes, nous avons aussi noté, au voisinage de l'origine, la grande régularité des données, probablement induite par l'analyse objective préalable.

L'étude de la structure spatiale de la SST brute, où les variations saisonnières dominent le processus, fait apparaître un caractère local très marqué de la variabilité. En revanche, la même analyse de la SST désaisonnalisée nous informera sur la structure spatiale de la variabilité interannuelle des anomalies de SST. C'est ce que nous allons voir maintenant.

4-3-3 Analyse de la structure **spatiale** de la SST désaisonnalisée.

Comme pour la SST brute, les variogrammes bruts ont été calculés sur l'ensemble du domaine (Fig. 4-12) puis sur les hémisphères Nord et Sud (Fig 4-13 et 4-14 resp.). Sauf dans la direction N-S où il est gaussien, le modèle utilisé est de type exponentiel. Le tableau 4-2 résume les valeurs des paramètres des différents variogrammes modélisés.

4-3-3-1 *Comportement au voisinage de l'origine.*

Le modèle exponentiel, s'il est toujours continu à l'origine, n'est plus dérivable, contrairement au modèle gaussien. La régularité de la SST désaisonnalisée est donc moins grande que celle de la SST brute, ce qui paraît logique puisqu'on travaille sur des anomalies. Cette moins grande régularité de la donnée de base, qui peut impliquer une variabilité se développant sur une plus petite échelle, peut-être du même ordre de grandeur que celle du lissage.., peut contribuer à atténuer, voire occulter l'effet de ce dernier. Une modélisation respectant les règles classiques (c'est-à-dire cherchant à bien reproduire le variogramme aux faibles distances) sera donc plus fiable.

4-3-3-2 *Comportement au voisinage de l'infini.*

i) Comme pour la SST brute, les variogrammes calculés sur l'ensemble du domaine présentent un palier situé entre 3000 et 7000 Km, avant de voir, à nouveau apparaître une dérive, qui, cette fois, s'avère plus difficile à interpréter puisque la moyenne climatologique des champs est nulle. Ce qu'on peut dire, c'est que, comme précédemment, ce sont des couples situés de part et d'autre de l'équateur qui sont mis en jeu.

Dans toutes les directions, la variabilité est équivalente (entre 0.49 et 0.58°C). Les portées, quant à elle, varient selon les directions et sont plus grandes dans les directions W-E et NW-SE que N-S ou NE-SW. (Mais, compte tenu de la géométrie du domaine, ce sont

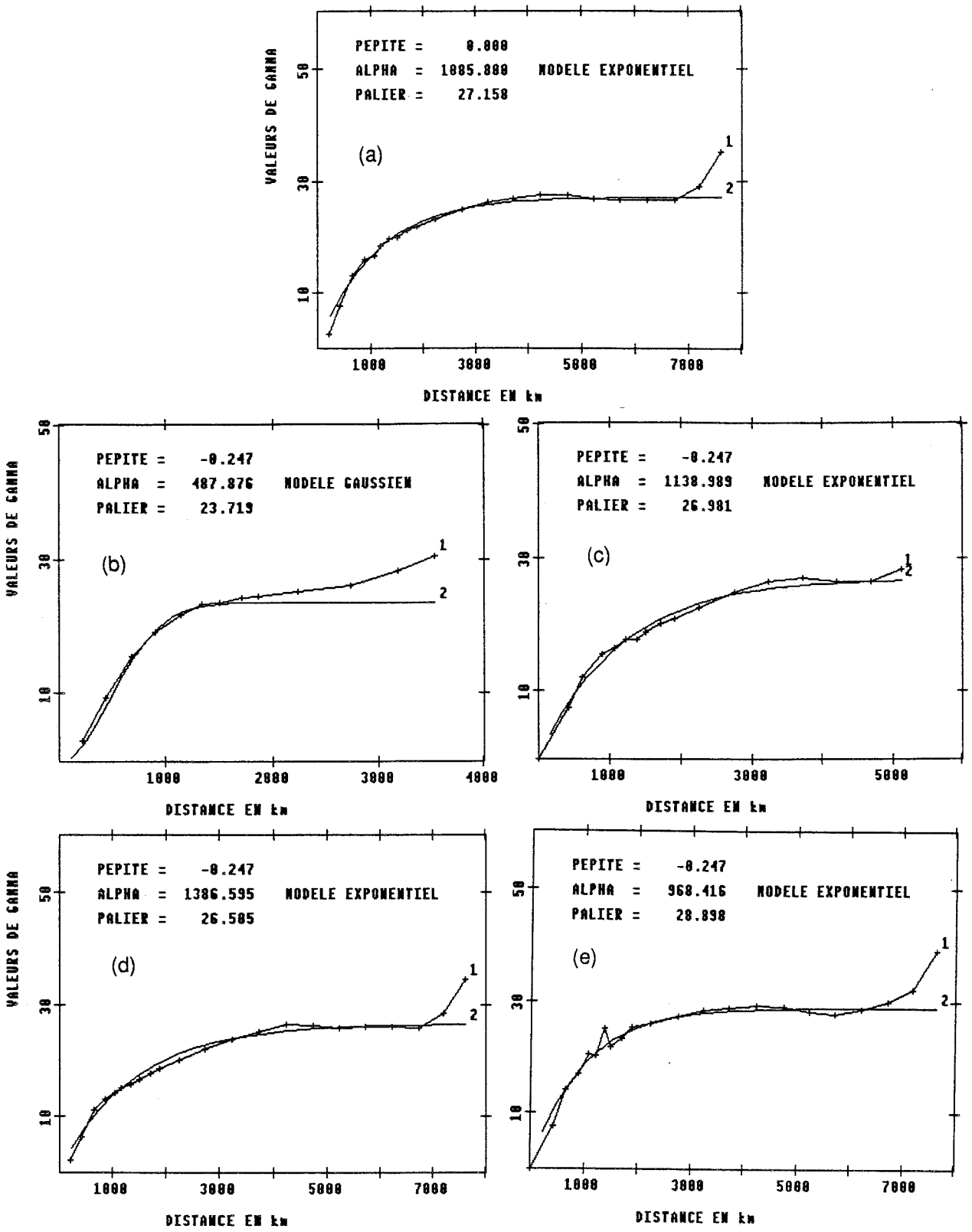


Fig. 4-13:Variogrammes climatologiques de la SST désaisonnalisée calculés sur l'hémisphère Nord.

(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

aussi les directions où les mouvements des courants peuvent prendre la plus grande extension.)

		Global exponentiel	Direction N-S gaussien	Direction NE-SW exponentiel	Direction W-E exponentiel	Direction NW-SE exponentiel
Ensemble du domaine	ae* ou ag*	1586. (5015.)	462. (1033.)	950. (3004.)	1263. (3994.)	1504. (4756.)
	σ^2*	0.34	0.24	0.28	0.28	0.33
	σ*	0.58	0.49	0.53	0.53	0.57
Hémisphère Nord	ae* ou ag*	1086. (3434.)	487. (1091.)	1139. (3602.)	1386. (4383.)	968. (3061.)
	σ^2*	0.27	0.24	0.27	0.26	0.29
	σ*	0.52	0.49	0.52	0.51	0.54
Hémisphère Sud	ae* ou ag*	982. (3105.)	447. (999.)	903. (2855.)	995. (3146.)	876. (2770.)
	σ^2*	0.32	0.26	0.34	0.30	0.28
	σ*	0.56	0.51	0.58	0.55	0.53

Tableau 4-2: Paramètres σ^{2*} et a_e* ou a_g* des variogrammes climatologiques modélisés de la SST désaisonnalisée Pour a_e* ou a_g*, la valeur entre parenthèses est la portée sphérique équivalente . Les valeurs de σ^{2*} ont été divisées par 100 par rapport aux graphiques pour avoir des valeurs en °C.

ii) Si on essaie maintenant de moduler ces premiers résultats en regardant ce qui se passe hémisphère par hémisphère, les inhomogénéités sont beaucoup moins flagrantes que sur la SST brute. Notons aussi une plus grande régularité des variogrammes qui ne présentent pas de dérive, notamment sur l'hémisphère Nord (ça n'est pas vrai au Sud où elle apparaît entre 5000 et 7000 Km dans le direction W-E, peut-être à cause des forts gradients observés sur l'écart-type au large des côtes angolaises (Fig. 4-4)). Pour expliquer ce comportement plus "régulier" des variogrammes, on peut peut-être aussi évoquer la normalité des données désaisonnalisées, qui respectent donc mieux les hypothèses de la géostatistique classique. Par ailleurs, les portées sont plus élevées au Nord qu'au Sud, ce qui voudrait dire que la variabilité interannuelle se développe sur des zones moins étendues au Sud, mais en revanche, elle y est en général plus forte (même si cela se mesure en 1/100ème de °C, et que la qualité des données sur l'hémisphère Sud est bien inférieure à celle de l'hémisphère Nord....)

Si nous cherchons maintenant à relier ces résultats à la physique de la donnée étudiée: la SST à laquelle on a retiré le cycle saisonnier, nous pouvons, à travers l'analyse variographique, décrire la variabilité interannuelle de la SST. Contrairement aux variations

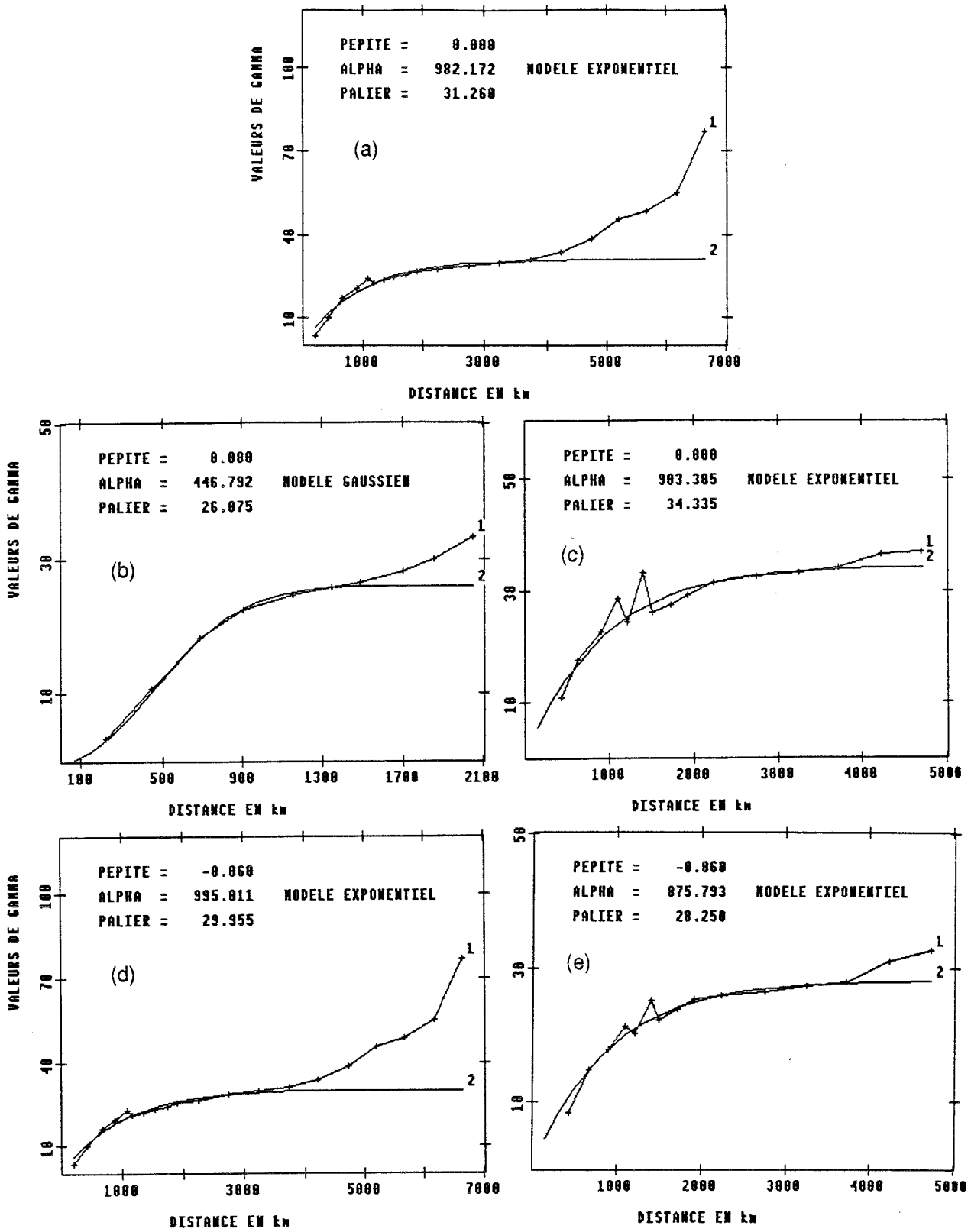


Fig. 4-14:Variogrammes climatologiques de la SST désaisonnalisée calculés sur l'hémisphère Sud.
(a) global (b) Direction N-S (c) Direction NE-SW
(d) Direction W-E (e) Direction NW-SE

saisonnières, il ne semble pas y avoir de région qui se démarque par une variabilité plus forte et les zones d'extension sont comparables. L'utilisation d'un même variogramme sur l'ensemble de la zone ne serait donc pas aussi pénalisante qu'elle ne le serait pour la SST brute où nous avons souligné les grandes différences entre les deux hémisphères. Nous pensons, entre autres, à une utilisation de ces modèles en simulation. Nous verrons aussi, au § 4-5 comment on peut s'en servir pour faire l'ACPP des données en essayant de s'affranchir des erreurs d'estimation des covariances induites par un échantillonnage statistique réduit.

Mais, pour l'heure, après cette analyse structurale classique, nous allons passer à l'application de l'ACPP aux données de SST.

4-4 APPLICATION DE L'ACPP AUX DONNEES DE SST ET COMPARAISON AVEC L'ACP CLASSIQUE.

4-4-1 Résultats déjà obtenus par ACP classique des données.

Le premier fichier ayant servi de base aux résultats présentés par la suite ne couvrait que la période 1964-1979, soit 16 ans de mesures. Nous verrons donc si avec 8 ans supplémentaires, les résultats obtenus à l'époque restent stables.

L'analyse par ACP a, par ailleurs, été réalisée sur trois types de données:

i) le cycle moyen $X_s(t_i, \underline{x}_j)$ défini par (4-3) (Piedelièvre, 1985)

ii) les données brutes $X(t_i, \xi_l, \underline{x}_j)$ (Piedelièvre, 1985; Servain et Legler, 1986)

iii) les données désaisonnalisées $X_d(t_i, \xi_l, \underline{x}_j)$ (Piedelièvre, 1985).

Nous allons rappeler les principaux résultats obtenus et, pour les deux derniers cas, discuter leur stabilité lorsqu'on dispose de 8 ans de mesures supplémentaires.

4-4-1-1 *ACP du cycle moyen.*

D'après Piedelièvre (1985), le poids de la première F.P. est écrasant (91,5% de variance expliquée). Sa cartographie se confond quasiment avec celle de l'écart-type des données brutes (cf Fig. 4-3-b), avec deux zones à fort gradient: les îles du Cap Vert et la région de remontée du Courant Sud Equatorial (CSE) au large des côtes angolaises. La composante principale (abrégée par C.P. dans la suite) associée dans le domaine temporel est une sinusoïde décrivant le cycle saisonnier.

La F.P. 2 explique seulement 4,5% de la variance. Sa structure est plus complexe, mais on peut admettre qu'elle décrive les zones d'upwellings équatorial et côtiers (golfe de Guinée, côte angolaise). La C.P. associée est quasiment nulle jusqu'en Mai et présente deux maxima de signes opposés en Juillet et Novembre.

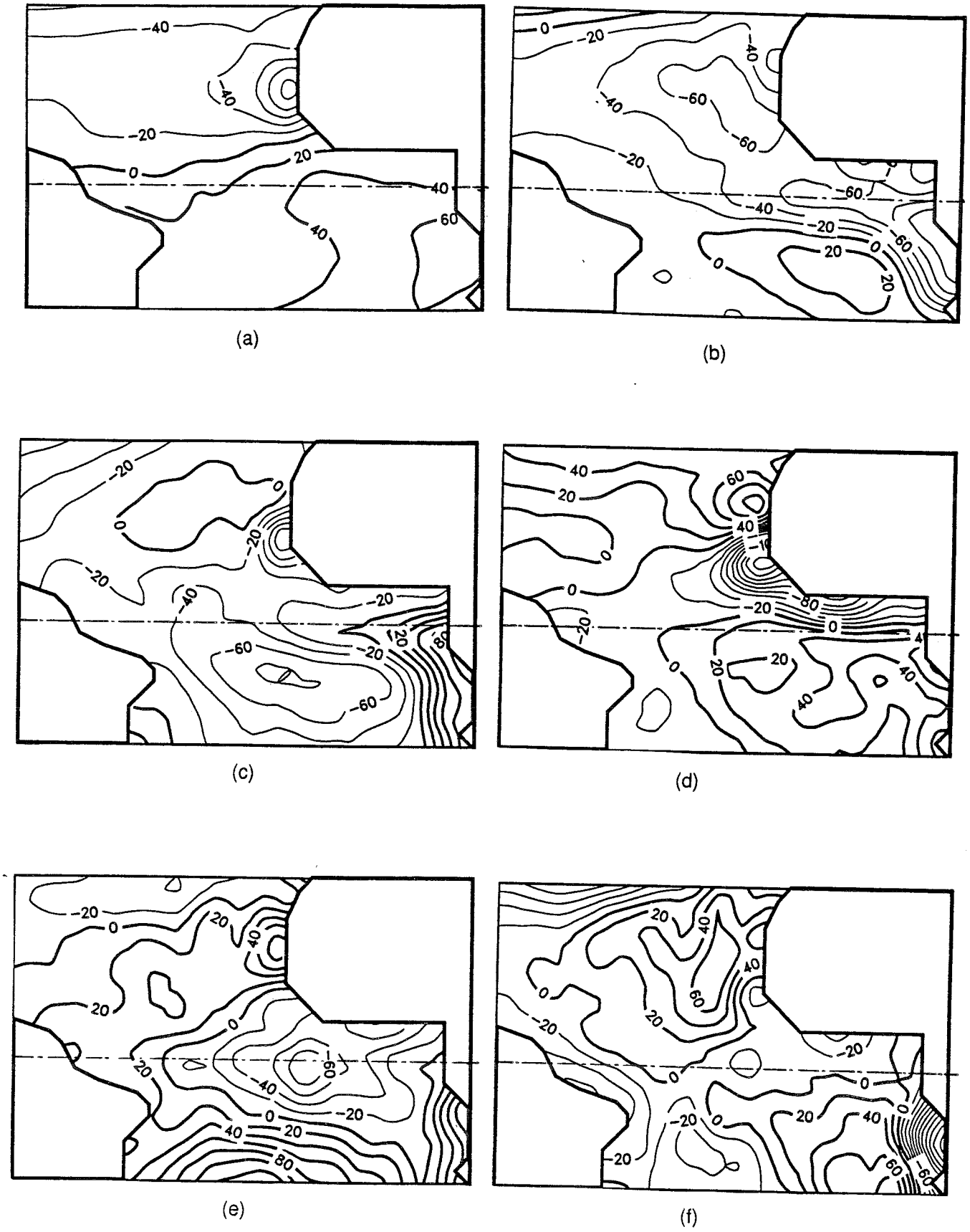


Fig. 4-15:Fonctions propres calculées par ACP classique de la SST brute entre 1964 et 1987.
(a) F.P. 1 (b) F.P. 2 (c) F.P. 3 (d) F.P. 4 (e) F.P. 5 (f) F.P. 6

La F.P. 3 (2% de variance expliquée) ne présente pas de changement de signe notable, sauf quelques zones de gradient près de Cap Vert, du golfe de Guinée et le la côte angolaise. La C.P. correspondante présente une variation biannuelle: deux minima et deux maxima pouvant correspondre à la périodicité du jet nord du Courant Sud Equatorial (CSE).

4-4-1-2 *ACP des données brutes.*

Elle a été décrite par Piedelièvre (1985) et Servain et Legler (1986). Le tableau 4-3 donne les pourcentages de variance expliqués par les F.P. 1à 6 pour les deux périodes 1964-1979 et 1964-1987.

	F.P.	1	2	3	4	5	6
1964-1979	% variance expliquée	81.0	5.1	3.8	1.6	0.9	0.7
	% cumulé	81.0	86.1	89.9	91.5	92.4	93.1
1964-1987	% variance expliquée	80.5	5.1	3.6	1.6	0.8	0.8
	% cumulé	80.5	85.6	89.2	90.8	91.6	92.4

Tableau 4-3: % de variance expliqués et % cumulés pour les six premières F.P. obtenues par ACP classique de la SST brute pour 1964-1979 et 1964-1987.

On voit donc qu'il y a une bonne stabilité des résultats qui se manifeste aussi sur la cartographie des F.P. (Fig. 4-15). La Fig. 4-16 donne l'évolution temporelle des C.P. correspondantes.

Au niveau de l'interprétation qui est faite de ces F.P., on peut reprendre ce qui a été dit pour le cycle moyen. En particulier, l'analyse temporelle des C.P. montre (à travers le calcul du périodogramme, approximation de la densité spectrale d"énergie) des pics correspondant à, respectivement une périodicité de 12 mois pour la C.P. 1, 6 et 12 mois pour la C.P. 2, 12 mois pour la C.P. 3, et 6 mois pour la C.P. 4 (Fig. 4-17).

4-4-1-3 *ACP classique de données désaisonnalisées.*

Nous reprenons ici les résultats de Piedelièvre (1985). Le tableau 4-4 suivant donne les pourcentages de variance expliqués pour les dix premières F.P. sur les périodes 19-64-1979 et 1964-1987.

Là encore, on observe une très grande stabilité des résultats qui se retrouve sur la cartographie des F.P. (Fig. 4-18). La Fig. 4-19 présente les C.P. correspondantes. Notons que, maintenant, ces C.P. ne présentent plus des périodicités comme avec les données brutes, l'autocorrélation temporelle décroissant plutôt de façon exponentielle (Fig. 4-20).

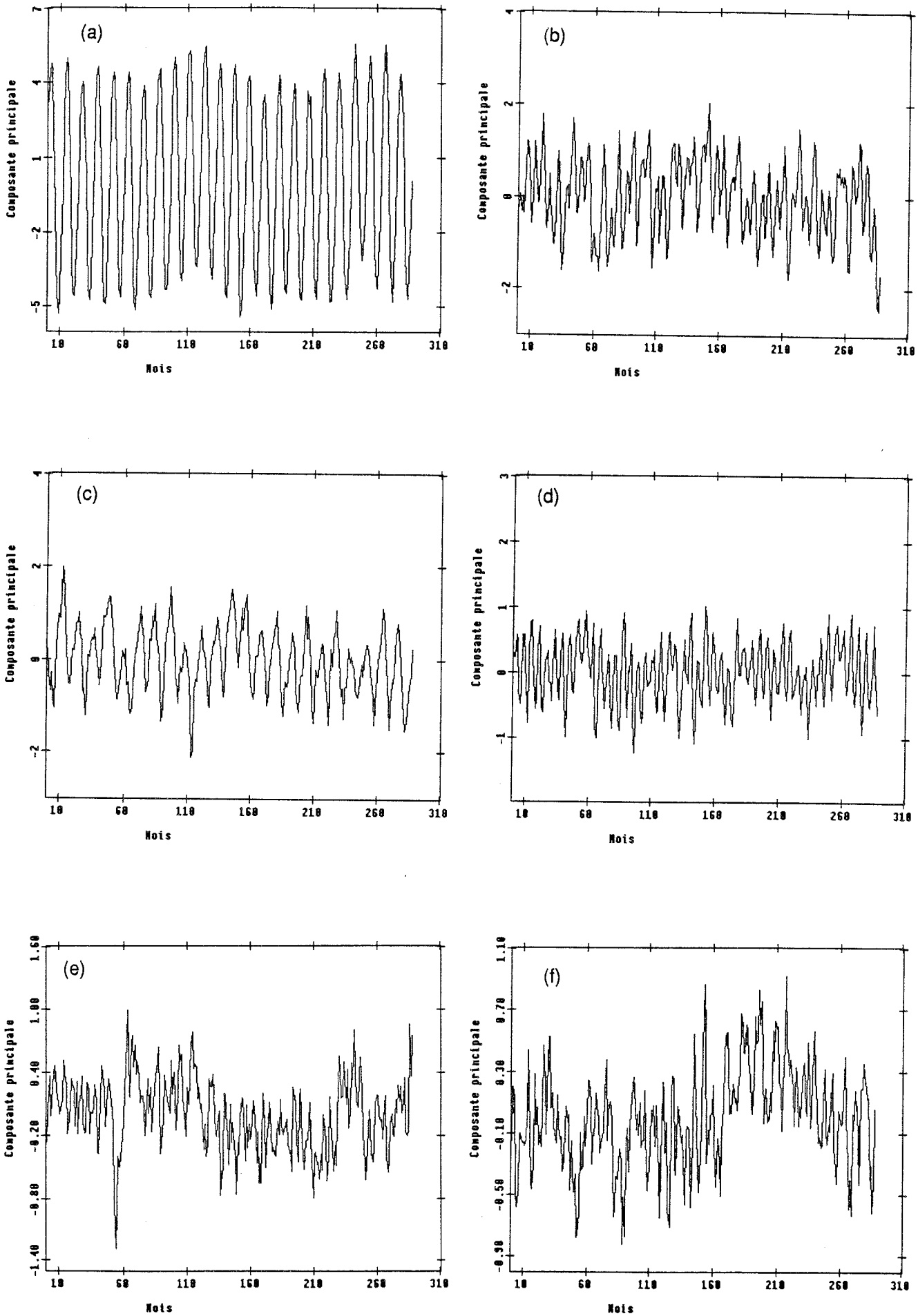


Fig. 4-16:Evolution temporelle des Composantes Principales obtenues par ACP classique de la SST brute entre 1964 et 1987.
(a) C.P. 1 (b) C.P. 2 (c) C.P. 3 (d) C.P. 4 (e) C.P. 5 (f) C.P. 6

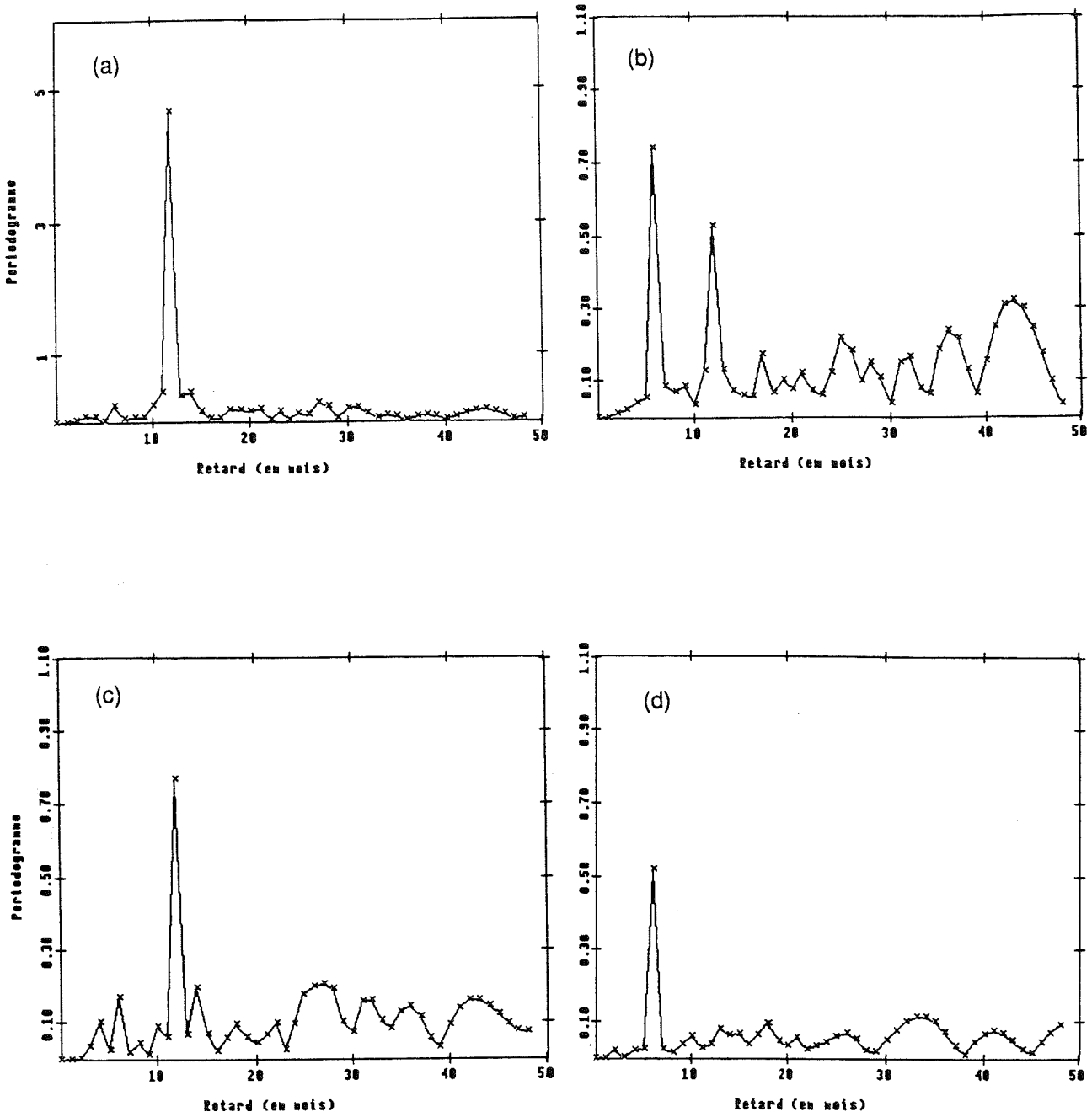


Fig. 4-17: Périodogrammes des Composantes Principales obtenues par ACP classique de la SST brute entre 1964 et 1987.
(a) C.P. 1 (b) C.P. 2 (c) C.P. 3 (d) C.P. 4

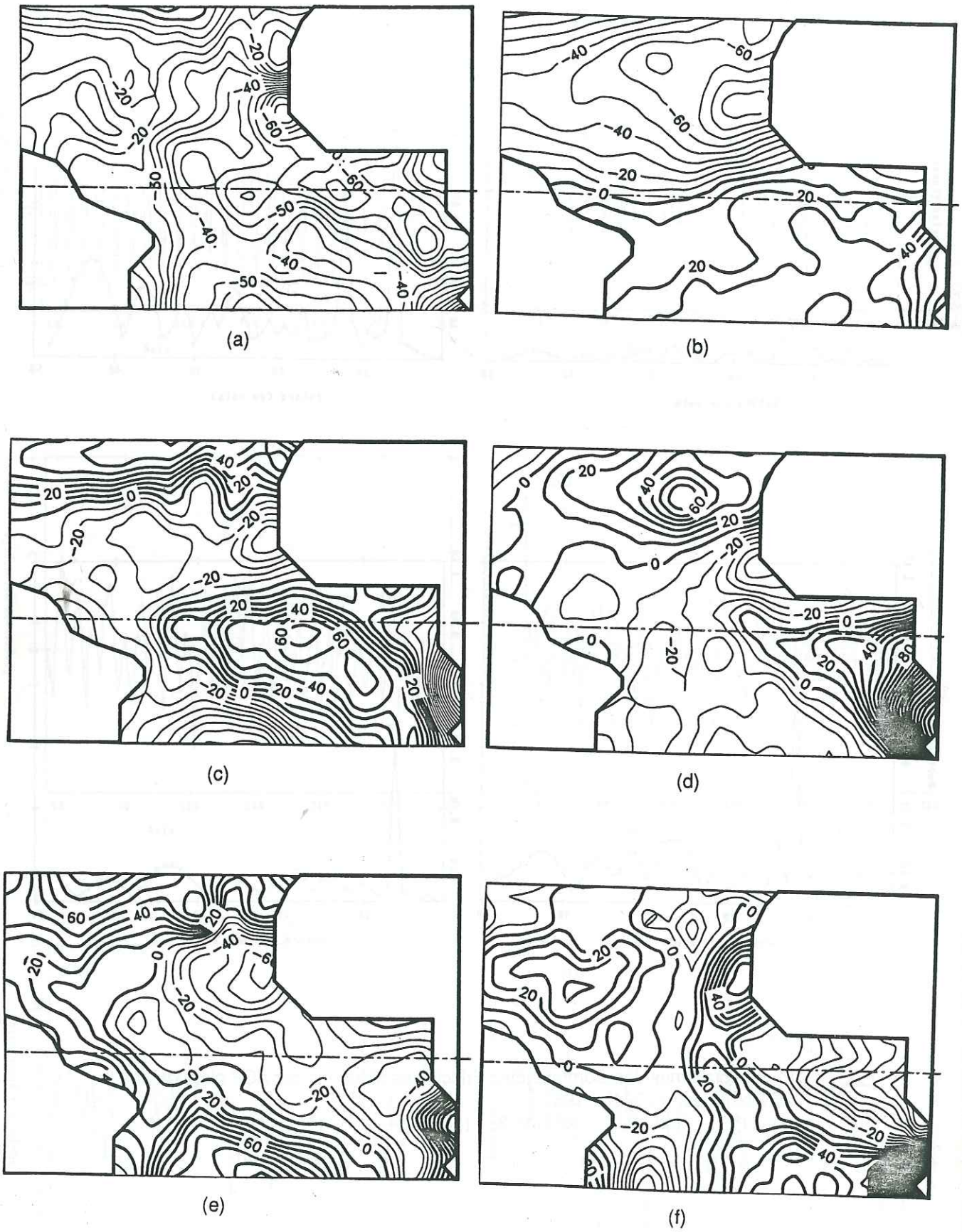


Fig. 4-18:Fonctions propres calculées par ACP classique de la SST désaisonnalisée entre 1964 et 1987.
(a) F.P. 1 (b) F.P. 2 (c) F.P. 3 (d) F.P. 4 (e) F.P. 5 (f) F.P. 6

Si on regarde la F.P. 1, elle garde un signe constant, avec des extrema au voisinage du Cap Vert et de l'upwelling équatorial. On peut de plus dire que le signe de la C.P. 1 traduit un réchauffement ou un refroidissement par rapport à l'année moyenne.

La F.P. 2 change de signe au niveau de l'équateur et traduit l'opposition entre les deux hémisphères. Elle est aussi très proche de la F.P. 1 des données brutes.

	F.P.	1	2	3	4	5	6	7	8	9	10
1964-1979	% variance expliquée	22.1	17.0	6.4	5.6	3.9	3.5	3.1	2.8	2.3	2.0
	% cumulé	22.1	39.1	45.5	51.1	55.0	58.5	61.6	64.4	66.7	68.7
1964-1987	% variance expliquée	20.7	17.0	6.2	5.2	3.9	3.3	3.1	2.7	2.4	2.1
	% cumulé	20.7	37.7	43.9	49.1	53.0	56.3	59.4	62.1	64.5	66.4

Tableau 4-4: % de variance expliqués et % cumulés pour les dix premières F.P. obtenues par ACP classique de la SSTdésaisonnalisée pour 1964-1979 et 1964-1987.

Ce premier paragraphe nous a donc permis de faire le point sur les études menées par ACP classique sur les données. Nous allons maintenant aborder l'application de l'ACPP proprement dite, en voyant tout d'abord comment on peut la mettre en oeuvre sur la sphère, et avec un fichier contenant plus de points de mesure que d'observations. Ensuite, nous regarderons, d'une part la stabilité des F.P. calculées par ACP ou par ACPP lorsque la grille régulière initiale de 2° par 2° est dégradée aléatoirement et, d'autre part, nous comparerons les deux approches avant de tester l'utilisation de l'ACPP pour la reconstitution des champs de 1986 et 1987.

4-4-2 Application pratique de l'ACPP aux données de SST.

4-4-2-1 *Les problèmes posés.*

L'approximation numérique décrite au § 1-3 a, bien sûr, été utilisée, mais son application pratique, dans le cas de nos données de SST, nécessite les deux adaptations suivantes:

i) Pour le calcul de la matice **E** des produits scalaires, il faut tenir compte du fait que le domaine d'étude n'est plus plan, mais une portion de surface sphérique, ce qui modifie le calcul des surfaces des triangles lorsqu'on utilise des fonctions linéaires par facette.

ii) Nous travaillons avec un fichier possédant plus de stations que de réalisations. Pour alléger les calculs et éviter de diagonaliser une matrice **C.E** de dimension (P,P), P étant le nombre de stations, dont P-N valeurs propres seraient nulles (N est le nombre de réalisations), nous avons adapté la méthode consistant à travailler avec la matrice

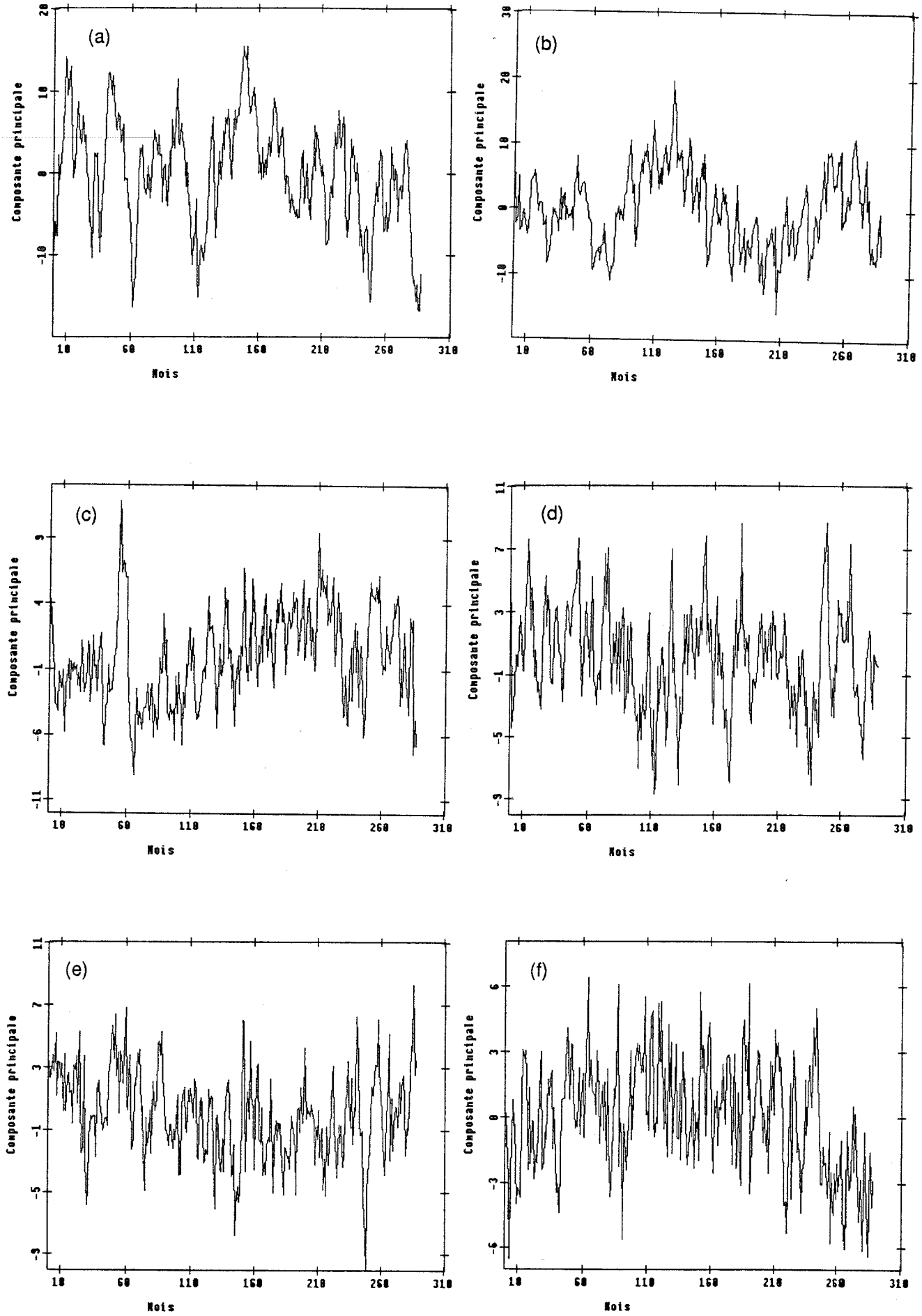


Fig. 4-19: Evolution temporelle des Composantes Principales obtenues par ACP classique de la SST désaisonnalisée entre 1964 et 1987.
(a) C.P. 1 (b) C.P. 2 (c) C.P. 3 (d) C.P. 4 (e) C.P. 5 (f) C.P. 6

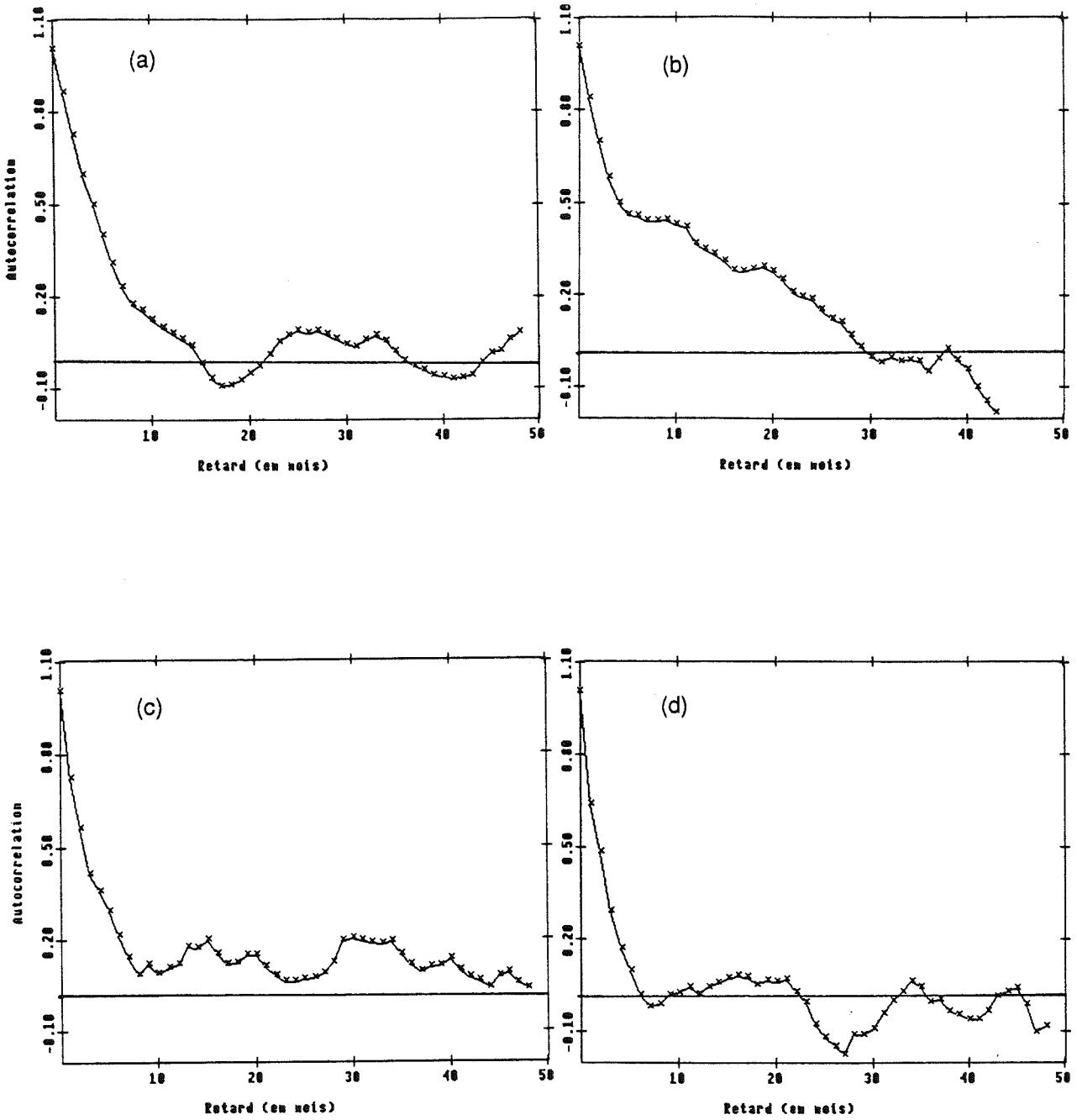


Fig. 4-20:Autocorrélation temporelle des Composantes Principales obtenues par ACP classique de la SST désaisonnalisée entre 1964 et 1987.
(a) C.P. 1 (b) C.P. 2 (c) C.P. 3 (d) C.P. 4

transposée $\mathbf{X}^T$, de la matrice (N,P) des observations **X** (Lebart et Fenelon, 1973; Piedelièvre, 1985; Lagerloeff et Bernstein, 1988) pour prendre en compte l'introduction de **E**.

Le calcul des éléments E_{ij} de la matrice **E** sur la sphère est décrit en Annexe C2. Nous allons maintenant expliciter le second point.

4-4-2-2 *Mise en oeuvre pratique de l'ACPP lorsqu'on dispose de plus de points de mesure que de réalisations.*

On considère toujours un processus aléatoire $X(\xi, \underline{x})$ où $\underline{x}$ est le vecteur des coordonnées spatiales et ξ l'indice de la réalisation aléatoire.

Soit **X**, de dimension (N,P), la matrice des N observations ξ_i, i=1,..N de $X(\xi, \underline{x})$ échantillonné en P points de mesure $\underline{x}_j$, j=1,..P.

α- *Cas où N > P*

Nous avons vu au § 1-3 que l'ACPP de $X(\xi, \underline{x})$ conduisait à résoudre le problème matriciel suivant:

$$\mathbf{C} \,.\, \mathbf{E} \,.\, \mathbf{F} = \mathbf{F} \,.\, \Lambda \qquad (4\text{-}12)$$

avec
- $\mathbf{C} = 1/N \,.\, \mathbf{X}^T \,.\, \mathbf{X}$ = matrice (P,P) de variance-covariance.
- **E** = matrice (P,P) des produits scalaires
- **F** = matrice (P,P) des vecteurs propres
- Λ = matrice diagonale (P,P) des valeurs propres.

On calcule aussi la matrice **Z** (N,P) des composantes principales:

$$\mathbf{Z} = \mathbf{X} \,.\, \mathbf{E} \,.\, \mathbf{F} \qquad (4\text{-}13)$$

Lorsque N>P, le rang de **C** est au plus P et, en général, les P valeurs propres sont non nulles. En revanche, si N<P, le rang de **C** est au plus N et P-N valeurs propres sont nulles. Dans le cas de l'ACP classique, (où rappelons-le, on résout $\mathbf{C} \,.\, \mathbf{F} = \mathbf{F} \,.\, \Lambda$), on travaille alors avec la matrice associée $\mathbf{C}^* = 1/N \,.\, \mathbf{X}.\, \mathbf{X}^T$, de dimension (N,N), ce qui permet de réduire les temps de calcul. Les vecteurs propres de $\mathbf{C}^*$ sont alors, à une constante multiplicative près, les composantes principales de **C**, et réciproquement.

Voyons maintenant comment nous pouvons adapter cette technique en introduisant la matrice **E**.

β- *Cas où N < P*

Considérons la matrice $\mathbf{C^* = 1/N \,.\, X.\, E \,.\, X^T}$, de dimension (N,N). Ses éléments C_{ij}^* peuvent être définis comme une "covariance spatiale" inter-événement. En effet, si nous calculons $A=\int_D X(\xi_i, \underline{x}) \,.\, X(\xi_j, \underline{x}) \,.\, d\underline{x}$ pour approcher cette covariance, mais en approximant $X(\xi, \underline{x})$ comme au § 1-3 par

$$X^*(\xi, \underline{x}) = \sum_{m=1}^{P} X(\xi, \underline{x}_m) \,.\, e_m(\underline{x}) \qquad (4\text{-}14)$$

nous obtenons:

$$A = \int_D \sum_{l=1}^{P} X(\xi_i, \underline{x}_l) \,.\, e_l(\underline{x}) \,.\, \sum_{m=1}^{P} X(\xi_j, \underline{x}_m) \,.\, e_m(\underline{x}) \; d\underline{x}$$

$$= \sum_{l=1}^{P} \sum_{m=1}^{P} X(\xi_i, \underline{x}_l) \,.\, \int_D e_l(\underline{x}) \,.\, e_m(\underline{x}) \,.\, d\underline{x} \; X(\xi_j, \underline{x}_m)$$

$$= \sum_{l=1}^{P} \sum_{m=1}^{P} X(\xi_i, \underline{x}_l) \,.\, E_{lm} \,.\, X(\xi_j, \underline{x}_m) = C_{ij}^*$$

i) **C*** possède les mêmes valeurs propres que **C** et la matrice **F*** (N,N) des vecteurs propres normés est:

$$\mathbf{F^* = 1 / N^{1/2} \,.\, Z_N \,.\, \Lambda_N^{-1/2}} \qquad (4\text{-}15)$$

où
- $\mathbf{Z_N = X \,.\, E \,.\, F_N}$ = matrice (N,N) des N premières composantes de **C.**
- $\mathbf{F_N}$ = matrice (P,N) des N premiers vecteurs propres de **C.**
- Λ_N = sous-matrice diagonale (N,N) des N valeurs propres non nulles de **C.**

Démonstration: $C^* . F^* = 1/ N^{3/2} . X . E . X^T . Z_N . \Lambda_N^{-1/2}$

$= 1/ N^{3/2} . X . E . X^T . X . E . F_N . \Lambda_N^{-1/2}$

$= 1/ N^{1/2} . X . E . C . E . F_N . \Lambda_N^{-1/2}$

$= 1/ N^{1/2} . X . E . F_N . \Lambda_N . \Lambda_N^{-1/2}$

d'après (4-12) restreinte à F_N

$= 1 / N^{1/2} . Z_N . \Lambda_N^{-1/2} . \Lambda_N = F^* . \Lambda_N$

ii) On peut calculer une matrice Z^* (P,N) des "composantes" de C^*.

$$Z^* = X^T . F^* = N^{1/2} . F_N . \Lambda_N^{1/2} \qquad (4\text{-}16)$$

Démonstration: $Z^* = X^T . F^* = 1/ N^{1/2} . X^T . Z_N . \Lambda_N^{-1/2}$

$= 1/ N^{1/2} . X^T . X . E . F_N . \Lambda_N^{-1/2}$

$= N^{1/2} . C . E . F_N . \Lambda_N^{-1/2}$

$= N^{1/2} . F_N . \Lambda_N . \Lambda_N^{-1/2} = N^{1/2} . F_N . \Lambda_N^{+1/2}$

iii) F^* est bien orthogonale:

Démonstration: $F^{*T} . F^* = 1/N . \Lambda_N^{-1/2} . Z_N^T . Z_N . \Lambda_N^{-1/2}$

$= \Lambda_N^{-1/2} . \Lambda_N . \Lambda_N^{-1/2} = I_N$ où I_N est la matrice unité (N,N)

Nous avons de plus utilisé le fait que les Z_k sont des variables décorrélées.

iv) On peut déduire F_N et Z_N de F^* et Z^* (ce qui est le but final de tout ce calcul)

$$F_N = 1/ N^{1/2} . Z^* . \Lambda_N^{-1/2} \qquad (4\text{-}17)$$

$$Z_N = N^{1/2} . F^* . \Lambda_N^{1/2} \qquad (4\text{-}18)$$

v) On peut vérifier que les vecteurs propres F_j sont bien orthonormés pour le produit scalaire défini par les $e_i(\underline{x})$, c'est-à-dire que:

$$F_N^T . E . F_N = I_N$$

Démonstration:

$$F_N^T . E . F_N = 1/N \Lambda_N^{-1/2} . Z^{*T} . Z^* . \Lambda_N^{-1/2}$$
$$= 1/N \Lambda_N^{-1/2} . F^{*T} . X . E . X^T . F^* . \Lambda_N^{-1/2}$$
$$= \Lambda_N^{-1/2} . F^{*T} . C^* . F^* . \Lambda_N^{-1/2}$$
$$= \Lambda_N^{-1/2} . F^{*T} . F^* . \Lambda_N . \Lambda_N^{-1/2}$$
$$= \Lambda_N^{-1/2} . I_N . \Lambda_N . \Lambda_N^{-1/2} = I_N$$

4-4-3 Stabilité des F.P. calculées par ACP ou par ACPP lorsqu'on dégrade aléatoirement le réseau.

Lors de la présentation de la technique d'ACPP, nous avons souligné que son intérêt, par rapport à une ACP classique, est de prendre en compte explicitement la forme et la taille du domaine, mais aussi d'approcher objectivement l'intégration analytique en affectant des poids différents aux diverses stations, lorsque le réseau de mesure est irrégulier. Nous avons voulu voir si cela se vérifiait pour les données de SST désaisonnalisées.

Les données initiales étant déjà sous forme de grille régulière, il nous était difficile d'évaluer cet aspect. Nous avons donc construit arbitrairement un réseau irrégulier, pour voir si nos techniques étaient bien capables de compenser cette irrégularité. Pour cela, partant de la grille régulière, présentée à la Fig. 4-1, nous avons retiré aléatoirement 254 stations sur les 628 initiales afin de constituer un réseau, à la fois moins dense et plus irrégulier. Sa discrétisation est donnée Fig. 4-21.

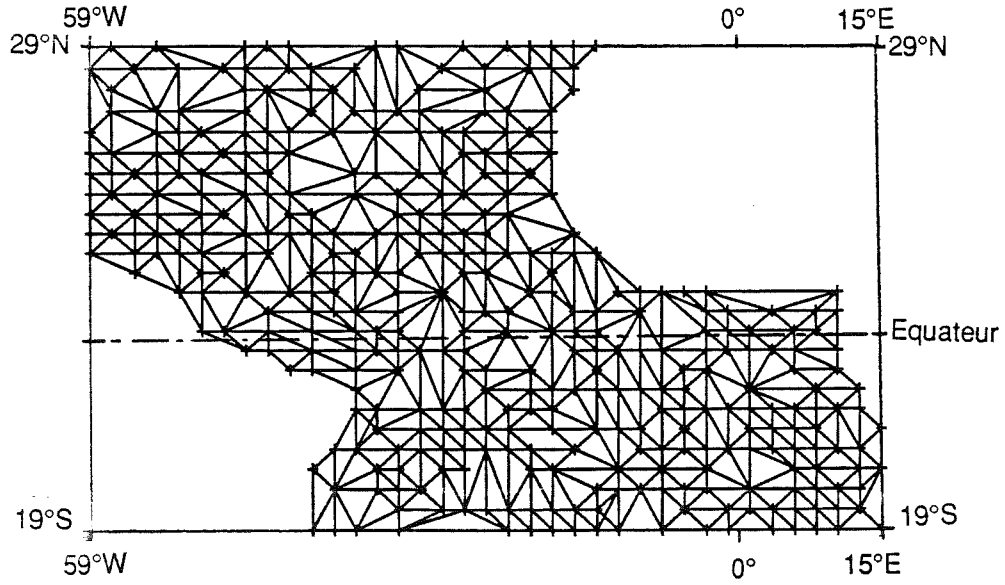


Fig. 4-21:Discrétisation du réseau de P=374 stations obtenu après dégradation aléatoire de la grille 2° par 2° initiale de P=628 points.

Nous avons ensuite réalisé une ACP classique et une ACPP sur la grille régulière de 2° par 2° et ce réseau aléatoire. Le tableau 4-5 donne les % de variance expliqués par les dix premières F.P. dans chacun des cas étudiés.

On remarque que, pour l'ACP classique, les résultats sont stables et on atteint 66.4% de variance expliquée avec la grille régulière et 66.8% avec le réseau aléatoire. En revanche, sur le réseau aléatoire, l'ACPP fait gagner 3% de variance expliquée sur les dix premières F.P., notamment sur la première et la troisième. Néanmoins, les résultats restent comparables.

Afin de quantifier la ressemblance entre les F.P., nous avons calculé les coefficients de détermination ρ^2 entre les valeurs des F.P. aux noeuds de la grille régulière de P=628 points, les valeurs du réseau aléatoire y ayant été obtenues par interpolation spline, même si cela peut contribuer à modifier légèrement les résultats. Le tableau 4-6 donne ces coefficients. Comme au chapitre 2, nous avons comparé une F.P. calculée sur la grille 2° par 2° (première colonne) avec plusieurs F.P. obtenues sur le réseau aléatoire (deuxième colonne), afin de déceler les dégénérescences. Les colonnes 3 et 4 contiennent respectivement les ρ^2 entre grille régulière et aléatoire pour un calcul par ACP classique, puis par ACPP.

		F.P.	1	2	3	4	5	6	7	8	9	10
ACP classique	grille 2° *2°	% variance expliquée	20.7	17.0	6.2	5.2	3.9	3.3	3.1	2.7	2.4	2.1
		% cumulé	20.7	37.7	43.9	49.1	53.0	56.3	59.4	62.1	64.5	66.4
	Rés. aléa.	% variance expliquée	20.7	17.2	6.0	5.3	3.8	3.5	3.3	2.5	2.4	2.1
		% cumulé	20.7	37.9	43.9	49.2	53.0	56.5	59.8	62.3	64.7	66.8
ACPP	grille 2° *2°	% variance expliquée	23.4	18.5	6.2	5.1	3.9	3.2	2.9	2.6	2.3	1.8
		% cumulé	23.4	41.9	48.1	53.2	57.1	60.3	63.2	65.8	68.1	69.9
	Rés. aléa.	% variance expliquée	25.6	18.1	7.6	5.3	3.6	3.4	3.0	2.5	2.0	2.0
		% cumulé	25.6	43.7	51.3	56.6	60.2	63.6	66.6	69.1	71.1	73.1

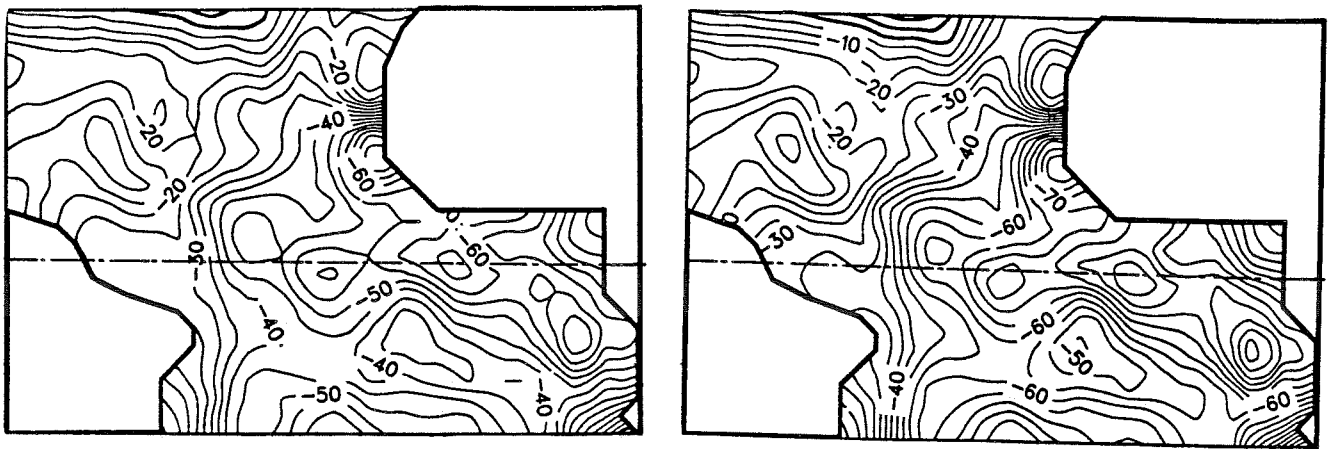
Tableau 4-5: % de variance expliqués et % cumulés des dix premières F.P. des données désaisonnalisées calculées par ACP classique ou ACPP sur la grille 2° par 2° ou le réseau aléatoire.

Ces coefficients montrent, malheureusement, une plus grande stabilité des résultats par ACP classique, où de réels mélanges entre plusieurs F.P. ne se produisent qu'à partir de la F.P. 8. En revanche, par ACPP, sauf pour la F.P. 2, les résultats sont très décevants puisque, même la F.P. 1 ne coïncide pas entre les deux essais. Pour les autres F.P., des mélanges se produisent, mais néanmoins, on arrive à retrouver, en combinant l'ensemble des

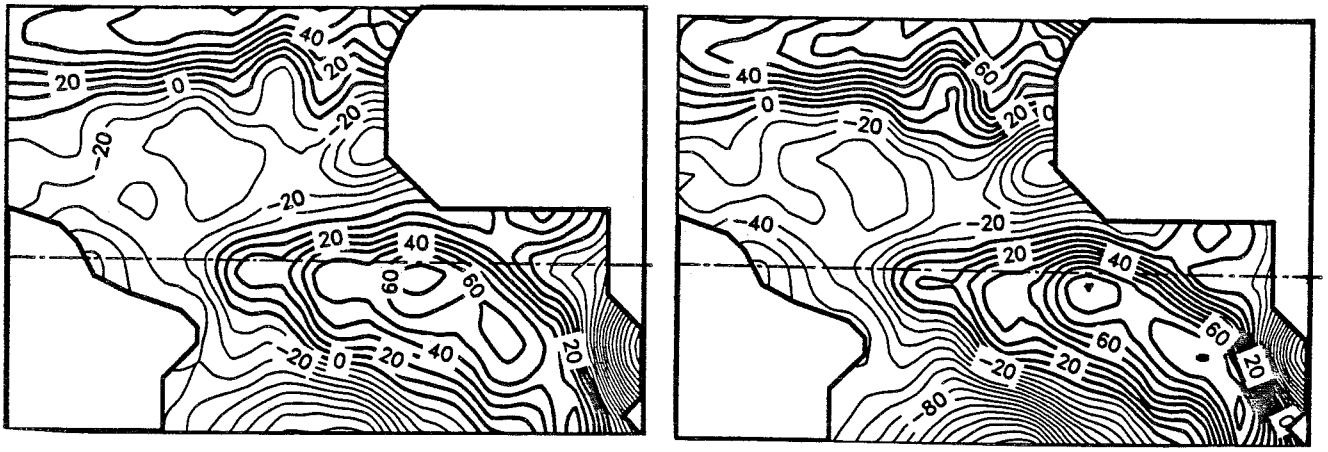
n° FP P=628	n° FP comparée P=374	ACP	ACPP
1	1	0.97	0.24
	2	0.18	0.35
	3		0.14
	7		0.14
2	1	0.38	0.32
	2	0.996	0.96
3	3	0.87	0.69
	4	0.13	0.42
	5		0.20
4	3		0.15
	4	0.86	0.42
	5		0.21
	6		0.40
	7		0.12
5	1	0.17	0.30
	5	0.84	0.58
	6	0.10	
	7		0.30
6	5	0.09	
	6	0.85	0.46
	8		0.36
7	7	0.92	0.31
	8		0.44
	9		0.12
8	8	0.40	0.11
	9	0.55	0.69
9	8	0.56	
	9	0.33	
	10		0.41
10	10	0.87	0.21

Tableau 4-6: Coefficients de détermination ρ^2 entre F.P. calculées sur la grille régulière 2° par 2° (P=628) et le réseau aléatoire (P=374) pour une estimation par ACP classique ou par ACPP.

grille 2*2 | réseau aléatoire



(a) F.P. 1



(b) F.P. 4

(c) F.P. 5

Fig. 4-22:F.P. calculées par ACP classique, à gauche sur la grille régulière de 2° par 2°, à droite sur le réseau aléatoire.
(a) F.P. 1 (b) F.P. 4 (c) F.P. 5

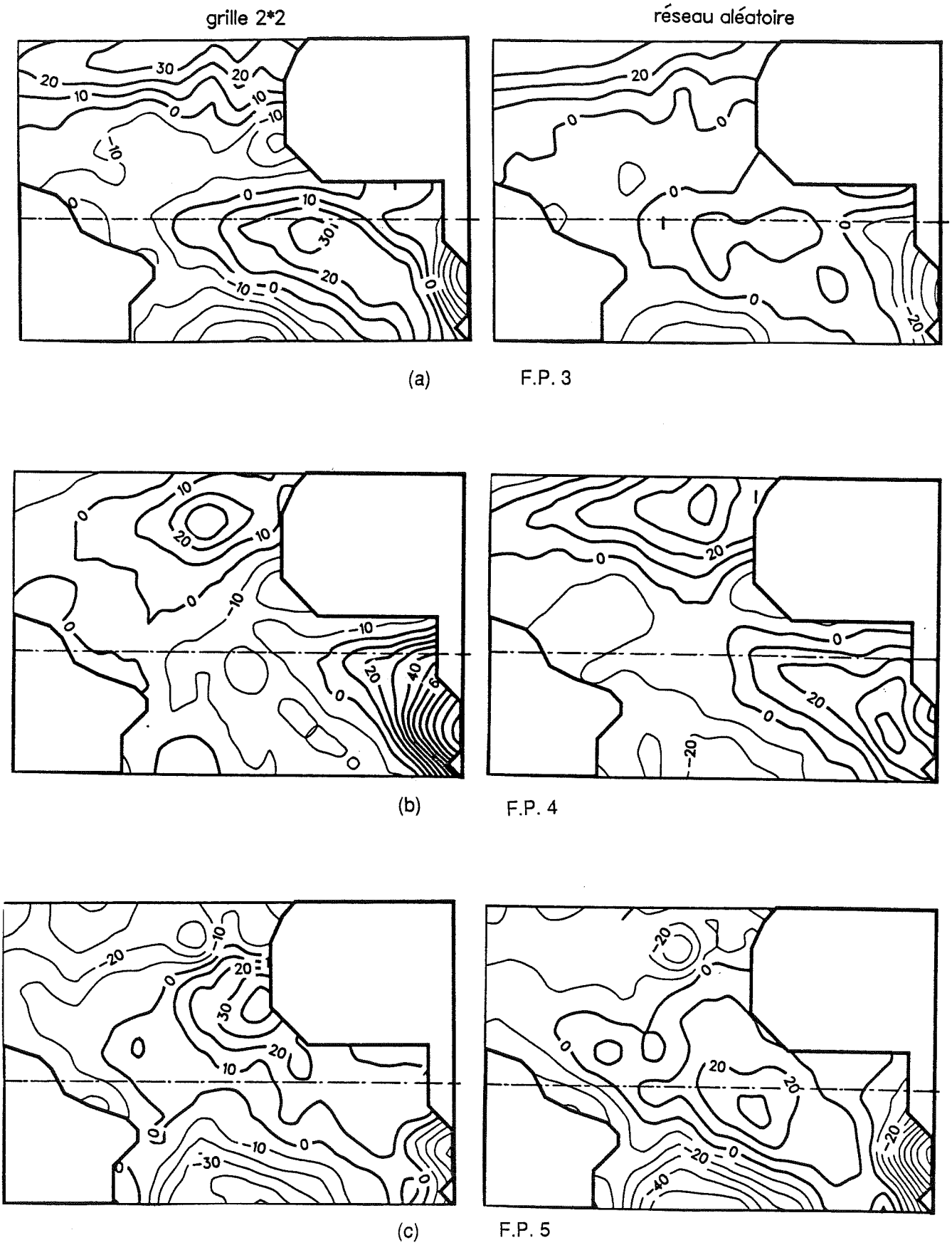


Fig. 4-23:F.P. calculées par ACPP, à gauche sur la grille régulière de 2° par 2°, à droite sur le réseau aléatoire.
(a) F.P. 3 (b) F.P. 4 (c) F.P. 5

F.P., les différentes zones à fort gradient. La Fig. 4-22 illustre les bons résultats par ACP pour les F.P. 1, 3 et 5. La Fig. 4-23 montre quant à elle les mélanges entre les F.P. 3,4 et 5 qui se produisent par ACPP.

Néanmoins, et contrairement aux essais du chapitre 2, nous ne connaissons pas les vraies F.P. et nous ne pouvons pas dire si les instabilités observées par ACPP condamnent les résultats qu'elle donne pour les F.P. En effet, nous avons vu au chapitre 2 que la taille finie de l'échantillon utilisé pour estimer la matrice de variance-covariance induisait les erreurs les plus importantes sur l'estimation des F.P. Or, avec N=288 observations, de plus autocorrélées, pour P=374 ou 628 stations, nous sommes sûrement dans un cas très défavorable. Et il est fort probable que nous ne soyons pas en mesure de bien séparer les différentes valeurs propres. Dans ces conditions, les F.P. estimées sont des combinaisons linéaires des vraies F.P. et, si nous retrouvons ce phénomène, plus visible en ACPP, c'est parce que la matrice **E** des produits scalaires change lorsqu'on passe de la grille régulière au réseau aléatoire. La matrice **C.E** diagonalisée sur le réseau aléatoire est donc très différente d'une sous-matrice de son homologue sur la grille régulière. En revanche, avec l'ACP classique, la matrice **C** pour le réseau aléatoire, est extraite de **C** pour la grille régulière. Ceci pourrait expliquer que les F.P. estimées par ACP classique soient peu modifiées lorsqu'on dégrade le réseau, contrairement à ce qui se passe en ACPP.

Cependant, pour peu qu'on interprète l'ensemble des F.P., on réussira à déterminer toutes les zones à fort gradient, qui sont assez bien individualisées et correspondent toujours aux zones d'upwellings (îles du Cap Vert, golfe de Guinée, côtes de l'Angola, équateur). Mais, on peut avoir plus ou moins de mal à faire émerger de ces cartes les connexions entre réchauffement ou refroidissement de telle ou telle zone. On retrouve donc ici le problème de l'interprétation des résultats d'une analyse EOF, avec ici une forte suspicion pour les effets d'échantillonnage (nombre d'observations)....

4-4-4 Comparaison entre F.P. calculées par ACP classique et par ACPP.

Cette comparaison a été réalisée à la fois sur les données brutes (grille 2° par 2°) et les données désaisonnalisées (grille 2° par 2° et réseau aléatoire). Le tableau 4-5 contient, rappelons-le les pourcentages de variance expliqués et cumulés pour les données désaisonnalisées. Le tableau 4-7 les présente pour les données brutes.

Les résultats montrent que l'ACPP permet d'extraire plus de variance des données, notamment sur les premières F.P. En effet, si on ne gagne que 0.8% sur le pourcentage cumulé des dix premières F.P. issues de l'analyse des données brutes, on atteint 3.5% (resp. 6.3%) pour celle des données désaisonnalisées sur la grille 2° par 2° (resp. le réseau

aléatoire). Ce résultat est intéressant pour le stockage des données et la condensation de l'information.

	F.P.	1	2	3	4	5	6	7	8	9	10
ACP classique	% variance expliquée	80.5	5.1	3.6	1.6	0.8	0.8	0.6	0.5	0.5	0.3
	% cumulé	80.5	85.6	89.2	90.8	91.6	92.4	93.0	93.5	94.0	94.3
ACPP	% variance expliquée	81.2	5.4	3.7	1.5	0.8	0.7	0.5	0.5	0.5	0.3
	% cumulé	81.2	86.6	90.3	91.8	92.6	93.3	93.8	94.3	94.8	95.1

Tableau 4-7: % de variance expliqués et cumulés pour les dix premières F.P. calculées , respectivement par ACP classique ou ACPP de la SST brute.

Si on regarde maintenant les coefficients de détermination ρ^2 entre F.P. calculées par ACP classique ou ACPP (tableau 4-8 , voir § précédent pour la façon dont il doit être lu), on observe que, pour une même densité de points, la cohérence des F.P. estimées par ACP classique et ACPP est plus grande sur les données brutes jusqu'à la F.P. 5. Puis, c'est le contraire pour les F.P. suivantes. Cependant, pour les F.P. 6 et 7, on note une inversion entre ACP et ACPP, pour les données désaisonnalisées. En revanche, sur le réseau aléatoire, la cohérence est nettement moins bonne et les F.P. estimées par chacune des techniques se mélangent comme l'illustre la Fig. 4-24 pour les F.P 3 et 4.

La bonne cohérence entre ACP et ACPP sur la grille régulière peut se comprendre assez facilement. En effet, à cause de cette régularité (et malgré l'effet de la latitude), on peut estimer que les surfaces représentatives de chaque station sont comparables. L'introduction de la matrice **E** ne jouera donc pas une rôle capital, d'où la similitude des résultats. En revanche, lorsqu'on travaille sur le réseau aléatoire, la matrice **E** prend tout son poids, ce qui explique la divergence des résultats.

Mais alors, que conclure? Où est la vérité? On peut probablement reprendre ici les remarques faites au § précédent. Ne connaissant pas les vraies F.P., nous ne pouvons pas déterminer la meilleure méthode. En revanche, les instabilités constatées devraient inciter à beaucoup de prudence lors d'une éventuelle interprétation des F.P. car elles nous prouvent que les conditions d'échantillonnage sont telles que les erreurs introduites sur l'estimation des F.P. sont très importantes et justifient une exploitation très circonspecte des résultats.

n° FP ACPP étudiée	n° FP ACP comparée	Données Brutes P = 628	Données désaisonnalisées ACP et ACPP sur grille P = 628	grille P = 374
1	1	0.9997	0.991	
	2			0.38
	5			0.33
2	1		0.23	0.58
	2	0.992	0.997	0.92
	3	0.38		
3	1			0.19
	2	0.51		
	3	0.98	0.93	0.55
	4			0.20
4	3			0.52
	4	0.99	0.90	0.30
5	3			0.23
	4			0.22
	5	0.97	0.97	0.37
	6			0.11
6	4			0.40
	6	0.85		
	7	0.14	0.87	0.36
	8		0.11	
7	1			0.11
	5			0.44
	6	0.12	0.78	
	7	0.65		
	8	0.17	0.13	0.11
	9			0.22
8	6		0.20	0.56
	7	0.15	0.09	0.19
	8	0.57	0.58	
	9	0.22		
9	6			0.28
	7			0.27
	8	0.17		
	9	0.72	0.89	0.27
10	8			0.22
	10	0.77	0.79	0.44

Tableau 4-8: Coefficients de détermination ρ^2 entre F.P. calculées par ACP ou par ACPP.

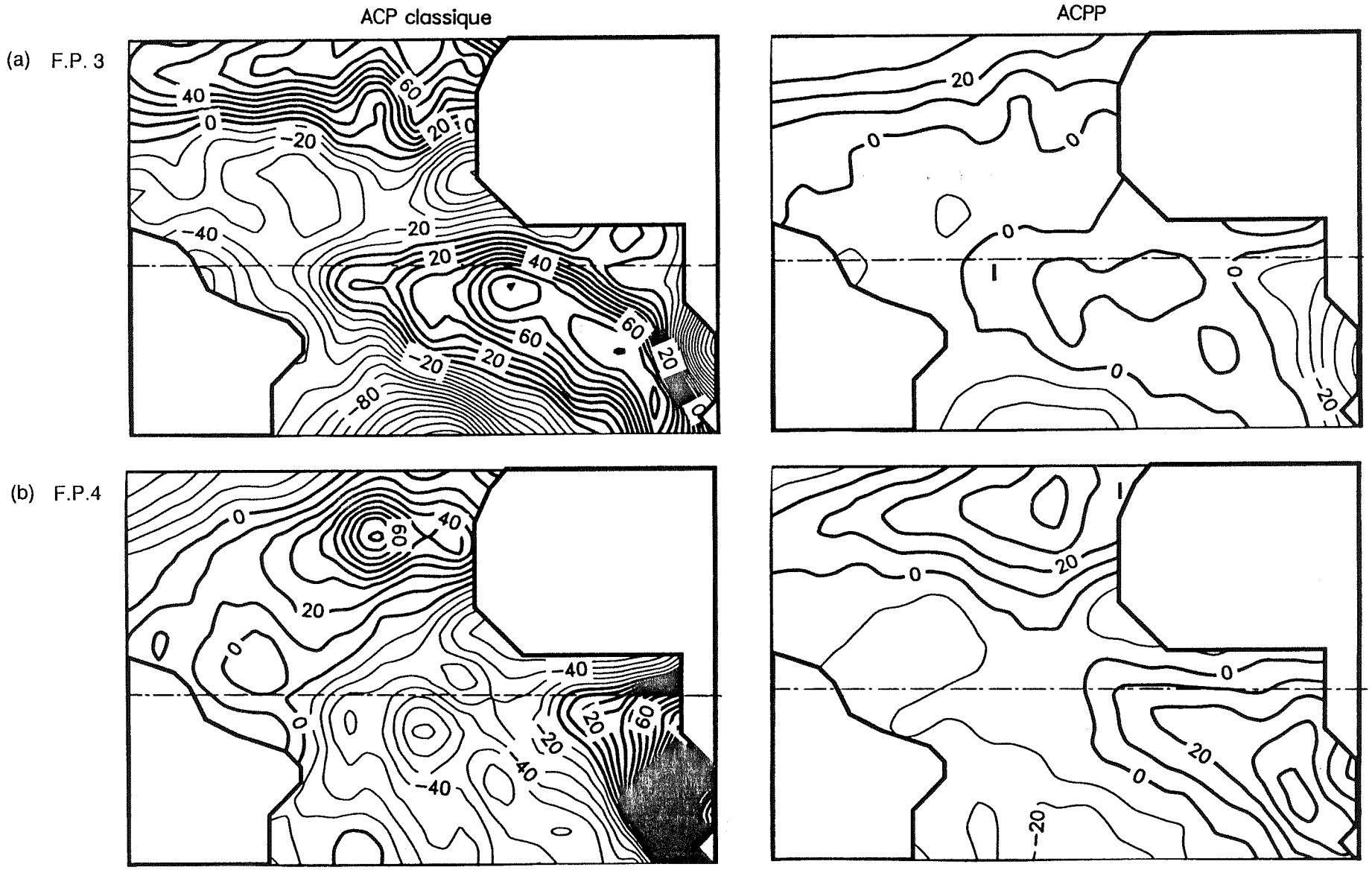


Fig. 4-24:F.P. calculées sur le réseau aléatoire, à gauche par ACP classique, à droite par ACPP. (a) F.P. 3 (b) F.P. 4.

4-5 COMPARAISON DE DIFFERENTES TECHNIQUES DE RECONSTITUTION BASEES SUR L'ACP OU L'ACPP POUR LA RECONSTITUTION DES CHAMPS DE 1986 ET 1987.

4-5-1 Introduction.

Aux paragraphes 4-4-3 et 4-4-4 précédents, nous avons vu qu'il était difficile de comparer efficacement les F.P. estimées par ACP classique et par ACPP car nous ne disposions pas des vraies F.P. Néanmoins, nous allons voir maintenant que, malgré les problèmes d'instabilité des F.P. mis en évidence plus haut, la méthode d'ACPP peut être utilisée avec succès pour la reconstitution de champs. Il suffit pour cela de faire abstraction de toute interprétation physique des F.P. et de les considérer dans leur globalité. Alors, la méthode devient très intéressante, notamment pour le stockage de données: on stocke en effet $K<P$ F.P. et, pour chaque champ ξ, $K<P$ $Z_k(\xi)$ qui seront utilisées pour reconstituer les champs (voir § suivant).

Pour clore ce chapitre, nous avons donc, dans cette optique, testé et validé diverses techniques de reconstitution basées sur l'ACP ou l'ACPP (voir chapitres 1 et 3), sur les 24 champs de SST désaisonnalisée, correspondant aux années 1986 et 1987, qui n'auront pas servi à l'ACP ou l'ACPP, effectuée sur les années 1964 à 1985 ou à partir d'un modèle moyen de covariance ajusté au § 4-3-3.

Dans ce cas, la comparaison entre les méthodes peut vraiment être "objective" puisqu'on dispose, d'une part des "vraies" valeurs observées, constituant le fichier de référence, et, d'autre part, des valeurs reconstituées par l'une des techniques de reconstitution testées.

Dans ce paragraphe, toujours grâce aux 24 champs de 1986 et 1987, nous essaierons aussi de répondre à une question qui a été soulevée au chapitre 2: vaut-il mieux faire l'ACPP (ou l'ACP) de la matrice de covariance empirique ou utiliser, pour la remplir, un modèle moyen, ajusté sur le nuage de corrélation, comme ceux que nous avons obtenus au § 4-3-3?

Nous présenterons donc tout d'abord les différentes techniques de reconstitution utilisées, puis les critères de comparaison retenus, avant de passer à l'analyse des résultats.

4-5-2 Les différentes techniques de reconstitution testées.

Elles s'appuient toutes sur la formulation proposée au chapitre 3 en simulation. Si nous notons $X^*_d(t, \xi, \underline{x})$ la SST désaisonnalisée reconstituée, nous aurons:

$$X^*_d(t, \xi, \underline{x}) = \sum_{k=1}^{K} Z_k(t, \xi) \,.\, F_k(\underline{x}) + \varepsilon(t, \xi, \underline{x}) \qquad (4\text{-}19)$$

ou encore pour le mois t_i, l'année ξ_l et le point $\underline{x}_j$:

$$X^*_d(t_i, \xi_l, \underline{x}_j) = \sum_{k=1}^{K} Z_k(t_i, \xi_l) \,.\, F_k(\underline{x}_j) + \varepsilon(t_i, \xi_l, \underline{x}_j) \qquad (4\text{-}20)$$

Ce qui distinguera chaque méthode testée sera <u>le mode de calcul des fonctions propres</u>, qui pourra être de quatre types:

1- *ACP classique de la SST désaisonnalisée de 1964 à 1985*

2- *ACPP de la SST désaisonnalisée de 1964 à 1985* en utilisant une discrétisation *constante par morceau* (CPM)

3- *ACPP de la SST désaisonnalisée de 1964 à 1985* en utilisant une discrétisation *linéaire par facette* (LPF) (cf Fig. 4-1).

4- *ACPP en LPF* en remplissant la matrice **C** de variance covariance en utilisant le *modèle exponentiel ajusté sur l'ensemble du domaine* (cf § 4-3-3):

$$C(h) = 34.0 \exp(- h/1586.) \qquad (4\text{-}21)$$

<u>Remarque</u>: Nous aurions aimé exploiter l'information sur les anisotropies de structure révélées par l'analyse variographique par azimut, en utilisant les divers modèles ajustés dans chacune des quatre directions d'investigation. Mais, dans ce cas, la matrice **C** n'est plus forcément définie positive et, effectivement, la tentative a été vouée à l'échec car nous obtenions la moitié des valeurs propres négatives et, de plus, du même ordre de grandeur, en valeur absolue, que les valeurs propres positives. C'est pourquoi nous avons choisi de n'utiliser qu'un seul modèle, le modèle moyen, tout en sachant bien qu'il simplifiait considérablement la structure du phénomène. Pour tenir compte des anisotropies, il aurait fallu caler des modèles ayant même palier dans chaque direction et des portées éventuellement différentes, ce qui aurait permis de conserver la positivité de la matrice **C**.

Les $Z_k(t_i, \xi_l)$ ont été calculés grâce au produit matriciel :

$$\mathbf{Z = X \,.\, E \,.\, F} \qquad (4\text{-}22)$$

où **X** est, ici, la matrice (24,P) des 24 champs de 1986 et 1987 observés aux P stations $\underline{x}_j$, j=1..P.

Remarque: Notons que nous pouvons intégrer l'ACP classique dans cette formulation, en prenant pour la matrice **E**, la matrice identité $\mathbf{I_P}$, de dimension (P,P).
De même, le système à résoudre pour calculer la matrice des fonctions propres aux P stations, s'écrira pour toutes les techniques **C . E . F = F .** Λ.

Ce qui distinguera chacune sera le mode de calcul de **C** (à l'aide des données de 1964 à 1985 ou du modèle (4-21)) et celui de **E** (ACP classique ou de processus, en CPM ou en LPF).

Enfin, le résidu $\varepsilon(t_i, \xi_l, \underline{x}_j)$, si nous nous référons au chapitre 3 sera représenté par un bruit blanc gaussien, ajouté pour compenser le déficit en variance dû à la troncature à K F.P. Dans cette étude, nous avons choisi K expliquant 90% de la variance totale, ce qui donne K=39 pour l'ACP classique et l'ACPP en CPM; K=34 pour l'ACPP en LPF et K=44 pour l'ACPP utilisant le modèle, mais dans ce dernier cas nous avons pu calculer 628 F.P.(c'est-à-dire autant que de points de grille) contre 264 dans les trois autres cas (où la matrice de covariance était de rang N=264 < P=628). Nous avons vu au chapitre 3 que nous pouvions calculer la variance du résidu en chaque point. A cause de problèmes numériques, nous n'avons pas pu le faire ici et nous avons dû nous contenter de générer un bruit blanc gaussien de même variance partout, égale à 10% de la variance totale.

Nous avons aussi testé chaque méthode sans ajouter ce résidu ε, ce qui nous donne, au total, 8 techniques de reconstitution à comparer.

4-5-3 Les critères de comparaison.

Il s'agit essentiellement de critères statistiques sur les écarts $e(t_i, \xi_l, \underline{x}_j)$ entre valeurs observées et valeurs reconstituées:

$$e(t_i, \xi_l, \underline{x}_j) = X_d(t_i, \xi_l, \underline{x}_j) - X^*_d(t_i, \xi_l, \underline{x}_j) \qquad t_i,\ i=1,..12;\ \underline{x}_j\ j=1..P,\ \xi_l\ l=86,87 \tag{4-23}$$

i) *Erreur relative normée par la moyenne* E.R. : il s'agit de la moyenne des écarts divisée par la moyenne des valeurs observées. E.R. nous renseigne sur l'existence d'un biais systématique.

ii) *Erreur absolue normée par la moyenne* E.A. : c'est la moyenne des valeurs absolues des écarts, toujours normée par la moyenne des valeurs de référence. Ce critère donne des indications sur la dispersion des écarts tout comme le critère suivant:

iii) *Erreur quadratique moyenne normée par la variance des valeurs observées* E.Q. Ce dernier critère donne beaucoup de poids aux fortes erreurs (à cause du carré).

iv) On peut aussi tracer l'histogramme de ces écarts afin de déceler d'éventuelles particularités de leur distribution.

Une analyse des corrélations entre valeurs observées et reconstituées a, bien sûr, été menée. Nous avons calculé la pente a et le coefficient b de la droite de régression:

$$X^*_d(t_i, \xi_l, \underline{x}_j) = a \,.\, X_d(t_i, \xi_l, \underline{x}_j) + b \qquad (4\text{-}24)$$

et le coefficient de corrélation R correspondant, pour l'ensemble des 24x628 = 15072 valeurs reconstituées, mais aussi champ par champ, ce qui nous permet, en classant les 24 coefficients $R^2(t_i, \xi_l)$, i=1..12, l=86,87 obtenus par ordre croissant, de tracer des courbes dites "d'efficacité" de chacune des méthodes. Si ces dernières étaient parfaites, les critères a, R, $R^2(t_i, \xi_l)$ vaudraient un et b serait nul.

4-5-4 Analyse des résultats globaux.

Ils sont résumés dans les tableaux 4-9 (calcul des erreurs) et 4-10 (analyse des corrélations).

Technique	Moyenne	Variance	E.R.	E.A.	E.Q
Données Observées	1.69	37.22			
ACP class. 64.85 S.R.	1.67	30.23	1.16	120.10	18.90
ACPP 64.85 CPM S.R.	1.67	30.20	1.14	120.00	18.90
ACPP 64.85 LPF S.R.	1.61	29.44	4.43	123.80	21.00
ACPP LPF modèle S.R.	1.64	31.33	2.50	95.80	13.10
ACP class. 64.85 A.R.	1.67	34.04	1.49	150.80	28.40
ACPP 64.85 CPM A.R.	1.67	33.42	1.26	148.90	27.70
ACPP 64.85 LPF A.R.	1.62	33.11	4.15	156.40	31.20
ACPP LPF modèle A.R.	1.66	34.85	1.73	131.30	22.00

Tableau 4-9: Moyenne, variance, erreur relative E.R., absolue E.A. et quadratiques E.Q. calculées sur les 15072 valeurs reconstituées. Dans la désignation des techniques, S.R. signifie "sans résidu" et A.R. "avec résidu".

Plusieurs choses sont frappantes. Tout d'abord, quelle que soit la technique, les critères sont meilleurs lorsqu'on ne rajoute pas de résidus, sauf parfois l'erreur relative qui, d'ailleurs, varie peu. Ceci voudrait dire que le résidu (dont l'espérance est nulle) ne biaise pas la reconstitution, mais augmente sensiblement la dispersion du nuage. Nous n'avions pas observé un tel phénomène au chapitre 3 lors de l'utilisation de l'ACPP en simulation où, là aussi, nous ajoutions un résidu ε. Mais alors, nous travaillions avec des données de variance

identique en tout point, ce qui permet d'expliquer que nous n'ayons pas eu de problème particulier d'instabilité. Au contraire, nous avions pu observer l'effet bénéfique de cet ajout sur la variance totale.

Ici, la Fig. 4-25 de l'histogramme des écarts entre valeurs observées et reconstituées par ACP classique des données de 1964 à 1985 illustre bien la dispersion plus grande avec le résidu.

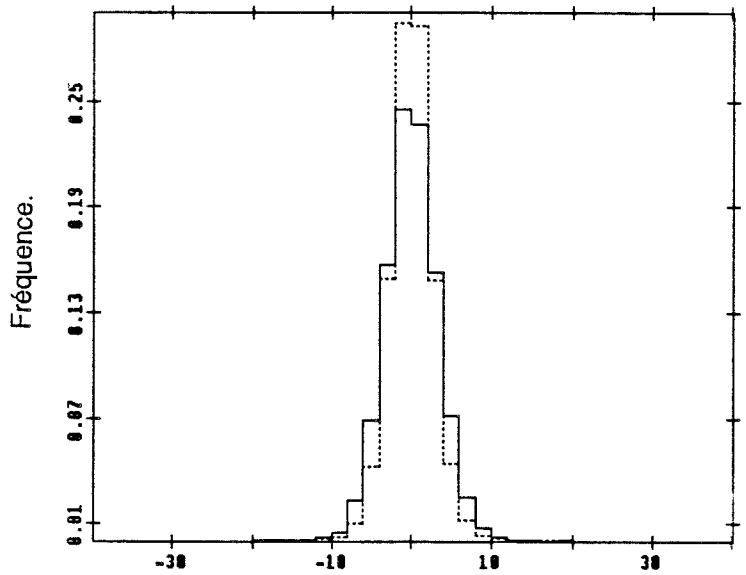


Fig 4-25: Histogramme des écarts entre valeurs observées et reconstituées par ACP classique des données de 1964 à 1985
- - - - sans ajouter de résidu. Moyenne des écarts=.019, variance des écarts= 7.12
——— en ajoutant un résidu. Moyenne des écarts=.021, variance des écarts= 10.42.

Néanmoins, c'est lorsqu'on ajoute les résidus qu'on est plus proche de la vraie variance des données (colonne 2 du tableau 4-9), la reconstitution avec uniquement les K premières F.P. fournissant des champs a priori trop lisses. Mais, la plus grande dispersion des écarts, traduite par l'augmentation de E.A. et E.Q. et la chute de la corrélation, nous montre que l'ajout d'un résidu non structuré spatialement augmente artificiellement la variance des données en bruitant les champs, ce qui les écarte des champs réels, relativement lisses. Peut-être aussi que les résidus ajoutés, non corrélés spatialement, s'écartent trop de la structure des vrais résidus (donnés par les N-K fonctions propres laissées de côté) qui pourraient être corrélés spatialement. Ceci voudrait alors dire que nous avons tronqué trop rapidement la série des F.P. et que de l'information structurée était encore contenue dans les résidus.

Cependant, si on voulait utiliser les champs reconstitués pour étudier l'influence de leur variabilité sur des sorties de modèles par exemple, peut-être faudrait-il privilégier les champs avec résidus dont la variabilité est plus grande et donc plus proche de celle des données réelles (qui, déjà, ont été fortement lissées par l'interpolation optimale à laquelle elles ont été soumises).

Si l'on cherche maintenant à classer les quatre modes de calcul des F.P., indépendamment de l'ajout du résidu, c'est lorsque **C** est remplie à l'aide du modèle que les résultats en corrélation sont les meilleurs. L'ACP classique et l'ACPP en CPM sont strictement équivalentes, mais compte-tenu de la régularité de la grille, et malgré les déformations des surfaces induites par la latitude, on peut admettre que les surfaces associées à chacune des stations soient équivalentes, ce qui explique le résultat car la matrice **E** est alors pratiquement la matrice identité, à une constante multiplicative près. On peut espérer que, sur un réseau plus irrégulier, l'ACPP en CPM aurait donné de meilleurs résultats que l'ACP classique.

Technique	R	R^2	a	b	variance résiduelle
ACP class. 64.85 S.R.	0.90	0.81	0.811	0.296	5.70
ACPP 64.85 CPM S.R.	0.90	0.81	0.811	0.298	5.71
ACPP 64.85 LPF S.R.	0.89	0.79	0.79	0.278	6.20
ACPP LPF modèle S.R.	0.93	0.87	0.855	0.202	4.12
ACP class. 64.85 A.R.	0.85	0.73	0.815	0.286	9.32
ACPP 64.85 CPM A.R.	0.86	0.73	0.811	0.297	8.95
ACPP 64.85 LPF A.R.	0.84	0.70	0.789	0.285	9.95
ACPP LPF modèle A.R.	0.89	0.79	0.858	0.21	7.45

Tableau 4-10: Coefficients de corrélation R, % de variance expliquée R^2, coefficients a et b de la droite de régression et variance résiduelle calculés sur les 15072 valeurs reconstituées. Dans la désignation des techniques, S.R. signifie "sans résidu" et A.R. "avec résidu".

Enfin, c'est l'ACPP en LPF qui est la moins performante, mais, on perd seulement 2% de variance expliquée par rapport à l'ACP classique, contre 8% par rapport à l'ACPP (toujours en LPF) du modèle.... La supériorité de cette dernière technique, utilisant un modèle lisse, vient donc confirmer ce que nous avancions au chapitre 2: le calcul des F.P. est alors meilleur qu'avec une matrice de covariance empirique. Nous avions d'ailleurs aussi noté que l'approche LPF était beaucoup plus sensible que l'approche CPM à un échantillonnage insuffisant, ce qui pourrait expliquer son moins bon comportement ici, où le nombre d'observations est réduit par rapport au nombre de stations. (Mais ajoutons aussi au crédit de l'approche LPF qu'elle permet de concentrer plus de variance sur moins de composantes que l'ACP classique K90% = 34 contre 39).

On peut néanmoins être surpris que l'utilisation d'un modèle de covariance unique donne les meilleurs résultats. En effet, les paragraphes 4-2 et 4-3 nous avaient montré la

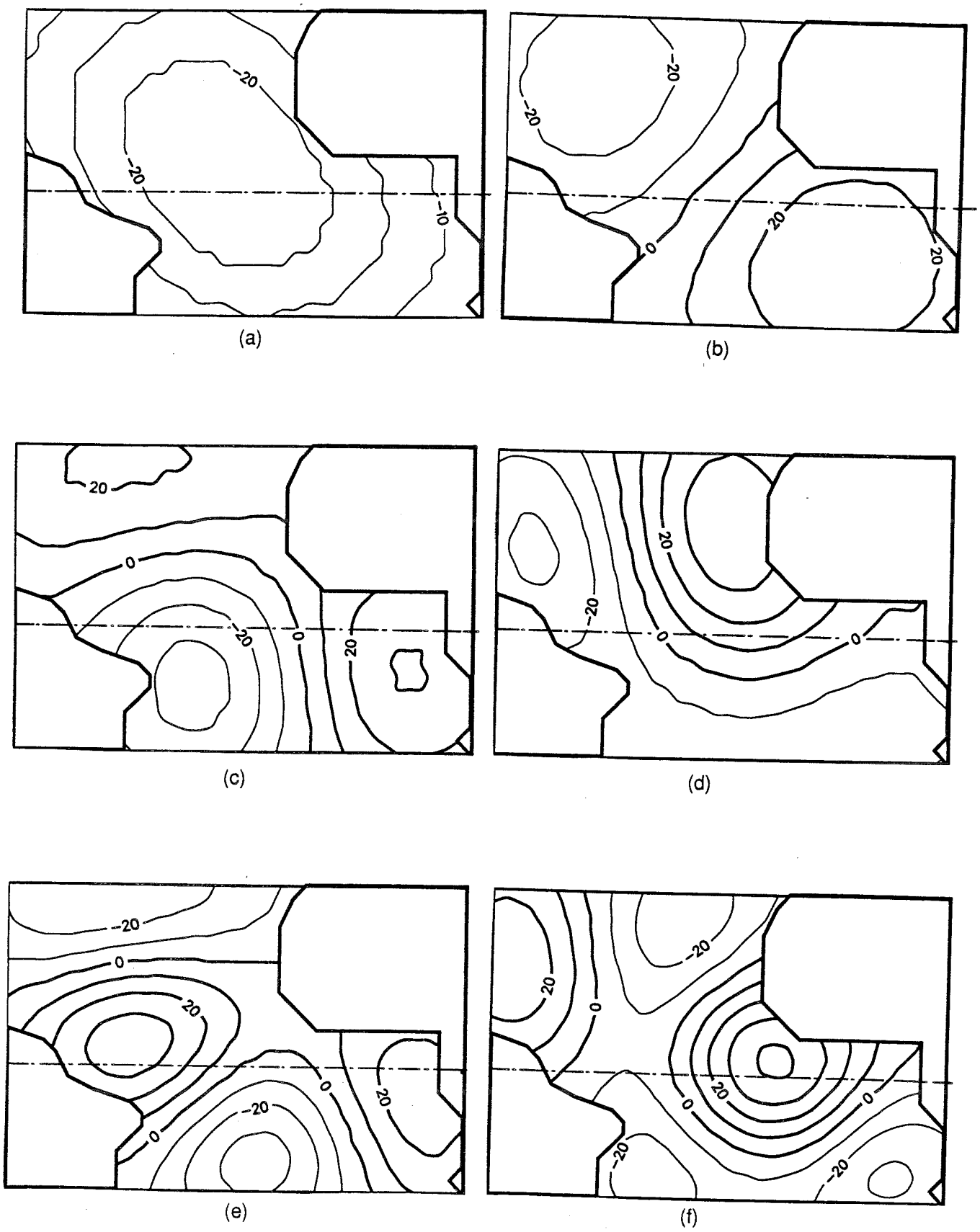


Fig. 4-26:Fonctions propres calculées par ACPP (LPF) de la SST désaisonnalisée en remplissant la matrice de covariance **C** à l'aide du modèle:

C(h) = 34.0 exp(-h / 1586.)

(a) F.P. 1 (b) F.P. 2 (c) F.P. 3 (d) F.P. 4 (e) F.P. 5 (f) F.P. 6

non-stationnarité de la SST désaisonnalisée et on pouvait craindre qu'un modèle homogène et isotrope soit incapable de bien reproduire ces inhomogénéités.

Le classement des différentes techniques est confirmé par les valeurs des pentes de la droite de régression qui est plus proche de un lorsque **C** est remplie avec le modèle et c'est encore avec cette méthode que les erreurs absolues et quadratiques ont les valeurs les plus faibles.

Avant de passer à l'analyse champ par champ et aux courbes d'efficacité, nous aimerions détailler un peu les résultats de l'ACPP lorsque **C** est remplie avec le modèle. Ce qui est frappant, c'est la grande différence entre les F.P. estimées avec le modèle ou avec les covariances empiriques. La Fig. 4-26 présente les six premières F.P. obtenues à partir du modèle que l'on pourra comparer avec leurs homologues données par exemple par ACP classique du fichier 1964-1987 (Fig. 4-18). En fait, on n'a pas le sentiment d'observer des fonctions sensées décrire le même phénomène: les F.P. calculées avec le modèle sont beaucoup plus lisses et les gradients sont bien moins élevés. Par ailleurs, les différents noyaux positifs et négatifs semblent plutôt s'organiser selon la géométrie un peu particulière du domaine, dont nous retrouvons encore le rôle prépondérant sur l'estimation des F.P. De plus, contrairement aux F.P. estimées à l'aide des covariances empiriques, on ne peut plus relier les F.P. aux zones de forte variabilité de la SST, puisque, en imposant un modèle unique sur tout le domaine, nous en avons "oublié" l'existence. Néanmoins, lorsque l'ordre des F.P. augmentera, les différents noyaux positifs et négatifs verront leur zone d'influence diminuer progressivement, décrivant alors des phénomènes plus locaux, notamment près des frontières. Mais, ceci est plutôt un effet géométrique et n'a pas vraiment de lien avec la donnée elle-même. Cependant, on peut admettre que, lors du calcul des $Z_k(t_i, \xi_l)$, qui fait intervenir les valeurs observées, s'il y a des zones à plus forte variabilité, elles induiront, pour certaines valeurs de k d'ordre élevé, des valeurs des Z_k plus fortes, venant alors "exciter" une F.P. décrivant une zone réduite. Mais ceci resterait à vérifier!....

Passons maintenant à l'analyse champ par champ et à l'étude des courbes d'efficacité.

4-5-5 Analyse des résultats champ par champ.

Ils viennent confirmer les résultats que nous avons déjà soulignés globalement:

i) la supériorité d'une reconstitution sans résidu plutôt qu'avec résidu. La Fig. 4-27 des courbes d'efficacité pour l'ACPP des données de 1964 à 1985 en LPF avec ou sans résidu l'illustre clairement et il en va de même pour toutes les techniques. De même, la Fig. 4-28 qui présente le champ de Février 1986 reconstitué par la même méthode avec ou sans

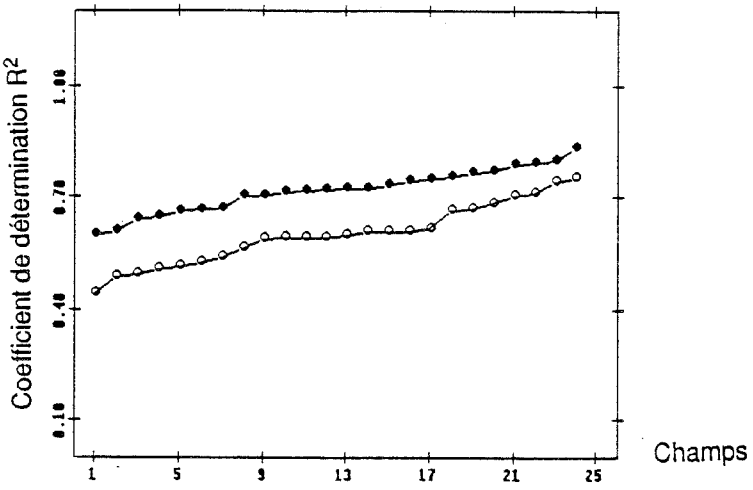


Fig. 4-27:Courbe d'efficacité (coefficient de détermination $R^2(t_i,\xi_i)$ classés) des 24 champs reconstitués par ACPP des données de 1964 à 1985 en LPF.

—●— sans résidu

—○— avec résidu.

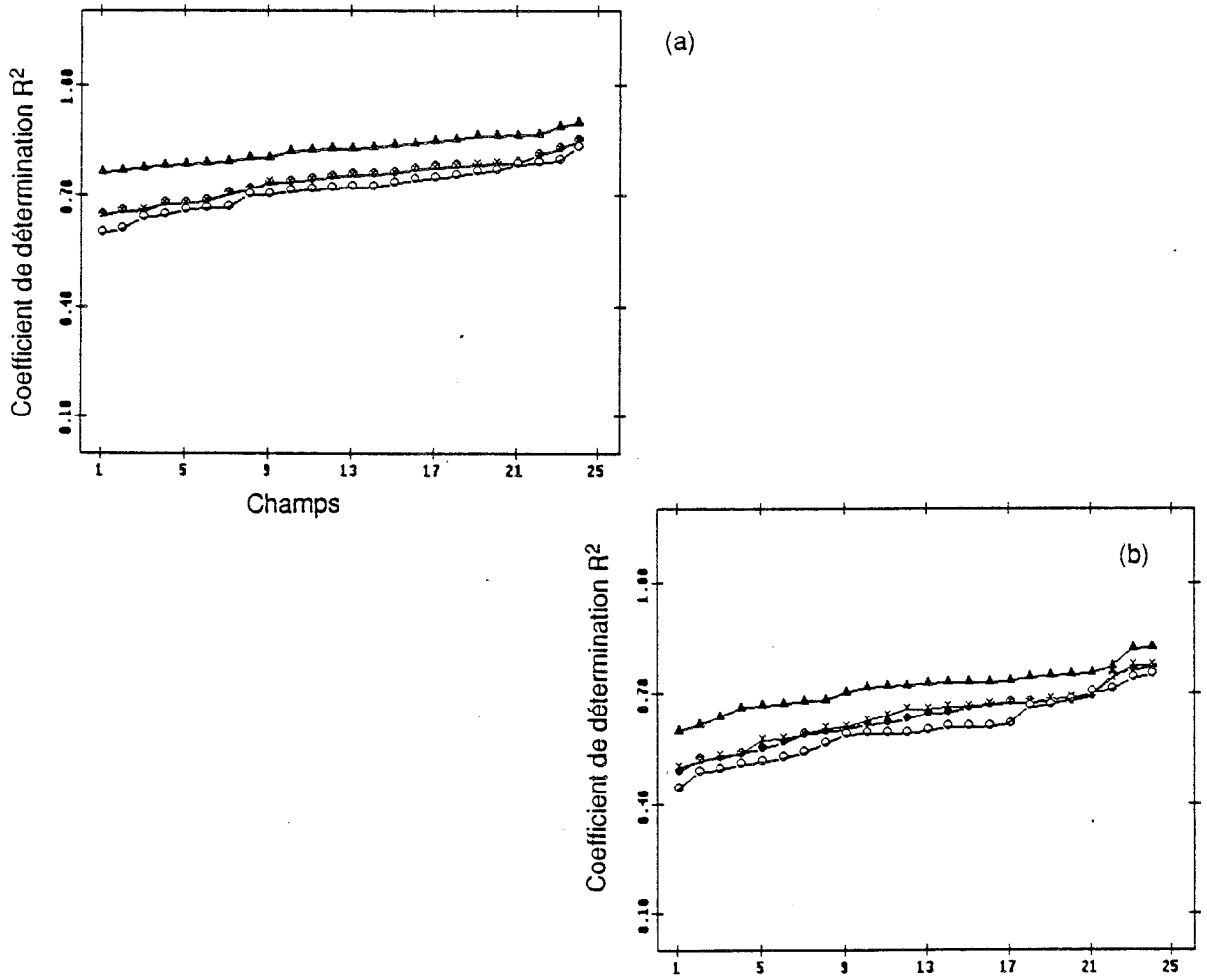


Fig. 4-29:Courbes d'efficacité des 24 champs reconstitués à l'aide des quatre techniques étudiées. (a) sans résidu (b) avec résidu.

—●— ACP classique des données de 1964 à 1985.

—✕— ACPP des données de 1964 à 1985 en CPM.

—○— ACPP des données de 1964 à 1985 en LPF.

—▲— ACPP en LPF en remplissant **C** à l'aide du modèle

$C(h) = 34.0 \exp(-h / 1586.)$

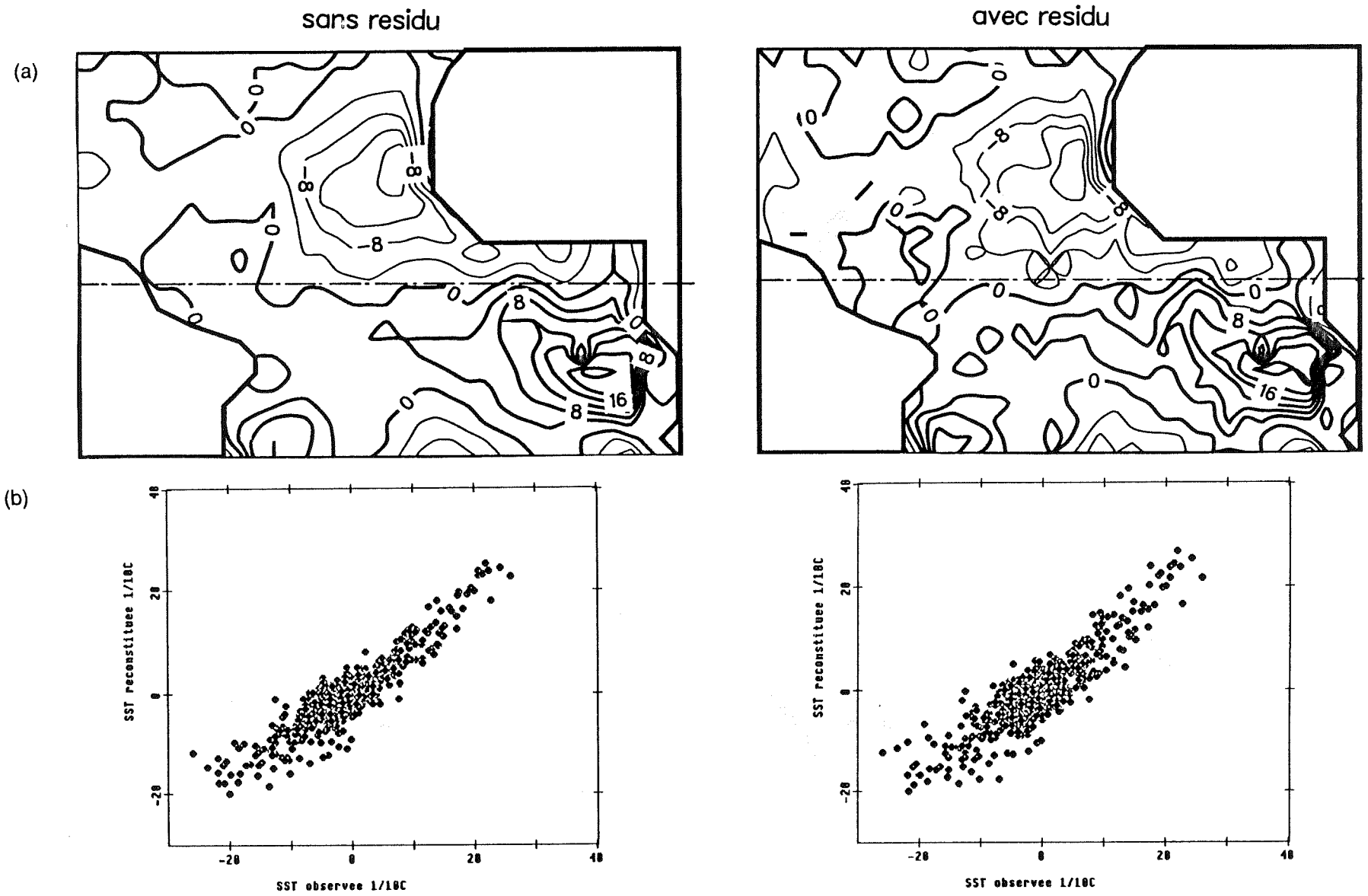


Fig. 4-28:(a) Cartographie du champ de Février 1986 reconstitué par ACPP de la SST désaisonnalisée de 1964 à 1985 en CPM, à gauche sans résidu, à droite avec résidu.

(b) Nuages de corrélation valeurs observées/ valeurs reconstituées

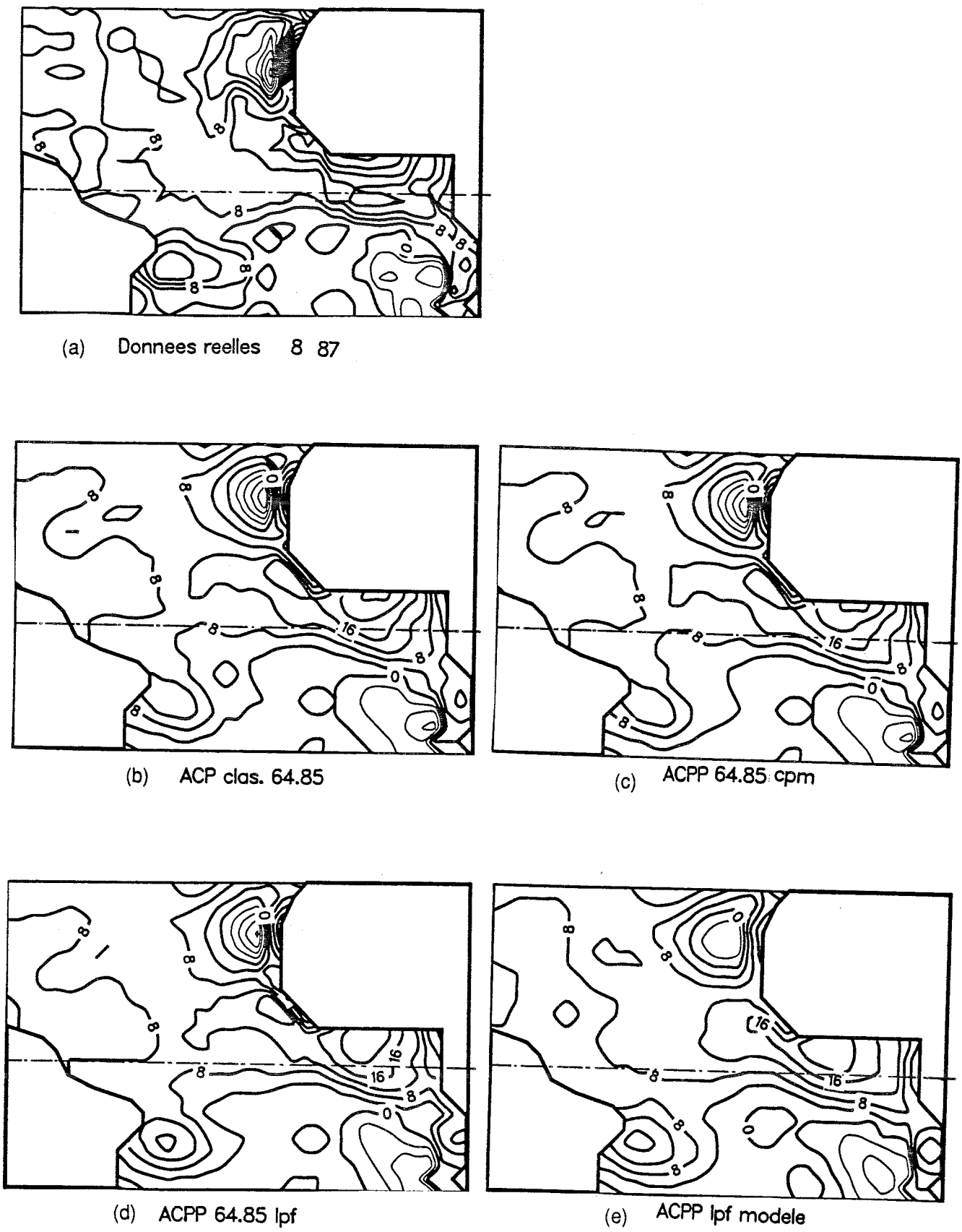


Fig. 4-30:Cartographie du champ de SST désaisonnalisée d'Aout 1987

(a)	Champ observé	
(b)	ACP classique des données de 1964 à 1985.	R=.90
(c)	ACPP des données de 1964 à 1985 en CPM.	R=.90
(d)	ACPP des données de 1964 à 1985 en LPF.	R=.89
(e)	ACPP en LPF en remplissant **C** à l'aide du modèle $C(h) = 34.0 \exp(-h / 1586.)$	R=.91

résidu, ainsi que les nuages de corrélation correspondants montrent la plus grande dispersion du nuage lorsqu'on ajoute les résidus, et la plus grande irrégularité des champs simulés.

ii) les meilleurs résultats sont obtenus lorsque **C** est remplie avec le modèle. L'ACP classique et l'ACPP en CPM sont équivalentes (les deux courbes d'efficacité se superposent) alors que l'ACPP en LPF est légèrement moins bonne (Fig. 4-29-a pour les courbes d'efficacité sans résidu et 4-29-b avec résidu).

Si on observe quels sont les champs bien ou mal reconstitués par les diverses méthodes, il y a cohérence entre elles: les champs de 1986 (correspondant à des champs de moyenne faible) sont mieux reconstitués que ceux de 1987 (où les moyennes sont plutôt fortement positives; 4 à 5 1/10 °C). Mais cet effet "de série" peut s'expliquer par l'autocorrélation temporelle entre les champs que nous avions évoquée à la fin du § 4-2.

La Fig. 4-30 présente enfin le champ d'Aout 1987 réel et reconstitué par les quatre techniques, sans résidu.

4-6 CONCLUSIONS

Ce dernier chapitre, et notamment le dernier paragraphe, par un essai de reconstitution des champs de SST désaisonnalisée de 1986 et 1987, ne nous a pas permis de montrer de façon probante la supériorité de l'approche de type processus sur l'ACP classique, surtout dans sa version utilisant des fonctions linéaires par facette. Néanmoins, la prise en compte de la continuité spatiale de la SST à travers l'analyse variographique nous a permis d'ajuster un modèle moyen de covariance, dont l'utilisation s'est avérée performante. Il semblerait donc bien que, même sur des données non homogènes, l'utilisation d'un modèle soit préférable à celle des covariances empiriques, au moins en termes de reconstitution sur l'exemple traité. Ceci demanderait à être confirmé par l'étude d'autres jeux de données.

De plus, c'est peut-être moins l'utilisation des covariances empiriques en elle-même qui est à incriminer, mais plutôt l'insuffisance de l'échantillonnage statistique (nombre d'observations) qui ne permet pas d'assurer la robustesse des F.P. estimées, comme nous l'avons vu au § 4-3 lors de l'étude de stabilité des F.P. en dégradant aléatoirement le réseau de base. Dans ce cas, l'interprétation des F.P. est délicate et le modèle moyen, même s'il ne donne aucune information sur les hétérogénéités peut être préférable.

Cette application, si elle ne nous a certes pas permis de prouver la supériorité de l'ACPP sur l'ACP classique, nous a au moins confirmé certains résultats annoncés au chapitre 2 lors de la comparaison des solutions analytiques et numériques: la grande importance des effets d'échantillonnage sur l'estimation des F.P. et la plus grande stabilité donnée par l'utilisation d'un modèle ajusté.

Par ailleurs, le choix d'un exemple avait été orienté vers un problème présentant un domaine bien défini (ici le bassin Atlantique intertropical), puisque nous avions mis en évidence l'importance de ce domaine d'étude. Nous pouvons regretter que la régularité de la grille et les lissages induits par l'utilisation de données par blocs ne nous aient pas permis d'exploiter pleinement, sur ce domaine, les possibilités de l'ACPP sur un réseau irrégulier. Néanmoins, nous pouvons citer les résultats d'un autre travail, réalisé à l'IMG par Berlin et Cendrier (1986), sur des champs journaliers de géopotentiels couvrant l'Europe Occidentale. Dans ce cas, assez typique des exemples atmosphériques, le réseau des stations de radiosondage est irrégulier (plus dense sur le continent que sur l'océan) mais le domaine d'étude doit être défini a priori car il n'y a pas de frontières "naturelles" évidentes. Sur ce réseau, on a pu mettre en évidence une supériorité de l'ACPP utilisant des fonctions linéaires par facette (LPF) sur l'ACP classique lorsque les composantes principales déduites de chacune des analyses sont utilisées pour la prévision de la pluviométrie des Cévennes. Le gain est certes faible (3% de variance expliquée supplémentaires) mais confirmé par plusieurs essais. Bien sûr, il ne s'agit pas d'une démonstration au sens strict, mais seulement d'une suggestion: l'ACPP extrait des mêmes données initiales une information qui se révèle plus explicative que celle extraite par ACP classique, pour prévoir une variable exogène: la précipitation...

L'idéal serait donc, pour tester efficacement les apports de l'approche ACP de processus de pouvoir faire la comparaison avec les techniques classiques sur un exemple où:

i) le domaine d'étude serait défini par la géographie, car nous avons montré le rôle de ce domaine sur la forme des fonctions propres.

ii) le réseau serait irrégulier, et où on disposerait de données brutes (non affectées par des lissages ou des agrégations préalables qui peuvent dénaturer la fonction de covariance au voisinage de l'origine, et gommer en partie des hétérogénéités)

ii) l'échantillonnage statistique serait suffisant, pour permettre aux deux effets précédents, de nature plutôt analytique, voire géométrique, d'être sensibles par rapport au "bruit" statistique qui tend vite à devenir prédominant lorsque le nombre de champs disponibles est faible par rapport au nombre de "stations" ou points de grille utilisé.

CONCLUSIONS GENERALES

Le travail qui a été présenté dans ce mémoire s'inscrit donc dans la ligne de recherche initiée à Grenoble par D. Duband au début des années 1970. A l'IMG, les travaux d'Obled (1979), poursuivis par Creutin et Obled (1982), Bouhaddou (1984) et Obled et Creutin (1986) ont cherché à étendre les principes de l'Analyse en Composantes Principales (ACP), connue et utilisée depuis très longtemps en traitement de données, à des variables présentant une continuité spatiale et/ ou temporelle, donc à des processus. On parle alors d'ACP de processus (ACPP).

Dans le premier chapitre, nous avons rappelé les fondements de cette extension, qui conduit à rechercher, sur un domaine D borné, les fonctions propres d'un noyau de covariance et donc à résoudre une équation intégrale de Fredholm. Puis, une approximation numérique de ce problème a été présentée avant de développer comment cette technique pouvait se rattacher à la recherche de modes normaux de systèmes dynamiques d'une part, à l'approche géostatistique et aux méthodes d'interpolation optimale d'autre part. Nous avons alors pu montrer que, dans sa version continue, l'ACPP pouvait s'interpréter comme un cas particulier de krigeage de données entachées d'erreurs de mesure. Cette analogie s'appuie sur le fait que l'ACPP fournit un modèle de covariance, s'exprimant à l'aide des seules fonctions propres et valeurs propres, possèdant par ailleurs la propriété de ne pas être nécessairement homogène et isotrope.

Compte tenu des utilisations nombreuses de l'ACPP en Géophysique, et de toutes les interprétations qui sont généralement recherchées et proposées a posteriori pour expliquer l'allure des fonctions propres, nous nous sommes interrogés, dans le deuxième chapitre, sur le degré de confiance que l'on pouvait accorder à ces fonctions lorsqu'elles sont estimées sur un échantillon de taille finie et un domaine discrétisé spatialement. Pour y répondre, dans la lignée des travaux de Fortus (1973, 1975) ou de Bouhaddou (1984), nous avons proposé, pour un modèle de covariance homogène, isotrope, à densité spectrale rationnelle et sur un cercle de R^2, une solution analytique du problème intégral de Fredholm, en le transformant en système d'équations aux dérivées partielles avec conditions aux limites. Dans ce traitement, qui exclut les effets statistiques on a pu mettre en relief la grande importance de la taille et de la forme du domaine sur le résultat final.

Cette solution analytique a ensuite été exploitée pour quantifier les erreurs introduites, dans les cas réels, par l'approximation numérique, ce qui nous a montré la grande sensibilité de l'estimation des fonctions propres à l'échantillonnage statistique. En

revanche, il s'est avéré que la discrétisation spatiale, pour peu que la densité du réseau soit cohérente avec la longueur de corrélation, avait un rôle moins prépondérant. Enfin, nous avons aussi mis en évidence l'intérêt de l'approche ACPP lorsqu'il est nécessaire d'interpoler les fonctions propres en des points où elles n'ont pas été calculées.

Dans le troisième chapitre, nous avons essayé de formaliser l'utilisation de l'ACPP pour la simulation de champs bidimensionnels, utile lorsque l'on cherche, par exemple, à tester la sensibilité d'un modèle géophysique à la variabilité des entrées. A travers une étude de sensibilité de la méthode au choix des paramètres de la simulation, en termes de restitution d'une covariance cible, nous avons proposé une gamme de choix pour ces derniers afin d'obtenir une bonne reconstitution du modèle de covariance. Une comparaison avec d'autres techniques de simulation (bandes tournantes, méthode LU) nous a montré que l'ACPP se révélait tout aussi performante, malgré quelques problèmes dus à la troncature de la série des fonctions propres. Enfin, un petit exemple simplifié de simulation de champs hétérogènes nous a montré que l'ACPP permettait de bien répondre à ce type de problèmes. La méthode peut même servir à détecter les inhomogénéités, pourvu que l'on soit dans un contexte de multiréalisations avec des conditions d'échantillonnage statistique raisonnables.

Enfin, dans le quatrième et dernier chapitre, nous nous sommes consacrés à une application à un cas réel. Celui que nous avons retenu s'appuyait sur le fichier des températures mensuelles de surface de la mer (SST), établi de 1964 à 1987 par J. Servain sur l'Atlantique intertropical. Une comparaison entre les F.P. estimées par ACP classique ou par ACPP et une étude de sensibilité des résultats lorsque le réseau de mesure est dégradé aléatoirement, a mis en évidence que les incertitudes introduites par l'échantillonnage statistique étaient probablement trop grandes pour accorder beaucoup de confiance aux fonctions propres estimées. Ceci a été confirmé par les résultats de la validation de plusieurs techniques basées sur l'ACP ou l'ACPP pour la reconstitution des champs de 1986 et 1987. En effet, toutes les techniques utilisant des fonctions propres estimées à partir d'une matrice de covariance empirique avaient de moins bonnes performances que lorsque **C** était remplie en utilisant un modèle moyen obtenu par l'analyse de la structure spatiale de la SST.

L'exemple choisi pour illustrer le cas d'un domaine bien défini mais quelconque était donc, en fait, trop "handicapé" au départ par le manque d'observations pour que les conclusions sur les différentes méthodes soient probantes.

Globalement cependant, l'ensemble des résultats obtenus dans ce mémoire a mis en évidence la grande sensibilité de l'ACPP à l'échantillonnage statistique, c'est-à-dire au nombre d'observations disponible. La comparaison entre solutions analytiques et numériques a montré que les instabilités induites sur la structure des F.P. ne pouvaient être atténuées

qu'en présence d'un nombre très élevé de réalisations, permettant alors d'assurer la robustesse de la matrice de covariance empirique. Or, lors de l'application pratique de la méthode, ces conditions sont rarement satisfaites et on ne peut plus assurer la stabilité des F.P. estimées, ce qui est renforcé, le plus souvent, par la proximité des valeurs propres. Les contraintes d'orthogonalité des F.P. et le choix du domaine d'étude contribuent alors à rendre l'interprétation physique de ces fonctions propres fortement sujette à caution, notamment si l'on veut les relier à des inhomogénéités réelles du processus physique sous-jacent.

Néanmoins, si l'on accepte de considérer les fonctions propres comme un outil mathématique, pratique pour condenser l'information multiple contenue dans une masse importante de données, l'approche continue, développée ici, donne un cadre théorique à une technique (l'ACP classique) souvent utilisée comme une "boîte noire"... Par ailleurs, même si la difficulté d'interprétation des résultats est frustrante pour l'esprit, la méthode s'avère tout à fait adaptée à l'estimation et à la simulation de champs, où ses performances n'ont rien à envier à celles d'autres techniques reconnues.

Enfin, dans tous les développements méthodologiques présentés dans ce mémoire, nous avons posé comme hypothèse l'indépendance statistique des différents champs. Cette hypothèse n'était pas, il est vrai, vérifiée sur les données de SST et l'on sait par ailleurs que cela revient à réduire encore le nombre de champs indépendants "équivalent", et donc à aggraver les effets de l'échantillonnage statistique.

Il serait donc intéressant d'étendre encore les possibilités de la méthode pour permettre la description et la simulation de la variabilité spatio-temporelle des champs. En effet, la technique d'ACPP permet de "séparer" la variabilité spatiale, décrite par les fonctions propres, d'une évolution temporelle traduite sur les composantes principales. Au chapitre 3, nous avons vu comment la méthode permettait de simuler la structure spatiale. En générant, cette fois, les composantes avec une corrélation temporelle tirée de l'analyse de séries, on pourrait espérer restituer les variations spatio-temporelles de la variable étudiée. C'est une voie de recherche ouverte pour le futur que nous n'avons pas eu le temps d'approfondir sur les données de SST, qui pourtant s'y prêtaient....

REFERENCES BIBLIOGRAPHIQUES

ABRAMOWITZ M., STEGUN I.A. (1970)
Handbook of mathematical functions with formula graphs and mathematical tables, 9th édition, Dover publications, Inc, New-York, 1046 pp.

ALABERT F. (1987)
The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, **19** (5), 369-385.

ALFARO M. (1979)
Etude de robustesse des simulations des fonctions aléatoires. Thèse de docteur-ingénieur. Ecole des Mines de Paris, 161 pp.

AMOROCHO J., WU B. (1977)
Mathematical models for the simulation of cyclonic storms sequences and precipitation fields. *Journal of Hydrology*, **32** (3/4), 329-345.

APPERT N., LOUVEAU C. (1973)
Principales méthodes de calcul sur ordinateur. Rapport du centre de recherche Pont-à-Mousson n°219.73. France, 198 pp.

BARANCOURT C. (1990)
Etude de l'intermittence et de la variabilité des champs de pluie par une approche stochastique. Thèse de Doctorat de l'UJF, Grenoble, 268 pp.

BARNIER B. (1986)
Investigaton of seasonal variability of the wind stress curl over the north atlantic ocean by means of Empirical Orthogonal Function analysis. *Journal of Geophysical Research* **91** (C1), 863-868.

BASS J. (1962)
Eléments de calcul des probabilités théorique et appliqué. Masson, Paris, 219 pp.

BASS J. (1964)
Cours de mathématiques. Tome 1: Matrices et tenseurs. Courbes et surfaces. Intégrales simples et multiples. Séries et intégrale de Fourier. Fonctions orthogonales. Fonctions eulériennes. 3ème édition.Masson, Paris, New-York, Barcelone, Milan.

BASS J. (1978)
Cours de mathématiques. Tome 2 : Equations différentielles et aux dérivées partielles. Optimisation. Groupes de transformations. Méthodes numériques. 5ème édition. Masson, Paris, New-York, Barcelone, Milan, 466 pp.

BERLIN V., CENDRIER D. (1986)
Typologie de stations de mesures pluviométriques dans le bassin de la moyenne Durance. Analyse en Composantes Principales de Processus et fonctions empiriques orthogonales (EOF) appliquées à un réseau de stations de radiosondage. Rapport interne de l'Ecole Nationale de la Météorologie, Toulouse, 264 pp.

BOUHADDOU O. (1984)
Analyse en composantes principales et interpolation de processus. Méthode et simulation. Thèse de 3ème cycle USMG. Grenoble, 165 pp.

BOUHADDOU O., OBLED Ch., PHAM DINH TUAN (1987)
Principal component analysis and interpolation of stochastic processes: methods and simulation. *Journal of Applied Statistics*, **14** (3), 251-267.

BOULANGER F. (1989)
Géostatistique et processus autorégressifs : une nouvelle méthode de modélisation. *Geostatistics*, Vol 1 , M. Armstrong Ed, Kluwer Academic Publishers,Dordrecht, Boston, London, 283-294.

BRAS R.L., RODRIGUEZ-ITURBE I. (1976)
Rainfall generation : a nonstationnary time varying multidimensional model. *Water Resources Research,* **12** (3), 450-456.

BRAUD I. OBLED Ch. (1989)
A comparison between analytical eigenfunctions of a covariance kernel and their numerical approximations.Preprints of the 11th Conference on Probability and Statistics in Atmospheric Sciences. October 1-5 1989. Monterey (Ca). American Meteorological Society. 26-29.

BRAUD I. OBLED Ch. (1990)
Utilisation de l'Analyse en Composantes Principales de Processus pour la simulation de champs bidimensionnels. A paraître dans *Sciences de la Terre.*

BRILLINGER D.R. (1975)
Time series. Data analysis and theory. International series in decision processes. Holt, Reinhart et Windstar Ed, 500 pp.

BUELL E.C. (1971)
Integral equation representation for factor analysis.*Journal of the Atmospheric Sciences*, **28** ,1502-1505.

BUELL E.C. (1975)
The topography of the empirical orthogonal functions. Preprints of the 4th Conference on Probability and Statistics in Atmospheric Sciences. November 18-21 1975. Tallahasse (Florida). American Meteorological Society. 188-193.

BUELL E.C. (1978)
The number of significant proper functions of two dimensional fields. *Journal of Applied Meteorology*, **17** (6), 717-722.

BUELL E.C. (1979)
On the physical interpretation of empirical orthogonal functions. Preprints of the 6th Conference on Probability and Statistics in Atmospheric Sciences. Banff. American Meteorological Society, 112-117.

CHRISTAKOS G. (1984)
On the problem of permissible covariance and variogram models. *Water Resources Research*, **20** (2), 251-265.

CHRISTAKOS G. (1987)
Stochastic simulation of spatially correlated geo-processes. *Mathematical Geology,* **19** (5), 369-385.

CRESSMAN G.P. (1959)
An operational objective analysis system. *Monthly Weather Review*, **87** (10), 367-374.

CREUTIN J.D. (1979)
Méthodes d'interpolation optimale de champs hydrométéorologiques. Comparaison et application à une série d'épisodes pluvieux cévenols. Thèse de docteur-ingénieur. Grenoble USTMG-INPG.

CREUTIN J.D., OBLED Ch. (1982)
Objective analysis and mapping techniques for rainfall fields. An objective comparison. *Water Resources Research*, **18**, 413-431.

DAVENPORT W.B, ROOT W.L. (1958)
An introduction to the theory of random signal and noise. Mc graw Hill.

DAVIS M. (1987-a)
Production of conditional simulations via the LU triangular decompositon of the covariance matrix. *Mathematical Geology*, **19** (2), 91-98.

DAVIS M. (1987-b)
Generating large stochastic simulations. The matrix polynomial approximation method.*Mathematical Geology*, **19** (2), 99-109.

DAVIS R.E. (1976)
Predictability of sea surface temperature and sea level pressure anomalies over the north Pacific ocean. *Journal of Physical Oceanography*, **6**, 249-266.

DE FOUQUET C., BEUCHER H., GALLI A., RAVENNE C. (1989)
Conditional simulation of random sets. Application to an argilaceous sandstone reservoir. *Geostatistics*, Vol 2 , M. Armstrong Ed, Kluwer Academic Publishers, Dordrecht, Boston, London, 517-530.

DELHOMME J.P. (1979)
Spatial variability and uncertainty in groundwater flow parameters. A geostatistical approach. *Water Resources Research*, **15** (2), 269-280.

DEQUE M. (1986)
On the comparison between two sets of 500mb geopotential height eigenfunctions. *Tellus*, **38A**, 448-452.

DEQUE M., SERVAIN J. (1989)
Teleconnections between tropical atlantic SST and midlatitude 50kpa height during 1964-1986. *Journal of Climate*, **2**, 929-944.

DER MEGREDITCHIAN G. (1988)
Le traitement statistique des données multidimensionnelles. Direction de la Météorologie Nationale. Etablissement d'Etudes et de Recherches Météorologiques. Ecole Nationale de la Météorologie. Paris. Tome 1, 300 pp., Tome 2, 324 pp.

DEVILLE J.C. (1974)
Méthodes statistiques et numériques de l'analyse harmonique. *Annales de l'INSEE* n°15, 1-101.

DUBRULE O. (1982)
Krigeage et splines en cartographie automatique.Application à des exemples pétroliers. Thèse de l'Ecole des mines de Paris, 141 pp.

DUCHON J. (1976)
Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Revue Française d'Automatique, d'Informatique et de Recherche Opérationnelle, Analyse Numérique,* **10** (12), 5-12.

DYER T.G. (1975)
Assigment of rainfall stations into homogeneous groups : an application of Principal Component Analysis. *Quarterly Journal of the Royal Meteorological Society*, **101**, 1005-1013.

ECMWF (1977)
Workshop on the use of Empirical Orthogonal Functions in Meteorology. Bracknell. European Centre for Medium Range Weather Forecast. 155 pp.

ESCOUFFIER Y., LOBERT A., CASTELLANI X. (1968)
Interpolation de fonctions aléaoires en Hydrologie. Application aux pluies mensuelles de l'Allier. Note n°22, Faculté des Sciences de Montpellier.

FAYE J.P. (1982)
Des logiciels efficaces de calcul et de contrôle des valeurs propres et vecteurs propres. Thèse de doctorat 3ème cycle. Université Pierre et Marie Curie. Paris VI.

FORTUS M.I. (1973)
Statistically orthogonal functions for finite segments of a random process. *Izvestia Academic Sciences USSR. Atmospheric Oceanic Physics*, **9** (1), 34-36.

FORTUS M.I. (1975)
Statistically orthogonal functions for a random field specified in a finite region of the plane.*Izvestia Academic Sciences USSR. Atmospheric Oceanic Physics*, **11** (11), 1107-1112.

FORTUS M.I. (1980)
Method of Empirical Orthogonal Functions and its application in Meteorology. *Meteorologiya i Gidrologiya*, **4**, 113-119. (en Russe)

GALLE S. (1987)
Analyse des champs spatiaux par utilisation de la télédétection. Estimation de la durée quotidienne d'insolation en France à l'aide d'images du satellite Météosat et de mesures sol. Thèse de doctorat INPG. Grenoble. 257 pp.

GANDIN L.S. (1965)
Objective analysis of meteorological fields. Leningrad 1963, Trans. Israël Program for Scientific Translation, 242 pp.

GOUSSEBAILE J. (1977)
Analyses statistiques de données journalières d'insolation et d'irradiation. Application au Sud-Est de la France. Thèse de Docteur-Ingénieur, USMG-INPG, Grenoble, 215pp.

HOLMSTROM I. (1963)
On a method for parametric representation of the state of the atmosphere. *Tellus*, **15** (2), 127-149.

HOLMSTROM I. (1977)
Optimization of atmospheric models. *Tellus*, **29** (5), 415-427.

HOREL J.D. (1981)
A rotated Principal Component Analysis of the interanual variability of the northern hemisphere 500mb height field.*Monthly Weather Review*, **109**, 2080-2092.

JOURNEL A.G., HUIJBREGTS Ch. (1978)
Mining Geostatistics. Academic Press, 600 pp.

KELLY K.A. (1985)
The influence of winds and topography on the sea surface temperature patterns over the northern California slope. *Journal of Geophysical Research* . **90** (C6), 11-783, 11-798.

KETTANI-WAGNER S. (1986)
Etude comparative de techniques de simulation de fonctions aléatoires gaussiennes appliquées à l'analyse de la sécurité des structures. Thèse de doctorat 3ème cycle. Université de Clermont III.

KRAJEWSKY W.F., DUFFY C.J. (1988)
Estimation of correlation structure for a homogeneous isotropic random field: a simulation study. *Computers and Geosciences*, **14** (1), 113-122.

KUENY J.L. (1977)
Contribution au traitement statistique des données météorologiques. Thèse de Docteur Ingénieur, INPG, Grenoble.

LAGERLOEFF G., BERNSTEIN R.L. (1988)
Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara channel.*Journal of Geophysical Research* , **93** (C6), 6863-6873.

LEBART L.,FENELON J.P. (1973)
Statistique et Informatique Appliquées, Dunod, 2ème édition, 439 pp.

LOBERT A., DAUTY J. (1969)
Application des méthodes d'interpolation d'une fonction aléatoire aux problèmes régionaux de pluie. Note n°31, Faculté des Sciences de Montpellier.

LOEVE M. (1963)
Probability theory (3rd edition), D. Van Nostrand, New-York.

LORENTZ E.N. (1956)
Empirical Orthogonal Functions and statistical weather prediction. Sci. Rept n°1, Statistical Forecasting Weather Project, Mass. Inst. Tech, Dept. of Meteorology, Cambridge, Mass, 49 pp.

LUMLEY J.L. (1967)
The structure of inhomogeneous turbulent flows. *Atmospheric Turbulence and Radio Wave Propagation*, Ed. A.M. Yaglom et V.I. Tatarsky, Moscou, USSR, 166 pp.

LUMLEY J.L. (1970)
Stochastic tools in turbulence, Academic press, New-York, London, 194 pp.

MAC VEIGH J.P., BARNIER B. , LE PROVOST C. (1987)
Spectral and EOF analysis of four years of ECMWF wind stress curl over the north atlantic ocean. *Journal of Geophysical Research* , **92** , 13-141,13-152.

MANTOGLOU A.(1987)
Digital simulation of multivariate two and three dimensional stochastic processes with spectral turning band methods. *Mathematical Geology*, **19** (2), 129-150.

MANTOGLOU A., GELHAR L.W. (1987)
Effective hydraulic conductivities of transient unsaturated flow in stratified soils. *Water Resources Research*, **23** (1), 57-67.

MANTOGLOU A., WILSON J.L. (1981)
Simulation of random fields with the turning bands method. Report n°264, Mass. Inst. Tech., Department of civil engineering, 199 pp.

MANTOGLOU A., WILSON J.L. (1982)
The turning band method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, **18** (5), 1379-1394.

MATHERON G. (1965)
Les variables régionalisées et leur estimation. Ed Masson, Paris, 305 pp.

MATHERON G. (1970)
La théorie des variables régionalisées et ses applications. Ecole des Mines de Paris. Cahiers du Centre de Morphologie Mathématique, fascicule 5. Fontainebleau, France 212 pp.

MATHERON G. (1972-a)
Leçons sur les fonctions aléatoires d'ordre deux. Notes de cours . Centre de Géostatistique. ENSMP. Fontainebleau. 80 pp

MATHERON G. (1972-b)
Théorie des variables régionalisées. *Traité d'informatique géologique.* P. Laffite Ed. Masson. 306-377.

MATHERON G. (1973)
The intrinsic random functions and their applications, *Advanced Applied Probability*, **5**, 439-468

MEIJA G., RODRIGUEZ-ITURBE I. (1974)
On the synthesis of random fields from the spectrum. An application to the generation of hydrologic spatial processes. *Water Resources Research*, **10** (4), 705-711.

MOIN P., MOSER R.D. (1989)
Characteristic-eddy decomposition of turbulence in a channel, *Journal of Fluid Mechanics,* **200**, 471-509.

MOLTENI F. (1987)
Empirical Orthogonal Function analysis of the zonal and eddy components of 500mb height fields. Technical report n°61. European Centre for Medium Range Weather Forecast. 25 pp.

MOLTENI F., SUTERA A., TRONCI N. (1988)
The EOFs of the geopotential eddies at 500mb in winter and their probability density distributions. *Journal of the Atmospheric Sciences*, **45** (21), 3063-3080.

MORLIERE A., REBERT J.P. (1986)
Rainfall shortage and El-Nino southern oscillation in New-Caledonia southwestern Pacific. *Monthly Weather Review*, **114** (6), 1131-1137.

MUNOZ-PARDO J.F. (1987)
Approche géostatistique de la variabilité spatiale des milieux géophysiques. Application à l'échantillonnage de phénomènes bidimensionnels par simulation d'une fonction aléatoire. Thèse de docteur-ingénieur. INPG. Grenoble, 250 pp.

NORRIE D.H., DE VRIES G. (1973)
The finite element method. Academic Press.

NORTH G.R. (1984)
Empirical Orthogonal Functions and normal modes. *Journal of the Atmospheric Sciences*, **41** (5), 879-887.

NORTH G.R., BELL T.L., CAHALAN R.F., MOENG F.J. (1982)
Sampling errors in the estimation of Empirical Orthogonal Functions. *Monthly Weather Review*, **110**, 699-706.

OBLED Ch. (1979)
Contribution à l'analyse des données en Hydrométéorologie: la prévision des phénomènes accidentels et l'analyse des champs spatiaux. Thèse de Doctorat d'Etat ès Sciences Physiques, INPG-USMG, 256 pp.

OBLED Ch. (1987)
Introduction au krigeage à l'usage des hydrologues. 2ème journées hydrologiques de l'ORSTOM. Ed ORSTOM. Collection Colloques et Séminaires, 174-222.

OBLED Ch., BRAUD I. (1989)
Analogies entre géostatistique et analyse en composantes principales de processus ou analyse EOFs. *Geostatistics*, Vol 1 , M. Armstrong Ed, Kluwer Academic Publishers, Dordrecht, Boston, London, 237-248.

OBLED Ch., CREUTIN J.D. (1986)
Some developments in the use of Empirical Orthogonal Functions for mapping meteorological fields. *Journal of Applied Meteorology*, **25** (9), 1189-1204.

OBUKHOV A.M. (1960)
The statistically orthogonal expansion of empirical functions. *Izvestia Geophy. Ser.USSR. Atmospheric Oceanic Physics*, **60**, 280-291.

PAPOULIS A. (1962)
The Fourier integral and its applications. Mc Graw Hill, New-York.

PAPOULIS A. (1965)
Probability random variables and stochastic processes. Mc Graw Hill, 583 pp.

PICAUT J., SERVAIN J., LECOMTE P., SEVA M., LUKAS S., ROUGIER G. (1985)
Climatic atlas of the tropical Atlantic wind-stress and sea surface temperature 1964-1979. Université de Bretagne Occidentale, University of Hawaii, 467 pp.

PIEDELIEVRE J.P. (1985)
Téléconnexions entre la température de l'Atlantique intertropical et le géopotentiel du domaine tempéré "Europe- Atlantique". Rapport interne de l'Ecole Nationale de la Météorologie (E.N.M.), Toulouse, 201 pp.

PREISENDORFER R.W. (1979)
Principal components and the motions of simple dynamical systems. SIO Ref. Ser. 79-11, Scipps Institution of Oceanography, 79 pp.

PREISENDORFER R.W. (1988)
Principal component analysis in Meteorology and Oceanography. Developments in atmospheric sciences 17. Elsevier Amsterdam, Oxford, New-York,Tokyo. 425pp.

RICHMAN M.B. (1986)
Review article: rotation of Principal Component Analysis, *Journal of Climatology*, **6**, 293-335.

RINNE J., KARHILA V. (1975)
A spectral barotropic model in horizontal empirical orthogonal functions. *Quarterly Journal of the Royal Meteorological Society*, **101**, 365-382.

RINNE J., KARHILA V. (1979)
Empirical functions of 500mb height in the northern hemisphere determined from a large data sample. *Quarterly Journal of the Royal Meteorological Society*, **105**, 873-884.

ROYER J.J., VIERA P.C. (1984)
Dual formalism or kriging. G. Verly et all (Ed). *Geostatistics for natural resources characterization. Part 2.* D Reidel Publishing Company. 691-702.

SCHUBERT S.D. (1985)
A statistical dynamical study of empirically determined modes of atmospheric variability.*Journal of the Atmospheric Sciences*, **42**, 3-17.

SERVAIN J., LEGLER D. (1986)
Empirical orthogonal function analysis of tropical atlantic sea surface temperature and wind stress : 1964-1979. *Journal of Geophysical Research* , **91** (C12), 14-181, 14-191.

SERVAIN J., PICAUT J., BUSALACHI A.J. (1985)
Interanual and seasonal variability of the tropical Atlantic ocean depicted by 16 years of sea surface temperature and wind stress. *Coupled ocean atmosphere models*, J.J. Nihou Ed, Elsevier Sciences Publisher, 211-237.

SERVAIN J., SEVA M., LUKAS S., ROUGIER G. (1987)
Climatic atlas of the tropical Atlantic wind-stress and sea surface temperature : 1980-1984. *Ocean-Air interactions*, Vol. 1, Gordon and Breach Science Publishers, Inc, 109-182.

SHINUZOKA M. (1971)
Simulation of multivariate and multidimensional random processes. *The Journal of the Acoustical Society or America*, **49** (1), 357-368.

SHINUZOKA M., JAN C.M. (1972)
Digital simulation of random processes and its applications. *Journal of Sound Vibration* **25** (1), 111-128.

SHUMAN F.G. (1957)
Numerical methods in weather prediction. II Smoothing and filtering. *Monthly Weather Review*, **85**, 357-361.

SIMMONS A.J., WALLACE J.M., BRANDSTATOR G.W. (1983)
Barotropic wave propagation and instability in atmospheric teleconnection patterns. *Journal of the Atmospheric Sciences*, **40** (6), 1363-1392.

SMITH G.L. (1987)
Relations between Empirical Orthogonal Functions and normal modes. Preprints of the 10th conference on probability and statistics in atmospheric sciences. October 16-8 1987. Edmonton Alta (Canada). American Meteorological Society. 208-211.

TABONY R.C. (1981)
A principal component and spectral analysis of european rainfall. *Journal of Climatology*, **1**, 283-294.

THIEBAUX H.J., PEDDER M.A. (1987)
Spatial objective analysis, with applications in atmospheric sciences. Academic Press. 299 pp.

TOMPSON A.F., ABABOU R., GELHAR L.W. (1989)
Implementation of the three dimensional Turning Bands Random Fields Generator. *Water Resources Research*, **25**(10), 2227-2243.

TOURASSE P. (1981)
Analyse spatiale et temporelle de précipitations et utilisation opérationnelle dans un système de prévison des crues. Application aux régions cévenoles. Thèse de docteur-ingénieur, USTMG-INPG. Grenoble, 189 pp.

VON STORCH H., HANNOSCHOCK G. (1985)
Statistical aspects of estimated principal vectors (EOFs) based on small samples sizes. *Journal of Climate and Applied Meteorology*, **24**, 716-724.

WALLACE J.M. GUTZLER D.S (1981)
Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, **109**, 784-812.

ANNEXE A : OBTENTION DES FONCTIONS PROPRES ET DES COMPOSANTES PRINCIPALES PAR MINIMISATION D'UNE FONCTIONNELLE

On cherche donc les $Z_k(\xi)$ et les $F_k(\underline{x})$ minimisant pour $K=1,2....,\infty$ la fonctionnelle définie au chapitre 1:

$$A_K = E\left[\int_D \{ X(\xi,\underline{x}) - \sum_{k=1}^{K} Z_k(\xi).F_k(\underline{x}) \}^2.d\underline{x}\right] \qquad \text{(A-1)}$$

avec la condition de normalisation (A-2) des $F_k(\underline{x})$ qui assure l'unicité de la solution:

$$\int_D F_k^2(\underline{x}).d\underline{x} = 1 \qquad \text{(A-2)}$$

La démonstration sera faite par récurrence, en reprenant une approche de type variationnel proposée par Holmstöm (1963,1977) et utilisant le théorème général suivant (Bass,1978).

Théorème: Si $\delta f(\underline{x})$ est un accroissement élémentaire d'une fonction f de la variable $\underline{x} \in R^n$, et D un domaine fixé de R^n
si la fonctionnelle $F(\underline{x},f(\underline{x}))$ est telle que:

$$\forall\ \delta f(\underline{x}) \quad \int_D F(\underline{x},f(\underline{x})) \,.\, \delta f(\underline{x}).\, d\underline{x} = 0 \text{ alors } F(\underline{x},f(\underline{x})) = 0 \qquad \text{(A-3)}$$

Nous étendrons ce théorème au cas où $\int_D$ est remplacée par l'espérance mathématique E (qui peut d'ailleurs s'écrire sous forme d'intégrale) et nous supposerons que nous pouvons échanger les signes E et $\int_D$. Enfin, pour alléger les notations, nous écrirons X au lieu de $X(\xi,\underline{x})$; Z_k au lieu de $Z_k(\xi)$ et F_k pour $F_k(\underline{x})$.

Si $\forall\ m<l$, Z_m et F_m satisfont à (A-1) et (A-2), nous allons montrer, par récurrence, pour K=1,2,... la propriété (P) suivante:

$$(P) \quad \begin{cases} \int_D X \,.\, F_l \,.\, d\underline{x} = Z_l & \text{(A-4)} \\ E[\, X \,.\, Z_l \,] = E[\, Z_l^2 \,] \,.\, F_l = \lambda_l \,.\, F_l & \text{(A-5)} \\ \forall\ m<l \quad \int_D F_m \,.\, F_l \,.\, d\underline{x} = \delta_{lm} & \text{(A-6)} \\ \forall\ m<l \quad E[\, Z_m \,.\, Z_l \,] = E[\, Z_l^2 \,] \,.\, \delta_{lm} = \delta_{lm} \,.\, \lambda_l & \text{(A-7)} \end{cases}$$

<u>K =1</u> : On cherche Z_1 et F_1 minimisant A_1

$$A_1 = E[\int_D (X - Z_1 . F_1)^2 . d\underline{x}]$$

En faisant varier de façon élémentaire Z_1 et F_1 , on obtient:

$$E[-2 \int_D (X - Z_1 . F_1) . F_1 . d\underline{x} \ \delta Z_1] = 0 \qquad \forall \delta Z_1$$

$$\int_D -2 \ E[(X - Z_1 . F_1) . Z_1] \delta F_1 \ . d\underline{x} = 0 \qquad \forall \delta F_1$$

L'utilisation conjointe de (A-3) et (A-2) conduit alors à:

$$\int_D X . F_1 . d\underline{x} = Z_1 \qquad \text{(A-8)}$$

$$E[X . Z_1] = E[Z_1^2] . F_1 \qquad \text{(A-9)}$$

La propriété (P) est donc vraie pour K=1.

<u>K=2</u> : (P) est supposée vraie au rang 1. Z_1 et F_1 sont alors définis par les relations (A-8) et (A-9) ci-dessus. On cherche alors Z_2 et F_2 minimisant:

$$A_2 = E[\int_D (X - Z_1 . F_1 - Z_2 . F_2)^2 . d\underline{x}]$$

Une variation par rapport à Z_2 et F_2 conduit à :

$$E[-2 \int_D (X - Z_1 . F_1 - Z_2 . F_2) . F_2 . d\underline{x} \ \delta Z_2] = 0 \qquad \forall \delta Z_2$$

$$\int_D -2 \ E[(X - Z_1 . F_1 - Z_2 . F_2) . Z_2] \delta F_2 \ . d\underline{x} = 0 \qquad \forall \delta F_2$$

soit en utilisant (A-3) et (A-2):

$$\int_D X . F_2 . d\underline{x} = Z_1 . \int_D F_1 . F_2 . d\underline{x} + Z_2 \qquad \text{(A-10)}$$

$$E[X . Z_2] = E[Z_1 . Z_2] \ . F_1 + E[Z_2^2] . F_2 \qquad \text{(A-11)}$$

Pour montrer l'orthogonalité de F_1 et F_2 , on multiplie (A-8) par Z_2 et (A-11) par F_1, ce qui donne:

$$\int_D X . Z_2 . F_1 . d\underline{x} = Z_1 . Z_2$$

$$E[X . Z_2] . F_1 = E[Z_1 . Z_2] . F_1^2 + E[Z_2^2] .F_1 . F_2$$

On prend ensuite l'espérance de la première et on intègre la seconde sur D.En utilisant aussi la condition de normalisation de F_1, on obtient:

$$\int_D E [X . Z_2] . F_1 . d\underline{x} = E[Z_1 . Z_2]$$

$$\int_D E[X . Z_2] . F_1 . d\underline{x} = E[Z_1 . Z_2] + E[Z_2^2] . \int_D F_1 . F_2 . d\underline{x}$$

En faisant la différence entre ces deux équations, il vient nécessairement:

$$\int_D F_1 . F_2 . d\underline{x} = 0.$$

De la même façon, en multipliant (A-9) par F_2 et (A-10) par Z_1, il vient:

$$E[X . Z_1] . F_2 = E[Z_1^2] . F_1 . F_2$$

$$\int_D X . Z_1 . F_2 .d\underline{x} = Z_1^2 . \int_D F_1 . F_2 . d\underline{x} + Z_1 . Z_2$$

L'intégration de la première sur D et l'espérance sur la seconde conduisent à:

$$\int_D E[X . Z_1] . F_2 .d\underline{x} = E[Z_1^2] . \int_D F_1 . F_2 . d\underline{x}$$

$$\int_D E[X . Z_1] . F_2 .d\underline{x} = E[Z_1^2] . \int_D F_1 . F_2 . d\underline{x} + E[Z_1 . Z_2]$$

ce qui montre l'orthogonalité de Z_1 et Z_2, puisque, par différence, on obtient: $E[Z_1 . Z_2] = 0$

(P) est donc vérifiée au rang 2.

<u>K = k+1</u> : Supposons maintenant Ⓟ vérifiée jusqu'au rang k, montrons qu'elle est vraie au rang k+1.

$$A_{k+1} = E\left[\int_D \{ X(\xi,\underline{x}) - \sum_{m=1}^{k+1} Z_m \cdot F_m \}^2 . d\underline{x} \right]$$

En prenant la variation par rapport à Z_{k+1} et F_{k+1} puis en utilisant (A-3) et (A-2), on aura:

$$\int_D X \cdot F_{k+1} \cdot d\underline{x} = \sum_{m=1}^{k} Z_m \cdot \int_D F_m \cdot F_{k+1} \cdot d\underline{x} + Z_{k+1} \qquad \text{(A-12)}$$

$$E[\, X \cdot Z_{k+1} \,] = \sum_{m=1}^{k} E[\, Z_m \cdot Z_{k+1} \,] \cdot F_m + E[\, Z_{k+1}^2] \cdot F_{k+1} \qquad \text{(A-13)}$$

Montrons maintenant l'orthogonalité de F_{k+1} et F_l pour l<k+1. Pour cela, comme Ⓟ est vraie jusqu'au rang k, pour tout l<k, on peut multiplier (A-4) par Z_{k+1} et prendre l'espérance. De même, on multiplie (A-13) par F_l et on intègre sur D. D'où:

$$\int_D E[\, X \cdot Z_{k+1} \,] \cdot F_l \cdot d\underline{x} = E[\, Z_l \cdot Z_{K+1} \,] \qquad \forall\, l<k+1$$

$$\int_D E[\, X \cdot Z_{k+1} \,] . F_l \cdot d\underline{x} = \underbrace{\sum_{m=1}^{k} E[Z_m \cdot Z_{k+1}] . \int_D F_m \cdot F_l \cdot d\underline{x}}_{= E[Z_l \cdot Z_{k+1}] \text{ en utilisant (A-6)}} + E[\, Z_{k+1}^2] . \int_D F_l \cdot F_{k+1} \cdot d\underline{x}$$

Donc $\forall\, l<k+1 \quad \int_D F_l \cdot F_{k+1} \cdot d\underline{x} = 0$.

Pour montrer l'orthogonalité de Z_{K+1} avec tout Z_l , l<k+1, on multiplie (A-5) par F_{k+1} et on intègre sur D puis on multiplie (A-12) par Z_l et on en prend l'espérance soit:

$$\int_D E[\, X \cdot Z_l \,] \cdot F_{k+1} \cdot d\underline{x} = E[\, Z_l^2] \cdot \int_D F_l \cdot F_{k+1} \cdot d\underline{x} \qquad \forall\, l<k+1$$

$$\int_D E[\, X \cdot Z_l \,] \cdot F_{k+1} \cdot d\underline{x} = \underbrace{\sum_{m=1}^{k} E[\, Z_l \cdot Z_m] \cdot \int_D F_m \cdot F_{k+1} \cdot d\underline{x}}_{= E[\, Z_l^2] . \int_D F_l \cdot F_{k+1} \cdot d\underline{x} \text{ en utilisant (A-7)}} + E[\, Z_l \cdot Z_{k+1} \,]$$

Donc $\forall\ l<k+1\ E[\ Z_l\ .\ Z_{k+1}\ ]\ =\ 0.$

La propriété Ⓟ est donc vraie au rang k+1 et par extension pour tout K.

Pour obtenir tous les résultats annoncés au chapitre 1 et en particulier les relations (1-5), il suffit de remplacer dans (A-5) l'expression de Z_l par (A-4) soit:

$$E[\ X(\underline{x})\ .\int_D X(\underline{x}')\ .\ F_l(\underline{x}')\ .\ d\underline{x}'\ ]\ =\ \lambda_l\ .\ F_l(\underline{x})$$

soit encore $\int_D C(\underline{x},\underline{x}')\ .\ F_l(\underline{x}')\ .\ d\underline{x}'\ =\ \lambda_l\ .\ F_l(\underline{x}).$

ANNEXE B: QUELQUES PROPRIETES DES FONCTIONS DE COVARIANCE

ANNEXE B1 : La fonction de covariance $C^*(\underline{x},\underline{x}')$ du processus interpolé est de type positif.(cf 1-3)

Nous avons:

$$C^*(\underline{x},\underline{x}') = \sum_{i=1}^{P} \sum_{j=1}^{P} C(\underline{x}_i,\underline{x}_j).e_i(\underline{x}).e_j(\underline{x}')$$

Une fonction $C(\underline{x},\underline{x}')$ est de type positif si elle vérifie la propriété suivante (Matheron, 1972-a; Christakos, 1984):

Propriété: $\forall M, \forall \alpha_1, \ldots\ldots, \alpha_M \in \mathbb{C}, \forall \underline{x}_1, \ldots\ldots \underline{x}_M \in R^2$

$$A = \sum_k \sum_l \alpha_k . C(\underline{x}_k,\underline{x}_l). \overline{\alpha_l} > 0 \qquad (\overline{\alpha_l} \text{ est le complexe conjugué})(B\text{-}1)$$

Formons $A^* = \sum_k \sum_l \alpha_k . C^*(\underline{x}_k,\underline{x}_l). \overline{\alpha_l}$

$$A^* = \sum_k \sum_l \alpha_k .\left\{ \sum_{i=1}^{P} \sum_{j=1}^{P} C(\underline{x}_i,\underline{x}_j).e_i(\underline{x}_k).e_j(\underline{x}_l)\right\} . \overline{\alpha_l}$$

$$A^* = \sum_{i=1}^{P} \sum_{j=1}^{P} \left\{\sum_k \alpha_k .e_i(\underline{x}_k)\right\} .C(\underline{x}_i,\underline{x}_j).\left\{\sum_l \overline{\alpha_l} . e_j(\underline{x}_l)\right\} .$$

Posons $\beta_i = \sum_k \alpha_k .e_i(\underline{x}_k)$

On a donc $A^* = \sum_{i=1}^{P} \sum_{j=1}^{P} \beta_i . C(\underline{x}_i,\underline{x}_j). \overline{\beta_j}$ qui est une quantité > 0

car $C(\underline{x},\underline{x}')$ est une fonction de covariance et vérifie la propriété pour M=P. $C^*(\underline{x},\underline{x}')$ est donc de type positif et vérifie la propriété (B-1).

ANNEXE B2 : La fonction de covariance $C_s(\underline{x},\underline{x}')$ du processus simulé est de type positif.(cf § 3-3)

La fonction de covariance $C_s(\underline{x},\underline{x}')$ du processus simulé est donnée par (3-10):

$$C_s(\underline{x},\underline{x}') = \sum_{k=1}^{K} \lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}') + \delta(\underline{x}-\underline{x}') \cdot \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x})^2 \qquad (3\text{-}10)$$

Montrons que $\forall$ k, $\lambda_k \cdot F_k(\underline{x}) \cdot F_k(\underline{x}')$ vérifie la propriété (B-1) (voir Annexe B1)

$$A_k = \sum_m \sum_l \alpha_m \cdot \lambda_k \cdot F_k(\underline{x}_m) \cdot F_k(\underline{x}_l) \cdot \overline{\alpha_l}$$

$$A_k = \sum_m \alpha_m \cdot \sqrt{\lambda_k} \cdot F_k(\underline{x}_m) \cdot \sum_l \overline{\alpha_l} \cdot \sqrt{\lambda_k} \cdot F_k(\underline{x}_l)$$

Posons $\quad \beta_k = \sum_m \alpha_m \cdot \sqrt{\lambda_k} \cdot F_k(\underline{x}_m)$

Alors $A_k = \beta_k \cdot \overline{\beta_k} = |\beta_k|^2 > 0$

Si on forme:

$$A = \sum_m \sum_l \alpha_m \cdot C_s(\underline{x}_m,\underline{x}_l) \cdot \overline{\alpha_l}$$

$$= \sum_{k=1}^{K} \sum_m \sum_l \alpha_m \cdot \lambda_k \cdot F_k(\underline{x}_m) \cdot F_k(\underline{x}_l) \cdot \overline{\alpha_l}$$

$$+ \sum_m \sum_l \alpha_m \delta(\underline{x}_m-\underline{x}_l) \cdot \overline{\alpha_l} \sum_{k=K+1}^{\infty} \lambda_k \cdot F_k(\underline{x}_m)^2$$

$$A = \sum_{k=1}^{K} A_k + \sum_m |\alpha_m|^2 \cdot \sum_{k+K+1}^{\infty} \lambda_k \cdot F_k(\underline{x}_m)^2$$

Tous les A_k sont positifs ou nuls. De même, le second terme est une somme de termes positifs ou nuls (les λ_k sont positives). Donc, la quantité A est positive ou nulle. La propriété (B-1) et donc vérifiée.

ANNEXE C : CALCUL PRATIQUE DE LA MATRICE E DES PRODUITS SCALAIRES DANS LE CAS D'UTILISATION DE FONCTIONS DE BASE LINEAIRES PAR FACETTE.

ANNEXE C1 : Cas d'une surface plane.

Nous reprenons les développements proposés par Obled et Creutin (1986), dans le cas de fonctions linéaires par facette (cf § 1-3-3 et Fig. 1-1-b)

Pour évaluer

$$E_{im} = \int_D e_i(\underline{x}).e_m(\underline{x}).d\underline{x} \qquad \text{(C-1)}$$

on procédera de la manière suivante.

Un point $\underline{x}_i$ appartient à un nombre de triangles variable NT(i). Considérons maintenant deux points du réseau $\underline{x}_i$ et $\underline{x}_m$. Quatre cas sont possibles: ils n'ont aucun triangle commun, un seul triangle commun (à la frontière du domaine), deux triangles communs ou tous les triangles communs ($\underline{x}_i = \underline{x}_m$).

Supposons $\underline{x}_i \neq \underline{x}_m$. Soit Δ_{imk} un triangle commun à ces deux points Calculons maintenant la contribution de Δ_{imk} à E_{im} (Fig. C-1-a).

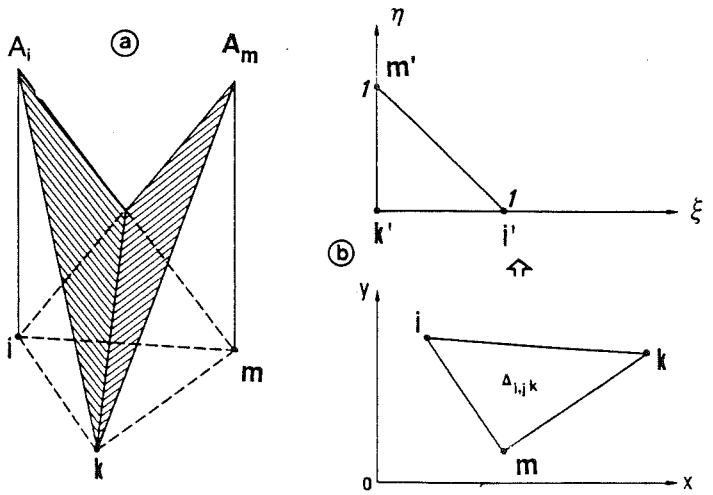


Fig. C-1: (a) Prismes mis en jeu dans le calcul de $\int_{\Delta imk} e_i(\underline{x}).e_m(\underline{x}).d\underline{x}$

(b) Transformation des coordonnées (x,y) aux coordonnées (η,ξ) (d'après Obled et Creutin, 1986).

Soit (x_i,y_i), (x_m,y_m), (x_k,y_k), les coordonnées des trois sommets de ce triangle.

$e_i(\underline{x})$ est définie par:

$$e_i(\underline{x}) = [\ (y_m - y_k)\ .\ (x - x_k) - (y - y_k)\ .\ (x_m - x_k)\ ]\ /\ \text{Den} \qquad \text{(C-2)}$$

avec $$Den = (x_i - x_k) . (y_m - y_k) - (x_m - x_k) . (y_i - y_k) \qquad (C\text{-}3)$$

De même $e_m(\underline{x})$ est donnée par:

$$e_m(\underline{x}) = [\ (y_i - y_k) . (x - x_k) - (y - y_k) . (x_i - x_k)\ ] / Den \qquad (C\text{-}4)$$

Pour calculer $\int_{\Delta imk} e_i(\underline{x}).e_m(\underline{x}).d\underline{x}$, on peut s'inspirer des techniques des éléments finis, utilisant le changement de coordonnées suivant (Fig. C-1-b)

$$\begin{cases} x - x_k = (x_i - x_k) . \eta + (x_m - x_k) . \xi \\ y - y_k = (y_i - y_k) . \eta + (y_m - y_k) . \xi \end{cases} \qquad (C\text{-}5)$$

η et ξ appartiennent à [0, 1] et le Jacobien de la transformation est $J=|Den|$ d'où:

$$\int_{\Delta imk} e_i(\underline{x}).e_m(\underline{x}).d\underline{x} = 1/Den^2 \int_{\Delta' imk} Den . \eta . \xi . |Den| .d\eta . d\xi$$

où $\Delta' i_{mk}$ est le triangle (1,0), (0,0), (0,1) dans le système de coordonnées (η,ξ) (Fig. C-1-b).

Finalement:

$$\int_{\Delta imk} e_i(\underline{x}).e_m(\underline{x}).d\underline{x} = |Den| / 24 = S_{imk} / 12 \qquad (C\text{-}6)$$

avec $|Den| = 2. S_{imk}$ où S_{imk} est la surface du triangle Δ_{imk}.

La contribution du triangle Δ_{imk} à E_{im} est donc S_{imk} / 12, qui est ajoutée à E_{im} et E_{mi}. En itérant sur tous les triangles communs à $\underline{x}_i$ et $\underline{x}_m$, on obtient la valeur finale de $E_{im} = E_{mi}$.

Si i=m, $$\int_{\Delta imk} e_i(\underline{x})^2 .d\underline{x} = S_{imk} / 6 \qquad (C\text{-}7)$$

En résumé, la procédure suivie sera la suivante:

i) Calcul de l'aire S_{imk} du triangle Δ_{imk} pour $i < m < k$.

ii) Ajout de S_{imk} / 6 aux termes diagonaux E_{ii}, E_{mm}, E_{kk}.

iii) Ajout de S_{imk} / 12 aux termes non diagonaux E_{im}, E_{mk}, E_{ik}.

La matrice **E** finale est alors obtenue en balayant ainsi, successivement, tous les triangles du réseau.

ANNEXE C2: Cas d'une surface sphérique.

Il s'agit toujours d'évaluer:

$$E_{im} = \int_D e_i(\underline{x}).e_m(\underline{x}).dx$$

mais, contrairement au cas précédent, le domaine n'est plus plan, mais sphérique.

Nous adopterons la même démarche qu'à l'Annexe C1 (balayage des triangles Δ_{imk} du réseau), mais il nous faut calculer $\int_{\Delta imk} e_i(\underline{x}).e_m(\underline{x}).d\underline{x}$ sur la sphère.

Nous nous contenterons en fait d'une approximation qui supposera des sommets suffisamment proches pour admettre que, localement, on puisse travailler dans un plan tangent au triangle. Dans ce cas, le raisonnement emprunté aux éléments finis, décrits en Annexe C1, peut être réutilisé. La contribution d'un triangle Δ_{imk} sera alors:

$S'_{imk} / 6$ pour E_{ii}, E_{mm}, E_{kk}

$S'_{imk} / 12$ pour E_{im}, E_{mk}, E_{ik}

où S'_{imk} est la surface du triangle projeté dans le plan tangent considéré. Son calcul s'avérant très compliqué, nous avons choisi de la remplacer par la surface réelle, sur la sphère du triangle ($\underline{x}_i$, $\underline{x}_m$, $\underline{x}_k$), donnée par Bass (1964):

$$S_{imk} = R_T^2 \cdot (\mathbf{I} + \mathbf{J} + \mathbf{K} - \Pi) \qquad \text{(C-8)}$$

où R_T est le rayon terrestre moyen et **I**, **M**, **K** sont les angles suivants (Fig. C-2):

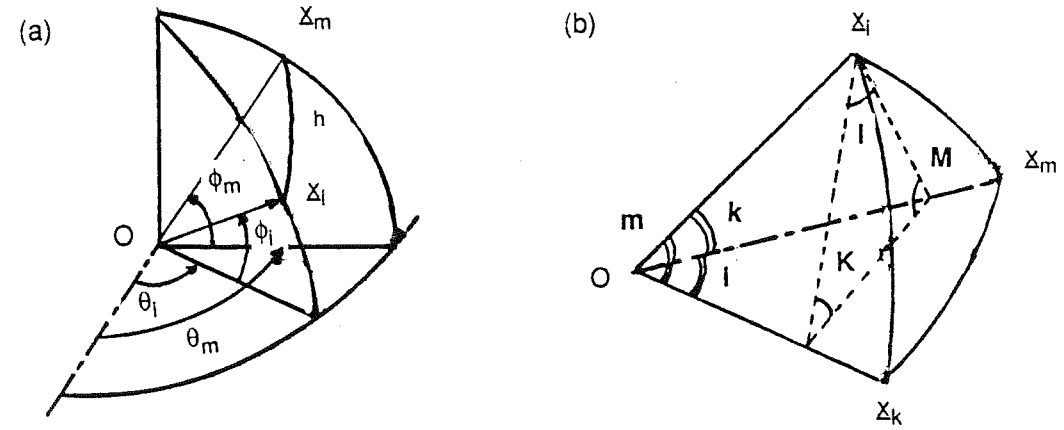


Fig. C-2: (a) Définition de la longitude et de la latitude (θ_i ,ϕ_i) d'un point $\underline{x}_i$

(b) Représentation des côtés (**i**, **m**, **k**) et des angles (**I**, **M**, **K**) d'un triangle sphérique qui sont respectivement les faces et les dièdres d'un trièdre.

I = angle entre les plans (O, $\underline{x}_i$, $\underline{x}_k$) et (O, $\underline{x}_i$, $\underline{x}_m$)

M = angle entre les plans (O, $\underline{x}_m$, $\underline{x}_i$) et (O, $\underline{x}_m$, $\underline{x}_k$)

K = angle entre les plans (O, $\underline{x}_k$, $\underline{x}_i$) et (O, $\underline{x}_k$, $\underline{x}_m$)

Si on définit par ailleurs, **i**, **m**, **k** comme respectivement, les angles entre les droites:

- $(O, \underline{x}_m)$ et $(O, \underline{x}_k)$
- $(O, \underline{x}_i)$ et $(O, \underline{x}_k)$
- $(O, \underline{x}_i)$ et $(O, \underline{x}_m)$

Abramowitz et Stegun (1970) donnent (formule 4-3-139):

$$\cos \mathbf{i} = \cos \mathbf{m} \,.\, \cos \mathbf{k} + \sin \mathbf{m} \,.\, \sin \mathbf{k} \,.\, \cos \mathbf{l} \qquad \text{(C-9)}$$

Si $(\underline{x}_i, \underline{x}_m, \underline{x}_k)$ ne sont pas alignés, on en déduit alors:

$$\begin{cases} \mathbf{l} = \text{Arcos} \{ [\cos \mathbf{i} - \cos \mathbf{m} \,.\, \cos \mathbf{k}] / (\sin \mathbf{m} \,.\, \sin \mathbf{k}) \} & \text{(C-10)} \\ \mathbf{M} = \text{Arcos} \{ [\cos \mathbf{m} - \cos \mathbf{k} \,.\, \cos \mathbf{i}] / (\sin \mathbf{k} \,.\, \sin \mathbf{i}) \} & \text{(C-11)} \\ \mathbf{K} = \text{Arcos} \{ [\cos \mathbf{k} - \cos \mathbf{i} \,.\, \cos \mathbf{m}] / (\sin \mathbf{i} \,.\, \sin \mathbf{m}) \} & \text{(C-12)} \end{cases}$$

Il faut, pour terminer, donner les expressions des angles **i**, **m**, **k** en fonction des latitudes et longitudes (θ_i, ϕ_i), (θ_m, ϕ_m) et (θ_k, ϕ_k) des trois sommets.

i est l'angles entre deux droites $(O, \underline{x}_m)$ et $(O, \underline{x}_k)$. Il est donc tel que:

$$\cos \mathbf{i} = \overrightarrow{O\,\underline{x}_m} \,.\, \overrightarrow{O\,\underline{x}_k} / \| \overrightarrow{O\,\underline{x}_m} \| \,.\, \| \overrightarrow{O\,\underline{x}_k} \|$$

Or le passage des coordonnées sphériques aux coordonnées cartésiennes se fait par:

$$\begin{cases} x = R_T \,.\, \cos\theta \,.\, \cos\phi \\ y = R_T \,.\, \sin\theta \,.\, \cos\phi \\ x = R_T \,.\, \sin\phi \end{cases}$$

D'où on en déduit facilement:

$$\begin{cases} \mathbf{i} = \text{Arcos} [\cos\phi_m \,.\, \cos\phi_k \,.\, \cos(\theta_m - \theta_k) + \sin\phi_m \,.\, \sin\phi_k] & \text{(C-13)} \\ \mathbf{m} = \text{Arcos} [\cos\phi_k \,.\, \cos\phi_i \,.\, \cos(\theta_k - \theta_i) + \sin\phi_k \,.\, \sin\phi_i] & \text{(C-14)} \\ \mathbf{k} = \text{Arcos} [\cos\phi_i \,.\, \cos\phi_m \,.\, \cos(\theta_i - \theta_m) + \sin\phi_i \,.\, \sin\phi_m] & \text{(C-15)} \end{cases}$$

Une remarque pour terminer ce paragraphe. Pourquoi ne pas avoir utilisé un calcul plus simple de surfaces planes dans le plan obtenu par projection Mercator? (voir Fig. 4-7 et Thiébaux et Pedder (1987) pour le rappel de la définition de cette projection) Un calcul indicatif pour le point (31°S, 17°W) montre que l'erreur relative commise serait alors de 10%, avec une surestimation de la surface avec la projection Mercator. Nous avons donc décidé de ne pas introduire ces erreurs supplémentaires.

ANNEXE D: OBTENTION DES FONCTIONS PROPRES ANALYTIQUES DU NOYAU DE COVARIANCE PROPOSE PAR FORTUS (1975)

Nous recherchons les solutions du problème suivant:

$$\begin{cases} 2\Pi.\ A\ .\ (\alpha^2 - \beta^2)\ \phi^a(r,\theta) = \lambda\ .\ (\Delta - \alpha^2)\ .\ (\Delta - \beta^2)\ \phi^a(r,\theta) & r \leqslant R \quad \text{(D-1)} \\ (\Delta - \alpha^2)\ .\ (\Delta - \beta^2)\ \phi^b(r,\theta) = 0 & r > R \quad \text{(D-2)} \\ F(r,\theta) = (\Delta - \alpha^2)\ .\ (\Delta - \beta^2)\ \phi(r,\theta) & \quad \text{(D-3)} \end{cases}$$

sous la forme $\phi^l(r,\theta) = \phi_m{}^l(r)\ .\ \exp(im\theta)$ $\qquad l=a,b$ $\qquad$ (D-4)

En polaires, le Laplacien s'écrit $\partial^2 / \partial r^2 + 1/r\ .\ \partial / \partial r + 1/r^2\ .\ \partial^2 / \partial\theta^2$

Si on appelle B_m, l'opérateur différentiel:

$$B_m = d^2/dr^2 + 1/r\ .\ d/dr - m^2/r^2$$

les équations (D-1) et (D-2) s'écrivent:

$$\begin{cases} 2\Pi.\ A\ .\ (\alpha^2 - \beta^2)\ \phi_m{}^a(r) = \lambda\ .\ (B_m - \alpha^2)\ .\ (B_m - \beta^2)\ \phi_m{}^a(r) & r \leqslant R \quad \text{(D-5)} \\ (B_m - \alpha^2)\ .\ (B_m - \beta^2)\ \phi_m{}^b(r) = 0 & r > R \quad \text{(D-6)} \end{cases}$$

ANNEXE D1: Rappels sur les fonctions de Bessel.

Si z est un nombre complexe , $C_m(z)$ désignera l'une quelconque des fonctions de Bessel suivantes: $J_m(z)$, $Y_m(z)$, $H_m{}^1(z)$, $H_m{}^2(z)$ (notation de Abramowitz et Stegun, 1970, qui sera abrégé par la suite par A-S, 1970).

D'après (A-S, 1970, formule 9-1-1), $C_m(z)$ vérifie l'équation différentielle:

$$[\ d^2/dz^2 + 1/z\ .\ d/dz + (1 - m^2/z^2)\ ]\ C_m(z) = 0$$

Si on considère une constante complexe μ et un réel r, il est aisé de vérifier que $C_m(\mu r)$ vérifie:

$$(\ B_m + \mu^2\ )\ C_m(\mu r) = 0 \qquad \text{(D-7)}$$

ANNEXE D2: Solution à l'intérieur du cercle:

En utilisant (D-7), on montre facilement que pour que $C_m(\mu r)$ vérifie (D-5), μ^2 doit vérifier:

$$2\Pi.\ A\ .\ (\alpha^2 - \beta^2) = \lambda\ .\ (\mu^2 + \alpha^2)\ .\ (\mu^2 + \beta^2) \qquad \text{(D-8)}$$

soit encore:

$$\mu^4 + (\alpha^2 + \beta^2) . \mu^2 + \alpha^2. \beta^2 - [2\Pi. A . (\alpha^2 - \beta^2)] / \lambda = 0 \qquad \text{(D-9)}$$

(D-9) est une équation du second degré en μ^2, dont le discriminant est:

$$\det = (\alpha^2 - \beta^2)^2 + [8\Pi. A . (\alpha^2 - \beta^2)] / \lambda$$

Fortus (1975) montre que, pour tout modèle de covariance stationnaire et isotrope, toutes les valeurs propres sont positives et que, par ailleurs, elles sont inférieures où égales à $(2\Pi)^2.S(w_o)$ où w_o est le maximum absolu de la densité spectrale S(w). On en déduit d'une part que det>0, c'est-à-dire que les deux racines μ_1^2 et μ_2^2 sont réelles et que, d'autre part, leur produit est négatif. En effet:

$$\mu_1^2 . \mu_2^2 = \alpha^2. \beta^2 - [2\Pi.A. (\alpha^2 - \beta^2)] / \lambda$$

$$= \{ \alpha^2. \beta^2 / [2\Pi. A . (\alpha^2 - \beta^2)] - 1/ \lambda\} . 2\Pi. A . (\alpha^2 - \beta^2)$$

$$= \{ 1/ ((2\Pi)^2.S(0)) - 1/\lambda\} . 2\Pi. A . (\alpha^2 - \beta^2) \qquad <0$$

puisque w=0 est le maximum absolu de S(w) et que $\lambda < (2\Pi)^2. S(0)$.

On posera $\begin{cases} \mu_1^2 = \chi^2 & \text{pour la racine positive} \\ \mu_2^2 = (i\kappa)^2 & \text{pour la racine négative.} \end{cases}$

Par ailleurs, (A-S, 1970), les fonctions de Bessel $J_m(\mu r)$ et $Y_m(\mu r)$ constituent deux solutions indépendantes de (D-7). On aura donc quatre solutions particulières pour (D-5): $J_m(\chi r)$, $J_m(i\kappa r)$, $Y_m(\chi r)$, $Y_m(i\kappa r)$. Or on veut que la solution $\phi_m^a(r)$ soit régulière en zéro, et comme $Y_m(\chi r)$ et $Y_m(i\kappa r)$ tendent vers l'infini quand r tend vers zéro (A-S, 1970), la forme de la solution à l'intérieur du cercle sera finalement:

$$\phi^a(r, \theta) = [C_1^a . J_m(\chi r) + C_2^a . I_m(\kappa r)] . \exp(im\theta) \qquad \text{(D-10)}$$

χ et κ vérifient (D-8), ce qui nous donne la relation (2-29) du § 2-2-3:

$$(\alpha^2 + \chi^2) . (\beta^2 + \chi^2) = (\alpha^2 - \kappa^2) . (\beta^2 - \kappa^2) \qquad \text{(D-11)}$$

que l'on peut encore écrire:

$$S(\chi) = S(i\kappa) \qquad \text{(D-12)}$$

(D-8) nous donne aussi les valeurs propres puisque:

$$\lambda = (2\Pi)^2 . S(\chi) = (2\Pi)^2 . S(i\kappa) \qquad \text{(D-13)}$$

ANNEXE D3: Solution à l'extérieur du cercle.

A l'aide de (D-7), on voit immédiatement que, pour que $C_m(\mu r)$ vérifie (D-5), il est nécessaire que:

$$(\alpha^2 + \mu^2) . (\beta^2 + \mu^2) = 0$$

ce qui donne deux solutions: $\mu_1{}^2 = (i\alpha)^2$ et $\mu_2{}^2 = (i\beta)^2$.

Par ailleurs, (A-S, 1970), $H_m{}^1(i\gamma r)$, $H_m{}^2(-i\gamma r)$ sont deux solutions indépendantes de (D-7), $\gamma = \alpha$ ou β; tendant vers zéro lorsque r tend vers l'infini.

De plus, (A-S, 1970, formule 9-6-4):

$$\begin{cases} K_m(x) = \Pi/2.\ i^{m+1} . H_m{}^1(ix) \\ K_m(x) = \Pi/2.\ i^{1-m} . H_m{}^2(ix) \qquad x \text{ réel.} \end{cases} \qquad \text{(D-14)}$$

Donc, on peut écrire la solution à l'extérieur du cercle sous la forme:

$$\phi_m{}^b(r, \theta) = [\ C_1{}^b . K_m(\alpha r) + C_2{}^b . K_m(\beta r)\ ] . \exp(im\theta) \qquad \text{(D-15)}$$

ANNEXE D4: Calcul des constantes $C_1{}^a$ et $C_2{}^a$.

i) Vérifions que les opérateurs différentiels suivants D_1 et D_2 satisfont: $D_1\ \phi^b = D_2\ \phi^b = 0$ en dehors du cercle.

$$\begin{cases} D_1 = \{\ \beta . K_{m-1}(\beta r) + K_m(\beta r)\ [\ d/dr + m/r\ ]\ \}\ (B_m - \alpha^2) & \text{(D-16)} \\ D_2 = \{\ \alpha . K_{m-1}(\alpha r) + K_m(\alpha r)\ [\ d/dr + m/r\ ]\ \}\ (B_m - \beta^2) & \text{(D-17)} \end{cases}$$

Démonstration:

Compte tenu de (D-14) et (D-7), si $\gamma = \alpha$ ou β, on aura:

$$(B_m - \gamma^2) K_m(\gamma r) = 0 \qquad \text{(D-18)}$$

Donc, $(B_m - \alpha^2) \phi_m{}^b(r) = C_1{}^b (B_m - \alpha^2) K_m(\alpha r) + C_2{}^b (B_m - \alpha^2) K_m(\beta r)$

$= C_2{}^b (B_m - \alpha^2) K_m(\beta r)$ d'après (D-18)

Par ailleurs, on sait (A-S, 1970, formule 9-2-26) que:

$$K_m'(z) = - K_{m-1}(z) - (m/z) . K_m(z)$$

soit pour $K_m(\gamma r)$:

$$1/\gamma . d/dr\, K_m(\gamma r) = - K_{m-1}(\gamma r) - (m/\gamma r) . K_m(\gamma r).$$

On aura donc $- \gamma . K_{m-1}(\gamma r) = (d/dr + m/r) K_m(\gamma r)$ (D-19)

Finalement, on a:

$$D_1\phi_m{}^b(r) = [\beta . K_{m-1}(\beta r) + K_m(\beta r) (d/dr + m/r)] C_2{}^b (\beta^2 - \alpha^2) K_m(\beta r)$$

$$= C_2{}^b (\beta^2 - \alpha^2) [\beta . K_{m-1}(\beta r) .K_m(\beta r) + K_m(\beta r) (d/dr + m/r) K_m(\beta r)]$$

$$= C_2{}^b (\beta^2 - \alpha^2) [\beta . K_{m-1}(\beta r) .K_m(\beta r) - \beta . K_m(\beta r) .K_{m-1}(\beta r)]$$

d'après (D-19).

Soit donc $D_1\phi_m{}^b(r) = 0$

On vérifie de la même façon que $D_2\phi_m{}^b(r) = 0$

ii) Calculons maintenant $D_1\phi_m{}^a(r)$ et $D_2\phi_m{}^a(r)$.

$$D_1\phi_m{}^a(r) = D_1 (C_1{}^a . J_m(\chi r) + C'_2{}^a . J_m(i\kappa r))$$

Or $(B_m - \alpha^2)J_m(\chi r) = - (\alpha^2 + \chi^2) J_m(\chi r)$

$(B_m - \alpha^2)J_m(i\kappa r) = (\kappa^2 - \alpha^2) J_m(i\kappa r)$ d'après (D-7)

$[d/dr + m/r] J_m(\chi r) = \chi . [J_m'(\chi r) + (m/\chi r) . J_m(\chi r)]$

D'après (A-S, 1970, formule 9-2-27):

$$J_m'(z) = J_{m-1}(z) - (m/z) \cdot J_m(z)$$

Donc: $[\, d/dr + m/r]\, J_m(\chi r) = \chi \cdot J_{m-1}(\chi r)$

et $[\, d/dr + m/r]\, J_m(i\kappa r) = i\kappa \cdot J_{m-1}(i\kappa r)$

$D_1\, \phi^a(R, \theta) = 0$ devient donc:

$$C_1^a \cdot [\, -(\chi^2 + \alpha^2)(\beta \cdot K_{m-1}(\beta R) \cdot J_m(\chi R) + \chi \cdot K_m(\beta R) \cdot J_{m-1}(\chi R)\,] +$$
$$C'^a_2 \cdot [\, (\kappa^2 - \alpha^2)(\beta \cdot K_{m-1}(\beta R) \cdot J_m(i\kappa R) + i\kappa \cdot K_m(\beta R) \cdot J_{m-1}(i\kappa R)\,] = 0$$

Par ailleurs, (A-S, 1970, formule 9-6-3), pour x réel:

$$J_m(ix) = i^m \cdot I_m(x) \qquad \text{(D-20)}$$

Si on pose
$$\begin{cases} C_2^a = -i^m \cdot C'^a_2 \\ d_1(\gamma, \chi) = [\, \gamma \cdot J_m(\chi R) \cdot K_{m-1}(\gamma R) + \chi \cdot J_{m-1}(\chi R) \cdot K_m(\gamma R)\,] / (\chi^2 + \gamma^2) & \text{(D-21)} \\ d_2(\gamma, \kappa) = [\, \gamma \cdot I_m(\kappa R) \cdot K_{m-1}(\gamma R) + \kappa \cdot I_{m-1}(\kappa R) \cdot K_m(\gamma R)\,] / (\kappa^2 - \gamma^2) & \text{(D-22)} \end{cases}$$

$D_1\, \phi^a(R, \theta) = 0$ devient :

$$d_1(\beta, \chi) \cdot C_1^a + d_2(\beta, \kappa) \cdot C_2^a = 0 \qquad \text{(D-23)}$$

en utilisant (D-11) pour simplifier l'équation.

En changeant α en β et réciproquement, on déduit la relation issue de l'égalité:

$D_2\, \phi^a(R, \theta) = 0$

$$d_1(\alpha, \chi) \cdot C_1^a + d_2(\alpha, \kappa) \cdot C_2^a = 0 \qquad \text{(D-24)}$$

ANNEXE D5: Calcul de la constante C_1^a.

On cherche C_1^a telle que $\int_0^{2\Pi} \int_0^R F^2(r,\theta).r.dr.d\theta = 1$ avec:

$$F(r,\theta) = C_1^a.(\alpha^2 + \chi^2).(\beta^2 + \chi^2).[\, J_m(\chi r) + d_1(\alpha, \chi) / d_2(\alpha, \kappa) \cdot I_m(\kappa r)\,].\cos(m\theta)$$
ou $\sin(m\theta)$

i) $\int_0^{2\Pi} \cos^2(m\theta).d\theta = 2\Pi$ si m=0 (D-25)

$\int_0^{2\Pi} \sin^2(m\theta).d\theta = \int_0^{2\Pi} \cos^2(m\theta).d\theta = \Pi$ si m≠0 (D-26)

ii) Calcul de $A_1(m) = \int_0^R J_m(\chi r) . J_m(i\kappa r) . r . dr$

A-S, 1970, formule 11-3-39, nous permet d'écrire que:

$$A_1(m) = R. \left[\gamma .J_{m+1}(\chi R) . J_m(i\kappa r) - i\kappa . J_m(\chi R) . J_{m+1}(i\kappa r)\right] / (\chi^2 + \kappa^2)$$

Or $J_{m+1}(z) = (2m/z) . J_m(z) - J_{m-1}(z)$ (A-S, 1970, formule 9-1-27)

Donc:

$$A_1(m) = (R/(\chi^2 + \kappa^2)). \{\gamma. [(2m/\chi R) .J_m(\chi R) -J_{m-1}(\chi R)] - J_m(i\kappa R)$$
$$-i\kappa . J_m(\chi R). [(2m/i\kappa R) .J_m(i\kappa R) -J_{m-1}(i\kappa R)] \}$$

$$A_1(m) = (R/(\chi^2 + \kappa^2)). [- \gamma. J_{m-1}(\chi R). J_m(i\kappa R) + i\kappa . J_m(\chi R). J_{m-1}(i\kappa R)]$$

$$A_1(m) = (Ri^m/(\chi^2 + \kappa^2)). [- \gamma. J_{m-1}(\chi R). I_m(\kappa R) + \kappa . J_m(\chi R). I_{m-1}(\kappa R)] \quad \text{(D-27)}$$

iii) Calcul de $A_2(m) = \int_0^R J_m^2(\chi r) . r . dr$

$$A_2(m) = \int_0^R J_m^2(\chi r) . r . dr = 1/\chi^2. \int_0^{\chi R} J_m^2(t) . t . dt = 1/\chi^2. B_2(m, \chi R)$$

<u>m=0</u> $B_2(0, \chi R) = (\chi^2R^2/2) . [J_0^2(\chi R) + J_1^2(\chi R)]$ (A-S, 1970, formule 11-3-34)

<u>m≠0</u> $B_2(m-1, z) - B_2(m+1, z) = 2m . J_m^2(z)$ (A-S, 1970, formule 11-3-33)

On en déduit les relations suivantes:

* si m est pair $B_2(m, z) = B_2(0, z) -2. \sum_{k=0}^{p-1} (2k+1) . J_{2k+1}^2(z)$ (D-28)

m=2p

* si m est impair $\quad B_2(m, z) = B_2(1, z) - 2. \sum_{k=0}^{p-1} 2(k+1) . J_{2(k+1)}^2(z)$ (D-29)

m=2p+1

$$B_2(1, z) = 2 \sum_{k=0}^{\infty} 2(k+1) . J_{2(k+1)}^2(z)$$ (A-S, 1970, formule 11-3-32)

On en déduit alors que:

* si m est pair $\quad A_2(m) = (R^2/2) . [J_0^2(\chi R) + J_1^2(\chi R)] - (2/\chi^2) \sum_{k=0}^{p-1} (2k+1) . J_{2k+1}^2(\chi R)$ (D-30)

m=2p

* si m est impair $\quad A_2(m) = - (2/\chi^2) \sum_{k=p}^{\infty} 2(k+1) . J_{2(k+1)}^2(\chi R)$ (D-31)

m=2p+1

iii) Calcul de $A_3(m) = \int_0^R J_m^2(i\kappa r) . r . dr$

$$A_2(m) = -1/ \kappa^2. \int_0^{i\kappa R} J_m^2(t) . t . dt = -1/ \kappa^2. B_2(m, i\kappa R)$$

En utilisant les relations (D-28), (D-29) et (D-20), on déduit que:

* si m est pair $\quad A_3(m) = (R^2/2) . [I_0^2(\kappa R) - I_1^2(\kappa R)] - (2/\kappa^2) \sum_{k=0}^{p-1} (2k+1) . I_{2k+1}^2(\kappa R)$ (D-32)

m=2p

* si m est impair $\quad A_3(m) = - (2/\kappa^2) \sum_{k=p}^{\infty} 2(k+1) . I_{2(k+1)}^2(\kappa R)$ (D-33)

m=2p+1

<u>Remarque</u>: En pratique, dans (D-31) et (D-33), les sommes convergent rapidement à 10^{-10} à et 10^{-5} près (une vingtaine de termes suffisent).

v) Valeur finale de C_1^a.

En regroupant les résultats précédents, on en déduit que:

* si m est pair $C_1^a = \Big\{ 2.\Pi . (\alpha^2 + \chi^2)^2 .(\beta^2 + \chi^2)^2.$

m=2p si m=0 uniquement

$$\Big\{ (R^2/2) . [J_0^2(\chi R) + J_1^2(\chi R)] - (2/\chi^2) \sum_{k=0}^{p-1} (2k+1) . J_{2k+1}^2(\chi R)$$

$$-2R.d_1(\alpha, \chi) /d_2(\alpha, \kappa) . [\gamma.J_{m-1}(\chi R) .I_m(\kappa R) - \kappa .J_m(\chi R) .I_{m-1}(\kappa R)]/(\chi^2 + \kappa^2)$$

$$+ d_1^2(\alpha, \chi) /d_2^2(\alpha, \kappa) [(R^2/2) . [I_0^2(\kappa R) - I_1^2(\kappa R)] - (2/\kappa^2) \sum_{k=0}^{p-1} (2k+1) . I_{2k+1}^2(\kappa R) \Big\} \Big\}^{-1/2}$$

* si m est impair $C_1^a = \Big\{ \Pi . (\alpha^2 + \chi^2)^2 .(\beta^2 + \chi^2)^2.$

m=2p+1

$$\Big\{- (2/\chi^2) \sum_{K=p}^{\infty} 2(k+1) . J_{2(k+1)}^2(\chi R)$$

$$-2R.d_1(\alpha, \chi) /d_2(\alpha, \kappa) . [\gamma.J_{m-1}(\chi R) .I_m(\kappa R) - \kappa .J_m(\chi R) .I_{m-1}(\kappa R)]/(\chi^2 + \kappa^2)$$

$$- (2/\kappa^2). d_1^2(\alpha, \chi) /d_2^2(\alpha, \kappa) - \sum_{k=p}^{\infty} 2(k+1) . I_{2(k+1)}^2(\kappa R) \Big\} \Big\}^{-1/2} \qquad \text{(D-35)}$$

ANNEXE E: OBTENTION DES FONCTIONS PROPRES ANALYTIQUES DU NOYAU DE COVARIANCE $C(r) = \gamma r \,.\, K_1(\gamma r)$.

Si on considère le modèle de covariance défini par:

$$C(r) = \gamma r \,.\, K_1(\gamma r) \qquad \text{(E-1)}$$

sa densité spectrale est (Mantoglou et Wilson, 1981):

$$S(w) = \gamma^2/\Pi \,.\, 1/\,(w^2 + \gamma^2)^2 \qquad \text{(E-2)}$$

Les polynômes $P(w^2)$ et $Q(w^2)$ sont:

$$\begin{cases} P(w^2) = \gamma^2/\Pi = \text{cste} \\ Q(w^2) = (w^2 + \gamma^2)^2 \end{cases}$$

On voit donc que, par rapport au cas précédent (modèle K_o de Fortus), on est, pour $Q(w^2)$, dans le cas particulier où $\alpha = \beta = \gamma$.

On peut obtenir la solution en partant de celle du cas précédent et en faisant tendre α vers β.

La solution pour $\phi^a(r, \theta)$ à l'intérieur du cercle a toujours la même forme:

$$\phi^a(r, \theta) = [\, C_1^a \,.\, J_m(\chi r) + C_2^a \,.\, I_m(\kappa r) \,] \,.\, \exp(im\theta) \qquad \text{(E-3)}$$

mais les valeurs de χ et κ vont changer.

Si $\alpha \neq \beta$, χ et κ vérifient:

$$\begin{cases} S(\chi) = S(i\kappa) & \text{(E-4)} \\ \begin{vmatrix} d_1(\alpha, \chi) & d_2(\alpha, \kappa) \\ d_1(\beta, \chi) & d_2(\beta, \kappa) \end{vmatrix} = 0 & \text{(E-5)} \end{cases}$$

où les fonctions d_1 et d_2 sont définies par les équations (2-36) et (2-37).

On peut transformer ce système en :

$$\begin{cases} S(\chi) = S(i\kappa) & \text{(E-6)} \\ \begin{vmatrix} d_1(\alpha, \chi) & d_2(\alpha, \kappa) \\ [\, d_1(\alpha, \chi) - d_1(\beta, \chi)\,]/(\alpha-\beta) & [\, d_2(\alpha, \kappa) - d_2(\beta, \kappa)\,]/(\alpha-\beta) \end{vmatrix} = 0 & \text{(E-7)} \end{cases}$$

En faisant tendre α et β vers γ, on en déduit que, pour les F.P. analytiques du modèle (E-1), χ et κ sont solutions de:

$$\begin{cases} S(\chi) = S(i\kappa) & \text{(E-6)} \\ \begin{vmatrix} d_1(\gamma, \chi) & d_2(\gamma, \kappa) \\ \partial/\partial\gamma\, d_1(\gamma, \chi) & \partial/\partial\gamma\, d_2(\gamma, \kappa) \end{vmatrix} = 0 & \text{(E-7)} \end{cases}$$

Par la suite, le calcul des constantes est analogue à ce qui a été décrit en Annexe D5.

ANNEXE F: NOTION D'ECHELLE INTEGRALE. CALCUL POUR QUELQUES MODELES DE COVARIANCE USUELS.

La notion d'échelle intégrale J, associée à un modèle de covariance homogène et isotrope C(r), est utilisée en turbulence. Muñoz-Pardo (1987): en rappelle la définition:

$$J = [\, 2/\sigma^2 \,.\, \int_0^\infty C(r)\,.r\,.\,dr\,]^{1/2}$$

Elle permet d'associer au modèle considéré une longueur de corrélation moyenne. Donc, si on impose l'égalité de cette échelle intégrale pour deux modèles, on pourra choisir les paramètres de façon à obtenir la même longueur de corrélation.

Ainsi, si on considère les modèles de covariance suivants:

sphérique: $\begin{cases} C(r) = \sigma^2 \,.\, [\, 1 - 3/2\,.\,(r/a_{sph}) + 1/2\,.(r/a_{sph})^3\,] & \text{si } r<a_{sph} \\ C(r) = 0 & \text{si } r>a_{sph} \end{cases}$

exponentiel $C(r) = \sigma^2 \,.\, \exp(-r/a_e)$

Fortus K_o $C(r) = \sigma^2/\mathrm{Ln}(\alpha/\beta)\ .\,[\, K_O(\,\beta.r\,) - K_O(\,\alpha.r)\,]$ $\alpha>\beta$

Gaussien $C(r) = \sigma^2 \,.\, \exp(\,-r^2/a_g^2)$

le calcul donne pour les différentes échelles intégrales:

$$J_{sph} = a_{sph}/\sqrt{5} \qquad \text{(F-1)}$$

$$J_e = \sqrt{2}\,.\,a_e \qquad \text{(F-2)}$$

$$J_{Ko} = 1/(\alpha/\beta)\,.\,[\, 2(\,\alpha^2 - \beta^2\,)/\mathrm{Ln}(\alpha/\beta)\,]^{1/2} \qquad \text{(F-3)}$$

$$J_g = a_g \qquad \text{(F-4)}$$

Démonstration de (F-3):

$$J^2_{Ko} = 2/Ln(\alpha/\beta)\ \{ \int_0^\infty K_O(\beta.r)rdr - \int_0^\infty K_O(\alpha.r)rdr \}$$

$t = \beta.r$ $\quad$ $t = \alpha.r$

$dt = \beta.dr$ $\quad$ $dt = \alpha.dr$

$$J^2_{Ko} = 2/Ln(\alpha/\beta)\ \{ 1/\beta^2 \int_0^\infty K_O(t)tdt - 1/\alpha^2 \cdot \int_0^\infty K_O(t)tdt \}$$

$$J^2_{Ko} = [\ 2/Ln(\alpha/\beta)\ ].\ [(\alpha^2 - \beta^2)/\alpha^2\beta^2]\ .\ \int_0^\infty K_O(t)tdt$$

Or d'après (Abramowitz et Stegun, 1970, formule 11-4-22), on a:

$$\int_0^\infty K_O(t)tdt = \Gamma(1)\ .\ \Gamma(1) = 1$$

Donc :

$$J_{Ko} = 1/(\alpha.\beta)\ .\ [\ 2(\alpha^2 - \beta^2)/Ln(\alpha/\beta)\ ]^{1/2}$$

Démonstration de (F-4):

$$J^2_g = 2\ .\int_0^\infty \exp(-r^2/a_g^2)\ .\ r\ .\ dr = [\ -a_g^2\ .\ \exp(-r^2/a_g^2)\ ] = a_g^2$$

soit $\quad J_g = a_g$

Si on pose l'égalité de (F-1), (F-2), (F-3), (F-4) on en déduit les relations suivantes pour la portée sphérique équivalente, respectivement aux modèles exponentiels,Ko de Fortus et gaussien:

$$\begin{cases} a_{sph} = a_e\ .\sqrt{10} & \text{(F-4)} \\ a_{sph} = \sqrt{5}/(\alpha.\beta)\ .\ [\ 2(\alpha^2 - \beta^2)/Ln(\alpha/\beta)\ ]^{1/2} & \text{(F-5)} \\ a_{sph} = a_g\ .\sqrt{5} & \text{(F-6)} \end{cases}$$

AUTORISATION de SOUTENANCE

VU les dispositions de l'Arrêté du 23 novembre 1988 relatif aux Etudes doctorales

VU les rapports de présentation de

- Monsieur M.DEQUE
- Monsieur PHAM DINH TUAN

Mademoiselle BRAUD Isabelle

est autorisé(e) à présenter une thèse en soutenance en vue de l'obtention du diplôme de DOCTEUR de L'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE, spécialité "Mécanique"

Fait à Grenoble, le 21 Mars 1990

Pour le Président de l'I.N.P.-G.
et par délégation,
le Vice-Président
P. VENNEREAU

Thèse de Doctorat de l'Institut National Polytechnique de Grenoble

Titre de l'ouvrage :

ETUDE MÉTHODOLOGIQUE DE L'ANALYSE EN COMPOSANTES PRINCIPALES DE PROCESSUS BIDIMENSIONNELS.
EFFETS DES APPROXIMATIONS NUMERIQUES ET DE L'ECHANTILLONNAGE ET UTILISATION POUR LA SIMULATION DE CHAMPS ALEATOIRES
Application au traitement des températures de surface de la mer sur l'Atlantique intertropical.

Nom de l'auteur :

BRAUD Isabelle

Etablissement :

Institut National Polytechnique de Grenoble
Institut de Mécanique de Grenoble

RESUME

Le point de départ de ce mémoire est la technique d'Analyse en Composantes Principales de Processus (ACPP), extension à des champs spatialement et/ou temporellement continus de l'Analyse en Composantes Principales (ACP) classique. Elle conduit à rechercher les fonctions propres d'un noyau de covariance, solutions d'une équation intégrale de Fredholm et permet ainsi de tenir compte explicitement de la forme et de la taille du domaine d'étude.
Dans le chapitre 1, après un rappel de la méthode et la présentation de l'approximation numérique utilisée lorsqu'on travaille avec un échantillon de taille finie et un processus discrétisé spatialement, on s'attache à montrer les liens de cette technique avec la recherche de modes normaux de systèmes dynamiques puis avec les techniques d'interpolation optimale et la géostatistique. On montre ainsi que l'ACPP peut s'interpréter comme un cas particulier de krigeage s'appuyant sur un modèle de covariance spécifique, non nécessairement stationnaire.
Dans le chapitre 2, afin de quantifier les erreurs sur l'estimation des fonctions propres, dues à l'échantillonnage statistique et à la discrétisation spatiale, une solution analytique, pour un modèle particulier de covariance et un domaine circulaire, est proposée puis comparée aux solutions numériques.
Le chapitre 3 pose ensuite les principes de l'utilisation de l'ACPP pour la simulation de champs bidimensionnels et analyse la sensibilité de la méthode au choix des paramètres en terme de restitution d'une covariance cible. La technique est aussi comparée à d'autres méthodes (Bandes tournantes, méthodes matricielles), puis utilisée pour la simulation de champs non stationnaires.
Enfin, le chapitre 4 illustre tous ces développements méthodologiques sur des données de température mensuelle de surface de la mer sur l'Atlantique intertropical entre 1964 et 1987.

MOTS CLES

Champs aléatoires ; Analyse en Composantes Principales ; Fonctions propres analytiques ; Approximations numériques ; Echantillonnage ; Erreurs d'estimation ; Géostatistique ; Simulation stochastique.