



HAL
open science

Validation et étalonnage d'images de télédétection à l'aide de mesures ponctuelles selon une approche géostatistique: application à la mesure des précipitations et de l'insolation

Jean-Dominique Creutin

► **To cite this version:**

Jean-Dominique Creutin. Validation et étalonnage d'images de télédétection à l'aide de mesures ponctuelles selon une approche géostatistique: application à la mesure des précipitations et de l'insolation. Météorologie. Université Scientifique et Médicale de Grenoble, 1987. Français. NNT: . tel-00686393

HAL Id: tel-00686393

<https://theses.hal.science/tel-00686393>

Submitted on 10 Apr 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.

THÈSE

présentée à

L'UNIVERSITÉ SCIENTIFIQUE, TECHNOLOGIQUE
ET MÉDICALE DE GRENOBLE

ET

L'INSTITUT NATIONAL POLYTECHNIQUE
DE GRENOBLE

pour obtenir le grade de

DOCTEUR ES-SCIENCES PHYSIQUES

par

Jean-Dominique CREUTIN

Ingénieur ENSG Nancy

S U J E T

**VALIDATION ET ÉTALONNAGE D'IMAGES DE
TÉLÉDÉTECTION A L'AIDE DE MESURES PONCTUELLES
SELON UNE APPROCHE GÉOSTATISTIQUE.
APPLICATIONS A LA MESURE
DES PRÉCIPITATIONS ET DE L'INSOLATION.**

Soutenu le 12 Mars 1987 devant la Commission d'Examen

M. A. BERROIR	Président
M. A. LEBRETON	
M. G. de MARSILY	Rapporteurs
M. C. OBLED	
M. J. CRUETTE	
M. J.P. DELHOMME	Examineurs
M. L. LLIBOUTRY	

AVANT-PROPOS

Ce mémoire fait la synthèse d'un ensemble de travaux menés au sein de l'équipe Hydrométéorologie de l'Institut de Mécanique de Grenoble dans le domaine de la télédétection de l'atmosphère. J'exprime mes remerciements les plus sincères à ceux qui ont pris part à ces travaux avec gentillesse et efficacité.

Mr. le Professeur A. BERROIR, Directeur de l'Institut National des Sciences de l'Univers, a accepté de présider le Jury de cette thèse. C'est pour moi, et aussi pour toute notre équipe, un honneur pour lequel je lui suis profondément reconnaissant.

Mr. G. De MARSILY, Professeur à l'Ecole des Mines de Paris, m'a guidé tout au long de mes recherches en jouant son rôle de "parrain CNRS" avec tact et lucidité. Je tiens à lui dire ici combien ses conseils m'ont été précieux et combien j'ai appris à son contact.

Mr. A. LEBRETON, Professeur à l'Université de Grenoble, a apporté à mon travail un peu de sa rigueur mathématique, bien nécessaire pour tempérer mon ardeur d'utilisateur trop soucieux du résultat. A l'inverse, Mr. J. CRUETTE, Directeur de Recherche à l'ORSTOM, Chargé de Mission au Ministère de la Recherche et de la Technologie, m'a invité à penser plus complètement aux possibilités d'application pratique des résultats obtenus. Tous deux ont donc contribué à élargir ma perception des problèmes : je les en remercie vivement.

Mr. J.P. DELHOMME, Ingénieur de Recherche de la Société Schlumberger, avait déjà largement défriché les possibilités d'application de la géostatistique à l'hydrologie, puis à la télédétection dans ses activités au Centre d'Informatique Géologique à Fontainebleau. Je voudrais qu'il sache le plaisir avec lequel j'ai toujours suivi sa lecture claire et pragmatique de la géostatistique. Je dois également le remercier pour sa disponibilité et sa gentillesse jamais démenties.

Mr. Charles OBLED, Professeur à l'Ecole d'Hydraulique de Grenoble, a dirigé mon travail pendant onze années. Il est difficile pour moi de mesurer l'intégrale des apports provenant des innombrables échanges que nous avons eus dans la plus agréable complicité. Je lui exprime ici ma plus sincère reconnaissance. Il serait injuste de ne pas associer à ces

remerciements, Mr. Philippe BOIS, Professeur à l'ENSHMG, qui fut souvent impliqué dans ces échanges, y apportant sa compétence et sa bonne humeur.

Enfin, Mr. L. LLIBOUTRY, Professeur à l'USTMG, a aimablement accepté de participer à ce jury. Je le remercie d'avoir ainsi apporté la hauteur de vue d'un géophysicien très averti des problèmes de télédétection dans les milieux naturels.

De manière plus large, il convient aussi de rappeler le rôle décisif joué par Mr. H.SAUVAGEOT, Professeur à l'Université de Toulouse, en tant qu'initiateur de nos études sur le radar météorologique. Il nous a mis en contact avec l'équipe radar de l'EERM, dirigée à l'époque par Mr. Marc GILET et ainsi nous a permis de prendre pied dans le domaine de la télédétection active. MM. Y. FLOCH, V. PIRCHER et J. TARDIEU ont été les artisans de la mise en forme des données utilisées.

De même, dans le domaine de la télédétection satellitaire, je tiens à mentionner les efforts consacrés par Mr. N. BERIOT, à l'époque adjoint au directeur du Centre de Météorologie Spatiale de Lannion, pour mettre à notre disposition les résultats du programme Cactus. Mr. C.PASTRE et Mme C. VILIEN ont également toujours facilité nos démarches auprès de la Météorologie Nationale pour cette étude.

Au sein de l'IMG, je n'oublie pas le soutien efficace que m'ont apporté Mme M. PIAU, Directeur du Laboratoire, Mr. G. VACHAUD, Responsable du Groupe Hydrologie, et de nombreuses personnes du laboratoire, en créant un environnement scientifique, technique et administratif très favorable.

Enfin, je souhaite remercier de manière toute particulière les chercheurs avec qui j'ai partagé de manière quotidienne les difficultés et les joies de ce travail - Guy DELRIEU, Sylvie GALLE et Philippe LACOMBA-, ainsi que Sylviane FABRY et Isabelle RICCIARDELLA à qui est due la forme de cette thèse.

SOMMAIRE

<u>INTRODUCTION</u> : LE PROBLEME ABORDE ET LES DONNEES DISPONIBLES	
- Les variables régionalisées et l'évolution récente des techniques de mesure	1
- Les mesures in-situ	2
- Les mesures par télédétection	3
- Exemple de la mesure des précipitations	4
- Objectifs de cette étude	9
- Fondements et apports de la méthodologie proposée	10
- Présentation des mesures utilisées	12
- Mesures de la pluviométrie dans le bassin parisien	13
- Mesures de la pluviométrie en Arabie Séoudite	16
- Mesures de la fraction d'insolation sur la France	18
<u>CHAPITRE I</u> : RAPPELS DE GEOSTATISTIQUE	23
- But du krigeage	23
- Estimateurs linéaires	24
- Fonctions aléatoires	26
- Système de krigeage	27
- Choix d'un modèle	31
- Inférence statistique dans le temps	33
- Inférence statistique dans l'espace	35
- Fonction aléatoire stationnaire	36
- Analyse objective	38
- Fonctions aléatoires intrinsèques d'ordre k	44
- Inférence des fonctions aléatoires intrinsèques d'ordre k	48
- Fonctions aléatoires intrinsèque d'ordre 0 : le variogramme	53
- Quasi-stationnarité : le krigeage en voisinage glissant	55
- Conclusion	57
Annexe I.A : Calcul de la covariance centrée des estimations de l'espérance ponctuelle d'une fonction aléatoire stationnaire	58
Annexe I.B : Krigeage à l'aide du variogramme	60
Annexe I.C : Annulation de facteurs de pondération et krigeage en voisinage glissant	62
<u>CHAPITRE II</u> : VALIDATION D'UN DISPOSITIF DE MESURES	65
1 - Choix des mesures de référence	65

2 - Choix des critères de validité	68
- Les "erreurs"	70
- La corrélation empirique	77
3 - Protocoles de validation retenus	81
- Cas des dispositifs de mesure non étalonnés	81
- Cas des dispositifs de mesure étalonnés	83
4 - Retour sur les critères de validité : critère théorique et critère expérimental	86
<u>CHAPITRE III</u> : EXEMPLES DE VALIDATIONS DE DISPOSITIFS DE MESURE NON ETALONNES	91
1 - Rappel	91
2 - Cas des pluies journalières dans le bassin parisien : validation du réseau de pluviomètres et du radar	93
a - Analyse de la structure statistique de la pluie au sol	94
b - Résultats de validation du réseau d'étalonnage	106
c - Résultats de validation du radar météorologique non étalonné	111
3 - Cas de la fraction journalière d'insolation en France : validation du réseau d'héliographes et de la méthode CACTUS	115
a - Analyse de la structure statistique de la fraction journalière d'insolation au sol	115
b - Résultats de validation du réseau d'étalonnage seul	129
c - Résultats de validation du modèle Cactus	133
<u>CHAPITRE IV</u> : LE COKRIGEAGE : RAPPELS THEORIQUES DANS LE CADRE DE L'ETALONNAGE D'IMAGES	137
- Buts du cokrigeage	137
- Système de cokrigeage	139
- Modèle stationnaire et corégionalisation	144
- Aspects nouveaux de l'inférence des moments d'ordre 1	145
- Aspects nouveaux de l'inférence des moments d'ordre 2	149
- Structure croisée et décalage spatial des variables	152

- Choix de modèles des moments d'ordre 2 : condition de positivité et modèle linéaire	157
- Assouplissement du modèle linéaire	161
- Choix d'un modèle pour l'étalonnage d'images	162
- Inférence empirique du modèle quasi-stationnaire retenu pour l'étalonnage en contexte spatial	166
- Proposition pour l'inférence automatique d'un modèle quasi-stationnaire pour l'étalonnage	167
- Conclusion	171
<u>CHAPITRE V</u> : EXEMPLES D'APPLICATION DU COKRIGEAGE : MESURE DE LA PLUVIOMETRIE EN ARABIE ET DE LA FRACTION JOURNALIERE D'INSOLATION EN FRANCE	173
1 - Mesure de la pluviométrie en Arabie	173
a - Généralités	173
b - Analyse de structure	174
c - Résultats	189
2 - Mesure de la fraction d'insolation en France	195
a - Analyse de structure	195
b - Résultats obtenus par cokrigeage	207
<u>CHAPITRE VI</u> : SIMPLIFICATIONS DU SYSTEME DE COKRIGEAGE DANS LE CADRE DU MODELE LINEAIRE	223
- Présentation générale	223
- Cokrigeage en voisinage glissant	225
- Quelques simplifications apportées par le modèle linéaire	227
- Corrélation intrinsèque et régression	228
- Décomposition valeur vraie-erreur de mesure et cokrigeage d'erreur	232
- Recalage d'une pré-estimation	234
- Fluctuations orthogonales autour d'une dérive commune	241
- Krigeage avec une dérive connue ou externe	245
- Eléments déterminant le choix d'une simplification	248
- Signification de quelques pratiques d'étalonnage d'images	249
Annexe VI.A : Cokrigeage et régression sous hypothèse de corrélation parfaite	256
Annexe VI.B : Krigeage de données de différentes qualités	263
Annexe VI.C : Cokrigeage de valeurs pré-krigées : une approche généralement sous- optimale	268

<u>CHAPITRE VII</u> : EXEMPLES D'APPLICATION DE SIMPLIFICATIONS DU SYSTEME DE COKRIGEAGE : MESURE DE LA PLUVIOMETRIE DANS LE BASSIN PARISIEN ET DE LA FRACTION JOURNALIERE D'INSOLATION EN FRANCE	273
1 - Mesure de la pluviométrie dans le Bassin Parisien : recalage des mesures radar consi- dérées comme pré-estimations de la pluvio- métrie	273
a - Généralités	273
b - Analyse de structure	275
c - Résultats	284
2 - Mesure de la fraction journalière d'insolation en France	291
a - Choix d'un étalonnage simplifié	291
b - Résultats obtenus	293
<u>CONCLUSION</u> : APPORTS DE CE TRAVAIL ET PERSPECTIVES OUVERTES	299
<u>BIBLIOGRAPHIE</u>	307

INTRODUCTION

LES PROBLEMES ABORDES ET LES DONNEES DISPONIBLES

LES VARIABLES REGIONALISEES ET L'EVOLUTION RECENTE DES TECHNIQUES DE MESURE.

La description quantitative de notre environnement naturel comporte une étape fondamentale, lourde et ingrate, de mesure de différentes grandeurs physiques dont la valeur varie dans l'espace et le temps.

Cette étape est fondamentale car elle sanctionne l'ensemble du travail de conceptualisation destiné à modéliser les phénomènes décrits : cycle hydrologique, circulation atmosphérique ou océanique par exemple. En effet, les modèles proposés ayant pour fonction de relier numériquement en tout point de l'espace et du temps les grandeurs physiques caractéristiques du phénomène, celles-ci devront être connues en chacun de ces points pour initialiser, forcer et finalement valider le fonctionnement du modèle. Cette étape est lourde puisqu'elle exige la répétition d'une opération physique plus ou moins complexe destinée à établir la mesure en un point, le contrôle de cette mesure, et, éventuellement, sa généralisation à d'autres points. Enfin, cette étape est ingrate car elle ne constitue pas un but en elle-même : la connaissance d'un champ pluviométrique, barométrique ou gravimétrique à un instant donné ne présente que peu d'intérêt.

L'évolution récente des technologies (développement de l'informatique, de l'électronique et des techniques de l'espace) a entraîné au cours des dernières décennies un bouleversement profond des méthodes de

mesure applicables aux Sciences de l'Environnement. Si l'on considère la dimension géographique comme une dimension privilégiée, il est possible de décrire ce bouleversement comme le passage de la mesure in-situ à la mesure de télédétection, en prenant en seule considération les positions géographiques relatives du dispositif de mesure et du point mesuré.

LES MESURES IN-SITU

Ces mesures nécessitent la mise en place d'un appareil de mesure au point géographique où la grandeur est mesurée.

La mesure in-situ est la plus classique car l'opération physique réalisée peut être très simple d'un point de vue technique : interception du flux de précipitations par la bague d'un pluviomètre ou utilisation du rayonnement solaire pour brûler la bande cartonnée d'un héliographe par exemple. La simplicité technique s'accompagne généralement d'une simplicité d'interprétation de ces mesures ; la valeur fournie par l'appareil est liée à la grandeur du phénomène de manière très directe (on parle de mesure directe) : la quantité d'eau recueillie par un pluviomètre est proportionnelle à l'intensité de pluie moyenne pendant un intervalle de temps donné, la longueur de la trace de brûlure sur une bande héliographique est proportionnelle à la durée d'insolation. Cette simplicité d'interprétation conduit le plus souvent à des erreurs de mesure faibles qui ont été bien étudiées.

Naturellement, le besoin d'accéder à des grandeurs physiques plus difficiles à mesurer (granulométrie des gouttes de pluie, rayonnement solaire direct ou global par exemple) ont entraîné une évolution des mesures in-situ vers des techniques plus sophistiquées : enregistrement de l'énergie fournie par l'impact des gouttes de pluie à l'aide d'un disdromètre ou bien de l'énergie fournie par le rayonnement solaire à l'aide de la cellule photoélectrique d'un pyranomètre ; la grandeur fournie par l'appareil n'est plus liée que de façon indirecte à celle de la variable mesurée (on parle de mesures indirectes) ; des hypothèses physiques plus lourdes sont nécessaires pour interpréter les mesures dont la précision peut devenir moins grande.

Par ailleurs, la mesure in-situ est ponctuelle au sens où le volume caractérisé par la valeur obtenue (volume d'échantillonnage) est petit par rapport à l'interdistance des appareils de mesure. Cette caractéristique peut constituer un avantage lorsque l'intérêt porte sur l'évolution d'un phénomène en un point précis de l'espace ; elle devient un inconvénient lorsque l'on souhaite apprécier l'évolution d'un phénomène dans l'espace. En effet, la connaissance de la grandeur mesurée en tous points d'une zone d'étude sera obtenue à l'aide de techniques d'interpolation s'appuyant sur un réseau d'appareils de mesure dont la densité devra être adaptée à la variabilité spatiale du phénomène. Aux erreurs commises ponctuellement par chaque appareil, s'ajoutent alors les erreurs d'interpolation. La gestion pratique de ces réseaux devient rapidement très lourde.

Enfin, la mesure in-situ, par la présence d'un appareil au point de mesure peut perturber le phénomène mesuré : un pluviomètre, par exemple, modifie localement le flux de précipitation en créant de petites turbulences qui conduisent à une sous-estimation des intensités.

LES MESURES PAR TELEDETECTION

Ces mesures sont effectuées à distance du lieu géographique où la grandeur est mesurée.

La télédétection est fondée sur les propriétés de transmission ou d'émission de rayonnements électromagnétiques de l'espace. Son développement est récent car l'opération physique nécessaire à la mesure s'appuie toujours sur une technologie complexe ; les dispositifs utilisés (radiomètres, systèmes radar) sont capables de capter et d'analyser des signaux de très faible puissance ; ces dispositifs sont souvent aéroportés ou placés en orbite spatiale afin de couvrir des domaines géographiques importants. Cette complexité rend la télédétection fragile : les sources de pannes sont importantes malgré les efforts mis dans le choix des composants et dans le suivi de ces appareils.

La mesure par télédétection est toujours indirecte ; les valeurs de réflectivité fournies par un radar météorologique ou de réflectance

fournies par un radiomètre sont reliées à des intensités de pluie ou des pourcentages d'ennuage à l'aide de relations souvent complexes sous un ensemble d'hypothèses très fortes qu'il est généralement impossible de vérifier. Les erreurs commises peuvent donc être importantes lorsque ces hypothèses sont mises en défaut.

Du fait de la distance séparant le capteur de télédétection du lieu mesuré, le signal enregistré caractérise toujours un volume d'espace important : il s'agit donc d'une mesure intégrée. En revanche la continuité d'analyse du signal et de balayage de la zone explorée permet la répétition de cette mesure de manière à ce que la juxtaposition des volumes d'échantillonnage (qui peuvent éventuellement chevaucher) assure une couverture continue. Les mesures par télédétection se présentent donc sous forme d'images dont les éléments ont la taille du volume échantillonné. Cette continuité présente un avantage déterminant pour apprécier l'évolution spatiale d'un phénomène puisqu'elle évite le recours à l'interpolation.

Notons enfin que la mesure par télédétection ne perturbe pas le phénomène étudié.

Cette distinction entre les mesures in-situ et par télédétection et les remarques qu'elle appelle s'applique à tous les domaines de la physique de l'environnement. Des capteurs spécifiques ont été développés pour répondre aux besoins des Sciences du Sol (programmes Landsat, ou Spot par exemple) et du Sous-sol (systèmes sismique ou gravimétrique par exemple), de l'Océanographie (infra-rouge thermique du programme Météosat ou projets Seasat et Topex/Poséidon par exemple) et de la Météorologie (programme Nimbus, Météosat ou techniques radar par exemple). Ces dispositifs de télédétection viennent compléter les techniques de mesures in-situ classiquement utilisées dans ces disciplines.

EXEMPLE DE LA MESURE DES PRECIPITATIONS

Le détail d'un exemple précis est utile pour placer les considérations générales qui précèdent dans un contexte pratique. Le cas de

la mesure des précipitations à l'aide d'un radar météorologique et de pluviomètres est à cet égard assez démonstratif.

Sans reprendre la description du fonctionnement d'un radar météorologique (les ouvrages de L.J. BATTAN, 1973 et H. SAUVAGEOT, 1982 décrivent la technique radar de manière très complète), nous rappellerons simplement que ce dispositif de télédétection active de l'atmosphère fonctionne comme tout radar en suivant un cycle d'émission et de réception:

- l'émission concentre un bref signal électromagnétique de puissance connue dans un faisceau aussi fin que possible ;
- la réception mesure la puissance du signal rétrodiffusé par les cibles contenues dans le faisceau.

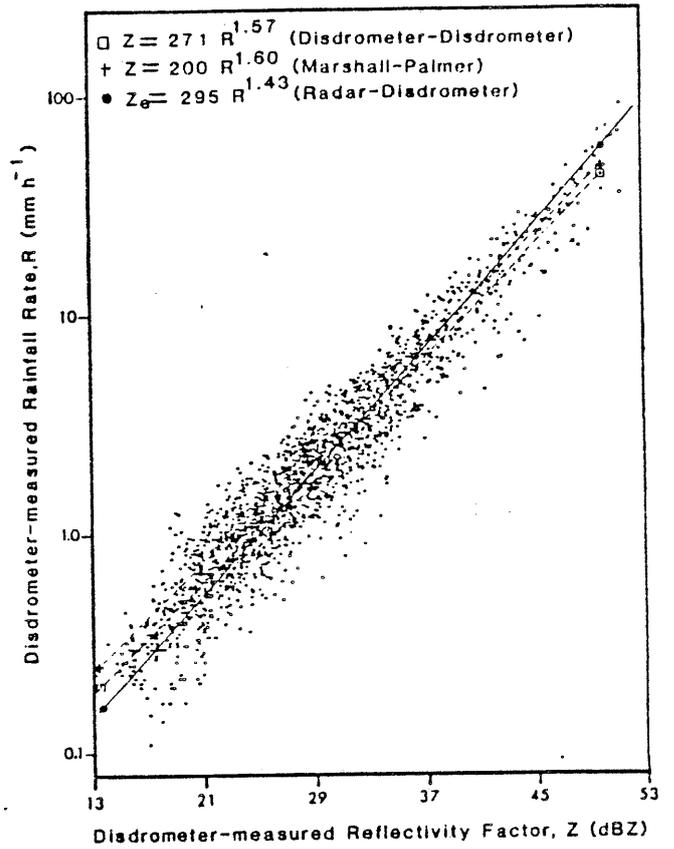
La durée écoulée entre l'émission et la réception du signal renseigne sur la distance à laquelle se trouve la cible ; cette information complétée par la donnée de l'orientation du faisceau permet de situer la cible dans l'espace. La mesure de la puissance reçue permet d'évaluer la réflectivité de la cible en utilisant l'équation du radar: cette équation relie linéairement la puissance reçue à la réflectivité en fonction des grandeurs caractéristiques du radar, de la cible et de l'atmosphère.

Le radar météorologique est simplement caractérisé par un choix de longueur d'onde approprié à la détection de l'eau liquide (ondes de 3 à 10 cm) dont la réflectivité est égale à la puissance sixième du diamètre des gouttes de pluie ; le faisceau radar éclairant généralement un ensemble de gouttes de diamètres variables, la réflectivité de cet ensemble se calcule par :

$$Z = \int_{D_{MIN}}^{D_{MAX}} N(D) D^6 dD \quad (\text{mm}^6/\text{m}^3) \quad (0.1)$$

(l'utilisation de la lettre Z pour la réflectivité, conforme à l'usage des radaristes, ne dépassera pas le cadre de cette introduction ; elle entrerait sinon en conflit avec la notation Z des variables aléatoires d'usage classique en géostatistique) où N(D) représente la distribution

Figure 0.1 : Diagramme des valeurs de réflectivité et d'intensité calculées d'après les mesures disdrométriques effectuées à Toronto pendant les étés 1977, 1978 et 1979. Chaque point concerne une période de 7 minutes.



D'après W. RICHARDS et C. CROZIER, 1983.

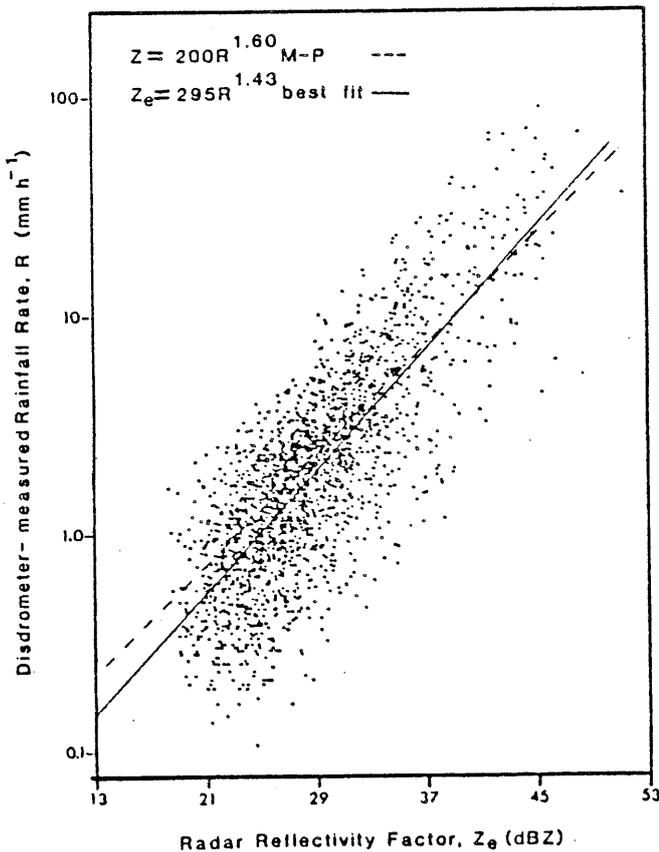


Figure 0.2 : Diagramme des réflectivités radar et des intensités calculées d'après des mesures disdrométriques lors de la même expérience.

granulométrique des gouttes entre les diamètres DMIN et DMAX des plus petites et des plus grosses gouttes.

Sous réserve de disposer d'un radar dont le fonctionnement électronique est stable et sous réserve de la vérification des hypothèses formulées concernant :

- l'état de l'atmosphère parcouru par les ondes,
- la nature liquide des précipitations,
- le remplissage homogène du faisceau radar par des précipitations aux caractéristiques homogènes,
- l'absence d'obstacle solide ou liquide interceptant une partie du faisceau ou l'atténuant,

nous sommes donc capables de mesurer une réflectivité moyenne de la pluie traversée par le faisceau. Il reste alors à interpréter ces mesures en terme d'intensité de pluie.

L'approche physique de cette interprétation consiste, dans un premier temps, à relier la réflectivité de la formule (0.1) à l'intensité de précipitation à l'intérieur du faisceau radar (notée R conformément à l'usage des radaristes) qui peut être exprimée sous la forme :

$$R = 3.6 \cdot 10^6 \int_{\text{DMIN}}^{\text{DMAX}} N(D) D^3 (V_t(D) - W) dD \quad (\text{mm/h}) \quad (0.2)$$

L'intensité est obtenue en sommant, pour chaque diamètre de gouttes, le flux d'eau liquide traversant une surface horizontale pendant une unité de temps ; on suppose que ces gouttes chutent à une vitesse $V_t(D)$, fonction de leur taille, diminuée éventuellement de la composante verticale W du mouvement de la masse d'air contenant ces gouttes. Il n'existe donc pas de relation univoque entre Z et R : une même intensité de pluie peut être calculée à l'aide de différentes distributions granulométriques et ces différentes distributions conduiront à des calculs de réflectivités différentes. Une illustration de cette dispersion est donnée, d'après W.RICHARDS et C. CROZIER (1983), en figure 0.1 : pour plus de 2000 périodes de 7 minutes, ces auteurs ont mesuré au sol des distributions granulométriques de gouttes de pluie à l'aide d'un disdromètre ; puis, ils

ont calculé par les formules (0.1) et (0.2) les réflectivités et les intensités correspondant à ces distributions. Il est aussi possible d'apprécier expérimentalement l'impact de la variabilité des distributions sur la dispersion de la relation intensité-réflectivité. Remarquons que le calcul des intensités a été conduit en faisant l'hypothèse d'un mouvement vertical nul de la masse d'air et en utilisant une relation simple liant le diamètre des gouttes à leur vitesse de chute. Il est clair que la connaissance simultanée de la réflectivité, de la distribution granulométrique ainsi que de la vitesse de chute des gouttes permettrait de résoudre physiquement ce problème d'interprétation. Malheureusement, l'espoir de disposer de telles mesures en tous points de l'image radar est lointain, même si les techniques doppler et de double polarisation sont prometteuses (R.J. DOVIAK, 1983, A.R. JAMESON, 1985). Il est donc exclu, dans le contexte technologique actuel, de traiter de manière déterministe l'interprétation d'une image radar en terme d'intensité pluvieuse.

Cette approche physique conforte cependant les résultats obtenus de manière empirique par de nombreux auteurs à la recherche d'une relation simple liant la réflectivité mesurée par un radar à l'intensité mesurée par des pluviomètres au sol. En effet, les nombreuses relations Z-R (dites relations de Marshall-Palmer, du nom des premiers auteurs d'une relation de ce type -J.S. MARSHALL et al. 1947, 1948) proposées au gré d'expériences diverses sont de la forme $Z = aR^b$ conformément aux résultats de W. RICHARDS et C. CROZIER.

Il faut alors bien voir qu'une approche fondée sur la comparaison de mesures radar et sol intègre un nombre important de nouvelles sources d'incertitudes dues aux transformations possibles de l'intensité de pluie entre le faisceau radar et le sol. Sans être exhaustifs (voir I. ZAWADZKI, 1984), nous citerons l'advection latérale et verticale des précipitations par le vent ou bien leur évaporation partielle. Une illustration de cet accroissement de la dispersion est donné en figure 0.2 par W. RICHARDS et C. CROZIER : l'intensité de précipitation est alors portée, pour la même période que sur la figure 0.1, en fonction de la réflectivité mesurée par un radar à la verticale du disdromètre. Il est clair que ces nouvelles sources d'incertitudes éloignent encore l'espoir d'une interprétation physique directe des mesures de réflectivité.

OBJECTIFS DE CETTE ETUDE

La disponibilité nouvelle des mesures de télédétection, dont nous venons de rappeler brièvement les caractéristiques essentielles, pose différentes questions pratiques aux utilisateurs potentiels de ces données:

i) - Quelle est la précision des mesures par télédétection ?

Il est difficile de répondre à cette question fondamentale, tout d'abord en raison de l'absence irrémédiable de référence absolue. Les mesures directes du phénomène étudié étant elles-mêmes entachées d'erreurs et présentant des caractéristiques différentes (volume d'échantillonnage, résolution spatiale et temporelle), il est en effet exclu de les prendre sans précaution comme référence. Une fois cette référence établie, il est ensuite difficile de caractériser l'erreur commise par le dispositif validé. De nombreux critères évaluent la distance entre la mesure validée et la référence de manières très différentes (erreurs moyennes, indices de co-fluctuation) : un choix est nécessaire. Enfin, quel que soit le critère de précision retenu, sa valeur ne prend un sens que par comparaison avec la précision obtenue à l'aide d'un autre dispositif de mesure ou bien avec la précision exigée par un utilisateur : un critère de précision ne peut être interprété que de manière relative.

ii) - Les mesures par télédétection peuvent-elles être corrigées à l'aide de mesures in-situ ?

Nous avons vu qu'une complémentarité existe entre ces deux types de mesures. La télédétection apporte en effet la finesse de résolution spatiale qui fait défaut aux mesures in-situ ; en revanche, la nature directe des mesures in-situ leur confère ponctuellement une précision plus grande. Il paraît donc séduisant de tirer parti de cette complémentarité afin de fournir une mesure combinant l'information des deux dispositifs.

iii) - Quelle est la précision des mesures de télédétection corrigées ?

Une fois retenu un mode de combinaison des mesures par télédétection et in-situ, il est nécessaire d'apprécier le gain de précision obtenu. Idéalement, ces mesures corrigées devraient être de meilleure qualité que celles obtenues individuellement par chaque dispositif.

Notre étude se fixe pour objectifs de proposer et d'appliquer une méthodologie afin de répondre à ces questions. Nous distinguerons dans cette proposition : i) la définition d'un protocole de validation (Chapitres I et II) destiné à caractériser la précision d'un dispositif de télédétection, éventuellement corrigé, et ii) la définition de méthodes d'étalonnage (Chapitres IV et VI) destinées à corriger les mesures de télédétection en les combinant à des mesures in-situ. Les exemples d'application (Chapitres III, V, VII) sont choisis dans le domaine météorologique : pluviométrie radar et satellite, mesure de l'insolation par satellite. Il est naturellement possible d'appliquer cette méthodologie à d'autres domaines.

L'intérêt d'apporter des réponses précises à ces questions est scientifique en raison de l'impact de l'incertitude des données d'entrée, de forçage ou de contrôles sur le fonctionnement et la validation des modèles numériques décrivant les phénomènes étudiés ; l'intérêt est aussi économique car le choix de remplacer, ou de compléter, un dispositif de mesure in-situ par un dispositif de télédétection devrait s'appuyer, au moins en partie, sur une comparaison entre les investissements réalisés et les gains de précision obtenus.

FONDEMENTS ET APPORTS DE LA METHODOLOGIE PROPOSEE

La comparaison et la combinaison des mesures de télédétection et in-situ à des fins de validation ou d'étalonnage se heurtent principalement au problème de l'hétérogénéité des caractéristiques de ces mesures.

L'hétérogénéité résulte, en premier lieu, de la différence de nature des variables mesurées (réflectivité radar et intensité de précipitations, radiance des masses nuageuses et durée d'insolation au sol par exemple). Nous utiliserons pour résoudre cette hétérogénéité une approche physique aussi complète que possible. Les variables mesurées par télédétection et nos variables d'intérêt mesurées in-situ possèdent généralement un terme commun (contenu en eau liquide de l'atmosphère pour les précipitations, nébulosité pour l'insolation) grâce auquel une analyse physique permet d'établir un lien quantitatif entre ces variables (voir l'exemple du radar). Il arrive que ce terme commun n'existe pas en raison de la simplicité des mesures de télédétection effectuées au regard de la complexité du phénomène (exemple de la pluviométrie à l'aide d'images satellitaires visible et infra-rouge) ; dans ce cas l'analyse physique sera limitée à un aspect qualitatif (reconnaissance des types de nuage a priori générateur de précipitations) ; une relation empirique sera ensuite établie en suivant une approche statistique pour lier quantitativement les observations par télédétection et in-situ sans préjuger de la physique réelle du phénomène (recherche d'une régression statistique entre la présence de tel type de nuage et la pluviosité au sol). Remarquons qu'il s'agit là d'une étape provisoire que les progrès des capteurs de télédétection permettra un jour d'éviter ; par exemple, la télédétection active par satellite (projet TRMM/BEST de la NASA et du CNES) permettra, par mesure de la réflectivité radar du sommet des nuages, de déduire un contenu en eau que des modèles microphysiques devraient permettre de lier aux précipitations au sol.

L'hétérogénéité résulte également de la différence des modes d'échantillonnage des dispositifs de mesures. Ces dispositifs opèrent, sur les variables que nous souhaitons comparer et combiner, une intégration dans l'espace ou le temps qui leur est propre. Etant donné l'absence de modèle physique assez fin pour décrire les fluctuations spatiales et temporelles de ces variables nous les assimilerons à un processus aléatoire en suivant ainsi une approche statistique dont le cadre mathématique est la Théorie des Fonctions Aléatoires.

En résumé, et pour situer les apports de cette étude en terme de méthode, nous ferons deux remarques :

i) nous avons considéré que les travaux importants consacrés au traitement du problème de l'hétérogénéité de nature des variables sont suffisants pour constituer une base de travail fiable. Nous utiliserons donc les méthodes proposées dans la littérature en les rappelant brièvement et en renvoyant le lecteur, soucieux de mieux les appréhender, à des travaux menés dans notre laboratoire sur des exemples précis (P.LACOMBA, 1986 ; G. DELRIEU, 1986 et S. GALLE, 1987) ou bien à des références classiques.

ii) il nous a semblé, par contre, que la question de l'hétérogénéité d'échantillonnage méritait d'être prise en compte de manière plus consistante qu'elle ne l'est généralement dans la littérature. Nous avons, pour cela, choisi d'utiliser un ensemble de méthodes statistiques connues sous les termes de Géostatistique et d'Analyse Objective. Notre travail a consisté, pour l'essentiel, à placer ces méthodes dans un cadre cohérent afin de satisfaire nos objectifs de validation et d'étalonnage de mesures télédétectées. Cette démarche nous a permis, parallèlement, de situer dans ce cadre les techniques classiques d'étalonnage de sorte à identifier de manière précise leurs hypothèses sous-jacentes, souvent mal perçues.

PRESENTATION DES MESURES UTILISEES

Une méthodologie n'ayant de valeur que si elle est applicable et efficace, il était important de la mettre en pratique sur différents ensembles de mesures, réalisés simultanément par un dispositif de télédétection et un réseau d'appareils au sol. Un effort cohérent de l'équipe Hydrométéorologie de l'IMG a permis de réunir de tels ensembles de mesures à l'occasion de différentes études menées entre 1981 et 1986 en collaboration avec d'autres organismes (Météorologie Nationale, Bureau de Recherches Géologiques et Minières). Deux ensembles concernent les précipitations et comportent :

- dans l'Est du Bassin Parisien un réseau de pluviomètres et des images de radar météorologique.

- dans le Nord de l'Arabie Séoudite un réseau de pluviomètre et des images de satellites météorologiques géostationnaire et défilant ;

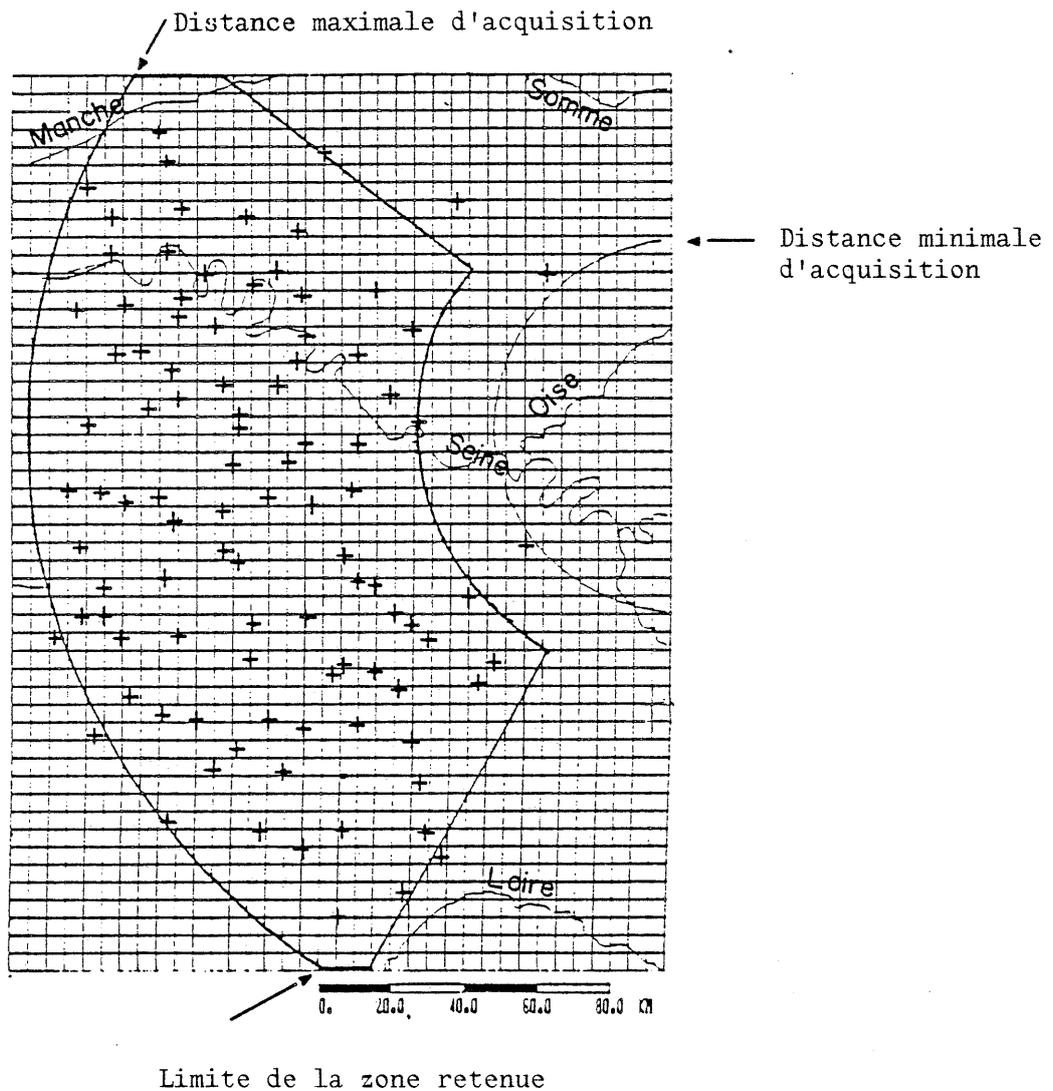
Le troisième ensemble concerne la mesure de la fraction journalière d'insolation sur la France à l'aide d'un réseau d'héliographes et des images du satellite Météosat.

MESURES DE LA PLUVIOMETRIE DANS LE BASSIN PARISIEN

Les mesures de télédétection sont issues d'une expérience pilote menée par la Météorologie Nationale pendant les mois de Mai 1980 à Octobre 1982 sous la responsabilité de M. GILET (responsable du Centre de Recherches en Physique de l'Atmosphère de l'EERM, aujourd'hui directeur adjoint du CTM de TRAPPES) et Y. FLOCH (responsable de l'installation radar de DAMMARTIN en GOELLE, aujourd'hui responsable du radar de BREST). Le point clé de cette expérience était l'installation sur le site du radar d'un dispositif électronique capable de numériser et d'enregistrer le signal radar afin de rendre les images radar propres à une utilisation informatique. Un premier projet de même type mais de durée plus courte (projet HYDROMEL du Printemps 1976, voir G. FROMENT, 1981) avait échoué en raison d'une pluviométrie exceptionnellement faible. On peut donc considérer qu'il s'agit du premier ensemble de mesures de ce type disponible en France.

i) Le détail des caractéristiques techniques du radar MELODI installé à DAMMARTIN en GOELLE se trouve dans M. GILET et al., 1982. Nous rappellerons simplement qu'il s'agit d'un radar météorologique de longueur d'onde 10.7 cm dont l'antenne de 4 mètres permet de concentrer la moitié des 700 kW de puissance émise dans un faisceau de 1.8°. Il s'agit donc d'un radar de forte puissance, peu sensible à l'effet d'atténuation, donc a priori bien adapté pour la mesure de fortes intensités. Sa situation géographique, sur une colline au Nord-Est de l'aéroport de ROISSY, pose deux problèmes importants : l'altitude de l'antenne permet aux lobes secondaires de produire des échos de sol assez éloignés du radar et la présence de l'agglomération parisienne dans un rayon de 50 km au Sud-Ouest

Figure 0.3 : Carte de la zone étudiée rappelant l'emplacement des 98 stations pluviométriques ainsi que la grille de discrétisation de l'information radar.



constelle l'image d'échos fixes au site sélectionné de 0.7° . Ces contraintes ont conduit à l'abandon des mesures effectuées entre 0 et 52km du radar.

ii) La chaîne électronique d'acquisition a été développée par la Météorologie Nationale sous la dénomination SAPHYR (M. GILET et al., 1980). Elle nous a permis d'obtenir des mesures numérisées sur 256 niveaux de réflectivité compris entre 0 et 63 dBz ; ces mesures sont réparties selon une grille sphérique comportant 180 rayons sur lesquels s'échelonnent 256 points entre 52 et 180 km. Nous avons développé à l'I.M.G. des programmes de conversion, de cumul et de redistribution suivant une grille rectangulaire de ces mesures permettant de disposer par exemple d'images de pluviométrie (calculée par une loi de Marshall-Palmer classique $Z=200R^{1.6}$) au pas de temps journalier (c'est à dire après cumul de 288 images enregistrées toutes les 5 minutes) et suivant un maillage carré de 5 km de côté ; le détail des algorithmes contenus dans ces programmes réalisés par J.D. CREUTIN et G. DELRIEU sont rappelés par G.DELRIEU, 1986.

Les mesures sol proviennent des stations du réseau des observateurs de la Météorologie Nationale ; elles résultent de la lecture quotidienne directe du contenu de pluviomètres classiques. Il s'agit de mesures assez fiables en raison de leur simplicité et du sérieux de leur collecte ; leur répartition spatiale est généralement bonne en régularité et en densité (chaque département compte 20 à 30 appareils de ce type).

Le choix des mesures finalement retenues pour cette étude (11 journées sur une région couvrant la Beauce, le Thimerais, le Roumois, le Pays de Caux et le Vexin) résulte de la conjonction de plusieurs contraintes dictées essentiellement par les mesures radar :

- malgré le nombre d'heures d'enregistrement assez élevé (environ 700 heures soit environ 9000 images) seules 12 journées complètes étaient disponibles (au sens de 24 heures d'enregistrement continu à partir de 8h du matin, heure de lecture des pluviomètres) ; parmi ces journées, certaines présentaient quelques interruptions de la mesure radar : nous nous sommes assurés qu'aucune pluie au sol n'était alors enregistrée par le réseau de pluviographes couvrant la zone d'étude (voir dans G. DELRIEU,

1986 la description de ce réseau) ; une de ces journées a dû être abandonnée à cause de l'existence de "propagations anormales" du faisceau.

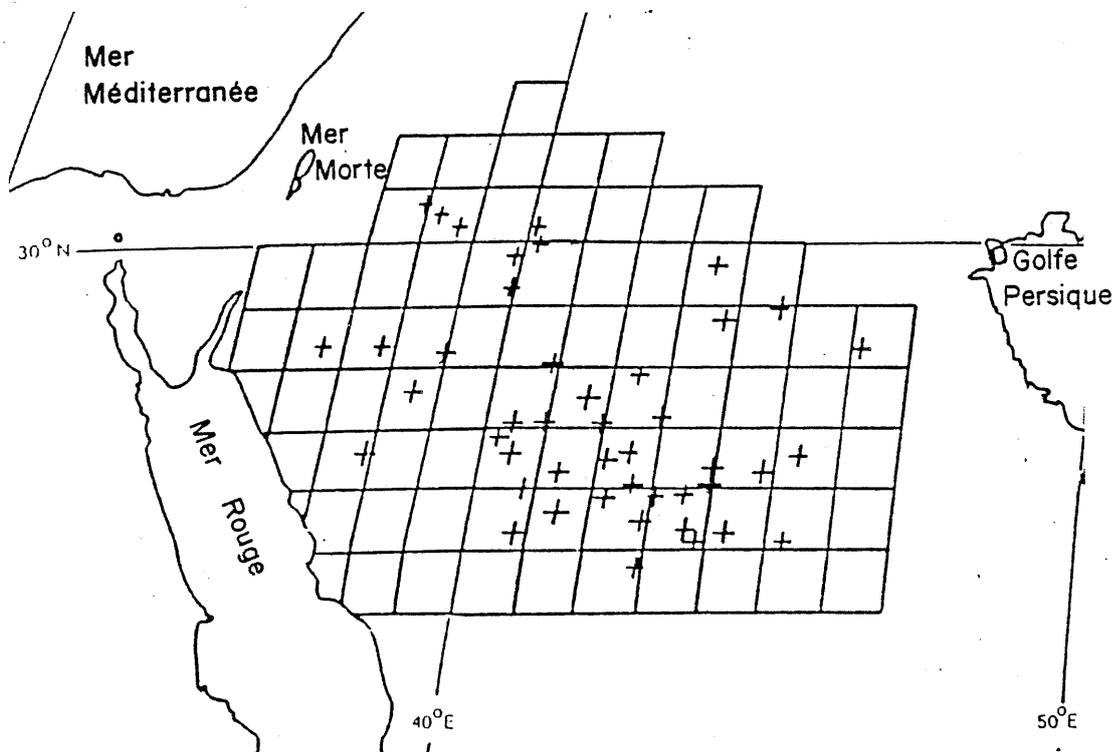
- en raison de la présence d'échos fixes entre 52 et 70 km (agglomérations de Versailles et de Saint Quentin en Yvelines par exemple) seules les mesures entre 70 et 180 km ont été retenues ; l'existence d'effets de masques importants nous ont par ailleurs conduit à limiter notre étude dans un secteur d'angle compris entre les azimuts N210E et N300E environ; pour être complet, mentionnons l'abandon préalable de portions saines d'images radar dans l'Est en raison de la moindre densité du réseau sol.

La figure 0.3 schématise la zone retenue ainsi que l'implantation du réseau de pluviomètres.

MESURES DE LA PLUVIOMETRIE EN ARABIE SEOUDITE

Les mesures de télédétection résultent de l'étude menée à l'I.M.G. par Ch. OBLED, P. LACOMBA et Y. FREIN dans le cadre d'un contrat BRGM (voir le rapport final Ch. OBLED et al., 1984). Il s'agissait de déterminer la pluviométrie sur un vaste aquifère de la partie Nord-Ouest de l'Arabie en utilisant l'information satellitaire. La méthode choisie, inspirée des travaux de E.C. BARRETT, 1977 et A. VAN DIJK, 1981, repose sur l'attribution de facteurs de production de précipitations à différents types de nuage ; cette attribution se fait par régression statistique entre les pourcentages d'ennuage et les hauteurs de précipitations mesurées au sol ; une néphanalyse permet de déterminer ces pourcentages sur les mailles d'une grille régulière en opérant une classification entre trois configurations nuageuses : les bandes nuageuses d'origine intertropicale, les nuages de convection et les perturbations cycloniques. Les détails concernant les satellites utilisés (GOES.10 et DMSP), la méthode de néphanalyse et les différents essais de régression effectués sont exposés dans P. LACOMBA, 1986. Ce travail nous a permis de disposer pour 5 périodes pluvieuses de l'année 1979 (la durée de ces périodes varie entre 4 jours et 1 mois) des estimations des hauteurs de précipitations sur 79 mailles de 1° de côté (soit environ 1200 km²) couvrant la zone étudiée (un million de km² environ).

Figure 0.4 : Carte de la zone d'étude : réseau de 43 pluviomètres et grille de discrétisation de l'information satellitaire.



Les mesures sol proviennent d'un réseau de 43 pluviomètres dont la répartition spatiale est très inégale (les 2/3 de ces appareils sont concentrés sur 1/4 de la surface).

La figure 0.4 schématise la grille de discrétisation des mesures satellite ainsi que l'implantation du réseau de pluviomètres dans la région concernée.

MESURES DE LA FRACTION D'INSOLATION SUR LA FRANCE

Dans ce troisième ensemble de données, les mesures de télédétection proviennent du Centre de Météorologie Spatiale de LANNION où un programme informatique de détermination des pourcentages journaliers d'ennuage sur la France à partir des images de Météosat a été utilisé en routine depuis le mois de Janvier 1984. Ce programme, dénommé CACTUS, repose sur un algorithme simple mis au point par C. PASTRE et N. BERIOT (voir C. PASTRE, 1981) ; pour chaque maille de 20 km de côté d'une grille couvrant la France (3300 mailles) et pour chaque image visible du satellite Météosat (en principe une image toutes les demi-heures), le programme compte le nombre de pixels dont la brillance excède un certain seuil (chaque maille contient environ 20 pixels de l'image Météosat) ; le seuil, dont la valeur varie en fonction de l'angle zénithal solaire, est choisi de manière à distinguer au mieux les pixels nuageux des pixels clairs ; on dispose ainsi pour chaque maille d'un pourcentage "instantané" d'ennuage ; la répétition de ce calcul au long de la journée et sur l'ensemble de la grille conduit à une estimation de la fraction journalière d'ennuage sur la France ; le complément à 1 de cette fraction donne naturellement la fraction d'insolation. Une description détaillée de cette méthode ainsi que des différentes opérations de réglage des seuils et de calage géographique est développée dans S. GALLE, 1987.

L'avantage déterminant de cette méthode est sa simplicité :

- son réglage ne dépendant que de la détermination de 3 seuils type peut être effectué sur une période test beaucoup plus courte que pour des modèles plus physiques mettant en oeuvre des paramètres plus nombreux.

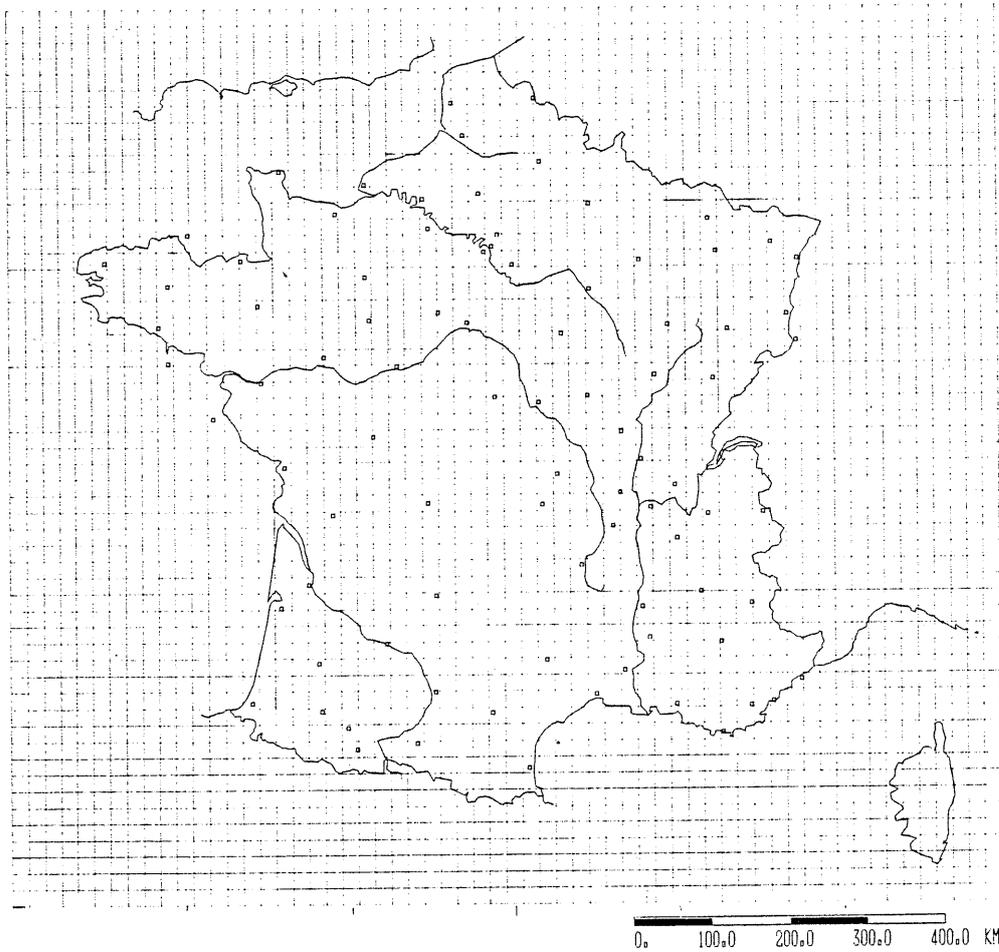


Figure 0.5 : Carte du réseau de 88 héliographes et de la grille utilisée par le programme Cactus (60x55 mailles carrées de 20 km de côté).

- son utilisation en routine ne nécessite aucune précaution particulière et s'effectue avec un temps de calcul raisonnable (environ une heure sur un ordinateur SOLAR dont la puissance de calcul est modeste).

Il a donc été possible au CMS de produire un ensemble d'images journalières unique en ce qui concerne la mesure du rayonnement solaire par télédétection ; nous avons pu bénéficier de 200 images journalières calculées entre Janvier et Décembre 1984 ; d'autres images ont été produites depuis.

En regard de ces images, nous disposons des mesures sol effectuées par les 88 stations du réseau synoptique de la Météorologie Nationale à l'aide d'héliographes. Ces appareils, très rudimentaires, fournissent la durée journalière pendant laquelle le rayonnement solaire au sol a dépassé 200 Watts par m^2 environ ; ce seuil énergétique, qui pratiquement coïncide avec l'apparition d'ombres portées, est un bon indicateur de la présence ou de l'absence de nuage ; la mesure obtenue est donc la durée journalière d'insolation. Cette durée, rapportée à la durée potentielle d'insolation, qui dépend de la période de l'année, de la latitude de la station mais aussi de son environnement topographique, sera appelée fraction journalière d'insolation (voir dans J.D. CREUTIN, 1983, une démarche pratique de calcul de ces fractions). C'est la variable sur laquelle nous travaillerons dans la mesure où elle est très comparable à celle qu'estime le programme CACTUS.

La figure 0.5 représente la grille CACTUS et le réseau héliographique français.

CHAPITRE I

RAPPELS DE GEOSTATISTIQUE

Résumé

Le rôle de ce chapitre est de rappeler brièvement le développement du système de krigeage, auquel nous ferons fréquemment allusion dans les chapitres suivants, et de définir un certain nombre de notations. Il est bien entendu indispensable de se reporter aux travaux de G.MATHERON (notamment 1965, 1970, 1971 et 1980) pour trouver une présentation exhaustive de cette méthode d'estimation dans le cadre de la théorie des variables régionalisées.

BUT DU KRIGEAGE

La connaissance quantitative d'un phénomène naturel se développant dans l'espace nécessite une première étape de mesure qui fournit un ensemble de valeurs $z(t_i, w)$ caractérisant la grandeur scalaire du phénomène mesuré au point de coordonnées t_i (dans le plan t_i représente les coordonnées du i ème point de mesure, soit le couple (x_i, y_i) dans un système cartésien) à l'instant w .

Pour rester au plus près des applications présentées, nous admettrons, d'une part, que le phénomène étudié ne peut pas être représenté fidèlement par une équation analytique simple en t_i et w ; ceci est fréquemment le cas dans les sciences de l'eau ou de l'atmosphère où l'hétérogénéité du milieu et la combinaison de différents mécanismes physiques générateurs rendent une approche déterministe infructueuse pour paramétrer simplement la fonction z . D'autre part, nous considérerons que le réseau de mesure utilisé est implanté de manière aléatoire ; les stations météorologiques sont en effet réparties suivant des préoccupations pratiques qui éloignent parfois fort du maillage régulier ; les coordonnées t_i des n points de mesures ne possèdent donc pas de propriétés géométriques

simples. Enfin, nous supposerons que le phénomène mesuré peut se répéter dans le temps auquel cas il est bon de prévoir un indice w permettant de différencier les observations successives ; assez propre à la météorologie, ce contexte "multiobservations" n'est pas toujours pris en compte par les techniques présentées, aussi l'indice w sera oublié lorsque la connaissance de plusieurs réalisations n'est pas explicitement prise en considération.

Compte-tenu de la complexité du phénomène et du nombre nécessairement limité de points t_i , cette seule étape de mesure ne permet pas d'accéder à la connaissance quasi-continue du phénomène dans l'espace que requièrent la cartographie où l'intégration sur des régions données ; le krigeage (terme générique compact sous lequel nous groupons bien entendu la méthode développée par G.MATHERON mais aussi les techniques telles que l'analyse objective développée par L.S. GANDIN, 1985 ou l'interpolation par certaines fonctions spline) est une méthode permettant l'estimation de la variable z en tout point ou ensemble de points du domaine de l'étude.

ESTIMATEURS LINEAIRES

Comme toutes les méthodes classiques d'interpolation, le krigeage propose un estimateur ou interpolateur z^* de la variable z , en un point t_0 du domaine D étudié, qui combine linéairement les différentes mesures disponibles :

$$z^*_w(t_0) = \sum_{i=1}^n \lambda^i_w(t_0) \cdot z_w(t_i) ; t_0 \in D \quad (I.1)$$

Les poids λ proposés par une méthode d'interpolation linéaire doivent dépendre idéalement :

i) du point t_i où est prise la mesure pondérée afin de privilégier les points de mesures les plus comparables au point t_0 ; cette notion de comparabilité reposera, d'une part, sur des critères géométriques de distance et, d'autre part, sur une appréciation statistique de la variabilité du phénomène et de la covariation de deux points voisins (éventuellement conditionnée par les autres points).

ii) de l'emplacement t_0 du point estimé ; de la même manière que ci-dessus, il paraît logique de faire varier le poids attribué à une mesure lorsque le point reconstitué t_0 se déplace ;

iii) de l'observation w considérée s'il est possible de montrer que la variabilité du phénomène diffère d'une observation à l'autre ou si la composition du réseau change.

Les fonctions $\lambda_w^i(t)$, engendrent l'espace vectoriel des fonctions d'interpolation $Z_w^*(t)$. Chacune d'elle est attachée à un point de mesure pour un type de phénomène, afin d'en fixer le pouvoir explicatif en un point t quelconque. Le choix de ces fonctions est classiquement effectué de manière arbitraire en imposant simplement à une forme analytique donnée (polynomiale par exemple), une condition d'interpolation :

$$\left\{ \begin{array}{l} \lambda_w^i(t_j) = 0 \quad \text{si } i \neq j \\ \lambda_w^i(t_i) = 1 \end{array} \right. \quad (\text{I.2})$$

garantissant que l'interpolateur $Z_w^*(t)$ passe par les points de mesure (on parle d'interpolateur exact) ; le nombre de points où cette condition doit être vérifiée fixe le niveau de complexité des fonctions choisies; on est conduit rapidement, notamment dans le cas des polynômes, à des fonctions dont les oscillations sont peu compatibles avec l'idée que l'on se fait du "pouvoir explicatif" d'un point de mesure ; en fait, hors de ces points, de telles fonctions ne vérifient que des critères de continuité qui ne suffisent pas pour reproduire de manière plausible le phénomène interpolé.

En conservant le même type d'approche déterministe, certaines méthodes (fonctions spline notamment) proposent d'ajouter à cette condition d'interpolation exacte une condition minimisant les oscillations des fonctions de base λ_w^i ; cette condition s'exprime par une minimisation de la courbure moyenne de chacune de ces fonctions de base c'est à dire :

$$\min \int_D (\partial^2 \lambda_w^i(t) / \partial t^2)^2 dt \quad (\text{I.3})$$

la satisfaction simultanée des conditions I.2 et I.3 conduit naturellement à des fonctions de pondération plus satisfaisantes, car beaucoup plus

lisses ; mais le choix de ces conditions, indépendant de l'indice w , implique cependant des hypothèses très fortes sur le comportement du phénomène étudié qu'il est difficile, a priori, de justifier : la condition d'exactitude sous-entend notamment l'absence d'erreur de mesure ponctuelle, la condition de minimum de courbure sous-entend une forte régularité des variations de la variable.

Les fonctions de pondération proposées par le krigeage sont choisies de manière plus pragmatique, dans la mesure où, en plus des caractéristiques géométriques du réseau, elles prennent en compte la structure spatiale du phénomène étudié.

Les conditions imposées à l'estimateur I.1 sont alors de nature statistique ; partant de l'idée que toute estimation sera entachée d'une erreur, le krigeage se fixe pour objectif de rendre, en moyenne quadratique ces erreurs aussi faibles que possible ; cet objectif se traduit par deux conditions : une condition de non biais, qui évite une erreur systématique en annulant l'erreur moyenne et une condition d'optimalité qui pénalise les fortes erreurs en minimisant la moyenne quadratique de l'erreur. Il est clair que la formalisation de ces conditions nécessite l'utilisation d'un cadre mathématique rigoureux: celui de la théorie des fonctions aléatoires.

FONCTIONS ALEATOIRES

Touté la difficulté réside en fait dans l'application, à une variable se développant dans l'espace (ou variable régionalisée), de la notion d'espérance mathématique qui sous-entend l' "espoir" de voir se réaliser, en moyenne, les conditions imposées si l'estimation est réalisée partout un grand nombre de fois dans des conditions égales.

Le concept de fonction aléatoire, généralisation à un espace du concept de variable aléatoire, permet cette application.

La fonction $z_w(t)$, $w^{\text{ième}}$ observation du phénomène étudié dont nous avons mesuré la valeur $z_w(t_i)$ en n points, est considérée comme une réalisation particulière de la fonction aléatoire $Z_w(t)$; cette fonction aléatoire est une famille de fonctions, prenant leurs valeurs dans le même

espace repéré par le paramètre t et telles que, pour N points t_1, t_2, \dots, t_N , il soit possible de définir la fonction de répartition suivante :

$$F(Z_1, Z_2, \dots, Z_N; t_1, t_2, \dots, t_N) = \Pr(Z_w(t_1) < Z_1, Z_w(t_2) < Z_2, \dots, Z_w(t_N) < Z_N) \quad (I.4)$$

qui caractérise le comportement statistique de la famille de fonction $Z_w(t)$ de la même manière que l'on caractérise le comportement d'une variable aléatoire Z_w par les différentes probabilités que Z_w dépasse des valeurs seuil z .

Ce concept permet d'écrire l'estimateur I.1 de manière plus générale, au sens où il s'applique non plus à une fonction particulière $z_w(t)$, mais à une famille de fonctions $z_w(t)$ dont les propriétés statistiques d'ensemble sont connues :

$$Z_w^*(t_0) = \sum_{i=1}^n i_w(t_0) Z_w(t_i) \quad (I.5)$$

Il est alors possible de substituer la notion d'espérance à celle de moyenne dans la condition de non-biais :

$$E(Z_w^*(t_0) - Z_w(t_0)) = 0 \quad (I.6)$$

et d'optimalité :

$$E(Z_w^*(t_0) - Z_w(t_0))^2 \text{ minimal} \quad (I.7)$$

en utilisant le symbole E pour noter le moment d'ordre 1 espérance de la fonction aléatoire $Z_w(t)$ qui se déduit de la fonction F_w .

SYSTEME DE KRIGEAGE

Avant de considérer en détail les contextes pratiques dans lesquels ce concept de fonction aléatoire peut être utilisé, développons les conditions imposées à l'estimateur que nous recherchons sans nous préoccuper du mode d'identification de la loi de probabilité de la fonction aléatoire $Z_w(t)$ et de ses premiers moments (voir § suivant les problèmes d'inférence statistique).

Lorsque l'on utilise la formule I.5, la condition de non-biais devient :

$$M = \sum_{i=1}^n \lambda_{w}^i(t_0) E Z_w(t_i) - E Z_w(t_0) = 0 \quad (I.8)$$

et, après quelques calculs, la condition d'optimalité s'écrit :

$$\begin{aligned} V = & \sum_{i=1}^n \sum_{j=1}^n \lambda_{w}^i(t_0) \lambda_{w}^j(t_0) E Z_w(t_i) \cdot Z_w(t_j) \\ & - 2 \sum_{i=1}^n \lambda_{w}^i(t_0) E Z_w(t_i) \cdot Z_w(t_0) \\ & + E Z_w(t_0) \cdot Z_w(t_0) \text{ minimal} \end{aligned} \quad (I.9)$$

La minimisation de la quantité V par rapport aux différents $\lambda_{w}^i(t_0)$ sous la contrainte d'annulation de M, se fait en introduisant un paramètre de Lagrange $\rho_w(t_0)$ et en annulant les dérivées partielles de $V+2 \rho_w(t_0) M$ par rapport aux différents $\lambda_{w}^i(t_0)$ et à $\rho_w(t_0)$; on obtient alors le système de krigeage :

$$\begin{aligned} \sum_{i=1}^n \lambda_{w}^i(t_0) E Z_w(t_i) \cdot Z_w(t_j) + \rho_w(t_0) E Z_w(t_j) &= E Z_w(t_j) \cdot Z_w(t_0) \text{ pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda_{w}^i(t_0) E Z_w(t_i) &= E Z_w(t_0) \end{aligned} \quad (I.10)$$

Lorsque ce système est vérifié, l'expression I.9 devient :

$$V = - \sum_{i=1}^n \lambda_{w}^i(t_0) E Z_w(t_i) \cdot Z_w(t_0) - \rho_w(t_0) E Z_w(t_0) + E Z_w(t_0) Z_w(t_0) \quad (I.11)$$

qui est la variance d'erreur de l'estimateur ou variance d'estimation.

Il est possible de vérifier que la linéarité de l'estimateur de krigeage permet d'obtenir directement la valeur du phénomène intégrée sur une zone D donnée :

$$\begin{aligned}
 Z^*(D) &= \int_D Z^*_W(t) dt = \int_D \sum_{i=1}^n \lambda^i_W(t) Z_W(t_i) dt \\
 &= \sum_{i=1}^n \int_D \lambda^i_W(t) dt Z_W(t_i) = \sum_{i=1}^n \lambda^i_W(D) Z_W(t_i)
 \end{aligned}
 \tag{I.12}$$

en substituant D à t_0 dans le système I.10 et l'équation I.11, l'espérance $E Z_W(t_i) \cdot Z_W(D)$ étant donnée par :

$$\int_D E Z_W(t_i) \cdot Z_W(t) dt$$

Notons enfin que le système I.10 peut se noter sous forme matricielle de la manière suivante :

$$\begin{bmatrix} S & D \\ D^T & 0 \end{bmatrix} \times \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} S_0 \\ D_0 \end{bmatrix}
 \tag{I.13}$$

$$\begin{aligned}
 \text{avec : } S &= |EZ_W(t_i) \cdot Z_W(t_j)|^{n \times n} \\
 D^T &= |EZ_W(t_1) \dots EZ_W(t_n)|^{1 \times n} \\
 \Lambda^T &= |\lambda^1_W(t_0) \dots \lambda^n_W(t_0)|^{1 \times n} \\
 S^T_0 &= |EZ_W(t_i) \cdot Z_W(t_0)|^{1 \times n} \\
 D^T_0 &= EZ_W(t_0)
 \end{aligned}$$

ce qui conduit à calculer l'estimateur $Z^*_W(t_0)$ par :

$$Z^*_W(t_0) = \begin{bmatrix} Z \\ 0 \end{bmatrix}^T \cdot \left(\begin{bmatrix} S & D \\ D^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} S_0 \\ D_0 \end{bmatrix} \right)
 \tag{I.14}$$

$$\text{avec } Z^T = |Z_W(t_1), Z_W(t_2) \dots Z_W(t_n)|^{1 \times n}$$

Ce produit étant associatif, il est possible d'évaluer l'estimateur par :

$$Z^*_w(t_0) = \begin{pmatrix} Z \\ 0 \end{pmatrix}^T \times \begin{pmatrix} S & D \\ D^T & 0 \end{pmatrix}^{-1} \times \begin{pmatrix} S_0 \\ D_0 \end{pmatrix} \quad (\text{I.15})$$

reportant ainsi dans le deuxième facteur de la multiplication le seul vecteur dépendant de l'emplacement t_0 du point où se fait l'estimation ; ce conditionnement est qualifié d'interpolateur (par rapport au conditionnement I.14 qualifié d'estimateur) ; la première multiplication de I.15 fournit en effet un nouveau vecteur :

$$\begin{pmatrix} A \\ b \end{pmatrix}^T = \begin{pmatrix} Z \\ 0 \end{pmatrix}^T \times \begin{pmatrix} S & D \\ D^T & 0 \end{pmatrix}^{-1} \quad \text{avec } A^T = |a^1_w \dots a^n_w|^{1 \times n} \quad (\text{I.16})$$

qui permet de calculer directement l'estimateur Z^*_w en tout point t_0 par :

$$Z^*_w(t_0) = \sum_{i=1}^n a^i_w EZ_w(t_i) \cdot Z_w(t_0) + b_w EZ_w(t_0) \quad (\text{I.17})$$

où les coefficients a^i_w et b_w sont indépendants de t_0 et donnés par le système suivant :

$$\begin{aligned} Z^*_w(t_j) &= Z_w(t_j) \quad \text{pour } j = 1 \text{ à } n \\ \sum_{i=1}^n a^i_w EZ_w(t_i) &= 0 \end{aligned}$$

La parfaite équivalence de ces deux conditionnements (G.MATHERON, 1981 et O.DUBRILE, 1982) autorise à choisir l'un ou l'autre suivant des critères pratiques de commodité de calcul ou de mémorisation des résultats. Dans la suite du texte nous privilégierons la formulation estimateur dont le lien avec les conditions de non-biais et d'optimalité est plus naturel. La formulation interpolateur s'adapte plus à des critères globaux tel que

I.3 ; elle est utilisée pour les fonctions spline où $EZ_w(t_i) \cdot Z_w(t_0)$ est calculée à l'aide de la fonction $(t_i - t_0)^2 \text{Log } |t_i - t_0|$ et où $EZ_w(t_0) = \alpha t_0 + \beta$ dans le cas des fonctions spline de type plaque mince (M. ATTEIA, 1966 ; J. DUCHON, 1975 ; L. PAIHUA-MONTES, 1978) $-(t_i - t_0)$ représente la distance séparant les points t_i et t_0 . Notons cependant que la formulation estimateur se rencontre dans le domaine des interpolateurs polynomiaux : l'interpolation de Lagrange propose de calculer les poids $\lambda^i(t_0)$ par :

$$\lambda^i(t_0) = \frac{(t_0 - t_1)(t_0 - t_2) \dots (t_0 - t_{i-1})(t_0 - t_{i+1}) \dots (t_0 - t_n)}{(t_i - t_1)(t_i - t_2) \dots (t_i - t_{i-1})(t_i - t_{i+1}) \dots (t_i - t_n)}$$

(voir N. APPERT et C. LOUVEAU, 1973)

Arrivés à ce moment de la description du krigeage, il est nécessaire de se poser la question de l'applicabilité d'un tel système d'estimation. En effet, pour formuler des conditions de nature statistique, nous avons supposé qu'il était possible de décrire le comportement du phénomène étudié à l'aide du concept de fonction aléatoire. Nous devons maintenant examiner dans quelle mesure ce modèle probabiliste peut être inféré, c'est à dire ajusté à la réalité. Notons que le terme de modèle constitue peut être ici un abus de langage ; le recours au concept de Fonction Aléatoire est en effet avant tout, dans un premier temps, un choix méthodologique.

CHOIX D'UN MODELE

Dans ce "dialogue" entre le problème traité et le modèle proposé d'une part et entre le modèle proposé et le système d'estimation d'autre part, dialogue qui traduit le pragmatisme de la méthode, G.MATHERON (1975) propose de retenir quatre critères pour choisir un modèle :

- a) - Possibilité d'en réaliser l'inférence c'est à dire de l'adapter numériquement aux données disponibles ;

b) - Capacité de ce modèle à rendre la résolution du système de krigeage possible ; nous entendons par résolution le calcul des pondérateurs mais aussi celui de la variance d'estimation ;

c) - Compatibilité du modèle avec les caractéristiques statistiques du phénomène étudié ;

d) - Efficacité du modèle en terme de précision des résultats.

Le modèle de fonction aléatoire que nous avons choisi au départ satisfait pleinement le critère b dans la mesure où la connaissance de la fonction de répartition I.4 suffit pour calculer en tout point t_k ;

le moment d'ordre 1 ou espérance :

$$M_1(t_k) = EZ_w(t_k) = \int_{-\infty}^{+\infty} Z_k f_w^{(1)}(Z_k ; t_k) dZ_k$$

avec : $f_w^{(1)}(Z_k; t_k) dZ_k = dF_w(Z_1, Z_2 \dots Z_N ; t_1, t_2 \dots t_N)$ et pour tout couple de points $(t_k, t_{k'})$.

le moment d'ordre 2 :

$$M_{1,1}(t_k, t_{k'}) = \iint_{-\infty}^{+\infty} Z_k Z_{k'} f_w^{(2)}(Z_k, Z_{k'} ; t_k, t_{k'}) dZ_k dZ_{k'}$$

avec $f_w^{(2)}(Z_k, Z_{k'}; t_k, t_{k'}) dZ_k dZ_{k'} = d^2F_w(Z_1, Z_2 \dots Z_N ; t_1, t_2 \dots t_N)$, qui permet le calcul des covariances simples $EZ_w(t_k) \cdot Z_w(t_{k'}) = M_{1,1}(t_k, t_{k'})$. La connaissance de la fonction de répartition suffit donc pour résoudre le système I.10.

De plus, le critère c à toutes chances d'être satisfait à cause de la grande souplesse du modèle.

Malheureusement, le critère a n'est jamais satisfait par un modèle aussi général qui suppose que l'on observe préalablement à son étude un phénomène pendant un temps infiniment long et en un nombre infini de

points afin de pouvoir identifier la loi de répartition ou tout au moins ses deux premiers moments.

Dans les paragraphes qui suivent, nous nous laisserons guider par le critère a pour réduire la complexité de ce modèle suivant les contextes expérimentaux rencontrés ; les critères b et c serviront de garde-fous pour conserver à ces modèles leurs propriétés opérationnelles. Enfin, même si le critère d n'est pas évoqué par la suite, Précisons qu'il fait l'objet du chapitre suivant.

Comme nous venons de l'évoquer, l'identification d'une fonction de répartition contient un aspect temporel dans l'opérateur P_r et un aspect spatial qui rend compte de l'évolution de ces probabilités en fonction du lieu t .

INFERENCE STATISTIQUE DANS LE TEMPS

Nous avons jusqu'à maintenant proposé de considérer chaque observation $z_w(t)$ du phénomène étudié comme une réalisation d'une fonction aléatoire $Z_w(t)$. Ceci conduit, si l'on dispose de P observations d'un phénomène à étudier P fonctions aléatoires distinctes.

Nous dirons que nous sommes dans un contexte climatologique lorsque ces P fonctions aléatoires possèdent la même fonction de répartition. Les P observations $z_w(t)$ peuvent alors être considérées comme des réalisations indépendantes d'une fonction aléatoire Z . D'un point de vue physique, cette approche sous-entend l'unicité du processus générateur des différents champs $z_w(t)$ observés à différents instants w sur le domaine décrit par t . D'un point de vue pratique, si nous disposons d'un nombre P "suffisamment grand" d'observations, il sera possible de les utiliser pour estimer par des moyennes d'ensemble les deux premiers moments $EZ(t_i)$ et $E(Z(t_i)-EZ(t_i))(Z(t_j)-EZ(t_j))$ aux différents points de mesure t_i .

Le terme climatologique est choisi par référence à la discipline météorologique qui consiste à collectionner au fil du temps les mesures d'une même variable dans l'espace.

Notons que l'hypothèse d'indépendance des réalisations $z_w(t)$ est forte lorsque le processus générateur est continu dans l'espace et le temps. Si l'on note $z(t,w)$ ce processus en indiquant le temps par w , les observations $z_w(t)$ à l'instant w résultent généralement d'une intégration

dans le temps : $z_w(t) = \int_w^{w+\Delta w} z(t,w) dw$ sur un pas d'échantillonnage Δw .

Il faudra que l'écart entre les instants d'échantillonnage w soit "suffisamment lâche" au regard du pas Δw pour que l'autocorrélation temporelle du processus n'intervienne plus. En d'autres termes, l'échantillonnage temporel est supposé tel que l'ordre des indices w n'ait pas lieu d'être pris en compte ; nous évoquerons, au paragraphe IV, une démarche susceptible de traiter le cas contraire où les observations sont organisées suivant w .

Nous dirons que nous sommes dans un contexte spatial pur lorsque les P fonctions aléatoires possèdent des fonctions de répartition distinctes. Cette approche est justifiée dans la pratique lorsque l'on dispose de "trop peu" d'observations pour établir des moyennes d'ensemble fiables. C'est naturellement un contexte fréquent dans les Sciences de la Terre où les observations d'un phénomène sont souvent uniques. Ce contexte est également utile lorsque l'on doute de l'unicité du processus générateur des champs observés.

La frontière entre les notions de "suffisamment grand" et de "trop peu" est évidemment floue. Il existe des tests mathématiques permettant de quantifier la robustesse et l'homogénéité d'un ensemble d'observations, mais leur maniement est lourd. Dans la pratique, nous trancherons souvent en utilisant la connaissance du phénomène étudié qui donne une idée du nombre d'observations à traiter pour arriver à une certaine stabilité des moments calculés.

Pour simplifier les notations, nous désignerons par la suite les fonctions aléatoires sans indice w en gardant en mémoire que, suivant le contexte choisi, la fonction Z est connue par une ou plusieurs de ses réalisations.

INFERENCE STATISTIQUE DANS L'ESPACE

D'un point de vue spatial, nos hypothèses de départ supposent le comportement probabiliste du phénomène connu "partout" ; sans entrer dans le problème délicat du passage de N à l'infini, il est au moins nécessaire d'admettre que la fonction de répartition F est connue pour tout ensemble de points $t_1 \dots t_N$, si l'on veut considérer le phénomène comme une fonction aléatoire ; or, il est parfaitement exclu dans un contexte opérationnel d'accéder à une telle connaissance ; les critères b et c évoqués ci-dessus nous invitent à examiner plus finement le problème traité : deux exemples peuvent illustrer des choix de modèles très différents.

Supposons que nous souhaitions appliquer le système de krigeage dans un contexte climatologique pour estimer le phénomène en un point t_0 où des mesures ont déjà été réalisées (cas de la reconstitution de données manquantes en un point de mesure, mais aussi cas de la cartographie d'un phénomène mesuré durant une longue période antérieure à l'aide d'un appareil dont la densité spatiale de mesure est comparable à celle des grilles de cartographie). Nous pouvons donc identifier la fonction F , ou tout au moins ses premiers moments, aux points t_i et au point t_0 ; il suffit alors pour résoudre le système I.10 de calculer les différents moments croisés d'ordre 2 entre les variables aléatoires $Z(t_i)$, $Z(t_j)$ et $Z(t_0)$ ainsi que leurs espérances ; le comportement du phénomène en dehors de ces points nous importe peu, donc le recours au modèle fonction aléatoire est inutile: le système I.10 est alors un système classique de régression multiple.

Au contraire, si nous supposons que le point t_0 n'a jamais fait l'objet de mesure (cas de la cartographie), nous allons être obligés de considérer la variable aléatoire $Z(t_0)$ comme faisant partie d'un ensemble plus large : une fonction aléatoire $Z(t)$ dont les propriétés en certains points t_i peuvent s'étendre à d'autres points t_0 . *Quelles sont les hypothèses les plus raisonnables que l'on puisse faire au sujet de cette extension ?* (question posée par le critère c). Différents modèles de fonction aléatoires proposent des réponses en faisant l'hypothèse que le phénomène étudié est stable dans l'espace (moments directs indépendants du point d'appui ou variant lentement) et présente par conséquent une certaine

structure (moments croisés liés au vecteur formé par les points considérés); cette stabilité et cette structure confèrent alors au phénomène de bonnes propriétés vis à vis de l'échantillonnage et permettent la résolution du système de krigeage.

FONCTION ALEATOIRE STATIONNAIRE

La stationnarité est une hypothèse commune à tous les modèles utilisés en géostatistique ; nous allons en décrire tout d'abord son application directe à la variable étudiée, puis montrer qu'elle s'avère souvent plus réaliste lorsqu'on l'applique après transformation de cette variable.

On dit que la fonction aléatoire $Z(t)$ est stationnaire si sa loi de répartition F reste inchangée pour toute translation de vecteur d imposée à l'ensemble des points t :

$$F(Z_1, Z_2 \dots Z_N ; t_1+d, \dots t_N+d) = F(Z_1, Z_2 \dots Z_N ; t_1, \dots t_N) \quad (I.19)$$

Cette hypothèse de stabilité des propriétés probabilistes d'une fonction dans l'espace a deux conséquences pratiques très importantes :

i) si l'on observe de façon "suffisamment" complète une seule réalisation de la fonction, il est possible de déterminer la loi de répartition de cette fonction en faisant une hypothèse d'ergodicité qui assure l'équivalence probabiliste entre l'observation de N points et N observations en un point, sous réserve de quelques précautions garantissant l'indépendance des N points échantillonnés ; nous voyons apparaître ici une solution au problème difficile de l'inférence dans un contexte spatial pur. Cette solution ne nous paraît pas exiger l'existence réelle d'une ergodicité ; en effet, supposons que la fonction étudiée soit construite de la manière suivante. :

$$Z_w(t) = A(w).Z'_w(t) + B(w)$$

où $Z'_w(t)$ est une fonction aléatoire stationnaire ergodique, $A(w)$ et $B(w)$ des variables aléatoires ; il est clair que la connaissance d'une seule

réalisation w_0 de cette fonction ne nous permettra pas d'accéder à sa loi de répartition puisque nous n'avons alors aucune information sur la distribution des A et B ; par contre, nous avons accès à la loi de répartition de cette fonction pour A_{w_0} et B_{w_0} fixés qui suffit pour résoudre nos problèmes d'estimation.

ii) Si la loi F possède des moments, la stationnarité s'y applique évidemment ; ceci implique d'une part que la valeur moyenne du phénomène (moment d'ordre 1) est constante dans l'espace :

$$EZ(t+d) = EZ(t) = m \quad (I.20)$$

et, d'autre part, que la covariance (moment d'ordre 2) ne dépend que du vecteur formé par les points t et t' :

$$EZ(t+d).Z(t'+d) = EZ(t).Z(t') = c'(tt') \quad (I.21)$$

Notons qu'il est souvent possible, par transformation géométrique appliquée aux coordonnées t, de se ramener d'une covariance anisotrope (où $c'(tt')$ dépend de l'orientation du vecteur) à une covariance isotrope $c(tt')$ ne dépendant plus que de l'interdistance des points ; nous considérerons toujours par la suite que nous nous sommes ramenés à un cas isotrope.

Si ces propriétés concernent les seuls deux premiers moments, on parle alors de stationnarité d'ordre 2 et les moments d'ordres supérieurs à 2 sont supposés quelconques. Elles suffisent pour résoudre le système de krigeage I.10 dans le cas où le point t_0 n'a pas fait l'objet de mesure ; en effet les différents moments d'ordre 2 peuvent être déduits de la fonction $c(h)$ et la valeur moyenne est connue en tout point.

Notons que la fonction $c(h)$ ne peut être quelconque ; en effet elle est utilisée pour calculer des moments d'ordre deux de combinaisons linéaires :

$$E\left(\sum_{i=1}^n \lambda^i Z(t_i)\right)^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda^i \lambda^j c(h_{ij}) \quad (I.22)$$

qui doivent donc quelles que soient les pondérations λ^i prendre des valeurs positives ; on dira qu'une telle fonction est conditionnellement (c'est à dire si les λ^i vérifient par ailleurs une relation du type $\sum_{i=1}^n \lambda^i EZ(t_i) = 1$ par exemple) définie positive ; nous ne rentrerons pas dans le détail des différents critères permettant d'établir cette propriété (G. CHRISTAKOS, 1984), nous nous limiterons simplement à n'utiliser que des fonctions classiques pour modéliser nos fonctions de covariance.

En résumé, le concept de fonction aléatoire stationnaire présente donc deux avantages majeurs : celui de rendre envisageable le traitement géostatistique de phénomènes connus par peu de réalisations, voire une seule et celui de permettre l'utilisation du système de krigeage en des lieux où le phénomène n'a jamais été mesuré.

Malheureusement, ces avantages sont acquis au prix d'hypothèses difficilement tenables dans la nature, où la valeur moyenne d'un phénomène et sa variabilité changent souvent d'un point à l'autre.

Deux voies ont été utilisées pour conserver les avantages de la stationnarité en assouplissant ses exigences : celle de la statistique classique qui opère par codage de la fonction $Z(t)$ dans un contexte climatologique (il s'agit de l'analyse objective proposée par L.S. GANDIN, 1965 et les hydrométéorologues russes) et celle de la statistique des processus qui opère en différenciant la fonction $Z(t)$ (il s'agit de la théorie intrinsèque proposée par G. MATHERON, 1971).

ANALYSE OBJECTIVE

Développée dans un contexte climatologique, cette méthode suppose i) dans un premier temps que l'on a accès en chaque point de mesure à la valeur moyenne $EZ(t_i)$ du phénomène et que le champ de ces valeurs moyennes $EZ(t)$ est, soit connu partout, soit suffisamment lisse pour qu'il soit possible d'en estimer la valeur en tout point t_0 .

On peut alors appliquer les hypothèses de la stationnarité d'ordre 2 à la fonction aléatoire centrée :

$$Y(t) = Z(t) - EZ(t) \quad (\text{I.23})$$

dont la moyenne est nulle et dont la covariance peut se noter :

$$\begin{aligned} C(t, t') &= E(Y(t).Y(t')) \\ &= E(Z(t).Z(t') - EZ(t).EZ(t')) \\ &= c(t, t') - m(t).m(t') \end{aligned} \quad (\text{I.24})$$

et se modéliser à l'aide d'une fonction $C(h)$; il est intéressant de noter que c , covariance de Z , peut alors avoir un comportement dépendant des points t et t' .

Le système de krigeage fournira alors un estimateur $Y(t_0)$ sans avoir recours à la condition de non-biais d'emblée satisfaite :

$$\sum_{i=1}^n \lambda^i(t_0) C(h_{ij}) = C(h_{j0}) \quad \text{pour } j = 1 \text{ à } n \quad (\text{I.25})$$

où h_{ij} représente la distance entre les points t_i et t_j . La variance de cet estimateur sera :

$$V = C(0) - \sum_{i=1}^n \lambda^i(t_0) C(h_{i0}) \quad (\text{I.26})$$

Mais, il faut remarquer que cette variance ne porte que sur l'erreur d'estimation de $Y(t_0)$; l'estimation de $Z(t_0)$ par :

$$Z^*(t_0) = E^*Z(t_0) + \sum_{i=1}^n \lambda^i(t_0) Y(t_i) \quad (\text{I.27})$$

nécessite la connaissance ou l'estimation de la valeur moyenne en t_0 ; lorsque cette valeur est connue, alors la variance d'estimation de Z est donnée par I.26 et $E^*Z(t_0)$ est remplacée par sa vraie valeur $EZ(t_0)$ dans I.27.

Par contre, lorsque $EZ(t_0)$ doit être estimée, il y a lieu de rajouter à la variance I.26 la variance d'estimation de cette valeur moyenne - ceci dans le cas où les espérances sont estimées sur un échantillon indépendant de la réalisation traité (dans le cas contraire, il

y a lieu de tenir compte d'une liaison valeurs de la réalisation-espérances estimée, ce qui complique le problème).

Dans l'hypothèse où les espérances $EZ(t)$ sont estimées à l'aide de p réalisations indépendantes, il est possible de montrer, en suivant l'approche de G.H. JOWETT, 1965 (voir annexe I.A), que la covariance centrée des estimations de l'espérance notées E^{\wedge} aux points t et t' se déduit de la covariance de la fonction étudiée par la formule :

$$C_{E^{\wedge}}(t,t') = \frac{1}{p} C(t,t') \quad (\text{I.28})$$

Il paraît alors séduisant de former un estimateur linéaire de l'espérance au point t_0 suivant la formule :

$$E^*Z(t_0) = \sum_{i=1}^n \lambda_E(t_0) E^{\wedge}Z(t_i) \quad (\text{I.29})$$

auquel on impose à nouveau des conditions de non-biais et de variance d'erreur minimale ; ceci conduit, selon les mêmes calculs qu'en I.10 au système :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda_E(t_0) E(E^{\wedge}Z(t_i) \cdot E^{\wedge}Z(t_j)) + \rho_E EZ(t_j) = EZ(t_j) \cdot EZ(t_0) \text{ pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda_E(t_0) EZ(t_i) = EZ(t_0) \end{array} \right.$$

Si l'on remarque que la condition de non-biais implique :

$$EZ(t_j) \cdot EZ(t_0) = EZ(t_j) \cdot \sum_{i=1}^n \lambda_E(t_0) EZ(t_i) = \sum_{i=1}^n \lambda_E(t_0) EZ(t_i) \cdot EZ(t_j)$$

Nous obtenons le système :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda_E(t_0) C_{E^{\wedge}}(t_i, t_j) + \rho_E EZ(t_j) = 0 \quad \text{pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda_E(t_0) EZ(t_i) = EZ(t_0) \end{array} \right. \quad (\text{I.30})$$

La variance d'estimation ayant pour forme :

$$V = -\rho_E EZ(t_0) \quad (\text{I.31})$$

Malheureusement, nous voyons réapparaître dans ce système les valeurs $EZ(t_0)$ auxquelles nous n'avons pas accès par hypothèse.

Il est donc nécessaire de formaliser mieux l'hypothèse d'un champ $EZ(t)$ "suffisamment lisse" en écrivant $EZ(t)$ sous la forme :

$$EZ(t) = \sum_{l=1}^L a_l f^l(t) \quad (\text{I.32})$$

où $f^l(t)$ sont des fonctions de base, polynomiales par exemple ; cette écriture, dont on peut restreindre la validité à un voisinage de t_0 en remplaçant éventuellement n par n' , nombre de voisins choisis, permet de calculer le système I.30 par :

$$\begin{aligned} \sum_{i=1}^n \lambda_E(t_0) C_E^{\wedge}(t_i, t_j) + \sum_{l=1}^L \rho_l f^l(t_j) &= 0 \quad \text{pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda_E(t_0) f^l(t_i) &= f^l(t_0) \quad \text{pour } l = 1 \text{ à } L \end{aligned} \quad (\text{I.33})$$

où la variance d'estimation devient :

$$C = - \sum_{l=1}^L \rho_l f^l(t_0) \quad (\text{I.34})$$

Nous obtenons ainsi un estimateur $E^*Z(t_0)$ qui nous permet d'écrire I.27 comme suit :

$$Z^*(t) = \sum_{i=1}^n \lambda_E(t_0) E^{\wedge}Z(t_i) + \sum_{i=1}^n \lambda_i(t_0) Y(t_i) \quad (\text{I.35})$$

et pour lequel la variance d'estimation vaut :

$$V = C(0) - \sum_{i=1}^n \lambda^i(t_0) C(h_{i0}) - \sum_{l=1}^L \rho_l f^l(t_0) \quad (\text{I.36})$$

Cette démarche, très lourde, n'a en fait qu'un intérêt "pédagogique" ; elle permet de montrer à quel niveau se fait sentir le besoin de formaliser l'hypothèse de champ de valeurs moyennes lisse et elle souligne les difficultés rencontrées lorsque l'on prend la décision de faire des estimations séparées d'un terme moyen et des fluctuations autour de cette moyenne. Nous verrons, au paragraphe suivant, que l'introduction de la formulation I.32 directement dans le système I.10, que l'on qualifie alors de krigeage universel, fournit une solution plus élégante sans restreindre beaucoup la généralité de ce type d'interpolateur. Cette "élégance" ne doit cependant pas masquer la difficulté de prendre en compte un champ de valeurs moyennes de la forme I.32 au moment de l'identification de la covariance centrée C.

Enfin, nous pouvons remarquer que cette application de l'analyse objective reprend la démarche classique en statistique qui consiste à distinguer, lors de la comparaison de variables aléatoires, la part de ressemblance due à la co-fluctuation (c'est à dire due à la similitude de l'évolution moyenne des valeurs que l'on appelle ici dérive) de la part due à la covariation (c'est à dire due à l'existence d'une liaison statistique linéaire entre les fluctuations autour de la dérive). Cette distinction consiste donc à "soustraire" l'effet des caractéristiques du phénomène présentant une tendance systématique dans l'espace.

Tout le problème est alors de savoir où trouver une bonne estimation de cette dérive ; il est clair que dans un contexte climatologique où les différentes observations d'un phénomène ne présentent pas d'évolution temporelle marquée (hypothèses formulées par L.S. GANDIN, 1965, notamment) le réseau de mesure peut fournir un bon estimateur de ces tendances ; cette estimation peut être confortée par l'introduction d'information extérieure si il existe une relation entre la dérive et une variable exogène, par exemple l'altitude, connue en tout point t_0 ; d'autre part, il peut être intéressant, dans le cas où l'on analyse une succession

de réalisations présentant une certaine continuité dans le temps, de considérer la dernière estimation réalisée comme une bonne estimation de la tendance moyenne du phénomène (voir T. SCHLATTER, 1974 et DURAND, 1975).

Nous voyons donc que le choix de l'estimateur de la dérive doit se faire à partir des caractéristiques du phénomène étudié et apparaît aussi comme une ouverture possible à une source d'information complémentaire.

ii) Dans un deuxième temps, il est possible de supposer que l'on a accès en chaque point de mesure à la variance $E(Z(t_i) - EZ(t_i))^2$ du phénomène.

Les hypothèses de la stationnarité d'ordre 2 sont alors appliquées à la fonction aléatoire centrée et réduite :

$$X(t) = \frac{Y(t)}{\sqrt{C(t,t)}}$$

dont la valeur moyenne est nulle et dont la covariance notée :

$$\begin{aligned} R(t,t') &= EX(t).X(t') \\ &= \frac{C(t,t')}{\sqrt{C(t,t).C(t',t')}} \end{aligned}$$

est le coefficient de corrélation que l'on modélisera par une fonction $R(h)$, dite corrélogramme. Ce modèle possède naturellement, encore plus de souplesse que le précédent puisqu'il n'exige plus à son tour l'homogénéité de la fonction C , qui peut dépendre des points t et t' . Contre-partie logique de cette plus grande souplesse : il sera nécessaire de pouvoir estimer la variance en tout point t_0 où l'on souhaite estimer Z la variable initiale ; il est bien entendu possible de faire des hypothèses sur le champ des variances, mais les difficultés rencontrées sont encore plus ardues que pour le champ des moyennes ce qui limite pratiquement l'application de ce type de modèle au cas où l'on connaît $C(t_0, t_0)$, variance du phénomène au point t_0 .

FONCTIONS ALEATOIRES INTRINSEQUES D'ORDRE k

Dans un contexte spatial pur, le nombre de réalisations mesurées est insuffisant pour estimer correctement l'évolution spatiale de la valeur moyenne ou de la variance du phénomène. Pourtant, une telle évolution, si elle existe, interdit l'application d'hypothèses de stationnarité et, d'un point de vue pratique lorsqu'on la néglige, on est conduit à des estimations erronées de la variable et de la variance d'erreur. Considérons deux exemples simples pour illustrer cette affirmation :

i) Supposons un phénomène stationnaire d'ordre 2 de moyenne non nulle m ; la fonction de covariance à laquelle nous avons accès expérimentalement est $c(h) = C(h) + m^2$ (en effet, si $Z(t) = Y(t) + m$, $c(h) = EZ(t).Z(t+h) = E(Y(t)+m).(Y(t+h)+m) = C(h) + m^2$ puisque $EY(t) = EY(t+h) = 0$) ; compte-tenu des conditions de non-biais (ici $\sum_{i=1}^n \lambda^i(t_0) = 1$) l'utilisation de la fonction c au lieu de C dans le système de krigeage I.10 n'aura pas d'effet sur l'estimateur obtenu ni sur la variance d'erreur.

ii) Par contre si le phénomène étudié présente une dérive de la forme $m(t) = at + b$, la fonction de covariance $c(h) = C(h) + a^2t^2 + 2abt + a^2th + abh + b^2$ contient des termes en t^2 que la condition de non-biais (ici $\sum_{i=1}^n \lambda^i(t_0)t_i = t_0$ et $\sum_{i=1}^n \lambda^i(t_0) = 1$) ne parvient pas à "filtrer"; l'estimateur obtenu ainsi que la variance d'erreur sont donc erronés car calculés à l'aide d'une fonction de structure biaisée.

Nous éprouvons donc à nouveau le besoin de distinguer dans la ressemblance de deux variables la part due à la co-fluctuation de celle due à la covariation, avec ici la contrainte supplémentaire de ne pas pouvoir ponctuellement effectuer une dichotomie valeur moyenne-fluctuation autour de cette valeur en raison du caractère purement spatial du problème posé.

La solution proposée par G. MATHERON (1971; voir aussi P. DELFINER et G. MATHERON, 1980) et par M.I. YUDIN (1961, cité par L.S. GANDIN, 1965, page 36) généralise au cas des processus spatiaux, mesurés à l'aide de réseaux répartis aléatoirement un codage utilisé en analyse de processus

temporels en présence d'une tendance (modèle ARIMA, voir BOX et JENKINS, 1976).

Ce codage consiste à combiner linéairement différentes valeurs de la variable étudiée $Z(t)$ pour former une nouvelle variable :

$$dZ(t; \vec{\lambda}^\alpha) = Z(t) - \sum_{\alpha} \lambda^\alpha Z(t_\alpha) \quad (\text{I.37})$$

à laquelle on va imposer des hypothèses de stationnarité d'ordre 2 quelques soient les points t_α choisis :

i) sa valeur moyenne doit être constante dans l'espace, ce qui conduit à choisir les poids λ^α de la manière suivante :

$$EdZ(t; \vec{\lambda}^\alpha) = EZ(t) - \sum_{\alpha} \lambda^\alpha EZ(t_\alpha) = 0 \quad (\text{I.38})$$

Il est toujours possible de fixer la valeur moyenne à zéro puisque nous avons le choix de l'ordre de la combinaison linéaire ; ainsi, lorsque $EdZ(t; \vec{\lambda}^\alpha) = m$, nous avons toujours la possibilité de travailler sur la variable $d'Z(t; \vec{\rho}) = dZ(t; \vec{\lambda}) - dZ(t'; \vec{\lambda}')$, qui est bien de la forme I.37 et dont la valeur probable est nulle.

Par contre, le choix de ce codage implique, pour un vecteur de poids donné, que le champ des valeurs moyennes $EZ(t)$ possède de bonnes propriétés vis à vis de la translation :

$$EdZ(t+h; \vec{\lambda}) = EdZ(t; \vec{\lambda}) \quad (\text{I.39})$$

Si l'on décompose la fonction $EZ(t)$ sur une base de dimension finie :

$$EZ(t) = \sum_{l=1}^L a_l f^l(t) \quad (\text{I.40})$$

la condition I.39 exige que les fonctions $f^l(t)$ soient invariantes par translation, c'est à dire que $f^l(t+h)$ doit pouvoir s'exprimer comme suit d'après P. DELFINER et G. MATHERON, 1980) :

$$f^1(t+h) = \sum_{s=1}^S b_s^1(h) f^s(t)$$

ce qui réduit le choix de ces fonctions dans l'ensemble des fonctions exponentielles-polynômes :

$$f^1(t) = P(t).e^{at}$$

Cette restriction est en fait peu contraignante car de telles fonctions permettent localement de modéliser un ensemble très large de comportements de la dérive $EZ(t)$; on se limite même dans la pratique à choisir les fonctions $f^1(t)$ sous forme de monômes :

$$f^0(t) = 1, f^1(t) = t, f^2(t) = t^2 \text{ etc...}$$

Gardons en mémoire que, dans le plan, t représente le couple de coordonnées (x,y) , donc $f^2(t)$ représente les trois fonctions :

$$f^2_1(t) = x^2, f^2_2(t) = xy \text{ et } f^2_3(t) = y^2$$

D'autre part, notons que le nombre L de fonctions de base retenues, lié au degré k du polynôme retenu, donc à la complexité de la dérive modélisée, va restreindre de manière croissante le nombre de combinaison linéaires disponibles pour caractériser la nouvelle variable dZ .

ii) sa covariance ne doit pas dépendre de la position des points t et t' ; on montre alors (P. DELFINER et G. MATHERON, 1980, pages 15 à 17) que cette covariance peut alors s'exprimer par :

$$\begin{aligned} EdZ(t; \lambda^\alpha).dZ(t'; \lambda^\beta) &= \sum_{\alpha} \sum_{\beta} \lambda^\alpha \lambda^\beta K(t_\alpha - t_\beta) + 2 \sum_{\alpha} \lambda^\alpha \lambda^{\alpha_0} K(t - t') \\ &+ 2 \sum_{\beta} \lambda^\beta \lambda^{\beta_0} K(t - t_\beta) + \lambda^{\alpha_0} \lambda^{\beta_0} K(t - t') \quad (I.41) \\ &= \sum_{\alpha'} \sum_{\beta'} \lambda^{\alpha'} \lambda^{\beta'} K(t_{\alpha'} - t_{\beta'}) \end{aligned}$$

(Les notations α' et β' indiquent que les poids λ^{α_0} et λ^{β_0} égaux à 1 et affectant $Z(t)$ et $Z(t')$ dans les dZ sont pris en compte dans les sommations générales en α et β). $K(h)$ est une fonction symétrique appelée covariance généralisée. Cette fonction est définie à un polynôme pair près de degré inférieur ou égal à $2k$ (k degré du polynôme retenu pour les $f^1(t)$, en raison des conditions de non-biais (I.38) vérifiées par les poids λ ; ces conditions imposent en effet aux poids de filtrer tout polynôme de degré inférieur ou égal à k ; on qualifie les combinaisons linéaires obtenues à l'aide de tels poids de combinaisons linéaires autorisées.

De la même manière que pour la covariance simple d'une fonction aléatoire stationnaire, la covariance généralisée doit vérifier des conditions de positivité pour garantir un calcul cohérent de la variance de dZ :

$$EdZ^2(t; \vec{\lambda}) = \sum_{\alpha'} \sum_{\beta'} \lambda^{\alpha'} \lambda^{\beta'} K(t_{\alpha'} - t_{\beta'}) > 0 \quad (I.42)$$

Compte-tenu de la contrainte I.38 imposée au poids $\lambda^{\alpha'}$, il est clair que cette contrainte de positivité sous condition I.38 permettra d'accéder à des familles de modèles de covariance d'autant plus riches que le nombre de fonctions de bases retenu est grand et donc que le nombre de combinaisons linéaires autorisées est restreint. Ainsi la fonction $|h|^\alpha$ avec $\alpha > 0$, n'est jamais un modèle de fonction de covariance simple convenable, alors que

$$K(h) = \sum_{p=0}^k (-1)^{p+1} b_p |h|^{2p+1} \text{ est un modèle de covariance généralisée}$$

acceptable lorsque la dérive est modélisée par un polynôme d'ordre k (voir P. DELFINER et G. MATHERON, 1980, pages 19 à 22).

Lorsque la variable dZ est stationnaire on dit que la fonction Z est une fonction aléatoire intrinsèque d'ordre k , k étant le degré du polynôme modélisant la dérive.

Une fois ce codage très sophistiqué choisi, il convient de remarquer d'abord qu'il répond bien aux préoccupations évoquées en tête de paragraphe ; en effet, la fonction de covariance généralisée caractérise la structure du phénomène d'une manière unique quelle que soit la fonction

polynomiale que l'on rajoute à la fonction Z ; donc, si l'on fait l'hypothèse que la dérive est correctement modélisée par une fonction polynomiale, nous voyons que la fonction de covariance identifiée est insensible à la dérive ; en d'autres termes, cette fonction caractérise les seules fluctuations autour de cette dérive sans avoir explicitement recours à la dichotomie dérive-fluctuations.

Ensuite, nous pouvons noter qu'un tel modèle permet la résolution du système de krigeage ; en effet, la condition de non-biais I.8 est satisfaite par le choix de pondérateurs effectué en I.38 (on dit que la combinaison linéaire dZ ainsi choisie est autorisée) et le calcul de la variance I.9 est directement accessible à l'aide de la covariance généralisée en utilisant I.42.

Le choix de ce modèle pose enfin le délicat problème de l'inférence c'est à dire du calcul pratique des paramètres définissant la covariance généralisée d'un phénomène donné.

INFERENCE DES FONCTIONS ALEATOIRES INTRINSEQUES D'ORDRE k

Rappelons tout d'abord les pratiques d'identification de la fonction de covariance d'une fonction aléatoire stationnaire dans un contexte purement spatial ; deux méthodes sont classiquement employées :

i) l'identification directe de la fonction de covariance point par point pour des classes de distance $h_\alpha \pm \Delta h$ "suffisamment" larges pour rassembler un nombre $N(h_\alpha)$ de couples de points de mesure qui conduit à une estimation de $C(h)$ par la formule :

$$C^*(h_\alpha) = \frac{1}{N(h_\alpha)} \sum_{i,j \in \mathcal{C}_{h_\alpha}} (Z(t_i) - \hat{E}Z(t_i)) \cdot (Z(t_j) - \hat{E}Z(t_j))$$

où \mathcal{C}_{h_α} représente l'ensemble des $N(h_\alpha)$ couples d'indices de points distants de $h_\alpha \pm \Delta h$. Un ajustement d'un modèle autorisé $C(h)$ à la fonction expérimentale $C^*(h)$ est ensuite réalisé en utilisant une méthode automatique plus ou moins sophistiquée ou plus simplement à la main.

ii) l'identification indirecte de la fonction de covariance en utilisant un modèle choisi a priori $C(h ; a, b, \dots)$ et en imposant à ses paramètres de forme (a, b, \dots) de produire dans le cadre d'une validation croisée des erreurs de reconstitutions minimales au sens d'un critère de moindres carrés par exemple (méthode MIE proposée par G. BASTIN et M. GEVERS, 1984 ou T. LEBEL, 1984).

D'un point de vue théorique, l'identification indirecte est la plus séduisante dans la mesure où elle pose clairement la finalité de l'identification (trouver une fonction de structure qui interpole de manière satisfaisante) et où il est possible de montrer (voir G. BASTIN, 1984) que la fonction de covariance identifiée converge, lorsque le nombre de points de mesure grandit, vers la vraie fonction de covariance du processus étudié ; tout le problème est alors de savoir si le nombre de points de mesures pratiquement disponible est suffisant, notamment en hydrologie où les réseaux de mesure sont souvent pauvres dans l'absolu (rarement plus de 100 points) et relativement à la variabilité des phénomènes mesurés.

D'un point de vue pratique, ces deux méthodes employées par T. LEBEL (1984), dans le cas de pluies horaires, l'ont amené à conclure sur l'équivalence des résultats obtenus dans le cas où l'identification directe est facile (ici 2 cas sur 3) ; par contre, la méthode MIE paraît capable de trancher de manière plus nette les cas plus litigieux où la fonction de covariance expérimentale ne laisse pas apparaître clairement le modèle que l'on recherche.

Le cas des covariances généralisées est plus retors au sens où l'identification directe est impossible et l'identification indirecte par une méthode MIE trop lourde. En effet, une identification directe de la fonction $K(h)$ pour A classes de distance $h_\alpha \pm \Delta h$ (α variant de 1 à A) supposerait que l'on soit capable de calculer le système suivant déduit de I.42 :

$$E|d^2Z(t, \lambda)| = E\left(\sum_{i=0}^n \lambda^i(t_0) Z(t_i)\right)^2 = \sum_{i=0}^n \sum_{j=0}^n \lambda^i(t_0) \lambda^j(t_0) K(t_i - t_j)$$

ce qui peut se noter de la manière suivante dans l'optique de l'identification de la fonction $K(h)$ pour chaque point h :

$$\Lambda^2 K = V \quad \text{avec} \quad \Lambda = |\lambda^i(t_0)| \quad i = 1 \text{ à } n \quad \text{et} \quad \alpha = 1 \text{ à } A$$

$$n \times A$$

$$K = |K(h_\alpha)| \quad \alpha = 1 \text{ à } A$$

$$A$$

$$V = \left[\frac{1}{n(\beta)} \sum_{k=1}^{n(\beta)} \left(\sum_{i=1}^n \lambda^i(t_0) Z(t_i) \right)^2 \right] \quad \beta = 1 \text{ à } A$$

$$A$$

la matrice V contenant une estimation expérimentale à partir de $n(\beta)$ cas de la variance des combinaisons linéaires autorisées contenues dans la matrice, ces combinaisons faisant intervenir systématiquement l'ensemble des valeurs de la covariance aux points h_α contenues dans le vecteur K ; il est clair que le calcul du vecteur V repose sur l'existence d'un grand nombre de configurations géométriques des points t_i conduisant à la même combinaison linéaire autorisée ; ce qui dans la pratique d'un réseau de mesures aléatoire et d'une dérive complexe est de probabilité nulle.

D'autre part, la méthode MIE est très efficace lorsque le minimum est recherché dans une dimension, c'est à dire lorsque la fonction de structure dépend d'un seul paramètre ; or, dans le cas d'une covariance généralisée où le degré de dérive et plusieurs paramètres de la fonction de structure sont à prendre en considération, le problème de minimisation se place dans un espace à plusieurs dimensions et devient très ardu.

En fait, les solutions pratiques apportées au problème d'identification de covariance généralisées relèvent d'une approche pragmatique combinant méthode directe et indirecte. La solution la plus développée est certainement celle proposée par J.P. CHILES (1978) et utilisée dans le logiciel Bluepack (Ecole des Mines de Paris). Nous ne détaillerons brièvement ici qu'une version simplifiée de cette solution (J.D. CREUTIN et Ch. OBLED, 1982).

Cette méthode (utilisée dans le programme DEDER de l'IMG) procède en deux étapes :

i) recherche par validation croisée du degré de dérive sous hypothèse de covariance nulle des fluctuations autour de cette dérive ; il s'agit simplement de reconstituer successivement chaque point de mesure à l'aide du reste du réseau (validation croisée) en interpolant par krigeage (système I.10) pour les hypothèses suivantes :

$$C(t-t') = 0 \quad \text{si } t \neq t'$$

$$EZ(t) = \sum_{l=1}^k a_l f^l(t) \quad \text{avec } f^l(t) = t^l \quad (\text{I.43})$$

l'interpolateur ainsi obtenu pour reconstituer la valeur au point t_j peut se noter de la manière suivante, en introduisant I.43 dans la formule I.17 où les a_l et b sont notés α_l et β pour éviter un conflit de notations avec les a_l) :

$$Z^*_j(t) = \sum_{i \neq j} \alpha_i \delta_{t_i, t} + \beta \sum_{l=1}^L a_l t^l \quad \text{avec } \delta_{tt'} = 0 \text{ si } t \neq t' \text{ et } =1 \text{ si } t=t'$$

il s'agit donc d'un polynôme de degré k ajusté par moindres carrés (contrainte d'optimalité) sur les points $t_i \neq t_j$; le calcul des erreurs d'interpolation pour différents degrés k en chaque point t_j :

$$\mathcal{E}_{(k)}(t_j) = Z^*_j(t_j) - Z(t_j) \quad (\text{I.44})$$

permet de choisir le degré k le plus approprié c'est à dire celui qui produit l'erreur moyenne la plus faible au sens des moindres carrés par exemple :

$$\sum_{j=1}^n \mathcal{E}^2_{(k)}(t_j) \quad \text{minimal lorsque } k \text{ varie}$$

ii) une fois choisi le degré de dérive k , il est possible de calculer la variance théorique de l'erreur $\mathcal{E}_{(k)}(t_j)$; en effet, cette erreur est une combinaison linéaire :

$$\mathcal{E}_{(k)}(t_j) = \sum_{i \neq j} \lambda^i(t_j) Z(t_i) - Z(t_j)$$

dont la variance s'exprime en appliquant le même calcul que pour l'équation I.9 :

$$\begin{aligned} E\mathcal{E}_{(k)}^2(t_j) &= \sum_{i \neq j} \sum_{i' \neq j'} \lambda^i(t_j) \lambda^{i'}(t_j) E Z(t_i) \cdot Z(t_{i'}) \\ &\quad - 2 \sum_{i \neq j} \lambda^i(t_j) EZ(t_i) \cdot Z(t_j) + EZ(t_j)^2 \quad (I.45) \\ &= \sum_{i=1}^n \sum_{i'=1}^n \lambda^i(t_j) \lambda^{i'}(t_j) K(t_i - t_{i'}) \quad \text{avec } \lambda^j(t_j) = -1 \end{aligned}$$

Si l'on se donne un modèle de fonction de covariance $K(h ; \alpha, \beta, \dots)$ dépendant de différents paramètres (α, β, \dots) , il est alors possible de rechercher les paramètres qui permettent de retrouver au mieux les erreurs expérimentales I.44 à l'aide du calcul théorique I.45, au sens des moindres carrés par exemple :

$$\sum_{j=1}^n \frac{1}{P_j} \mathcal{E}_{(k)}^2(t_j) - E\mathcal{E}_{(k)}^2(t_g))^2 \quad \text{minimal} \quad (I.46)$$

lorsque (α, β, \dots) varient ; les poids P_j sont destinés à normer ces écarts (variance expérimentale-variance théorique), dans la mesure où leur valeur est tributaire de la configuration des points explicatifs $(t_i \neq t_j)$; idéalement cette normalisation doit être effectuée par la variance théorique $E\mathcal{E}_{(k)}^2(t_j)$; malheureusement cela conduit à une optimisation non linéaire et l'on préfère normer par le terme prépondérant de la variance théorique ; de manière plus générale, cette recherche de minima sera d'autant plus simple que le nombre de paramètres (α, β, \dots) est faible et surtout qu'ils interviennent linéairement ; par exemple (voir J.D. CREUTIN et Ch. OBLED, 1982) une covariance linéaire avec constante (cette constante traduit la présence d'un bruit blanc que les géostatisticiens appelle couramment effet de pépite) :

$$K(h) = \alpha + \beta h$$

conduit à minimiser selon I.46 l'expression suivante :

$$\sum_{j=1}^n \frac{1}{S'_i} (\epsilon^{2(k)}(t_j) - \alpha S_i - \beta S'_i)^2$$

$$\text{avec } S_i = \sum_{i=1}^n \sum_{i'=1}^n \lambda^i(t_j) \cdot \lambda^{i'}(t_j) \text{ et } S'_i = \sum_{i=1}^n \sum_{i'=1}^n \lambda^i(t_j) \cdot \lambda^{i'}(t_j) h_{ii},$$

Cette méthode d'identification de covariance généralisée est bien entendu sous-optimale puisque d'une part le degré k de dérive n'est pas remis en cause une fois les paramètres (α, β, \dots) calculés et d'autre part le critère minimisé I.46 ne porte pas sur l'erreur d'estimation optimale de la variable mais sur l'erreur d'estimation de la variance d'une estimation non-optimale (par simple calage de dérive).

Notons enfin que la méthode d'identification contenue dans le logiciel Bluepack (option RECO) diffère de cette version simplifiée par une application de la validation croisée non plus à des points mais à des ensembles de points ou auréoles concentriques : l'auréole externe étant utilisée pour reconstituer les points de l'auréole interne et réciroquement ; cette façon de faire donne une plus grande robustesse à la méthode et est appliquée à des fonctions de covariance plus complexes (polynômes jusqu'au degré 5 par exemple) ; cette façon de faire sous-entend aussi une plus grande complexité de programmation.

On mesure donc ici la difficulté du critère a de choix d'un modèle évoqué plus haut (possibilité de réaliser l'inférence) lorsque la complexité, la richesse du modèle devient grande. Nous verrons par la suite que, dans la plupart des applications à l'hydrométéorologie, les réseaux de mesure sol sont trop restreints pour mettre rentablement en oeuvre des modèles de fonctions aléatoires intrinsèques d'ordre supérieur à 0.

FONCTIONS ALEATOIRES INTRINSEQUES D'ORDRE 0 : LE VARIOGRAMME.

Une fonction aléatoire intrinsèque d'ordre 0 est une fonction aléatoire dont les différences premières sont stationnaires ; pour reprendre les notations du paragraphe précédent, le codage d'une telle

fonction pourra se faire de manière très élémentaire suivant I.37 en retenant une seule valeur de λ_α égale à -1 quel que soit le point t' :

$$dZ(t; \bar{\lambda}) = Z(t) - Z(t')$$

L'hypothèse de valeur moyenne des dZ constante I.38 se traduisant par :

$$EZ(t) - EZ(t') = 0 \quad \text{pour tout } t'$$

correspond bien à une hypothèse de dérive constante $EZ(t) = m$; ensuite il est intéressant de noter que la covariance de dZ peut s'écrire d'après I.41:

$$1/2 E(Z(t) - Z(t'))^2 = K(0) - K(t-t')$$

en choisissant $dZ(t; \bar{\lambda}) = dZ(t'; \bar{\lambda}') = Z(t) - Z(t')$; cette fonction de structure $\gamma'(t-t') = 1/2 E(Z(t)-Z(t'))^2$, appelée variogramme (par G.MATHERON, 1965 mais appelée aussi fonction de structure par L.S. GANDIN, 1965 ou courbe de variation par G. JOWETT, 1955) peut alors se substituer à $-K(t-t')$ car elle n'en diffère que d'une constante, c'est à dire d'un polynôme pair de degré inférieur ou égal à $2k$ (ici 0).

Les avantages procurés par le variogramme sont de deux ordres :

i) par rapport à la covariance classique $C(t-t')$, le variogramme possède l'avantage d'être "modélisable" à l'aide d'une famille plus riche de fonctions ; en effet, $\gamma'(t-t')$ est une combinaison de covariances généralisées d'ordre 0, $K(t-t')$ et peut donc par exemple être modélisé par $\gamma'(h) = \alpha h^\beta$ avec $0 < \beta < 2$ qui n'est jamais un modèle de fonction de covariance simple convenable ; en d'autres termes le variogramme n'est pas nécessairement borné comme la covariance, il est seulement soumis à une condition de croissance à l'infini ($\lim_{|h| \rightarrow \infty} \gamma'(h)/|h|^2 = 0$) qui se rattache à des conditions plus générales (voir G. CHRISTAKOS, 1984) ; par contre, lorsque le variogramme a un comportement borné ($\lim_{h \rightarrow 0} \gamma'(h)/h = 0$), nous sommes dans le cas où la fonction aléatoire intrinsèque d'ordre 0 considérée possède une représentation stationnaire, c'est à dire que la covariance généralisée $K(h)$ telle que $\lim_{h \rightarrow \infty} K(h) = 0$ est une covariance

simple ; alors le variogramme tend vers une limite finie : la variance de la fonction considérée $K(0)$.

ii) par rapport à une covariance généralisée, le variogramme possède l'avantage de pouvoir être inféré directement comme une covariance simple :

$$\hat{\gamma}^*(h_\alpha) = \frac{1}{2N(h_\alpha)} \sum_{i,j \in \mathcal{C}(h_\alpha)} (Z(t_i) - Z(t_j))^2 \quad (\text{I.47})$$

Si l'on ajoute à ces avantages le fait que, localement au moins (voir paragraphe suivant), l'hypothèse de moyenne constante mais non nulle s'applique avec satisfaction à un grand nombre de phénomènes, on comprend pourquoi ce modèle de fonction aléatoire et cette fonction de structure sont plus couramment utilisés en géostatistique. (L'annexe I.B décrit les modifications qu'entraîne l'utilisation du variogramme dans le système de krigeage).

QUASI-STATIONNARITE : LE KRIGEAGE EN VOISINAGE GLISSANT

Pour terminer ces rappels de géostatistique, il est utile de rappeler la possibilité, dans la pratique, de réduire le domaine d'application des hypothèses et du système de krigeage au "voisinage" du point estimé.

Cette restriction est simplificatrice à deux niveaux :

- elle évite souvent le recours à des hypothèses sophistiquées concernant la dérive dans la mesure où, comme nous l'avons évoqué ci-dessus, l'hypothèse de moyenne constante est souvent très satisfaisante dans un domaine peu étendu, ce qui peut se noter :

$$E(Z(t) - Z(t+h)) = m(t) - m(t+h) \quad 0 \quad \text{si } h < r$$

où r est la dimension caractéristique du voisinage retenu ; l'utilisation du variogramme comme fonction de structure s'avère alors satisfaisante puisque le variogramme brut :

$$2.\hat{\gamma}(t,t+h) = E(Y(t)-Y(t+h))^2 + (m(t)-m(t+h))^2 \approx E(Y(t)-Y(t+h))^2 \quad \text{si } h < r$$

est alors très proche du variogramme des résidus ou variogramme vrai du phénomène, dans la gamme de distance qui nous intéresse ; enfin, il est possible, lorsque la variabilité du phénomène varie lentement dans l'espace, de modifier le modèle en fonction du lieu considéré :

$$\hat{\gamma}(t,t+h) = v(t) \gamma'(h) \quad \text{pour } h < r$$

où v est une fonction lentement variable.

- elle allège le système de krigeage en réduisant le nombre de points de mesure à prendre en compte ; cette réduction revient en fait à l'introduction de contraintes supplémentaires dans le système (annulation de certains poids - voir annexe I.C) dont l'effet est très acceptable dans le cas d'une bonne corrélation spatiale puisque le système complet accorde alors la quasi-totalité des poids significatifs aux points de mesure entourant le point reconstitué et des poids très faibles aux autres points (effet d'écran) ; elle est également justifiée dans le cas d'un phénomène mal structuré puisqu'elle substitue à l'estimation de chaque point par la moyenne globale du champ, une estimation par moyenne locale souvent plus sensée.

Tout le problème de l'application d'une telle restriction réside en fait dans le choix de la taille du voisinage à considérer ; ce choix est un compromis entre le nombre et la répartition des points de mesure à utiliser et l'idée que l'on a de la portée du phénomène : il paraît raisonnable en effet de traiter un nombre minimal de points de mesure et d'éviter l'utilisation de points trop mal corrélés avec le point estimé.

Enfin, une limite de ce travail en voisinage glissant est de ne pas permettre le krigeage dans son conditionnement interpolateur.

CONCLUSION

Nous avons examiné comment peut être traité dans le cadre de la théorie des fonctions aléatoires le problème de la reconstitution dans l'espace d'un phénomène mesuré ponctuellement.

En imposant à un estimateur linéaire des contraintes de non-biais et d'optimalité, nous avons défini le système de krigeage. Suivant le contexte climatologique ou spatial pur dans lequel est utilisé ce système, il a été possible d'allier aux hypothèses probabilistes de structure des hypothèses de nature plus déterministe concernant la dérive afin de disposer d'une gamme de modèles susceptibles de représenter des types de phénomènes très variés.

L'aspect le plus restrictif de ces modèles demeure néanmoins l'hypothèse d'homogénéité, parfaitement incontournable dans un contexte purement spatial, mais qu'une connaissance climatologique du phénomène peut permettre d'assouplir.

ANNEXE I.A

**CALCUL DE LA COVARIANCE CENTREE DES ESTIMATIONS
DE L'ESPERANCE PONCTUELLE D'UNE FONCTION
ALEATOIRE STATIONNAIRE**

Nous noterons $\hat{E}Z(t)$ l'estimation de l'espérance de la fonction aléatoire stationnaire Z au point t qui s'obtient à partir de p réalisations observées à l'aide de la formule :

$$\hat{E}Z(t) = \frac{1}{p} \sum_{w=1}^p Z_w(t) \quad (A1)$$

où $Z_w(t)$ représente la $w^{\text{ème}}$ mesure disponible en t .

Nous noterons $EZ(t)$ la valeur de l'espérance en t , à savoir :

$$EZ(t) = \int z dF(z)$$

où F est la fonction de répartition de $Z(t)$.

La covariance centrée des \hat{E} est donnée par :

$$C_{\hat{E}}(t-t') = E(\hat{E}Z(t) - EZ(t)) \cdot (\hat{E}Z(t') - EZ(t')) \quad (A2)$$

puisque $E(\hat{E}Z(t)) = EZ_w(t) = EZ(t)$.

D'après (A1) et (A2), on a :

$$\begin{aligned} C_{\hat{E}}(t-t') &= E \left[\frac{1}{p} \left(\sum_w Z_w(t) - EZ_w(t) \right) \cdot \frac{1}{p} \left(\sum_{w'} Z_{w'}(t') - EZ_{w'}(t') \right) \right] \\ &= \frac{1}{p^2} \sum_w \sum_{w'} E(Z_w(t) - EZ_w(t)) \cdot (Z_{w'}(t') - EZ_{w'}(t')) \quad (A3) \end{aligned}$$

Si l'on suppose que les différentes réalisations prises en considération sont indépendantes, alors le terme général de la somme dans (A3) vaut $C(t-t')$ ou 0 selon que $w=w'$ ou $w \neq w'$. En définitive, il vient :

$$C_E^{\wedge}(t-t') = \frac{1}{p} C(t-t')$$

ANNEXE I.B*KRIGEAGE A L'AIDE DU VARIOGRAMME*

La substitution du variogramme à la fonction de covariance dans le système de krigeage simple conduit à de petites modifications qui sont parfois source d'erreurs lors de la programmation de cette technique au moment du calcul des variances d'estimation.

Le système I.10 peut se noter dans le cadre de l'hypothèse intrinsèque d'ordre 0 à l'aide d'une fonction de covariance :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i_C(t_0) C(t_i, t_j) + C = C(t_j, t_0) \text{ pour } j = 1 \text{ à } m \\ \sum_{i=1}^n \lambda^i_C(t_0) = 1 \end{array} \right.$$

avec pour variance d'estimation $V_C = C(t_0, t_0) - \sum_{i=1}^n \lambda^i_C(t_0) C(t_i, t_0) - \mu_C$

ou bien d'un variogramme :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i_{\gamma}(t_0) \gamma(t_i, t_j) + \mu_{\gamma} = \gamma(t_j, t_0) \text{ pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda^i_{\gamma}(t_0) = 1 \end{array} \right. \quad (B1)$$

Comme le variogramme peut se noter $\gamma(t, t') = C(0) - C(t, t')$, le système B1 peut donc se transformer, en remarquant que $\sum_i \lambda^i_{\gamma}(t_0) \cdot C(0) = C(0)$:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i_{\gamma}(t_0) C(t_i, t_j) - \mu_{\gamma} = C(t_j, t_0) \text{ pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda^i_{\gamma}(t_0) = 1 \end{array} \right.$$

avec pour variance d'estimation $V_{\gamma} = C(t_0, t_0) - \sum_{i=1}^n \lambda_{\gamma}^i(t_0) C(t_i, t_0) + \mu_{\gamma}$

qui peut également s'écrire :

$$V_{\gamma} = \sum_{i=1}^n \lambda_{\gamma}^i(t_0) \cdot \hat{\gamma}(t_i, t_0) + \mu_{\gamma}$$

d'où :

$$V_{\gamma} = \hat{\gamma}(t_0, t_0) + \sum_{i=1}^n \lambda_{\gamma}^i(t_0) \cdot \hat{\gamma}(t_i, t_0) + \mu_{\gamma}$$

ANNEXE I.C

*ANNULATION DE FACTEURS DE PONDERATION
ET KRIGEAGE EN VOISINAGE GLISSANT*

Dans le développement du système de krigeage, il est possible d'ajouter aux contraintes I.8 et I.9 de non-biais et d'optimalité d'autres contraintes visant par exemple à annuler certains facteurs $\lambda_i(t_0)$.

L'introduction de ces contraintes supplémentaires que l'on peut formuler par :

$$M_k = \lambda^k(t_0) = 0 \quad \text{pour } K \text{ valeurs de } k \text{ choisies entre } 1 \text{ et } n$$

va nécessiter l'introduction de paramètres de Lagrange supplémentaires μ_k et la minimisation de V sous les contraintes M et $(M_k, k \text{ parmi } 1 \text{ à } n)$

se fera en annulant les dérivées partielles de $V + 2\mu M + 2 \sum_{k=1}^K \mu_k M_k$ par rapport aux différents λ et μ . On obtient alors au lieu du système I.10 :

$$\left. \begin{aligned} & \sum_{i=1}^n \lambda_{V}^i(t_0) EZ(t_i) \cdot Z(t_j) + \mu_{V}(t_0) EZ(t_j) = EZ(t_j) Z(t_0) \quad \text{pour } j \neq k \\ & \sum_{i=1}^n \lambda_{V}^i(t_0) EZ(t_i) Z(t_k) + \mu_{V}(t_0) EZ(t_k) + \mu_k = EZ(t_k) Z(t_0) \quad \text{pour } k \text{ choisis} \\ & \sum_{i=1}^n \lambda_{V}^i(t_0) EZ(t_i) = EZ(t_0) \\ & \lambda_{V}^k(t_0) = 0 \end{aligned} \right\} \quad (C1)$$

avec pour nouvelle variance d'estimation :

$$V_V = EZ(t_0) \cdot Z(t_0) - \sum_{i \neq k}^n \lambda_{V}^i(t_0) EZ(t_i) \cdot Z(t_0) - \mu_{V}(t_0) \cdot EZ(t_0)$$

qui peut se mettre sous la forme :

$$V_V = V - \sum_{i=k}^n (\lambda_V^i(t_0) - \lambda^i(t_0))EZ(t_0)Z(t_i) - H(\mu_V(t_0) - \mu(t_0)).EZ(t_0) \quad (C2)$$

les indices v des λ et μ distinguent ces coefficients après introduction des contraintes supplémentaires ; il en est de même pour V).

Cette formulation assez lourde du système de krigeage a pour mérite de formaliser l'effet de l'abandon de certains points de mesure ; la cartographie des quantités $(V - V_V)$ ou $(1 - V_V/V)$ tirées de la formule C2 peut servir à visualiser la perte d'information entraînée ; en d'autres termes, lorsque le voisinage choisi et la fonction de covariance utilisée sont tels que les poids $k(t_0)$ sont déjà presque nul en krigeage complet, il est évident que l'estimateur obtenu sur un voisinage a les mêmes caractéristiques qu'en krigeage complet ($\lambda_V^i(t_0) \approx \lambda^i(t_0)$, $\lambda_V(t_0) \approx \lambda(t_0)$ et donc $V_V \approx V$). Dans certains cas particuliers tels que le "bruit blanc" ($EZ(t_i).Z(t_0) = 0$ si $t_0 \neq t_i$), il est facile de montrer, en remplaçant $\mu(t_0)$ et $\mu_V(t_0)$ par leurs valeurs tirées de I.10 et (C1) :

$$\mu(t_0) = -EZ(t_0) / \sum_{j=1}^n (E^2Z(t_j) / E(Z(t_j)^2))$$

$$\mu_V(t_0) = -EZ(t_0) / \sum_{j \neq k} (E^2Z(t_j) / EZ(t_j)^2))$$

que cette différence de variance d'estimation vaut :

$$\begin{aligned} V_V - V &= (\mu(t_0) - \mu_V(t_0)) EZ(t_0) \\ &= (EZ(t_0))^2 \cdot (1 / \sum_{j \neq k} (EZ(t_j)^2 / E^2Z(t_j)) - 1 / \sum_{j=1}^n (EZ(t_j)^2 / E^2Z(t_j))) \end{aligned}$$

Elle est donc toujours positive (ou nulle lorsque $K=0$). Cette constatation rappelle simplement que dans le cas d'un bruit blanc, l'estimation par krigeage se réduit à celle de la dérive qui sera d'autant plus robuste que le nombre de points de mesure est grand ; le krigeage complet est alors toujours préférable à un krigeage en voisinage glissant.

CHAPITRE II

VALIDATION D'UN DISPOSITIF DE MESURE

Résumé

Ce chapitre propose une démarche pour valider les résultats fournis par un dispositif de mesure. Cette démarche nécessite d'une part le choix de mesures de référence supposées représenter les vraies valeurs de phénomène et d'autre part le choix de critères de validité permettant d'apprécier la ressemblance des résultats validés avec les valeurs de référence.

Les choix proposés concernent la validation d'un dispositif de mesure à forte résolution spatiale, fournissant des valeurs intégrées du phénomène (par exemple un capteur de télédétection étalonné ou non) à l'aide d'un dispositif de mesures ponctuelles de bonne précision (par exemple un réseau sol d'appareils de mesure directe).

Le procédé de validation ainsi défini sera utilisé dans les chapitres suivants comme outil de comparaison et d'évaluation.

1. CHOIX DES MESURES DE REFERENCE

La première étape pour valider une série de mesures d'un phénomène donné est de choisir une série de valeurs, supposée représenter les vraies valeurs du phénomène, afin de les utiliser par la suite comme valeurs de référence.

Ce choix doit satisfaire au moins trois impératifs :

i) les valeurs de référence doivent être "aussi précises que possible" ; bien qu'il soit illusoire d'imaginer avoir accès aux vraies valeurs du phénomène, il est en effet indispensable de choisir une série de valeurs de référence aussi proche que possible de cette vérité et de

pouvoir caractériser cette distance référence-vérité (par une analyse statistique de l'erreur par exemple) ;

ii) les valeurs de référence doivent être comparables aux valeurs testées ; un premier aspect de cette comparabilité concerne évidemment la nature même des données comparées : les mesures indirectes (réflectivité radar ou pourcentage de différents types de nébulosité vus par les satellites) doivent être transformées en valeurs correspondantes du phénomène étudié (hauteur de pluie au sol par exemple) ; un deuxième aspect concerne le volume d'espace échantillonné qui peut varier suivant le dispositif de mesure : une réflectivité radar traduit le contenu en pluie d'un volume approximativement tronconique alors qu'une mesure pluviométrique au sol mesure le contenu en pluie d'un cylindre vertical de volume infiniment plus petit ; un troisième aspect concerne le mode d'échantillonnage dans le temps qui peut être continu (cas de la mesure sol des hauteurs de précipitation par exemple) ou discret (cas des images de télédétection) ;

iii) la source des valeurs de référence doit être "aussi indépendante que possible" de la source des valeurs testées ; l'indépendance ainsi recherchée est en fait celle des erreurs commises sur les valeurs comparées ; en effet, nous venons de voir en i) que les erreurs sur les valeurs de référence devaient être aussi petites que possible, mais il paraît également nécessaire de s'assurer que ces erreurs ne confluent pas avec celles que nous cherchons à identifier sur les valeurs testées ; une bonne garantie de cette indépendance semble être l'utilisation de dispositifs de mesure différents qui ont généralement des sources d'erreur différentes (exemple de la validation d'un capteur de télédétection à l'aide d'un réseau de capteurs sol) ; malheureusement, d'une part, cette garantie n'est pas absolue (par exemple les régions montagneuses où l'identification des pourcentages de nébulosité par images satellite est perturbée par la présence de la neige sont aussi des régions où les effets de masque gênent la mesure héliographique) et d'autre part, certains contextes expérimentaux (validation de mesures de télédétection étalonnées) ne permettent pas de séparer les sources d'information.

Ces impératifs nous ont conduits à choisir comme référence les mesures sol des phénomènes étudiés intégrées par krigeage sur des surfaces identiques à celles des éléments d'images de télédétection correspondante. Les éléments sensibles de ce choix sont évidemment la nature des mesures utilisées et la méthode employée pour les mettre en forme.

L'utilisation des mesures sol pour établir une référence est essentiellement justifiée par la nature directe de ces mesures : qu'il s'agisse de pluie ou de durée d'insolation il est évident que, ponctuellement au moins, les mesures réalisées par des appareils de mesure au sol (pluviomètre, héliographe) procurent une précision satisfaisante (de l'ordre de quelques pourcents -voir J.C. RODDA, 1967). Ce choix est très souvent proposé dans la littérature (WILSON, 1970 pour les radars ou C.G. GRIFFITH, 1986 pour les images satellites), mais le caractère ponctuel de ces mesures rend leur utilisation directe impossible lorsqu'il s'agit de valider des mesures de télédétection dont le volume d'échantillonnage est plus grand et la résolution spatiale plus forte.

Il est alors nécessaire de mettre ces mesures en forme comparable mesures validées par intégration spatiale ; cette intégration peut se trouver simplifiée lorsque le réseau sol est de très forte densité : c'est le cas de certaines expériences de pluviométrie menées en Floride (WOODLEY, 1975 et HILDEBRAND, 1979), où une simple moyenne arithmétique conduit à des résultats satisfaisants comme le montrent différents essais de validation croisée ; devant des conditions expérimentales moins favorables, d'autres auteurs (HARROLD, 1974 par exemple) ont introduit dans leur valeur de référence l'information radar qu'ils souhaitaient valider : en recalant l'allure générale du phénomène perçue par le radar (dont ils veulent par la suite établir le bien-fondé des mesures) à l'aide des mesures ponctuelles -ils établissent en fait une référence dont les erreurs sont très dépendantes de celles de la mesure validée. Malgré leurs défauts ces deux démarches soulignent bien les difficultés de mise en forme des mesures sol; nous avons essayé de les résoudre par krigeage.

L'utilisation du krigeage pour régionaliser l'information sol apporte des éléments de réponse aux différents impératifs fixés :

i) les valeurs de référence obtenues sont d'une précision optimale au sens où certains critères sont vérifiés par l'estimateur calculé (voir Chapitre I en I.6 et I.7) ; de plus, compte-tenu de la configuration du réseau et de la variabilité du phénomène, la variance d'estimation de la valeur de référence (voir expression I.11) fournit un indicateur de la distance référence-vérité ; cette distance, comparée à la distance valeur de référence-valeur validée, permet de juger de l'intérêt de la validation.

ii) la comparabilité des valeurs de référence à des mesures échantillonnées dans des volumes différents est assurée par la possibilité d'obtenir des estimateurs intégrés sur une surface quelconque (voir formule I.12) ; d'autre part, seules des mesures directes du phénomène étant utilisées, la comparabilité concernant la nature des mesures et leur mode d'échantillonnage dans le temps devra être assurée par un traitement adéquat des valeurs validées ;

iii) l'indépendance de la source des valeurs de référence ainsi obtenue est bien entendu satisfaisante lorsque la source des valeurs validées est un autre dispositif de mesure (radar ou satellite par exemple) ; nous examinerons dans le paragraphe 3, consacré aux protocoles de validation, les précautions que l'on peut prendre pour préserver au mieux cette indépendance lorsque les valeurs validées contiennent une partie de l'information sol dans le cas d'un étalonnage.

Le krigeage apparaît donc capable de fournir une bonne estimation des valeurs de référence en même temps que des éléments (variance d'estimation, structure du phénomène) qui permettent de juger de la qualité de la validation entreprise.

2. CHOIX DES CRITERES DE VALIDITE

Pour juger la précision d'un dispositif de mesure, il est nécessaire d'apprécier la ressemblance entre les résultats qu'il fournit et la référence que l'on s'est fixée (voir paragraphe précédent). Le critère de ressemblance choisi peut être très variable et va dépendre d'un ensemble de facteurs très divers (propriétés des valeurs de référence disponibles ,

qualités que l'on veut plus précisément tester dans ce dispositif, etc...); nous aurons recours par la suite à trois types de critères :

i) - un critère qualitatif fondé sur la comparaison des cartes de la valeur de référence et de la mesure testée ; purement subjectif, cet examen permet de contrôler la cohérence des mesures fournies avec la référence en ayant simultanément sous les yeux leurs distributions spatiales respectives ; il est ainsi possible de mettre en évidence certains défauts majeurs comme des décalages dans l'espace et des tendances systématiques propres à certaines caractéristiques géographiques (effet d'éloignement de l'appareil de mesure, présence d'aberrations dans certaines régions, etc...). L'inconvénient majeur de ce critère résulte de la difficulté d'introduire dans cette comparaison la précision des mesures de référence ; en effet, il s'agit de pondérer les conclusions tirées de l'examen des cartes représentant la valeur du phénomène, par les renseignements tirés d'une troisième carte de précision de la référence (variance d'estimation par exemple).

ii) une famille de critères quantitatifs expérimentaux fondés sur la définition d'indices de ressemblance dont le calcul est reproductible donc objectif ; nous distinguerons, dans cette famille, les indices dont la valeur dépend à la fois de la ressemblance en moyenne et en variabilité des valeurs comparées (erreurs) des indices dont la valeur ne dépend que la covariation spatiale ou temporelle des valeurs comparées (corrélation).

Avant d'examiner plus en détail ces indices de similarité, nous évoquons pour mémoire le troisième type de critère qui, à la différence des deux premiers, ne nécessite pas la disponibilité d'un ensemble de valeurs de référence.

iii) un critère "pronostic" qui repose sur une modélisation du phénomène et donne, a priori, une estimation de l'erreur commise ; il est clair que ce type de critère n'est disponible que si des mesures directes de la variable estimée ont été prises en considération (étalonnage d'un capteur de télédétection par exemple).

Nous verrons, dans le chapitre suivant, quelles méthodes d'étalonnage peuvent fournir ce type de critère ; pour donner un exemple immédiatement compréhensible, citons le cas de l'interpolation des seules mesures sol : le comportement de la fonction de structure expérimentale et la valeur des variances d'estimation donnent, a priori et sous hypothèse d'une modélisation correcte de la réalité, une idée de la précision de ce type d'estimation ; il est bien entendu possible de vérifier cette précision de manière expérimentale à l'aide de critères du type ii) appliqués en des points de mesure de référence, c'est à dire n'ayant pas servi à l'interpolation (voir J.D. CREUTIN, 1979). L'avantage majeur de ce type de critère est, bien entendu, de pouvoir fonctionner sans nécessiter un ensemble de valeurs de référence ; sa faiblesse est de recourir à un modèle dont les hypothèses sont difficiles à vérifier.

LES ERREURS

Sous le terme d'erreurs sont en fait regroupés de critères de ressemblance appartenant à la famille ii) des critères quantitatifs. Pour clarifier l'examen des propriétés de ces différents critères, gardons en mémoire qu'ils reposent tous sur une combinaison des trois choix suivants :

i) *la variable qui sert à caractériser la ressemblance entre une valeur testée z^* et une valeur de référence z^{\wedge}* : ce premier choix est très important car il doit prendre tout d'abord en considération les caractéristiques des dispositifs de mesure utilisés : lorsque ces dispositifs produisent des erreurs indépendantes de la valeur du phénomène (cas de la mesure des durées d'insolation par exemple), il y aura lieu de choisir une variable de ressemblance basée sur l'écart simple $z^* - z^{\wedge}$, alors que si les erreurs sont fonction croissante de la valeur du phénomène (cas de la pluviométrie), il sera meilleur de choisir une variable de ressemblance sur un écart relatif $(z^* - z^{\wedge}) / z^{\wedge}$ ou plus simplement le rapport z^* / z^{\wedge} ; ce choix permet également de prendre en compte le sens de l'erreur commise en travaillant sur les valeurs relatives de ces écarts, ou bien de ne pas le faire en utilisant les valeurs absolues des écarts simples ou bien le plus grand des rapports z^* / z^{\wedge} ou z^{\wedge} / z^* ; enfin, ce choix peut servir à pénaliser de manière préférentielle les fortes erreurs par exemple en employant le carré des écarts simples.

ii) *la manière d'échantillonner cette variable* : dans l'espace, afin de mesurer l'erreur commise lors d'un événement ou bien dans le temps afin de mesurer l'erreur commise en un site donné ou bien sur l'ensemble des données disponibles pour qualifier globalement une méthode ; nous ne reviendrons pas systématiquement sur la nature de ce choix dans la mesure où les critères évoqués peuvent s'appliquer dans ces différents contextes.

iii) *les caractéristiques retenues pour résumer le comportement d'un échantillon* : valeur moyenne notée à l'aide d'une barre — (il s'agit d'une moyenne empirique et non d'une espérance) ou bien histogramme de distribution des valeurs par exemple.

Il est donc possible de regrouper ces différents critères de contrôle d'erreur dans un tableau à trois entrées comme le montre la table I.1 dans laquelle sont repérés les critères les plus usuels :

- la moyenne simple des écarts n'est sensible qu'aux biais systématiques puisqu'elle est égale à la différence des moyennes des deux séries de valeurs ; en effet, nous avons :

$$e_1 = \overline{z^* - z^\wedge} = \overline{z^*} - \overline{z^\wedge}$$

- la moyenne quadratique des écarts reflète également le biais mais aussi la covariation des valeurs comparées ; cette moyenne se note :

$$e_3 = \overline{(z^* - z^\wedge)^2} = \overline{z^{*2}} + \overline{z^{\wedge 2}} - 2 \overline{z^* \cdot z^\wedge}$$

qui peut se transformer en décomposant z^\wedge et z^* en somme de valeurs moyennes z^\wedge et z^* et de résidus y^\wedge et y^* de moyenne nulle :

$$e_3 = \overline{(y^* - y^\wedge)^2} + \overline{(z^* - z^\wedge)^2} \quad (\text{II.1})$$

Le premier terme sera d'autant plus faible que la covariation des valeurs centrées $y^\wedge \cdot y^*$ sera forte ; le deuxième terme sera d'autant plus faible que le biais sera faible.

Caratéristique statistique	Moyenne	Variance	Distribution
Variable de ressemblance			
Ecart $z^{\wedge}-z^*$	Erreur moyenne $e_1 = z^{\wedge}-z^*$	Erreur quadratique centrée $e_4 = e_3 - e_1^2$	Histogramme des erreurs
Valeur absolue d'écart $ z^{\wedge}-z^* $	$e_2 = (y^{\wedge}-y^*)^2 - e_1^2$		
Ecart relatif $\frac{z^{\wedge}-z^*}{z^{\wedge}}$			
Rapport maximum ou facteur de différence $\text{Max}(z^{\wedge}/z^*, z^*/z^{\wedge})$			Histogramme des facteurs de différence

Echantillonnage par événement

Echantillonnage par site

Echantillonnage global

Table II.1 : Critères quantitatifs d'erreurs usuels classés suivant les choix de variable de ressemblance, d'échantillonnage de cette variable et des caractéristiques statistiques analysées.

- le comportement de la moyenne des valeurs absolues d'écart $e_2 = |z^{\wedge}-z^*|$, plus difficile à expliciter en moyennes élémentaires, est elle aussi sensible au biais et à la covariation des valeurs comparées, sans "amplifier" les fortes erreurs comme le fait l'élévation au carré. Un calcul simple dans l'hypothèse où les écarts suivent une loi normale montre, que dans le cas d'un biais nul ($Ez = Ez^*$, la moyenne e_2 est

directement liée à l'écart-type des écarts ($e_2 = 2\sigma/\sqrt{2\pi}$) ; lorsque le biais augmente la figure II.1 montre que l'erreur e_2 tend vers e_1 .

- la variance des écarts fournit des renseignements très proches de ceux de la moyenne quadratique puisqu'elle peut se mettre sous la forme:

$$\begin{aligned} e_4 &= \overline{(z^{\wedge}-z^* - (z^{\wedge}-z^*))^2} \\ &= \overline{(z^{\wedge}-z^*)^2} - \overline{(z^{\wedge}-z^*)^2} = e_3 - e_1^2 \end{aligned}$$

Ces différentes erreurs moyennes conduisent à une appréciation globale de la ressemblance entre deux séries de mesures qui peut paraître trop synthétique. Afin d'analyser de manière plus fine les erreurs commises, il est possible d'en examiner la distribution en construisant un histogramme :

- l'histogramme des écarts permet de visualiser des disymétries éventuelles de la distribution que ne reflètent pas la moyenne et la variance des écarts ;

- l'histogramme cumulé des facteurs de différence :

$$FD = \max (z^*/z^{\wedge}, z^{\wedge}/z^*)$$

introduit par HARROLD et al. (1974) possède des propriétés liées à la variable rapport et à l'opérateur "max" utilisé ; le rapport permet de travailler sur une variable sans dimension, et a donc, une signification tout à fait identique à l'erreur relative (puisque $(z^*-z^{\wedge})/z^{\wedge} = z^*/z^{\wedge}-1$) qui admet une proportionalité entre l'erreur et la valeur de la variable ; l'opérateur max, appliqué à ces rapports a le même effet sur la distribution fréquentielle que l'opérateur valeur absolue appliquée aux différences (en effet : $\text{Log} (\max (z^*/z^{\wedge}, z^{\wedge}/z^*)) = |\text{Log } z^* - \text{Log } z^{\wedge}|$) ; dans l'hypothèse d'une distribution symétrique et centrée en 1, il rend plus robuste l'identification de cette distribution en mélangeant les deux demi-distributions ($y^*/y^{\wedge} < 1$ et $y^*/y^{\wedge} > 1$) ; cette pratique peut être très ambiguë si la distribution n'est pas symétrique ou si elle n'est pas centrée en 1.

Les différentes possibilités de la table II.1 ne sont pas toutes explorées dans la pratique ; nous constatons que seules les franges de la table, c'est à dire les critères les plus simples que nous venons de

Figure II.1 : Illustration schématique du calcul de la moyenne des valeurs absolues des écarts $z^* - z^{\wedge}$ dans l'hypothèse d'une distribution gaussienne de leurs valeurs ; (a) dans le cas du non-biais et (b) dans le cas d'un biais.

(a) si $\Pr(z^* - z^{\wedge} > u) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{x^2}{2\sigma^2}} dx$ alors $e_2 = E|z^* - z^{\wedge}| = \frac{2}{\sigma \sqrt{2\pi}} \int_0^{\infty} x e^{-\frac{x^2}{2\sigma^2}} dx = \frac{2\sigma}{\sqrt{2\pi}}$

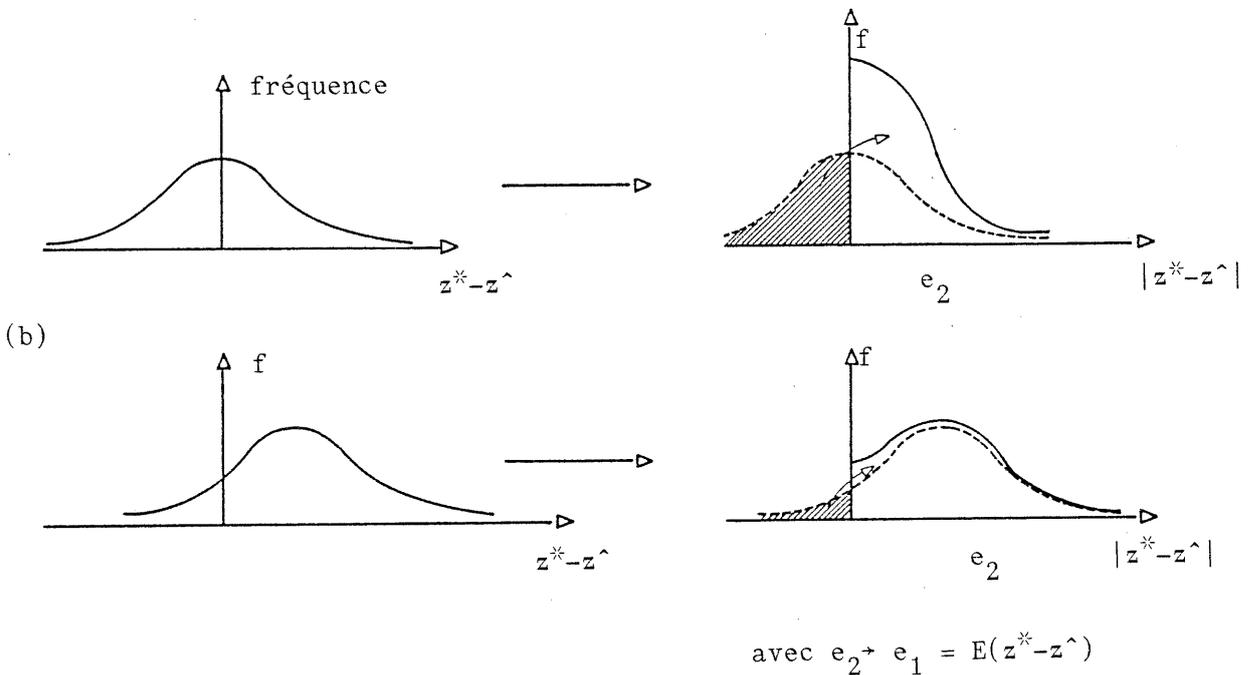
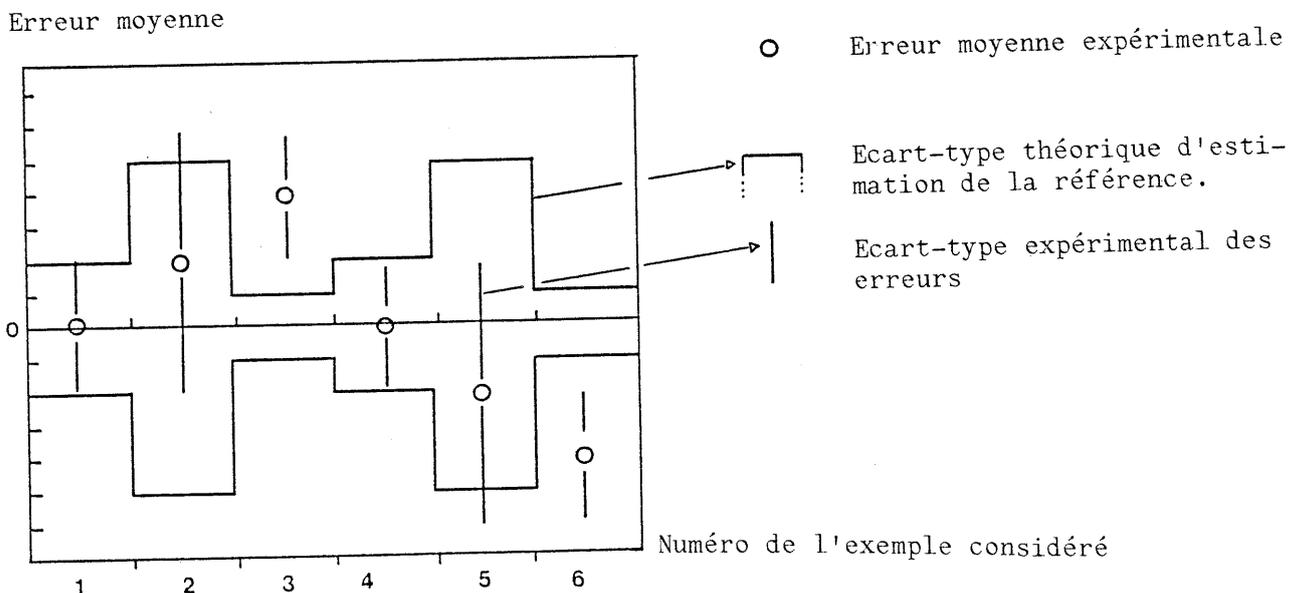


Figure II.2 : Exemple de diagramme d'efficacité permettant de juger de l'identité (1 et 4), de la différence non significative (2 et 5) ou de la différence significative (3 et 6) de deux ensembles de valeurs.



décrire, sont employées couramment. Le déplacement dans la troisième dimension, celle de l'échantillonnage, dépend du contexte expérimental et de la nature plus ou moins synthétique du résultat que l'on recherche :

- lorsque l'on veut mettre en évidence la capacité du dispositif testé à apprécier l'évolution spatiale du phénomène mesuré, il est naturel de caractériser l'erreur commise à propos de chaque événement (les moyennes — sont alors prises spatialement afin notamment de déceler d'éventuelles liaisons entre l'erreur commise et le type d'événement (différence de précision de la mesure d'un radar pour des situations météorologiques différentes par exemple) ; mais, il est évident que la valeur des critères d'erreur dépend, en général, des caractéristiques statistiques de l'événement traité ; il peut être alors judicieux d'opérer une normalisation de ces critères d'erreur, c'est à dire de les rendre adimensionnels donc plus aisément comparables ; on les divise pour cela par la moyenne ou la variance des valeurs de référence ; on obtient, par exemple :

$$e'_1 = \frac{e_1}{z} = 1 - \frac{\overline{z^*}}{\overline{z^{\wedge}}}$$

ou bien :

$$e'_3 = \frac{e_3}{z^{\wedge 2} - (\overline{z^{\wedge}})^2} = 1 + \frac{(\overline{z^{\wedge}})^2 + \overline{z^{*2}} - 2 \overline{z^{\wedge} z^*}}{z^{\wedge 2} - (\overline{z^{\wedge}})^2}$$

qui s'annulent respectivement dans le cas du non-biais ($e'_1=0$ si $\overline{z^{\wedge}}=\overline{z^*}$) et de la covariation totale ($e'_3=0$ si $\overline{z^{\wedge 2}} = \overline{z^{*2}} = \overline{z^{\wedge} z^*}$).

- lorsqu'une série temporelle de mesures suffisamment longue est disponible, il est intéressant de caractériser l'erreur commise en chaque site afin de déceler d'éventuelles régionalisations de ces erreurs (évolution de la précision de la mesure radar en fonction de la distance du point estimé au radar par exemple) ; les mêmes remarques que précédemment, concernant la normalisation des erreurs, peuvent être appliquées ; les valeurs moyennes — seront alors prises dans le temps au lieu de l'espace.

- enfin, il peut être utile pour qualifier un dispositif de résumer globalement sa précision en analysant l'ensemble des mesures disponibles en différents lieux pour différents événements ; les valeurs moyennes — sont alors calculées dans le temps et l'espace et la normalisation éventuelle des erreurs devient plus délicate dans la mesure où il faut alors soustraire ces indices au double effet des caractéristiques spatiales de chaque événement et temporelles de chaque lieu ; pour des raisons de simplicité nous n'opèrerons pas une telle normalisation, sur laquelle nous reviendrons à propos des calculs de corrélation.

Parmi les critères d'erreurs évoqués ci-dessus, nous attacherons, dans la suite, une importance prépondérante aux critères fondés sur les écarts simples :

- moyenne et histogramme, dans la mesure où ils reflètent un biais éventuel de manière directe et simple ;

- variance, dans la mesure où elle peut être comparable à la variance d'estimation fournie par le krigeage (voir plus loin une comparaison entre ces deux types de variances).

Pour des raisons de commodité de présentation, ces critères sont rassemblés sur un seul diagramme (d'après G. DELRIEU, 1986) que nous appelons diagramme d'efficacité (voir exemple, figure II.2) ; ce type de diagramme, en présentant simultanément la dispersion de la vraie valeur du phénomène et des mesures validées autour de la référence, permet de juger de l'importance des erreurs.

Nous éviterons d'utiliser des critères d'erreurs ambigus, dans la mesure où ils sont sensibles à la fois au biais et à la covariation des valeurs comparées (moyenne des valeurs absolues, moyenne quadratique, distribution des facteurs de différence) ; nous nous limiterons à les produire, dans quelques cas, puisqu'ils sont d'un usage courant dans la littérature, mais nous leur préfèrerons les critères de corrélation.

LA CORRELATION EMPIRIQUE

Afin d'illustrer le passage des critères d'erreurs aux critères de corrélation, il est plus imagé de raisonner en termes de géométrie affine.

Considérons l'espace affine à n-dimensions repéré par rapport aux n sites de mesure (dans le cas d'un échantillonnage par événement) ou bien par rapport aux n événements (dans le cas d'un échantillonnage en un site); dans cet espace l'ensemble des valeurs testées est représenté par un point P^* dont les coordonnées sont les valeurs produites par le dispositif testé en chaque site (respectivement pour chaque événement) ; il en est de même pour l'ensemble des valeurs de référence qui, idéalement, devrait être représenté par un point P confondu avec le premier.

Dans cet espace, la moyenne des erreurs quadratiques e_3 représente le carré de la distance euclidienne entre ces deux points ; la variance des écarts e_4 représente le carré de la distance euclidienne entre les transformées de ces points par une translation de vecteur de coordonnées constante \hat{z} et z^* respectivement ; l'erreur e_2 représente elle aussi une distance dans une métrique non euclidienne.

Le défaut majeur de ces critères est d'être sensibles à l'échelle choisie pour exprimer \hat{z} et z^* , donc d'être sensibles à des facteurs multiplicatifs appliqués à \hat{z} et z^* ; une manière efficace de contourner cette difficulté est de ne plus raisonner en terme de distance mais en terme d'angle.

L'utilisation des rapports z^*/\hat{z} ou des facteurs de différence constitue un premier pas dans cette direction dans la mesure où le rapport z^*/\hat{z} représente une tangente dans l'espace à deux dimensions repéré par rapport aux axes des valeurs test et de référence ; ce critère est bien insensible à un changement d'échelle commun aux deux variables \hat{z} et z^* , donc au choix des unités de mesure, mais il ne permet pas d'opérer un changement d'échelle approprié à chaque variable pour en normer la variabilité.

Nous préférons utiliser l'angle à l'origine des points \tilde{P} et \tilde{P}^* dans l'espace à n dimensions précédemment utilisé. Ces points sont les analogues des points P et P^* pour les valeurs centrées de \hat{z} et z^* (respectivement \hat{y} et y^*). La relation trigonométrique classique appliquée au triangle $(O, \tilde{P}, \tilde{P}^*)$ nous donne :

$$\overline{\tilde{P}\tilde{P}^*}^2 = \overline{O\tilde{P}}^2 + \overline{O\tilde{P}^*}^2 - 2\overline{O\tilde{P}}\overline{O\tilde{P}^*} \cos \widehat{PO\tilde{P}^*}$$

ce qui peut se noter sous la forme :

$$\overline{(\hat{y}-y^*)}^2 = \overline{\hat{y}}^2 - 2\sqrt{\overline{\hat{y}}^2} \cdot \sqrt{\overline{y^*}^2} \cos \widehat{PO\tilde{P}^*} + \overline{y^*}^2$$

où \hat{y} et y^* sont les valeurs centrées de \hat{z} et z^* ; or, la variance des écarts $\overline{(\hat{y}-y^*)}^2$ peut s'écrire :

$$\overline{(\hat{y}-y^*)}^2 = \overline{\hat{y}}^2 - 2\overline{\hat{y}y^*} + \overline{y^*}^2$$

le critère caractérisant la ressemblance des valeurs z et z^* en terme d'angle sera donc :

$$\cos \widehat{PO\tilde{P}^*} = \frac{\overline{\hat{y}y^*}}{\sqrt{\overline{\hat{y}}^2 \cdot \overline{y^*}^2}}$$

qui est le coefficient de corrélation empirique classique. Ce critère atteint bien l'objectif fixé d'être insensible à toute transformation linéaire opérée sur \hat{z} et z^* ; en effet, une transformation additive est corrigée par l'opération de centrage en conduisant simplement au choix d'une origine O différente ; une transformation multiplicative à facteur positif sur l'une ou l'autre des variables ne modifie pas l'angle $\widehat{PO\tilde{P}^*}$; par contre une transformation multiplicative à facteur négatif sur l'une des variables change le signe du critère (en modifiant l'angle $\widehat{PO\tilde{P}^*}$ de π) ; sans pour autant changer sa signification en terme de covariation : ceci conduit à travailler sur la valeur absolue du cosinus ou sur son carré (ce dernier choix est plus simple à interpréter d'un point de vue statistique puisque le carré du coefficient de corrélation exprime un pourcentage de variance expliquée).

Rappelons deux pièges classiques dans lesquels ce type de critère peut tomber :

i) la présence sur un axe de l'espace à n dimensions de deux valeurs \hat{y} et y^* exceptionnellement fortes et d'écart faible peut diminuer l'angle $\tilde{\text{POP}}^*$ d'une manière artificiellement forte qui ne préjuge en rien de la covariation sur les autres axes dont la valeur est obtenue par projection de $\tilde{\text{POP}}^*$ sur l'hyperplan orthogonal à l'axe en question (voir schéma en figure II.3).

ii) plus généralement, l'existence d'une "famille" d'axes de l'espace à n dimensions pour laquelle les moyennes des valeurs centrées \hat{y} et y^* sont significativement différentes de zéro, peut, par le même mécanisme que ci-dessus, diminuer fortement la valeur de l'angle $\tilde{\text{POP}}^*$; dans l'hypothèse où "cette famille" est identifiable (par exemple l'ensemble des observations d'une même réalisation, dans le cas où ce critère est utilisé pour qualifier globalement un dispositif tous événements et tous sites confondus) il est légitime et souhaitable de corriger cet effet systématique en opérant un deuxième centrage des \hat{y} et y^* par rapport aux valeurs moyennes sur les "familles" ($\hat{x} = \hat{y} - \bar{y}^w$ et $x^* = y^* - \bar{y}^{*w}$ où $\bar{\quad}^w$ représente la moyenne prise sur la réalisation w par exemple).

La corrélation constitue donc un critère dont les propriétés sont très satisfaisantes pour juger la qualité d'un dispositif de mesure en terme de "variabilité" ; suivant le contexte expérimental et la nature des caractéristiques que l'on veut mettre en évidence nous disposerons pour un dispositif validé d'un seul coefficient (résumant globalement les mesures disponibles aux différents sites pour différents événements) ou bien d'une série de coefficients (calculés pour différents événements à l'aide de l'ensemble des sites ou bien réciproquement). Se pose alors le problème de la comparaison de deux séries de coefficients lorsque l'on veut comparer les performances de deux dispositifs de mesure. Cette comparaison peut évidemment se faire événement par événement (ou bien site par site) mais lorsque le nombre de couples de coefficients est grand il est difficile de conclure ; il est préférable dans ce cas de considérer la courbe de distribution des coefficients de corrélation ou plutôt de leurs valeurs

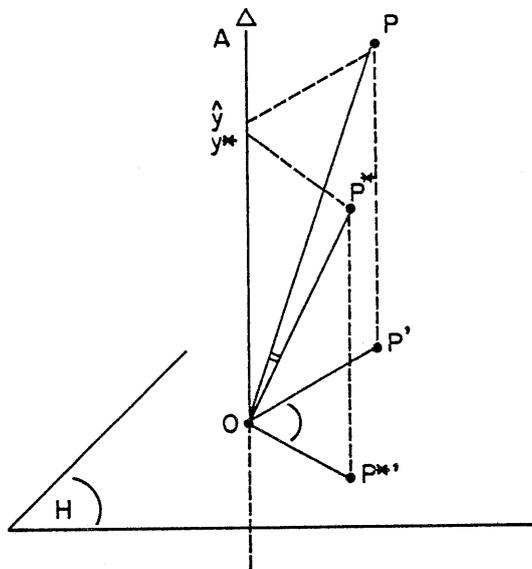


Figure II.3 : Schématisation de l'effet de valeurs exceptionnellement fortes sur un axe A.

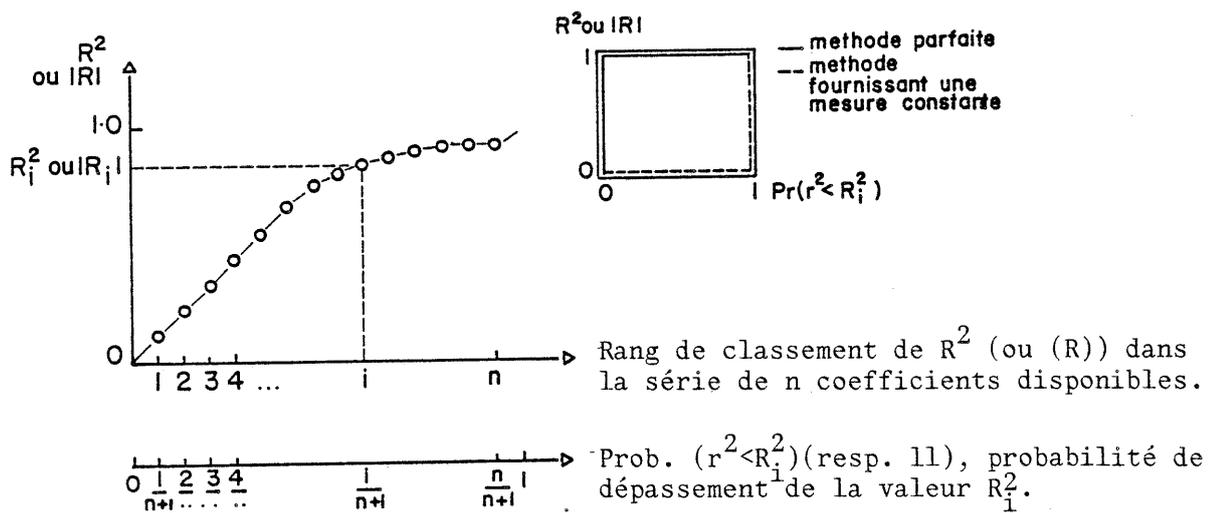


Figure II.4 : Exemples de courbes d'efficacité.

absolues ou de leurs carrés afin d'éviter le problème du signe du coefficient évoqué ci-dessus ; cette courbe (que nous appellerons "courbe d'efficacité" comme dans J.D. CREUTIN, 1979 et J.D. CREUTIN et Ch. OBLED, 1982) s'obtient en classant par ordre croissant les valeurs absolues ou bien les carrés des coefficients de corrélation ; dans la schématisation proposée, figure II.4, un dispositif sera d'autant plus efficace que sa courbe est située plus haut ; lorsque les valeurs classées sont les carrés des coefficients de corrélation l'aire située sous la courbe peut s'interpréter comme le pourcentage de la variance des valeurs de référence expliqué par les valeurs validées.

3. PROTOCOLES DE VALIDATION RETENUS

Les modalités pratiques d'application des "outils de base" définis ci-dessus pour établir une référence et la comparer aux mesures validées vont dépendre de la nature des mesures validées. En effet, nous avons fait le choix de calculer les valeurs de référence à l'aide des mesures réalisées par un réseau d'appareils de mesure au sol (pluviomètres, héliographes) en raison de la nature directe de ces mesures: ce choix est parfaitement légitime lorsque les mesures validées sont produites par un dispositif différent (capteur de télédétection) puisque les erreurs commises sur la référence sont d'une part d'origines généralement différentes de celles que nous cherchons à identifier dans les mesures validées et d'autre part d'amplitude identifiable si la méthode de calcul des valeurs de référence utilisée est bien choisie ; mais ce choix, nous l'avons vu, peut être discutable lorsque les mesures validées sont produites en utilisant, pour partie, le type de dispositif servant à établir la référence (cas de l'étalonnage) puisque nous n'avons plus de garantie sur l'indépendance des erreurs commises sur les mesures validées d'une part et de référence d'autre part. Nous distinguerons donc le cas des dispositifs non-étalonnés de celui des dispositifs étalonnés lors de la définition d'un protocole.

Cas des dispositifs de mesure non étalonnés

Ce cas de figure inclut de manière franche le cas des mesures de la hauteur de pluie par télédétection radar dans la mesure où l'on applique

une relation réflectivité-intensité standard (c'est à dire établie par ailleurs, sans utiliser les mesures sols correspondant aux événements considérés) et de la même manière, nous considèrerons comme faisant partie de ce cas, les mesures de fraction journalière d'insolation issues de l'analyse d'images satellites par la méthode CACTUS.

Par contre, il est difficile de ranger dans cette catégorie la mesure des hauteurs de pluie par néphanalyse d'images satellites dans la mesure où la régression nécessaire pour lier les valeurs brutes issues des images à la variable d'intérêt n'a pas pu être établie à propos d'événements distincts de ceux que nous traitons ; cet empêchement résulte d'un manque de données.

Le protocole de validation retenu dans ces exemples est simple car nous considérons que l'utilisation de dispositifs de mesure différents est une garantie suffisante d'indépendance des erreurs commises sur les valeurs de référence d'une part et validées d'autre part : nous calculerons donc les valeurs de référence à l'aide de l'ensemble des mesures sol disponibles comme indiqué au paragraphe 1 et nous les comparerons aux valeurs validées en utilisant les critères définis au paragraphe 2.

Suivant les applications envisagées il est possible de travailler sur différents types de cibles (ou surfaces d'intégration) :

i) dans cette étude nous avons choisi des cibles élémentaires correspondant à un élément d'image, afin d'apprécier la covariation des mesures comparées dans l'espace ; nous avons imposé à ces cibles de contenir un point de mesure sol afin de ne retenir que les cibles où l'erreur commise sur les valeurs de référence est a priori le plus faible (la mention a priori prévient le cas défavorable des phénomènes à structure pépitique pure ou bruits blancs pour lesquels la variance d'estimation est partout égale à la variance du champ). Ce type de cible a été utilisé dans le cas des mesures de pluie par radar et de durée quotidienne d'insolation par satellite pour lesquels nous disposions de réseaux sol suffisamment denses.

ii) des cibles de dimension plus grande, correspondant à l'agrégation de plusieurs éléments d'image peuvent également être choisies afin de mesurer l'effet de l'intégration sur la qualité des mesures. Cette démarche, liée par exemple à la préoccupation hydrologique de mesurer des lames d'eau sur des bassins versants, a été entreprise par G. DELRIEU, 1986, à propos du radar.

Cas des dispositifs de mesure étalonnés

Ce cas de figure regroupe les ensembles de mesures validées obtenues, au moins pour partie, à l'aide de type de dispositif servant à établir la référence ; il s'agit bien entendu des mesures radar étalonnées à l'aide d'un réseau héliographique ; il s'agit aussi des résultats de néphanalyse d'images satellites interprétées en hauteur de pluie par régression entre différents pourcentages de nébulosité et des mesures de pluie au sol.

Le protocole de validation retenu doit alors assurer au mieux l'indépendance des erreurs commises sur les valeurs comparées. La précaution la plus élémentaire et la plus efficace que l'on puisse prendre est de faire une distinction nette entre un réseau sol d'étalonnage et un réseau sol de validation, en d'autres mots, d'éviter la double utilisation d'un appareil de mesure afin i) d'étalonner le dispositif testé et ii) d'établir la référence. Cette précaution garantit l'indépendance des erreurs lorsque celles-ci sont propres à la situation géographique d'un appareil ou à ses conditions d'exploitation ; par contre, cette précaution est sans effet pour des erreurs liées aux caractéristiques du type d'appareil utilisé pour étalonner et établir la référence (par exemple les erreurs de sous-estimation commises par les pluviomètres pour certains types de pluie) ; nous serons cependant obligés d'accepter ce risque de biais en l'absence de précaution alternative ou complémentaire.

Cette séparation du réseau sol en réseau de validation et réseau d'étalonnage est bien adaptée à une validation ponctuelle (qui peut-être légitime dans le cas de la validation de technique d'interpolation de mesures ponctuelles - J.D. CREUTIN, 1979). Par contre lors d'une validation sur des cibles, nécessitée par la nature intégrée des mesures validées

(mesures de télédétection principalement), le calcul des valeurs de référence passe par l'intégration des mesures ponctuelles du seul réseau de validation ; ce réseau étant, par définition, moins dense que l'ensemble du réseau disponible, on peut craindre une réduction sensible de la précision de la référence. Un compromis entre l'indépendance des sources d'information et la précision de la référence peut être obtenu, dans le cas de cibles élémentaires, en ne retenant pour cibles de comparaison que celles qui contiennent une station de validation ; le compromis est plus difficile à trouver dans le cas de grandes cibles (voir G. DELRIEU, 1986, pages 145 et suivantes).

Les résultats de validation obtenues à l'aide des protocoles définis ci-dessus permettent de répondre dans un premier temps à la question "*Quelle est la précision d'un dispositif étalonné ou bien non étalonné ?*" Cette appréciation absolue d'un dispositif de mesure donné revêt en fait un intérêt mineur aux yeux d'un utilisateur potentiel de ce dispositif qui posera plutôt la question "*Qu'apporte l'utilisation d'un dispositif de télédétection par rapport à un réseau de mesure conventionnel ?*" ou bien "*Qu'apporte l'étalonnage ?*" afin d'apprécier de manière relative les performances de chaque dispositif. Nous présenterons donc toujours les critères de validité obtenus avant et après étalonnage (pour répondre à la deuxième question) ainsi que les critères obtenus pour la reconstitution des valeurs de référence à l'aide des seuls appareils d'étalonnage (pour répondre à la première question). Pour être comparables, ces critères doivent être obtenus suivant un protocole identique qui sera toujours le plus contraignant : celui des dispositifs étalonnés qui distingue les sous-réseaux de validation et d'étalonnage.

Cette présentation "comparative" s'applique aux différents critères de validité définis au paragraphe 2 ; elle peut être complétée par une application plus nuancée du critère de corrélation : *la corrélation partielle*.

En effet, si l'on veut répondre de manière synthétique à la question "*Qu'apporte un dispositif de télédétection par rapport à un réseau de mesure conventionnel ?*" , il est tout d'abord nécessaire de calculer le coefficient de corrélation totale entre la valeur de référence z^{\wedge} , les

valeurs fournies par le dispositif de télédétection éventuellement étalonné z^* , et par le seul réseau d'étalonnage $z^{\wedge*}$, ce coefficient de corrélation totale $r_{z^{\wedge}, z^* z^{\wedge*}}$ pouvant s'exprimer en pourcentage de la variance de z^{\wedge} expliquée par z^* et $z^{\wedge*}$:

$$r_{z^{\wedge}, z^* z^{\wedge*}}^2 = \frac{r_{z^{\wedge}, z^{\wedge*}}^2 + r_{z^{\wedge}, z^*}^2 - 2 r_{z^*, z^*} \cdot r_{z^{\wedge}, z^{\wedge*}} \cdot r_{z^{\wedge}, z^*}}{1 - r_{z^{\wedge*}, z^*}^2}$$

où la notation $r_{x,y}$ représente la corrélation simple entre les variables x et y au sens où nous l'avons définie au paragraphe précédent. Le pourcentage de la variance de z^{\wedge} déjà expliquée par le seul réseau conventionnel étant donné par $r_{z^{\wedge}, z^* z^{\wedge*}}^2$, le gain obtenu en pourcentage de variance de z^{\wedge} expliquée grâce à l'utilisation du dispositif de télédétection et à son éventuel étalonnage se calcule par :

$$dr^2 = r_{z^{\wedge}, z^* z^{\wedge*}}^2 - r_{z^{\wedge}, z^{\wedge*}}^2$$

ce gain est une estimation absolue de l'apport recherché par la question posée ci-dessus. Il est possible de donner une expression relative de ce gain en considérant qu'il sera d'autant plus significatif que le pourcentage de variance $r_{z^{\wedge}, z^{\wedge*}}^2$ est fort, c'est à dire que z^{\wedge} est déjà bien expliqué par $z^{\wedge*}$ seul ; on obtient alors le coefficient de corrélation partielle classique :

$$r_{z^{\wedge} z^*, z^{\wedge*}}^2 = \frac{dr^2}{1 - r_{z^{\wedge}, z^{\wedge*}}^2}$$

qui donne en fait la corrélation simple des résidus de z^{\wedge} sachant z^* avec les résidus de z^* sachant $z^{\wedge*}$.

Ces critères de gains absolu et relatif peuvent bien sûr être utilisés pour répondre à d'autres questions : apport d'un réseau d'étalonnage ($r_{z^{\wedge} z^*, z_T}^2$ si l'on note z_T la mesure fournie par le dispositif de télédétection non étalonné) par exemple.

4. RETOUR SUR LES CRITERES DE VALIDITE : CRITERE THEORIQUE ET CRITERE EXPERIMENTAL.

Après la description des protocoles de validation, il est possible de revenir sur une question évoquée lors de la description des critères de validité : comme nous l'avons souligné au paragraphe II.2 la comparaison de la variance des écarts entre la valeur de référence et la valeur testée avec la variance d'estimation de la valeur de référence, c'est à dire la variance des écarts référence-vérité, détermine l'intérêt de la validation entreprise ; en effet, la validation n'aura de sens que si la dispersion de la vraie valeur du phénomène autour de la référence choisie est significativement plus faible que la dispersion de la valeur testée autour de la référence.

D'autre part, comme la variance d'estimation de la valeur de référence varie d'un site à l'autre en fonction de la configuration du réseau de mesure, il peut paraître séduisant d'accorder à chaque site un poids fonction de cette variance afin de privilégier lors du calcul des critères d'erreurs ou de corrélation les sites pour lesquels la valeur de référence est la plus sûre. Une manière simple d'introduire cette pondération sera de diviser pour chaque site t_j les valeurs comparées ($z^{\wedge}(t_j)$ et $z^*(t_j)$) par l'écart-type d'estimation de z^{\wedge} au point t_j (formule I.11) ; ainsi nous utilisons localement des unités de mesure qui dépendent de la précision de la mesure de référence.

La comparaison et l'utilisation conjointe de critères d'origine différente apparaît donc très utile, mais il est nécessaire de se poser les questions de savoir i) *si une telle comparaison est légitime* et ii) *s'il est possible de l'éviter en utilisant des critères de nature comparable.*

i) La première question est la plus délicate puisqu'elle pose de manière générale le problème de la validité du modèle utilisé pour calculer la variance d'estimation théorique, et singulièrement celui de l'inférence de la fonction de structure.

Nous avons déjà écarté au chapitre I une première réponse possible qui consiste à mettre en oeuvre des tests statistiques de validité

de la fonction de structure ; en effet, cette fonction portant sur des moments d'ordre 2 nécessite pour sa validation de calculer des moments d'ordre 4 dont l'accès expérimental est exclu dans notre contexte où les données issues des réseaux sol sont trop peu nombreuses.

La réponse la plus satisfaisante pour juger de la qualité d'un modèle est la validation croisée qui consiste à comparer aux n points de mesure t_i la valeur effectivement mesurée $z(t_i)$ et la valeur estimée $z^{*-t_i}(t_i)$ à l'aide du réseau privé du point t_i). Evidemment, se repose alors le problème de la comparaison de deux séries de mesure : l'analyse de la distribution des écarts $z(t_i) - z^{*-t_i}(t_i)$ normés par l'écart-type d'estimation de z^{*-t_i} en t_i constitue un bon critère (solution retenue dans le programme Bluepack de l'ENSMP).

Malheureusement, cette validation ne peut pas porter sur les configurations de réseau ayant servi à établir les valeurs de référence où les valeurs estimées sont très proches des points de mesure (les cibles choisies contiennent une station de validation).

La démarche pour justifier la comparaison de critères théoriques et expérimentaux peut donc être résumée comme suit :

- une étape de validation croisée destinée à justifier le bien fondé du modèle choisi qui devra fournir des variances d'estimation expérimentales proches des variance théoriques annoncées ;

- une étape de comparaison entre variance expérimentale et variance d'estimation où l'on s'appuie sur le bien fondé du modèle pour justifier le mélange de critères de nature différente.

ii) Deux solutions sont a priori possibles pour éviter la comparaison de critère de nature différente :

- la première qui consiste à ne travailler qu'avec des critères expérimentaux est écartée d'une part du fait qu'il est impossible, comme nous l'avons mentionné ci-dessus, de retrouver expérimentalement des configurations de réseau permettant de valider des estimations réalisées

très près des points de mesure ; d'autre part une variance d'estimation expérimentale n'a de sens que ponctuellement, ce qui exclut le cas des cibles.

- la deuxième solution consiste à n'utiliser que des critères théoriques ; considérons par exemple la variance des écarts valeurs de référence-valeur validée décrite au paragraphe II.2 :

$$e_4 = E(\hat{Z} - Z^* - E(\hat{Z} - Z^*))^2$$

qui peut se noter en remplaçant \hat{Z} par sa forme linéaire $\sum_i \lambda_i(T)Z(t_i)$, et en considérant une cible T de la manière suivante :

$$e_4 = E(Z^*(T) - EZ^*(T) - \sum_i \lambda_i(T)(Z(t_i) - EZ(t_i)))^2$$

soit en développant :

$$\begin{aligned} e_4 &= E(Z^*(T) - EZ^*(T))^2 + \sum_i \sum_j \lambda_i(T) \lambda_j(T) E(Z(t_i) - EZ(t_i)) \cdot (Z(t_j) - EZ(t_j)) \\ &\quad - 2 \sum_i \lambda_i(T) E(Z^*(T) - EZ^*(T)) \cdot (Z(t_i) - EZ(t_i)) \\ &= C_{Z^*Z^*}(T-T) + \sum_i \sum_j \lambda_i(T) \lambda_j(T) C_{ZZ}(t_i - t_j) - 2 \sum_i \lambda_i(T) C_{ZZ^*}(t_i - T) \end{aligned}$$

où les fonctions $C_{Z^*Z^*}$ et C_{ZZ} sont les fonctions de covariance, décrites au chapitre I, établies respectivement à l'aide de la variable validée Z^* et de la variable de référence Z suivant des modalités propre à la nature intégrée ou non du signal considérée (voir chapitre IV) ; où la fonction C_{ZZ^*} représente la fonction de covariance liant ces deux variables, que nous appelons covariance croisée et dont nous parlerons plus en détail également au chapitre IV.

Lorsque \hat{Z} est le résultat d'une estimation par krigeage, le deuxième terme de la somme e_4 peut s'écrire :

$$\begin{aligned} \sum_i \sum_j \lambda_i(T) \lambda_j(T) C_{ZZ}(t_i - t_j) &= \sum_i \lambda_i(T) (C_{ZZ}(t_i - T) - \rho EZ(T)) \\ &= \sum_i \lambda_i(T) C_{ZZ}(t_i - T) - \rho EZ(T) \\ &= C_{ZZ}(T-T) - (V+2\rho EZ(T)) \end{aligned}$$

où V représente la variance d'estimation de $Z^*(T)$; ce terme fluctue entre 0 et $C_{ZZ}(T-T)$ (puisque $V < C_{ZZ}(T-T)$) ; sa valeur maximale correspond à une reconstitution parfaite du signal de validation sur la cible T ; une valeur nulle traduit l'inadéquation totale du réseau de validation utilisé. Ainsi la variance e_4 apparaît comme la somme d'un terme caractérisant la variabilité du signal validé et d'un terme caractérisant la variabilité du signal de validation effectivement prise en compte par son réseau de mesure, somme à laquelle on retranche une mesure de la co-fluctuation des deux signaux. Bien sûr, cette estimation de la variance e_4 dépend de la modélisation des fonctions de covariance et appelle le même type de validation que celui évoqué en i). D'autre part plusieurs "lectures" de cette variance sont possibles suivant le contexte dans lequel elle est calculée ; lorsque Z^* est le résultat d'un dispositif de mesure distinct du réseau de validation (mesures de télédétection par exemple) et lorsque les cibles choisies T sont des éléments d'image du dispositif validé e_4 s'écrit sous la forme :

$$e_4 = C_{Z^*Z^*}(T-T) + C_{ZZ}(T-T) - V - 2 \sum_i \lambda_i^1(T) C_{ZZ^*}(t_i-T)$$

où $C_{Z^*Z^*}(T-T')$ et $C_{ZZ^*}(t-T)$ sont les fonctions de covariance expérimentales ; nous verrons au chapitre IV la formulation de e_4 lorsque Z^* est un résultat d'étalonnage.

Il est maintenant nécessaire de mettre en pratique l'ensemble de la démarche que nous venons de présenter pour valider un dispositif de mesure : nous le faisons dans un premier temps au chapitre suivant à propos de dispositifs de télédétection non étalonnés et de réseaux de mesure sol.

CHAPITRE III

EXEMPLES DE VALIDATIONS DE DISPOSITIFS DE MESURE
NON ETALONNESRésumé

Ce chapitre propose plusieurs exemples de validation qui illustrent la démarche préconisée au chapitre précédent.

Ces exemples concernent des mesures de précipitations et d'insolation à l'aide de mesures sol et de télédétection.

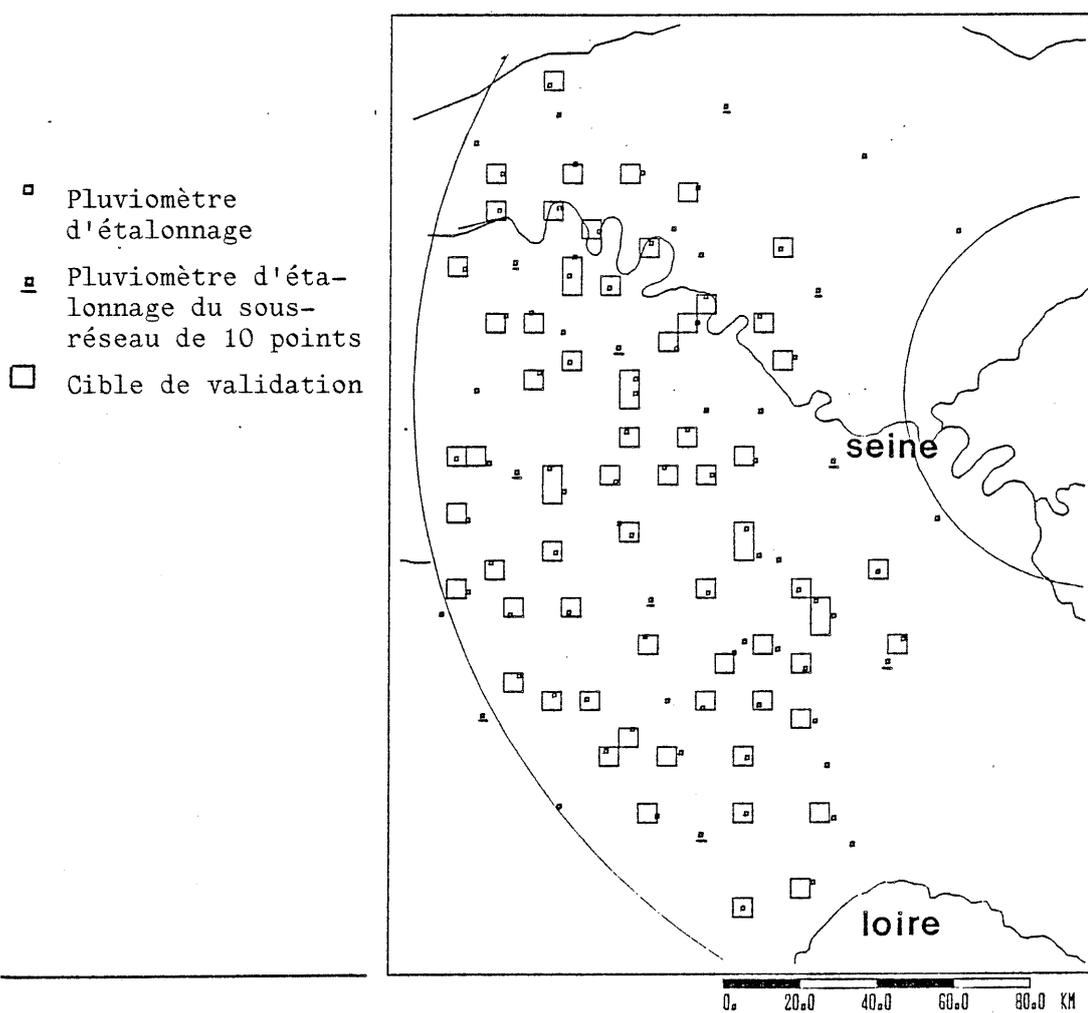
Pour chacun de ces phénomènes une partie du réseau de stations sol est utilisée pour établir des valeurs de référence sur des cibles de validation. Ensuite, ces valeurs de référence sont comparées aux estimations obtenues à l'aide de la partie restante du réseau : cette première validation permet d'une part d'apprécier l'apport d'un réseau sol et d'autre part, de tester expérimentalement la cohérence de ce dispositif de mesure. Enfin, les valeurs de référence servent à la validation des mesures fournies par des dispositifs de télédétection (radar météorologique et méthode CACTUS) ; cette seconde validation permet de situer les performances d'un dispositif de télédétection par rapport aux moyens plus conventionnels de mesure au sol.

1. RAPPEL

Comme nous l'avons évoqué au chapitre précédent, il est possible d'utiliser les dispositifs de mesure sol pour établir une valeur de référence mais aussi pour les valider eux-mêmes. En effet, ils sont formés, au contraire des appareils de télédétection, par la juxtaposition d'un ensemble d'appareils de mesure généralement indépendants que nous pouvons arbitrairement scinder en deux sous-ensembles :

- un réseau de validation, dont l'ensemble de mesures doit être suffisamment complet pour fournir une valeur de référence acceptable ;

Figure III.1 : Carte du réseau de 98 pluviomètres ; les stations de validation sont incluses dans une maille radar (69 points) ; les stations d'étalonnage (29 points) sont soulignées lorsqu'elles appartiennent au sous-réseau de 10 pluviomètres.



- un réseau que nous qualifions d'étalonnage par référence à l'usage que nous en ferons par la suite, qui dans le cas présent nous servira de réseau de base, auquel nous appliquerons une méthode d'interpolation (le krigeage) afin de fournir des estimations des valeurs du phénomène aux points de validation.

Nous l'avons dit également, il est possible de prendre pour valeur de référence les valeurs mesurées au points de référence lorsqu'il s'agit de valider des méthodes d'interpolation ; nous considèrerons ici ce travail comme déjà accompli (J.D. CREUTIN, 1979 et J.D. CREUTIN, 1983) et nous utilisons les résultats acquis, notamment quant au choix de la technique la plus performante (krigeage), pour préparer les validations exigées par la présente étude : validation d'images étalonnées ou non.

Comme ces images se composent de mesures intégrées et non ponctuelles, nous considèrerons comme valeur de référence le résultat de l'intégration des mesures de validation sur des cibles test de taille compatible avec celle des éléments d'images.

Pour les deux exemples présentés ici en détail (pluies journalières du Bassin Parisien et durée quotidienne d'insolation en France) nous considèrerons :

- a) tout d'abord les problèmes posés par l'analyse de la structure statistique du phénomène à l'aide des mesures sol ; le modèle retenu sera utilisé pour établir des valeurs de référence que nous comparerons ;
- b) aux estimations obtenues à l'aide du seul réseau d'étalonnage ;
- c) aux mesures établies par télédétection.

2. CAS DES PLUIES JOURNALIERES DANS LE BASSIN PARISIEN :

Validation du réseau de pluviomètres et du radar

Nous disposons pour réaliser ces validations d'un réseau total de 98 stations que nous avons séparé en 29 stations d'étalonnage et 69 stations de validation (voir figure III.1) ; ce réseau est considéré pendant les 11 journées où la chaîne d'acquisition numérique du radar de

Dammartin a fonctionné de manière continue sur 24 heures entre 1980 et 1982.

a) - Analyse de la structure statistique de la pluie au sol

Avant de présenter le détail de cette analyse, soulignons son importance en rappelant qu'elle va intervenir par la suite dans le calcul des valeurs de référence et surtout de leur variance d'estimation ; il est donc capital de mener cette analyse de manière très soignée si l'on veut pouvoir tirer des conclusions fondées quant à la précision des mesures que nous mettrons en regard de ces valeurs de référence. Notre approche a été pragmatique, au sens où nous avons mené parallèlement deux méthodes d'analyse sensiblement différentes ; la concordance des résultats obtenus est pour nous un gage de fiabilité.

Notons également qu'en raison du faible nombre d'événements considérés cette analyse a été menée événement par événement pour les 11 journées disponibles et pour l'ensemble des points de mesure ; le tableau III.1 regroupe l'ensemble des résultats numériques obtenus ; pour des raisons pratiques de mise en page, les résultats graphiques présentés ne portent que sur certaines de ces journées ; nous essaierons de justifier au cours de leur présentation le caractère représentatif de ces exemples (une présentation plus complète se trouve dans G. DELRIEU, 1986).

La première méthode utilisée (construction des variogrammes expérimentaux) est visuelle et interactive ; elle est concrète et laisse une grande part d'initiative à l'utilisateur. La seconde méthode (identification de la covariance généralisée) est automatique et n'exige une attitude critique qu'au moment du choix d'un critère d'identification.

- La construction des variogrammes expérimentaux est à notre avis une procédure de toute manière indispensable, puisqu'elle permet de visualiser la structure du phénomène. Comme nous l'avons vu au Chapitre I, le variogramme expérimental d'un champ de mesures s'obtient par calcul pour des classes de distance $h_\alpha \pm \Delta h$ de la moyenne du carré des écarts entre les mesures réalisées en des points distants de $h_\alpha \pm \Delta h$ (voir formule I.47 où nous avons noté $\mathcal{C}(h_\alpha)$ l'ensemble des couples de points dont

Tableau III.1 : Détail des résultats de la reconnaissance de structure obtenus empiriquement et automatiquement pour 11 journées.

Nature météorologique de l'événement	Date de l'événement	Coefficient de variation (m/σ)	Modèle ajusté empiriquement au variogramme expérimental	Modèle ajusté par Bluepack		Résultats de la validation croisée : moyenne quadratique des écarts de validation et de leur valeur normée par la variance d'estimation pour des modèles *		
				covariance	Degré de dérive	Ajustés empiriquement	Ajustés automatiquement	Choisis a priori pépétiques purs
FRONTAL	7/10/80	0.90	S 1800 50 0	CGP 0 -1.28 0.023	2	201. 44%	137. 84%	783. 5%
	8/10/80	0.82	S 8000 80 0	CGP 0 0 0.059	2	426. 40%	80. 77%	1699 4%
	10/10/80	2.61	S(1) 1200 60 0	L 48.4 0	1	1004. 165%	1048. 132%	955 5%
	14/12/80	2.67	S(1) 500 80 0	P 208	1	96. 87%	218. 93%	213 5%
	15/12/80	1.13	L 12 0	CGP 65.9 0 0.004	1	110. 93%	102. 99%	217 4%
	22/10/82	3.29	L 50 0	L 66 0	1	735. 114%	743. 99%	994 4%
CONVECTIF	2/8/81	0.34	L 3 900	L 25.9 0	0	1122. 103%	1496. 263%	1120 9%
	23/6/82	1.44	L 6 300	P 545	0	559. 114%	567. 96%	567 8%
	13/7/82	0.91	S 11000 50 0	L 484 0	1	6712. 130%	6577. 108%	8032 8%
	15/7/82	1.16	L 2.4 600	L 5.5 469	0	681. 97%	676. 104%	688 9%
	22/7/82	0.30	L 2.5 450	L 11.5 137	0	780. 118%	947. 164%	733 9%

Unités : pluviométrie en 1/10 mm ; distances en km

Description des modèles utilisés : S, modèle sphérique : palier, portée, pépite

L, modèle linéaire : pente, pépite

CGP, covariance généralisée polynomiale C, A_1, A_3 avec $K(h) = C + A_1(h) + A_3(h^3)$

P, modèle pépétique pur : pépite

* Ces deux valeurs peuvent s'interpréter comme un index de la qualité du choix d'un modèle (valeur du haut qui doit être aussi faible que possible) et du choix de ses paramètres (valeur du bas qui doit être aussi proche de 100 % que possible).

l'interdistance appartient à la classe des distances $h_\alpha \pm \Delta h$. Il est toujours souhaitable de distinguer, à l'intérieur de ces classes de distance, des classes d'azimut $a_\beta \pm \Delta a$ qui regroupent les couples des points dont l'orientation est comprise dans ce secteur d'angle ; cette distinction permet de détecter une éventuelle anisotropie dans la structure du champ étudié ; la formule utilisée se déduit simplement de I.47 en notant $\mathcal{C}(h_\alpha, a_\beta)$ les ensembles de couples de points retenus lors du calcul du variogramme dans la direction a_β .

L'ajustement de fonctions théoriques autorisées à ces variogrammes expérimentaux est nécessaire à la mise en oeuvre du système de krigeage (voir chapitre I). Nous avons retenu deux types de variogrammes théoriques :

le modèle linéaire avec pépite :

$$\hat{\gamma}(h) = C(1 - \delta(h)) + w|h| \quad \text{où } \delta(h) \text{ est la fonction de Dirac}$$

et un modèle de variogramme à portée, c'est à dire dont la pente tend vers 0 au-delà d'une certaine distance appelée portée, le variogramme ayant alors atteint une valeur b à peu près constante appelée palier. Nous avons retenu par "habitude" le modèle sphérique avec pépite :

$$\hat{\gamma}(h) = C(1 - \delta(h)) + b \left(\frac{3|h|}{2a} - \frac{1}{2} \frac{|h|^3}{a^3} \right) \quad \text{pour } h \leq a$$

$$\hat{\gamma}(h) = C + b \quad \text{pour } h > a$$

qui a des propriétés très proches d'autres modèles de variogramme à portée (exponentiel ou gaussien notamment) dont l'adéquation est parfois meilleure; nous ne reviendrons pas sur les propriétés respectives de ces modèles très bien décrites dans J.P. DELHOMME, 1976.

L'ajustement de ces fonctions théoriques a été réalisé en utilisant les valeurs expérimentales calculées par le programme GAMMA (J.P.CHILES) ; trois précautions ont été prises :

i) *aux courtes distances*, nous avons conclu à un effet de pépite, parfois très important, lorsqu'un ensemble de 3 ou 4 points significatifs (établis à l'aide de plus de 20 couples de valeurs) n'indique pas de

manière cohérente un départ du variogramme de l'origine ; cette partie du variogramme, comprise en deça du pas d'échantillonnage, mérite une attention particulière car elle a un effet très sensible sur le résultat obtenu par krigeage (voir les essais de sensibilité effectués par H. DARRICAU-BEUCHER, 1981) ; de plus, elle a également un effet sur la variance d'estimation que nous utilisons par la suite comme index de précision et qui ne doit pas être sous-estimée si l'on veut aller dans le sens de la prudence.

ii) *aux grandes distances*, nous avons pris en considération les variogrammes calculés par azimuth lorsque le variogramme global présente un comportement traduisant une dérive à l'échelle du domaine étudié (branche parabolique ou courbe en "cloche") ; il est en général possible d'identifier une direction dans laquelle l'effet de la dérive est moins sensible, ce qui permet de caler une fonction théorique plus réaliste. Mentionnons que cette précaution relève plus d'un souci de cohérence physique que d'une recherche d'efficacité ; en effet, dans le cadre du modèle quasi-stationnaire que nous avons retenu, nous allons pratiquer le krigeage sur des voisinages réduits et donc utiliser seulement pour de courtes distances la fonction de structure ajustée ; à ces distances le comportement du variogramme est insensible à une dérive à grande échelle qui peut être assimilée à une moyenne localement constante (donc filtrée par le variogramme) lentement variable dans l'espace ; par contre, il est plus satisfaisant, d'un point de vue physique, d'identifier une distance au-delà de laquelle la covariation du phénomène s'éteint (portée) pour laisser place à une co-fluctuation due à la tendance à grande échelle.

iii) *aux distances intermédiaires*, il est évident qu'une harmonisation est nécessaire pour conjuguer les décisions prises aux deux étapes précédentes ; il s'agit, dans le cas d'un variogramme à portée, de fixer le palier, c'est à dire la variabilité du phénomène, à un niveau qui respecte la pente du variogramme à l'origine et la valeur de la covariation retenue après examen de l'effet de dérive.

La difficulté de prendre en compte ces différentes précautions rend impossible une automatisation de l'ajustement pour au moins deux raisons ; d'une part, il est très difficile de traduire numériquement les

Figure III.2 : Variogrammes expérimentaux, modèles ajustés empiriquement et automatiquement pour 4 journées représentatives.

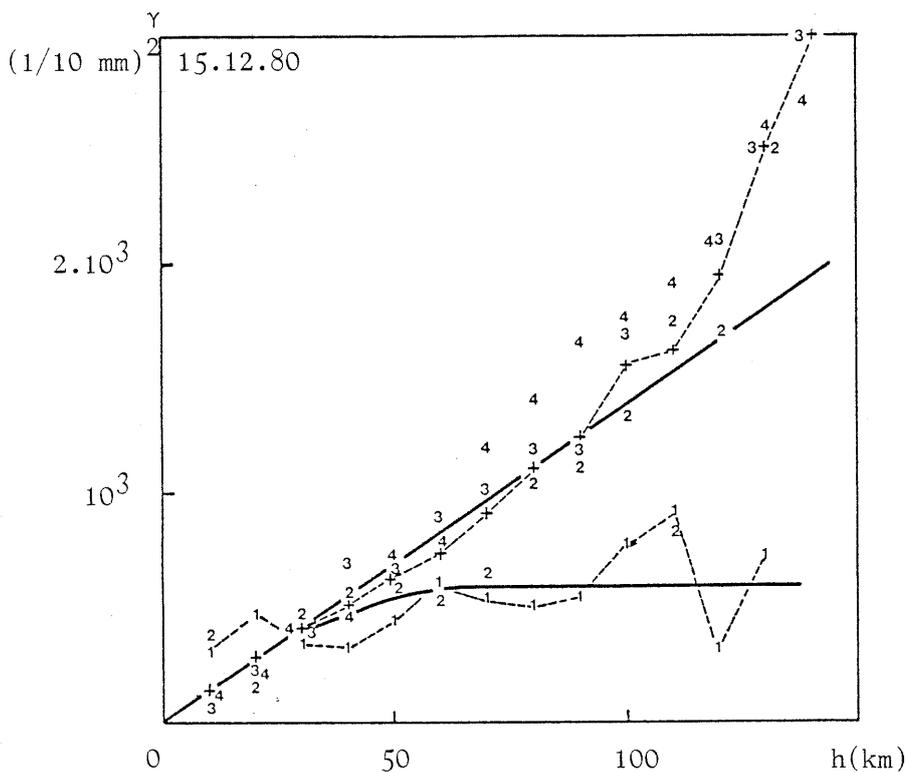
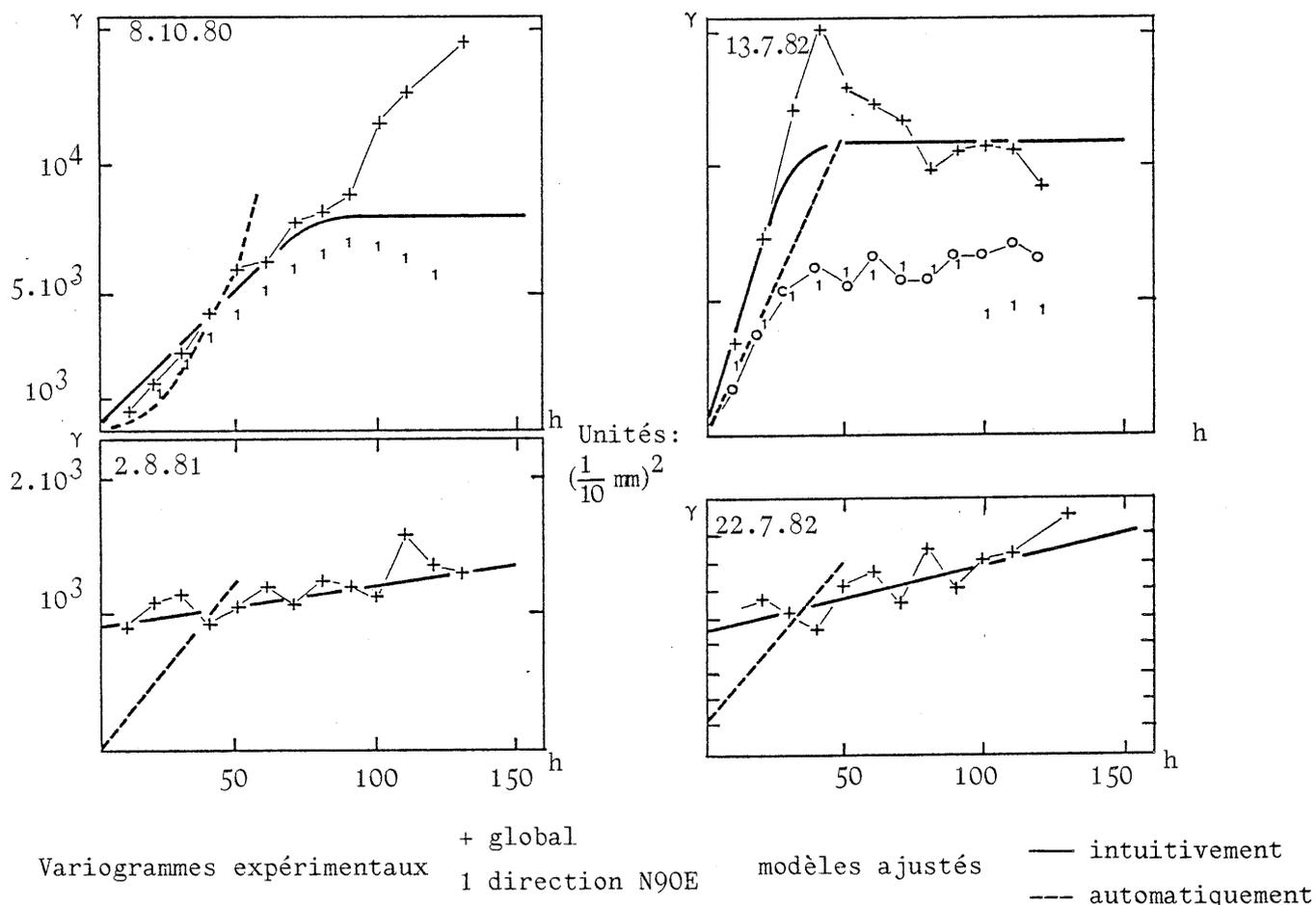


Figure III.3 : Variogrammes expérimentaux et modèles linéaires et sphérique "acceptables".

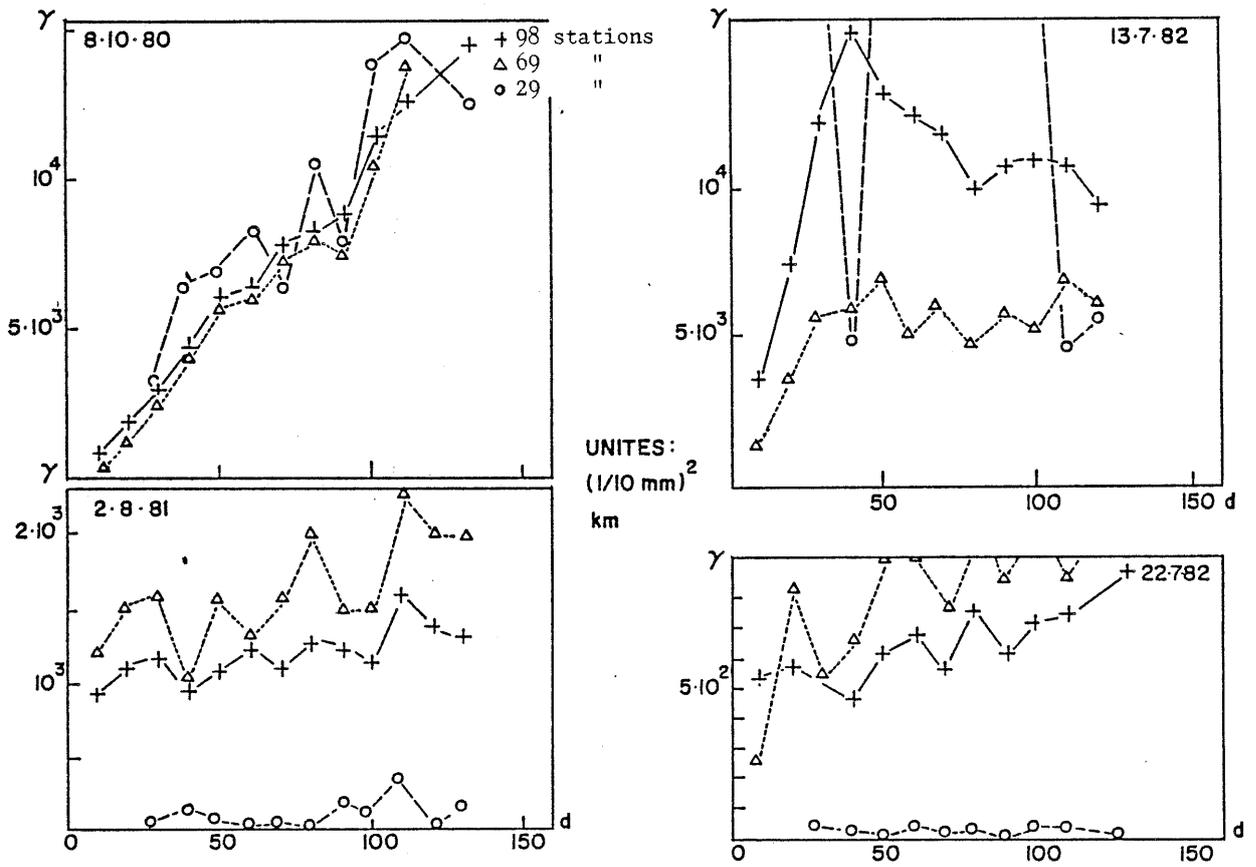
précautions évoquées pour extrapoler le comportement du variogramme au voisinage de 0 au bien pour choisir un palier et une portée à l'aide de variogrammes directionnels ; d'autre part, dans l'hypothèse où cette automatisation est réalisée, rien n'autorise à la juger plus objective qu'un ajustement empirique effectué en respectant les mêmes précautions : son seul avantage serait la rapidité d'exécution qui n'est pas ici un problème d'autant plus que l'ajustement manuel peut être très rapide.

Pour les 11 journées analysées, il est bon de distinguer, selon la situation météorologique, les champs pluvieux à répartition régulière dans l'espace dont la variabilité est généralement faible (situations frontales d'Automne et d'Hiver) des champs pluvieux à répartition plus atomisée produits en Eté par des situations convectives.

Parmi les 6 journées "frontales", 4 peuvent être modélisées à l'aide de variogrammes dont la portée est de 50 à 80 km comme l'illustre le cas du 8/10/80 représenté figure III.2 (le variogramme E-W a été utilisé pour identifier la portée masquée par une importante dérive N-S). Pour les deux journées restantes, il n'a pas été possible d'estimer la portée du phénomène masquée par une dérive complexe ; nous avons préféré utiliser un modèle de variogramme linéaire qui dans sa partie utile ($h < 50$ km) diffère très peu d'un modèle sphérique plausible (voir figure III.3 à propos du 15/12/80).

Parmi les 5 journées convectives, 4 présentent un variogramme expérimental linéaire à fort effet de pépite. La journée du 13/7/82, bien que possédant elle aussi un caractère convectif, présente une répartition beaucoup plus régulière dont le réseau de 98 stations a eu une bonne perception ; le variogramme obtenu a une portée de 50 km assez nette, mais il est difficile de fixer un palier en raison de l'impact très fort de la mesure exceptionnelle réalisée en une station (80.6 mm à Senonches pour une moyenne spatiale de 9.4 mm et un écart-type de 10.4 mm) ; par prudence, et sans raison majeure de remettre en cause cette mesure, le modèle retenu la prend en considération avec un palier à 110 mm^2 (au lieu de 60 comme l'indique le variogramme calculé sans Senonches).

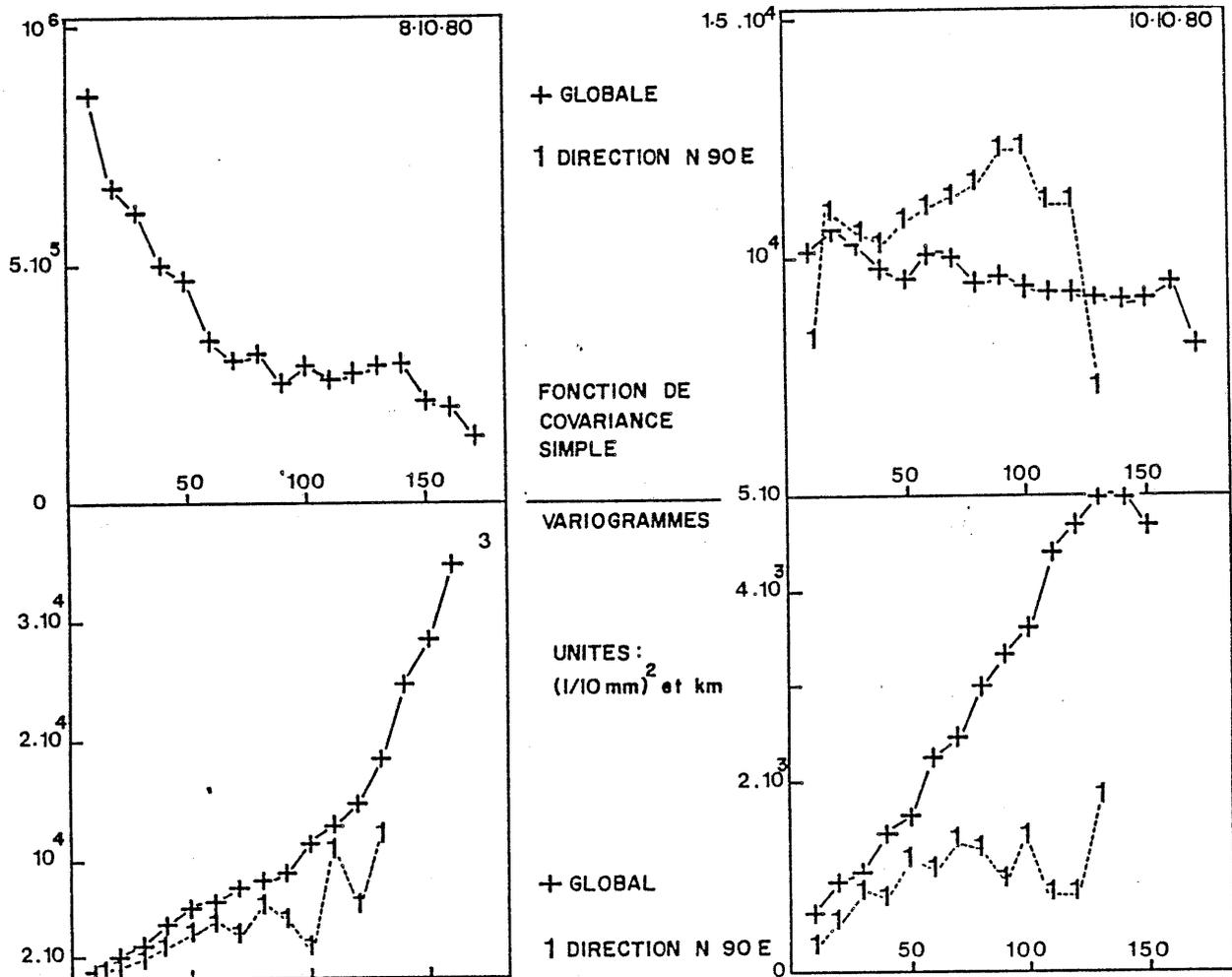
Figure III.4 : Variogrammes expérimentaux obtenus à l'aide de différents réseaux de mesure pour 4 journées représentatives.



Cette dernière remarque rappelle la sensibilité de la fonction de structure au réseau de mesure utilisé pour son inférence ; or, les différentes applications que nous devons faire des variogrammes modélisés s'appuient sur des réseaux différents et visent des buts différents : établir les valeurs de référence à l'aide de 69 stations de validation et évaluer les performances du krigeage appliqué aux 29 stations d'étalonnage; dans chacun de ces cas, *quel est le réseau le plus approprié pour réaliser l'inférence du variogramme ?*

D'un point de vue "méthodologique", le premier but fixé a pour seul impératif la précision (précision de la valeur de référence, mais aussi de sa variance d'estimation) ; il exige donc l'utilisation du variogramme le plus robuste possible, c'est à dire a priori celui inféré à l'aide du réseau le plus dense (98 stations). Le deuxième but doit au contraire avoir une cohérence "opérationnelle" : si l'on dispose de 29 stations pour réaliser une estimation, on est contraint d'inférer un variogramme à l'aide de ces seules stations ; les résultats obtenus souffriront peut-être des incertitudes de cette inférence, mais ils reflèteront de manière fidèle les performances du krigeage appliqué à ce réseau. D'un point de vue pratique, si l'on "compare" les variogrammes expérimentaux obtenus à l'aide des différents réseaux évoqués (réseau complet de 98 stations, réseaux test et d'étalonnage de 69 et 29 stations), on peut remarquer deux comportements très différents : dans le cas des journées "frontales" la sensibilité du variogramme à la densité du réseau est très faible, ce qui est lié à l'homogénéité spatiale du phénomène qu'un réseau peu dense peut identifier (même si, comme le montre la figure III.4 pour le 8/10/80, les fluctuations du variogramme sont plus fortes lorsque le nombre de points de mesure est faible) ; dans le cas des journées convectives, au contraire, la faiblesse du réseau de 29 stations est évidente en raison de la répartition très erratique des cellules pluvieuses; d'un point de vue qualitatif, la nature largement pépétique de la structure est identifiée par les différents réseaux pour 4 des 5 journées, mais la valeur de la variabilité du champ fluctue fortement y compris lors du passage de 69 à 98 stations ; le 13/7/82 est à nouveau pénalisé par la présence d'une valeur forte à Senonches, station faisant partie du réseau de 29 stations pour lequel le variogramme expérimental "explose" en prenant une allure nettement pépétique.

Figure III.5 : Comparaison des fonctions de covariance simple et des variogrammes expérimentaux obtenus pour deux journées frontales.



Au terme de cette description rapide des effets de la densité du réseau sur le variogramme, nous pouvons choisir de la manière suivante :

i) pour établir les valeurs de référence à l'aide du réseau de 69 stations le réseau complet ou le sous-réseau test peuvent être utilisés presque indifféremment pour calculer le variogramme ;

ii) pour évaluer les performances du krigeage appliqué au réseau d'étalonnage, il est préférable, en situation convective, de modéliser le variogramme obtenu à l'aide de 29 stations même si dans un contexte opérationnel on peut espérer que la connaissance climatologique du phénomène nous permette d'accéder à une fonction de structure plus robuste. Soulignons toutefois que la densité du réseau modifie l'amplitude et non la forme des variogrammes (sauf pour le 13/7/82) ; cette précaution ne concerne donc que le calcul de la variance d'estimation, puisque les estimations ne dépendent que de la forme du variogramme.

Pour terminer ce paragraphe consacré aux variogrammes expérimentaux, nous donnons, à titre d'illustration, deux exemples d'une autre fonction de structure expérimentale : la fonction de covariance simple (voir figure III.5) ; cette fonction est obtenue suivant le même mode de répartition en classe $\mathcal{C}(h_{\alpha}, a_{\beta})$ que le variogramme en appliquant la formule :

$$c(h) = EZ(t) \cdot Z(t+h)$$

Les valeurs de Z ne sont pas centrées dans la mesure où la seule estimation accessible de $EZ(t)$ est la moyenne spatiale : l'effet sur la fonction $c(h)$ de ce centrage est réduit à une translation $(EZ(t))^2$ n'affectant pas la forme de la fonction. Les deux exemples choisis montrent que si la fonction de covariance est potentiellement capable d'apprécier la structure du phénomène (cas du 8/10/80), son mode de construction peut être gravement pris en défaut par l'effet d'une moyenne spatiale très variable (cas du 10/10/80) ; le choix du variogramme comme fonction de structure expérimentale est donc pleinement justifié.

- Voyons dans un deuxième temps si une méthode d'identification automatique de la structure, explorant l'éventuelle adéquation de modèles beaucoup plus riches, est justifiée.

Les résultats présentés dans le tableau III.1 ont été obtenus à l'aide du programme Bluepack que le Centre de Géostatistique de l'Ecole des Mines à mis à notre disposition (P. DELFINER, J.P. DELHOMME et J.P. CHILES, 1978 et P. DELFINER, J.P. DELHOMME, J.P. CHILES, A. GALLI et D. RENARD, 1985). La méthode de fonctionnement de ce programme est décrite, sous une forme simplifiée, au Chapitre I.

Parmi les trois modèles proposés, à l'issue de chaque reconnaissance automatique de Bluepack, nous avons systématiquement choisi celui qui donne les meilleurs résultats par validation croisée (option THOMAS du programme) ; ce choix est criticable (il ne rejoint pas toujours celui que le programme effectue sous l'appellation de "best covariance fit", mais il nous est apparu le plus conforme à notre démarche ultérieure qui est de comparer les modèles ajustés empiriquement et automatiquement à l'aide d'une validation croisée : il est donc a priori favorable à l'ajustement automatique.

La description des résultats obtenus pour les 11 journées est à nouveau à mettre en relation avec la situation météorologique.

Pour les 6 journées en situation frontale, 2 journées paraissent significativement mieux décrites à l'aide d'un modèle à dérive complexe qu'à l'aide d'un modèle quasi-stationnaire (les 7 et 8 octobre 1980) ; en effet, pour ces 2 jours, le choix d'une dérive locale quadratique et d'une fonction de covariance généralisée polynomiale (par ailleurs très proche aux faibles distances du variogramme ajusté empiriquement -voir figure III.2 pour le 8/10/80-) a permis d'améliorer à la fois le critère de moyenne quadratique des écarts en validation croisée (moyenne plus faible pour Bluepack) et le critère de moyenne quadratique relative (plus proche de 1, c'est à dire sans biais important de la variance d'estimation). Pour les 4 journées restantes, le modèle quasi-stationnaire paraît tout aussi justifié et parfois meilleur (le 14/12/80, la covariance pépitique identifiée par Bluepack se révèle peu réaliste) ; pour ces 4 jours, la

dérive locale identifiée est linéaire (précisons que les voisinages retenus par Bluepack ont un rayon moyen de 54 km).

Pour les 5 journées convectives, les fonctions de structure ajustées sont généralement très voisines : linéaires avec dérive constante. Le 13/7/82 est la seule journée pour laquelle le degré de dérive n'est pas nul. Les résultats obtenus en validation croisée sont également très proches ; on peut simplement reprocher à Bluepack de sous estimer dans deux cas l'effet de pépité (2/8/81 et 22/7/82) ce qui le conduit à fournir de plus mauvais résultats ; notons que pour ces deux journées, une fonction de covariance purement pépitique est également proposée par le programme mais sa valeur trop forte conduit à des résultats encore plus mauvais en validation croisée.

Après avoir mené conjointement ces deux analyses de structure (empirique autour du modèle quasi-stationnaire et automatique à l'aide du modèle intrinsèque généralisé), nous pouvons proposer les conclusions suivantes :

- Du point de vue de l'efficacité des deux méthodes d'analyse, les résultats de validation croisée contenus dans le tableau III.1 peuvent être résumés par une comptabilisation sommaire de l'avantage que l'on peut donner à une méthode, ou bien de l'égalité de leur performance ; cette comptabilisation a été faite (tableau III.2) en recherchant la plus faible moyenne quadratique d'erreurs de validation (deux moyennes sont considérées distinctes si elles diffèrent de plus de 20 %) et la moyenne quadratique des erreurs normées par l'écart-type d'estimation la plus proche de 1 ; au sens de ces critères le trait dominant de cette comparaison est l'égalité des performances dans plus d'un cas sur deux ; par ailleurs, la méthode automatique prend un léger avantage en situation frontale et la méthode empirique semble la plus efficace en situation convective.

- Du point de vue de la lourdeur des méthodes, il est évident que la méthode empirique est la plus facile à mettre en oeuvre.

- Du point de vue de la nécessité de recourir à une analyse de structure, les résultats portés dans la dernière colonne du tableau III.1

confirment l'intérêt d'une méthode d'interpolation sophistiquée et aident à en situer les performances ; ces résultats ont été obtenus en utilisant un variogramme pépétique pur (de valeur égale à la variance du champ) et une dérive constante : il s'agit donc d'une simple moyenne arithmétique (mobile puisqu'effectuée en voisinage glissant). Lorsque le champ traité est bien structuré (situation frontale) les performances de cette méthode simpliste en terme de moyenne quadratique d'écarts, sont généralement très inférieures (5 jours sur 6). Naturellement pour les journées convectives, les résultats sont sensiblement identiques ; dans ce cas, pour l'ensemble des champs, la variance d'estimation annoncée (à peu près égale à la variance du champ) est beaucoup trop forte.

En résumé, le modèle quasi-stationnaire, par ses performances et sa simplicité, reste bien adapté pour traiter notre problème.

Tableau III.2 : Comptabilisation sommaire des performances en validation croisée des modèles inférés empiriquement et automatiquement.

	Nombre de journées où il y a		
	Avantage		Egalité des modèles
	au modèle empirique	au modèle "automatique"	
Situation frontale	1/0	2/2	3/4
Situation convective	2/0	0/1	3/2

En terme d'écart absolu
le plus faible

En terme d'écart relatif
le plus proche de 1

b) - Résultats de validation du réseau d'étalonnage

Une fois réalisée l'inférence d'un modèle (ici quasi-stationnaire) il est possible à l'aide du krigeage d'une part de calculer la valeur et l'écart-type d'estimation des lames de pluie de référence sur

Figure III.6 : Diagrammes d'efficacité comparant les résultats de deux réseaux de pluviomètres (10 et 29 stations d'étalonnage).

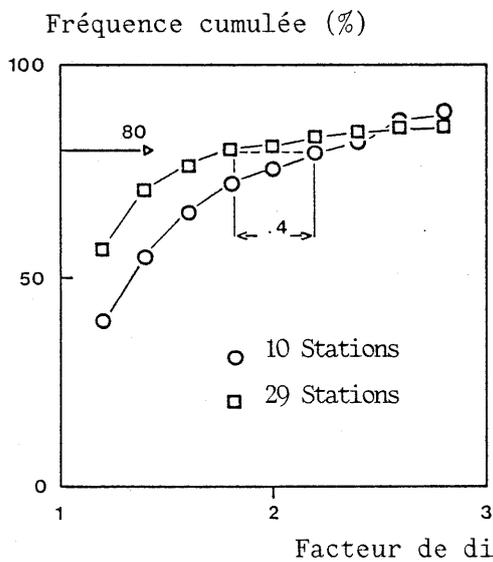
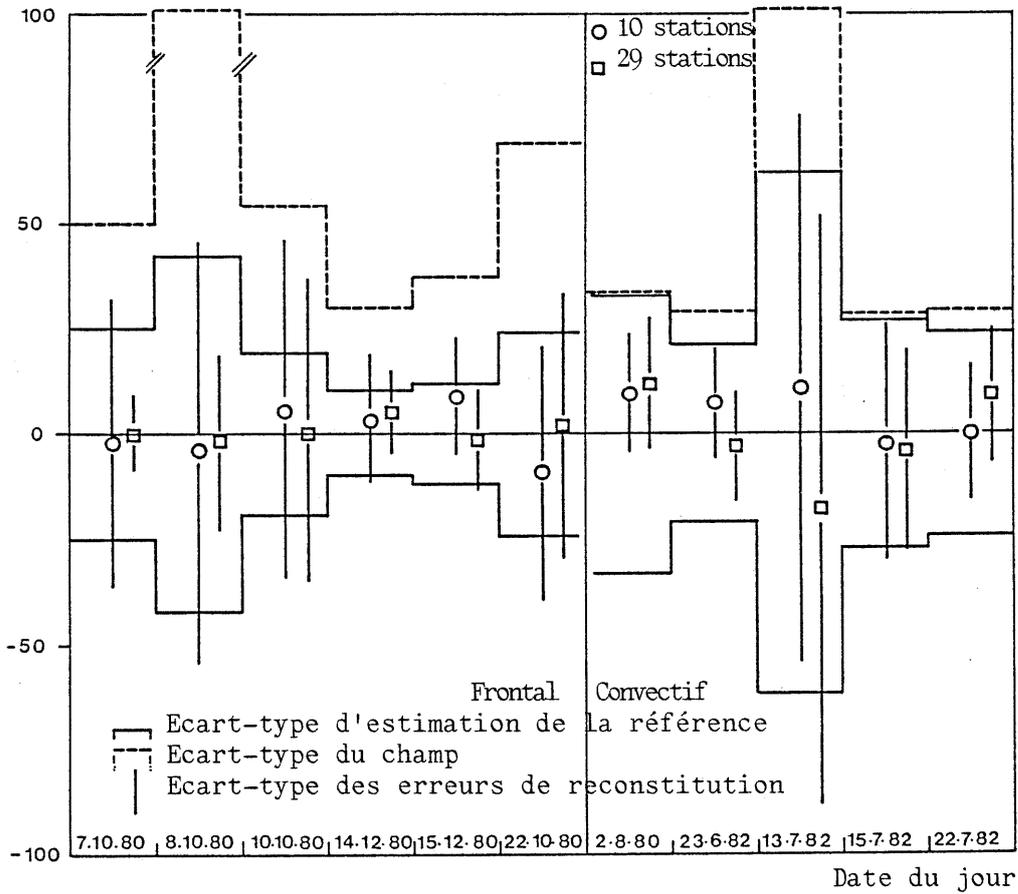
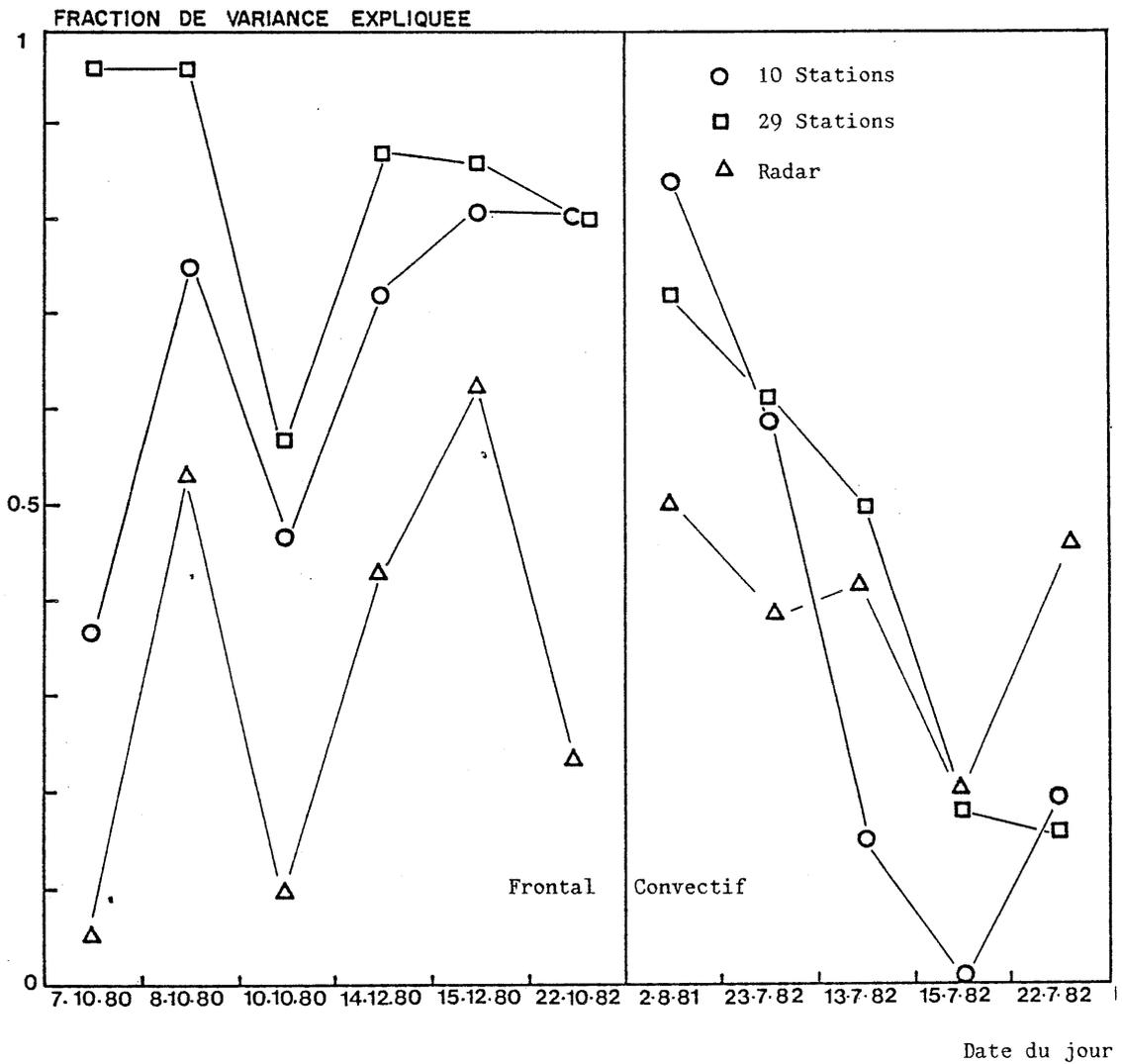


Figure III.7 : Histogramme cumulé des facteurs de différence pour deux réseaux de pluviomètres (10 et 29 stations d'étalonnage).

Figure III.8 : Courbe d'efficacité pour deux réseaux de pluviomètres (10 et 29 stations d'étalonnage) et pour les mesures radar (dont l'analyse est donnée au paragraphe -c-).



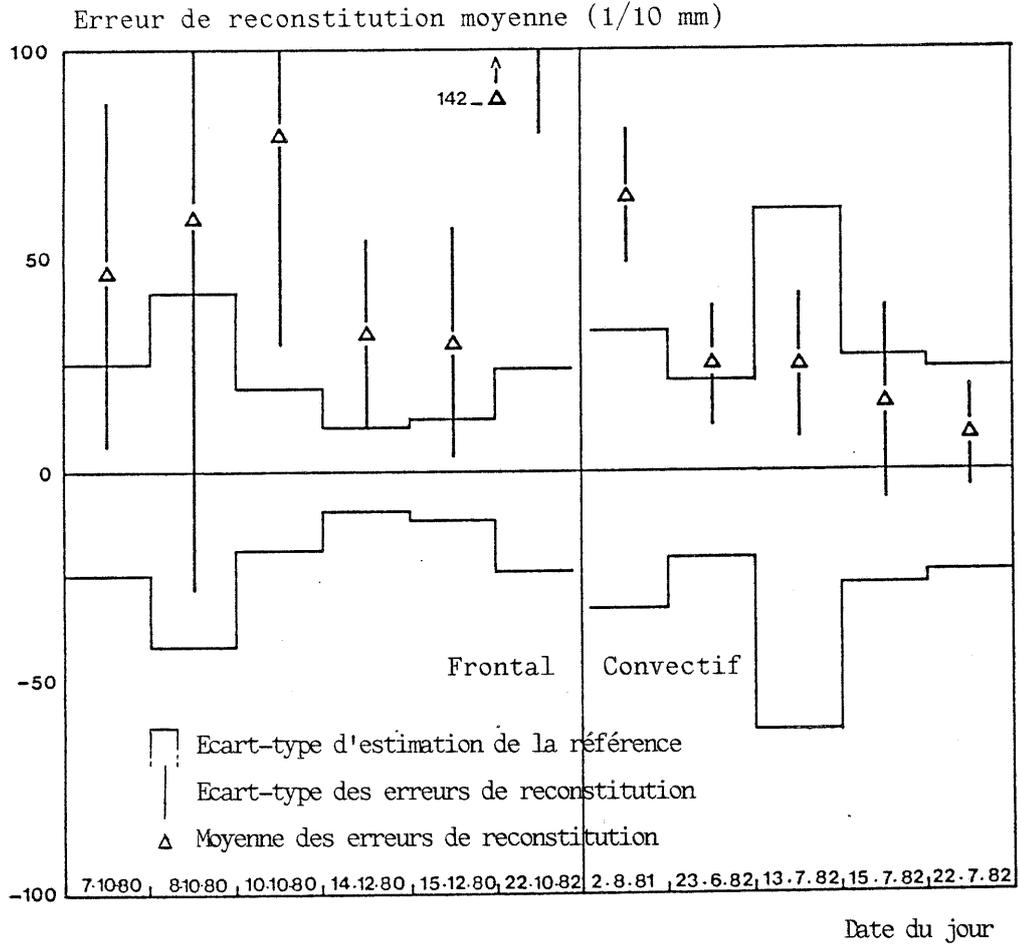
les 69 cibles test (mailles carrées de $5 \times 5 \text{ km}^2$) en utilisant le réseau de validation de 69 stations, d'autre part de calculer les lames de pluie sur ces cibles en utilisant d'autres stations (dites d'étalonnage par référence à leur usage ultérieur); cette démarche permet d'évaluer expérimentalement les performances d'un réseau de pluviomètres (29 dans un premier temps puis 10, afin d'apprécier l'effet de la densité : celle-ci passe de un pluviomètre pour 600 km^2 à un pour 1800 km^2 -voir figure III.1-). Comme les mêmes critères d'efficacité et les mêmes valeurs de référence seront utilisés par la suite pour juger le radar et son étalonnage, les résultats présentés ici constitueront un repère pour situer les performances de ce nouveau dispositif de mesure.

Comme nous l'avons dit, au chapitre précédent, le choix de réseaux de mesure distincts apporte certaines garanties quant à l'objectivité de cette validation ; nous devons cependant noter que pour des raisons de simplicité pratique nous avons krigé le réseau de 29 stations à l'aide du modèle inféré sur le réseau complet : il s'agit là sans aucun doute d'un choix qui favorise les performances du réseau sol; ce "favoritisme" n'est cependant pas excessif puisque nous avons vu précédemment que le type de fonction de structure retenu (variogramme pépitique, linéaire ou sphérique) n'est généralement pas dépendant du réseau choisi (une exception en 11 jours) ; seule fluctue l'appréciation de la variabilité du phénomène qui n'a d'incidence, comme nous l'avons dit plus haut, que sur le calcul des variances d'estimation (dont nous ne nous servons pas dans le cas des réseaux de 29 et 10 stations).

La comparaison des résultats à l'aide des critères d'efficacité retenus au chapitre précédent appelle les remarques suivantes :

- le diagramme d'efficacité (figure III.6) montre que la réduction du biais de reconstitution, lorsque la densité du réseau augmente, n'est sensible qu'en situation frontale ; la réduction de l'écart-type de reconstitution est rare et faible ; la fiabilité de la référence (donnée par l'écart-type d'estimation) est souvent du même ordre de grandeur que la dispersion des erreurs de reconstitution.

Figure III.9 : Diagramme d'efficacité des mesures radar sur les 69 cibles test.



Fréquence cumulée

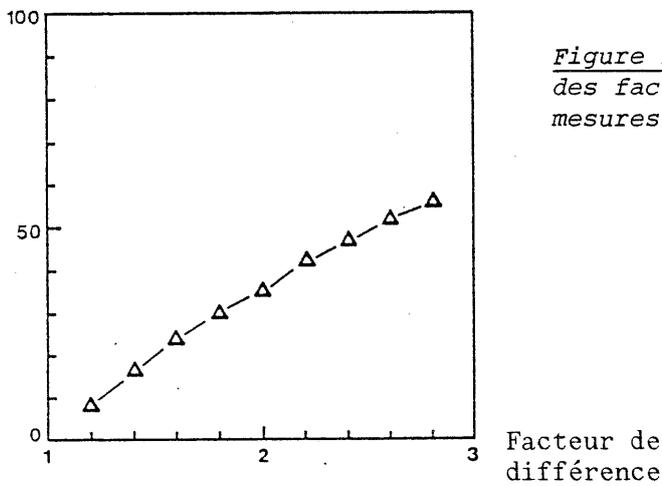


Figure III.10 : Histogramme cumulé des facteurs de différence des mesures radar.

- la distribution des facteurs de différence (figure III.7) montre que dans 80 % des cas, l'erreur commise par le réseau de faible densité est de 40 % plus fort, toutes journées confondues.

- enfin, les courbes d'efficacité (figure III.8) montrent une dégradation générale de la cofluctuation des valeurs reconstituées avec les valeurs de référence lorsque la densité de réseau diminue ; on peut également noter la faiblesse des performances d'un réseau sol en situation convective.

c) - Résultats de validation du radar météorologique non étalonné

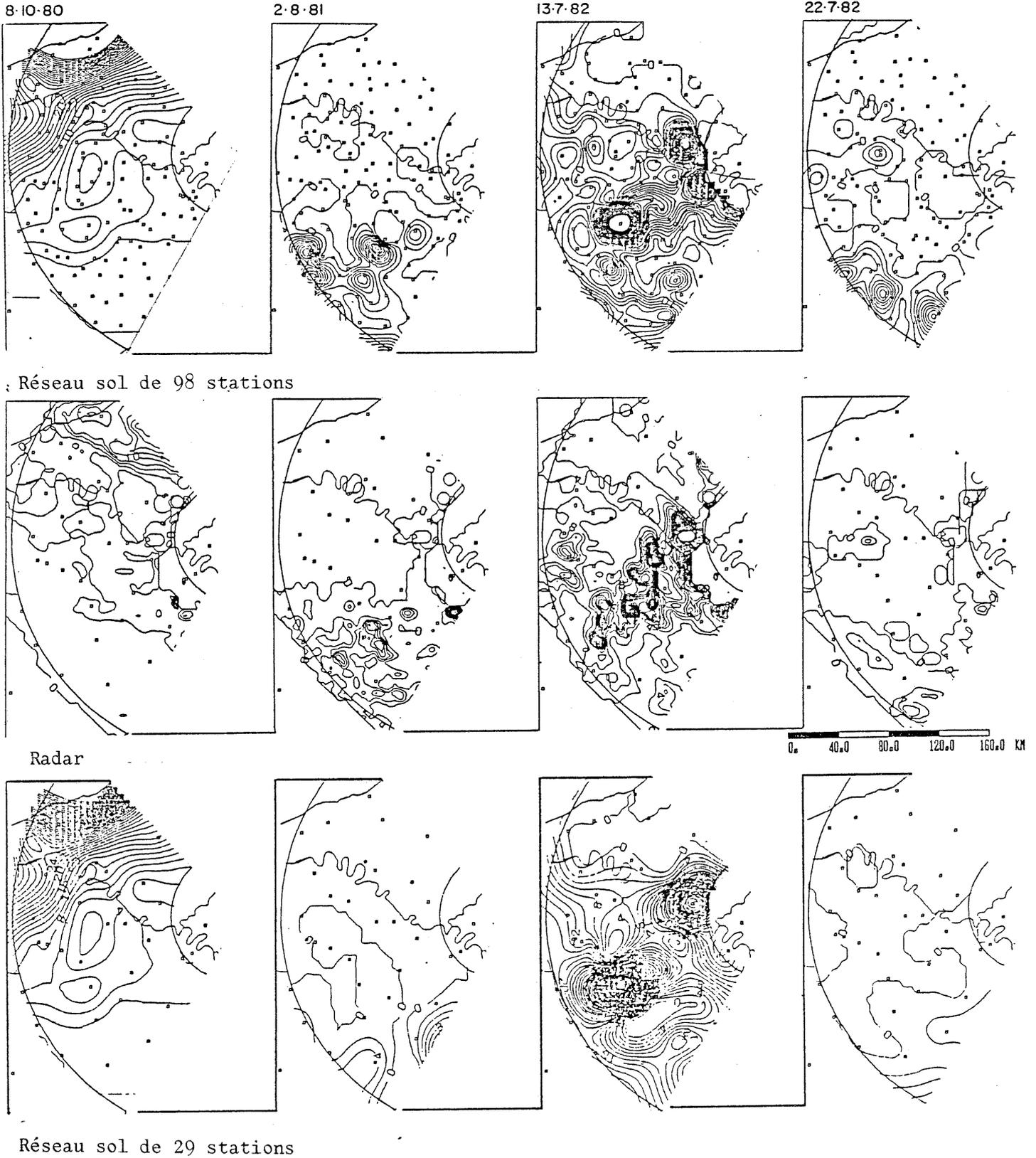
Les mesures radar validées ici ont été obtenues par simple application aux réflectivités enregistrées de la relation de Marshall-Palmer puis d'un cumul dans le temps (journalier) et l'espace (passage d'une grille sphérique à une image rectangulaire).

Les critères habituels d'efficacité appellent les remarques suivantes :

- le diagramme d'efficacité (figure III.9) montre une tendance systématique du radar à la sous-estimation ; cette tendance est plus forte pour les situations de type frontal, ce qui semble assez logique lorsqu'on rappelle que pour ce type de situation et aux grandes distances, la formation de la pluie à des altitudes plutôt faibles conduit à un remplissage partiel du faisceau ; mais il faut être très prudent sur l'interprétation de ces résultats qui reposent directement sur le réglage électronique de l'appareil dont nous savons qu'il a subi des modifications durant la période d'étude. Les écarts-types de reconstitution et d'estimation ne sont pas faciles à interpréter dans la mesure où ils paraissent moins forts en situation convective : en effet, il est surprenant que les fluctuations de la mesure radar autour de la référence soient les plus faibles lorsque cette référence est la moins sûre.

- la distribution des facteurs de différence (figure III.10) traduit essentiellement le fort biais de la mesure radar (dans 65 % des cas l'erreur dépasse 100 %).

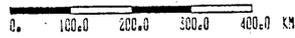
Figure III.11 : Cartes isohyètes obtenues pour quatre journées représentatives à l'aide du radar et d'un réseau sol de 98 puis 29 stations (l'écart entre deux isolignes est de 0.2 mm).



- La courbe d'efficacité (figure III.8) montre le très faible pouvoir explicatif du radar en terme de co-fluctuation spatiale ; cette constatation est assez gênante pour ce dispositif dont la qualité la plus attendue est une bonne perception des fluctuations spatiales du phénomène. En situation convective son comportement bien que médiocre (toujours moins de 50 % de variance expliquée) est plus stable que celui d'un réseau sol ; ceci est conforme à l'idée que ce type de situation est la plus défavorable pour un réseau de pluviomètres. En situation frontale, les performances du radar sont systématiquement inférieures et varient très parallèlement à celles d'un réseau sol de faible densité : cela pourrait indiquer que la qualité des mesures, indépendamment du dispositif utilisé, est sensible à certaines particularités du phénomène lui-même ; il est en effet difficile de proposer des explications communes à la défaillance des deux dispositifs. Cette dernière remarque mérite d'être confirmée par une étude plus robuste car ce critère de variance est sensible à la particularité de certaines mesures puisqu'il est construit à partir de formes quadratiques.

Pour terminer, la figure III.11 donne des éléments visuels de comparaison entre les mesures radar et sol en rapprochant les cartes d'isohyètes obtenues pour les 4 journées "représentatives" à l'aide du radar puis du réseau sol lorsque celui-ci compte 98 ou 29 pluviomètres. Ces cartes illustrent assez bien les causes de défaillance du radar qui restitue globalement de manière satisfaisante l'allure générale du phénomène mais qui commet quantitativement des erreurs grossières (dans le cas du 8/10/80 par exemple le réseau sol dense indique de manière très cohérente un maximum important de pluviométrie sur la côte Normande ; le radar, sans doute masqué dans cette zone, sous-détecte de manière importante). Les cartes obtenues à l'aide d'un réseau sol réduit illustrent la piètre appréciation de la variabilité et de la distribution spatiale du phénomène qu'a ce type de dispositif (voir en particulier, l'exemple du 2/8/81 ou du 22/7/82) ; il n'en demeure pas moins qu'il fournit statistiquement des résultats plus précis.

Figure III.12 : Cartes des sous-réseaux (a) d'étalonnage (23 stations) et (b) de validation (61 stations).



3. CAS DE LA FRACTION JOURNALIERE D'INSOLATION EN FRANCE

Validation du réseau d'héliographes et de la méthode CACTUS.

Nous avons utilisé pour réaliser ces validations une sélection de 15 journées parmi plus de 300 journées disponibles en 1984 (voir S. GALLE, 1986, pour des résultats plus complets). Le réseau héliographique compte 84 stations que nous avons séparées en 23 stations d'étalonnage et 61 stations de validation (voir figure III.12)

a) - Analyse de la structure statistique de la fraction journalière d'insolation au sol

Au contraire de l'exemple précédent sur la pluviométrie, nous possédons ici un nombre suffisant de réalisations du phénomène étudié pour utiliser une approche climatologique, c'est à dire considérer que les différentes journées observées constituent différentes réalisations indépendantes d'un même processus. Naturellement, il convient de rester prudent dans l'utilisation d'une telle approche :

- l'unicité du processus générateur n'est pas acquise puisque les données disponibles, bien que désaisonnalisées au sens astronomique portent sur une année entière, donc une période où des types de situations météorologiques différents se succèdent ; afin de travailler sur des ensembles de réalisations plus homogènes, nous avons divisé l'année en trois périodes : la période retenue ici (Mars, Avril, Mai, Juin) regroupe des champs d'insolation dont la variabilité spatiale apparaît importante au simple examen visuel des images CACTUS ;

- l'indépendance entre les champs journaliers successifs n'est pas assurée en raison de la persistance de certaines situations météorologiques ; la démarche la plus rigoureuse pour éviter cette dépendance serait de ne retenir qu'une journée sur n en choisissant pour valeur de n la portée de la fonction d'autocorrélation temporelle du phénomène ; malheureusement des études antérieures (J. GOUSSEBAILLE, 1974) montrent que cette valeur n (3 à 6 jours suivant les situations) conduirait à l'élimination de trop de journées compte-tenu de l'ensemble tout de même

réduit de réalisations disponibles ; nous avons préféré opérer une sélection visuelle d'après les images CACTUS supprimant simplement la répétition de champs trop ressemblants : cette opération empirique contient certes des risques de biais et n'est pas une garantie absolue d'indépendance ; elle nous a permis de ramener pour la période donnée l'échantillon de 120 à 71 journées. Ces 71 journées ont été utilisées pour analyser la structure du phénomène ; seules les 15 premières n'ont fait ici l'objet d'une validation.

Comme nous l'avons vu au Chapitre I, ce type d'approche climatologique se distingue de l'approche spatiale pure i) par un mode de calcul différent des fonctions de structure expérimentales et ii) par une plus grande richesse des possibilités de codage du processus étudié afin de mieux satisfaire les hypothèses de stationnarité.

i) Les fonctions de structure expérimentales s'obtiennent en calculant pour chaque couple de points de mesure la covariance des valeurs mesurées pour l'ensemble des réalisations, puis en reportant sur un graphe les résultats obtenus en fonction de l'interdistance des points. Si l'on dispose de n points de mesure on obtient ainsi un nuage de $(n+1).n/2$ points que l'on peut résumer par une courbe moyenne donnant pour différentes classes de distance $h_x \pm \Delta h$ la valeur moyenne des covariances obtenues pour cette classe. Ces deux formes expérimentales de la fonction de structure présentent un intérêt lors de la modélisation nécessaire à la mise en oeuvre du krigeage :

- le nuage permet de visualiser la dispersion des covariances qui est due pour partie à des fluctuations d'échantillonnage et pour partie à l'hétérogénéité spatiale des champs étudiés ; il permet donc d'apprécier le bien fondé des hypothèses d'homogénéité nécessaires à une modélisation de la structure ;

- la courbe moyenne, par construction beaucoup plus lisse, facilite l'ajustement d'un modèle suivant la même démarche que dans un contexte spatial.

ii) Afin de se ramener à un processus satisfaisant des hypothèses de stationnarité simples (d'ordre 2 par exemple), ce qui facilite le conditionnement du système de krigeage, il est possible de pratiquer différents types de codage (centrage des valeurs par rapport aux moyennes climatologiques ponctuelles ou réduction par rapport aux écarts-types ponctuels par exemple). Le choix d'un codage est naturellement propre à chaque cas particulier et repose sur l'analyse des caractéristiques statistiques des données disponibles ; ce choix doit être fait en gardant à l'esprit la nécessité, une fois l'estimation d'une valeur codée réalisée, de pouvoir "décoder" cette estimation afin de revenir aux valeurs "naturelles" du phénomène.

L'analyse des caractéristiques statistiques de la fraction d'insolation pendant la période choisie montre que :

- le champ des moyennes climatologiques présente une variabilité appréciable (écart-type compris entre 5 et 6 %) et une bonne structure spatiale (voir le variogramme obtenu figure III.13 : pépite nulle, portée de 400 km environ) ;

- le champ des écarts-types climatologiques présente une variabilité faible (écart-type entre 1 et 1.5 %) et une structure spatiale discutable (voir figure III.13 : pépite non nulle, portée sans doute inférieure à 200 km).

Ces deux éléments permettent de conclure d'une part à la nécessité de travailler sur des valeurs centrées (l'amplitude assez forte du champ des moyennes biaise certainement la covariance simple ; ce champ étant bien structuré, l'estimation de la moyenne en tout point reconstitué sera précise et permettra le "décodage" des valeurs reconstituées) ; d'autre part, il apparaît inutile (puisque l'amplitude du champ des écarts-types est faible) voire dangereux (la médiocrité de la structure du champ conduira à des reconstitutions ponctuelles peu précises donc à un "décodage" peu précis) de réduire les valeurs centrées.

On peut donc retenir dans un premier temps la décomposition suivante :

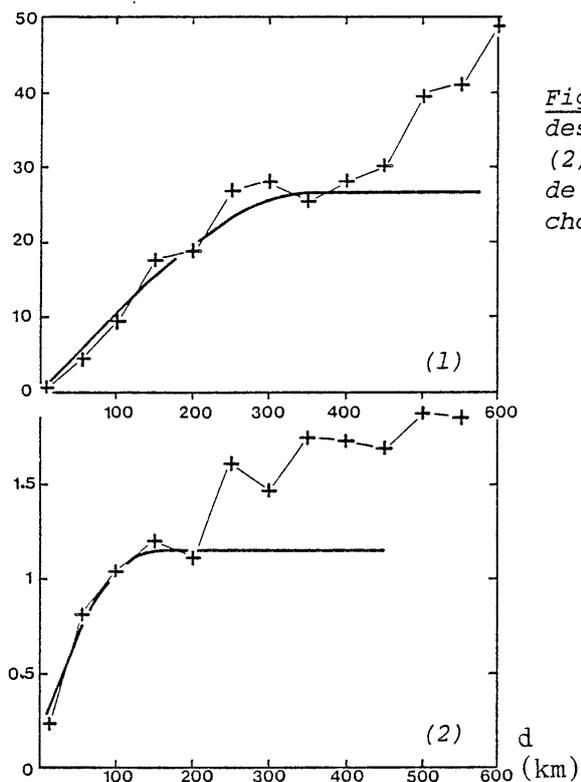
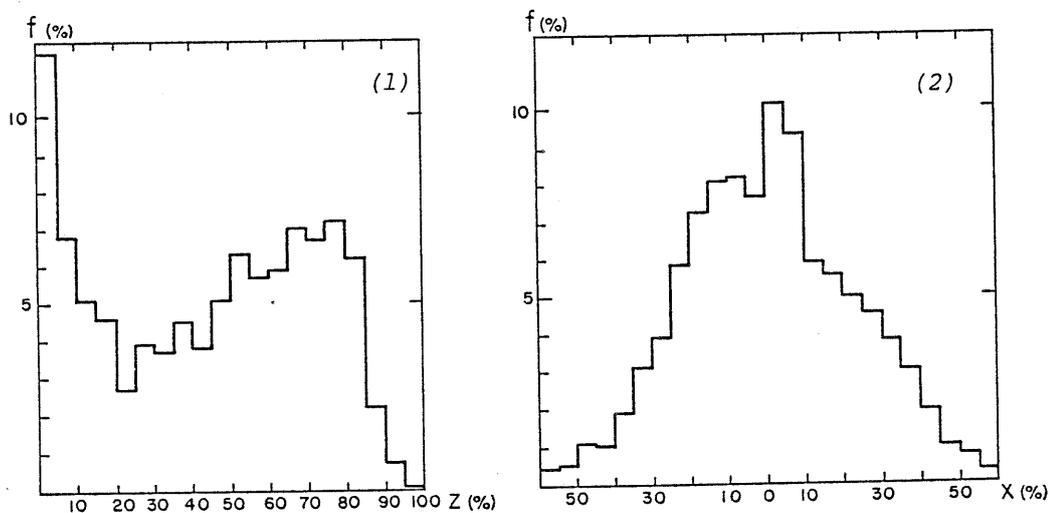


Figure III.13 : Variogramme du champ des moyennes (1) et des écarts-types (2) climatologiques calculés à l'aide de 84 héliographes pour la période choisie.

Figure III.14 : Histogrammes des fréquences de la fraction quotidienne d'insolation (1) brute et (2) après centrage par rapport aux moyennes climatologiques et spatiales.



$$Z^W(t) = EZ(t) + Y^W(t) \quad (\text{III.1})$$

où $Y^W(t)$ est d'espérance nulle.

Cependant, la moyenne spatiale de Y varie sensiblement d'une réalisation w à l'autre ; si l'on note cette moyenne :

$$U_w = \frac{1}{A} \int_A Y^W(t) dt \quad (\text{III.2})$$

il est facile d'avoir une estimation U_w^* de cette moyenne par :

$$U_w^* = \frac{1}{n} \sum_{i=1}^n Y^W(t_i)$$

et l'on peut faire deux remarques :

- d'une part l'écart-type d'estimation de U_w^* est faible ; cet écart-type se calcule aisément à l'aide de la fonction de covariance :

$$\sqrt{E(U_w - U_w^*)^2} = \sqrt{C_{YY}(A,A) - \sum_{i=1}^n \frac{1}{n} C_{YY}(t_i,A)}$$

ou :

$$C(t_i,A) = \int_A C(t_i,t) dt \quad \text{et} \quad C(A,A) = \iint_A C(t,t') dt dt'$$

et nous trouvons pour un réseau de 23 stations un résultat égal à 4 %.

- d'autre part, l'écart-type de la variable aléatoire U est estimé à 24 %.

La variabilité inter-réalisations de U est donc significative et reflète une cofluctuation assez forte de l'insolation à l'échelle de la France. Certes, cette cofluctuation ne nuit pas à l'estimation par krigeage à condition que l'on introduise une condition de non biais filtrant cette valeur moyenne spatiale ; par contre, elle entraîne une surestimation importante de la covariance.

En effet, si l'on note $X_W(t)$ la valeur du processus $Y_W(t)$ centrée par rapport à U_W :

$$Y^W(t) = X^W(t) + U_W \quad (\text{III.3})$$

il est facile de montrer que la covariance des Y s'écrit :

$$EY(t).Y(t') = EX(t).X(t') + EU^2 \quad (\text{III.4})$$

la surestimation est égale à la variance des moyennes spatiales ($X^W(t)$ est d'espérance nulle puisque $EY(t)=0$ et III.2 implique : $EU=1/A \int_A EY(t)dt=0$).

En combinant les décompositions III.1 et III.3, nous obtenons le codage qui semble approprié à notre exemple :

$$Z^W(t) = EZ(t) + X^W(t) + U_W \quad (\text{III.5})$$

où la fonction étudiée apparaît comme la somme d'une fonction commune à l'ensemble des réalisations, d'une fonction aléatoire propre à chaque réalisation et d'une variable aléatoire.

L'effet de ce codage est triple :

- il permet d'éliminer dans le calcul de la covariance une part importante de co-fluctuation due à des tendances systématiques spatiales et temporelles que l'on peut évaluer avec une bonne précision (si le nombre de réalisations est suffisant pour estimer $EZ(t)$ et si le nombre de points de mesures et la structure du phénomène permettent de calculer correctement U_W^*).

- il conduit à une utilisation simple du système de krigeage : estimation de $Y_W(t_0)$ par krigeage simple dans $Y_W(t_i)$ et krigeage universel des $EZ(t)$; la variance d'estimation est obtenue par addition des variances d'estimation fournies par ces krigeages.

- enfin, il s'agit là d'un effet "secondaire", puisque non recherché, mais d'une grande importance, ce codage modifie la distribution fréquentielle des valeurs ; la figure III.14 présente les histogrammes des fréquences de Z^W et X^W : il est clair que la distribution des X^W est plus gaussienne que celle des Z_W dont l'allure est sensiblement bimodale ; l'application du système de krigeage aux X_W plutôt qu'aux Z_W est donc plus satisfaisante puisque ce système repose sur un critère de moindres carrés "préférant" la normalité des valeurs traitées.

Pour terminer, il est intéressant, d'un point de vue pratique, de remarquer que si l'on note V_W la moyenne spatiale de Z^W :

$$V_W = 1/A \int_A Z^W(t) dt = 1/A \int_A (EZ(t) + Y^W(t)) dt = EV + U_W$$

la décomposition III.5 peut s'écrire :

$$\begin{aligned} Z^W(t) &= EZ(t) + X^W(t) + V_W - EV \\ &= E(Z(t) - V) + X^W(t) + V_W \end{aligned}$$

ce qui signifie que l'ordre dans lequel s'effectuent les deux opérations de centrage ne modifie pas la valeur de la variable codée $X^W(t)$.

La figure III.15 montre la fonction de covariance des $X^W(t)$ obtenue après codage des fractions journalières d'insolation suivant III.5, à l'aide des 71 journées de la période choisie et de 84 stations héliographiques ; cette figure présente également le corrélogramme des $X^W(t)$, c'est à dire la fonction de covariance des valeurs $X^W(t)/E(X(t)^2)$, pour laquelle la présentation du nuage expérimental nous était plus commode d'un point de vue informatique. Ces fonctions expérimentales appellent les remarques suivantes :

- leur forme générale est assez proche d'une fonction exponentielle-cosinus (modèle proposé par L.S. GANDIN, 1969 : $C(h) = e^{-ah} \cos bh$) s'annulant à environ 300 km ; une légère discontinuité à l'origine ($C(\varepsilon) \neq C(0)$ lorsque ε petit) indique la présence de microrégionalisation ou d'erreurs dont la structure échappe au réseau de

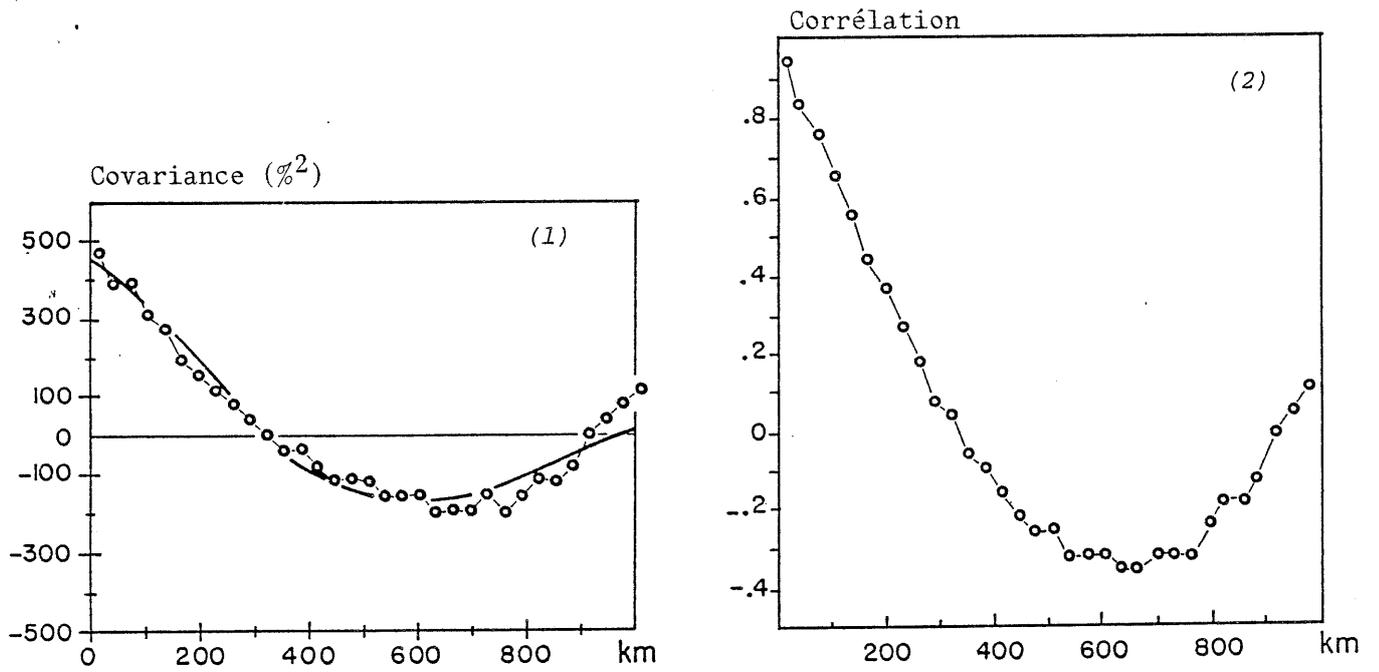


Figure III.15 : Fonction de covariance expérimentale des fractions journalières d'insolation après double centrage (1) puis centrage et réduction (2). Modèle exponentielle-cosinus ajusté empiriquement. Corrélogramme expérimental des fractions journalières d'insolation (3).

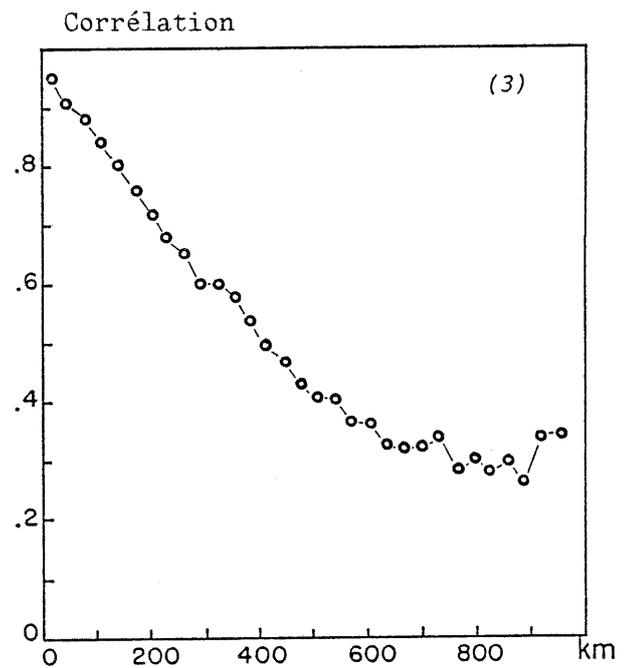
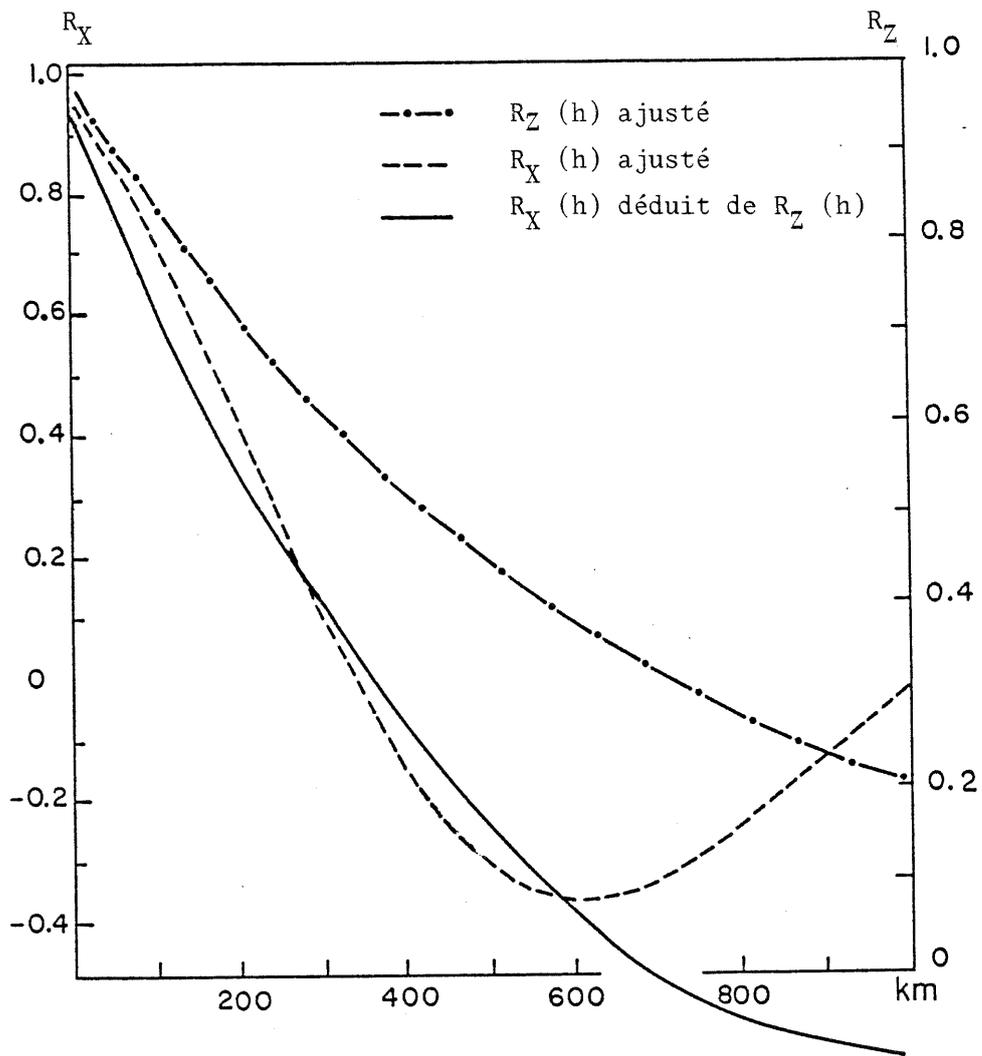


Figure III.16 : Comparaison des corrélogrammes des valeurs X ajusté au nuage expérimental et déduit du corrélogramme des valeurs Z .



mesure, ce qui conduit, en toute rigueur à modifier la décomposition III.5 en :

$$Z^W(t) = EZ(t) + X^W'(t) + e^W(t) + U_W \quad (\text{III.6})$$

avec : $E e^W(t) = 0$
 $1/A \int_A e^W(t) dt = 0$

et : $EX^W'(t).e^W(t') = 0$ si $t \neq t'$

ainsi, nous admettons que ce terme ne modifie pas le calcul de $EZ(t)$ et U_W , par contre la covariance de X_W se décompose en :

$$\begin{aligned} C_X(t, t') &= EX(t).X(t') = EX'(t).X'(t') + Ee(t).e(t') \\ &= C_{X'}(t, t') + C_e(t, t') \end{aligned}$$

avec : $C_e(t, t') = 0$ si $t \neq t'$

$$C_e(t, t) = kC_X(t, t)$$

une lecture directe du corrélogramme à l'origine montre qu'ici k est de l'ordre de 5 %.

- la stabilité des fonctions expérimentales moyennes est remarquable jusqu'aux plus grandes distances, ce qui traduit la robustesse de l'approche climatologique.

- la dispersion du nuage expérimental croît sensiblement avec la distance ; si l'on admet que l'écart-type d'échantillonnage des coefficients de corrélation peut être approché par l'expression :

$$\sigma_R \approx \frac{1-R^2}{\sqrt{N}}$$

où N représente le nombre de réalisations prises en compte d'après J.VIALAR, 1956, page 49) ; cette croissance s'explique puisque R décroît avec la distance ; de plus, la distribution des points autour du corrélogramme moyen semble très proche d'une distribution normale d'écart-type R (la partie la plus dense du nuage se trouve à moins d'un écart-type, la quasi-totalité des points est à moins de deux écarts-types). Il est donc possible d'imputer raisonnablement l'essentiel de la dispersion

aux fluctuations d'échantillonnage ce qui conforte les hypothèses faites sur la stationnarité du processus codé (la non-stationnarité étant une autre source de dispersion).

A titre d'illustration, la figure III.15 présente également le corrélogramme brut du processus (c'est à dire la fonction de covariance des Y normés par leurs écarts-types). La décroissance de la corrélation est alors beaucoup moins forte. Du fait que nous travaillons ici en corrélation, la relation III.4 se complique légèrement ; le corrélogramme de la figure III.14 est obtenu par :

$$R_z(t, t') = \frac{EY(t) \cdot Y(t')}{\sqrt{EY(t^2) \cdot EY(t')^2}}$$

soit en utilisant la décomposition III.3 et la formule III.4 et en admettant que la variance de Y est constante dans l'espace :

$$R_z(t, t') = \frac{EX(t) \cdot X(t') + EU^2}{EX(t)^2 + EU^2} = \frac{R_X(t, t') + EU^2/EX(t)^2}{1 + EU^2/EX(t)^2}$$

en notant $R_X(t, t')$ le coefficient de corrélation des X dont l'expression devient :

$$R_X(t, t') = R_z(t, t') + (R_z(t, t') - 1) \cdot EU^2/EX(t)^2$$

Il est alors très satisfaisant de constater que les courbes ajustées aux deux corrélogrammes vérifient numériquement cette formule ; en effet, les paramètres retenus pour R_z sont $a = -1.87 \cdot 10^{-3}$ et $b = 4.8 \cdot 10^{-3}$ et les estimations des variances donnent $EU^2 = 592 (\%)^2$ et $EX(t)^2 = 498 (\%)^2$; ce qui conduit pour $R_X(t, t')$ à une courbe déduite de $R_z(t, t')$ très voisine de sa forme expérimentale jusqu'aux distances moyennes (voir figure III.16).

Nous obtenons ainsi une vérification complémentaire de l'adéquation du modèle proposé en III.5.

Enfin, la figure III.15 présente les corrélogrammes des $X_w(t)$ pour différents azimuts afin de détecter une éventuelle anisotropie

Tableau III.2 : Caractéristiques statistiques globales des mesures sol et CACTUS pour les 15 journées de validation.
Description des modèles de variogramme ajustés aux mesures sol. (Résultats obtenus à l'aide de 84 points de mesures).

Dates	MESURES SOL			Modèles de variogramme ajustés aux mesures sol	MESURES CACTUS		
	Moyenne	Ecart-type	Coefficient de variation		Moyenne	Ecart-type	Coefficient de variation
02.03.84	26	24	0.92	S 500, 500, 0	17	27	1.59
04.03.84	60	23	0.38	S 530, 350, 100	46	33	0.72
08.03.84	57	28	0.49		38	32	0.84
10.03.84	52	34	0.65	S 1400, 550, 0	51	38	0.74
11.03.84	38	35	0.92	L .3, 0	42	42	1.0
12.03.84	46	32	0.70	S 500, 400, 0	45	39	0.87
14.03.84	57	29	0.51	S 600, 500, 0	69	38	0.55
15.03.84	45	32	0.71	S 1000, 750, 0	57	43	0.75
16.03.84	34	24	0.71	S 800, 600, 0	44	36	0.82
19.03.84	49	22	0.45	S 250, 250, 0	71	24	0.34
21.03.84	39	29	0.74	S 600, 450, 0	50	40	0.80
22.03.84	45	22	0.49	S 500, 500, 0	58	35	0.60
23.03.84	55	19	0.34	S 300, 300, 0	71	25	0.35
24.03.84	13	16	1.23	S 200, 600, 0	10	15	1.50
25.03.84	35	23	0.66	S 700, 800, 0	29	24	0.83

Unités : fractions d'insolation en %
distances en km
Description des modèles utilisés : S, modèle sphérique avec (palier, portée, pépité)
L, Modèle linéaire avec (pente, pépité)

géométrique des champs de ces résidus. Pour construire ces corrélogrammes, seuls sont pris en considération les couples de points de mesure orientés suivant une direction donnée du plan (ici NOE, N45E et N90E) ou ne s'écartant que faiblement de cette direction (ici moins de 15° d'écart). Ces corrélogrammes présentent un léger allongement de la portée (distance de corrélation nulle passant de 300 à 400 km) et une légère diminution de l'effet de pépite (passant de .12 à .03) lorsque l'on va d'une orientation NS à EW. Ces fluctuations ne sont pas surprenantes d'un point de vue météorologique, puisque la différenciation des régimes d'insolation est plus forte en latitude qu'en longitude. Néanmoins, en raison des codages opérés (notamment le centrage par rapport à la moyenne climatologique), ces fluctuations sont très faibles ; il paraît donc raisonnable de ne pas les prendre explicitement en considération par la suite.

Pour compléter cette analyse de structure, la table III.2 décrit les modèles de variogrammes ajustés empiriquement aux variogrammes expérimentaux des valeurs brutes de 15 réalisations (parmi les 72) choisies arbitrairement pour effectuer notre validation. Ces fonctions expérimentales présentent toutes, de manière globale ou bien suivant un azimut choisi, une forme avec palier et sans effet de pépite sensible, à l'exception du 11/03/84 dont l'allure nettement linéaire, quelque soit l'azimut considéré, peut refléter une portée plus longue que la zone d'observation ; pour les autres journées, les portées obtenues varient entre 300 et 800 km. La figure III.17 montre deux exemples des variogrammes expérimentaux obtenus. Afin de tester la sensibilité de ces variogrammes à la dérive du champ, nous avons également traité les valeurs $X(t)$ obtenues suivant III.5 ; en fait ces variogrammes sont identiques à ceux de $Y(t)$ puisqu'ils sont insensibles, par construction, à l'addition à X d'une valeur constante U ; comme le montre la figure III.17 les résidus à $EZ(t)$ ont une structure rigoureusement identique aux variables brutes $Z(t)$ pour les distances faibles et une structure peu différente aux distances supérieures ; cela indique que l'hypothèse quasi-stationnaire de dérive localement constante s'applique bien à ce phénomène. Enfin, il peut paraître surprenant que les variogrammes identifiés ne présentent jamais d'effet de pépite, alors que les corrélogrammes de la figure III.15 en présentent un de 5 % environ ; cela est simplement dû à la faiblesse de cet effet que les incertitudes de la construction des variogrammes ne

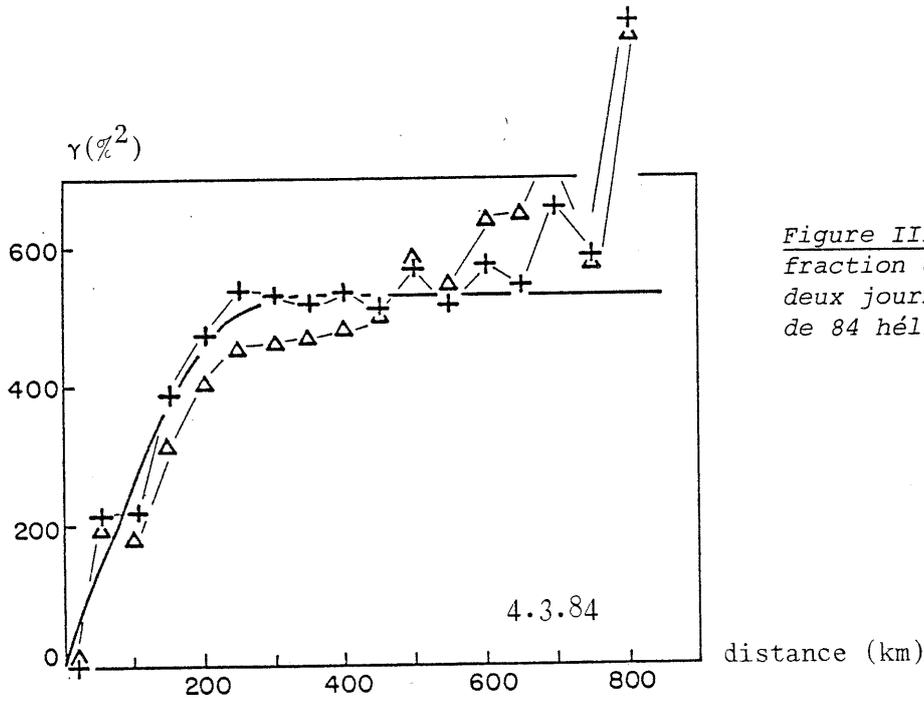
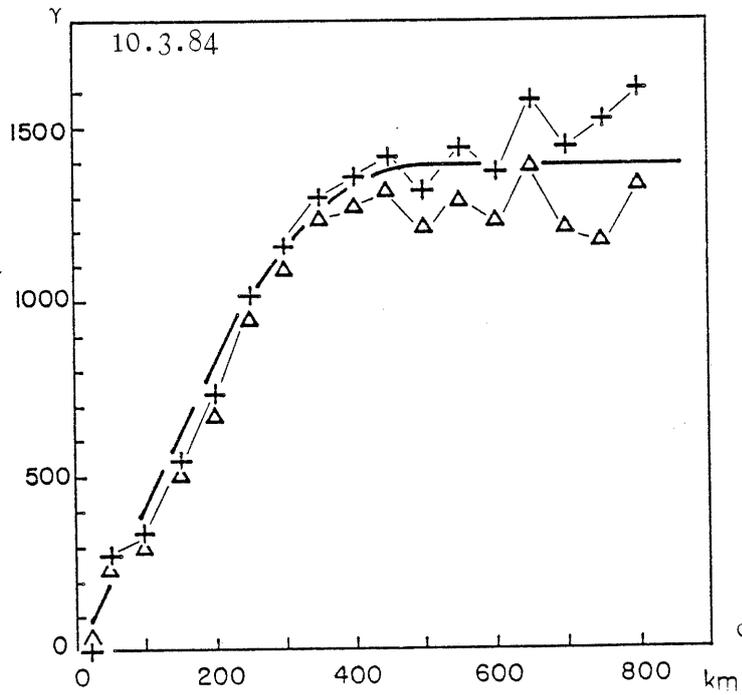


Figure III.17 : Variogrammes de la fraction d'insolation obtenus pour deux journées de Mars 1984 à l'aide de 84 héliographes.

+ Variogramme des valeurs brutes
 Δ Variogrammes des valeurs centrées deux fois.



d

permettent pas de cerner (le premier point des variogrammes pour ce réseau de 84 est calculé à l'aide de 2 couples de stations).

En conclusion, cette analyse de structure montre l'intérêt d'une approche climatologique en raison de sa simplicité et de sa robustesse (fonctions de structure expérimentales plus stables), même si certaines réserves peuvent être faites quant à l'homogénéité de l'ensemble des réalisations.

b) - Résultats de validation du réseau d'étalonnage seul

De la même manière que dans l'exemple précédent de la pluviométrie, nous avons entrepris de valider les reconstitutions de la fraction journalière d'insolation à l'aide d'un réseau d'héliographes afin que les performances de ce dispositif classique puisse servir de base de comparaison pour d'autres dispositifs tels que le modèle CACTUS.

Pour effectuer cette validation, nous avons, dans un premier temps, calculé par krigeage les valeurs de référence de la fraction d'insolation intégrée sur différentes cibles de validation ; le choix de ces cibles est lié au dispositif de télédétection que nous voulons valider par la suite (ici la méthode CACTUS) et au choix du réseau sol de validation (ici 61 stations héliographiques repérées sur la figure III.12). Ce krigeage peut être mené dans le cadre de deux modèles différents :

- le modèle climatologique dont nous avons présenté l'inférence ci-dessus ; les différentes étapes de calcul respectent alors comme nous l'avons vu la décomposition III.5 (krigeage des $EZ(t)$ puis des $Y^W(t)$) avec une économie de calcul intéressante puisque les systèmes de krigeage sont résolus une seule fois pour l'ensemble des réalisations ; malheureusement, cette approche fournit pour chaque cible une variance d'estimation des $Z^W(t)$ valable pour l'ensemble des réalisations w ce qui rend peu utilisables les critères fondés sur la comparaison des variances d'estimation réalisation par réalisation (diagramme d'efficacité par exemple) ;

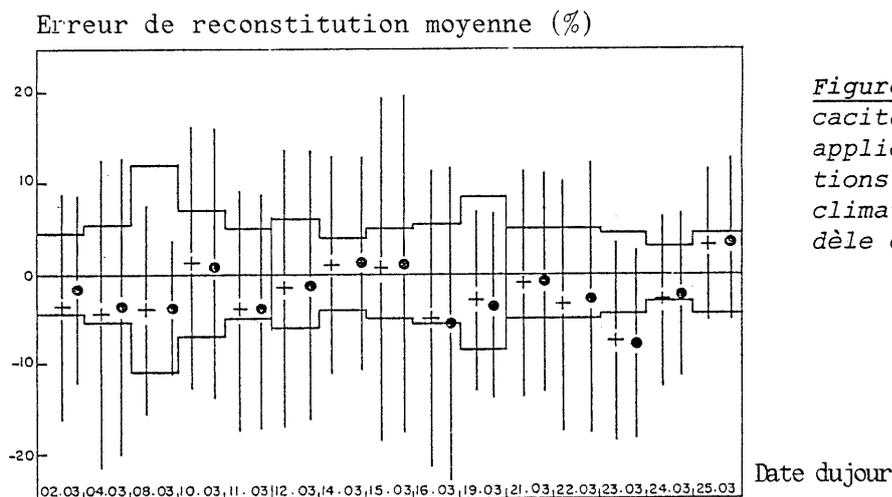


Figure III.18 : Diagramme d'efficacité du système de krigeage appliqué à un réseau de 13 stations dans le cadre d'un modèle climatologique (●) et d'un modèle quasi-stationnaire (+)

Figure III.19 : Diagramme d'efficacité du système de krigeage dans le cadre d'un modèle climatologique pour un réseau de 23 stations (o) puis 13 stations (●).

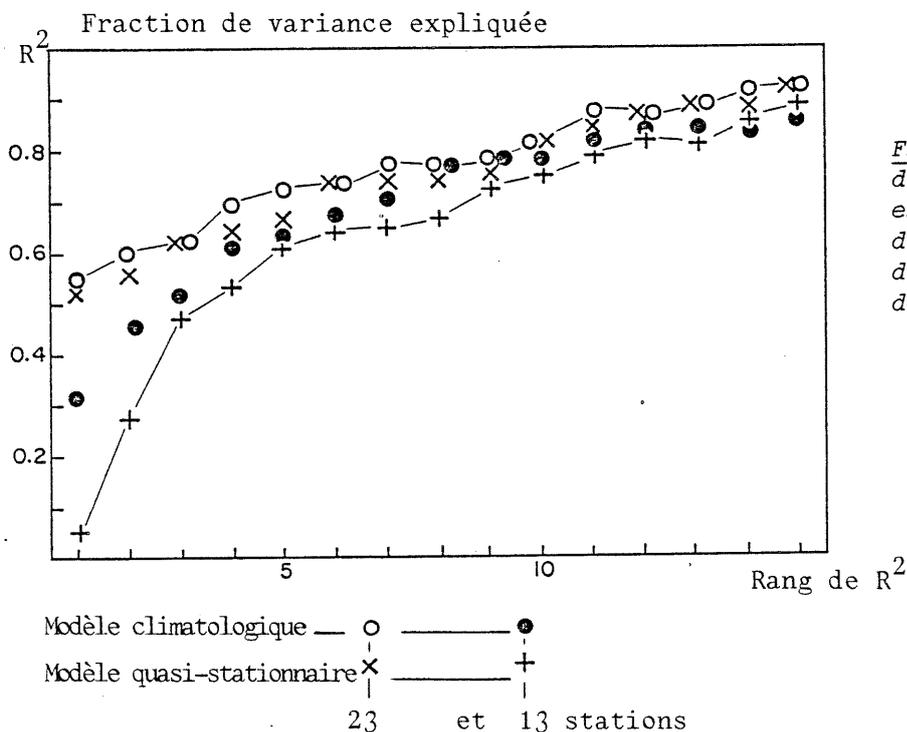
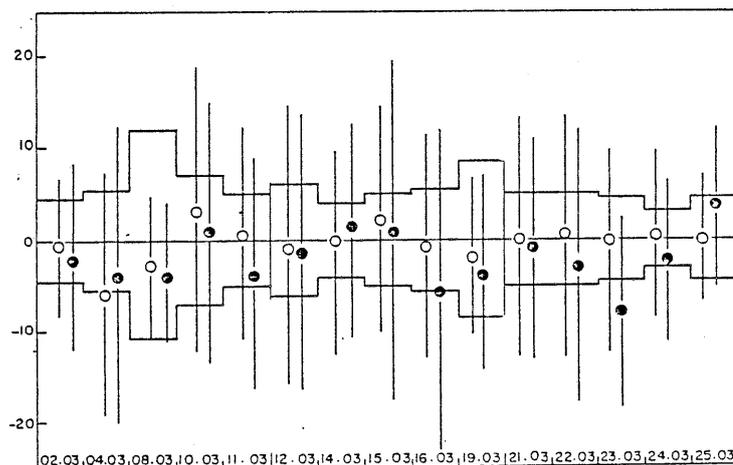


Figure III.20 : Courbes d'efficacité des différents essais de reconstitution de la fraction journalière d'insolation sur 61 cibles de validation.

- le modèle quasi-stationnaire, adapté à chaque réalisation par inférence d'un variogramme (voir table III.2) dont l'utilisation est plus lourde, mais qui donne en chaque cible une variance d'estimation propre à chaque réalisation.

Etant donné la grande similitude des valeurs de référence calculées par ces deux modèles (corrélation de .98 entre les valeurs fournies) dans le contexte où nous les appliquons ici (intégration sur une cible contenant un point de mesure), nous avons choisi le second modèle afin de préserver plus de "nuances" au critère du diagramme d'efficacité ; pour mémoire, nous porterons cependant sur ces diagrammes les variances d'estimation fournies par le premier modèle.

A l'aide des seules stations héliographiques restantes (23 stations dites d'étalonnage -voir figure III.12-) nous avons dans un deuxième temps réalisé différents essais de reconstitutions des valeurs de référence :

- en appliquant le système de krigeage dans le cadre du modèle climatologique puis quasi-stationnaire ;

- en réduisant le nombre de stations d'étalonnage de 23 à 13 (c'est à dire en diminuant la densité des points de mesure de $1/(200)^2$ à $1/(300)^2$ station par km^2 environ).

Les résultats obtenus à l'aide des critères habituels appellent les remarques suivantes :

- le diagramme d'efficacité figure III.18, comparant les performances des deux modèles appliqués à un réseau de 13 stations (réseau peu dense qui est a priori le plus susceptible de révéler des différences), montre une très forte co-fluctuation des performances avec un très léger avantage au modèle climatologique qui se traduit globalement par une erreur absolue moyenne de 22 % de la valeur moyenne de la référence contre 24 % pour le modèle quasi-stationnaire ; les écarts-types des erreurs sont très comparables.

Figure III.21 : Diagramme d'efficacité du modèle Cactus (\square) sur 61 cibles de validation ; en référence sont portés les résultats du modèle climatologique (\circ) appliqué à 23 stations.

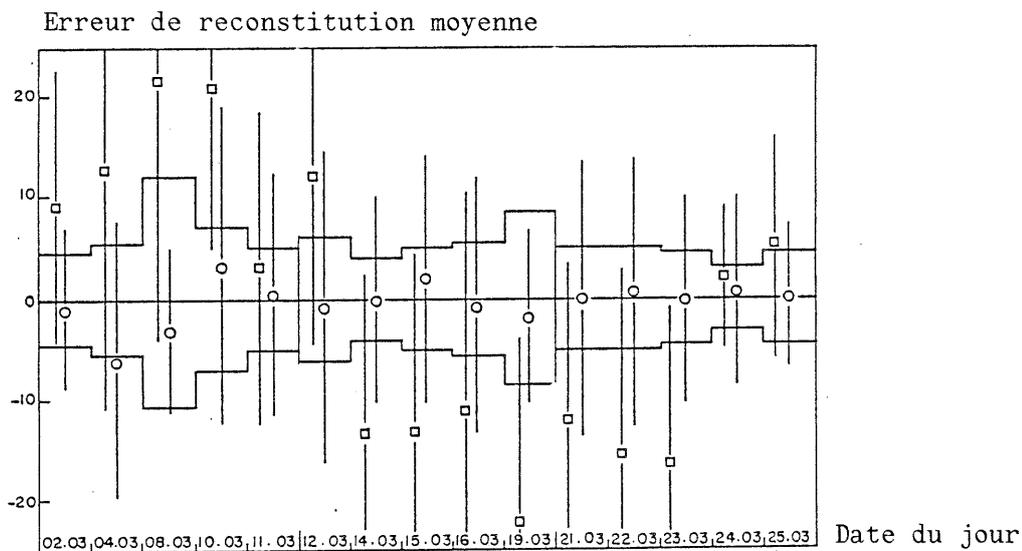
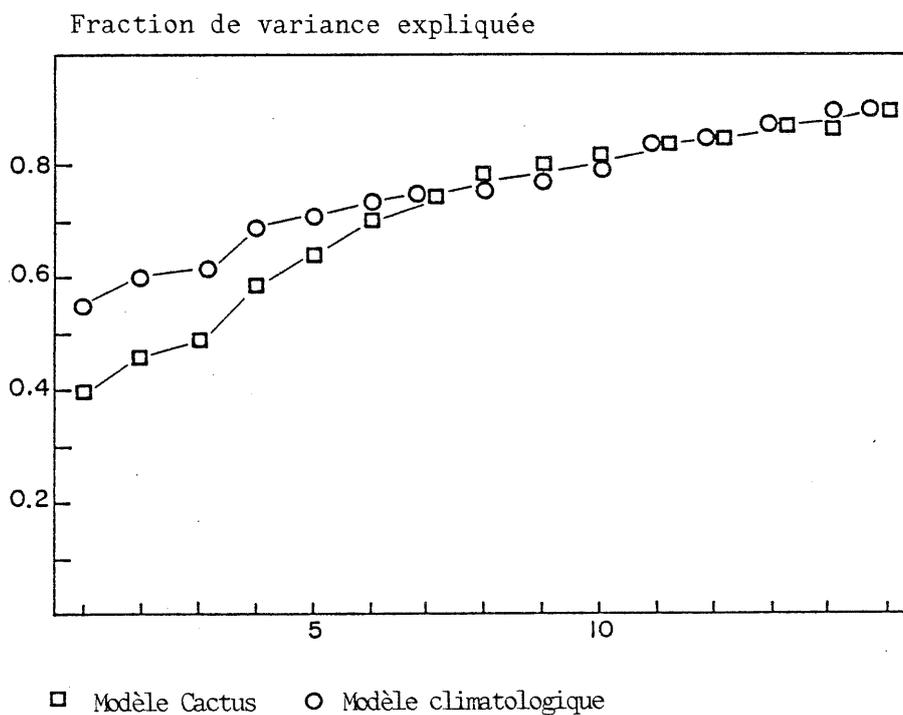


Figure III.22 : Courbe d'efficacité du modèle Cactus sur 61 cibles de validation ; en référence sont portés les résultats du modèle climatologique appliqué à 23 stations.



- le diagramme d'efficacité figure III.19, comparant les performances des réseaux de 13 et 23 stations dans le cadre du modèle climatologique montre qu'une augmentation de la densité provoque une réduction légère de la moyenne des erreurs sans réduire sensiblement leur dispersion.

- les courbes d'efficacité figure III.20 montre tout d'abord que pour les 10 meilleures reconstitutions, les différents essais conduisent à des pourcentages de variance expliquée très comparables ; pour les 5 plus mauvaises reconstitutions, seuls les essais à l'aide de 13 stations montrent nettement une faiblesse du modèle quasi-stationnaire.

La fraction d'insolation apparaît donc comme une variable dont la mesure à l'aide de dispositifs sol classiques est généralement fiable même avec des réseaux de très faible densité.

c) - Résultats de validation du modèle CACTUS

En comparant les valeurs de la fraction journalière d'insolation calculées par le modèle CACTUS avec les valeurs de référence sur les 61 cibles de validation décrites ci-dessus, nous avons obtenu les résultats suivants :

- les mesures du modèle CACTUS présentent, par rapport aux valeurs de référence, un biais significatif de signe et de valeur variables pour 13 journées sur 15 ; en effet, la figure III.21 montre que l'erreur moyenne de reconstitution par CACTUS sort 13 fois sur 15 de la "fourchette" théorique de variabilité de la référence (cette erreur se trouve 10 fois à plus de deux écarts-types d'estimation de la référence) ; les mesures de ce modèle présentent de plus une dispersion autour des valeurs de référence systématiquement plus forte que celle du modèle climatologique présenté sur le même diagramme. D'une manière générale, l'erreur absolue moyenne commise par les mesures CACTUS est de 38 % de la valeur moyenne de la référence.

- en revanche, les mesures du modèle CACTUS coïncident spatialement de manière très satisfaisante avec les valeurs de référence comme l'indique la courbe d'efficacité figure III.22 ; les 10 meilleures

reconstitutions de CACTUS, au sens de ce critère de variance de la référence expliquée, équivalent aux meilleures reconstitutions obtenues à l'aide d'un réseau sol de 23 héliographes ; d'un point de vue plus global, le coefficient de corrélation entre valeur CACTUS et valeur de référence, centrées par rapport à la moyenne de leurs réalisations respectives, est de .84 (contre .83 pour le modèle climatologique), ce qui confirme l'équivalent de leurs performances.

Enfin, le calcul de différents coefficients de corrélation partielle permet d'apprécier la complémentarité des deux dispositifs de mesure : d'une part, lorsque l'on dispose déjà d'un réseau de 23 (resp. 13) stations, le recours à l'information CACTUS permet d'expliquer 8% (resp. 7%) de variance de la référence supplémentaire ; d'autre part, l'utilisation des seules mesures CACTUS laisse "échapper" 14% (resp. 19%) de variance de la référence qu'un réseau de 23 (resp. 13) héliographes permettrait d'expliquer. Ces pourcentages, traduits en terme de corrélation partielle (c'est à dire normés par la variance résiduelle totale) montrent que dans les deux cas décrits ci-dessus, les gains que l'on peut attendre de l'autre dispositif sont substantiels (entre .32 et .61), ceci traduit clairement une forte complémentarité de ces deux types de mesures.

En résumé, le modèle CACTUS satisfait donc notre attente vis à vis d'un dispositif de télédétection à savoir une bonne perception de la variabilité spatiale du phénomène en dépit d'une mauvaise précision absolue de la mesure.

CHAPITRE IV

LE COKRIGEAGE :
RAPPELS THEORIQUES DANS LE CADRE DE L'ETALONNAGE D'IMAGES

Résumé

Ce chapitre présente une généralisation du système de krigeage au cas où plusieurs variables peuvent être utilisées lors de l'estimation d'une variable principale, dite variable d'intérêt. Compte-tenu de la diversité potentielle des propriétés statistiques de ces variables, le système de cokrigeage doit être capable de prendre en compte une certaine hétérogénéité de structure, ce qui le différencie du krigeage.

Après la description de quelques cas d'applications, nous développerons ce système d'abord, pour des raisons de clarté, dans le contexte simple de deux variables qui nous intéressent en étalonnage d'images; nous en donnerons ensuite la formulation générale, puis nous évoquerons les problèmes nouveaux de modélisation de dérive et des fonctions de structure ; enfin, nous indiquerons le mode d'application que nous avons choisi pour l'étalonnage.

Cette présentation s'appuie pour l'essentiel sur les documents suivants : D. FRANCOIS-BONGARCON, 1981, J.P. MARBEAU, 1978, A. MARECHAL, 1970, G. MATHERON, 1970, P. CHAUVET, 1977 et D.E. MYERS, 1982, auxquels il est indispensable de se reporter pour échapper au cadre volontairement restreint de ce chapitre.

BUT DU COKRIGEAGE

Nous avons vu au Chapitre I que le krigeage est une méthode qui permet l'estimation d'une variable régionalisée $z(t,w)$ en tout point ou ensemble de points à partir de la connaissance d'un certain nombre de résultats de mesures $z(t_i,w)$ réalisées aux noeuds d'un réseau non nécessairement régulier ($t_i, i=1$ à n) à un instant w .

Dans la pratique de l'étude quantitative de phénomènes naturels, il existe parfois une relation entre la variable que l'on cherche à

quantifier, que nous appellerons variable d'intérêt, et d'autres variables caractérisant un autre aspect de ce phénomène (exemple de la liaison entre la pression atmosphérique et le vent, voir P. CHAUVET et al., 1976) ou bien caractérisant un phénomène différent (exemple du contenu en eau d'un sol et de sa composition granulométrique, voir M. VAUCLIN et al., 1983) ; il peut également exister une relation entre les différentes mesures d'un phénomène à des instants successifs (exemple des champs de pression successifs, voir également P. CHAUVET et al. 1976), mais cet aspect temporel ne sera pas examiné ici.

Lorsque cette relation est analytiquement simple (cas de l'hypothèse géostrophique qui permet de relier par dérivation des composantes du vent au géopotential), il est bien entendu nécessaire de la prendre en compte et on peut espérer se ramener à l'étude d'une seule variable.

Par contre, lorsque la complexité physique du ou des phénomènes étudiés rendent cette relation difficile à paramétrer simplement, une approche de type stochastique semble la plus appropriée pour la prendre en compte dans une estimation : le cokrigeage se propose de réaliser cette approche, c'est à dire d'estimer une variable régionalisée $z(t,w)$ à l'aide des mesures d'autres variables $z_k(t,w)$.

Un second contexte d'application du cokrigeage peut être le cas où l'on veut effectuer l'estimation simultanée de plusieurs variables liées entre elles, soit pour obtenir la valeur d'une combinaison de ces variables sur un domaine donné (exemple, dans le domaine minier, de l'estimation de la minéralisation cumulée de plusieurs éléments sur un panneau ; voir A. MARECHAL, 1970) soit pour respecter une contrainte globale portant sur différentes variables (exemple de l'estimation de la puissance totale d'une couche géologique et des puissances des sous-couches qui la composent ; voir G. MATHERON, 1979).

Ces deux grandes utilisations du cokrigeage présentent en fait des caractéristiques d'application assez différentes :

- la première, qui consiste à compléter l'information acquise à propos d'une variable d'intérêt, n'a généralement pas recours à plus d'une variable secondaire et tire surtout son efficacité de la plus forte densité des mesures de cette variable secondaire (les réseaux utilisés étant de densités différentes on qualifie ce cas d'hétérotopique) ;

- la seconde, qui consiste à traiter simultanément plusieurs sources d'information, pourra éventuellement porter sur plusieurs variables mesurées par des réseaux de densité comparable voir confondus (cas isotopique).

Les applications du cokrigeage proposées dans cette étude concernant toutes la première utilisation du cokrigeage, les développements présentés ci-après se limitent au cas d'une seule variable secondaire qui permet sans trop de lourdeur de garder les mêmes systèmes de notations qu'au Chapitre I ; pour le cas où plusieurs variables secondaires seraient disponibles, il est évident que les notations matricielles proposées par D.E. MYERS, 1982 sont plus agréables à manipuler parce que plus synthétiques.

Nous noterons z_T la variable secondaire par référence à son origine qui dans notre cas est la Télédétection (mesure de la pluie par radar ou satellite, mesure de l'insolation par satellite). Cette variable, mesurée aux noeuds d'une grille régulière (image) que nous noterons $(t_g,$ $g=1,G)$, représente dans le cas des mesures de télédétection une intégration du processus z_T sur la maille T_g centrée en t_g et nous noterons :

$$z_T(t_g, w) = 1/T_g \int_{T_g} z_T(t, w) dt$$

la valeur intégrée du processus.

SYSTEME DE COKRIGEAGE

La démarche qui conduit à ce système est naturellement parallèle à celle décrite dans le cas du krigeage : nous nous proposons d'imposer des propriétés de nature statistique (non biais et optimalité) à un estimateur linéaire de la forme :

$$z^*(t_0, w) = \sum_{i=1}^n \lambda^i(t_0, w) z(t_i, w) + \sum_{g=1}^G \lambda^g(t_0, w) z_T(t_g, w) \quad (\text{IV.1})$$

Si l'on considère les fonctions z et z_T comme des réalisations particulières des fonctions aléatoires Z et Z_T , cet estimateur peut s'appliquer à l'ensemble des fonctions des familles Z et Z_T :

$$Z^*(t_0) = \sum_{i=1}^n \lambda^i(t_0) Z(t_i) + \sum_{g=1}^G \lambda^g(t_0) Z_T(t_g) \quad (\text{IV.2})$$

ce qui permet de traduire la condition de non-biais :

$$E(Z^*(t_0) - Z(t_0)) = 0 \quad (\text{IV.3})$$

par la formule :

$$M = \sum_{i=1}^n \lambda^i(t_0) EZ(t_i) + \sum_{g=1}^G \lambda^g(t_0) EZ_T(t_g) - EZ(t_0) = 0 \quad (\text{IV.4})$$

et la condition d'optimalité :

$$E(Z^*(t_0) - Z(t_0))^2 \text{ minimal} \quad (\text{IV.5})$$

par l'expression :

$$\begin{aligned} V = & \sum_{i=1}^n \sum_{j=1}^n \lambda^i(t_0) \lambda^j(t_0) EZ(t_i) \cdot Z(t_j) + \sum_{g=1}^G \sum_{g'=1}^G \lambda^g(t_0) \lambda^{g'}(t_0) EZ_T(t_g) \cdot Z_T(t_{g'}) \\ & + EZ(t_0)^2 + 2 \sum_{i=1}^n \sum_{g=1}^G \lambda^i(t_0) \lambda^g(t_0) EZ(t_i) \cdot Z_T(t_g) \\ & - 2 \sum_{i=1}^n \lambda^i(t_0) EZ(t_i) Z(t_0) - 2 \sum_{g=1}^G \lambda^g(t_0) EZ_T(t_g) \cdot Z(t_0) \text{ minimal} \end{aligned} \quad (\text{IV.6})$$

La minimisation de cette expression, sous la contrainte de non biais exprimée par IV.4, s'effectue en dérivant par rapport aux différents

paramètres λ et ρ l'expression $V + 2\rho M$; on obtient ainsi le système de cokrigage de deux variables :

$$\begin{aligned}
 & \sum_{i=1}^n \lambda^i(t_0)EZ(t_i).Z(t_j) + \sum_{g=1}^G \lambda^g(t_0)EZ_T(t_g).Z(t_j) \\
 & - EZ(t_j)Z(t_0) + \rho.EZ(t_j) = 0 \quad \text{pour } j = 1 \text{ à } n \\
 & \sum_{g=1}^G \lambda^g(t_0)EZ_T(t_g).Z_T(t_{g'}) + \sum_{i=1}^n \lambda^i(t_0)EZ(t_i).Z_T(t_{g'}) \\
 & - EZ_T(t_{g'})Z(t_0) + \rho.EZ_T(t_{g'}) = 0 \quad \text{pour } g' = 1 \text{ à } G \\
 & \sum_{i=1}^n \lambda^i(t_0)EZ(t_i) + \sum_{g=1}^G \lambda^g(t_0)EZ_T(t_g) = EZ(t_0)
 \end{aligned} \tag{IV.7}$$

Lorsque λ et ρ vérifient ce système, l'expression IV.6 devient :

$$\begin{aligned}
 V = & - \sum_{i=1}^n \lambda^i(t_0)EZ(t_i).Z(t_0) - \sum_{g=1}^G \lambda^g(t_0)EZ_T(t_g).Z(t_0) \\
 & + EZ(t_0)^2 - \rho.EZ(t_0)
 \end{aligned}$$

qui est la variance d'estimation par cokrigage.

La linéarité de l'estimateur IV.2 permet, comme dans le cas du krigeage, d'obtenir directement la valeur intégrée du phénomène sur une zone A en substituant A à t_0 dans le système IV.7.

Comme nous l'avons souligné plus haut, la notation matricielle du système VI.7 permet de traiter de manière plus synthétique le cas de plusieurs variables où l'estimateur IV.2 se note (si la variable d'intérêt est indiquée par 0) :

$$Z^*_0(t_0) = \sum_{k=0}^K \sum_{i=1}^{n_k} \lambda^i_k(t_0) Z_k(t^k_i) \tag{IV.8}$$

et pour lequel les conditions de non-biais et d'optimalité conduisent selon la même démarche que pour 2 variables au système :

$$\begin{bmatrix} |S_{k,k'}|_{K \times K} & D \\ D^T & 0 \end{bmatrix} \times \begin{bmatrix} \hat{\mu} \end{bmatrix} = \begin{bmatrix} S_0 \\ D_0 \end{bmatrix} \tag{IV.9}$$

avec :

$$S_{k,k'} = |EZ_k(t_i^{k_i}) \cdot Z_{k'}(t_j^{k'})|$$

$$D^T = ||EZ_1(t_1^1)|, \dots, |EZ_K(t^K_i)|| \quad \begin{matrix} K \\ 1 \times n_k \\ k=0 \end{matrix}$$

$$\Lambda^T = ||\lambda_1^i(t_0)|, \dots, |\lambda_K^i(t_0)|| \quad \begin{matrix} K \\ 1 \times n_k \\ k=0 \end{matrix}$$

$$S_0 = |EZ_k(t_i^{k_i}) \cdot Z_0(t_0)| \quad \begin{matrix} K \\ 1 \times n_k \\ k=0 \end{matrix}$$

et $D_0 = EZ_0(t_0)$

L'estimateur $Z^*(t_0)$ peut alors comme dans le cas du krigeage se noter sous deux formes suivant l'ordre de multiplication des différents facteurs qui le composent :

$$Z^*_{t_0}(t_0) = \begin{bmatrix} Z \\ 0 \end{bmatrix}^T \times \left(\begin{bmatrix} |S_{k,k'}|_{K \times K} & D \\ D^T & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} S_0 \\ D_0 \end{bmatrix} \right) \tag{IV.10}$$

avec :

$$Z^T = ||Z_0(t_1)|, \dots, |Z_K(t^K_i)|| \quad \begin{matrix} K \\ 1 \times n_k \\ k=0 \end{matrix}$$

ou bien :

$$Z^*_{t_0}(t_0) = \left(\begin{bmatrix} Z \\ 0 \end{bmatrix}^T \times \begin{bmatrix} |S_{k,k'}|_{K \times K} & D \\ D^T & 0 \end{bmatrix}^{-1} \right) \times \begin{bmatrix} S_0 \\ D_0 \end{bmatrix} \tag{IV.11}$$

Ce dernier conditionnement (interpolateur) fait apparaître un vecteur de poids issu de la première multiplication :

$$\begin{bmatrix} A \\ b \end{bmatrix}^T = \begin{bmatrix} Z \\ 0 \end{bmatrix}^T \times \begin{bmatrix} |S_{k,k'}| & D \\ D^T & 0 \end{bmatrix}^{-1} \quad (\text{IV.12})$$

avec : $A^T = ||a^1_i||, \dots, ||a^K_i||$ $\begin{matrix} K \\ 1 \times n_k \\ k=0 \end{matrix}$

qui caractérise globalement l'interpolateur $Z^*(t_0)$:

$$Z^*(t_0) = \sum_{i=1}^n a_i^k EZ(t_i) \cdot Z(t_0) + \sum_{g=1}^G a_g EZ_T(t_g) \cdot Z(t_0) + bEZ(t_0) \quad (\text{IV.13})$$

les coefficients a et b étant solutions du système suivant :

$$Z^*(t_j) = Z(t_j) \quad \text{pour } j = 1 \text{ à } n$$

$$\begin{aligned} & Z^*(t_{g'}) + \sum_{i=1}^n a_i (EZ(t_i) \cdot Z_T(t_{g'}) - EZ(t_i) \cdot Z(t_{g'})) \\ & + \sum_{g=1}^G a_g (EZ_T(t_g) \cdot Z_T(t_{g'}) - EZ(t_g) \cdot Z_T(t_{g'})) = Z(t_{g'}) \quad \text{pour } g'=1 \text{ à } G \\ & \sum_{i=1}^n a_i EZ(t_i) + \sum_{g=1}^G a_g EZ(t_g) = 0 \end{aligned}$$

L'application de tels systèmes à des phénomènes réels, suppose bien sûr le même effort de modélisation que celui décrit à propos du krigeage. Le modèle requis doit même être a priori plus riche puisqu'il est nécessaire d'étendre les propriétés statistiques de stationnarité à la deuxième variable utilisée ainsi qu'à la liaison entre ces variables.

Nous ne reprendrons pas ici la description des critères de choix de modèle qui restent valides quelle que soit la complexité du modèle, de

même que la description des différents aspects de l'inférence dans le temps et dans l'espace.

Par contre, il est nécessaire de préciser les hypothèses supplémentaires qu'exige l'extension au cas de deux variables du modèle stationnaire.

MODELE STATIONNAIRE ET COREGIONALISATION

Le système de cokrigeage, par la linéarité de l'estimateur qu'il propose et la nature statistique des contraintes qu'il impose à cet estimateur, ne recourt en fait qu'aux différents moments d'ordre 1 et 2 des variables étudiées :

$$Ez(t), Ez_T(t), Ez(t).z(t'), Ez_T(t).z_T(t') \text{ et } Ez(t).z_T(t')$$

Les hypothèses de stationnarité nécessaires à l'inférence de ce système peuvent donc se réduire à ces différents moments ; on parlera alors de stationnarité d'ordre 2 de la corégionalisation, qui peut se résumer comme suit :

$$EZ(t) = m$$

$$EZ_T(t) = m_T$$

$$EZ(t).Z(t') = c_{ZZ}(t-t') (= c(t-t')) \quad (\text{IV.15})$$

$$EZ_T(t).Z_T(t') = c_{Z_T, Z_T}(t-t') (= c_{T T}(t-t'))$$

$$EZ(t).Z_T(t') = c_{Z, Z_T}(t-t') (= c_{T Z}(t-t'))$$

Cette stationnarité contient naturellement les hypothèses de stationnarité d'ordre 2 des variables Z et Z_T considérées séparément auxquelles s'ajoute l'hypothèse de stationnarité des moments croisés : on admet ainsi que la covariance entre les mesures de Z et Z_T ne dépend que de la distance $t-t'$ séparant les points de mesure, indépendamment du lieu de ces points ; la fonction $c_T(h)$ représentant l'évolution de cette covariance est appelée covariance croisée ou rectangle.

Il est important de noter que cette hypothèse supplémentaire n'est pas contenue dans la stationnarité des variables Z et Z_T , en effet la stationnarité d'ordre 2 de Z et Z_T n'implique pas la stationnarité des moments croisés. Un contre-exemple simple peut être construit de la manière suivante : imposons à deux processus stationnaires d'ordre 2 d'être liés algébriquement en un point t_0 par la relation $Z(t_0) = aZ_T(t_0)$; en ce point la covariance $c_T(0) = EZ(t_0) \cdot Z_T(t_0)$ atteint une valeur maximale égale à $EZ^2(t_0)$ que l'on n'a aucune chance de retrouver partout ailleurs dans le champ : elle n'est donc pas stationnaire.

Notons au passage que la fonction différence $Z(t) - Z_T(t)$ est alors stationnaire d'ordre 2 puisque :

$$E(Z(t) - Z_T(t)) = m - m_T \quad (\text{IV.16})$$

$$E(Z(t) - Z_T(t))(Z(t+h) - Z_T(t+h)) = c(h) - 2c_T(h) + c_{TT}(h) = c_d(h)$$

cette propriété sera utilisée lors de certaines simplifications apportées au système de cokrigeage.

Ce modèle rencontre évidemment le même type de limites dans le cas de plusieurs variables que celles évoquées à propos d'une seule variable : la valeur moyenne et la variabilité des phénomènes naturels et de leur différentes mesures varient souvent dans l'espace, ce qui rend les hypothèses de stationnarité d'ordre 2 difficilement tenables dans de nombreux cas pratiques.

Les deux voies de l'analyse objective et de la théorie intrinsèque peuvent à nouveau être utilisées pour adapter ces hypothèses. La description des grands traits de ces deux démarches présentée au Chapitre I ne sera pas reprise ici ; nous soulignons simplement dans les paragraphes suivants les aspects nouveaux de l'inférence des moments d'ordre 1 et 2 dans le cas d'une corégionalisation de deux variables.

ASPECTS NOUVEAUX DE L'INFERENCE DES MOMENTS D'ORDRE 1.

La nouveauté majeure introduite par la présence de plusieurs variables au moment de l'inférence des moments d'ordre 1 est due à la

Tableau IV.1 : Ecriture de la formule de non-biais IV.4 pour des comportements particuliers des moments d'ordre 1 des fonctions aléatoires.

Propriétés des moments d'ordre 1	Ecriture de la formule de non-biais
Moments d'ordre 1 nul ou connus (cas des F.A. Stationnaire et contexte climatologique)	
$EZ(t) = EZ_T(t) = 0$ <p>ou $EY(t) = EY_T(t) = 0$ pour tout t</p> <p>Les variables Z et Z_T ou les variables centrées Y et Y_T ont une espérance nulle.</p>	La condition de non-biais disparaît (= 0)
Moments d'ordre 1 inconnus mais constants (Cas des F.A. Stationnaire et des F.A. Intrinsèques d'ordre 0)	
$EZ(t) = EZ_T(t) = m$ pour tout t <p>Les variables ont même espérance</p> $EZ(t) = m \quad EZ_T(t) = m^T$ <p>La variable secondaire a une espérance différente de celle de la variable d'intérêt</p>	$\sum_{i=1}^n \lambda^i(t) + \sum_{g=1}^G \lambda^g(t) = 1$ $\sum_{i=1}^n \lambda^i(t) = 1$ $\sum_{g=1}^G \lambda^g(t) = 0$ <p>On retient la moyenne de la variable d'intérêt et on filtre la moyenne de la variable secondaire</p>
Non-stationnarité des moments d'ordre 1 (Cas des F.A. Intrinsèques d'ordre K)	
$EZ(t) = EZ_T(t) = \sum_{l=1}^L a_l f^l(t)$ <p>Les fonctions ont même dérive</p> $EZ(t) = \sum_{i=1}^L a_i f^l(t)$ $EZ_T(t) = \sum_{k=1}^K b_k g^k(t)$ <p>Les variables présentent des dérivées différentes et indépendantes</p> $EZ(t) = \sum_{l=1}^L a_l f^l(t)$ $EZ_T(t) = \sum_{l=1}^{L-1} a_{l+1} \cdot (l+1) \cdot f^l(t)$ <p>Les dérivées sont liées algébriquement (ici Z_T est le gradient de Z, les fonctions f sont supposées polynomiales)</p>	$\sum_{i=1}^n \lambda^i(t) f^l(t_i) + \sum_{g=1}^G \lambda^g(t) f^l(t_g) = f^l(t)$ <p style="text-align: right;">pour l = 1</p> $\sum_{i=1}^n \lambda^i(t) f^l(t_i) = f^l(t) \quad \text{pour } l = 1 \text{ à } L$ $\sum_{g=1}^G \lambda^g(t) g^k(t_g) = 0 \quad \text{pour } k = 1 \text{ à } K$ <p>On retient la dérivée de la variable d'intérêt et on filtre la moyenne de la variable secondaire</p> $\sum_{i=1}^n \lambda^i(t) f^l(t_i) = f^l(t)$ $\sum_{i=1}^n \lambda^i(t) f^l(t_i) + \sum_{g=1}^G \lambda^g(t) l f^{l-1}(t_g) = f^l(t)$ <p style="text-align: right;">pour l = 2 à L</p>

possibilité d'établir une distinction entre l'espérance $EZ(t)$ de la variable d'intérêt et l'espérance $EZ_T(t)$ de la variable secondaire. Il peut arriver que l'espérance de la variable secondaire ait une valeur (ou une évolution spatiale si cette espérance n'est pas constante dans l'espace) différente de celle de la variable d'intérêt, auquel cas la formule de non-biais IV.4 doit être séparée en deux équations (dans le cas d'une corégionalisation de deux variables) :

$$M1 = \sum_{i=1}^n \lambda^i(t_0)EZ(t_i) - EZ(t_0) = 0$$

$$M2 = \sum_{g=1}^G \lambda^g(t_0)EZ_T(t_g) = 0$$

ce qui alourdit légèrement l'écriture du système IV.7 et nécessite l'introduction de deux paramètres de Lagrange μ_1 et μ_2 dans le système de cokrigage IV.9.

Le tableau IV.1 résume pour quelques modélisations classiques des moments d'ordre 1 la traduction dans le système de cokrigage d'une telle distinction entre les valeurs moyennes de deux variables.

Cette distinction n'a bien sûr pas lieu d'être mentionnée dans le cas de fonctions aléatoires stationnaires de moyenne nulle ; seul, dans la pratique, un contexte climatologique permet de se ramener à ce cas très favorable, qui laisse par la suite subsister le problème de l'estimation de la valeur moyenne en des points non mesurés (voir les limites de l'analyse objective évoquées au Chapitre I).

Dans le cas de fonctions aléatoires d'espérance inconnue (constante ou lentement variable dans l'espace), la modélisation des dérivées à l'aide de fonctions polynomiales dans le cadre de l'hypothèse intrinsèque conduit à une condition de non-biais très proche de celle du krigage universel lorsque ces dérivées sont supposées identiques pour les deux variables ; cette condition de non-biais devient plus lourde lorsque les dérivées sont supposées distinctes ou liées algébriquement.

Tableau IV.2 : Fonctions de structure croisée pour différents modèles de fonctions aléatoires.

Modèle de fonction aléatoire utilisé	Contexte habituel	Codage utilisé	Fonction de structure croisée
F.A. Stationnaire	quelconque	aucun	Covariance croisée simple : $C_{Z,Z_T}(t-t') = EZ(t) \cdot Z_T(t')$
	climatologique	Centrage : $Y(t) = Z(t) - EZ(t)$	Covariance croisée centrée : $C_{Z,Z_T}(t-t') = C_{Z,Z_T}(t-t') - EZ(t) \cdot EZ_T(t')$
	climatologique	Centrage et réduction : $X(t) = Y(t) / C_Z(0)$	Corrélation croisée : $R_{Z,Z_T}(t-t') = C_{Z,Z_T}(t-t') / C_Z(t-t) \cdot C_{Z_T}(t'-t')$
F.A. Intrinsèque d'ordre 0	spatial	Différence première : $Z(t) - Z(t+h)$	Variogramme croisé : $\gamma_{Z,Z_T}(t-t') = \frac{1}{2} E(Z(t)-Z(t')) \cdot (Z_Z(t)-Z_T(t'))$
F.A. Intrinsèque d'ordre k	spatial	Différence d'ordre k : $dZ(t_i, \lambda) = Z(t) - \sum_{\alpha} \lambda^{\alpha} Z(t_{\alpha})$	Covariance croisée généralisée : $EdZ(t_i, \lambda) \cdot dZ_T(t_i, \lambda') = \sum_i \sum_j \lambda_i \lambda'_j K_{Z,Z_T}(t_i - t'_j)$

Ces modélisations classiques, assimilant la tendance systématique des mesures analysées à une fonction polynomiale, ne sont en fait qu'un remède approximatif apporté par les statisticiens pour se ramener à des variables stationnaires. Il est important de rappeler que le système de cokrigeage, comme le système de krigeage, est tout à fait capable d'accepter une description plus précise de cette tendance systématique si des considérations physiques sont à même de la lui fournir : par exemple, lorsque cette tendance est due en partie à une imperfection du capteur utilisé, il est possible de retenir parmi les fonctions de dérive la fonction de correction des mesures issue de l'analyse physique du fonctionnement du capteur.

ASPECTS NOUVEAUX DE L'INFERENCE DES MOMENTS D'ORDRE 2

La prise en considération simultanée de plusieurs variables se justifie, comme nous l'avons dit, par l'existence de liaisons statistiques entre ces différentes variables ; il est donc nécessaire, en plus de la structure directe de chaque variable, d'identifier les structures croisées liant les différents couples de variables ; cette modélisation est réalisée à l'aide de fonctions de structure proches de celles utilisées à propos des structures directes.

Réaliser l'inférence simultanée de plusieurs fonctions de structure directes et croisées pose d'une part un problème spécifique nouveau dû à l'existence possible de liaisons statistiques décalées entre les variables et d'autre part complique fortement la vérification des conditions de positivité que doit vérifier un modèle de covariance.

Avant d'examiner en détail dans les paragraphes suivants ces deux aspects essentiels de la modélisation de structures croisées, il est nécessaire de décrire les différentes fonctions classiquement utilisées. Le tableau IV.2 résume pour divers modèles de fonctions aléatoires les codages applicables dans le contexte climatologique et spatial et les fonctions de structure croisée qui en découlent.

La covariance croisée simple et le variogramme croisé apparaissent comme des outils de base en raison de leur simplicité d'estimation empirique et de la possibilité de les appliquer en tout contexte sans préjuger des difficultés de modélisation qu'ils peuvent susciter notamment en présence de dérive ou de décalage.

Le centrage et la réduction climatologiques conduisent certainement à des fonctions très représentatives de la covariation des variables étudiées dans la mesure où une large partie de l'hétérogénéité en moyenne et en variabilité est soustraite par ce codage, mais leur contexte d'application est réduit du fait de la nécessité de séries climatologiques longues pour estimer la variabilité climatologique.

Enfin, nous mentionnons le recours possible au modèle intrinsèque d'ordre k ; en raison du caractère abstrait des covariances ainsi définies (aucune estimation empirique directe n'est possible comme nous l'avons dit au Chapitre I) il nous paraît a priori difficile d'utiliser un modèle aussi sophistiqué si l'on considère l'aspect combinatoire que prennent les problèmes d'identification déjà rencontrés dans le cas monovarié et les nouvelles contraintes de cohérence entre modèles imposés par le traitement multivarié.

Pour compléter cette description, rappelons quelques propriétés géométriques et statistiques des fonctions de covariance et des variogrammes croisés.

Au contraire de la covariance simple directe, la covariance croisée n'est pas a priori symétrique en $(h, -h)$ et en Z, Z_T ; en effet :

$$\begin{aligned}
 C_{Z, Z_T}(h) &= EZ(t) \cdot Z_T(t+h) \\
 &= EZ(t-h) \cdot Z_T(h) \quad (\text{en raison de l'hypothèse de} \\
 &\quad \text{stationnarité}) \quad \quad \quad \quad \quad \quad \quad \quad \quad (IV.18) \\
 &= EZ_T(t) \cdot Z(t-h) \quad (\text{en raison de la commutativité} \\
 &\quad \text{de la multiplication}) \\
 &= C_{Z_T, Z}(-h) \quad (\text{par définition de la covariance croisée})
 \end{aligned}$$

par contre, le variogramme croisé est symétrique en $(h, -h)$ et (Z, Z_T) puisque :

$$\begin{aligned} \gamma_{Z, Z_T}(h) &= 1/2 E(Z(t)-Z(t+h)) \cdot (Z_T(t)-Z_T(t+h)) \\ &= C_{Z, Z_T}(0) - 1/2(C_{Z, Z_T}(h)+C_{Z_T, Z}(h)) \end{aligned} \quad (\text{IV.19})$$

Cette expression se simplifie lorsque la covariance croisée est symétrique pour redonner la formule classique : $\delta_{Z, Z_T}(h) = C_{Z, Z_T}(0) - C_{Z, Z_T}(h)$.

Pour une distance nulle, $C_{Z, Z_T}(0) = EZ(t) \cdot Z_T(t)$ représente la covariance ponctuelle des deux variables que l'on peut normer par :

$C_{ZZ}(0) \cdot C_{Z_T, Z_T}(0)$ pour obtenir un coefficient de corrélation ; le variogramme s'annule lorsque $h = 0$.

Enfin, mentionnons que, d'un point de vue statistique, les matrices suivantes :

$$|C(h)| = \begin{bmatrix} C_{ZZ}(h) & C_{Z, Z_T}(h) \\ C_{Z, Z}(h) & C_{Z_T, Z_T}(h) \end{bmatrix} \quad \text{et} \quad |\delta(h)| = \begin{bmatrix} \delta_{ZZ}(h) & \delta_{Z, Z_T}(h) \\ \delta_{Z, Z}(h) & \delta_{Z_T, Z_T}(h) \end{bmatrix}$$

sont des matrices de variance-covariance et doivent donc être semi-définies positives.

Il est facile de s'assurer que les valeurs expérimentales de ces matrices vérifient cette propriété puisqu'elles sont obtenues à l'aide du produit scalaire d'un vecteur par son transposé. Le cas du variogramme est le plus simple :

$$|\delta(h)| = 1/2 |\vec{dZ}(h) \vec{dZ}_T(h)| \cdot |\vec{dZ}(h) \vec{dZ}_T(h)|^T$$

où les vecteurs ligne $\vec{dZ}(h)$ et $\vec{dZ}_T(h)$ représentent toutes les différences premières des fonctions Z et Z_T entre des points distants de h expérimentalement accessibles dans le temps, en contexte climatologique, et dans l'espace.

Dans le cas de la covariance, la valeur expérimentale de la matrice $|C(h)|$ doit être extraite d'une matrice plus complexe :

$$|C_1(h)| = |\vec{Z}(t) \vec{Z}(t+h) \vec{Z}_T(t) \vec{Z}_T(t+h)|^T \cdot |\vec{Z}(t) \vec{Z}(t+h) \vec{Z}_T(t) \vec{Z}_T(t+h)|$$

où les vecteurs $\vec{Z}(t)$ et $\vec{Z}(t+h)$ (respectivement $\vec{Z}_T(t)$, $\vec{Z}_T(t+h)$) contiennent tous les couples de mesures de Z (respectivement Z_T) établis en des points distants de h expérimentalement accessibles. La matrice $|C_1(h)|$ est une matrice (4x4) semi-définie positive ; donc ses différents mineurs le sont et en particulier $|C(h)|$, mineur d'ordre 2 obtenu en retirant les colonnes 1 et 2 et les lignes 2 et 4.

STRUCTURE CROISEE ET DECALAGE SPATIAL DES VARIABLES

La dissymétrie d'une covariance croisée se traduit par l'existence d'une direction et d'un sens privilégiés dans lesquels pour un pas d'espace $h=d$ donné la fonction $C_{Z,Z_T}(h)$ prend sa valeur maximale qui en toute logique devrait être atteinte pour $h=0$; il existe donc suivant cette direction et ce sens un décalage d'amplitude d dans la liaison statistique des variables ; par analogie au domaine temporel on appelle parfois ce décalage, effet retard, ce qui signifie que les variations de l'une des variables sont en retard par rapport à l'autre.

Afin de visualiser de manière simple la prise en compte par la covariance et le variogramme croisés d'un tel décalage, considérons dans un espace à une dimension le cas d'une fonction $Z_T(t)$ déduite par translation et amplification de la fonction stationnaire $Z(t)$:

$$Z_T(t) = \alpha \cdot Z(t+d) \quad (\text{IV.20})$$

Si la fonction de covariance de la fonction $Z(t)$ se note :

$$C_{ZZ}(h) = EZ(t) \cdot Z(t+h) \quad (\text{IV.21})$$

il est évident que la fonction $Z_T(h)$ aura pour covariance :

$$C_{Z_T,Z_T}(h) = EZ_T(t) \cdot Z_T(t+h) = \alpha^2 \cdot EZ(t+d) \cdot Z(t+d+h) = \alpha^2 \cdot C_{ZZ}(h) \quad (\text{IV.22})$$

en raison de la stationnarité de la fonction Z.

La covariance croisée $C_{Z,Z_T}(h)$ prendra pour valeur :

$$C_{Z,Z_T}(h) = EZ(t).Z_T(t+h) = \alpha.EZ(t).Z(t+d+h) = \alpha.C_{ZZ}(d+h) \quad (\text{IV.23})$$

alors que $C_{Z,Z_T}(-h)$ se note :

$$C_{Z,Z_T}(-h) = \alpha.C_{ZZ}(d-h) \quad (\text{IV.24})$$

La figure IV.1 illustre de manière schématique la construction de la covariance croisée par simple translation de la covariance directe ; on voit ainsi que l'estimation de la covariance croisée dans les deux sens (h et $-h$) pour le cas à une dimension permet de déceler un éventuel décalage ; à deux dimensions, il sera nécessaire d'identifier la direction dans laquelle l'assymétrie ($-h$, $+h$) est maximale.

Pour sa part, le variogramme croisé est sensible à la valeur du décalage mais par sa construction symétrique en (h , $-h$), il est inapte à en déceler géométriquement la valeur ; en effet, il est aisé de montrer en linéarisant la formule IV.19 lorsque $Z_T(t) = \alpha.Z(t+d)$ que :

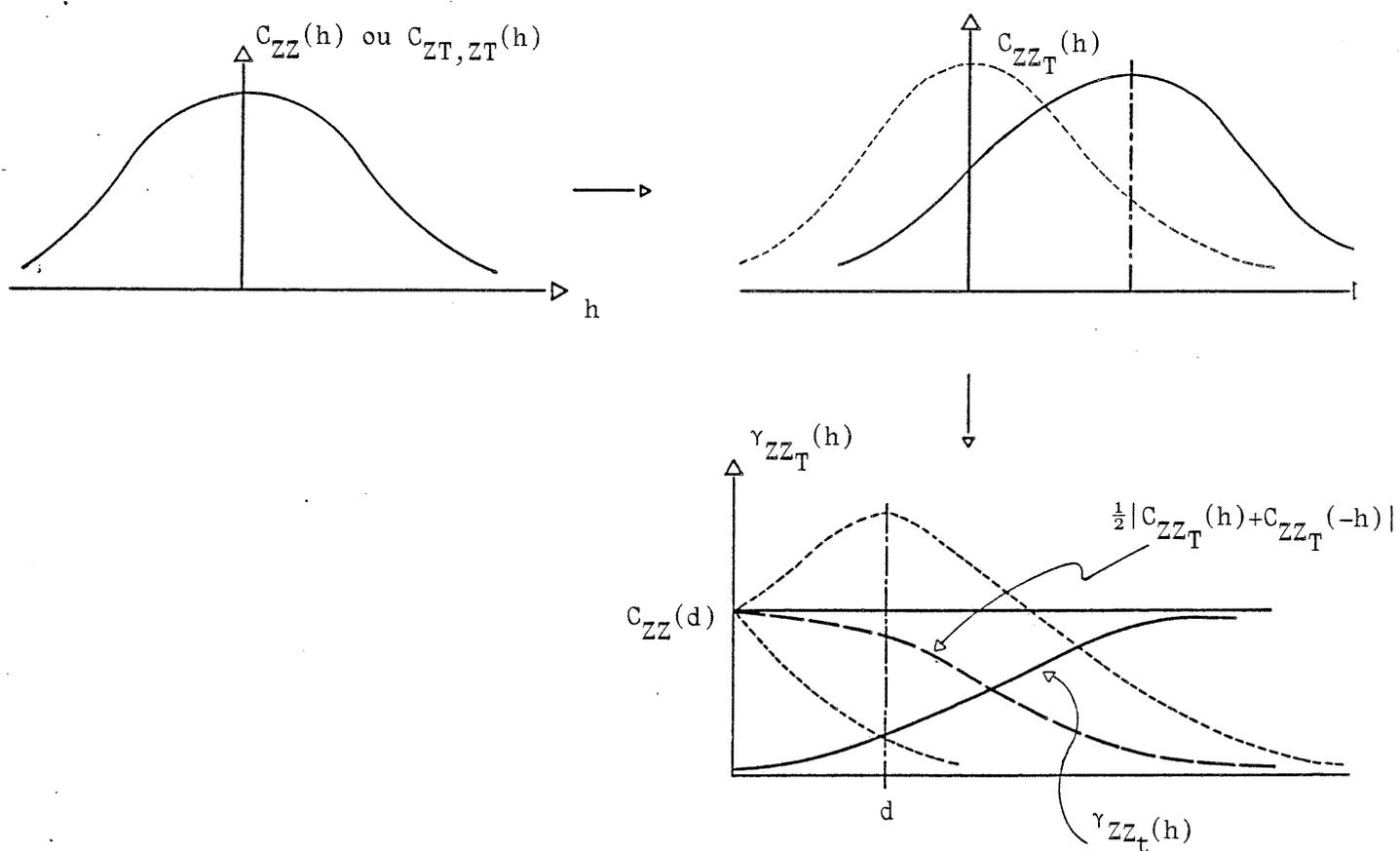
$$\delta'_{Z,Z_T}(h) = \alpha.C_{ZZ}(d) - \alpha/2 (C_{ZZ}(d+h) + C_{ZZ}(d-h)) \quad (\text{IV.25})$$

ce qui montre clairement que la prise en compte du sens de déplacement (h , $-h$) n'apporte aucune indication sur d (puisque $\delta'_{Z,Z_T}(-h) = \delta'_{Z,Z_T}(h)$) ; par contre, la valeur de δ'_{Z,Z_T} dépend de d ; si l'on suppose d nul :

$$\delta'_{Z,Z_T}(h) = C_{ZZ}(0) - C_{ZZ}(h) = \alpha.\delta'_{ZZ}(h)$$

alors en rapprochant les deux dernières formules nous voyons pour h grand par rapport à d , $C_{ZZ}(d \pm h)$ tendant vers 0, le variogramme croisé atteint en présence d'un décalage une valeur moins forte $\alpha.C_{ZZ}(d)$ que sans décalage ($C_{ZZ}(0) \geq C_{ZZ}(d)$ pour $d \neq 0$) ; pour h petit, le comportement du variogramme croisé dépend à la fois de d , de α et de la forme de la fonction de

Figure IV.1 : Illustration schématique de la prise en compte d'un décalage par une fonction de covariance et un variogramme croisés dans l'hypothèse où $Z_T(t) = Z(t+d)$.



covariance C_{ZZ} : des exemples de fonctions de covariance particulières permettent d'illustrer ce comportement :

- si $C_{ZZ}(h) = 1 - a|h|$ pour $|h| < H$ et $C_{ZZ}(h) = 0$ pour $|h| > H$ (covariance linéaire avec une portée H , qui n'est un modèle autorisé qu'à une dimension), alors :

$$\begin{aligned} \hat{\gamma}_{Z,Z_T}(h) &= 0 && \text{si } h < d < H \\ &= \alpha \cdot a(h-d) && \text{si } d < h < H \\ \text{et :} &= \alpha && \text{si } d < H < h \end{aligned}$$

il s'agit d'un des seuls modèles de covariance pour lequel le variogramme croisé permet d'estimer directement le décalage d .

- si $C_{ZZ}(h) = 1 - \cos h/a$ (fonction de covariance d'un signal sinusoïdal), alors :

$$\hat{\gamma}_{Z,Z_T}(h) = \alpha \cdot \cos d/a (1 - \cos h/a) = \alpha \cdot \cos d/a \hat{\gamma}_{ZZ}(h)$$

- enfin, si $C_{ZZ}(h) = \omega \cdot (1 - e^{-|h|/a})$ (modèle exponentiel fréquemment utilisé pour des phénomènes dont la portée est d'environ $3a$), alors :

$$\hat{\gamma}_{Z,Z_T}(h) = \alpha \cdot \omega \cdot e^{-|d|/a} (1 - e^{-|h|/a}) = \alpha \cdot e^{-|d|/a} \hat{\gamma}_{ZZ}(h)$$

Ces deux derniers exemples montrent que le variogramme croisé entre un signal et son translaté par un vecteur de décalage d amplifié par un facteur α , a généralement même forme que le variogramme direct du signal mais que ses valeurs sont diminuées par un facteur dépendant de d et de α ($\alpha \cdot \cos d/a$ ou $\alpha \cdot e^{-|d|/a}$ pour les covariances choisies). L'appréciation de d à l'aide de ce facteur n'est alors possible que lorsque le paramètre est identifiable, par comparaison des covariances directes par exemple.

Dans la pratique d'une reconnaissance de structure croisée, l'effet d'un éventuel décalage est souvent mélangé à celui de la cofluctuation imparfaite des deux variables analysées ; il est donc plus

réaliste de compliquer le modèle volontairement simpliste IV.20 de la manière suivante :

$$Z_T(t) = \alpha \cdot Z(t+d) + Y(t)$$

où $Y(t)$ représente une fonction aléatoire non corrélée à $Z(t)$. Il est alors facile de voir (puisque $EZ(t+d) \cdot Y(t) = 0$) que la covariance croisée $C_{Z,Z_T}(h)$ vérifie toujours les formules IV.23 et IV.24, ce qui signifie que l'estimation de la covariance croisée dans les deux sens (h , $-h$) permet toujours de mettre en évidence le décalage d .

Le variogramme croisé se calcule toujours suivant la formule IV.25 ; mais l'identification de α devient impossible puisque la formule IV.22 devient :

$$C_{Z_T,Z_T}(h) = \alpha^2 C_{Z,Z}(h) + C_{Y,Y}(h)$$

où la covariance de la variable Y est inconnue ; il n'est donc plus possible de distinguer dans le facteur de diminution du variogramme croisé l'effet de α de celui de d : le variogramme ne peut plus être utilisé pour identifier un éventuel décalage.

En résumé, nous dirons que l'étude de la structure croisée de deux variables régionalisées doit commencer par une vérification de la symétrie en (h , $-h$) de la covariance croisée qui est le seul outil capable de détecter géométriquement un décalage spatial des variables ; si ce décalage est physiquement justifiable, il est nécessaire de le corriger par translation d'une des variables avant de poursuivre l'étude.

Nous nous plaçons dans ce qui suit dans l'hypothèse de variables non décalées dans l'espace, c'est à dire dont la fonction de covariance croisée est une fonction monotone décroissante de h , quels que soient le sens et la direction de déplacements choisis.

CHOIX DE MODELES DES MOMENTS D'ORDRE 2 :*condition de possibilité et modèle linéaire*

Nous avons brièvement évoqué au Chapitre I la nécessité de choisir des fonctions vérifiant des conditions mathématiques de positivité pour modéliser les fonctions de structure directe (covariances ou variogrammes par exemple) ; ce problème de choix de modèle autorisé se pose à nouveau dans le cas de la modélisation de la structure d'une corégionalisation, afin de garantir que le calcul de la variance de toute combinaison linéaire des variables donne un résultat positif. Ainsi, l'expression suivante, calculée à l'aide de fonctions de covariance doit toujours prendre une valeur positive, quels que soient les pondérateurs choisis (afin de bénéficier de notations plus synthétiques, nous traiterons le cas de K variables qui peut sans difficulté se ramener à 2):

$$E\left(\sum_{k=0}^K \sum_{i=1}^{n_k} \lambda_{ik}^i Z_k(t_i)\right)^2 = \sum_k \sum_{k'} \sum_i \sum_j \lambda_{ik}^i \lambda_{jk}^j C_{kk'}(t_i, t_j) \quad (\text{IV.26})$$

avec : $C_{kk'}(t_i^{k'}, t_j^{k'}) = E Z_k(t_i) \cdot Z_{k'}(t_j)$

La vérification d'une telle condition est évidemment plus lourde que celle du cas monovarié I.22 puisque $K \cdot (K+1)/2$ modèles au lieu d'un entrent en jeu, parmi lesquels les covariances croisées dont le comportement est à la fois différent des autres (elles peuvent prendre des valeurs toujours négatives en présence d'une anticorrélation) et dépendant des autres (nous avons vu plus haut que la matrice $|C_{kk'}(h)|$ ne peut être quelconque).

Pour contourner la difficulté quasi inextricable d'une telle vérification, différentes hypothèses simplificatrices sont possibles ; elles supposent l'existence d'une famille de fonctions aléatoires orthogonales susceptibles d'engendrer linéairement les fonctions étudiées.

Si l'on note $U_l(t)$ ces fonctions de base, telles que $C_{ll'}(h) = 0$ pour tout h si $l \neq l'$, alors chaque variable Z_k peut s'écrire :

$$Z_k(t) = \sum_{I=1}^L a_{k'}^I U_I(t) \quad (\text{IV.27})$$

Les fonctions de covariance croisée des fonctions étudiées peuvent alors s'exprimer de la manière suivante :

$$\begin{aligned} C_{kk'}(h) &= E\left(\sum_{I=1}^L a_{k'}^I U_I(t) \cdot \sum_{I=1}^L a_{k'}^I U_I(t+h)\right) \\ &= \sum_{I=1}^L \sum_{I'=1}^L a_{k'}^I a_{k'}^{I'} E U_I(t) \cdot U_{I'}(t'+h) \\ &= \sum_{I=1}^L a_{k'}^I a_{k'}^I C_{II}(h) = \sum_{I=1}^L b_{kk'}^I C_{II}(h) \end{aligned} \quad (\text{IV.28})$$

Les expressions des covariances directes se déduisent immédiatement en remplaçant k' par k .

Ainsi, il apparaît que toutes les fonctions de covariance des variables Z_k s'expriment linéairement à l'aide d'un ensemble commun de fonctions de covariances ; la matrice $|C(h)|$ peut donc se noter :

$$\begin{aligned} |C(h)| &= \sum_{I=1}^L A_I \cdot |C_{II}(h)|^{K \times K} \cdot A_I^T \quad \text{avec } A_I = \begin{bmatrix} a_{k'}^I & & 0 \\ & a_{k'}^I & \\ 0 & & a_{k'}^I \end{bmatrix}^{K \times K} \\ &= \sum_{I=1}^L C_{II}(h) B_I \quad \text{avec } B_I = |b_{kk'}^I|^{K \times K} \end{aligned} \quad (\text{IV.29})$$

où $B_I = A_I \cdot A_I^T$ est une matrice de variance-covariance.

Alors il est possible de montrer que si les fonctions utilisées pour modéliser les différentes covariances C_{U_I, U_I} sont de type positif, la variance IV.26 est positive ; en effet, cette équation peut se décomposer en utilisant la formule IV.28 :

$$V = E\left(\sum_{k=0}^K \sum_{i=1}^{n_k} \lambda_k^i z_k(t_i)\right)^2 = \sum_{l=1}^L \sum_{k} \sum_{k'} \sum_i \sum_j \lambda_k^i \lambda_{k'}^j b_{kk'}^l C_{ll}(t_i, t_j) \quad (\text{IV.30})$$

qui peut se noter aussi :

$$\begin{aligned} V &= \sum_{l=1}^L \sum_{i} \sum_{j} C_{ll}(t_i, t_j) \cdot \sum_{k} \sum_{k'} \lambda_k^i \lambda_{k'}^j b_{kk'}^l \\ &= \sum_{l=1}^L \sum_{i} \sum_{j} C_{ll}(t_i, t_j) \Lambda_i B_l T_j \quad \text{avec } \Lambda_i = |\lambda_k^i|^{K \times 1} \end{aligned}$$

Comme la matrice B_l se décompose en un produit d'une matrice par sa transposée, le produit matriciel de l'expression ci-dessus se transforme comme suit :

$$\begin{aligned} \Lambda_i B_l \Lambda_i^T &= (\Lambda_i \cdot A_l) \cdot (A_l^T \cdot \Lambda_i^T) \\ &= \lambda_{i_l}^i \cdot \lambda_{i_l}^j \end{aligned}$$

La variance IV.30 peut alors se noter :

$$V = \sum_{l=1}^L \sum_{i} \sum_{j} \lambda_{i_l}^i \cdot \lambda_{i_l}^j C_{ll}(t_i, t_j) \quad (\text{IV.31})$$

Si les modèles de covariance appliqués à $C_{ll}(h)$ sont de type positif, alors les différents termes de la somme en l seront positifs, ainsi que la somme elle-même.

Faisons immédiatement une remarque concernant la symétrie des fonctions utilisées : le produit $a_{k'}^l a_k^l$, étant évidemment autotransposé et la fonction $C_{U_l, U_l}(h)$ étant symétrique en h , la matrice C est donc symétrique. Ce type de modèle de corégionalisation, dit "modèle linéaire" s'adresse donc, sous la forme présentée ici, à des variables non décalées dans l'espace (nous avons supposé au paragraphe précédent qu'il en est ainsi ; il nous semble inutilement lourd de compliquer le modèle IV.27 pour qu'il prenne en compte un décalage). Lorsque les covariances sont symétriques le calcul mené ci-dessus se généralise directement aux

variogrammes lorsque les pondérations λ_k^i forment des combinaisons linéaires autorisées ; l'équation IV.29 devient :

$$|\hat{\gamma}(h)| = \sum_{I=1}^L A_I \cdot |\hat{\gamma}_{II}(h)| \cdot A_I^T \quad (\text{IV.32})$$

et la variance IV.31 se note :

$$V = \sum_{I=1}^L \left(\sum_i \sum_j \lambda_I^i \lambda_I^j C_{II}(0) - \sum_i \sum_j \lambda_I^i \lambda_I^j \hat{\gamma}_{II}(t_i, t_j) \right)$$

il suffit alors de remarquer que le vecteur de poids $\{\lambda_I^i, i=1, n\}$, obtenu par combinaison linéaire de pondérations autorisées $\{\lambda_k^i, i=1, n_k\}$, forme lui même une combinaison linéaire autorisée ; la variance s'écrira alors :

$$V = - \sum_i \sum_j \lambda_I^i \lambda_I^j \hat{\gamma}_{II}(t_i, t_j) \quad (\text{IV.33})$$

qui est une valeur positive lorsque $\hat{\gamma}_{II}$ est calculée à l'aide d'un modèle de fonction conditionnellement négative (l'aspect conditionnel étant lié aux propriétés de la combinaison linéaire autorisée).

En résumé, le modèle linéaire de corégionalisation proposé permet, par le jeu de la linéarité des fonctions de structure, de transposer les propriétés de positivité établies dans le domaine monovarié au cas de plusieurs variables ; il suffira donc, pour obtenir différents modèles de structures directe et croisée d'une corégionalisation de combiner des modèles de fonction de structure autorisés à une variable.

Malheureusement, cette forme de modèle linéaire est d'une grande rigidité dans la mesure où les paramètres b_{kk}^I caractérisant les structures directes conditionnent totalement les structures croisées qui sont paramétrées par la moyenne quadratique $b_{k'k}^I$, de ces valeurs :

$$(|b_{k'k}^I| = \sqrt{b_{kk}^I \cdot b_{k'k}^I}).$$

ASSOUPLISSEMENT DU MODELE LINEAIRE

L'assouplissement du modèle linéaire décrit ci-dessus passe par un enrichissement de la décomposition orthogonale des variables ; les équations IV.27 peuvent être transformées en :

$$Z_k(t) = \sum_{l=1}^L \sum_{\alpha=1}^{A_l} a_{k\alpha}^l U_l^\alpha(t) \quad (\text{IV.34})$$

avec des fonctions de base U_l^α ayant même fonction de covariance pour un l donné :

$$E(U_l^\alpha(t).U_{l'}^\alpha(t')) = C_{ll'}(t,t') \quad (=0 \text{ si } l \neq l' \text{ et quel que soit } \alpha)$$

Ce nouveau modèle conduit à une décomposition de la fonction de covariance croisée :

$$C_{kk'}(h) = \sum_{l=1}^L b_{kk'}^l C_{ll}(h)$$

où la valeur $b_{kk'}^l$, s'obtient par un produit scalaire $\sum_{\alpha=1}^{A_l} a_{k\alpha}^l a_{k'\alpha}^l$ au lieu d'un produit simple $a_{k\alpha}^l a_{k'\alpha}^l$; alors la relation entre les coefficients b devient moins contraignante (d'après l'inégalité de CAUCHY-SCHWARZ appliquée au produit scalaire) :

$$|b_{kk'}^l| \leq \sqrt{b_{kk}^l \cdot b_{k'k'}^l} \quad (\text{IV.36})$$

Dans le cas de deux variables ($K=2$) cette condition garantit que la matrice H_l des coefficients b est semi-définie positive, donc représente bien une matrice de variance-covariance ; en présence de plus de variables ($K > 2$) cette condition ne vérifiant que la positivité des mineurs d'ordre 1 de la matrice H_l , il convient de s'assurer de la positivité des mineurs d'ordre supérieur à 2 avant d'affirmer que H_l est une matrice semi-définie positive, donc de variance-covariance. Cette précaution prise, la décomposition de H_l en produit d'une matrice par sa transposée ($H_l = A_l \cdot A_l^T$) permet de reprendre sans changement le raisonnement du paragraphe précédent à partir de l'équation IV.29.

De manière générale, le mécanisme pour obtenir un ensemble de fonctions de structure cohérent dans le cadre du modèle linéaire consiste simplement à approcher les différentes fonctions de structure expérimentales par combinaison linéaire de modèles autorisés suivant la formule IV.35 et à vérifier que les matrices des poids $|b_{kk'}^I|$ utilisés sont pour tous les I des matrices de variance-covariance.

La gamme des modèles de corégionalisation obtenus est très large, mais reste tout de même soumise à une condition assez forte, partiellement illustrée par la condition IV.36, qui restreint notamment l'ensemble des fonctions de structure de base disponibles pour la structure croisée à l'intersection des ensembles utilisés pour les structures directes correspondantes ($b_{kk}^I = 0 \Rightarrow b_{kk'}^I = 0$ pour tout k').

Notons enfin que le terme de corrélation intrinsèque, fréquemment rencontré dans la littérature, s'applique au cas où toutes les fonctions de base $U_I(t)$ ont même fonction de covariance ; les différents modèles de structures directes et croisées se déduisent alors de cette fonction de covariance par simple affinité.

CHOIX D'UN MODELE POUR L'ETALONNAGE D'IMAGES

Si l'on se réfère aux critères de choix de modèles évoqués au Chapitre I (d'après G. MATHERON, 1975), les différentes possibilités de modélisation que nous venons d'examiner de manière générale paraissent, par leur richesse, s'adapter à une gamme très large de réalités (satisfaction du critère c) en rendant possible la résolution du système de cokrigeage (critère b).

Afin de traiter un cas particulier tel que l'étalonnage d'images, il va être nécessaire de choisir l'un de ces modèles en réalisant un compromis entre son degré de sophistication et la possibilité de l'inférer (satisfaction du critère a). Etant donné la difficulté de mesurer l'adéquation d'un modèle, nous proposons ici d'examiner conjointement les caractéristiques des mesures disponibles et les difficultés pratiques

d'inférence que l'on peut prévoir afin de justifier a priori le choix de modèle que nous avons fait.

La sophistication d'un modèle est fondamentalement liée à sa faculté de prendre en compte une dérive plus ou moins complexe. Dans les paragraphes précédents, nous avons vu que lorsque les mesures analysées ne présentent pas de tendance systématique, il est relativement aisé de réaliser l'inférence en utilisant des fonctions de covariance (mesures de valeurs moyennes constantes dans l'espace) ou des variogrammes (moyennes localement constantes) ; par contre la présence de dérives non constantes (identiques ou distinctes) appelle l'utilisation de fonctions de covariance généralisées, dont l'inférence ne peut se réaliser que de manière algorithmique automatique, donc lourde et abstraite.

Dans le cas de l'étalonnage d'images, nous avons affaire à des mesures d'un même phénomène dont les caractéristiques (densité, degré d'intégration, précision) sont différentes. Il est donc raisonnable de penser que l'existence d'une dérive peut avoir deux origines :

- le phénomène lui-même peut présenter une tendance systématique que les deux dispositifs de mesure perçoivent de la même façon. Compte-tenu de l'expérience maintenant longue acquise dans le domaine de la pluviométrie et de l'insolation (P. DELFINER et J.P. DELHOMME, 1973 ; P. TOURASSE, 1981 ; J.D. CREUTIN et Ch. OBLED, 1982 ; H. DARRICAU-BEUCHER, 1982 ; J.D. CREUTIN, 1983 ; T. LEBEL, 1984) on peut considérer qu'un modèle localement stationnaire est généralement suffisant pour décrire ce type de phénomène.

- la différence de nature des dispositifs de mesure utilisés peut conduire à des perceptions différentes d'une même réalité et cette différence peut revêtir un caractère systématique, une évolution quasi-déterministe dans l'espace qui s'ajoute à l'évolution propre du phénomène. Il faut alors considérer séparément les caractéristiques des dispositifs utilisés et essayer d'apprécier leur effet en terme de dérive :

- i) la précision des mesures réalisées au sol peut être considérée comme à peu près constante dans l'espace, ce qui n'est pas le cas des

Figure IV.2 : Illustration schématique de l'effet du pas d'échantillonnage sur l'appréciation de la dérive : le réseau lâche t_i a une appréciation globalement linéaire de $z(t)$, alors que le réseau dense t_g perçoit sa composante périodique.

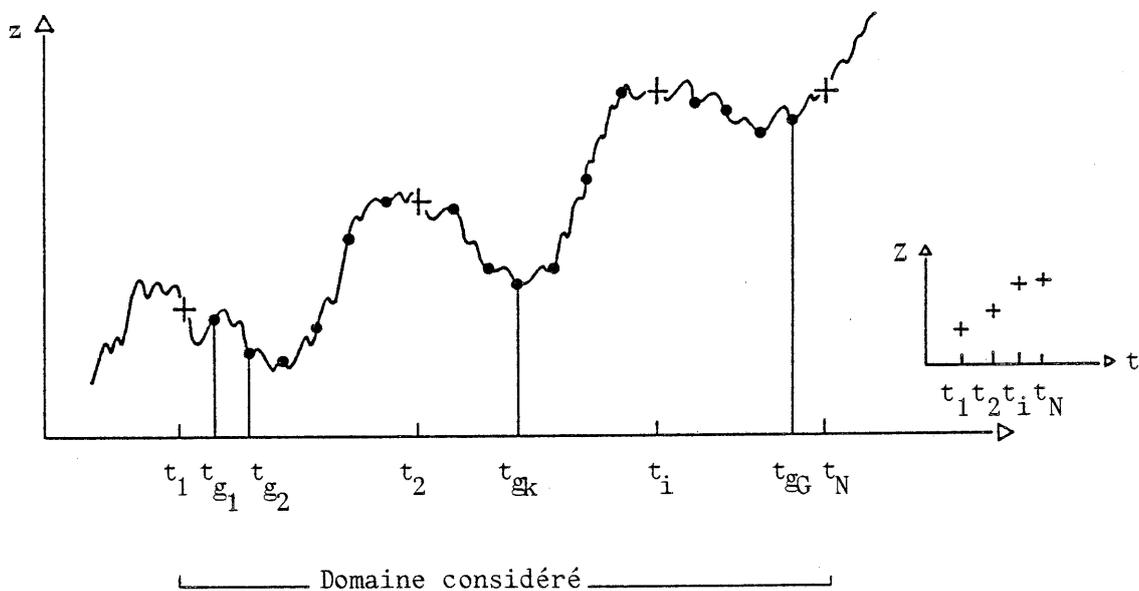
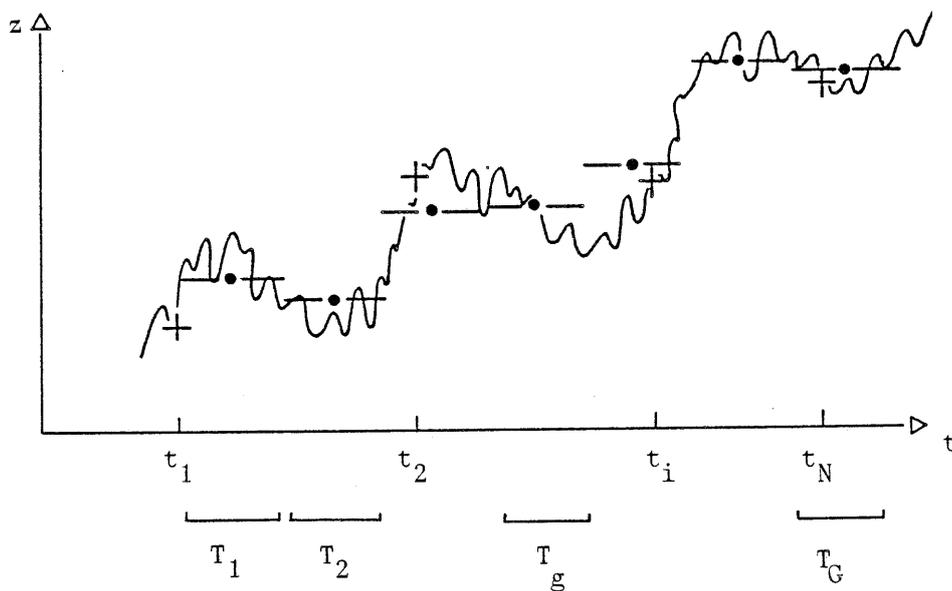


Figure IV.3 : Illustration schématique de l'effet d'intégration du signal par un appareil de mesure sur l'appréciation de la dérive : mesure ponctuelle et mesure intégrées ont une appréciation commune de la dérive.



capteurs de télédétection (dégradation de la mesure radar lorsqu'on s'éloigne du radar, mauvaise appréciation de l'insolation à l'aide du satellite lorsqu'il y a de la neige au sol par exemple) ; il est donc nécessaire d'aller aussi loin que possible dans la correction de ces erreurs systématiques (par l'utilisation de modèles physiques les prenant en considération ou par un pré-étalonnage statistique spécifiquement destiné à compenser ce type de biais systématique ou plus radicalement en évitant l'utilisation des zones concernées) ; puis, il est indispensable de vérifier que leur effet est localement négligeable ou au moins constant afin de le filtrer par une hypothèse de moyenne constante inconnue.

ii) la différence de densité des "réseaux" de mesure utilisés est très importante lorsque l'on considère des mesures sol et des images (une image présente dans la pratique une densité de mesure de 10 à 50 fois supérieure à un réseau sol) ; l'échelle d'observation des deux dispositifs peut donc être sensiblement différente ; or, la notion de dérive est directement liée à l'échelle d'observation, ce qui plaide en faveur d'une modélisation séparée des dérives lorsque la différence de densité de mesure est forte comme l'illustre la figure IV.2.

iii) cette dernière remarque doit être tempérée en rappelant la nature intégrée des mesures contenues dans les images de télédétection ; en effet, l'intégration d'un signal a pour effet de le lisser ce qui équivaut à un changement d'échelle d'observation puisque de nombreux détails disparaissent alors (voir figure IV.3).

Ces considérations sur le comportement des phénomènes étudiés et sur les caractéristiques des dispositifs de mesure utilisés semblent indiquer qu'un modèle quasi-stationnaire (c'est à dire localement stationnaire) peut suffire sous certaines réserves.

De fait, cette conclusion est bienvenue car la densité de mesure des images rendrait impraticable une recherche automatique du degré de dérive et des fonctions de covariance généralisée associées par l'approche classique de validation croisée (voir Chapitre I).

INFERENCE EMPIRIQUE DU MODELE QUASI-STATIONNAIRE RETENU
POUR L'ETALONNAGE EN CONTEXTE SPATIAL

Afin de réaliser l'inférence de ce type de modèle, nous avons suivi la démarche suivante :

i) le calcul de la fonction de covariance expérimentale croisée s'impose dans un premier temps afin de déceler d'éventuels décalages et de préciser le comportement de la structure croisée aux faibles distances, ce que le variogramme ne permet pas. En effet, nous avons vu que, de par sa construction symétrique, le variogramme croisé ne permet pas d'identifier un décalage spatial. D'autre part, le calcul du variogramme croisé ne prend en compte que les points où sont effectués simultanément les deux types de mesure z et z_T , ce qui limite sa "perception" de la structure croisée à l'échelle du réseau le moins dense ; or, le principal intérêt de l'image Z_T est d'apporter en tout point d'estimation une information géographiquement très proche, ce qui se traduit dans le système de cokrigage par l'utilisation de valeurs de γ_{z, Z_T} pour des distances inférieures aux interdistances du réseau le moins dense, c'est à dire pour des distances où son identification expérimentale est impossible ; la fonction de covariance, qui ne présente pas ce défaut apporte donc, au moins qualitativement, une information dont on doit tenir compte (présence d'un effet de pépite sur la structure croisée notamment).

ii) le calcul des variogrammes expérimentaux directs et croisés et leur modélisation, satisfaisant les conditions de positivité examinées précédemment, sont bien entendu nécessaires dans un deuxième temps. Certaines précautions sont à prendre en raison de la nature intégrée des mesures z_T et de la présence fréquente d'une dérive à l'échelle du domaine d'étude. Le degré d'intégration de la variable Z_T (on attribue au point t_g non pas la valeur $Z_T(t_g)$ mais le résultat de l'intégration $1/T_g \int_{T_g} Z_T(t) dt$) c'est à dire la taille du pavé d'intégration T_g affecte de manière plus ou moins importante la valeur expérimentale des variogrammes γ_{z, Z_T} et γ_{Z_T, Z_T} ; nous avons en fait accès à des valeurs intégrées de ces variogrammes :

$$\tilde{\gamma}_{z, Z_T}(t-t') = E(Z(t) - Z(t')) \cdot (1/T \int_T Z_T(t) dt - 1/T' \int_{T'} Z_T(t') dt')$$

qui peut se noter, lorsque les éléments d'intégration T et T' sont de même géométrie :

$$\begin{aligned}\tilde{\gamma}_{Z, Z_T(t-t')} &= 2.1/T \int_T EZ(t).Z_T(t) - 2.1/T \int_T EZ(t').Z_T(t) dt \\ &= 2.1/T \int_T EZ(t).Z_T(t) - EZ(t')Z_T(t) dt \\ &= 1/T \int_T \gamma_{Z, Z_T(t-t')} dt'\end{aligned}$$

et d'autre part : $\tilde{\delta}_{Z_T, Z_T(t-t')} = 1/T \int_T \int_{T'} \delta_{Z_T, Z_T(t-t')} dt dt'$

L'effet de l'intégration qui lisse ces fonctions de structure pourra masquer certaines propriétés de ces variogrammes (notamment de proportionnalité avec $\hat{\delta}_Z$ dans le cas de corrélation intrinsèque) bien utiles pour réaliser une modélisation simple ; néanmoins ces formes intégrées sont à terme nécessaire pour le système de cokrigage qui traite les mesures intégrées de Z_T seules disponibles. La présence d'une dérive à grande échelle n'affecte que très modérément les premiers pas des variogrammes directs et croisés ; pour donner une allure plausible à nos modèles aux grandes distances, nous nous limiterons à rechercher une direction dans laquelle la dérive est peu sensible ; cette précaution, importante pour vérifier la cohérence des modèles retenus, n'a pas d'incidence sur les résultats obtenus par cokrigage sur des voisinages réduits et qui s'appuient donc sur les premiers pas des fonctions de structure.

**PROPOSITION POUR L'INFERENCE AUTOMATIQUE D'UN MODELE
QUASI-STATIONNAIRE POUR L'ETALONNAGE**

Les différentes mises en garde sur la complexité d'une reconnaissance automatique de structure dans le cas du cokrigage peuvent être nuancées.

Il est vrai qu'une généralisation au cas d'une corégionalisation de l'approche classique d'inférence automatique par validation croisée telle que celle décrite au Chapitre I (dans le cadre des modèles intrinsèques généralisés) devient très complexe : le nombre de possibilités de modélisation est très grand et les conditions de positivité sont lourdes. Une telle approche peut néanmoins être envisagée dans certains cas

de figure favorables, notamment lorsque les différentes variables sont mesurées par un réseau unique (cas isotopique, traité par le Centre de Géostatistique pour un problème à une dimension -A. GALLI, communication personnelle, janvier 1986). Par contre, les caractéristiques des images sont très peu favorables à ce type d'approche : la densité très forte de ces mesures comparée à celle des mesures sol conduirait à des systèmes trop lourds et à une gestion complexe des voisinages glissants.

Il est cependant possible d'envisager une approche simplifiée s'appuyant sur certaines hypothèses simplificatrices et un critère de validation mieux adapté aux images que l'erreur de reconstitution ponctuelle du phénomène obtenue par validation croisée.

Le critère retenu repose sur l'écart entre l'estimation réalisée à l'aide des mesures sol et celle réalisée par télédétection :

$$\varepsilon(t_g) = z^*(t_g) - z_T(t_g)$$

Il est possible en chaque point t_g de l'image d'obtenir une valeur expérimentale ε^* de cet écart. De plus, la variance théorique de cet écart peut s'exprimer de la manière suivante (développement déjà proposé au Chapitre II où cette variance est donnée comme valeur théorique de l'erreur e_4) :

$$\begin{aligned} W &= E(Z^*(t_g) - Z_T(t_g) - EZ^*(t_g) + EZ_T(t_g))^2 \\ &= E(\sum_i \lambda^i(t_g)(Z(t_i) - EZ(t_i)))^2 + E(Z_T(t_g) - EZ_T(t_g))^2 \\ &\quad - 2E(\lambda^i(t_g)(Z(t_i) - EZ(t_i)) \cdot (Z_T(t_g) - EZ_T(t_g))) \end{aligned}$$

en supposant que $Z^*(t_g)$ est obtenue linéairement à l'aide des mesures $Z(t_i)$; si les poids $\lambda^i(t_g)$ vérifient le système de krigeage en covariance, alors :

$$\begin{aligned} W &= \sum_i \lambda^i(t_g) C_{ZZ}(t_i - t_g) - \rho_C(t_g) EZ(t_g) + C_{ZT, ZT}(0) \\ &\quad - 2 \sum_i \lambda^i(t_g) C_{Z, ZT}(t_i - t_g) \end{aligned}$$

Avant de bâtir un critère de moindres carrés rapprochant cette variance théorique du carré des écarts expérimentaux, voyons quelles hypothèses peuvent permettre le calcul de W et l'utilisation de ces écarts expérimentaux.

A priori, le calcul de W dépend des modèles de fonctions de covariance directes et croisées ainsi que de la dérive choisie pour Z . Compte-tenu du rôle particulier joué par la modélisation de Z , qui conduit au calcul des λ et μ et comme il est possible de réaliser cette modélisation séparément, nous supposons que celle-ci est déjà effectuée à l'aide d'une méthode classique. La modélisation de Z_T , qui peut être réalisée indépendamment de celle de Z , est également nécessaire pour fixer la covariance ponctuelle $C_{Z_T, Z_T}(0)$ sous hypothèse d'un type de dérive ; mais il faut voir alors que si la dérive de Z_T est différente de celle de Z les écarts expérimentaux ε^* perdent leur sens et ne peuvent plus être utilisés pour estimer W : en effet, une partie de leur valeur sera due à un écart de dérive que nous aurons beaucoup de difficulté à déterminer. Nous ferons donc l'hypothèse de l'identité locale des dérivées de Z et Z_T . Enfin, nous devons respecter une certaine cohérence des modèles choisis pour les différentes fonctions de covariance: une précaution minimale consiste à se placer dans le cadre du modèle linéaire où différentes fonctions de base $C_I(h)$ vont nous servir à exprimer sous certaines contraintes de positivité les fonctions de covariance directes et croisées. Pour simplifier la méthode d'identification, nous nous proposons ici de ne vérifier les conditions de positivité qu'a posteriori.

Pour résumer, et par cohérence avec le choix de modèle quasi-stationnaire effectué au paragraphe précédent, les hypothèses formulées sont les suivantes :

- les fonctions Z et Z_T sont quasi-stationnaires
- les fonctions Z et Z_T ont même moyenne
- une modélisation correcte de Z et Z_T a permis d'identifier les fonctions de covariance ou bien les variogrammes de ces deux fonctions.

La variance W peut alors s'exprimer en remplaçant $C(h)$ par $C(0) - \hat{\gamma}(h)$ et μ_C par $-\mu_K$ (pour respecter les notations du krigeage en variogramme ; cf. Chapitre I) :

$$\begin{aligned} W &= - \sum_i \lambda^i(t_g) \hat{\gamma}_{ZZ}(t_i - t_g) + \hat{\gamma}_{ZZ}(\infty) + \mu_K(t_g) + \hat{\gamma}_{ZT, ZT}(\infty) \\ &\quad - 2 \hat{\gamma}_{Z, ZT}(\infty) + 2 \sum_i \lambda^i(t_g) \hat{\gamma}_{Z, ZT}(t_i - t_g) \\ &= \hat{\gamma}_{ZZ}(\infty) + \hat{\gamma}_{ZT, ZT}(\infty) - 2 \hat{\gamma}_{Z, ZT}(\infty) + \mu_K(t_g) \\ &\quad + \sum_i \lambda^i(t_g) (2 \hat{\gamma}_{Z, ZT}(t_i - t_g) - \hat{\gamma}_{ZZ}(t_i - t_g)) \end{aligned}$$

En raison de l'hypothèse de linéarité, les variogrammes peuvent se noter :

$$\hat{\gamma}_{ZZ}(h) = \sum_{l=1}^L a^l \hat{\gamma}_l(h), \quad \hat{\gamma}_{Z, ZT}(h) = \sum_{l=1}^L a^{lT} \hat{\gamma}_l(h) \quad \text{et} \quad \hat{\gamma}_{ZT, ZT}(h) = \sum_{l=1}^L a^{lTT} \hat{\gamma}_l(h)$$

où les a^{lT} sont les coefficients recherchés.

Le critère de moindres carrés à minimiser s'écrit :

$$\begin{aligned} \text{Crit} &= \sum_{g \in P_g} \frac{1}{g} (\varepsilon^{*2}(t_g) - E \varepsilon^2(t_g)) \\ &= \sum_{g \in P_g} \frac{1}{g} (\varepsilon^{*2}(t_g) - (\sum_{l=1}^L (a^l + a^{lTT}) - 2a^{lT}) \hat{\gamma}_l(\infty) + \mu_K(t_g) \\ &\quad + \sum_i \lambda^i(t_g) \sum_{l=1}^L (2a^{lT} - a^l) \hat{\gamma}_l(t_i - t_g))^2 \end{aligned}$$

En annulant les dérivées de cette expression par rapport aux a^{lT} , nous obtenons le système suivant :

$$\sum_{g \in P_g} \frac{1}{g} (E_1(t_g) + 2 \sum_l a^{lT} E_2(t_g, l)) = 0 \quad \text{pour } l = 1 \text{ à } L$$

en posant :

$$E_1(t_g) = \varepsilon^{*2}(t_g) - \sum_I (a^I + a^I_{TT}) \delta_I(\infty) - \mu_K(t_g) + \sum_i \lambda^i(t_g) \sum_I a^I \delta_I(t_i - t_g)$$

$$E_2(t_g, I) = \delta_I(\infty) + \sum_i \lambda^i(t_g) \delta_I(t_i - t_j)$$

Le choix du pondérateur P_g doit idéalement porter sur la variance théorique W dont la valeur dépend des paramètres recherchés ; pour éviter une optimisation non-linéaire nous préférons attribuer pour valeur approchée à P_g la variance d'estimation de $z^*(t_g)$.

L'avantage de la méthode proposée ci-dessus est sa simplicité d'application lorsque l'on dispose de réseaux de densités différentes (simplicité par rapport à une approche en validation croisée). Ses inconvénients sont sa grande dépendance de la modélisation des variables Z et Z_T supposée acquise au départ et le caractère restrictif des hypothèses faites sur les dérivées et la linéarité des fonctions de structure.

CONCLUSION

Nous avons examiné comment se généralise au cas de plusieurs variables la technique de krigeage présentée au premier Chapitre : sans décrire à nouveau les différents contextes (climatologique ou spatial pur) dans lesquels peuvent s'appliquer les contraintes de non-biais et d'optimalité, nous nous sommes attachés à souligner les aspects nouveaux que revêt l'inférence statistique des moments d'ordre 1 et 2 lorsque plusieurs variables sont analysées simultanément. Etant donné la richesse des modélisations potentiellement utilisables nous avons ensuite précisé notre choix de modèle pour traiter l'étalonnage d'images de télédétection en le justifiant comme un compromis entre la nécessité qu'il prenne en compte les propriétés du phénomène mesuré et des dispositifs de mesure utilisés et la possibilité pratique de le mettre en oeuvre.

Dans le chapitre suivant, nous décrivons deux cas d'application de la technique de cokrigeage.

CHAPITRE V

EXEMPLES D'APPLICATION DU COKRIGEAGE :
MESURE DE LA PLUVIOMETRIE EN ARABIE ET DE LA FRACTION
JOURNALIERE D'INSOLATION EN FRANCE

Résumé

Ce chapitre regroupe deux exemples d'application du cokrigeage à des mesures conjointes par dispositif de télédétection et par réseau d'appareils au sol.

Le premier exemple concerne la pluviométrie dans le Nord de l'Arabie Séoudite, déduite par néphanalyse et régression multiple d'un ensemble d'images de satellites météorologiques, et mesurée au sol par un réseau de pluviomètres. Après une analyse détaillée de la structure spatiale directe et croisée de ces mesures, nous présentons les résultats obtenus sous forme de cartes isohyètes.

Le deuxième exemple concerne la mesure de la fraction journalière d'insolation fournie, à partir des images Météosat, par la méthode CACTUS et par le réseau d'héliographes de la Météorologie Nationale. En nous appuyant sur l'analyse climatologique de la structure directe de ces mesures présentée au chapitre III et en complétant cette analyse par celle de la structure croisée, nous validerons les résultats obtenus par cokrigeage suivant le protocole défini au chapitre III.

1. MESURE DE LA PLUVIOMETRIE EN ARABIEa) - Généralités

Comme le laisse apparaître la description des données disponibles pour cette étude (se reporter à l'Introduction), cet exemple offre assez peu de possibilités de validation quantitative. En effet, le nombre d'événements pluvieux est très réduit (pour des raisons évidentes de lourdeur de la méthode d'interprétation des images satellites) et le réseau de pluviomètres est peu fourni (43 stations) mais surtout très mal

distribué géographiquement (la moitié des stations se regroupent sur le cinquième de la zone d'étude).

Néanmoins, l'assez bonne co-fluctuation des mesures sol et télédétection permet d'illustrer les différentes étapes d'une analyse de structure en présence de deux variables ; cette analyse est menée dans un contexte spatial pur c'est à dire en considérant chaque événement pluvieux individuellement.

Nous nous limiterons à une validation qualitative des résultats obtenus au cours de différents essais (mesures sol et satellite considérées individuellement ou combinées, utilisation de sous-réseaux) en comparant les cartes d'isohyètes.

b) - Analyse de structure

L'analyse de la structure spatiale de ces mesures est rendue délicate par au moins deux éléments :

- d'un point de vue météorologique, la zone étudiée paraît être soumise à deux régimes pluviométriques assez différents (si tant est que les 5 épisodes pluvieux disponibles soient représentatifs de la climatologie locale) : la moitié sud-est recueille l'essentiel des précipitations alors que la moitié nord-ouest n'est que très faiblement arrosée ; nous sommes donc apparemment en présence d'une assez importante non-stationnarité qu'il est difficile de traiter en raison de la faible densité du réseau de mesure (une analyse par zone conduit à travailler avec moins de 20 stations, ce qui n'est pas réaliste).

- les mesures comparées (sol et satellite) résultent de procédés d'intégration très différents : les mesures sol, très ponctuelles, sont intégrées sur des périodes allant de la semaine au mois, alors que les mesures satellite sont des cumuls de mesures instantanées à pas de temps lâches et irréguliers (1 à 3 photos par jour essentiellement l'après-midi) fortement intégrés dans l'espace (mailles de 1° de côté, soit environ 10000 km² ; il est dans ces conditions très difficile d'apprécier l'effet de ces

intégrations très disparates, étant entendu que nous ne connaissons pas la structure "élémentaire" du phénomène.

Les réponses apportées à ces difficultés sont rudimentaires et recourent à des simplifications peut être discutables. Nous examinerons tout d'abord quelques aspects de la structure individuelle de chaque type de mesure, puis nous analyserons leur structure croisée.

Structure directe

Le tableau V.1 regroupe les résultats des différents essais d'analyse de la structure directe.

Pour ce qui concerne les mesures sol nous avons tout d'abord calculé les variogrammes expérimentaux des valeurs brutes fournies par les 43 pluviomètres. Bien que les fluctuations de ces fonctions expérimentales soient souvent importantes (en raison du faible nombre de points de mesure) et malgré l'effet sensible de la non-stationnarité à grande échelle (soulignée plus haut) qui perturbe ces fonctions au-delà de 500 km le modèle sphérique semble convenir pour chacune des 5 réalisations. Puis, nous avons calculé les variogrammes expérimentaux des valeurs intégrées par krigeage sur des cibles correspondant à des mailles satellite afin d'apprécier l'effet d'une telle intégration ; en raison de la distribution géographique des 43 pluviomètres, nous n'avons pu prendre en compte que 29 cibles puisque certaines contiennent plusieurs pluviomètres. Il est bien sûr nécessaire de faire certaines réserves sur la comparabilité de variogrammes calculés avec des nombres de points de mesure faibles et différents ; néanmoins, les variogrammes ainsi obtenus présentent des modifications allant bien dans le sens attendu d'une réduction de la valeur du palier et d'un allongement de la portée (à l'exception de l'événement de Mars 79 où la valeur du palier, très sensible à une valeur extrême de 25 mm soit presque 5 écarts-types, ne suit pas la règle indiquée). Enfin, nous avons essayé de conforter cette analyse empirique en utilisant la procédure de reconnaissance automatique du programme Bluepack. Il est satisfaisant de constater d'une part que le degré de dérive proposé est en général nul, ce qui justifie l'hypothèse quasi-stationnaire (ce programme a choisi dans notre cas des voisinages de rayon moyen égal à 370 km) ; l'exception du 15-

Figure V.1 : Exemples de variogrammes obtenus à l'aide de mesures pluviométriques au sol : mesures brutes des 43 pluviomètres puis mesures intégrées par krigeage sur 29 cibles.

- o mesures brutes
- mesures intégrées
- covariance généralisée identifiée automatiquement

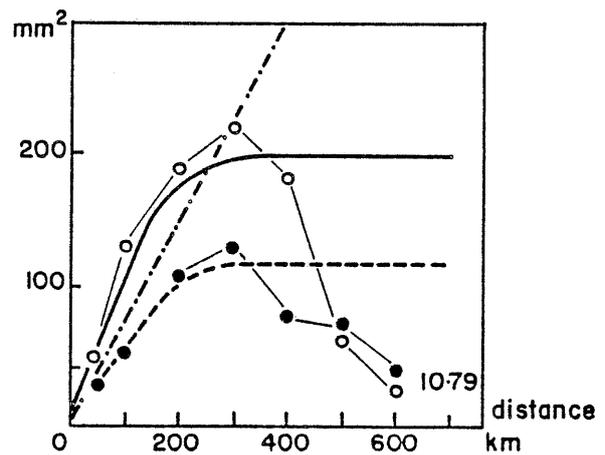
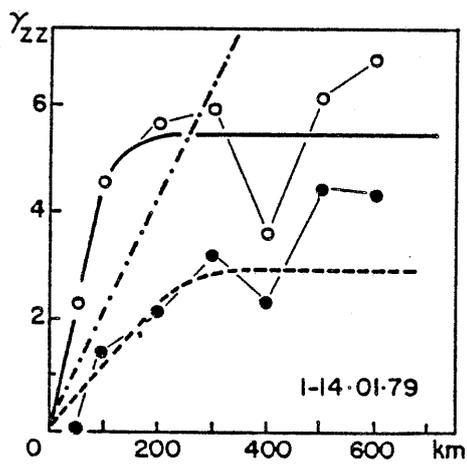
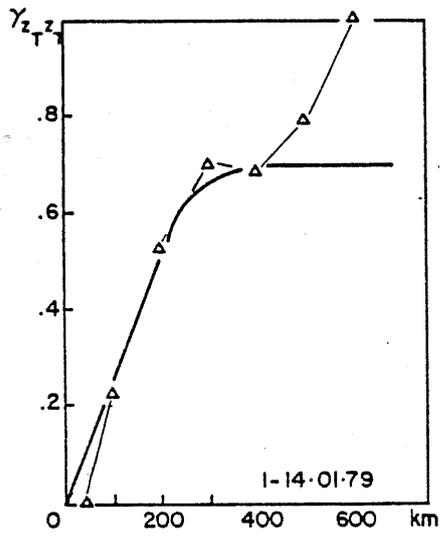
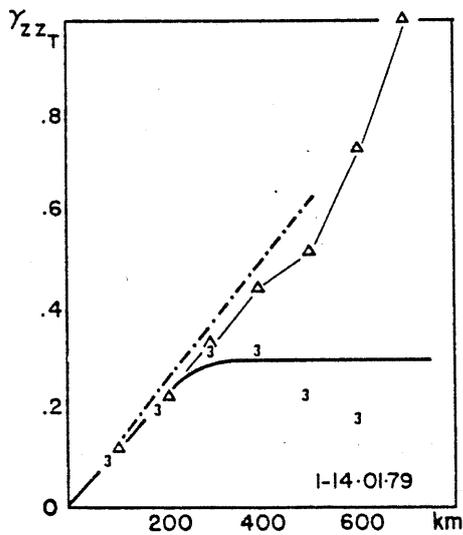
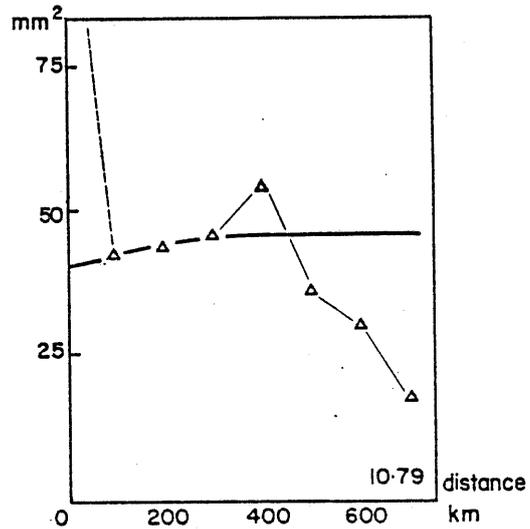


Figure V.2 : Exemples de variogrammes obtenus à l'aide des mesures satellites de la pluviométrie sur les seules mailles contenant des stations sol (a) puis sur l'ensemble des mailles (b).

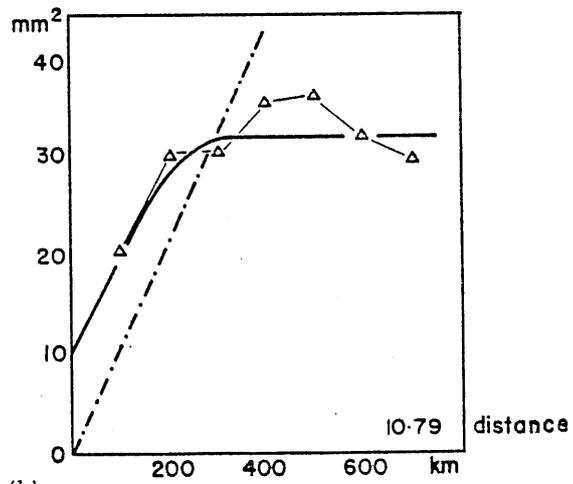
- Δ Variogramme global
- 3 Variogramme dans la direction NOE
- Modèle ajusté empiriquement
- - - Covariance généralisée identifiée automatiquement



(a)



(b)



26/01/79 où le degré est 1 correspond en fait à des performances presque identiques des degrés 0 et 1. D'autre part, les modèles linéaires calés par le modèle sont parfaitement compatibles avec nos modèles empiriques aux distances faibles. La figure V.1 illustre pour deux événements choisis, ces différentes étapes de l'analyse des mesures sol.

Pour ce qui concerne les mesures satellites, les variogrammes expérimentaux obtenus ont à nouveau été modélisés à l'aide de fonctions sphériques (voir tableau V.1) dont les caractéristiques diffèrent de celles ajustées pour les mesures sol : ces différences portent surtout sur la valeur du palier ce qui s'explique par la différence entre les moyennes et les écarts-types spatiaux des deux ensembles de mesures ; elles portent également sur la portée ce qui est dû en grande partie à l'effet d'intégration qui tend à allonger la portée (le tableau V.1 montre que les portées obtenues après intégration des mesures sol sont beaucoup plus proches des portées des variogrammes des mesures satellite).

Lors de ces comparaisons, il convient toutefois de rester prudent en raison de la forte sensibilité d'échantillonnage de ces variogrammes à la densité et à la répartition géographique du réseau de mesure ; à titre d'illustration, la figure V.2 compare les variogrammes des mesures satellites lorsque l'on utilise 29 mailles (celles contenant des stations sol) ou bien l'ensemble des 79 mailles disponibles : pour le 1-14/01/79, la portée qui paraissait assez naturelle avec 29 points doit être recherchée dans la direction NS avec l'ensemble des points et pour le 10/79 le variogramme apparemment pépétique pur obtenu avec 29 points prend une allure nettement sphérique pour l'ensemble des points de la grille.

De manière générale l'existence d'une différence de structure entre ces deux types de mesure n'est pas surprenante lorsque l'on considère la différence radicale de nature de ces deux dispositifs de mesure. Il est au contraire très satisfaisant de noter le bon rapprochement qualitatif des fonctions de structure par simple intégration des mesures sol.

Tableau V.1 : Résultats principaux de l'analyse de la structure spatiale des mesures sol et satellite.

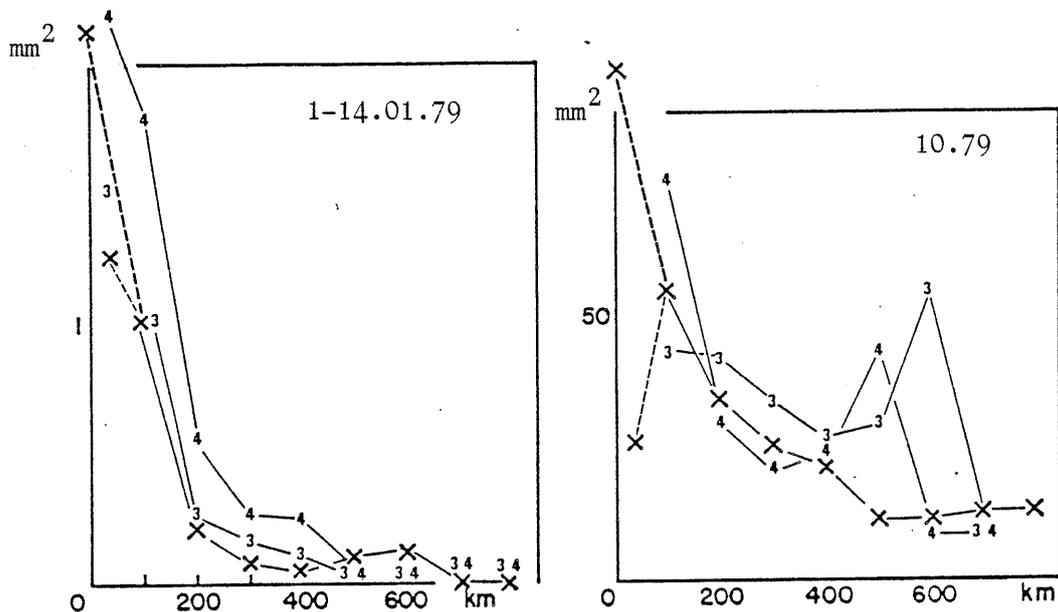
DATE DE L'ÉVÉNEMENT	MESURES SOL					MESURES SATELLITES					
	Moyenne Ecart-type Variation	Variogrammes empiriques des valeurs		Reconnaissance automatique en 43 points		Moyenne Ecart-type Variation		Variogrammes empiriques des valeurs brutes		Reconnaissance automatique sur 79 mailles	
		brutes en 43 points	intégrées sur 29 mailles	dérive	covariance	29 mailles	79 mailles	29 mailles	79 mailles	dérive	covariance
1-14 JANV. 79	1.2 2.3 1.9	S 5.5 200 0	S 3 350 0	0	CGL $2.16 \cdot 10^{-2}$ 0	.4 .8 2.	0.3 0.7 2.3	S .7 400 0	S 0.3 350 0	1	L $0.12 \cdot 10^{-2}$ 0
15-26 JANV. 79	14.4 18.2 1.3	S 360 150 0	S 200 (160) 350 (300) 50 (0)	1	P 343	35 20 0.6	29.6 22.5 0.8	S 600(500) 600(350) 0 (0)	S 450 550 0	1	CGP 0 -1.1 10^{-4}
27-31 JANV. 79	1.9 35. 1.8	S 15 150 10	S 3 350 0	0	CGL $6.8 \cdot 10^{-2}$ 0	6. 7. 1.2	3.8 5.8 1.5	S 70 350 0	S 40 400 0	1	CGP 0 -0.13 $0.9 \cdot 10^{-5}$
MARS 79	1.4 4.9 3.5	S 25 150 0	S 40 300 0	0	CGL $8.2 \cdot 10^{-2}$ 0	4. 6. 1.5	2.6 4.8 1.8	S 40 300 0	S 25 300 0	1	CGP 0 -0.048 $1.3 \cdot 10^{-4}$
OCTOBRE 79	9.1 12.1 1.3	S 200 300 0	S 120 300 0	0	CGL 0.77 0	6.3 6.3 1.0	.5 5.6 11.2	P 45	S 32 350 10	0	CGL 0.11 0

Unités : Pluviométrie en mm
Distance en km

Modèles de fonction de structure : S : variogramme sphérique (palier, portée, pépité)
P : variogramme pépitique pur (pépité)
CGL: covariance généralisée linéaire (pente, pépité)
CGP: covariance généralisée polynomiale (pépité, coefficients A1 et A3)

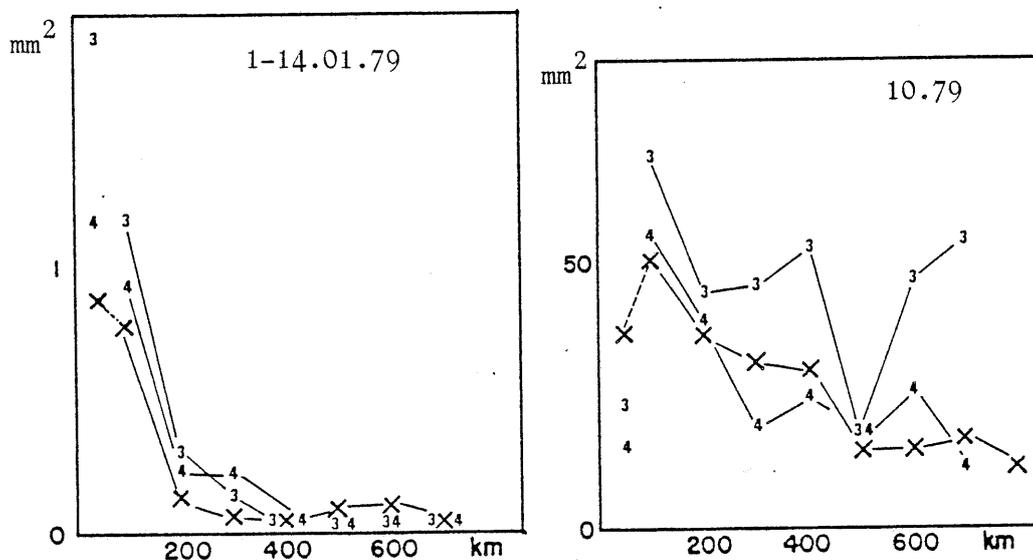
Figure V.3 : Exemples de covariance simple croisée obtenue à l'aide des mesures pluviométriques sol et satellite en se limitant aux points de mesure communs : covariance globale sur le champ et covariances suivant l'azimut N45E en distinguant l'ordre d'intervention des points.

- x covariance globale
- 3 covariance suivant N45E pour les couples tels que la mesure sol se trouve à l'Ouest de la mesure satellite
- 4 covariance suivant N45E pour les autres couples



a - avant translation de l'image satellite

b - après translation de l'image satellite



Structure croisée

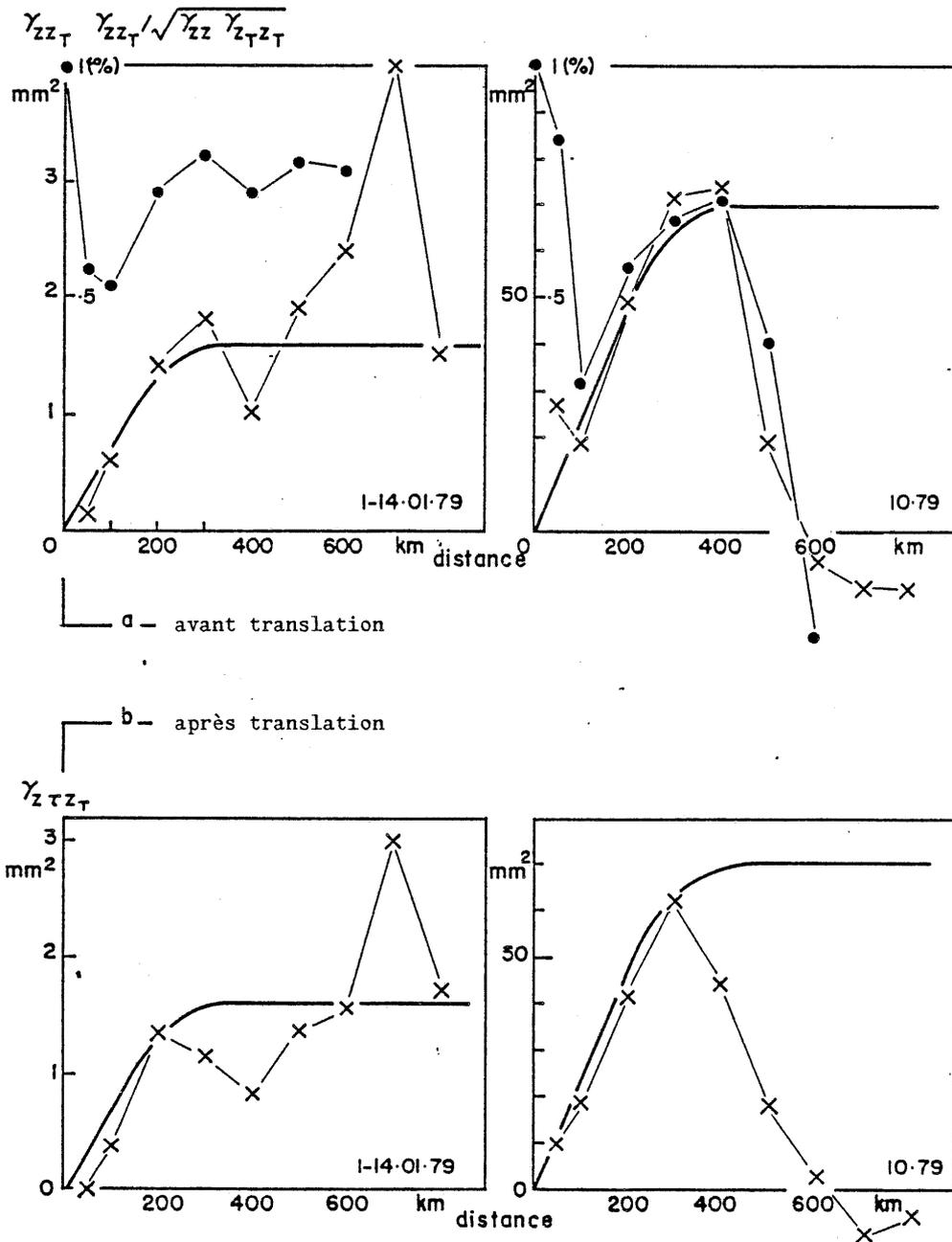
La première étape de cette analyse est l'examen de la symétrie de la fonction de covariance. En effet, nous avons vu au Chapitre IV que seule la fonction de covariance était susceptible par construction de révéler la présence de décalages systématiques d'un ensemble de mesures par rapport à l'autre. Nous avons donc recherché une direction pour laquelle le sens dans lequel sont prises les mesures a le plus d'effet sur la valeur de la covariance ; par sens de prise de mesures nous entendons la distinction, pour une orientation donnée, entre les couples de mesures pour lesquels le point de mesure satellite est plus proche de l'origine que le point de mesure sol et les couples pour lesquels l'ordre de proximité est inverse (toutes les coordonnées des points de mesure étant positives). L'examen des fonctions de covariance par sens suivant 4 directions (NOE, N45E, N90E et N135E) montre que la sensibilité au sens se fait sentir plutôt pour les directions NOE (le 27-31/01/79) et N45E (les 1-14/01/79, 03/79 et 10/79), sans toutefois être très sensible comme le montre la figure V.3-a. La différence de comportement se fait surtout sentir aux courtes distances (< 200 km). Un essai de translation de l'image satellite afin de corriger le décalage diagnostiqué par la fonction de covariance a été réalisé. L'examen des cartes sol et satellite (figure V.5) montre qu'une telle translation (environ 100 km vers le NE) améliore plutôt la cohérence spatiale des maxima de précipitation. D'autre part, la construction des fonctions de covariance par direction et sens après translation (voir figure V.3-b) montre un léger rapprochement des courbes expérimentales aux courtes distances. Cependant, cette démarche appelle plusieurs remarques :

- il est difficile de justifier une telle translation d'un point de vue physique ; en effet, s'il est légitime de craindre des erreurs de positionnement des images satellite, il est surprenant qu'un cumul de ces images n'annule pas des erreurs successives qui ont toutes chances d'être indépendantes ; en deuxième lieu, même si ces erreurs sont liées, il est surprenant de voir que leur "intégrale" change d'une période à l'autre.

- la construction des fonctions de covariance sans distinction de sens et suivant la direction de la translation montre (voir figure V.3) que cette translation a en fait assez peu d'effet.

Figure V.4 : Exemples de variogrammes croisés avant et après translation de l'image satellite.

- x Variogramme croisé global
- Variogramme croisé normé
- Modèle ajusté empiriquement



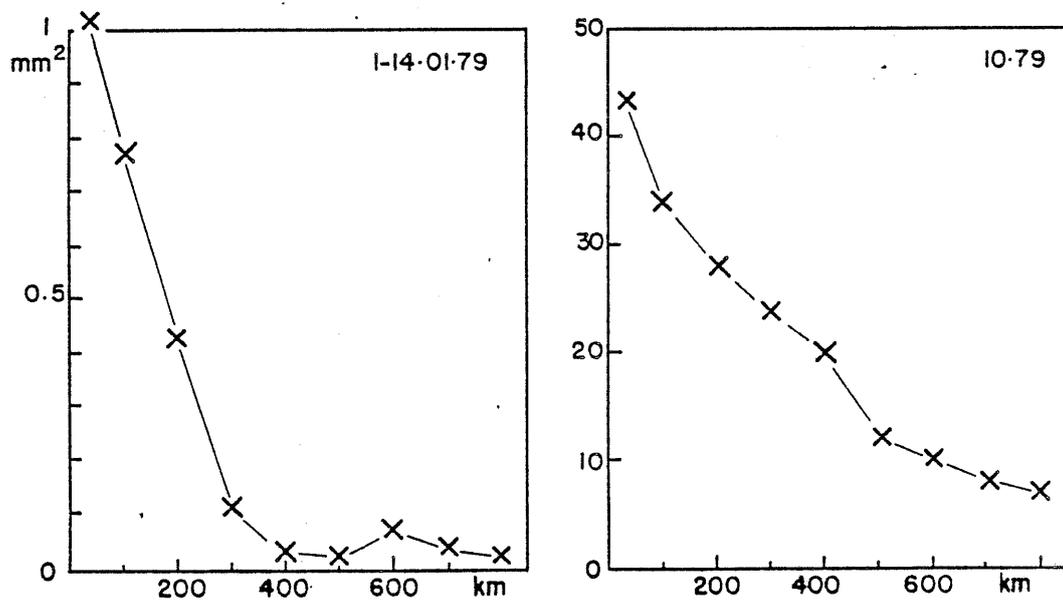
- enfin, l'effet de cette translation sur les variogrammes croisés est lui aussi mineur (voir figure V.4).

En fait, il nous semble que dans le cas présent les fonctions de covariance simple par sens et azimut sont des outils assez fragiles : leur sensibilité à l'évolution spatiale moyenne du champ est très forte, y compris aux faibles distances et leurs fluctuations d'échantillonnage sont sans doute considérables. De plus, la présence de décalage, physiquement improbable, n'est pas manifeste à l'examen des cartes si l'on considère l'ensemble du champ. Enfin, d'un point de vue pratique, les translations envisagées sont d'amplitude modeste et ont peu d'effet sur les fonctions de structure globales. Nous avons donc décidé de considérer par la suite ces décalages comme inexistant, ce qui nous permet d'utiliser des variogrammes dont le comportement, au moins aux courtes distances, est plus sain.

Le tableau V.2 résume les principaux résultats de l'analyse de la structure croisée. Ils appellent les trois remarques suivantes :

i) Les variogrammes croisés obtenus se comportent d'une manière assez proche d'un modèle sphérique pour les distances inférieures à 500 km; nous devons, cependant noter que le faible nombre de points de mesure au sol rend le comportement de ces variogrammes expérimentaux assez instables ; aux très courtes distances cette instabilité est due à l'incapacité du variogramme croisé à prendre en considération l'ensemble des mesures disponibles (puisqu'il ne considère par construction que les points où les deux types de mesure existent - dans notre cas les points de mesure sol) ; afin de mieux apprécier la structure aux courtes distances, il est possible de construire la covariance croisée à l'aide de toutes les mesures sol et satellite ; la figure V.5 montre que la valeur de la covariance au premier pas de calcul (ici 34 km) se trouve bien dans la continuité des valeurs aux plus grandes distances : les valeurs obtenues figures V.3 et V.4 pour ce premier pas de calcul, sans signification par manque de robustesse, pouvaient parfois laisser penser à une discontinuité (voir le 10/79), mais notons que cette remarque sur la robustesse du premier point est vraie pour l'ensemble des champs) ; remarquons, pour finir, que les fonctions de covariance calculées pour l'ensemble des points ont des valeurs

Figure V.5 : Exemples de fonctions de covariance simple croisée obtenues à l'aide de l'ensemble des mesures sol et satellite.



sensiblement plus faibles que celles de la figure V.3 : ceci est dû à la différence de variabilité des deux types de mesure.

ii) La vérification des conditions de positivité : $(\gamma^2_{Z,Z_T}(h) < \gamma_{ZZ}(h) \cdot \gamma_{Z_T,Z_T}(h))$ est rendue délicate i) par la nature intégrée des mesures satellite Z_T et ii) par l'incertitude d'ajustement des modèles théoriques. Pour simplifier cette vérification, nous avons fait une hypothèse de corrélation intrinsèque (voir Chapitre IV) justifiée a priori par le fait que Z et Z_T sont des mesures d'un même phénomène ; cette hypothèse consiste à écrire :

$$\gamma_{ZZ}(t-t') = a \cdot \gamma(t-t') ; \gamma_{Z,Z_T}(t-t') = a_T \cdot \gamma(t-t') \text{ et } \gamma_{Z_T,Z_T}(t-t') = a_{TT} \gamma(t-t')$$

Ce qui ramène la condition de positivité à $a^2_T < a \cdot a_{TT}$. Or, nous n'avons pas accès expérimentalement à $\gamma_{Z,Z_T}(t-t')$ mais à :

$$\hat{\gamma}_{Z,Z_T}(T-T') = \int_{T'} \gamma_{Z,Z_T}(t-t') dt'$$

de même que nous disposons d'une forme intégrée de γ_{Z_T,Z_T} :

$$\hat{\gamma}_{Z_T,Z_T}(T,T') = \int_T \int_{T'} \gamma_{Z_T,Z_T}(t-t') dt dt'$$

Par contre, nous pouvons approcher expérimentalement $\hat{\gamma}_{ZZ}(T-T')$ et $\hat{\gamma}_{Z,Z_T}(T-T')$ qui peuvent se noter sous l'hypothèse faite :

$$\hat{\gamma}_{ZZ}(T-T') = a \cdot \hat{\gamma}(T-T') \quad \text{et} \quad \hat{\gamma}_{Z,Z_T}(T-T') = a_T \cdot \hat{\gamma}(T-T')$$

$$\text{avec : } \hat{\gamma}(T-T') = \int_T \int_{T'} \gamma(t-t') dt dt'$$

Alors si :

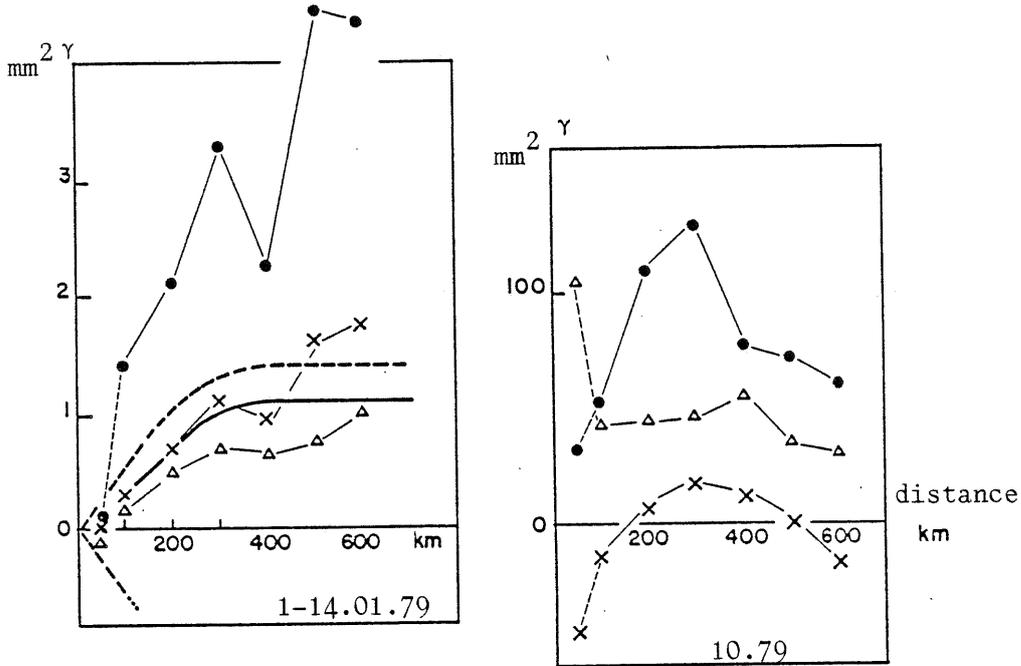
$$\hat{\gamma}^2_{Z,Z_T}(T-T') < \hat{\gamma}_{ZZ}(T-T') \cdot \hat{\gamma}_{Z_T,Z_T}(T-T') \quad (V.1)$$

nous aurons vérifié la condition de positivité $a^2_T < a \cdot a_{TT}$.

Tableau V.2 : Résultats principaux de l'analyse de la structure spatiale croisée des mesures sol et satellite.

DATES DE L'EVENEMENT	Statistiques ponctuelles		Variogrammes croisés ajustés empiriquement		covariance croisée centrée identifiée automatiquement	Enveloppe "théorique" du variogramme croisé
	Covariance croisée simple	Corrélation croisée	Mesures sol ponctuelles	Mesures sol intégrées		
1-14 JANV. 79	2.2 1.5	.74	S 1.6 350 0	S 1.3 400 0	0.21	S 1.4 350-400 0
15-26 JANV. 79	691.0 205.7	.59	S 300 500 0	S 300 (200) 500 (350) 0 (0)	203.	S 350. (280) 350-600 (300-350) 0 (0)
27-31 JANV. 79	22.3 10.5	.46	S 16. 400 6	S 13. 450 0	9.4	S 14. 350 0
MARS 79	21.9 15.6	.51	S 20. 250 0	S 25. 300 0	12.9	S 40 300 0
OCTOBRE 79	98.9 41.73	.53	S 70 450 0	P 0	52.7	P 0

Figure V.6 : Exemples de variogrammes directs et croisés obtenus à l'aide de mesures pluviométriques sol intégrées et satellite sur 29 mailles.



Variogramme expérimental
 direct ● des valeurs sol
 et ▲ des valeurs satellite
 puis croisé x

— modèle ajusté empiriquement au variogramme croisé
 --- enveloppe "théorique" du variogramme croisé

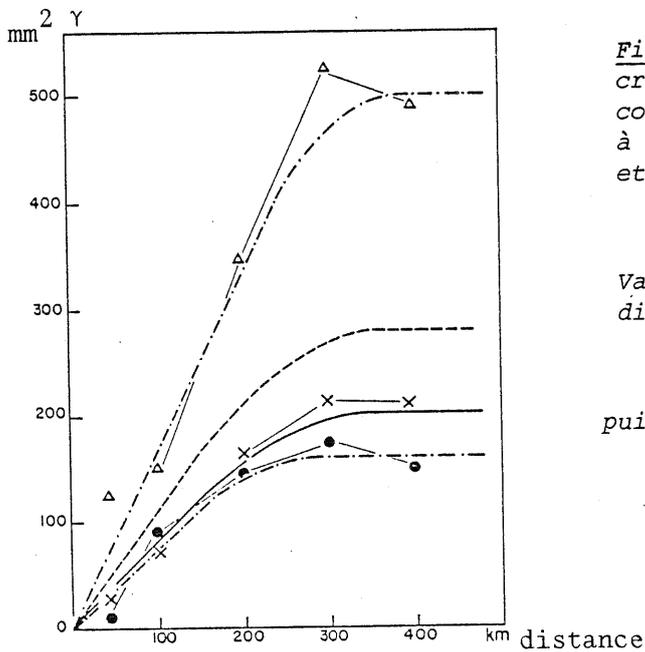


Figure V.7 : Variogrammes directs et croisés pour l'événement 10.79 en considérant la seule direction NS et à l'aide des mesures sol intégrées et satellite sur 29 mailles.

Variogramme expérimental dans la direction N.S.

direct ● des valeurs sol
 et ▲ des valeurs satellite
 puis croisé x

— modèle ajusté empiriquement au variogramme croisé
 --- enveloppe théorique du variogramme croisé

Pour effectuer cette vérification, nous avons porté dans la deuxième colonne du tableau V.2, l'enveloppe dans laquelle $\gamma_{Z,Z_T}(T-T')$ doit être contenue d'après la formule V.1 ; pour cela, nous avons considéré les variogrammes des valeurs sol intégrées par krigeage sur 29 mailles qui contiennent toutes une ou plusieurs stations sol ; les variogrammes des valeurs satellites ont également été calculés sur ces mêmes mailles pour des raisons de comparabilité ; les variogrammes croisés sol intégré-satellite portent également sur ces mailles. Soulignons immédiatement que le faible nombre de points de mesure soumet ces variogrammes à des fortes fluctuations d'échantillonnage dont nous devons tenir compte notamment à l'examen de variogrammes directionnels.

Trois événements (1-4/01/79, 27-31/01/79 et 03/79) ne posent pas de problème pour la détermination de cette enveloppe puisque les variogrammes directs sont sphériques et sensiblement de même portée ; par ailleurs, les variogrammes croisés sont eux aussi de mêmes caractéristiques et contenus dans l'enveloppe (voir par exemple le cas du 1-14/01/79 présenté figure V.6).

L'existence d'une dérive assez sensible lors de l'événement du 15-21/10/79 (identifiée automatiquement sur les données sol) explique sans doute que les portées des variogrammes directs et croisés s'échelonnent entre 350 et 600 km (fluctuation que ne suffit pas à expliquer l'effet d'échantillonnage) ; en effet, les variogrammes limités à la direction NS moins affectée par la dérive se révèlent beaucoup plus nettement homothétiques et satisfont la condition de positivité (voir figure V.7 et les modèles donnés entre parenthèses dans les tableaux V.1 et V.2).

Le cas de l'événement du 10/79 est plus délicat ; le variogramme des mesures satellite calculé à l'aide de 29 mailles est nettement pépitique (alors que celui construit à l'aide de l'ensemble des 79 mailles est sphérique) ; celui des mesures sol est sphérique ; l'hypothèse de corrélation intrinsèque n'est donc pas satisfaite par les mesures intégrées ; il est alors logique que le variogramme croisé soit nul puisqu'il doit constituer le "dénominateur commun" de ces deux modèles directs radicalement distincts (voir paragraphe sur le modèle linéaire au Chapitre IV).

iii) Enfin, le recours à une méthode automatique d'identification de la structure croisée n'apporte rien par rapport au calcul tout simple de la covariance croisée centrée ponctuelle ; cela est dû en grande partie au caractère simpliste et à certains égards approximatif de la recherche automatique entreprise (hypothèse de corrélation intrinsèque dans laquelle on s'est borné à identifier le palier du variogramme croisé en supposant les portées des variogrammes directs identiques).

En résumé, compte-tenu de notre but qui est la reconstitution ponctuelle par cokrigage de la pluie au sol en utilisant les mesures sol ponctuelles et les estimations satellite, nous avons retenu les modèles de variogrammes suivant :

- pour les mesures sol, le variogramme ajusté empiriquement aux mesures brutes sur l'ensemble des points disponibles ;
- pour les mesures satellite, le variogramme des mesures sur l'ensemble des mailles disponibles ;
- le variogramme croisé entre les mesures sol ponctuelles et les mesures satellite correspondantes.

Ce choix s'explique par la nécessité de remplir le système de cokrigage à l'aide des fonctions de structure appropriées et par le souci que ces fonctions soient aussi robustes que possible. Par ailleurs, nous pouvons supposer que ces fonctions satisfont les conditions de positivité dans la mesure où i) leurs formes intégrées vérifient ces conditions et ii) une hypothèse de corrélation intrinsèque est plausible (à l'exception de l'événement 10/79).

c) - Résultats

Comme nous l'avons dit en introduction le nombre réduit de points de mesure au sol et d'événements disponibles ne nous a pas permis d'entreprendre une validation quantitative de cette application du cokrigage.

La seule possibilité d'apprécier l'apport de l'information satellite à l'interpolation des données sol est de comparer qualitativement les cartes d'isohyètes obtenues par cokrigeage avec celles que fournit le krigeage des seules mesures sol. Il est évident, a priori, que cet apport est d'autant plus fort que la densité du réseau sol est faible, c'est pourquoi nous présentons en plus des résultats obtenus avec le réseau sol complet (43 pluviomètres) les résultats que donnent un réseau de 13 puis de 8 stations (ces sous-réseaux ont été choisis de manière purement aléatoire pour assurer une certaine équirépartition et non pas, comme cela pourrait se faire, en fonction des valeurs mesurées afin d'obtenir des résultats probants.).

Le système de cokrigeage utilisé suppose une dérive constante des mesures sol et satellite sur le voisinage utilisé. Les contraintes sur les poids λ^i des valeurs sol et λ^g des valeurs satellites sont telles que l'espérance de la valeur estimée soit égale à l'espérance des valeurs sol ($\sum \lambda^i = 1$ et $\sum \lambda^g = 0$) en raison du biais très important que présentent les mesures satellites (voir tableau V.1). D'autre part les voisinages utilisés comportent toutes les stations sols et les mailles satellite dont le centre est situé à moins de 200 km du point reconstitué et au moins 8 stations sol (éventuellement éloignées de plus de 200 km) ; ce choix empirique essaie de réaliser un compromis entre la qualité des reconstitutions et les temps de calcul nécessités compte-tenu de la géométrie des réseaux et de la structure du phénomène. Enfin, quelle que soit la configuration de réseau (43, 13 ou 8 stations), les fonctions de structure utilisées sont celles obtenues au paragraphe précédent à l'aide de l'ensemble des 43 stations ; cette façon de faire, imposée par une nécessité évidente de robustesse, ne rend pas forcément irréalistes les essais en réseau réduit; il est en effet possible d'imaginer dans un contexte opérationnel une première étape d'apprentissage assez longue pour pouvoir identifier avec un réseau dense des fonctions de structure climatologiques (voir exemple dans la deuxième partie de ce chapitre), puis une étape opérationnelle où seul un réseau réduit permet de corriger les estimations satellite par cokrigeage.

Les figures V.8 et V.9 présentent les résultats obtenus pour quatre des cinq événements (1-4/01/79, 15-26/01/79, 17-31/01/79 et 10/79).

L'événement du 03/79 n'a été exclu que pour des raisons pratiques de mise en page ; il présente en fait des caractéristiques générales très proches du 1-14/01/79.

La figure V.8 présente d'une part l'ensemble des données de base: mesures sol en 43 stations interpolées par krigeage et grille d'estimations satellite. Bien sûr, comme le laissait entrevoir l'examen des moyennes et écarts-types spatiaux, les estimations satellite diffèrent assez sensiblement de celles du sol ; néanmoins la répartition spatiale du phénomène est appréciée avec une cohérence satisfaisante par les deux dispositifs. Notons que nous accordons dans cette comparaison et les suivantes un crédit plus important aux mesures sol, en raison de leur nature directe ; il faut cependant rappeler que les isohyètes obtenues à l'aide de ces mesures sont très sensibles à la répartition géographique des points de mesure ; ainsi, le maxima indiqué par le réseau sol pour le 27-31/01/79 n'exclut pas vraiment celui perçu par le satellite à 100 km vers le NE puisque le réseau sol présente à cet endroit une lacune importante.

La figure V.8 présente également les cartes obtenues par cokrigeage à l'aide des 43 stations sol : ces cartes sont bien entendu très proches des résultats de krigeage puisque le système de cokrigeage est conditionné de manière à reconstituer des valeurs sol (vecteur de droite de système rempli à l'aide de $\gamma_{ZZ}(t-t')$ et $\gamma_{Z,ZT}(T-t')$); des écarts n'apparaissent que dans les zones de faible densité de mesures sol (voir les modifications de l'isohyète 0 mm pour les 15-26/01/79 et 27-31/01/79 ou bien, pour cette dernière période, la légère extension vers le NE du maxima évoqué ci-dessus).

La figure V.9 regroupe les résultats du krigeage et du cokrigeage pour ces mêmes événements à l'aide de sous-réseaux de 13 et 8 stations (soient de densités respectives de 1/30000 et 1/50000 station par km²).

Les résultats de krigeage se dégradent très brutalement lorsque l'on passe de 13 à 8 stations au sens où les cartes obtenues avec la plus faible densité perdent totalement la perception de la distribution spatiale de la pluie qu'avait encore approximativement le réseau de 13 pluviomètres.

Figure V.8 : Exemples de cartes isohyètes obtenues (a) par krigeage des seules données sol, (b) par utilisation des images satellite et (c) par cokrigeage.

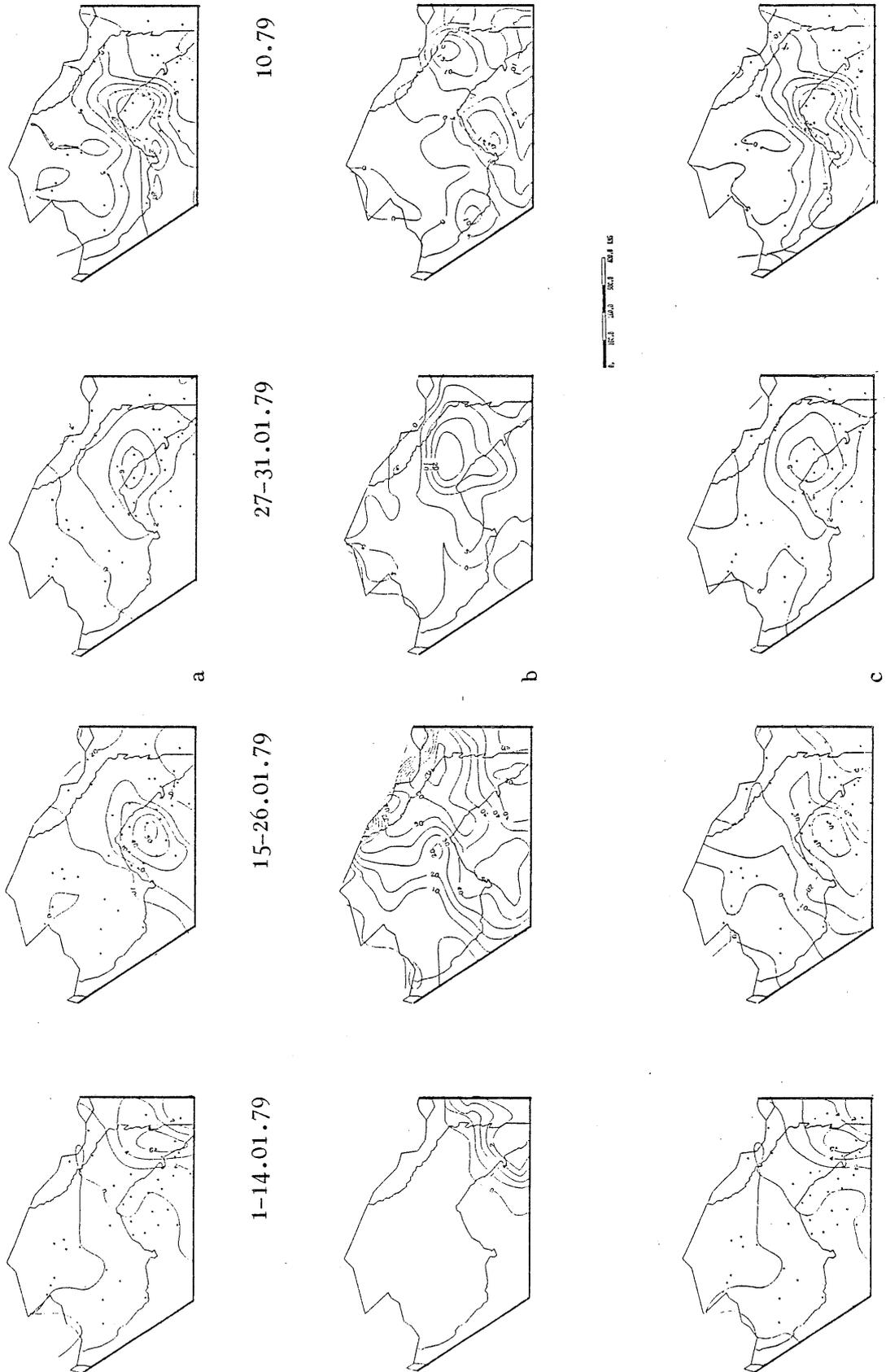
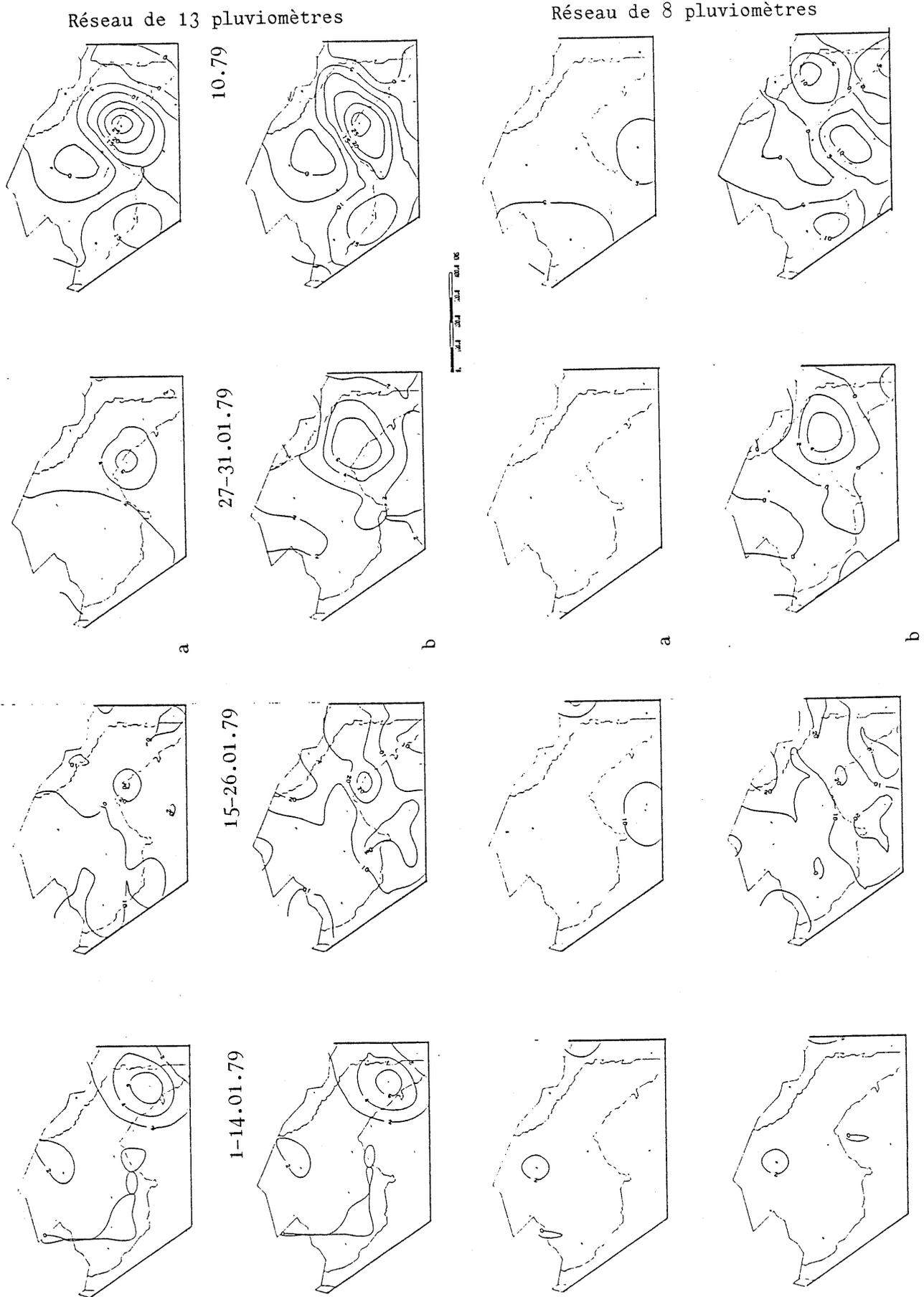


Figure V.9 : Exemples de cartes isohyètes obtenues (a) par krigeage et (b) cokrigeage à l'aide de sous-réseaux de 13 et 8 pluviomètres.



Appliqué au réseau de 13 stations, le cokrigeage se limite à modifier la forme des isohyètes krigées dans un sens souvent très réaliste si l'on se réfère aux résultats du krigeage de 43 stations : voir cellules de forte intensité des 27-31/01/79 et 10/79.

Appliqué au réseau de 8 stations, le cokrigeage est d'effet plus inégal : il peut se révéler parfaitement inefficace (pour le 1-14/01/79 par exemple) ou bien au contraire fournir des cartes d'allure très satisfaisante. Cet exemple de sous réseau très lache permet en fait de bien illustrer le fonctionnement du système de cokrigeage choisi ; en raison des conditions de non-biais choisies, ce système a pour effet de redistribuer spatialement, suivant l'allure générale perçue par le satellite, l'estimation moyenne faite par les pluviomètres. Ainsi, lorsque le réseau sol sous-estime notablement la valeur moyenne de la pluviométrie (exemple du 1-14/01/79), la fonction de redistribution du cokrigeage est inefficace ; il est cependant difficile de remédier à cet échec partiel car il est délicat de détecter a priori une sous-estimation du réseau sol qui conduirait à changer les conditions de non-biais et apporterait peut-être de meilleurs résultats.

En conclusion, malgré toute la prudence que doit inspirer l'analyse purement subjective d'un nombre restreint d'exemples, l'application du cokrigeage à ce type de mesures paraît intéressante, surtout si l'on se place dans la perspective de l'utilisation des estimations réalisées par des modèles hydrologiques ; en effet, la sensibilité de tels modèles à la distribution spatiale des précipitations (il n'est pas équivalent, au regard de l'infiltration ou de l'évaporation, de recueillir 3 mm de pluie uniformément répartis sur un bassin versant ou bien 30 mm concentrés sur 1/10 de la superficie du bassin) rend très précieuse l'effet de redistribution spatiale obtenu par cokrigeage.

2. MESURE DE LA FRACTION D'INSOLATION EN FRANCE

a) - Analyse de structure

L'analyse de la structure des mesures sol (présentée au Chapitre III) a montré la nécessité de coder les valeurs mesurées par centrage climatologique puis spatial afin de s'affranchir de diverses tendances systématiques et de se ramener à la fois à une distribution plus gaussienne des valeurs traitées et une meilleure stationnarité spatiale des quantités étudiées.

Afin de préparer l'application du cokrigeage à ces mesures sol et aux mesures du modèle CACTUS, nous devons maintenant vérifier que l'approche climatologique, employée pour les mesures sol, convient également à celles de CACTUS. Il est ensuite nécessaire d'analyser, dans ce même cadre, la structure croisée de ces deux types de mesures.

Rappelons que cette approche climatologique s'appuie sur les 71 journées de mesures sélectionnées pendant les mois de Mars à Juin 1984 (voir Chapitre III).

Analyse de la structure statistique des mesures CACTUS

Bien que les mêmes réserves, sur l'unicité du processus générateur et l'indépendance des réalisations traitées, formulées à propos des mesures héliographiques, soient applicables aux mesures CACTUS, nous avons suivi le même type d'approche climatologique. L'analyse des caractéristiques statistiques de ces mesures pendant la période choisie nous a conduit aux mêmes constatations que pour les mesures sol : variabilité spatiale significative des moyennes climatologiques (voir le variogramme de ces moyennes, figure V.10) et variabilité climatologique significative des moyennes spatiales des résidus à la moyenne climatologique (écart-type de 28 %). Nous avons donc choisi le même type de codage que celui décrit par la formule III.5 :

$$Z^W_T(t) = EZ_T(t) + X^W_T(t) + U^W_T \quad (V.2)$$

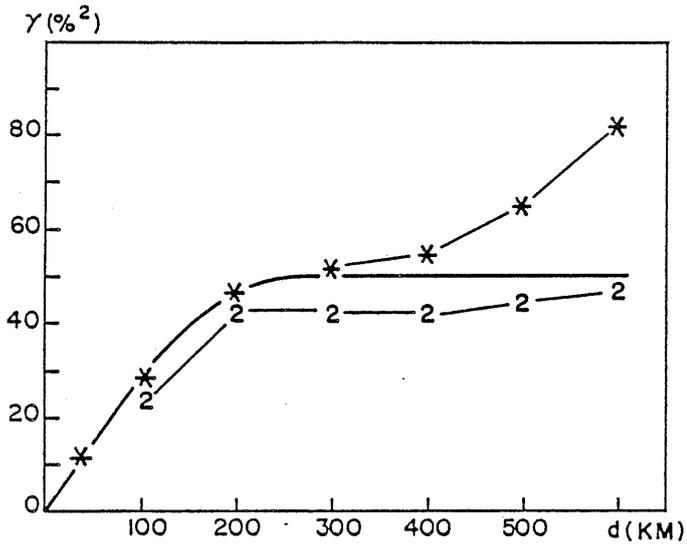


Figure V.10 : Variogramme du champ des moyennes climatologiques des mesures Cactus calculé à l'aide de 309 points pour la période choisie.

* Variogramme global
 2 Variogramme dans la direction N45E

Figure V.11 : Fonction de covariance expérimentale des mesures Cactus. après double centrage (1) ;

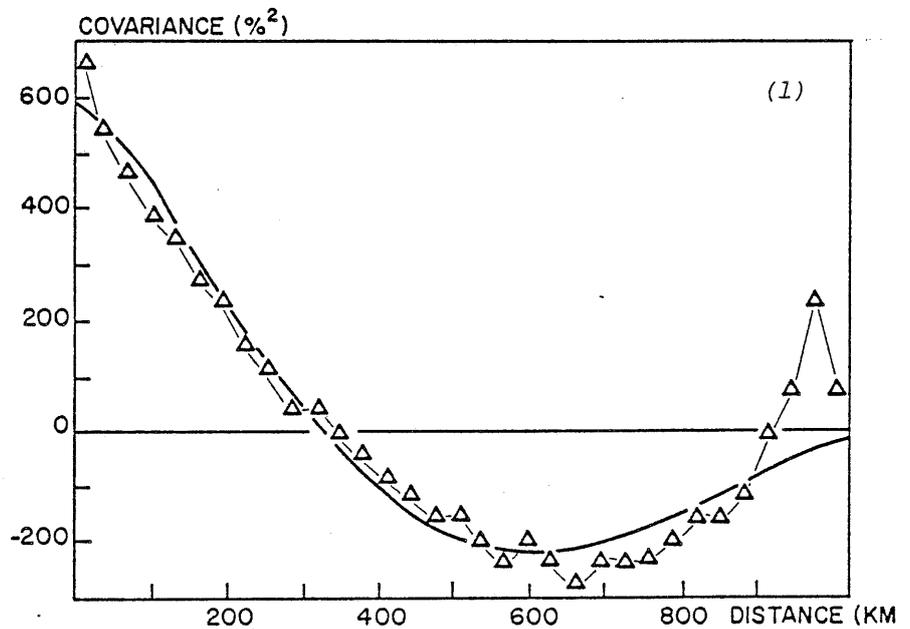
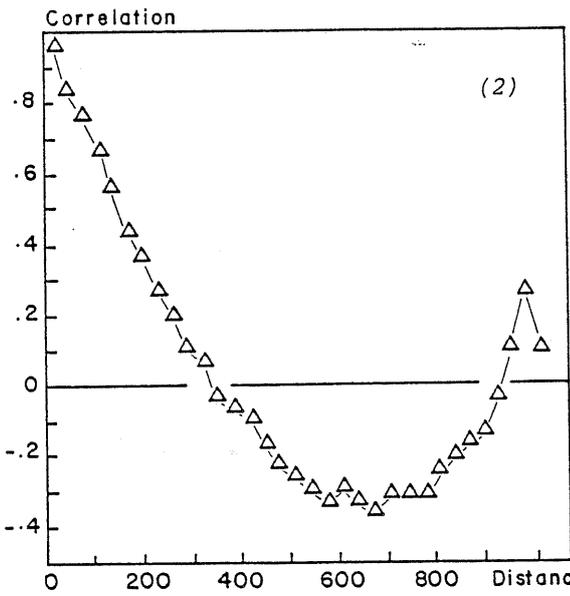
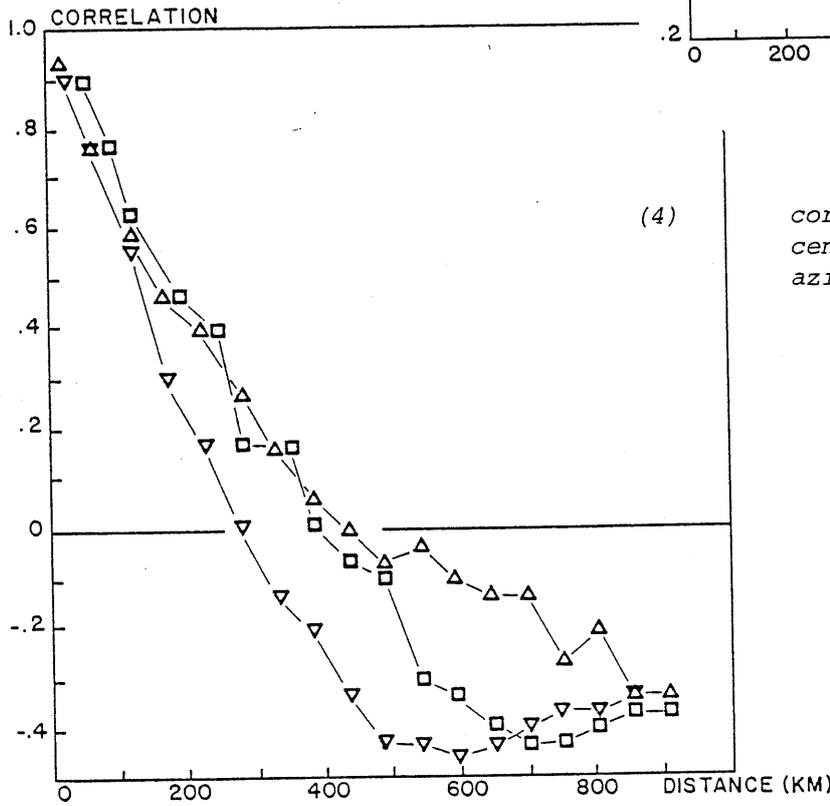
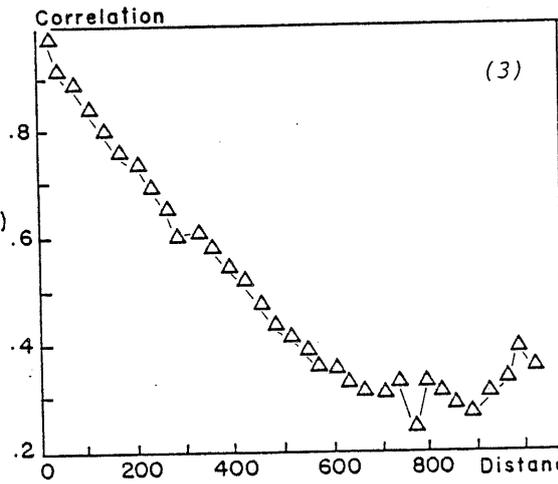


Figure V.11



corrélogrammes expérimentaux après double centrage (2)

après centrage climatologique seul (3)



corrélogramme après double centrage pour différents azimuts (4)

Direction ▽ N·0·E
 □ N·45·E
 △ N·90·E

La figure V.11 présente les différentes fonctions de structure obtenues, qui se rapprochent fortement de celles des mesures sol (voir figure III.14) :

- la fonction de covariance des valeurs $X^W_T(t)$ est de forme identique à celle des $X_w(t)$ (modèle exponentiel - cosinus s'annulant à 300 km environ) et d'amplitude légèrement supérieure (rapport d'échelle verticale sensiblement égale à 0.8).

- les corrélogrammes des valeurs $X^W_T(t)$ et $Y^W_T(t)$ sont eux rigoureusement confondus avec ceux des résidus sol pour les distances supérieures à 100 km ; aux faibles distances les corrélogrammes CACTUS sont plus proches de 1. ce qui peut indiquer que certaines microrégionalisations perçues par les mesures héliographiques ont été "gommées" par l'intégration spatiale réalisée par le modèle CACTUS ; il faut cependant être prudent dans ce type d'interprétation car les mesures comparées sont de natures très différentes et les comportements des corrélogrammes sont très voisins.

Pour compléter cette approche climatologique réalisée à l'aide de 71 journées de mesures, nous avons analysé champ par champ la structure des 15 journées de validation. Le détail des modèles ajustés empiriquement aux variogrammes de ces champs est donné par le tableau V.3. Le rapprochement des modèles ajustés aux variogrammes des mesures CACTUS et sol montre :

- une grande cohérence des modèles choisis ; généralement des modèles sphériques sans pépite et de portées sensiblement identiques (à l'exception des 12 et 19 mars pour lesquels la portée des mesures CACTUS est plus longue ;

- une variabilité systématiquement plus forte des mesures CACTUS (rapport d'échelle verticale moyen sensiblement égal à 0.7) ;

- la nécessité de rechercher un azimuth moins sensible à l'effet de dérive dans les deux tiers des cas, l'azimuth retenu pour la modélisation étant toujours le même pour les deux types de mesure.

Tableau V.3 : Détails de la reconnaissance de structures directes et croisées effectuée champ par champ pour les 15 journées de validation en utilisant 84 points de mesure.

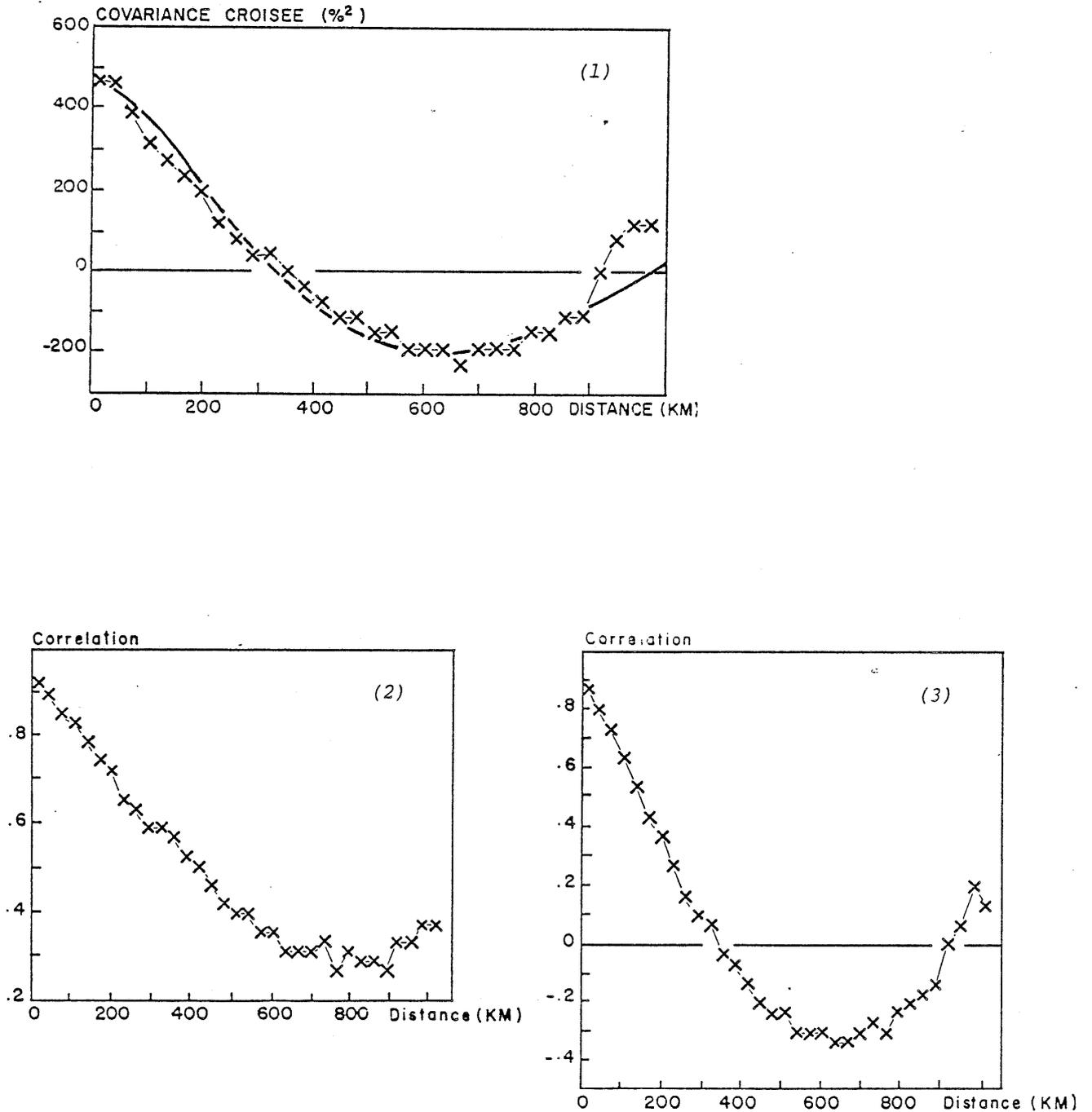
DATES	Modèle de variogramme ajusté aux mesures sol	Modèle de variogramme ajusté aux mesures Cactus	Modèle de variogramme croisé	Coefficient de corrélation intrinsèque*
02.03.84	S 500,500,0	S 600,500,0	S 500,500,0	.91
04.03.84	S 530,350,100	S 1250,350,100	S 550,350,100	.68
08.03.84	S(1) 400,250,0	S(1) 1000,450,0	S(1) 500,350,0	(.79)
10.03.84	S 1400,550,0	S 1650,500,0	S 1400,550,0	.92
11.03.84	L(2) .13, 0	L(2) .2, 0	L(2) .15, 0	.93
12.03.84	S(2) 500,400,0 (L.24,0)	S(2) 1500,750,0 (L.36,0)	S(2) 800,700,0 (L.28,0)	(.92)
14.03.84	S(1) 600,500,0	S(1) 700,450,0	S(1) 500,400,0	.77
15.03.84	S(1) 1000,750,0	S(1) 1200,750,0	S(1) 1000,750,0	.91
16.03.84	S(4) 800,600,0	S(4) 1200,500,0	S(4) 600,500,0	.61
19.03.84	S(2) 250,250,0	S(2) 450,500,100	S(2) 200,300,0	(.60)
21.03.84	S(2) 600,450,0	S(2) 1200,450,0	S(2) 800,450,0	.94
22.03.84	S(1) 500,500,0	S(1) 1000,500,0	S(1) 500,500,0	.71
23.03.84	S 300,300,0	S 600,300,0	S 300,300,0	.71
24.03.84	S(3) 250,600,0	S(3) 200,600,0	S(3) 170,600,0	.76
25.03.84	S(1) 500,700,0	S(1) 500,700,0	S(1) 400,700,0	.80

Unités : fractions d'insolation en % ; Distances en km

Description des modèles utilisés : S, (azimut modélisé : 1=EW, 2=SW-NE, 3=NS, 4=NW-SE),
modèle linéaire avec (palier, portée, pépite)
P, modèle pépitique pur avec (pépite)
L, modèle linéaire avec (pente, pépite)

* sous l'hypothèse d'une corrélation intrinsèque ; les champs s'écartant trop de cette hypothèse sont portés entre parenthèses)

Figure V.12 : Fonction expérimentale de covariance croisée entre les mesures sol et Cactus après double centrage (1) ; corrélogrammes croisés expérimentaux après double centrage (2) et centrage climatologique seul (3) ; corrélogrammes après double centrage pour différents azimuts (4).



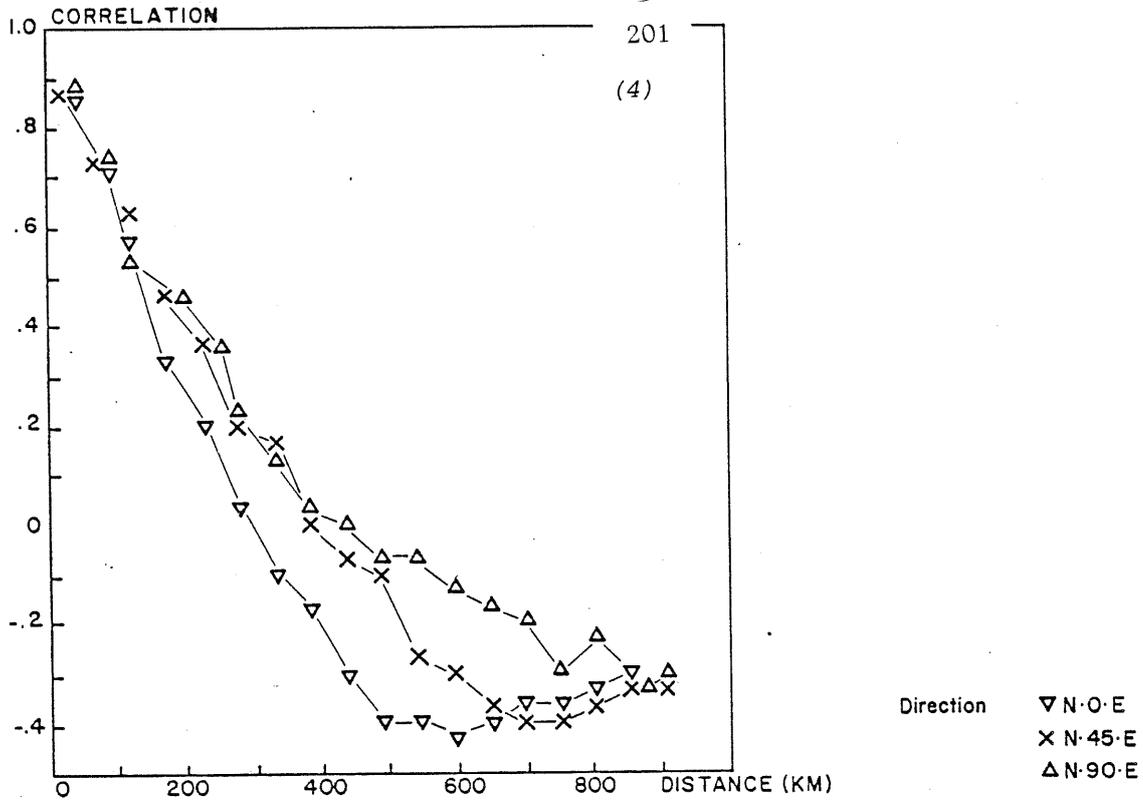
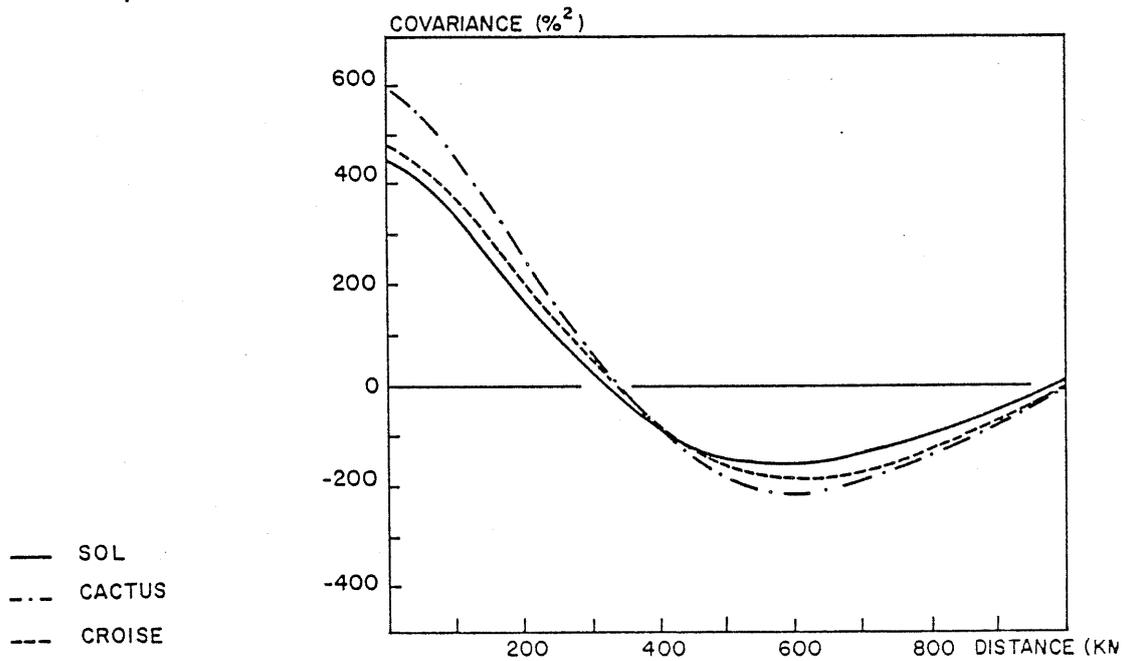


Figure V.13 : Modèles ajustés empiriquement aux fonctions de covariance directes et croisée des mesures sol et Cactus après double centrage (71 journées en 84 points de mesure).



En résumé, ces deux types d'approche climatologique et spatiale montrent que les mesures héliographiques et CACTUS ont une perception très cohérente de la structure spatiale de la variable fraction d'insolation.

Analyse de la structure croisée des mesures sol et CACTUS

Cette analyse va naturellement pouvoir s'appuyer dans un premier temps sur l'ensemble des réalisations disponibles, avec la réserve habituelle quant à l'unicité du processus générateur et l'indépendance des réalisations. Notons toutefois que l'on est autorisé ici à considérer un seul processus générateur pour les deux variables étudiées puisque les mesures traitées, bien que de natures différentes, s'appliquent à un même phénomène.

La figure V.12 présente les fonctions de structure croisée obtenues pour les différents codages utilisés pour la structure directe : nous sommes à l'évidence en présence d'un cas de corrélation intrinsèque. En effet, les différentes fonctions de covariance directe (voir figures V.11 et III.14) et croisée se déduisent d'un modèle commun unique $C(h)$ par simple affinité. La figure V.13 en donne l'illustration en rapprochant les modèles ajustés aux fonctions de covariance directes et croisées des résidus par rapport aux moyennes climatologiques puis spatiales. Le coefficient de corrélation "intrinsèque"

$$\rho(h) = \frac{C_{X,XT}(h)}{\sqrt{C_{XX}(h) \cdot C_{XT,XT}(h)}} = \frac{b_{X,XT}}{\sqrt{b_{XX} \cdot b_{XT,XT}}} \quad (V.3)$$

(avec $C_{X,XT}(h) = b_{X,XT} C(h)$, etc...).

vaut approximativement .9, ce qui traduit une bonne co-fluctuation des deux types de mesure. Ce coefficient diffère peu du coefficient de corrélation ponctuelle $\rho(0)$ des deux ensembles de mesures centrées dont la valeur est de .83 ; cette différence trouve certainement ici son explication dans les incertitudes d'ajustement des modèles (difficultés rencontrées notamment à propos des mesures CACTUS au voisinage de l'origine où la composante cosinus du modèle est mal adaptée, bien qu'elle soit par ailleurs indispensable pour représenter la partie négative de la fonction). Une

telle différence doit également faire penser à la présence possible d'un fort effet de pépite sur l'un des signaux ou sur les deux ; en effet dans ce cas, la fonction $C_{XX}(h)$ par exemple peut se noter :

$$C_{XX}(h) = b_{XX}C(h) + (1-\delta_h) C_{ee}(0)$$

où $C_{ee}(0)$ représente la variance du bruit blanc $e(t)$ responsable de l'effet de pépite ; alors, la covariance croisée n'étant pas affectée par ce bruit (puisque $C_{XT,e}(h) = 0$ pour tout h), le coefficient $\rho(h)$ gardera l'expression V.3 pour $h = 0$; $\rho(0)$ aura pour valeur :

$$\rho(0) = \frac{b_{X,XT}}{\sqrt{\left(b_{XX} - \frac{C_{ee}(0)}{C(0)}\right) \cdot b_{XT,XT}}}$$

et sera d'autant plus faible que la variance du bruit blanc est forte. Cette remarque montre au passage qu'un bon coefficient de corrélation intrinsèque, donc une bonne co-fluctuation spatiale, peut s'accompagner d'un faible coefficient de corrélation ponctuelle. Dans le cas présent, il n'en est rien ; il suffit pour s'en convaincre de considérer les corrélogrammes directs des résidus X et X_T (regroupés figure V.14) qui révèlent des effets de pépite très faibles (de l'ordre de 5 %).

Remarquons enfin que nous ne sommes pas dans un cas de corrélation intrinsèque parfaite ($\rho(h)=1$) qui traduirait l'existence d'une relation linéaire simple entre les résidus :

$$X_T^W(t) = \alpha \cdot X^W(t)$$

où α^2 représente le rapport des variances ; en fait ici ces résidus diffèrent d'un terme aléatoire e , sans doute indépendant de X puisque la covariance croisée est plus proche de la covariance des X que celle des X_T :

$$X_T^W(t) = X^W(t) + e^W(t) \quad (V.4)$$

Il est difficile de caractériser précisément le terme e en raison des fluctuations des fonctions de structure expérimentales ; nous l'avons vu, e

Figure V.14 : Corrélogrammes expérimentaux directs et croisé des résidus à la moyenne climatologique et à la moyenne spatiale pour 71 journées en 84 points de mesure.

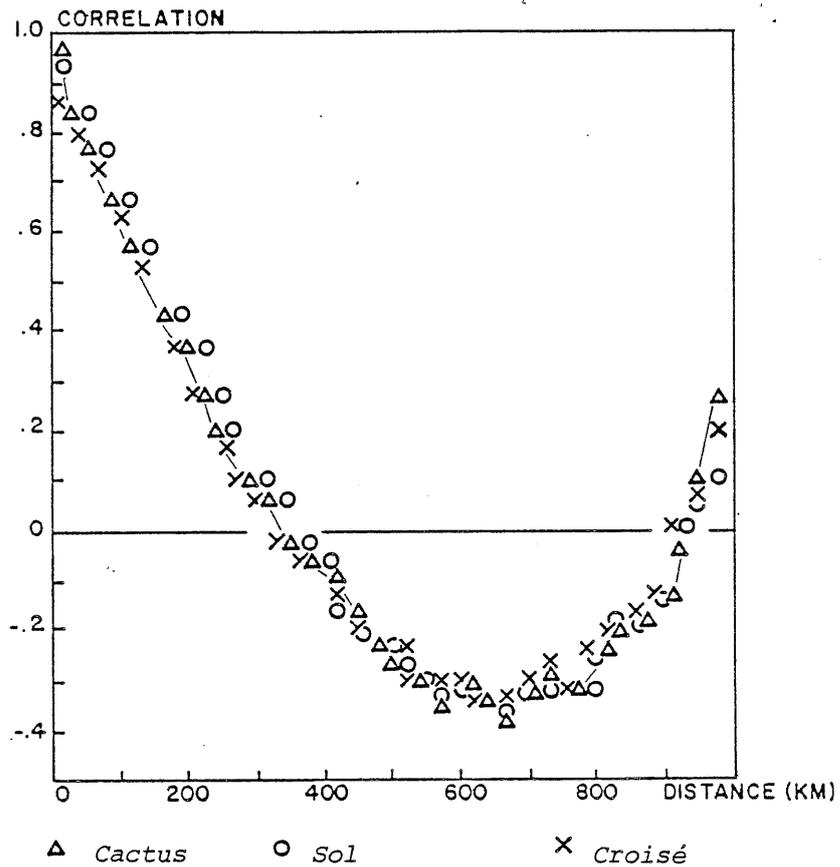
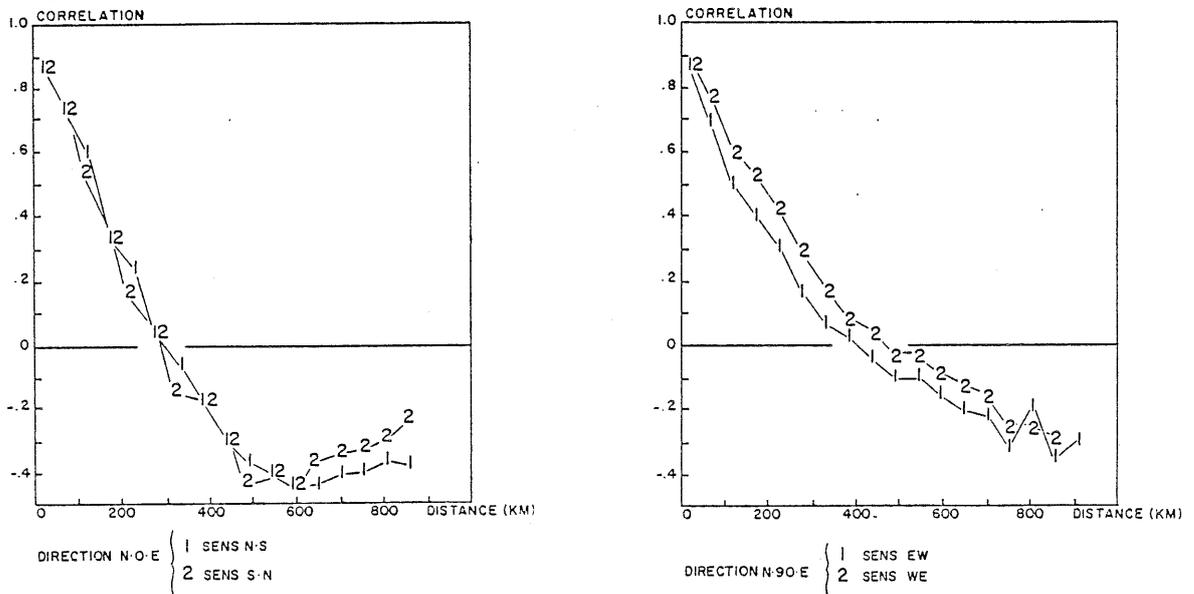


Figure V.15 : Corrélogrammes croisés, pour différentes directions et différents sens, des mesures sol et Cactus centrées par rapport à la moyenne climatologique et par rapport à la moyenne spatiale pour 71 journées et 84 points de mesure.



n'est pas un bruit blanc (sinon le corrélogramme des X_T serait plus pépétique que celui des X) mais au contraire une fonction autocorrélée suivant le même modèle que X (puisque X_T et X ont des corrélogrammes confondus) ; son amplitude est faible puisque, si on peut l'estimer à l'aide des modèles de covariance présentés figure V.13, le coefficient α égal à 1.07 (car $C_{X, X_T}(h) = \alpha \cdot C_{XX}(h)$ si l'on adopte le modèle V.4), alors la variance de e peut être estimée par $C_{ee}(0) = 88 \text{ } \mu\text{m}^2$ (car $C_{X_T, X_T}(h) = \alpha^2 C_{XX}(h) + C_{ee}(h)$). D'un point de vue physique, il est possible de penser que le terme e résulte de la position géographique variable des points de mesure au sol t_i par rapport aux points de la grille t_g les plus proches auxquels on les assimile.

L'ensemble de l'analyse qui précède repose sur une reconnaissance globale de la structure. Lors de l'analyse de la structure directe des deux types de mesures, nous avons souligné la présence d'une faible anisotropie que nous avons décidé de négliger : nous pouvons vérifier, figure V.12, qu'il en est de même pour la structure croisée.

Enfin, il est nécessaire d'examiner la symétrie des fonctions de covariance suivant les directions analysées afin de mettre en évidence d'éventuels décalages dans l'espace de ces ensembles de mesures. Compte-tenu de l'approche climatologique choisie, il faut bien voir que l'on se limite à la recherche d'un décalage d commun à l'ensemble des champs considérés ; en reprenant le modèle V.4, ce type de décalage peut se noter :

$$X^W_T(t) = X^W(t+d) + e^W(t)$$

Nous sommes ici dans un cas où il est vraisemblable que les images de télédétection soient décalées par rapport aux mesures sol (le positionnement des images satellite n'est en effet jamais très rigoureux) ce qui conduirait à une fonction de covariance croisée :

$$C_{X, X_T}(h) = C_{XX}(h-d) + C_{eX}(h)$$

contenant un terme dont la dissymétrie dépend de d (C_{XX} étant une fonction symétrique de h) ; alors si les fluctuations de e ne sont pas prépondérantes sur celles de X il serait possible d'identifier aisément d .

Dans le cas où l'on peut craindre que le décalage d varie d'une réalisation à l'autre, l'approche climatologique est impossible. Il est bien sûr envisageable de construire pour chaque réalisation la fonction de covariance simple croisée ; les mêmes remarques sur la fragilité de cet outil (trop sensible au comportement moyen des variables) peuvent être faites ici comme au paragraphe précédent. On peut également penser à construire cette fonction de covariance champ par champ à l'aide des résidus à la moyenne climatologique ; mais il est alors permis d'avoir des doutes sur la validité de la moyenne climatologique en raison même des différents décalages d^w . Enfin, le nombre de réalisations à traiter est un facteur dissuasif supplémentaire.

Devant ces difficultés, nous avons préféré tenir un raisonnement physique simple à propos de nos mesures : compte-tenu de leurs caractéristiques (voir en Introduction) seul un mauvais positionnement des images satellite peut entraîner une liaison décalée entre mesures sol et CACTUS. Ces décalages peuvent parfaitement varier d'un jour à l'autre et sont donc particulièrement difficiles à identifier. Nous nous remettons donc au contrôle très strict du positionnement réalisé par l'examen continu des cartes d'albédo (voir le détail dans S. GALLE, 1987) pour faire l'hypothèse d'une absence de décalage dans la série des images satellite. Nous nous bornons à vérifier, par l'examen de la symétrie du corrélogramme climatologique des résidus à la moyenne climatologique et à la moyenne spatiale, que nous n'avons pas commis d'erreur systématique dans le positionnement des images CACTUS par rapport au réseau sol. La figure V.15 regroupe ces corrélogrammes croisés pour les deux directions d'anisotropie maximale (NOE et N90E), et pour des sens opposés : aucune dissymétrie significative n'est identifiable.

De la même façon que pour les structures directes, nous pouvons compléter l'approche climatologique ci-dessus par l'analyse champ par champ de la structure croisée des mesures héliographiques et CACTUS. Sous l'hypothèse de l'absence de décalages géométriques entre ces ensembles de mesure, nous pouvons construire pour chaque journée le variogramme croisé et le rapprocher des variogrammes directs (voir tableau V.3 et quelques exemples figure V.16).

Aux incertitudes de modélisation près, il est clair que, pour la presque totalité des champs décrits dans le tableau V.3, nous sommes bien en présence de corrélation intrinsèque ; le calcul du coefficient de corrélation intrinsèque (rapport du paramètre d'échelle verticale de la structure croisée à la racine du produit des paramètres de la structure directe) montre une bonne co-fluctuation des mesures. En fait seuls deux champs (12 et 19/03/84) pour lesquels la portée des mesures CACTUS est nettement supérieure à celle des mesures sol, s'écartent de ce modèle ; il s'agit de journées pour lesquelles existe une forte dérive que la construction des variogrammes par directions n'arrive pas à gommer. Notons, pour finir, que pour huit autres champs, cette analyse par direction s'est avérée fructueuse, la direction privilégiée étant toujours commune aux trois variogrammes.

b) - Résultats obtenus par cokrigage

Les résultats présentés ci-dessous portent sur la sélection de 15 journées de Mars 1984 que nous avons décrite au Chapitre III. Parmi les 84 stations du réseau sol, 61 stations de validation ont été utilisées pour calculer les valeurs de référence de la fraction journalière d'insolation sur les mailles de la grille CACTUS contenant ces stations. Les 23 stations restantes (puis 13 et 5 d'entre elles) sont utilisées pour étalonner les images CACTUS par cokrigage.

La mise en oeuvre du système de cokrigage repose sur un ensemble de choix d'hypothèses qu'il est nécessaire de rappeler avant de commenter les résultats obtenus :

Figure V.16 : Exemples de variogrammes croisés entre mesures héliographiques et Cactus ; en référence sont portés les variogrammes directs.

- x Variogramme croisé expérimental
- .- Modèle sphérique ajusté empiriquement
- Variogrammes directs sol
- ▲ et Cactus

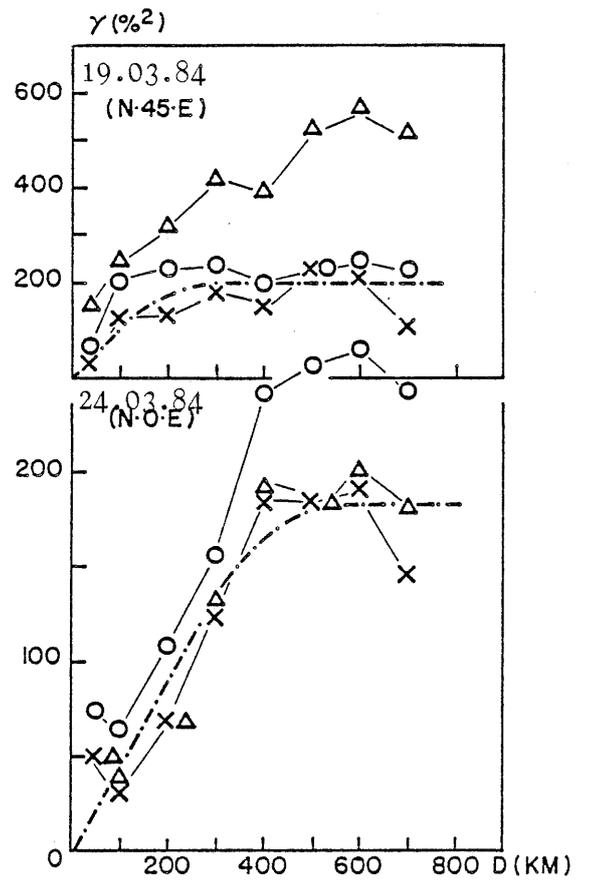
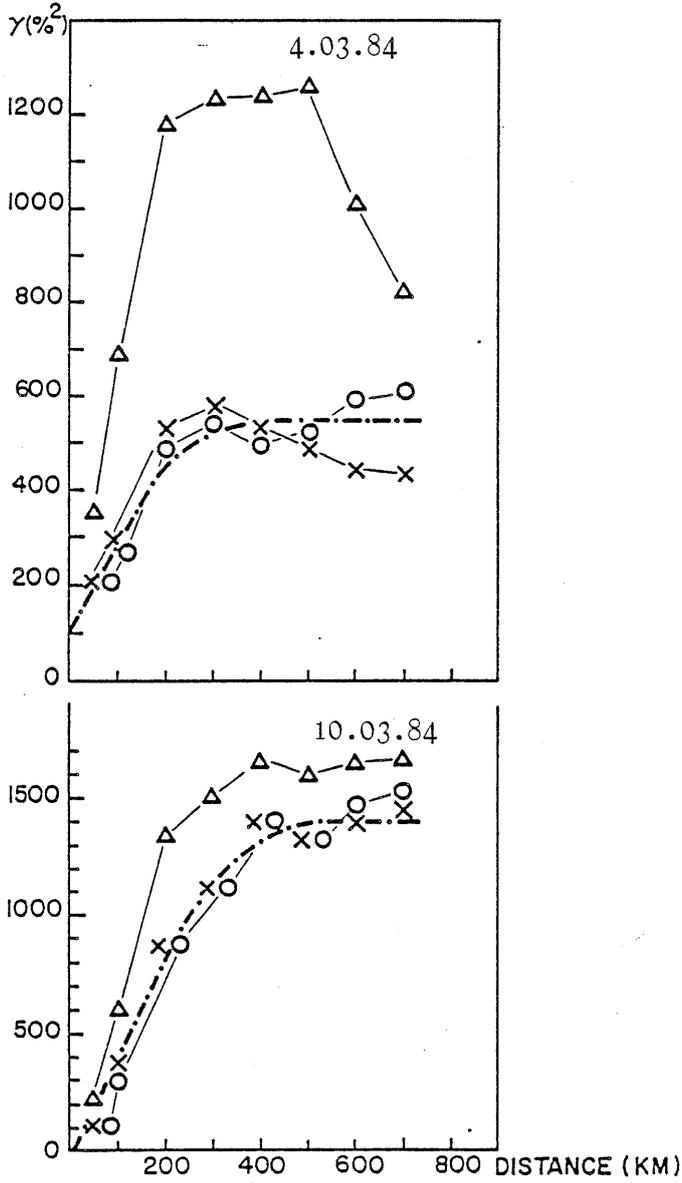


Figure V.17 : Variogrammes directs et croisé des champs de moyenne climatologique de la fraction journalière d'insolation calculée à l'aide des mesures sol et Cactus en 84 points pour 71 journées.

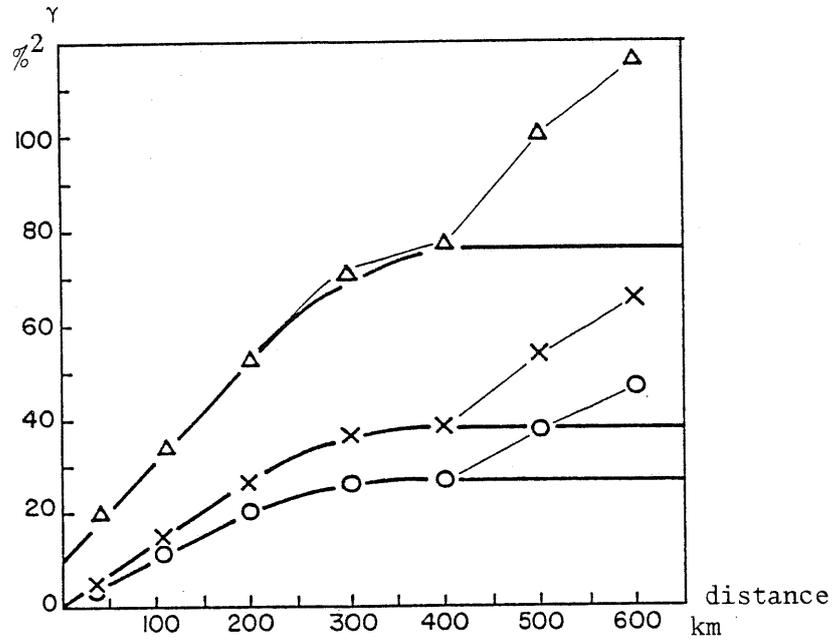
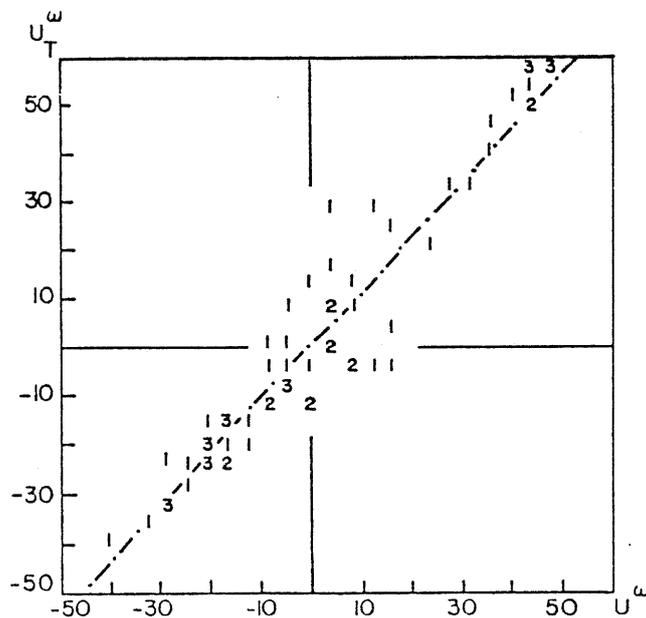


Figure V.18 : Diagramme de corrélation entre les moyennes spatiales des résidus climatologiques des mesures sol et Cactus pour 71 journées en 84 points.



- contexte climatologique.

Nous avons considéré que les 15 journées traitées font partie de l'ensemble des 71 journées analysées précédemment et supposées être des réalisations indépendantes d'un processus unique.

- normalité et stationnarité d'ordre 2 des résidus à la moyenne climatologique et à la moyenne spatiale.

La décomposition des fonctions $Z(t)$ et $Z_T(t)$ suivant les formules III.5 et V.2, nous a permis d'identifier les fonctions de structure directes et croisée par modélisation des fonctions de covariance des résidus X et X_T .

Etant donné ces hypothèses, deux problèmes pratiques se posent pour fournir une estimation de Z : i) celui de la reconstitution de la moyenne climatologique $EZ(t_0)$ au point t_0 d'estimation et ii) celui de l'estimation pour chaque champ w des moyennes spatiales U^w et U^w_T des résidus climatologiques $Y^w(t)$ et $Y^w_T(t)$.

i) Nous disposons pour estimer $EZ(t_0)$ à la fois des moyennes climatologiques aux points de mesure sol et aux noeuds de la grille CACTUS ; il est donc logique de réaliser cette estimation par cokrigeage. Etant donné le faible biais des estimations CACTUS de la moyenne climatologique (moins de 9 % d'erreur absolue) et la bonne co-fluctuation des deux champs (voir les variogrammes obtenus figure V.17), ce cokrigeage apporte des résultats sensiblement meilleurs que le seul krigeage des valeurs sol de ces moyennes tel que nous l'avons pratiqué au Chapitre III (le pourcentage de la variance des moyennes climatologiques sur les 61 cibles test expliqué par les valeurs estimées à l'aide de 13 stations d'étalonnage passe de 75 à 84). Cette estimation contribue donc déjà à améliorer, en terme de co-fluctuation, les résultats du cokrigeage par rapport à ceux du krigeage.

ii) Le problème de l'estimation des moyennes spatiales U^w et U^w_T des résidus Y^w et Y^w_T pose en fait la question de savoir sur quelles variables va porter le cokrigeage. Le premier choix possible est celui d'estimer $X^w(t_0)$ à l'aide des $X^w(t_1)$ et $X^w(t_g)$ puis d'y ajouter une estimation de U^w ; les variables X^w et X^w_T étant d'espérance nulle, le

ystème de cokrigage ne nécessite alors aucune condition de non-biais, par contre, l'estimation de U^W peut être délicate lorsque le nombre de points d'étalonnage décroît ; de plus, l'accès aux résidus X^{WT} nécessite également l'estimation de U^{WT} . Un deuxième choix peut consister à estimer directement $Y^W(t_0)$ à l'aide des $Y^W(t_i)$ et $Y^{WT}(t_g)$; le système de cokrigage permet alors de s'affranchir des estimations de U^W et U^{WT} qui seront "filtrées" par des conditions de non-biais appropriées ; étant donné que ces moyennes sont, par définition, constantes dans l'espace et, d'après l'analyse de leurs fluctuations pour les 71 journées disponibles, fortement ressemblantes ($U^{WT} = 1.12 \times U^W + \varepsilon^W$ avec un écart-type des ε^W inférieur à 6 %, voir figure V.18), nous avons été amenés à choisir une condition de non-biais unique :

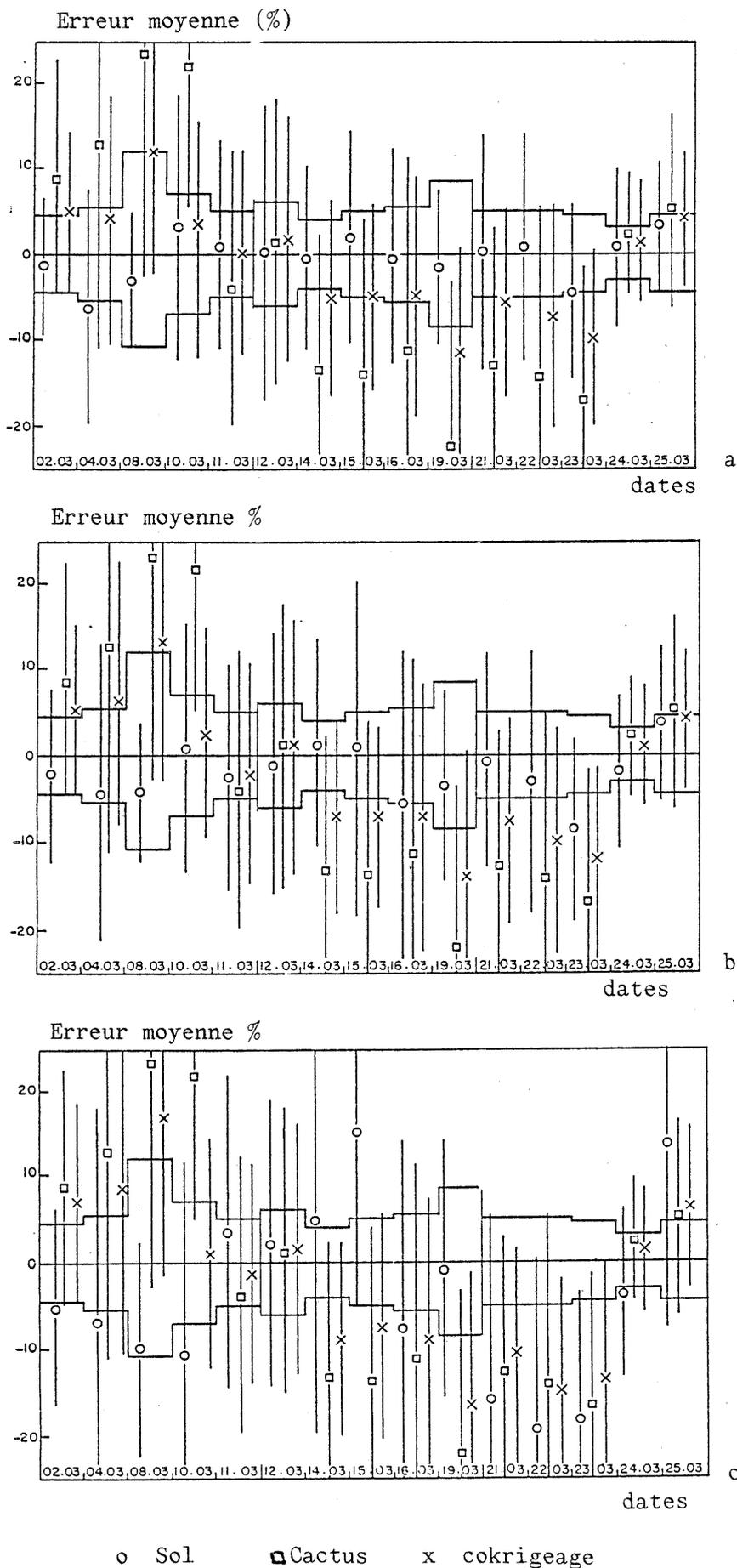
$$\sum_{i=1}^n \lambda^i Y(t_0) + \sum_{g=1}^G \lambda^g Y_T(t_0) = 1$$

pour les poids de cokrigage appliqués aux résidus climatologiques Y et Y_T . Ce deuxième choix étant plus simple d'un point de vue pratique, nous l'avons préféré au premier.

La démarche proposée ci-dessus, appliquée à un réseau d'étalonnage de 23 puis 13 et 5 stations (densités de $1/(200)^2$ puis $1/(300)^2$ et $1/(500)^2$ station par km^2 environ), nous a conduit à des résultats que nous avons comparés à ceux obtenus à l'aide des seules mesures héliographiques et CACTUS ; les critères habituels de comparaison appellent les remarques suivantes :

- les diagrammes d'efficacité présentés, figure V.19, traduisent tout d'abord l'évolution jour par jour de l'erreur moyenne expérimentale de cokrigage qui prend de manière générale des valeurs intermédiaires aux erreurs moyennes des dispositifs de base. Cela s'explique simplement en rappelant que ces erreurs reflètent de manière très directe l'erreur d'estimation de la moyenne climatologique du phénomène et l'erreur d'estimation de la moyenne des résidus à la moyenne climatologique : la première erreur se répercute de manière constante sur l'ensemble des journées ; elle est faible et peu sensible au nombre de stations d'étalonnage dans le cas du cokrigage (toujours inférieure à 2 %) ; elle

Figure V.19 : Diagrammes d'efficacité du cokrigage en modèle climatologique sur 61 cibles test ; en référence sont portés les résultats du modèle Cactus et du krigeage en modèle climatologique appliqué aux seules stations d'étalonnage. Réseau d'étalonnage de (a) 23 stations, (b) 13 stations et (c) 5 stations.



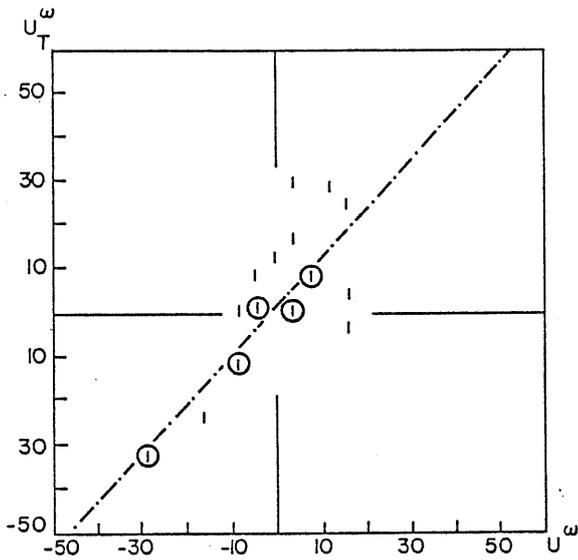
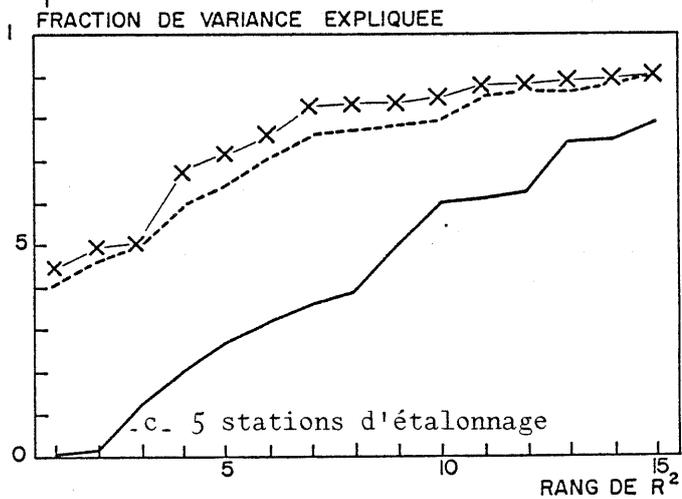
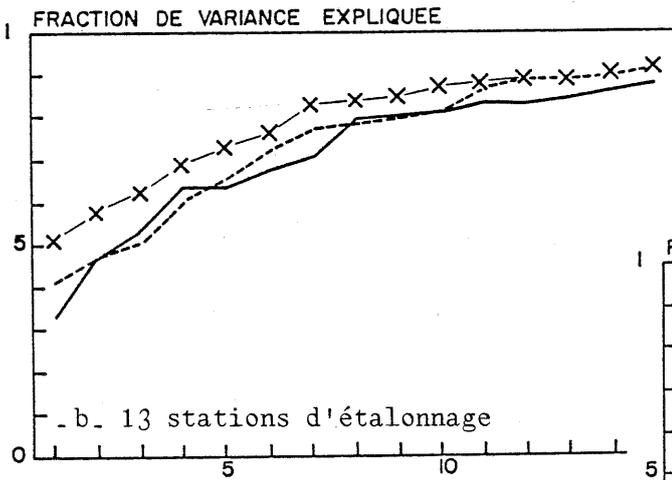
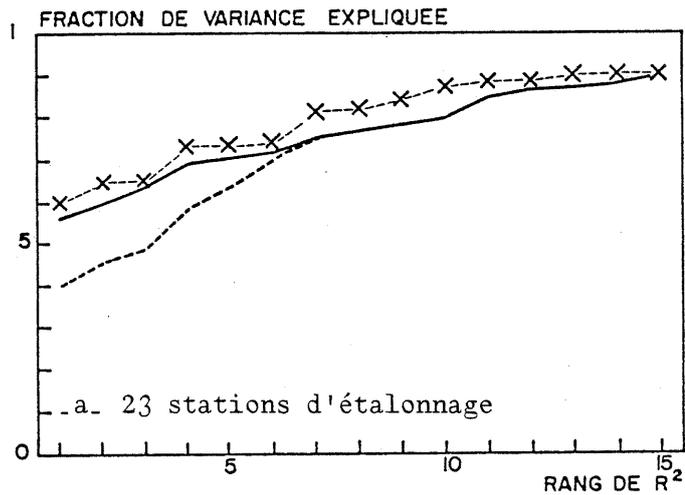


Figure V.20 : Diagramme de corrélation entre les moyennes spatiales des résidus climatologiques des mesures héliographiques et Cactus pour les 15 journées testées (les 5 meilleures reconstitutions sont cerclées : 10, 11, 12 puis 24 et 25.03.84.

Figure V.21 : Courbes d'efficacité du cokrigeage en modèle climatologique sur 61 cibles test ; en référence sont portés les résultats du modèle Cactus seul et du cokrigeage en modèle climatologique appliqué aux seules stations d'étalonnage.



x cokrigeage
 — Krigage en modèle climatologique
 --- Modèle Cactus

est plus forte dans le cas du krigeage des seules mesures sol (passant de 2 à 5 % lorsque la densité du réseau décroît) ; elle est forte dans le cas du modèle CACTUS (-9 %) ; la deuxième erreur varie d'un jour à l'autre ; dans le cas du cokrigeage avec condition de non-biais unique, tel que nous l'avons pratiqué ici, elle dépend de façon prépondérante de la moyenne U^W_T des mesures CACTUS (ce qui explique la grande stabilité des résultats globaux du cokrigeage lorsque la densité du réseau d'étalonnage varie) ; les meilleures reconstitutions par cokrigeage, au sens de l'erreur moyenne, sont obtenues pour les 5 journées (10, 11, 12 puis 24 et 25/03/84) pour lesquelles la relation $U^W_T = 1.2 U^W$ donne les plus faibles résidus (voir figure V.20).

Ces diagrammes d'efficacité fournissent ensuite l'écart-type expérimental des erreurs qui traduit une dispersion des estimations par cokrigeage assez peu sensible au nombre de points d'étalonnage et presque systématiquement inférieure à celle des dispositifs de base (sol seul et CACTUS).

De manière globale, l'erreur absolue moyenne par cokrigeage (exprimée en pourcentage de la valeur moyenne de la référence pour chaque jour) passe de 23 % à 25 % puis 30 % lorsque le réseau d'étalonnage décroît en densité (23 puis 13 et 5 stations sur la France) alors que l'erreur moyenne par krigeage des seules mesures héliographiques passe de 21 % à 24 % puis 38 % et l'erreur moyenne de CACTUS est de 38 % ; ainsi l'intérêt du cokrigeage en terme d'erreur moyenne n'apparaît nettement qu'aux très faibles densités de réseau sol.

- les courbes d'efficacité regroupées, figure V.21, marquent toutes une amélioration des performances du meilleur dispositif de mesure considéré individuellement lorsque le cokrigeage est employé ; ceci signifie qu'en terme de co-fluctuation, le cokrigeage tire parti, pour les densités de réseau sol utilisées, de la complémentarité des deux dispositifs de mesure pour établir des estimations de qualité supérieure. Cette constatation est confirmée de manière globale par les coefficients de corrélation entre valeurs reconstituées et valeurs de référence comme l'indique le tableau V.4. Les améliorations paraissent très modestes à l'examen des critères ci-dessus, qui s'appuient sur la corrélation simple ;

afin d'en faire apparaître l'importance relative, il est nécessaire de recourir au critère de corrélation partielle ; si l'on utilise opérationnellement pour effectuer l'estimation de la fraction journalière d'insolation (ici pour les 15 journées considérées et sur 61 cibles test) le seul réseau héliographique de 5 stations (resp. 13 et 23), on peut attendre de l'information CACTUS traitée par cokrigeage un gain d'explication de 19 % (resp. 15 % et - 1 %) de la variance des valeurs de référence ; normé par la variance résiduelle globale (c'est à dire traduit en carré du coefficient de corrélation partielle), ce gain s'élève à .86 (resp. .59 et -.03), tenant ainsi compte de la difficulté d'améliorer les performances d'un dispositif déjà très efficace. S'ils sont confirmés par une analyse plus large (voir S. GALLE, 1987), ces gains montrent une limite très nette au-delà de laquelle l'information apportée par la télédétection devient inutile : lorsque l'on dispose de 23 stations au moins, c'est à dire lorsque la densité du réseau sol est telle que l'interdistance des stations devient inférieure à la "portée" du phénomène. Un dernier critère de co-fluctuation peut être obtenu en comparant jour par jour les performances du cokrigeage avec celles des dispositifs sol et CACTUS en terme de fraction expliquée de la variance de référence ; la figure V.22, construite dans le cas d'un réseau héliographique de 13 stations, permet de distinguer 3 cas de figure :

i) Celui des journées où la co-fluctuation des deux types de mesure est excellente (coefficient de corrélation intrinsèque proche de 1.); dans ce cas les performances du cokrigeage sont très proches de celles de CACTUS ; ce résultat est sans doute à rapprocher de l'équivalence formelle entre cokrigeage et régression dans le cas de corrélation intrinsèque parfaite.

ii) Celui des journées où la co-fluctuation est moins bonne mais où l'hypothèse de corrélation intrinsèque demeure valide (les 4, 14, 22, 23, 24 et 25/03/84) ; il s'agit du cas le plus démonstratif des propriétés de cokrigeage qui améliore systématiquement les performances des dispositifs qu'il combine.

iii) Celui des journées pour lesquelles l'hypothèse de corrélation intrinsèque est prise en défaut (les 8 et 19/03/84 de manière

Figure V.22 : Fraction, journée par journée, de la variance des valeurs de référence en 61 cibles expliquée par les résultats de cokrigage à l'aide de 13 stations héliographiques. En référence sont portés les coefficients de corrélation intrinsèque extraits du tableau V.3.

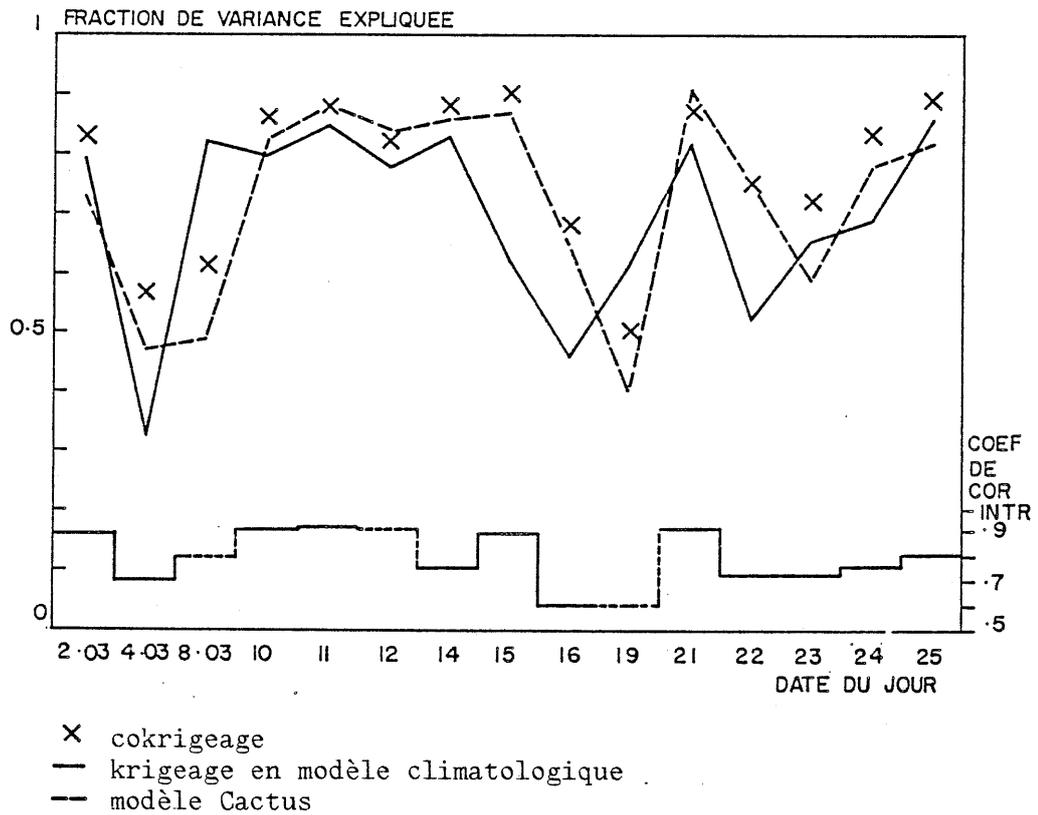
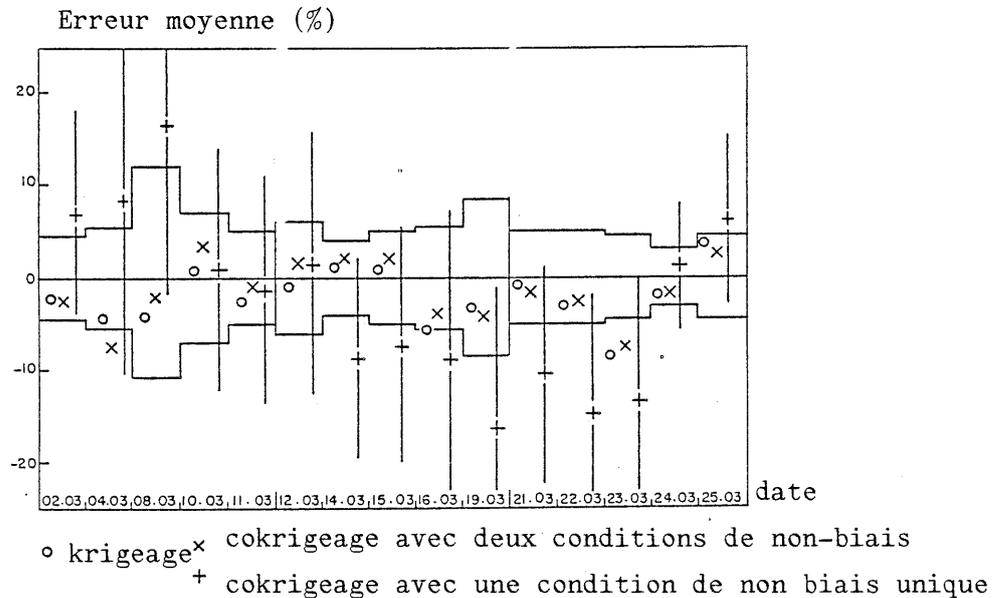


Figure V. 23 : Diagrammes d'efficacité du cokrigage à l'aide de 13 stations sol pour deux conditions de non-biais différentes ; en référence sont portées les erreurs moyennes obtenues par krigeage des seules mesures sol.



très nette et peut-être le 12/03/84) ; le modèle de structure climatologique utilisé étant erroné, il est logique que le cokrigeage ne sache pas tirer parti des mesures disponibles ; l'utilisation d'un modèle de structure mieux approprié (utilisant une covariance croisée plus faible) conduirait à égaler les performances du réseau sol seul et améliorerait les résultats ; il s'agit là d'un défaut de l'approche climatologique, qui court toujours le risque de mélanger des réalisations aux caractéristiques différentes.

Tableau V.4 : Coefficients de corrélation entre les valeurs de référence en 61 cibles et les valeurs reconstituées par le modèle CACTUS et pour différentes densités de stations héliographiques, par krigeage et cokrigeage (pour 15 journées).

	Krigeage			CACTUS	Cokrigeage		
	5	13	23		5	13	23
nombre de stations sol	5	13	23		5	13	23
Corrélation	.671	.876	.903	.833	.859	.885	.900
Pourcentage de variance expliquée	.450	.767	.815	.690	.738	.783	.810
Corrélation après centrage par rapport aux moyennes	.565	.798	.837	.838	.844	.857	.867
Pourcentage de variance expliqué correspondant	.319	.637	.701	.700	.712	.734	.752
Erreur absolue normée par la moyenne	37.9	23.6	20.5	38.4	29.6	25.4	23.1

En résumé, les choix d'hypothèses effectués au début de ce paragraphe nous ont apporté des résultats satisfaisants au sens de leur cofluctuation spatiale avec les mesures de référence mais décevants quant à leur précision moyenne. Il semble donc que la condition de non-biais choisie est mal adaptée au problème traité ; ce choix a en fait deux alternatives :

Figure V.24 : Courbes d'efficacité du cokrigeage à l'aide de 13 stations sol pour deux conditions de non-biais différentes ; en référence est portée la courbe d'efficacité du krigeage des seules mesures sol.

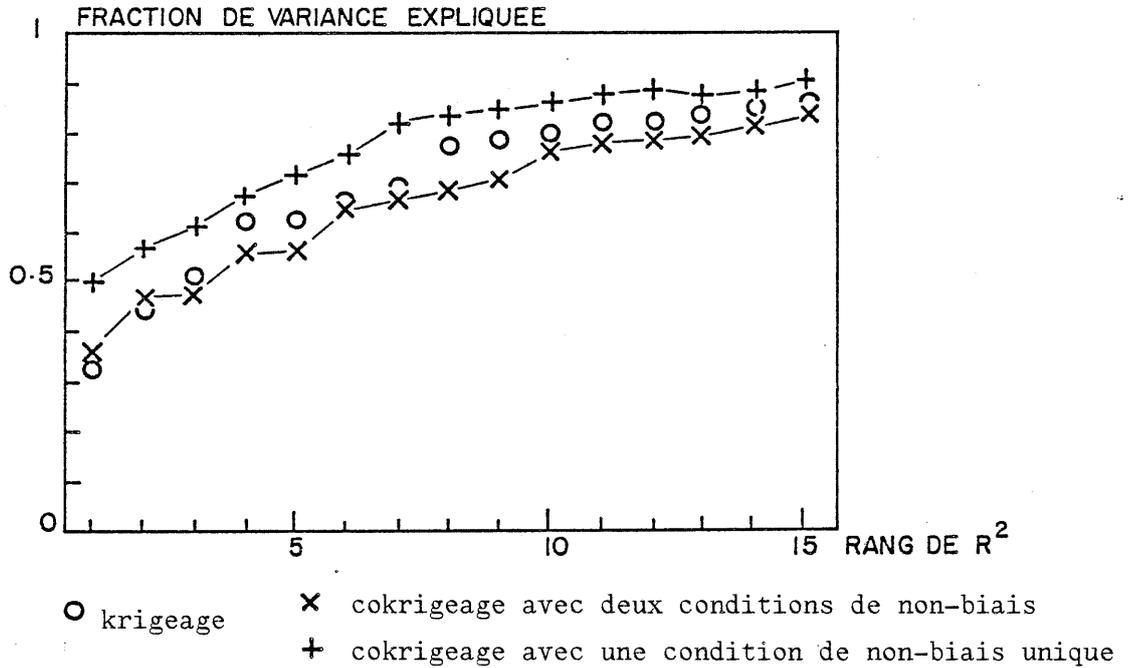
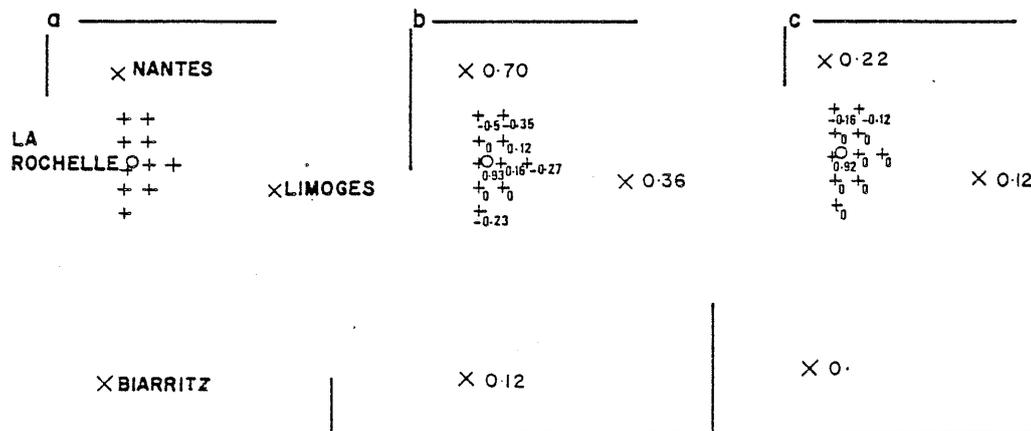


Figure V.25 : Pondérations calculées par le système de cokrigeage pour reconstituer l'insolation à la Rochelle (b) en appliquant deux conditions de non biais et (c) en appliquant une condition unique (la situation géographique des points de mesure sol (x) et Cactus (+) est rappelée en (a) sans tenir compte des échelles exactes.



i) Nous décidons de maintenir l'application du cokrigeage aux résidus climatologiques Y, mais nous considérons que ces résidus sont d'espérance différente selon que l'on considère les mesures héliographiques ou CACTUS. Deux conditions de non-biais sont alors nécessaires :

$$\sum_{i=1}^n \lambda^i_Y(t_0) = 1 \quad \text{et} \quad \sum_{g=1}^G \lambda^g_T(t_0) = 0$$

Un essai de cokrigeage à l'aide de 13 stations d'étalonnage en utilisant cette hypothèse nous permet de constater :

- une réduction sensible du biais de reconstitution jour par jour (voir le diagramme d'efficacité figure V.23) ; cette réduction correspond en fait simplement à un rapprochement des erreurs de cokrigeage avec celles de krigeage ; cette amélioration sera donc beaucoup moins nette lorsque la densité du réseau diminuera.

- une détérioration sensible de la cofluctuation spatiale des résultats de cokrigeage avec la référence (voir les courbes d'efficacité figure V.24) ; cette détérioration conduisant à une qualité de reconstitution inférieure à celle apportée par le seul réseau sol.

Ces résultats s'expliquent de manière qualitative en considérant l'aspect contradictoire des contraintes imposées aux poids de cokrigeage par ces hypothèses ; en effet, l'existence d'une forte cofluctuation entre les deux types de mesure conduit, en l'absence de condition de non-biais à une pondération forte des points explicatifs les plus proches du point reconstitué, qui sont toujours dans notre cas des points de grille CACTUS; imposer à la somme de ces poids d'être nulle rend le système très instable: les poids affectés aux mesures sol sont très proches de ceux du système de krigeage et ceux affectés aux mesures de télédétection présentent de fortes variations ; les résultats obtenus peuvent donc s'interpréter comme des résultats de krigeage auxquels s'ajoute un bruit provenant d'une pondération "fantaisiste" des mesures CACTUS. Au contraire, une condition de non-biais unique trouble beaucoup moins la distribution naturelle des poids obtenue par le sous-système de covariance; il faut remarquer que ce

sous-système vise à reconstituer une valeur sol (vecteur de droite rempli à l'aide de C_{ZZ} et $C_{Z,ZT}$) ce qui conserve aux points sol, même éloignés un certain pouvoir explicatif. L'ensemble de ces considérations est illustré par la figure V.25 qui présente dans les deux cas de figures les pondérations choisies par le système de cokrigeage pour reconstituer la valeur du phénomène à LA ROCHELLE.

Cette première alternative n'est donc pas praticable car elle compromet les performances de cokrigeage en terme de variabilité spatiale, sans apporter de garantie de précision lorsque la densité du réseau d'étalonnage diminue.

ii) Nous décidons d'appliquer le cokrigeage, sans contrainte de non-biais, aux résidus climatologiques centrés, X . Mais se pose alors le problème de l'estimation des moyennes U^W et U^{WT} . La qualité des résultats obtenus dépendra alors de la qualité d'estimation de U^W qui se dégrade lorsque le nombre de stations d'étalonnage diminue. Il est donc peu probable d'obtenir une amélioration des résultats à l'aide de cette seconde alternative.

En conclusion, cette application du cokrigeage à des données d'insolation nous permet de tirer deux enseignements pratiques sur l'apport du cokrigeage :

- tout d'abord la forte incidence du choix des conditions de non-biais sur la qualité des résultats obtenus notamment en terme de variabilité spatiale ; il semble en effet indispensable de traiter des variables ayant même moyenne spatiale afin de préserver la capacité du cokrigeage à intégrer aux résultats fournis par la perception fine des fluctuations spatiales qu'est sensée apporter la télédétection. Cette constatation remet bien sûr en cause l'application présentée à propos de la pluviométrie en Arabie ; il était toutefois impossible de négliger le fort biais que présentaient les mesures de télédétection et douteux de les corriger préalablement par régression dans la mesure où elles provenaient déjà d'une régression.

- ensuite la modestie de l'effet de complémentarité attendu des dispositifs de mesure combinés ; certes, plusieurs facteurs permettent d'expliquer cette modestie parmi lesquels les performances déjà très satisfaisantes de chaque dispositif considéré individuellement ou bien certaines faiblesses des hypothèses climatologiques. En fait, l'approche elle-même n'est pas en cause : l'analyse de la structure directe et croisée des mesures traitées est un révélateur efficace et objectif de leur cofluctuation, et le système de cokrigeage conduit de manière parfaitement rationnelle à une combinaison de ces mesures. Notre "déception" est plus liée à la confiance exagérée accordée au départ de façon purement subjective au capteur de télédétection pour compléter l'échantillonnage jugé a priori trop lâche de l'information sol ; en fait, la continuité du phénomène est telle que la densité des réseaux sol en-deça de laquelle cette complémentarité ne devient vraiment efficace est bien inférieure à celle escomptée intuitivement.

CHAPITRE VI

SIMPLIFICATIONS DU SYSTEME DE COKRIGEAGE DANS LE CADRE
DU MODELE LINEAIRERésumé

Ce chapitre examine différentes hypothèses de décomposition linéaire des variables traitées à l'aide de fonctions aléatoires orthogonales. Ces hypothèses permettent des simplifications du système de cokrigeage qu'elles rendent ainsi plus opérationnel.

Après avoir décrit quatre simplifications possibles et leur lien avec des pratiques statistiques ou géostatistiques connues, nous indiquerons les éléments de choix de ces simplifications en fonction de la nature des mesures et des réseaux disponibles afin de justifier le traitement appliqué à l'étalonnage d'images.

Ce chapitre utilise principalement la note établie par J.P.DELHOMME (1979) dans le cadre d'une étude géophysique mais aussi celles de A. MARECHAL (1970) et D. FRANCOIS-BONGARCON (1981).

PRESENTATION GENERALE

En rappelant les buts du cokrigeage au Chapitre IV, nous avons distingué deux contextes d'application de cette méthode : dans le premier, la prise en compte d'une (ou rarement plusieurs) variable complémentaire complète l'information concernant la variable d'intérêt grâce à une densité de points de mesure plus grande ; c'est le contexte dans lequel se situe cette étude ; l'autre cas de figure est celui de plusieurs variables mesurées par des réseaux de densités comparables (voir des réseaux confondus) pour lesquelles le cokrigeage permet un traitement global de l'information disponible qui respecte les relations existant entre ces variables.

Dans les deux cas l'application de la méthode se heurte à des difficultés au moment de l'inférence du modèle et lors de la résolution du système ; mais, en raison de la différence des caractéristiques de chaque contexte, les "remèdes" apportés à ces difficultés seront différents, notamment en ce qui concerne la simplification du système. Si l'on traite (i), le cas de la corégionalisation de plusieurs variables (K variables par exemple) mesurées par des réseaux de densité comparable, voir confondus (n points de mesure par exemple) l'analyse structurale nécessite l'identification de $K(K+1)/2$ fonction de structure ; cette analyse peut devenir rapidement un facteur limitant de l'application de la méthode lorsque K est important (par exemple K=15 dans le cas de données géochimiques traitées par H. WACKERNAGEL, 1985) ; cette difficulté a suscité la mise au point de démarches simplificatrices pour l'analyse structurale qui débouchent secondairement sur des simplifications du système de cokrigeage lui-même dont la taille K^2n^2 peut également être limitative. Si l'on traite (ii) le cas de la corégionalisation de peu de variables (en général K=2) mesurées par des réseaux de densités très différentes (on complète l'information d'un réseau de n points à l'aide d'un réseau de G points avec par exemple, dans notre cas des mesures de télédétection radar, $G/n=20$) il est évident que la préoccupation principale est d'obtenir une diminution de la taille du système de cokrigeage (égale à $(n+G)^2$), l'analyse structurale ne présentant pas de difficulté de calcul particulière.

Le but de ce chapitre est de présenter des simplifications de la méthode de cokrigeage applicables au cas (ii), dans lequel se range l'étalonnage d'images. Nous consacrerons pour cela dans un premier temps un paragraphe rappelant une démarche simplificatrice déjà évoquée au chapitre IV : le cokrigeage en voisinage glissant. Ensuite, nous présenterons un ensemble de trois simplifications reposant sur le modèle linéaire également introduit au Chapitre IV ; puis, nous détaillerons chacune de ces simplifications indiquant leur liaison avec des méthodes plus classiques. Nous verrons enfin la possibilité d'utiliser la variable Z_T pour la seule identification de la dérive du phénomène ; deux approches seront évoquées : celle qui suppose que les mesures Z et Z_T fluctuent de manière indépendante autour d'une dérive commune puis celle qui admet que la variable Z_T elle-même est assimilable à la dérive du phénomène ; cette deuxième approche

nous permet de revenir au simple système de krigeage. Pour terminer ce chapitre, nous tenterons de clarifier les éléments de choix d'une simplification, puis nous interpréterons dans le cadre de la géostatistique quelques pratiques courantes de l'étalonnage d'images.

COKRIGEAGE EN VOISINAGE GLISSANT

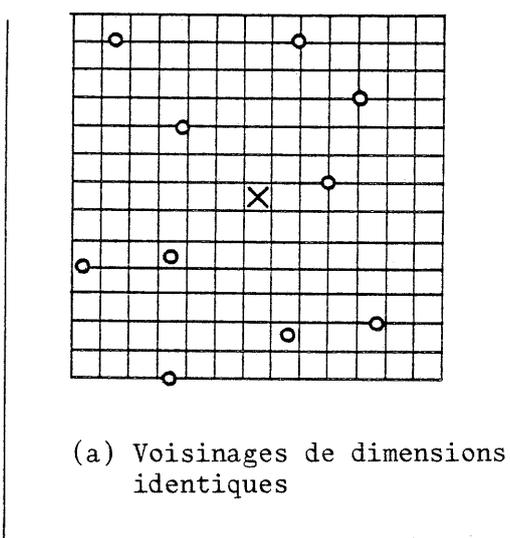
Comme dans le cas du krigeage, il est possible de limiter le nombre de points de mesure pris en considération lors d'une reconstitution en se limitant aux mesures effectuées dans un voisinage du point (ou du domaine) à reconstituer. Si l'on fait abstraction de la question de la dimension du voisinage choisi (dont l'effet n'est pas négligeable sur le résultat), il est évident que la simplification obtenue en terme de taille du système est appréciable. Malheureusement, dans le cas de l'étalonnage d'images, l'effet de la différence de densité des réseaux peut rendre cette démarche inopérante ; deux cas de figure peuvent se présenter (voir illustration schématique figure VI.1) :

a) - On choisit un même voisinage géographique pour les deux variables traitées ; ce voisinage doit contenir un minimum de points de mesures t_i du réseau le moins dense, notamment si les conditions de non-biais appliquées supposent que seule la variable sous-échantillonnée permet d'apprécier la dérive ($EZ(t) = m(t) \quad EZ_T(t)$); alors si l'on fixe par exemple à $n=10$ le nombre de points t_i retenus, le système de cokrigeage aura pour taille $(10+200)^2$ dans l'hypothèse réaliste où $G/n=20$: ce système demeure lourd pour un ordinateur classique.

b) - On choisit des voisinages géographiquement différents afin de retenir autant de points de mesure de chaque variable. Si l'on fait l'hypothèse que les dérives de Z et Z_T sont distinctes, les poids affectés à la variable Z_T seront tels que :

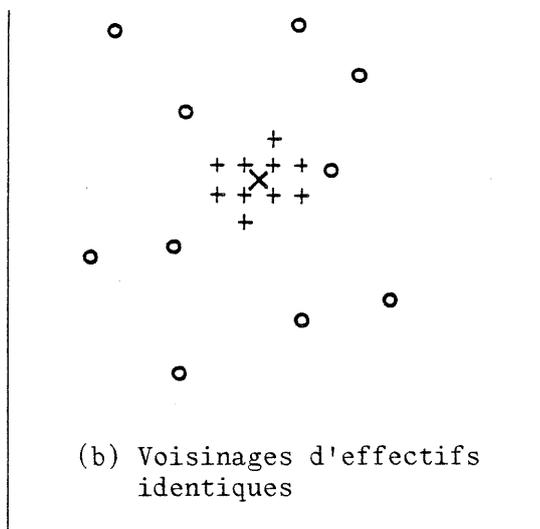
$$\sum_{g=1}^G \lambda g(t) = 0$$

sans préjuger de la forme de la dérive Z_T si l'on note m_T la moyenne spatiale des valeurs de Z_T sur le voisinage choisi, nous pouvons écrire la contribution de cette variable dans la reconstitution d'un point :



- o réseau sol t_i
- + grille de télédétection t_g
- X point à reconstituer t

Figure VI.1 : Illustration schématique des choix de voisinage glissant pour l'étalonnage d'images.



$$\sum_{g=1}^G \lambda g(t) Z_T(t) = \sum_{g=1}^G \lambda g(t) (m)_{T+Y_T(t)} = \sum_{g=1}^G \lambda g(t) Y_T(t)$$

les mesures de Z_T ne seront donc prises en considération que si elles varient sensiblement dans le voisinage et, généralement, plus ce voisinage est réduit et plus cette variabilité est faible. Par contre, si l'on fait l'hypothèse d'une dérive commune le choix de voisinages distincts ne présente pas ce travers systématique ; il convient simplement de rester prudents sur le bien-fondé de l'hypothèse de dérive commune sur des domaines de tailles différentes. Dans un contexte climatologique, cette hypothèse est raisonnable puisque nous travaillons sur des valeurs centrées donc sans dérive ; nous avons utilisé cette simplification au chapitre précédent dans le cas des mesures d'insolation. Dans un contexte purement spatial cette hypothèse est plus délicate à vérifier : en effet, la notion de dérive est alors relative à l'échelle d'échantillonnage (comme le rappelle le schéma de la figure IV.2) et il est difficile de rapprocher la dérive déduite de l'échantillonnage lâche des Z de celle perçue plus finement par Z_T . Remarquons enfin que la nature intégrée des mesures Z_T ou bien l'existence d'une relation analytique entre les variables (exemple des champs de vent déduits des champs de géopotential, T.SCHLATTER, 1974 ou P.CHAUVET, 1976) peuvent fournir une solution à ce problème d'identité de dérive en contexte spatial.

QUELQUES SIMPLIFICATIONS APORTEES PAR LE MODELE LINEAIRE

La plupart des simplifications du système de cokrigeage présentées dans la littérature reposent sur le modèle linéaire introduit au chapitre IV. A la manière d'autres méthodes que nous n'utiliserons pas dans ce travail (l'analyse en composantes principales de processus par exemple, Ch. OBLED, 1979 et 1986), ce modèle suppose que les différentes variables étudiées Z_k peuvent se décomposer linéairement à l'aide d'une famille de fonctions aléatoires orthogonales U_l , de même fonction de covariance pour un l donné suivant :

$$Z_k(t) = \sum_{l=1}^L \sum_{\alpha=1}^{A_l} a_{k\alpha}^l U_{\alpha}^l(t) \quad (\text{VI.1})$$

Comme nous l'avons vu, ce modèle permet de simplifier notablement l'analyse structurale en fournissant des règles simples de vérification des conditions de positivité ; cependant, sa formulation générale laisse entier le problème de la multiplicité des identifications de fonctions de structure (lorsque K est grand) et le problème de la lourdeur du système de cokrigeage. Il faudra donc choisir, en fonction des caractéristiques particulières de chaque cas, des fonctions U dont les propriétés conduisent à des simplifications supplémentaires.

Le paragraphe suivant présente les simplifications apportées par l'hypothèse, également évoquée au Chapitre IV, de corrélation intrinsèque qui suppose que toutes les fonctions U_l possèdent la même fonction de structure. En s'appuyant sur les travaux de A. MARECHAL (1970), nous verrons dans quelles conditions le cokrigeage peut alors être équivalent à une régression multiple.

Les deux paragraphes suivants seront consacrés à des décompositions linéaires du signal étudié en deux termes, l'un correspondant à la valeur mesurée du phénomène, l'autre à une erreur de mesure :

$$Z_k(t) = Z(t) + \varepsilon_k(t)$$

où Z est la fonction représentant le phénomène et ε_k est la fonction des erreurs de mesure commises par le capteur k, dont les valeurs sont supposées indépendantes de la valeur du phénomène ($Z \perp \varepsilon_k$). J.P.DELHOMME (1979) propose différentes utilisations de cette décomposition dont nous rappellerons qu'elles s'apparentent aux techniques classiques de krigage de données de qualité incertaine ou bien de recalage de "guess-field".

CORRELATION INTRINSEQUE ET REGRESSION

Lorsque l'on suppose que toutes les fonctions U_l ont la même structure quel que soit l, le modèle VI.1 peut s'écrire :

$$Z_k(t) = \sum_{\alpha=1}^A a_{k\alpha} U^\alpha(t) \quad (\text{VI.2})$$

et nous obtenons alors très directement la fonction de covariance croisée entre les variables Z_k et $Z_{k'}$:

$$C_{kk'}(h) = \sum_{\alpha=1}^A a_{k\alpha} a_{k'\alpha} C(h) \quad (\text{VI.3})$$

si C représente la fonction de covariance commune aux différentes fonctions U^α . Il y a donc proportionnalité des différentes fonctions de covariances directes et croisées avec la fonction C .

Si les fonctions U sont réduites (c'est à dire $C(0) = 1$, ce qui ne nuit pas à la généralité du modèle), l'équation VI.3 prend la forme :

$$C_{kk'}(h) = C_{kk'}(0) \cdot C(h) \quad (\text{VI.4})$$

ce qui signifie que les coefficients de proportionnalité sont égaux aux covariances ponctuelles directes et croisées des différents signaux. Une propriété remarquable de cette proportionnalité est son caractère intrinsèque au sens où elle n'est pas modifiée lorsqu'on s'intéresse, par exemple, à des valeurs intégrées des fonctions Z_k et $Z_{k'}$ sur des domaines D et D' :

$$\begin{aligned} C_{kk'}(D-D') &= \int_D \int_{D'} C_{kk'}(t-t') dt dt' \\ &= C_{kk'}(0) \int_D \int_{D'} C(t-t') dt dt' \\ &= C_{kk'}(0) C(D-D') \end{aligned}$$

De la même manière, la corrélation ponctuelle de ces fonctions ne dépend pas de leur éventuelle intégration ; en effet, si l'on note leur corrélation croisée :

$$R_{kk'}(t-t') = C_{kk'}(t-t') / \sqrt{C_{kkkk}(t-t') \cdot C_{k'k'}(t-t')}$$

le coefficient de corrélation $\rho_{kk'} = R_{kk'}(0)$ s'écrit après intégration :

$$\begin{aligned}
R_{kk'}(D-D) &= \iint_D R_{kk'}(t-t') dt dt' \\
&= \iint_D \frac{C_{kk'}(0)}{\sqrt{C_{kk}(0) \cdot C_{k'k'}(0)}} \cdot \frac{C(t-t')}{C(t-t')} dt dt' \\
&= \rho_{kk'} = R_{kk'}(t-t)
\end{aligned}$$

C'est dans ce sens que l'on parle de corrélation intrinsèque.

Dans le cadre de ce modèle très commode, le système de cokrigeage se simplifie de manière plus ou moins importante suivant ses conditions d'applications.

Examinons tout d'abord le cas de l'isotopie où le cokrigeage en corrélation intrinsèque se ramène tout simplement à un krigeage de la variable d'intérêt Z . En effet, si l'on prend en compte la relation VI.4 dans le système IV.7 et si l'on admet que les points t_i et t_g sont confondus, il est possible de combiner linéairement les deux premières équations du système afin de séparer l'estimation des pondérateurs de Z et de Z_T ; l'équation concernant les pondérateurs de Z_T s'obtient en multipliant les deux équations de départ respectivement par $C_{Z,Z_T}(0)$ et $-C_{ZZ}(0)$; cette combinaison transforme les covariances entre un point de mesure t_i et le point d'estimation t_0 , figurant dans le vecteur de droite du système IV.7, comme suit :

$$C_{Z,Z_T}(0) \cdot EZ(t_i) \cdot Z(t_0) - C_{ZZ}(0) \cdot EZ_T(t_i) \cdot Z(t_0)$$

soit en appliquant la relation VI.4 :

$$= (C_{Z,Z_T}(0) \cdot C_{ZZ}(0) - C_{ZZ}(0) \cdot C_{Z,Z_T}(0)) C(t_i - t_0) = 0$$

ce qui conduit à choisir des poids nuls pour les mesures de Z_T .

Une deuxième propriété importante est l'équivalence entre cokrigeage et régression lorsque le coefficient de corrélation $\rho_{kk'}$ vaut 1 quelle que soit la configuration des réseaux utilisés. En effet, il est

possible de montrer (voir la démonstration de A. MARECHAL, 1970, reprise en annexe VI.A) que le cokrigeage de la variable Z avec Z_T est équivalent au krigeage de cette variable à l'aide des valeurs de Z aux points de mesure t_i et des valeurs $E(Z(t_g)/Z_T(t_g))$, résultats de la régression Z sachant Z_T aux points t_g où seule la variable secondaire est mesurée. Comme le krigeage est un interpolateur exact, l'estimation, par cokrigeage, de la variable d'intérêt Z en un point t_g est donc égale à $E(Z(t_g)/Z_T(t_g))$, c'est à dire à son estimation par régression à l'aide de $Z_T(t_g)$. Cette simplification est particulièrement intéressante dans le cas de l'étalonnage d'images dont le but est précisément de fournir des estimations de Z aux points de mesure de l'image Z_T c'est à dire en t_g . Il est vrai que le caractère très restrictif de l'hypothèse de corrélation ρ_{kk} , égale à ± 1 n'en permet pas une application fréquente dans la pratique.

Notons que cette propriété d'équivalence cokrigeage-régression peut s'étendre lorsque $\rho_{kk} \neq \pm 1$ à condition que les interdistances entre tous les points de mesure t_i et t_g soient plus grandes que la portée de la fonction $C(h)$; ce qui revient à supposer les mesures statistiquement indépendantes, ce qui ne préjuge en rien de la dépendance entre point de mesure et point reconstitué. Cette hypothèse portant sur l'échelle d'étude du phénomène conduit en effet à un système de cokrigeage très particulier dont la solution se confond avec celle de la régression statistique ordinaire (voir A. MARECHAL, 1970, pages 15 à 17). Malheureusement cette nouvelle hypothèse n'est pas réaliste dans le cadre de l'étalonnage d'images qui présentent une densité de points de mesure t_g telle que les points contigus sont toujours très liés.

Rappelons pour terminer, que l'hypothèse de corrélation intrinsèque autorise la généralisation au cokrigeage du théorème dit de la transitivité du krigeage qui est utile lors de l'estimation de valeurs intégrées. On montre (voir A. MARECHAL, 1970, pages 19 à 26) que le cokrigeage de l'intégrale $Z(D) = \int_D Z(t) dt$ par $Z^*(D) = \sum_i \lambda^i(D) Z(t_i) + \sum_g \lambda^g(D) Z_T(t_g)$ équivaut au krigeage de cette valeur à partir des valeurs de Z en t_i et de l'estimation $Z^*(t_g)$ de Z en t_g par cokrigeage, soit :

$$Z^*(t_g) = \sum_i \lambda^i(t_g) Z(t_i) + \sum_{g'} \lambda^{g'}(t_g) Z(t_{g'})$$

Cette propriété ne conduit pas à une simplification du système puisque au cokrigage de l'intégrale sont substitués des cokrigages ponctuels d'importance identique. Seules des hypothèses complémentaires (corrélation $\rho_{kk'} = \pm 1$, interdistance des points de mesure grande par rapport à la portée) permettent de tirer parti de cette décomposition du calcul d'une intégrale.

DECOMPOSITION VALEUR VRAIE-ERREUR DE MESURE ET COKRIGEAGE D'ERREUR.

Bien qu'il ne produise pas de simplification intéressante dans le cadre de l'étalonnage d'images, il est légitime d'évoquer brièvement ce type de décomposition dans la mesure où (i) il tient une place importante dans la littérature (J. SERRA, 1967 cité par A. MARECHAL, 1970, D. FRANCOIS-BONGARCON, 1981 et J.P.DELHOMME, 1975 et 1979), (ii) il suggère des décompositions intéressantes l'étalonnage d'images, présentées par la suite et (iii) il permet, dans un cas monovarié, de traiter les mesures de différentes qualités.

Cette décomposition suppose que les mesures Z_T de la variable d'intérêt Z sont entachées d'une erreur e dont l'espérance est nulle en un point donné et dont les variations sont indépendantes de Z ; Z_T peut donc se noter :

$$Z_T(t) = Z(t) + e(t) \quad (\text{VI.5})$$

avec $Ee(t) = 0$ et $C_{ze}(h) = 0$

Les différentes fonctions de covariance directes et croisée s'expriment alors par :

$$\begin{aligned} C_{ZZ}(h) \\ C_{Z,Z_T}(h) &= C_{ZZ}(h) \\ C_{Z_T,Z_T}(h) &= C_{ZZ}(h) + C_{ee}(h) \end{aligned}$$

où l'on peut remarquer l'identité covariance directe de la variable d'intérêt-covariance croisée et la nécessité pour la différence $C_{ZZ}(h)$ -

$C_{Z_T, Z_T}(h)$ d'être un modèle de fonction de covariance acceptable pour $C_{ee}(h)$.

De plus, la condition $Ee(t)=0$ conduit à $EZ(t)=EZ_T(t)$, c'est à dire à une hypothèse d'identité des dérivées de Z et Z_T . Il est à noter que si cette identité n'est pas respectée (ce qui est souvent le cas lorsque les variables Z et Z_T ne sont pas de même nature), il n'est pas possible de la rétablir en corrigeant préalablement Z_T par régression sur Z ; cette démarche conduirait à une solution incompatible avec l'hypothèse d'orthogonalité $C_{Ze}(h)=0$.

Le modèle VI.5 paraît donc assez contraignant à inférer. De plus, bien que le système de cokrigeage obtenu soit très proche d'un système de krigeage appliqué à l'ensemble des mesures Z et Z_T , aucune simplification de ce système n'est apparemment possible.

L'application la plus intéressante de ce modèle se trouve dans le contexte monovarié sous le nom de cokrigeage d'erreur ; supposons, en effet, que la variable Z soit le résultat de la mesure de la variable Y entachée d'une erreur e :

$$Z(t) = Y(t) + e(t) \quad \text{avec} \quad Ee(t) = 0 \quad \text{et} \quad C_{eY}(h) = 0$$

Alors le krigeage de la variable Z en ignorant cette décomposition conduit à une variance d'estimation :

$$V_Z(t_0) = C_{ZZ}(0) - \sum_{i=1}^n \lambda_Z^i(t_0) C_{ZZ}(t_i - t_0) + \mu_Z$$

Le krigeage de la variable Y , seule variable intéressante, à partir de Z et en prenant en considération la décomposition orthogonale conduit à une variance d'estimation :

$$V_Y(t_0) = C_{YY}(0) - \sum_{i=1}^n \lambda_Y^i(t_0) C_{ZZ}(t_i - t_0) + \mu_Y$$

Enfin, le krigeage de l'erreur e à partir de Z donne une variance d'estimation :

$$V_e(t_0) = C_{ee}(0) - \sum_{i=1}^n \lambda^i_e(t_0) C_{ZZ}(t_i - t_0) + \mu_e$$

Si l'on remarque que le système de krigeage de Z peut se décomposer en somme des systèmes de krigeage de Y et e (en effet, la matrice $C_{ZZ}(t_i - t_j)$ étant identique pour les trois systèmes, la relation $C_{ZZ}(h) = C_{YY}(h) + C_{ee}(h)$ implique $\lambda^i_Z(t_0) = \lambda^i_Y(t_0) + \lambda^i_e(t_0)$), les variances d'estimation sont telles que :

$$V_Z(t_0) = V_Y(t_0) + V_e(t_0)$$

La connaissance de l'existence d'une erreur et sa prise en compte conduit donc à une diminution de la variance d'estimation ($V_Y(t_0) < V_Z(t_0)$). Le cas le plus simple de cokrigeage d'erreur ($C_{ee}(h) = 0$ si $h \neq 0$ et $C_{ee}(0) = C_{ee}$) est un simple krigeage avec effet de pépite. Un cas plus nuancé est obtenu en posant $C_{ee}(h) = 0$ si $h > 0$ et $C_{ee}(t-t) = C_{ee}(t)$; il s'agit du krigeage de données de qualités différentes (J.P. DELHOMME, 1975 et M. SLIMANI, 1985) qui offre la possibilité d'affecter aux différents points de mesure des variances d'erreurs différentes (voir une description rapide de ces systèmes en annexe VI.B).

RECALAGE D'UNE PRE-ESTIMATION

Une approche en quelque sorte "duale" de la précédente peut être obtenue en considérant que la variable Z_T constitue une pré-estimation (souvent appelée "guess-field") de la variable d'intérêt qu'il sera nécessaire de corriger à l'aide d'un terme de recalage r de la manière suivante :

$$Z(t) = Z_T(t) + r(t) \quad (\text{VI.6})$$

Cette décomposition de Z permet, dans le cas où le réseau $\{t_i\}$ est inclus dans $\{t_g\}$, d'écrire l'estimateur de cokrigage IV.2 comme suit:

$$\begin{aligned} z^*(t_0) &= \sum_{i=1}^n \lambda^i(t_0) Z(t_i) + \sum_{g=1}^G \lambda^g_T(t_0) Z_T(t_g) \\ &= \sum_{i=1}^n \lambda^i(t_0) Z_T(t_i) + \sum_{i=1}^n \lambda^i(t_0) r(t_i) + \sum_{g=1}^G \lambda^g_T(t_0) Z_T(t_g) \end{aligned}$$

soit en groupant les termes en Z_T :

$$z^*(t_0) = \sum_{i=1}^n \lambda^i(t_0) r(t_i) + \sum_{ig=1}^G \lambda^{ig}(t_0) Z_T(t_{ig}) \quad (\text{VI.7})$$

avec $\lambda^{ig}(t_0) = \lambda^g_T(t_0) + \lambda^i(t_0)$ avec $i = g = ig$ si $ig < n$
 T

et $\lambda^{ig}(t_0) = \lambda^g_T(t_0)$ avec $g = ig - n$ si $ig > n$
 T

Si l'on suppose que les termes Z_T et r sont indépendants ($C_{r,Z_T}(h) = 0$, ce qui entraîne les relations : $C_{ZZ}(h) = C_{Z_T,Z_T}(h) + C_{rr}(h)$ et $C_{Z,Z_T}(h) = C_{Z_T,Z_T}(h)$) l'estimation de Z est alors ramenée à un problème de cokrigage de deux variables indépendantes dont le système peut s'écrire d'après VI.7 :

$$\begin{aligned} \sum_{i=1}^n \lambda^i(t_0) C_{rr}(t_i - t_j) + \rho E r(t_j) &= C_{rZ}(t_j - t_0) \quad \text{pour } j=1 \text{ à } n \\ \sum_{g=1}^G \lambda^{ig}(t_0) C_{Z_T,Z_T}(t_{ig} - t_{jg}) + \rho E Z_T(t_{jg}) &= C_{Z,Z_T}(t_{jg} - t_0) \quad \text{pour } jg=1 \text{ à } G \quad (\text{VI.8}) \\ \sum_{i=1}^n \lambda^i(t_0) E r(t_i) + \sum_{ig=1}^G \lambda^{ig}(t_0) E Z_T(t_{ig}) &= E Z(t_0) \end{aligned}$$

avec pour variance d'estimation :

$$V_Z = C_{ZZ}(0) - \sum_{i=1}^n \lambda^i(t_0) C_{rZ}(t_i - t_0) - \sum_{ig=1}^G \lambda^{ig}(t_0) C_{Z,Z_T}(t_{ig} - t_0) + \mu EZ(t_0)$$

Notons ici que ce système VI.8 est équivalent au système de cokrigeage IV.7 dans toute sa généralité aux seules conditions que la décomposition orthogonale VI.6 soit vérifiée et que le réseau $\{t_i\}$ soit inclus dans $\{t_g\}$; aucune hypothèse n'a été formulée quant à la nature de la stationnarité des variables Z et Z_T . Sans compromettre fortement cette généralité, il est possible de modifier légèrement le système VI.8 en remarquant que l'espérance $EZ(t)$ peut se noter d'après VI.6 : $EZ(t) = EZ_T(t) + Er(t)$ et en imposant aux poids λ^i et λ^{ig} de filtrer séparément ces deux dérivées ; cela revient à remplacer la condition de non-biais "globale" contenue dans la dernière équation du système VI.8 par des conditions séparées :

$$\begin{aligned} & \sum_{i=1}^n \lambda^i(t_0) Er(t_i) = Er(t_0) \\ & \sum_{ig=1}^G \lambda^{ig}(t_0) EZ_T(t_{ig}) = EZ_T(t_0) \end{aligned}$$

ce qui entraîne dans les équations précédentes le recours à deux paramètres de Lagrange distincts, μ_r et μ_{Z_T} , au lieu d'un seul ; la variance d'estimation se trouve alors alourdie :

$$\begin{aligned} V'_Z = C_{ZZ}(0) - \sum_{i=1}^n \lambda^i(t_0) C_{rZ}(t_i - t_0) - \sum_{ig=1}^G \lambda^{ig}(t_0) C_{Z,Z_T}(t_{ig} - t_0) \\ + \mu_r Er(t_0) + \mu_{Z_T} EZ_T(t_0) \end{aligned}$$

le nombre de contraintes sur les λ^i et λ^{ig} étant supérieur dans ce dernier cas, nous aurons $V'_Z > V_Z$.

Cette hypothèse supplémentaire permet de séparer le système VI.8 en deux sous-systèmes indépendants lorsque l'on remarque que si $C_{r,Z_T}(h) = 0$ alors $C_{rZ}(h) = C_{rr}(h)$ et $C_{Z,Z_T}(h) = C_{Z_T,Z_T}(h)$; le premier sous-système permet d'estimer par krigeage Z_T au point t_0 :

$$\sum_{ig=1}^G \lambda^{ig}(t_0) C_{ZT, ZT}(t_{ig}-t_{jg}) + \mu_{ZT}^E(t_{jg}) = C_{ZT, ZT}(t_{jg}-t_0) \text{ pour } jg=j \text{ à } G$$

$$\sum_{ig=1}^G \lambda^{ig}(t_0) E_{ZT}(t_{ig}) = E_{ZT}(t_0) \quad (\text{VI.9})$$

avec pour variance d'estimation :

$$V_{ZT} = C_{ZT, ZT}(0) - \sum_{ig=1}^G \lambda^{ig}(t_0) C_{ZT, ZT}(t_{ig}-t_0) + \mu_{ZT} E_{ZT}(t_0)$$

et le second estime de la même façon le terme r de recalage :

$$\sum_{i=1}^n \lambda^i(t_0) C_{rr}(t_i-t_j) + \mu_r E_r(t_j) = C_{rr}(t_j-t_0) \text{ pour } j = 1 \text{ à } n$$

$$\sum_{i=1}^G \lambda^i(t_0) E_r(t_i) = E_r(t_0) \quad (\text{VI.9'})$$

avec pour variance d'estimation :

$$V_r = C_{rr}(0) - \sum_{i=1}^n \lambda^i(t_0) C_{rr}(t_i-t_0) + \mu_r E_r(t_0).$$

Ces variances sont telles que :

$$V'_Z = V_{ZT} + V_r \text{ puisque } C_{ZZ}(0) = C_{ZT, ZT}(0) + C_{rr}(0).$$

Du simple fait de cette séparation de l'estimation de Z en deux étapes indépendantes, il est possible d'obtenir une simplification intéressante du système de cokrigage en pratiquant les deux krigeages VI.9 et VI.9' sur des voisinages très différents, que l'on choisira bien sûr de taille compatible avec une charge de calcul raisonnable. Les craintes évoquées au premier paragraphe concernant l'utilisation de voisinages différents n'ont plus cours ici du fait de l'indépendance des variables cokrigées.

Mais il est possible d'obtenir une simplification encore plus grande dans le cas où l'estimation de Z doit se faire aux points t_g (cas

précis de l'étalonnage d'images) ; le système VI.9 devient alors trivial puisque le krigeage est un interpolateur exact $Z^*_T(t_g) = Z_T(t_g)$, et le cokrigeage se résume alors au simple krigeage des facteurs de recalage en t_g et à l'addition $Z^*(t_g) = Z_T(t_g) + r^*(t_g)$; la variance de cokrigeage vaut alors $V'_Z = V_r$. Cette approche est donc rigoureusement équivalente à la procédure de recalage proposée par A. EDDY (1964) et H.B. KRUGER (1964), pour les champs prévus de géopotential (cités par I.D. RUTERFORD, 1972) et dans le même contexte par T. SCHLATTER (1974).

Naturellement, tous les modèles de fonctions aléatoires évoqués au Chapitre I sont acceptables pour Z_T et r , ce qui permet à cette approche d'offrir une gamme très large de possibilités ; malheureusement, cette richesse complique les problèmes d'inférence : en effet, si l'on admet que Z_T et r peuvent présenter des dérivées complexes, la vérification des hypothèses d'orthogonalité et l'identification des fonctions de structure deviennent très délicates. Aussi, dans la pratique nous nous limiterons (comme J.P. DELHOMME, 1979) à une hypothèse de stationnarité d'ordre 2 des termes de recalage. Nous pourrions supposer :

i) que ces termes sont de moyenne nulle ($E(r(t)) = 0$) ; dans ce cas la condition de non-biais du système VI.9' est superflue, l'hypothèse supplémentaire de séparation des dérivées n'a plus d'objet et la variance d'estimation V'_Z égale V_Z . Bien qu'a priori peu réaliste cette hypothèse de moyenne nulle peut être justifiée par une régression préliminaire appliquée à Z_T ; l'ensemble de la démarche proposée ci-dessus peut, en effet, porter sur $Z^{\wedge}_T = E(Z/Z_T)$ au lieu de Z_T ; de plus, cette régression préliminaire est pratiquement compatible avec l'hypothèse de base d'indépendance entre Z_T et r (au contraire de ce que nous avons souligné au paragraphe précédent à propos des erreurs). Les points les plus délicats de l'application de ce pré-traitement sont en fait le choix d'une relation linéaire appropriée (une relation avec ordonnée à l'origine peut être choquante d'un point de vue physique ; une relation passant par l'origine peut conduire à modifier de manière déraisonnable la variabilité de Z_T) ou bien d'une relation non linéaire (indiquée par des considérations physiques) et l'établissement de cette relation (l'indépendance nécessaire entre les valeurs utilisées pour la régression est compromise, notamment en contexte purement spatial, par la corrélation spatiale des Z et des Z_T).

En raison de ces difficultés et compte-tenu du fait que des valeurs Z_T traitées résultent souvent déjà d'un pré-traitement plus ou moins voisin d'une régression qu'il paraît inutile de renouveler (application de la loi semi-empirique de Marshall-Palmer aux réflectivités radar, régression multiple destinée à affecter des coefficients de production à des pourcentages de nébulosité) nous considèrerons plutôt :

ii) que les termes de recalage sont de moyenne constante ($Er(t)=m_T$) ; dans ce cas le système VI.9' est un système de krigeage simple et le variogramme sera une fonction de structure mieux appropriée que la covariance.

Du fait que les estimations seront toujours réalisées en des points de mesures t_g de Z_T , nous n'évoquons pas ici les différentes possibilités de modélisation de Z_T . Remarquons cependant que les hypothèses faites ci-dessus à propos de $Er(t)$ impliquent bien évidemment l'identité des dérivées de Z et Z_T éventuellement à une constante près.

Pour terminer ce paragraphe, nous illustrerons le mode de fonctionnement de l'estimateur fourni par cette approche en nous plaçant dans le cas simple de deux fonctions aléatoires intrinsèques Z et Z_T en corrélation intrinsèque. L'estimateur VI.7 peut se noter en t_g :

$$Z^*(t_g) = Z_T(t_g) + \sum_{i=1}^n \lambda^i(t_g)(Z(t_i) - Z_T(t_i))$$

S'il y a corrélation intrinsèque les fonctions de structure de Z , Z_T et r sont homothétiques, donc les systèmes de krigeage simple de ces différentes variables fournissent des poids $\lambda^i(t_g)$ identiques ; la formule VI.7 peut donc également se noter :

$$Z^*(t_g) = Z_T(t_g) + Z^{**}(t_g) - Z^{**}_T(t_g)$$

où $Z^{**}_T(t_g)$ représente l'estimation par krigeage de Z_T en t_g à partir des seules valeurs $Z_T(t_i)$. Soit encore :

Figure VI.2 : Illustration schématique du mode de fonctionnement du recalage d'une pré-estimation dans le cas d'une corrélation intrinsèque

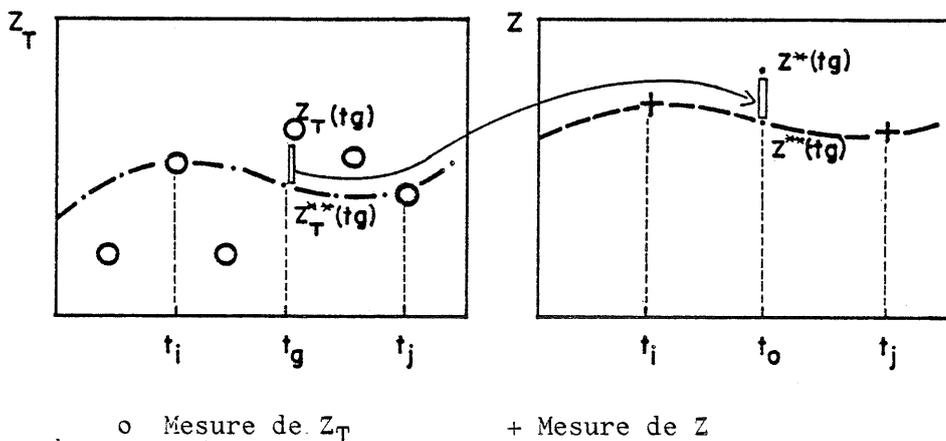
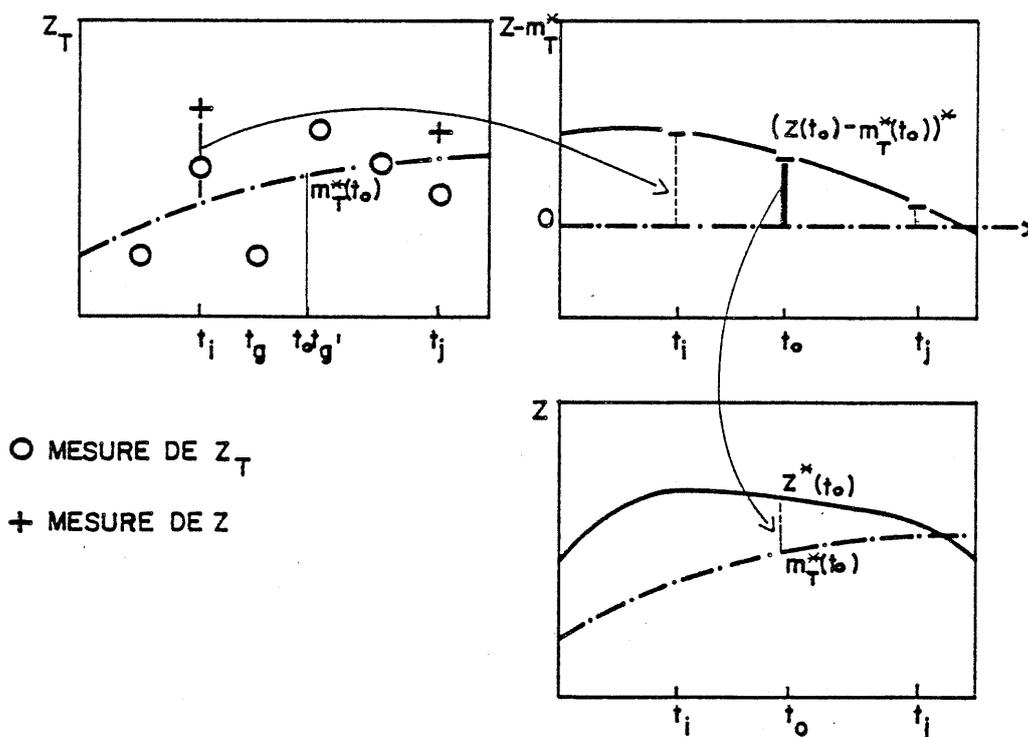


Figure VI.3 : Illustration schématique du mode de fonctionnement de l'approche "fluctuations orthogonales" suivant la première décomposition du système.



$$Z^*(t_g) = Z^{**}(t_g) + (Z_T(t_g) - Z^{**}_T(t_g))$$

Cette dernière expression montre que l'estimation Z^* par cette forme de cokrigeage simplifié s'obtient en corrigeant l'estimation par krigeage à l'aide des seules valeurs $Z(t_i)$ par l'erreur expérimentale commise sur les mesures radar considérées aux noeuds t_i d'un réseau identique ; la figure VI.2 illustre schématiquement le report de cette erreur expérimentale, accessible lorsque l'on considère les mesures de Z_T disponibles en t_i et t_g ; cette méthode utilise donc Z_T comme un "banc d'essai" de l'interpolation appliquée ensuite à Z .

FLUCTUATIONS ORTHOGONALES AUTOUR D'UNE DERIVE COMMUNE.

Toujours dans le but d'obtenir des simplifications du système de cokrigeage pour estimer Z en t_0 , il est possible de considérer que les fonctions Z et Z_T fluctuent de manière indépendante autour d'une dérive $m(t)$ commune :

$$\begin{aligned} Z(t) &= Z'(t) + m(t) \\ Z_T(t) &= Z'_T(t) + m(t) \end{aligned} \quad (\text{VI.10})$$

avec $C_{Z, Z_T}(h) = 0$

Le système de cokrigeage VI.7 se note alors :

$$\begin{aligned} \sum_{i=1}^n \lambda^i(t_0) C_{ZZ}(t_i - t_j) + m(t_j) &= C_{ZZ}(t_j - t_0) \quad \text{pour } i = 1 \text{ à } n \\ \sum_{g=1}^G \lambda^g(t_0) C_{Z_T, Z_T}(t_g - t_g') + m(t_g') &= 0 \quad \text{pour } g = 1 \text{ à } G \quad (\text{VI.11}) \\ \sum_{i=1}^n \lambda^i(t_0) m(t_i) + \sum_{g=1}^G \lambda^g(t_0) m(t_g) &= m(t_0) \end{aligned}$$

avec comme variance d'estimation :

$$V_Z = C_{ZZ}(0) - \sum_{i=1}^n \lambda^i(t_0) C_{ZZ}(t_i - t_0) + \mu \cdot m(t_0)$$

Mais, contrairement au cas du système VI.8, la décomposition en deux sous-systèmes n'est pas immédiate en raison de l'unicité de la condition de non-biais imposée par l'hypothèse de dérive commune qui par ailleurs est la justification primordiale de cette approche (en effet, si $EZ(t) = E Z_T(t)$ l'hypothèse d'orthogonalité conduit à ignorer totalement Z_T et donc à pratiquer un krigeage des Z).

J.P. DELHOMME, 1979 donne de manière très claire les deux seules décompositions possibles de VI.11 (nous ne reprendrons pas ici ses calculs mais nous en mentionnerons simplement les résultats).

La première s'obtient en "privilégiant" les G équations en C_{Z_T, Z_T} :

$$\begin{aligned} \sum_{g=1}^G \lambda^g_D(t_0) C_{Z_T, Z_T}(t_g - t_0) + \mu_D m(t_0) &= 0 \\ \sum_{g=1}^G \lambda^g_D(t_0) m(t_g) &= m(t_0) \end{aligned} \quad (\text{VI.12})$$

Ce sous-système conduit en fait à une estimation de la dérive en t_0 à partir des seules valeurs de Z_T :

$$m^*_{Z_T}(t_0) = \sum_{g=1}^G \lambda^g_D(t_0) Z_T(t_g)$$

et le sous-système complémentaire, obtenu en éliminant $\lambda^g_D(t_0)$ de VI.11, se note :

$$\begin{aligned} \sum_{i=1}^n \lambda^i_R(t_0) (C_{ZZ}(t_i - t_j) + C_{m^*_{Z_T}, m^*_{Z_T}}(t_i - t_j)) \\ = C_{ZZ}(t_j - t_0) + C_{m^*_{Z_T}, m^*_{Z_T}}(t_j - t_0) \text{ pour } j = 1 \text{ à } n \end{aligned} \quad (\text{VI.12'})$$

où C_{m^*T, m^*T} est la fonction de covariance des valeurs estimées de la dérive à partir des seules valeurs de Z_T ; ce sous-système fournit une estimation de l'écart à la dérive $m^*T(t)$; l'estimateur issu de la résolution des systèmes VI.12 et VI.12' se note :

$$Z^*(t) = m^*T(t_0) + \sum_{i=1}^n \lambda^i_R(t_0)(Z(t_i) - m^*T(t_i)) \quad (\text{VI.13})$$

La deuxième décomposition s'obtient en "privilégiant" les n premières équations de VI.11 :

$$\sum_{i=1}^n \lambda^i_U(t_0) C_{ZZ}(t_i - t_j) + \rho_U m(t_j) = C_{ZZ}(t_j - t_0) \text{ pour } j = 1 \text{ à } n \quad (\text{VI.14})$$

$$\sum_{i=1}^n \lambda^i_U(t_0) \cdot m(t_i) = m(t_0)$$

Ce sous-système conduit à une estimation de Z par krigeage universel :

$$Z^*_U(t_0) = \sum_{i=1}^n \lambda^i_U(t_0) Z(t_i)$$

et le sous-système complémentaire est obtenu en éliminant $\lambda^i(t_0)$ de VI.11 et se note :

$$\sum_{g=1}^G \lambda^g_D(t_0) (C_{Z_T, Z_T}(t_g - t_{g'}) + C_{m^*, m^*}(t_g - t_{g'})) - \rho_U m(t_{g'}) = 0 \quad (\text{VI.14'})$$

pour $g' = 1 \text{ à } G$

où C_{m^*, m^*} est la fonction de covariance des valeurs de la dérive estimées avec Z ; ce sous-système fournit une estimation de l'écart entre les dérivées calculées avec Z et Z_T ; l'estimateur fourni par VI.14 et 14' se note :

$$Z^*(t_0) = Z^*_U(t_0) + \sum_{g=1}^G \lambda^g_D(t_0) (Z_T(t_g) - m^*(t_g))$$

La figure VI.3 illustre le fonctionnement de la première décomposition qui se prête le mieux à une telle schématisation.

Quelle que soit la décomposition choisie pour le système VI.11, quatre critiques s'appliquent à cette approche :

i) bien que les hypothèses de départ soient très restrictives (orthogonalité et dérive unique), la décomposition du système qu'elles amènent est complexe sans parvenir à être totale puisque la mémorisation des covariances de dérive estimée (C_{m^*T, m^*T}) et $C_{m^*m^*}$) est nécessaire lors du passage d'un sous-système à l'autre ; cette décomposition complique donc sensiblement la démarche classique de krigeage.

ii) la subsistance de sous-système de dimension G dont la solution n'est pas évidente (même pour le cas simple de l'étalonnage où $t_0=t_g$) conduit à envisager de nouveau une réduction de la dimension de ces systèmes par le choix de voisinages réduits. Mais se repose alors la question de la compatibilité entre l'hypothèse de dérive unique et sa prise en considération à des échelles différentes ; J.P. DELHOMME, 1979, propose d'assurer cette compatibilité en prenant pour dérive sur un voisinage réduit les premiers termes du développement de Taylor de la dérive à plus grande échelle ; sans doute bien appropriée, cette démarche demande cependant de fixer le nombre de termes retenus dans le développement.

iii) les deux décompositions du système évoquées ci-dessus font apparaître le lissage très fort subi par la variable Z_T dans ce traitement qui n'est en fait utilisée que pour produire une estimation de la dérive commune ; l'intensité de ce lissage prend bien sûr en considération la variabilité spatiale de Z_T (ceci apparaît clairement dans le système VI.12) ; mais dans tous les cas de figure l'information locale apportée par Z_T est perdue ce qui est choquant dans le cadre de l'étalonnage des images dont la caractéristique la plus intéressante devrait être une appréciation fine de la variabilité du phénomène étudié.

iv) enfin, notons que l'inférence d'un tel modèle n'est pas simple dans la mesure où l'identification de la dérive commune exige un traitement simultané des deux fonctions.

Il apparaît donc que les hypothèses prises par cette approche n'atteignent que partiellement leur but de simplification du système de cokrigeage et s'adaptent assez mal aux exigences de l'étalonnage d'images.

KRIGEAGE AVEC UNE DERIVE CONNUE OU EXTERNE

Bien qu'il ne s'agisse pas d'une simplification du système de cokrigeage, nous évoquons ici cette utilisation particulière du système de krigeage en raison de ses ressemblances avec les approches présentées aux deux paragraphes précédents.

Dans sa version générale I.10 le système de krigeage fait apparaître explicitement la valeur moyenne $EZ(t)$ du phénomène aux points de mesure t_i et au point estimé t_0 . Nous avons évoqué au Chapitre I plusieurs façons d'accéder à ces valeurs moyennes soit par une connaissance climatologique du phénomène (c'est à dire par la connaissance de plusieurs réalisations de la fonction aléatoire Z), soit par une hypothèse sur la forme algébrique de la fonction $EZ(t)$.

Sous réserve que la nature de $Z_T(t)$ se prête à ce type d'assimilation certains auteurs (J.P. DELHOMME 1979, H. DARRICAU-BEUCHER, 1981) proposent d'utiliser directement cette variable complémentaire (respectivement le temps sismique pour l'étude de l'épaisseur d'une couche et la transmissivité déduite de la géologie pour l'étude de la transmissivité d'une nappe, dans les études citées ci-dessus) pour fournir une estimation de $EZ(t)$ en notant :

$$EZ(t) = m(t) = a Z_T(t) + b \quad (\text{VI.15})$$

où a et b constants quel que soit t ($m(t)$ peut alors être vue comme une espérance conditionnelle de Z).

Le système de krigeage s'écrit alors :

$$\begin{aligned} & \sum_{i=1}^n \lambda^i(t_0) C_{ZZ}(t_i, t_j) + \mu_a Z_T(t_j) + \mu_b = C_{ZZ}(t_j, t_0) \text{ pour } j = 1 \text{ à } n \\ & \sum_{i=1}^n \lambda^i(t_0) Z_T(t_i) = Z_T(t_0) \quad \text{(VI.16)} \\ & \sum_{i=1}^n \lambda^i(t_0) = 1 \end{aligned}$$

avec pour variance d'estimation :

$$V = C_{ZZ}(t_0) - \sum_{i=1}^n \lambda^i(t_0) C_{ZZ}(t_i, t_0) + \mu_a Z_T(t_0) + \mu_b$$

Si l'on décompose $Z(t) = Y(t) + aZ_T(t) + b$, l'estimateur fourni sera donc tel que :

$$\begin{aligned} Z^*(t_0) &= \sum_{i=1}^n \lambda^i(t_0) Z(t_i) \\ &= \sum_{i=1}^n \lambda^i(t_0) (Y(t_i) + aZ_T(t_i) + b) = \sum_{i=1}^n \lambda^i(t_0) Y(t_i) + aZ_T(t_0) + b \quad \text{(VI.17)} \end{aligned}$$

ce résultat est dû au fait que les poids $\lambda^i(t_0)$ interpolent exactement $Z_T(t)$ et sa forme est très proche de celui obtenu par recalage d'une pré-estimation lorsque cette pré-estimation a été corrigée par une régression préliminaire. Les résultats obtenus sont pourtant différents puisque :

i) les coefficients a et b sont d'origines très différentes : régression préliminaire les faisant apparaître explicitement dans le cas du recalage, calcul intégré au système VI.16 ne permettant pas de les expliciter dans le cas de la dérive connue ; cette dernière démarche est à cet égard plus satisfaisante puisqu'elle prend automatiquement en compte les liaisons statistiques spatiales qui faussent certainement la régression préliminaire directe.

ii) les écarts $Y(t)$ ne sont pas égaux au $r(t)$ de recalage en conséquence de la différence des coefficients de régression.

iii) les poids $\lambda^i(t_0)$ ne sont pas identiques car issus de systèmes différents VI.9' et VI.16.

Sans oublier l'avantage important que le système VI.16 apporte en réglant automatiquement le problème de la régression préliminaire nous devons noter trois inconvénients de cette approche :

i) l'inférence de la fonction C_{ZZ} est délicate car la valeur de $EZ(t)$ n'est pas accessible puisque les coefficients a et b dépendent de cette inférence et ne sont, de toutes manières, par explicités ; on peut alors penser à réaliser une inférence automatique suivant la démarche choisie pour traiter les fonctions aléatoires intrinsèques d'ordre k (voir Chapitre I) en choisissant pour fonction de base de la dérive $f^0(t)=1$ et $f^1(t)=Z_T(t)$; malheureusement cette démarche suppose que les fonctions f sont invariantes par translation, ce qui n'est généralement pas une propriété de $Z_T(t)$; H. DARRICAU-BEUCHER (1981) résout ce problème en réalisant son inférence à l'aide d'un variogramme dans une zone où $Z_T(t)$ peut être supposée constante : cette solution a peu de chance de s'appliquer à des images.

ii) la fonction Z_T doit être suffisamment lisse pour représenter de façon "plausible" une estimation de $EZ(t)$; certes il faut être très prudent sur la nécessité de représenter une dérive par une fonction lisse: il existe des exemples pour lesquels au contraire la dérive peut être cahotique (si l'on estime l'altitude de la cime des arbres d'une forêt, un excellent estimateur de la dérive, à une constante près, sera donné par la topographie du sol qui n'est pas nécessairement lisse) ; par contre dans le cas de phénomènes dont la connaissance climatologique indique une certaine régularité de la distribution spatiale des moyennes, il sera choquant d'approcher $EZ(t)$ par une fonction très irrégulière. Cette remarque peut conduire à opérer un lissage préalable de la fonction Z_T qui dans le cas de certaines images se fait au détriment de l'information fine qu'elles contiennent.

iii) enfin, cette approche exige d'une part l'inclusion du réseau $\{t_i\}$ de mesure de Z dans le réseau $\{t_g\}$ de mesure de Z_T et d'autre part le choix de points reconstitués t_0 parmi le réseau $\{t_g\}$; ce reproche, partagé avec l'approche "recalage", ne s'applique pas dans le cas d'un étalonnage d'images.

ELEMENTS DETERMINANT LE CHOIX D'UNE SIMPLIFICATION

A l'exception du cokrigeage avec erreur, les différents modèles présentés ci-dessus produisent tous des simplifications du système de cokrigeage en terme de diminution des volumes de calculs nécessaires. Mais chacun de ces modèles fait des hypothèses très fortes i) quant aux caractéristiques statistiques des mesures et ii) quant aux caractéristiques géométriques des réseaux de mesure.

Nous allons rapidement passer en revue les caractéristiques particulières que peuvent présenter des cas pratiques d'étalonnage d'images en indiquant les simplifications qu'elles suggèrent mais aussi surtout celles qu'elles interdisent.

i) Caractéristiques statistiques

L'identité des dérivées de Z et Z_T n'est exigée que par l'approche "fluctuations orthogonales", néanmoins cette propriété statistique sera fortement souhaitable dans le cas du "recalage d'une pré-estimation" afin de simplifier l'inférence de la fonction de structure des résidus.

La proportionnalité des différentes fonctions de structures directes et croisées (ou corrélation intrinsèque) ne conduit à elle seule à une simplification intéressante que dans le cas où cette corrélation est parfaite ($\rho_{kk'} = 1$, alors cokrigeage et régression sont équivalents) ou bien dans le cas où points de mesure et point estimé sont tous interdistantes de plus que la longueur de décorrélation ou bien enfin dans le cas isotopique (cokrigeage équivalent à un krigeage) ; dans les autres cas la corrélation intrinsèque facilite l'inférence des fonctions de structure mais doit posséder des propriétés complémentaires pour pouvoir entraîner des simplifications. L'égalité entre la fonction de structure

croisée et l'une des fonctions directes conditionne l'utilisation du cokrigeage avec erreur ($C_{ZZ}(h) = C_{Z,Z_T}(h)$) ou bien le "recalage de pré-estimation" ($C_{Z_T,Z_T}(h) = C_{ZZ_T}(h)$). Enfin, la nullité de la fonction de covariance croisée ($C_{Z,Z_T}(h) = 0$ pour $h \neq 0$) suggère l'approche "fluctuations orthogonales".

A côté de ces éléments caractéristiques du comportement moyen et de la variabilité du phénomène la nature plus ou moins intégrée des mesures de Z_T peut constituer également un élément de choix ; en effet, lorsque les mesures de Z_T sont le résultat d'une intégration sur une aire suffisamment importante (dont la dimension caractéristique est du même ordre de grandeur que la longueur de décorrélation du phénomène) il est légitime d'assimiler ces mesures à une estimation de la dérive de Z_T ; cette assimilation suggère, bien sûr, un krigeage en dérive connue mais aussi un "recalage de pré-estimation" qui dans ce cas fournira des résultats très proches de l'approche "fluctuations orthogonales" (puisque le système VI.12 fournira : $m^*_T(t_0) \approx Z_T(t_0)$).

ii) Caractéristiques géométriques

Notons tout d'abord qu'à l'exception des simplifications entraînées par un cas particulier de corrélation intrinsèque, toutes les autres simplifications exigent l'inclusion du réseau $\{t_i\}$ de mesure de Z dans $\{t_g\}$ celui de Z_T . De plus, ces autres méthodes ne sont effectivement simplificatrices que dans le cas d'estimations aux points t_g qui, nous l'avons déjà souligné, cadre bien avec l'étalonnage d'images.

SIGNIFICATION DE QUELQUES PRATIQUES D'ETALONNAGE D'IMAGES

Pour terminer ce chapitre, nous allons essayer de situer quelques pratiques d'étalonnage décrites dans la littérature par rapport aux différentes approches proposées ci-dessus. Les exemples que nous connaissons sont en fait peu nombreux et traitent pour la plupart d'étalonnage d'images radar ; nous ne rappellerons naturellement pas ici les exemples déjà cités dans le domaine de l'exploration géophysique (J.P.DELHOMME, 1979) ou de l'étalonnage de champs de géopotentiels (T.SCHLATTER, 1974).

La méthode d'étalonnage la plus simple et la plus fréquemment utilisée est celle de la correction uniforme de l'image par un facteur multiplicatif moyen déduit des rapports $Z(t_i)/Z_T(t_i)$ des valeurs mesurées aux n points du réseau de mesure de Z . Ce type de correction s'interprète facilement lorsque le facteur multiplicatif F est calculé suivant J.KOISTINEN et T.PAHUKKA, 1984, par :

$$F = \exp \left(\frac{1}{n} \sum_{i=1}^n \ln \frac{Z(t_i)}{Z_T(t_i)} \right)$$

en effet, la valeur corrigée en un point t_g se note alors :

$$Z^*(t_g) = F \cdot Z_T(t_g)$$

soit après transformation des variables Z et Z_T en logarithmes ($IZ(t) = \ln(t)$ et $IZ_T(t) = \ln Z_T(t)$) :

$$IZ^*(t_g) = \frac{1}{n} \sum_{i=1}^n (IZ(t_i)) + IZ_T(t_g)$$

Sous hypothèse de stationnarité des fonctions $IZ(t)$ et $IZ_T(t)$ et si n est suffisamment grand pour former un bon estimateur de l'espérance de ces fonctions au point t_g nous obtenons :

$$IZ^*(t_g) = IZ_T(t_g) + EIZ(t_g) - EIZ_T(t_g)$$

ce qui garantit le non-biais de l'estimateur IZ^* :

$$EIZ^*(t_g) = EIZ_T(t_g) + EIZ(t_g) - EIZ_T(t_g) = EIZ(t_g)$$

Cette méthode est donc une correction additive du biais des mesures Z_T sous hypothèse d'une distribution log-normale des mesures Z et Z_T et de la stationnarité des fonctions IZ et IZ_T .

La méthode d'étalonnage la plus simple et la plus fréquemment utilisée est celle de la correction uniforme de l'image par un facteur multiplicatif moyen déduit des rapports $Z(t_i)/Z_T(t_i)$ des valeurs mesurées aux n points du réseau de mesure de Z . Ce type de correction s'interprète facilement lorsque le facteur multiplicatif F est calculé suivant J.KOISTINEN et T.PAHUKKA, 1984, par :

$$F = \exp \left(- \frac{1}{n} \sum_{i=1}^n \log \frac{Z(t_i)}{Z_T(t_i)} \right)$$

en effet, la valeur corrigée en un point t_g se note alors :

$$Z^*(t_g) = F \cdot Z_T(t_g)$$

soit après transformation des variables Z et Z_T en logarithmes ($IZ(t) = \text{Ln}(t)$ et $IZ_T(t) = \text{Ln} Z_T(t)$) :

$$IZ^*(t_g) = - \frac{1}{n} \sum_{i=1}^n (IZ(t_i)) + IZ_T(t_g)$$

Sous hypothèse de stationnarité des fonctions $IZ(t)$ et $IZ_T(t)$ et si n est suffisamment grand pour former un bon estimateur de l'espérance de ces fonctions au point t_g nous obtenons :

$$IZ^*(t_g) = IZ_T(t_g) + EIZ(t_g) - EIZ_T(t_g)$$

ce qui garantit le non-biais de l'estimateur IZ^* :

$$EIZ^*(t_g) = EIZ_T(t_g) + EIZ(t_g) - EIZ_T(t_g) = EIZ(t_g)$$

Cette méthode est donc une correction additive du biais des mesures Z_T sous hypothèse d'une distribution log-normale des mesures Z et Z_T et de la stationnarité des fonctions IZ et IZ_T .

D'autres auteurs (J.W.WILSON et E.A.BRANDES, 1970 ; C.G.COLLIER et al., 1983) proposent d'utiliser un autre mode de calcul du facteur F :

$$F = \frac{1}{n} \sum_{i=1}^n Z(i) / \frac{1}{n} \sum_{i=1}^n Z_T(t_i)$$

qui ne peut plus s'interpréter à l'aide d'une transformation logarithmique mais considère plutôt que Z_T est le résultat d'une régression (non linéaire à travers la relation de Marshall-Palmer par exemple) sachant Z ; alors cette pratique apparaît comme une correction du facteur multiplicatif de la régression et nous nous plaçons de ce fait dans le cadre d'une corrélation intrinsèque parfaite (c'est à dire avec corrélation égale à 1) si nous voulons admettre que l'estimation fournie est optimale au sens des critères de la géostatistique. En d'autres mots, cette pratique d'une correction multiplicative n'a aucune chance d'être optimale si Z_T n'est pas le résultat d'une régression sachant Z ou aussi si Z et Z_T ne sont pas en corrélation intrinsèque parfaite.

Notons au passage que si l'on ne fait pas cet effort de correction, c'est à dire si l'on considère que la régression préliminaire (simple dans le cas de la transformation des réflectivités radar ou multiple dans le cas de la méthode de E.C.BARRETT, 1977) est satisfaisante pour fournir une estimation de Z on se place également dans un cas supposé de corrélation intrinsèque parfaite.

Une autre méthode d'étalonnage, plus complexe, consiste à corriger l'image par un facteur multiplicatif variant dans l'espace en interpolant les valeurs $Z(t_i)/Z_T(t_i)$ expérimentalement accessibles aux points t_i (E.A.BRANDES, 1975) ; cette démarche peut s'interpréter comme une adaptation locale de la formule de régression ayant servi à établir Z_T ; mais elle est peu crédible d'un point de vue statistique dans la mesure où cette correction s'appuie en chaque point t_i sur un seul couple de valeurs ; elle prendrait plus d'intérêt en utilisant un ensemble de réalisations des fonctions Z et Z_T qui permettrait d'estimer les rapports $EZ(t_i)/EZ_T(t_i)$ sans doute plus stables. Notons aussi que cette méthode appliquée au logarithme des rapports permet d'effectuer une correction additive des

variables IZ et IZ_T qui est à rapprocher de la méthode précédemment évoquée de "recalage d'une pré-estimation".

Enfin, nous évoquerons la méthode proposée par W.KRAJEWSKY (1986) qui recherche par cokrigeage l'estimateur :

$$Z^*_W(t_0) = \sum_{g=1}^G \lambda^g_T(t_0) Z_T(t_g) + \sum_{g=1}^G \lambda^g_K(t_0) Z^*_K(t_g) \quad (\text{VI.18})$$

où $Z^*_K(t_g)$ représente l'estimation de Z en t_g (point de mesure de Z_T) par krigeage des seules mesures de Z .

Le système de cokrigeage préconisé est le suivant :

$$\begin{aligned} \sum_{g=1}^G \lambda^g_T(t_0) C_{Z_T, Z_T}(t_g - t_{g'}) + \sum_{g=1}^G \lambda^g_K(t_0) C_{Z^*_K, Z_T}(t_g - t_{g'}) + \rho_T \\ = K_T \cdot C_{Z^*_K, Z_T}(t_{g'} - t_0) \text{ pour } g'=1 \text{ à } G \\ \sum_{g=1}^G \lambda^g_T(t_0) C_{Z^*_K, Z_T}(t_g - t_{g'}) + \sum_{g=1}^G \lambda^g_K(t_0) C_{Z^*_K, Z^*_K}(t_g - t_{g'}) + \rho \\ = K \cdot C_{Z^*_K, Z^*_K}(t_{g'} - t_0) \text{ pour } g'=1 \text{ à } G \end{aligned} \quad (\text{VI.19})$$

$$\sum_{g=1}^G \lambda^g_K(t_0) = 1$$

$$\sum_{g=1}^G \lambda^g_T(t_0) = 0$$

ce qui suppose que les covariances entre Z et Z_T d'une part et la variable reconstituée Z_W d'autre part peuvent s'écrire :

$$C_{Z^*_K, Z_W}(h) = K \cdot C_{Z^*_K, Z^*_K}(h) \text{ et } C_{Z_T, Z_W}(h) = K_T \cdot C_{Z^*_K, Z_T}(h) \quad (\text{VI.20})$$

où K et K_T représentent des constantes.

L'avantage principal de cette démarche est de permettre la réduction du système de cokrigeage complet à un voisinage très réduit dont la géométrie se répète dans le cas où le réseau t_g est un maillage régulier (cas de l'étalonnage d'image) ; ainsi, après l'étape préliminaire de krigeage des Z (qui ne met en jeu qu'un système de dimension $n+1$ dans le cas d'un krigeage simple), il sera possible de résoudre le système de cokrigeage réduit (on choisira un nombre G' de mailles t_g égal par exemple à 5 ou 9) une seule fois et d'en appliquer les pondérateurs pour étalonner l'ensemble des points de l'image (seules les mailles situées en bordure de l'image présentent des configurations de voisinage particulières et exigent une nouvelle résolution de système).

Malheureusement, cette économie de calcul est obtenue au prix d'hypothèses et de concessions très fortes non formulées clairement par W.KRAJEWSKY :

i) Le système VI.19 obtenu en appliquant des conditions de non-biais et d'optimalité à l'estimateur VI.18 contient des fonctions de covariance directe et croisée de la fonction $Z^*_K(t)$; or ces covariances peuvent s'exprimer de la manière suivante (voir formule C7 et C7' de l'annexe VI.C) :

$$C_{Z^*K, Z^*K}(t_g, t_{g'}) = \sum_{i=1}^n \sum_{j=1}^n \lambda^i(t_g) \cdot \lambda^j(t_{g'}) C_{ZZ}(t_i, t_j)$$

$$C_{Z^*K, ZT}(t_g, t_{g'}) = \sum_{i=1}^n \lambda^i(t_g) C_{Z, ZT}(t_i, t_{g'})$$

ces fonctions dépendent donc de l'emplacement des points t_g et $t_{g'}$ et ne sont donc pas stationnaires même si les fonctions C_{ZZ} et $C_{Z, ZT}$ le sont ; il sera donc très délicat d'en effectuer une modélisation, ce que W.KRAJEWSKY semble avoir perçu puisqu'il propose simplement de calculer ces covariances par :

$$C_{Z^*K, Z^*K}(t_g, t_{g'}) = \frac{1}{N(g, g')} \sum_{(g, g')} (Z^*K(t_g) - \frac{1}{G} \sum_{h=1}^G Z^*K(t_h)) \cdot (Z^*K(t_{g'}) - \frac{1}{G} \sum_{h=1}^G Z^*K(t_h))$$

et de même pour C_{Z^*K, Z^*T} en choisissant les $N(g, g')$ couples de points t_g et $t_{g'}$ formant un même vecteur $t_{gt_{g'}}$; ce faisant, il perd le bénéfice d'une modélisation de la structure et il ne règle pas le problème de la non-stationnarité pour autant.

ii) En écrivant les relations VI.20, afin de ne pas aboutir au résultat trivial $Z^*_W(t_g) = Z^*_K(t_g)$ obtenu si $K=K_T=1$, l'auteur admet que la vraie valeur Z_W du phénomène peut se décomposer en :

$$Z_W(t) = Z^*_K(t) + e(t)$$

où $e(t)$ représente l'erreur commise en krigeant ; ainsi, la covariance C_{Z_W, Z^*K} s'écrit :

$$C_{Z_W, Z^*K}(h) = C_{Z^*K, Z^*K}(h) + C_{ee}(h) - 2C_{Z^*K, e}(h)$$

et l'on aura la possibilité d'écrire :

$$C_{Z_W, Z^*K}(h) = KC_{Z^*K, Z^*K}(h)$$

dans le seul cas où Z^*_K et e sont en corrélation intrinsèque.

iii) Enfin, cette démarche, si elle conduit bien à un estimateur non-biaisé, n'est généralement pas équivalente à un cokrigeage classique (voir Annexe VI.C) ; elle est donc en général sous-optimale.

ANNEXE VI.A

*COKRIGEAGE ET REGRESSION SOUS HYPOTHESE DE CORRELATION
INTRINSEQUE AVEC $\rho_{Z,Z_T} = \pm 1$ ET DE STATIONNARITE D'ORDRE 2*

Nous reprenons dans cette annexe la démonstration de l'équivalence entre cokrigeage et régression donnée par A.MARECHAL, 1970 en essayant d'être moins rapide lors de certains passages délicats.

Cette démonstration s'appuie tout d'abord sur une simplification du système de cokrigeage grâce à l'hypothèse de corrélation intrinsèque que nous traduirons dans le cas de deux variables Z et Z_T par :

$$\begin{aligned} C_{ZZ}(h) &= C_{ZZ}(0) \cdot C(h) = C_{ZZ} \cdot C(h) \\ C_{Z,Z_T}(h) &= C_{Z,Z_T}(0) \cdot C(h) = C_{Z,Z_T} \cdot C(h) \\ \text{et} \quad C_{Z_T,Z_T}(h) &= C_{Z_T,Z_T}(0) \cdot C(h) = C_{Z_T,Z_T} \cdot C(h) \end{aligned}$$

$$\text{avec : } \rho_{Z,Z_T} = \frac{C_{Z,Z_T}^2}{C_{ZZ} C_{Z_T,Z_T}} = 1$$

Nous simplifierons l'écriture de C(t_i-t_j) par C_{ij}.

Le système de cokrigeage dans le cas de fonctions aléatoires Z et Z_T de moyennes distinctes, mesurées par des réseaux au moins partiellement distincts se note alors :

$\begin{matrix} C_{ZZ} & C_{ij} & C_{Z,Z_T} & C_{gj} & 1 & 0 \\ C_{Z,Z_T} & C_{ig'} & C_{Z_T,Z_T} & C_{gg'} & 0 & 1 \\ 1 & & 0 & & & \\ 0 & & 1 & & 0 & \end{matrix}$	$\times \begin{matrix} \lambda^i \\ \lambda^g \\ \mu_T \end{matrix}$	$\begin{matrix} C_{ZZ} & C_{jo} \\ C_{Z,Z_T} & C_{g'o} \\ 1 \\ 0 \end{matrix}$	pour j=1 à n et g'=1 à G (A1)
---	--	--	---

Malheureusement, du fait que les indices i, j et g, g' ne décrivent pas les mêmes ensembles, il est mal commode de tirer parti des particularités de la matrice de gauche dans le cadre de l'hypothèse de corrélation intrinsèque. Il est alors préférable de choisir un estimateur qui suppose que Z et Z_T sont mesurées par un même réseau de $n+G$ points :

$$z^*(t_0) = \sum_{i=1}^{n+G} \lambda^i z(t_i) + \sum_{g=1}^{n+G} \lambda^g z_T(t_g)$$

en imposant $\lambda^i = 0$ si $i > n$ et $\lambda^g = 0$ si $g < n$. L'introduction de ces poids supplémentaires et des contraintes qu'ils doivent vérifier conduit à modifier le système A1 de la manière suivante :

$C_{ZZ} \ C_{ij}$	$C_{Z, ZT} \ C_{ij}$	$\begin{array}{ c c } \hline 1 & 0 \\ \hline \hline \hline 1 & 0 \\ \hline \end{array}$	$\begin{array}{ c c } \hline & 0 \\ \hline 1 & 0 \\ \hline & \diagdown \\ \hline 0 & 1 \\ \hline \end{array}$	0	\times	$=$	λ^i	$C_{ZZ} \ C_{io}$	$(A2)$
$C_{Z, ZT} \ C_{ij}$	$C_{ZT, ZT} \ C_{ij}$	$\begin{array}{ c c } \hline 0 & 1 \\ \hline \hline \hline 0 & 1 \\ \hline \end{array}$	0	$\begin{array}{ c c } \hline 1 & 0 \\ \hline 0 & \diagdown \\ \hline & 1 \\ \hline \end{array}$			λ^g	$C_{Z, ZT} \ C_{io}$	
$1 \text{ --- } 1$	$0 \text{ --- } 0$	0					μ	1	
$0 \text{ --- } 0$	$1 \text{ --- } 1$	0					μ_T	0	
$\begin{array}{ c c } \hline 1 & 0 \\ \hline & \diagdown \\ \hline 0 & 1 \\ \hline \end{array}$	0	0					μ_i	0	
0	$\begin{array}{ c c } \hline 1 & 0 \\ \hline 0 & \diagdown \\ \hline & 1 \\ \hline \end{array}$	0			μ_g	0			

où les paramètres de Lagrange μ_i et μ_g résultent de la prise en compte des contraintes d'annulation de certains poids λ^i et λ^g .

Il est tout d'abord possible, sans changer le système, de diviser les matrices en C_{ij} et C_{i0} de la première ligne par C_{ZZ} et celles de la deuxième ligne par $C_{Z,ZT}$; les paramètres de Lagrange μ et μ_i sont alors divisés par C_{ZZ} et μ_T et μ_g par $C_{Z,ZT}$; les poids λ sont inchangés.

Une fois cette modification opérée le système n'est pas modifié si l'on remplace les lignes $(n+G+1)$ à $2(n+G)$ par leurs différences avec les lignes 1 à $n+G$; ces lignes sont identiques dans leur $2(n+G)$ premières colonnes puisque :

$$\text{d'une part } \frac{C_{ZZ}}{C_{ZZ}} C_{ij} = \frac{C_{Z,ZT}}{C_{Z,ZT}} C_{ij} \text{ et d'autre part } \frac{C_{Z,ZT}}{C_{ZZ}} C_{ij} = \frac{C_{ZT,ZT}}{C_{Z,ZT}} C_{ij}$$

$$(\text{car si l'on a } \rho_{Z,ZT} = 1, \text{ alors } \frac{C_{Z,ZT}}{C_{ZZ}} = \frac{C_{ZT,ZT}}{C_{Z,ZT}})$$

Le système A2 s'écrit alors :

$C_{ij} \frac{C_{Z,ZT}}{C_{ZZ}} C_{ij}$	$\begin{array}{c} 1 \ 0 \\ \ \\ 1 \ 0 \end{array}$	$\begin{array}{c} 0 \\ 1 \ 0 \\ \diagdown \\ 0 \ 1 \end{array}$	0		λ^i	C_{j0}		
$\begin{array}{cc} 0 & 0 \end{array}$	$\begin{array}{c} -1 \ 1 \\ \ \\ -1 \ 1 \end{array}$	$\begin{array}{c} 0 \\ -1 \ 0 \\ \diagdown \\ 0 \ -1 \end{array}$	$\begin{array}{c} 1 \ 0 \\ 0 \ 1 \\ \diagdown \\ 0 \end{array}$	0	λ^g	0		
$\begin{array}{cc} 1-1 & 0-0 \\ 0-0 & 1-1 \end{array}$	$\begin{array}{c} 0-0 \\ 1-1 \end{array}$					1	0	
$\begin{array}{c} 1 \ 0 \\ \ \\ 0 \ 1 \end{array}$	0					μ/C_{ZZ} $-\mu_T/C_{Z,ZT}$	0	
$\begin{array}{c} 1 \ 0 \\ \diagdown \\ 0 \ 1 \end{array}$	0					μ_i/C_{ZZ}	0	
0	$\begin{array}{c} 1 \ 0 \\ \diagdown \\ 0 \ 1 \end{array}$	0					$\mu_g/C_{Z,ZT}$	0

(A3)

Les lignes $n+G+1$ à $2(n+G)$ ne donnent plus que des relations entre paramètres de Lagrange :

pour $g < n$, nous avons :
$$-\frac{\mu_g}{C_{ZT,ZT}} = -\frac{\mu}{C_{ZZ}} + \frac{\mu_T}{C_{Z,ZT}} = \mu_1 = \text{Cte}$$

et

pour $i > n$, nous avons :
$$\frac{\mu_i}{C_{ZZ}} = -\frac{\mu}{C_{ZZ}} + \frac{\mu_T}{C_{Z,ZT}} = \mu_2 = \text{Cte}$$

Il est donc possible de supprimer ces $n+G$ lignes à condition de conserver en mémoire les relations liant les paramètres de Lagrange.

Les solutions de A1 sont donc solution de :

$C_{ij} \begin{matrix} C_{Z,ZT} \\ \hline C_{ZZ} \end{matrix} C_{ij}$	$\begin{matrix} 1 & 0 & 0 \\ \parallel & \parallel & \\ 1 & 0 & 1 \end{matrix}$	<div style="border: 1px solid black; width: 20px; height: 20px; background: repeating-linear-gradient(45deg, transparent, transparent 2px, black 2px, black 4px); margin: 0 auto;"></div>	λ^i	C_{j0}	
$\begin{matrix} 1 & \text{---} & 1 & & 0 & \text{---} & 0 \\ 0 & \text{---} & 0 & & 1 & \text{---} & 1 \end{matrix}$	x		λ^g	1	
$\begin{matrix} 1 & 0 \\ 0 & \diagdown & 1 \\ 0 & 1 \end{matrix}$			0	0	0
$\begin{matrix} 1 & 0 \\ 0 & \diagdown & 1 \end{matrix}$			0	μ / C_{ZZ}	0
$\begin{matrix} 0 & & & & 0 & & & & 0 \end{matrix}$			0	$\mu_T / C_{Z,ZT}$	0
			$-\mu_1$		
			<div style="border: 1px solid black; width: 20px; height: 20px; background: repeating-linear-gradient(45deg, transparent, transparent 2px, black 2px, black 4px); margin: 0 auto;"></div> μ_2		

(A4)

Les parties hachurées, correspondant à n contraintes d'annulation de λ^g , disparaissent.

Finalement, si l'on tient compte de la proportionnalité des deux matrices C_{ij} restantes, le système peut s'écrire :

$$\begin{array}{|c|c|c|} \hline & 1 & 0 \\ \hline C_{ij} & 1 & \\ & 1 & 1 \\ \hline 1 & 1 & 0 \\ \hline \end{array} \cdot \begin{array}{c} \varphi^j \\ \mu/C_{ZZ} \\ 1 \end{array} = \begin{array}{c} C_{jo} \\ 1 \\ 0 \end{array} \quad (A5)$$

avec : $\varphi^j = \lambda^j + \frac{C_{Z,ZT}}{C_{ZZ}} \lambda^g$

et $\sum \lambda^j = 1 ; \lambda^j = 0$ si $j > n$
 $\sum \lambda^g = 0 ; \lambda^g = 0$ si $g < n$

Les vecteurs de ce système peuvent être décomposés en somme de trois vecteurs de la manière suivante :

$$\begin{array}{|c|c|c|} \hline & 1 & 0 \\ \hline C_{ij} & 1 & 0 \\ & 1 & 1 \\ \hline 1 & 1 & 0 \\ \hline \end{array} \cdot \begin{pmatrix} \varphi^{jz} \\ \mu\varphi \\ 0 \end{pmatrix} + k \begin{pmatrix} \alpha^j \\ 0 \\ \mu\alpha \\ 0 \end{pmatrix} - k \begin{pmatrix} 0 \\ \beta^g \\ 0 \\ \mu\beta \end{pmatrix} = \begin{pmatrix} C_{jo} \\ 1 \end{pmatrix} + k \begin{pmatrix} 0 \\ 1 \end{pmatrix} - k \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (A6)$$

c'est à dire en posant :

$$\varphi^j = \varphi^{jz} + k \alpha^j - k \beta^g \quad (A7)$$

L'intérêt de l'écriture (A6) est de faire apparaître d'une part un système de krigeage simple, fournissant des poids φ_j^Z , comme si la variable Z était mesurée aux (n+G) points considérés ; d'autre part, deux autres systèmes de krigeage simple, portant sur les n premiers points puis les G suivants, ayant des covariances nulles au second membre : les α^j et β^g permettent donc de former des estimateurs de la moyenne de Z et Z_T .

L'équation (A7) peut être écrite suivant que j est plus ou moins grand que n en utilisant les conditions annexes à (A5) :

$$\begin{aligned} \varphi_j &= \lambda_j = \lambda_j^Z + k \cdot \alpha^j & \text{si } j < n \\ & \cdot C_{Z,Z_T} & \\ \varphi_g &= \frac{\lambda_g}{C_{ZZ}} = \varphi_g^Z - k \cdot \beta^g & \text{si } g > n \end{aligned} \quad (A8)$$

Les différentes conditions sur les sommes de poids permettent d'identifier le coefficient k :

$$\sum_{i=1}^n \lambda^i = 1 \text{ et } \sum_{i=1}^n \alpha^i = 1 \quad k = 1 - \sum_{i=1}^n \varphi^i_Z$$

ou bien encore :

$$\sum_{g=n+1}^{n+G} \beta^g = 1 \text{ et } \sum_{g=n+1}^{n+G} \varphi^g = 0 \Rightarrow k = \sum_{g=n+1}^{n+G} \varphi^g_Z$$

L'estimateur de cokrigeage :

$$Z^*(t_0) = \sum_{i=1}^n \lambda^i Z(t_i) + \sum_{g=n+1}^{n+G} \lambda^g Z_T(t_g)$$

peut donc s'écrire en tirant λ^i et λ^g de (A8) :

$$Z^*(t_0) = \sum_{i=1}^n (\varphi^i_Z + \sum_{g=n+1}^{n+G} \varphi^g_Z \alpha^i) Z(t_i) + \sum_{g=n+1}^{n+G} \frac{C_{ZZ}}{C_{Z,Z_T}} (\varphi^g_Z - \beta^g \sum_{g'=n+1}^{n+G} \varphi^{g'}_Z) Z_T(t_g)$$

soit :

$$Z^*(t_0) = \sum_{i=1}^n \varphi^i_Z Z(t_i) + \sum_{i=1}^{n+1} \sum_{g=n+1}^{n+G} \varphi^g_Z \alpha^i Z(t_i) + \sum_{g=n+1}^{n+G} \varphi^g_Z \frac{C_{ZZ}}{C_{Z,Z_T}} Z_T(t_g) - \sum_{g=n+1}^{n+G} \beta^g \frac{C_{ZZ}}{C_{Z,Z_T}} \sum_{g'=n+1}^{n+G} \varphi^{g'}_Z Z_T(t_g)$$

si l'on note $m^* = m(t_0) = \sum_{i=1}^n \alpha^i Z(t_i)$ l'estimation de la moyenne de Z et

$m^*_T = m_T(t_0) = \sum_{g=n+1}^{n+G} \beta^g Z_T(t_g)$, l'estimation de la moyenne de Z_T , nous

obtenons :

$$Z^*(t_0) = \sum_{i=1}^n \varphi^i_Z Z(t_i) + \sum_{g=n+1}^{n+G} \varphi^g_Z \left[\frac{C_{ZZ}}{C_{Z,Z_T}} (Z_T(t_g) - m^*_T) + m^* \right]$$

Nous voyons alors apparaître dans cette formule l'estimateur de Z en t_g par une régression classique :

$$EZ(t_g)/Z_T(t_g) = \frac{C_{Z,Z_T}}{C_{Z_T,Z_T}} (Z_T(t_g) - m^*_T) + m^*$$

Dans le cas de la corrélation intrinsèque avec $\rho_{Z,Z_T} = 1$ et de la stationnarité d'ordre 2, le cokrigeage est donc équivalent au krigeage simple des mesures de la variable Z aux points où elle est mesurée et des résultats de la régression de Z sachant Z_T aux points où seul Z_T est mesuré.

ANNEXE VI.B*KRIGEAGE DE DONNEES DE DIFFERENTES QUALITES*

Le résultat Z de la mesure de la variable Y est obtenu par superposition à Y d'une erreur de mesure e assimilée à un bruit blanc dont l'amplitude dépend du point de t où est effectuée la mesure :

$$Z(t) = Y(t) + e(t)$$

- avec :
1. $Ee(t) = 0$
 2. $C_{ee}(t-t') = Ee(t).e(t') = 0$ si $t \neq t'$
 3. $C_{ee}(t-t) = C_{ee}(t)$

La variable d'intérêt étant Y et les mesures disponibles $Z(t_i)$, il est logique de former l'estimateur suivant :

$$Y^*(t_0) = \sum_{i=1}^n \lambda^i(t_0) Z(t_i)$$

Pour que cet estimateur ne soit pas biaisé, nous devons avoir :

$$EY^*(t_0) = EY(t_0)$$

$$\text{avec : } EY^*(t_0) = \sum_{i=1}^n \lambda^i(t_0) EZ(t_i) = \sum_{i=1}^n \lambda^i(t_0) EY(t_i) \text{ puisque } Ee(t) = 0.$$

Ceci nous ramène aux conditions classiques de non biais qui dépendent des hypothèses faites sur $EZ(t)$; sans nuire à la généralité de ce qui suit, nous nous plaçons dans le cas simple où $EZ(t) = m$ est constante, ce qui

réduit les conditions de non biais à $\sum_{i=1}^n \lambda^i(t_0) = 1$.

Pour que cet estimateur minimise la variance d'estimation :

$$E(Y^*(t_0) - Y(t_0))^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda^i(t_0) C_{ZZ}(t_i - t_j) - 2 \sum_{i=1}^n \lambda^i(t_0) C_{YZ}(t_i - t_0) + C_{YY}(t_0 - t_0)$$

en vérifiant la condition de non-biais, les poids $\lambda^i(t_0)$ doivent satisfaire les $n+1$ équations :

$$\begin{cases} \sum_{i=1}^n \lambda^i(t_0) C_{ZZ}(t_i - t_j) + \rho = C_{YZ}(t_j - t_0) & \text{pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda^i(t_0) = 1 \end{cases}$$

ce qui se note, en rappelant que $C_{YZ}(h) = C_{YY}(h)$; $C_{ZZ}(h) = C_{YY}(h)$ pour $h \neq 0$ et $C_{ZZ}(t-t) = C_{YY}(t-t) + C_{ee}(t)$:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i(t_0) C_{YY}(t_i - t_j) + \lambda^j(t_0) C_{ee}(t_j) + \rho = C_{YY}(t_j - t_0) \text{ pour } j = 1 \text{ à } n \\ \sum_{i=1}^n \lambda^i(t_0) = 1 \end{array} \right. \quad (B1)$$

Nous obtenons ainsi un système de krigeage dont les n premiers termes de la diagonale de la matrice de gauche sont "perburbés" par la variance ponctuelle d'erreur. A l'optimum, la variance d'estimation se note :

$$E(Y^*(t_0) - Y(t_0))^2 = C_{YY}(t_0 - t_0) - \sum_{i=1}^n \lambda^i(t_0) C_{YY}(t_0 - t_i) - \rho$$

L'application de ce système exige d'une part l'identification de la fonction $C_{YY}(h)$; cela est relativement aisée pour $h = 0$ puisqu'elle se confond alors avec $C_{ZZ}(h)$ qui est expérimentalement accessible. Il est d'autre part nécessaire de connaître la variance d'erreur $C_{ee}(t_j)$ en chaque point de mesure et si l'on veut pouvoir calculer la variance d'estimation,

il faut également connaître la variance vraie du phénomène $C_{YY}(t_0-t_0)$ au point estimé.

Naturellement l'ensemble de ce développement peut être réalisé en utilisant le variogramme comme fonction de structure ; le système B1 se note alors :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i(t_0) \hat{\gamma}_{YY}(t_i-t_j) - \lambda^j(t_0) C_{ee}(t_j) - \rho = \hat{\gamma}_{YY}(t_j-t_0) \\ \sum_{i=1}^n \lambda^i(t_0) = 1 \end{array} \right. \quad (B2)$$

et la variance d'estimation :

$$E(Y^*(t_0)-Y(t_0))^2 = \sum_{i=1}^n \lambda^i(t_0) \hat{\gamma}_{YY}(t_0-t_i) - \rho$$

en utilisant la relation :

$$\gamma_{YY}(h) = C_{YY}(0) - C_{YY}(h)$$

Cette notation, ne faisant pas intervenir explicitement $C_{YY}(0)$ dans la formule de la variance d'estimation, simplifie donc le problème de l'inférence ; la fonction $\hat{\gamma}_{YY}(h)$ est par ailleurs égale à $\hat{\gamma}_{ZZ}(h)$ lorsque $h \neq 0$, donc expérimentalement accessible.

Remarquons alors que, si $\sum_{i=1}^n \lambda^i(t_0) = 1$, il est possible de remplacer $-\lambda^j(t_0) C_{ee}(t_j)$ dans B2 par $C_{ee}(t_j) (\sum_{i \neq j} \lambda^i(t_0) - 1)$ ou encore $\sum_{i=1}^n \lambda^i(t_0) (1 - \delta_{ij}) C_{ee}(t_j) - C_{ee}(t_j)$; soit en confondant les sommes de la première équation de (B2) :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda^i(t_0) (\hat{\gamma}_{YY}(t_i-t_j) + (1 - \delta_{ij}) C_{ee}(t_j)) - \rho = \hat{\gamma}_{YY}(t_j-t_0) + C_{ee}(t_j) \\ \sum_{i=1}^n \lambda^i(t_0) = 1 \end{array} \right. \quad (B3)$$

Ce qui nous ramène, lorsque $C_{ee}(t) = C_{ee}$ pour tout t , à un système de krigeage dont le variogramme se note :

$$\hat{\gamma}(h) = (1 - \delta_h) C_{ee} + \hat{\gamma}_{YY}(h)$$

c'est à dire à un krigeage avec effet de pépite.

Pour terminer, rappelons que l'estimateur $Y^*(t_0)$ peut être conditionné sous forme d'interpolateur suivant la démarche déjà évoquée aux Chapitres I et IV ; en effet, le système B1 peut se noter en utilisant les mêmes notations matricielles qu'au Chapitre I :

$$\begin{bmatrix} S+dS & D \\ D^T & 0 \end{bmatrix} \times \begin{bmatrix} \wedge \\ \mu \end{bmatrix} = \begin{bmatrix} S_0 \\ D_0 \end{bmatrix}$$

avec ici $S+dS = |C_{YY}(t_i-t_j) + \delta_{ij}C_{ee}(t_j)|^{n \times n}$
 et $S^T_0 = |C_Y(t_i-t_0)|^{1 \times n}$

et :

$$Y^*(t_0) = \begin{bmatrix} Z \\ 0 \end{bmatrix}^T \cdot \left(\begin{bmatrix} S+dS & D \\ D^T & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} S_0 \\ D_0 \end{bmatrix} \right)$$

$$= \left(\begin{bmatrix} Z \\ 0 \end{bmatrix}^T \cdot \begin{bmatrix} S+dS & D \\ D^T & 0 \end{bmatrix}^{-1} \right) \cdot \begin{bmatrix} S_0 \\ D_0 \end{bmatrix}$$

$$= \begin{bmatrix} A \\ b \end{bmatrix}^T \begin{bmatrix} S_0 \\ D_0 \end{bmatrix}$$

ce qui permet de calculer $Y^*(t_0)$ par :

$$Y^*(t_0) = \sum_{i=1}^n a_i C_{YY}(t_i - t_0) + bEZ(t_0)$$

(en redonnant une notation générale à $EZ(t)$ qui sous-entend un choix approprié de la fonction C_{YY}).

Les facteurs a et b sont alors solutions du système :

$$\begin{cases} Y^*(t_j) = a_j C_{ee}^j = Z(t_j) & \text{pour } j = 1 \text{ à } n \\ a_j EZ(t_j) = 0 \end{cases}$$

qui est identique, lorsque $C_{YY}(h) = h^2 \text{Log } h$ et $EZ(t)$ linéaire en t , au système définissant une fonction spline de lissage de type plaque mince (voir O. DUBRULE, 1984).

ANNEXE VI.C

COKRIGEAGE DE VALEURS PRE-KRIGÉES :
UNE APPROCHE GÉNÉRALEMENT SOUS-OPTIMALE

Soient Z et Z_T deux fonctions aléatoires intrinsèques d'ordre 0 mesurées respectivement par les réseaux $(t_i, i=1 \text{ à } n)$ et $(t_g, g = 1 \text{ à } G)$.

Nous souhaitons former un estimateur Z^* de Z en un point t_0 de la forme :

$$Z^*(t_0) = \sum_g \lambda_K^g(t_0) Z_K(t_g) + \sum_g \lambda_T^g(t_0) Z_T(t_g) \quad (C1)$$

où $Z_K(t_g)$ représente une estimation de Z en t_g par krigeage des Z ce qui peut se noter :

$$Z_K(t_g) = \sum_i \lambda^i(t_g) Z(t_i) \quad (C2)$$

L'estimateur Z^* sera non-biaisé si :

$$E(Z^*(t_0) - Z(t_0)) = \sum_g \lambda_K^g(t_0) E Z_K(t_g) + \sum_g \lambda_T^g(t_0) E Z_T(t_g) = 0 \quad (C3)$$

Si l'on suppose par exemple que $EZ(t) = m$ et $EZ_T(t) = m_T$ avec $m \neq m_T$, le non-biais de l'estimateur (C1) sera acquis si :

$$\sum_g \lambda_K^g(t_0) = 1 \quad \text{et} \quad \sum_g \lambda_T^g(t_0) = 0 \quad (C4)$$

Nous retrouvons ici les conditions classiques de non-biais d'un estimateur par cokrigeage ; la formule (C1) fournit un estimateur sans biais si les λ^g_K et λ^g_T d'une part et les λ^i d'autre part vérifient respectivement les conditions de non-biais contenues respectivement dans le système de cokrigeage des Z^* et de krigeage des Z_K : l'estimateur (C1) obtenu par cokrigeage de valeurs pré-krigées et donc sans biais.

Par contre, il est possible de montrer que cet estimateur n'est pas optimal en général ; pour ce faire, nous nous placerons dans le cas simple de fonctions Z et Z_T de moyenne nulle ($m = m_T = 0$) pour lequel les conditions de non-biais sont toujours satisfaites.

Dans ce cas, l'estimateur Z^* sera optimal (et sans biais, donc équivalent à un estimateur de cokrigeage classique) si :

$$V = E(Z^*(t_0) - Z(t_0))^2 \text{ minimale} \quad (C5).$$

Or, cette variance d'estimation V peut s'écrire :

$$\begin{aligned} V &= E\left(\sum_g \lambda^g_K(t_0) \cdot \sum_i \lambda^i(t_g) Z(t_i) + \sum_g \lambda^g_T(t_0) Z_T(t_g) - Z(t_0)\right)^2 \\ &= \sum_{gg'ij} \lambda^g_K(t_0) \lambda^{g'}_K(t_0) \lambda^i(t_g) \lambda^j(t_{g'}) C_{ZZ}(t_i, t_j) \\ &\quad + \sum_{gg'} \lambda^g_T(t_0) \lambda^{g'}_T(t_0) C_{Z_T, Z_T}(t_g, t_{g'}) + C_{ZZ}(t_0, t_0) \\ &\quad + 2 \sum_{gg'i} \lambda^g_K(t_0) \lambda^i(t_g) \lambda^{g'}_T(t_0) C_{Z, Z_T}(t_i, t_{g'}) \\ &\quad - 2 \sum_{gi} \lambda^g_K(t_0) \lambda^i(t_g) C_{ZZ}(t_i, t_0) - 2 \sum_g \lambda^g_T(t_0) C_{Z_T, Z}(t_g, t_0) \end{aligned}$$

Cette variance sera minimale lorsque ses dérivées par rapport aux différents paramètres s'annulent c'est à dire lorsque le système suivant est vérifié :

$$2 \sum_{g'j} \lambda^{g'K}(t_0) \lambda^i(t_g) \lambda^j(t_{g'}) C_{ZZ}(t_i, t_j) + 2 \sum_{g'i} \lambda^i(t_g) \lambda^{g'}(t_0) C_{Z, ZT}(t_i, t_{g'})$$

$$- 2 \sum_i \lambda^i(t_g) C_Z(t_i, t_0) = 0 \text{ pour } g = 1 \text{ à } G$$

$$2 \sum_g \lambda^{gT}(t_0) C_{ZT, ZT}(t_g, t_{g'}) + 2 \sum_{gi} \lambda^{gK}(t_0) \lambda^i(t_g) C_{Z, ZT}(t_i, t_{g'}, (t_i, t_{g'}))$$

$$- 2 C_{ZT, Z}(t_{g'}, t_0) = 0 \text{ pour } g' = 1 \text{ à } G \quad (C6)$$

$$2 \sum_{gg'j} \lambda^{gK}(t_0) \lambda^{g'K}(t_0) \lambda^j(t_{g'}) C_{ZZ}(t_i, t_j) + 2 \sum_{gg'} \lambda^{gK}(t_0) \lambda^{g'T}(t_0) C_{Z, ZT}(t_i, t_{g'})$$

$$- 2 \sum \lambda^{gK}(t_0) C_{ZZ}(t_i, t_0) = 0 \text{ pour } i = 1 \text{ à } n$$

Ce système peut se simplifier d'une part en prenant en compte les relations satisfaites par les facteurs $\lambda^i(t_g)$ qui vérifient le système de krigeage des Z_K :

$$\sum_i \lambda^i(t_g) C_{ZZ}(t_i, t_j) = C_{ZZ}(t_j, t_g) \text{ pour } j = 1 \text{ à } n \text{ et } g = 1 \text{ à } G$$

et d'autre part en remarquant que la covariance des valeurs Z_K s'exprime à l'aide de la covariance des Z :

$$C_{ZK, ZK}(t_g, t_{g'}) = E(Z_K(t_g) - EZ_K(t_g)) \cdot (Z_K(t_{g'}) - EZ_K(t_{g'}))$$

$$= E\left(\sum_i \lambda^i(t_g) (Z(t_i) - EZ(t_i))\right) \cdot \left(\sum_j \lambda^j(t_{g'}) (Z(t_j) - EZ(t_j))\right)$$

$$= \sum_{ij} \lambda^i(t_g) \cdot \lambda^j(t_{g'}) C_{ZZ}(t_i, t_j) \quad (C7)$$

et de même :

$$C_{ZK, ZK}(t_g, t_{g'}) = \sum_i \lambda^i(t_g) C_{Z, ZT}(t_i, t_{g'}) \quad (C7')$$

et
$$C_{ZK, Z}(t_g, t_0) = \sum_i \lambda^i(t_g) C_{ZZ}(t_0, t_i)$$

Le système (C6) se note alors :

$$\sum_{g'} \lambda^{g'}_K(t_0) C_{ZK, ZK}(t_g, t_{g'}) + \sum_{g'} \lambda^{g'}_T(t_0) C_{ZK, ZT}(t_g, t_{g'}) = C_{ZK, Z}(t_g, t_0) \quad \text{pour } g = 1 \text{ à } G$$

$$\sum_g \lambda^{g'}_T(t_0) C_{ZT, ZT}(t_g, t_{g'}) + \sum_g \lambda^{g'}_K(t_0) C_{ZK, ZT}(t_g, t_{g'}) = C_{ZT, Z}(t_{g'}, t_0) \quad \text{pour } g' = 1 \text{ à } G \quad (C8)$$

$$\sum_g \lambda^{g'}_K(t_0) \cdot \left(\sum_{g'} \lambda^{g'}_K(t_0) C_{ZZ}(t_i, t_{g'}) + \sum_{g'} \lambda^{g'}_T(t_0) C_{Z, ZT}(t_i, t_{g'}) - C_{ZZ}(t_i, t_0) \right) = 0 \quad \text{pour } i = 1 \text{ à } n$$

Les deux premières équations de (C8) sont bien les équations de cokrigage de l'estimateur Z^* ; mais la troisième équation de (C8) impose aux coefficients λ_K et λ_T de vérifier :

$$\sum_{g'} \lambda^{g'}_K(t_0) C_{ZZ}(t_i, t_{g'}) + \sum_{g'} \lambda^{g'}_T(t_0) C_{Z, ZT}(t_i, t_{g'}) = C_{ZZ}(t_i, t_0) \quad \text{pour } i=1 \text{ à } n$$

ce qui ne sera effectif que dans le cas où les réseaux (t_i) et (t_g) sont confondus (car nous aurons alors $Z_K(t_g) = Z(t_g)$ et cette troisième égalité sera équivalente à la première).

Il apparaît donc que le cokrigage de valeurs pré-krigées n'est pas équivalent en général au cokrigage classique.

CHAPITRE VII

EXEMPLES D'APPLICATION DE SIMPLIFICATIONS DU SYSTEME DE
COKRIGEAGE. MESURE DE LA PLUVIOMETRIE DANS LE BASSIN PARISIEN
ET DE LA FRACTION JOURNALIERE D'INSOLATION EN FRANCE.

Résumé

Ce chapitre propose deux exemples d'application de simplifications du système de cokrigeage.

Le premier exemple concerne la pluviométrie dans le bassin parisien: les images radar seront considérées comme une pré-estimation du phénomène que les mesures sol permettront de recalculer.

Le deuxième exemple concerne les mesures d'insolation par la méthode CACTUS dont la forte co-fluctuation spatiale avec les mesures héliographiques permet d'envisager une simple correction par régression.

1 - MESURE DE LA PLUVIOMETRIE DANS LE BASSIN PARISIEN :

Recalage des mesures radar considérées comme pré-estimations de la pluviométrie.

a - Généralités

Le choix de traiter nos mesures pluviométriques radar et sol de la région parisienne à l'aide d'un "recalage de pré-estimation", nécessite, bien sûr, d'être justifié par une analyse détaillée des caractéristiques statistiques de ces mesures que nous présentons au paragraphe suivant. Il s'agira d'une justification bien-fondée du modèle proposé qui montrera souvent un écart certain entre le modèle et la réalité.

Mais avant d'entamer cette démarche, il est bon de rappeler qu'un tel choix repose également sur des considérations pratiques. En effet, notre but d'hydrologues est de fournir une estimation de la pluviométrie en

combinant ces deux types de mesure de manière simple ; cette simplicité s'entend à la fois à propos de la mise en oeuvre, qui doit être aussi automatique que possible, et à propos de la puissance de calcul nécessaire, qui doit être compatible avec les moyens informatiques disponibles en contexte opérationnel. Il faut bien entendu être prudent à propos de ce dernier point en raison de l'évolution rapide des appareils de calcul, mais les sites de traitement opérationnels de mesure radar actuellement les mieux équipés (Observatoire radar de Saint-Anne de Bellevue au Québec ou bien d'Hameldon Hill en Grande-Bretagne par exemple) sont dotés de mini-ordinateurs dont la disponibilité pour des tâches d'étalonnage est réduite.

Dans ce contexte, le recours au cokrigeage paraît a priori très difficile en raison de la nécessité de réaliser une analyse de structure assez délicate pour chaque champ et de l'importance du volume de calculs nécessaire ; seule une bonne connaissance climatologique pourrait conduire à un traitement automatique rapide : malheureusement, nous ne disposons pas (au contraire du cas de l'insolation vu au Chapitre V) d'une série de mesures suffisamment longue pour dire si une telle solution est possible.

Parmi les simplifications évoquées au Chapitre VI, celle qui allège le plus le système de cokrigeage dans le cadre de l'étalonnage d'images est le recalage des mesures radar considérées comme pré-estimations. Cette simplification procure la même qualité d'estimation que le cokrigeage lorsque les hypothèses formulées sont vérifiées ; dans le cas contraire, les estimations obtenues seront sous-optimales d'un point de vue statistique, mais ce mode de correction demeure néanmoins fondé si l'on prend en compte l'organisation spatiale des termes de recalage (cette pratique est courante en météorologie sous l'appellation de "recalage de guess-field").

Se pose alors la question de savoir sur quels termes de recalage travailler. Nous avons dit, Chapitre VI, qu'il est possible de considérer directement les écarts entre valeur mesurée au sol et valeur brute de la mesure radar ou bien de considérer ces écarts après une première correction (additive, multiplicative ou plus généralement linéaire) de la mesure radar ; nous avons également souligné les défauts respectifs de ces corrections préalables (la correction additive conduit à des résultats

physiquement choquant lorsqu'on transforme des mesures de pluie nulles en mesures positives ou négatives, la correction multiplicative amplifie artificiellement la variance des mesures). Dans notre cas, G. DELRIEU, 1986 (pages 153 à 173) a montré de plus la sensibilité de ce type de correction à la composition du réseau d'étalonnage et l'effet négatif de la correction multiplicative sur la structure des écarts. Enfin, nous pouvons remarquer qu'une correction additive pure n'apporte rien puisque la fonction de structure utilisée (variogramme) et le système de krigeage appliqué pour interpoler ces écarts sont insensibles à la valeur moyenne pourvu qu'elle soit localement constante. Nous avons donc choisi de ne pas effectuer de correction préalable des mesures radar.

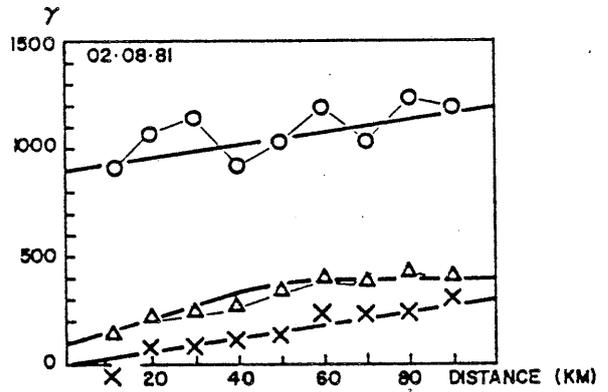
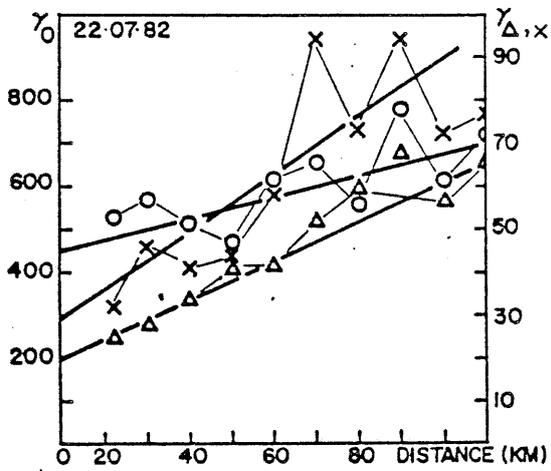
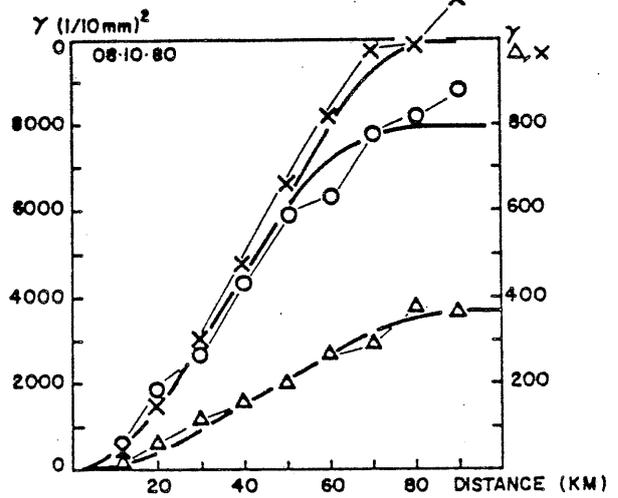
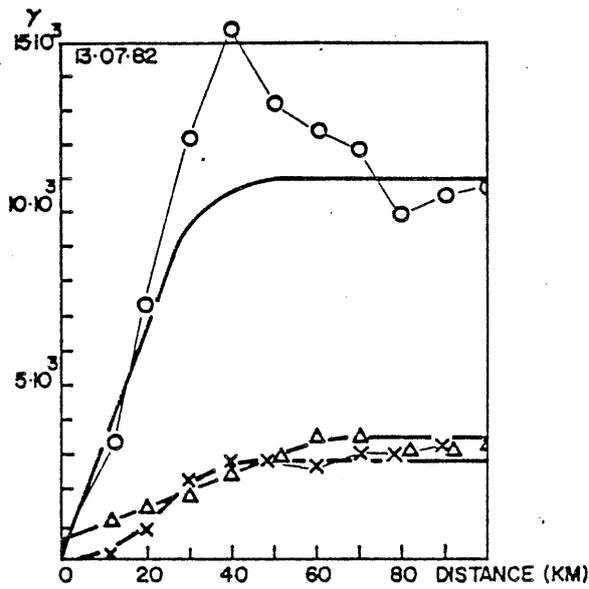
b - Analyse de structure

Cette analyse pourrait se limiter à la structure des écarts entre les mesures sol et radar si l'on considère les seuls besoins pratiques de l'approche retenue de recalage d'une pré-estimation radar. Nous consacrerons cependant le début de ce paragraphe à l'analyse de la structure directe des mesures radar puis de la structure croisée entre mesures sol et radar ; cela nous permettra d'acquérir une idée plus synthétique de la cofluctuation de ces mesures que par le simple examen des cartes de la figure III.11 ; de plus, nous pourrions apprécier le bien-fondé des hypothèses d'orthogonalité des résidus formulées par ce type de recalage.

Analyse de la structure statistique des mesures radar.

De manière très parallèle à l'analyse des mesures sol présentée en III.2, les mesures radar ont été traitées événement par événement pour les 11 journées disponibles dans le cadre d'une hypothèse de quasi-stationnarité. Les variogrammes présentés (tableau VII.1 pour les différentes journées disponibles et figure VII.1 pour quatre journées représentatives) ont été calculés à l'aide de 98 mesures correspondant à des pixels radar contenant les stations de mesure sol ; ce choix nous paraît raisonnable, car il permet d'obtenir des fonctions expérimentales suffisamment stables, très proches des fonctions obtenues à l'aide de l'ensemble des mesures. Les modèles théoriques ont été ajustés

Figure VI.1 : Variogrammes directs et croisés des mesures radar et sol en 98 points pour 4 journées représentatives.



- △ Variogramme expérimental des mesures radar
- Variogramme expérimental des mesures sol
- × Variogramme croisé expérimental

empiriquement en prenant des précautions identiques à celles évoquées au paragraphe III.2.

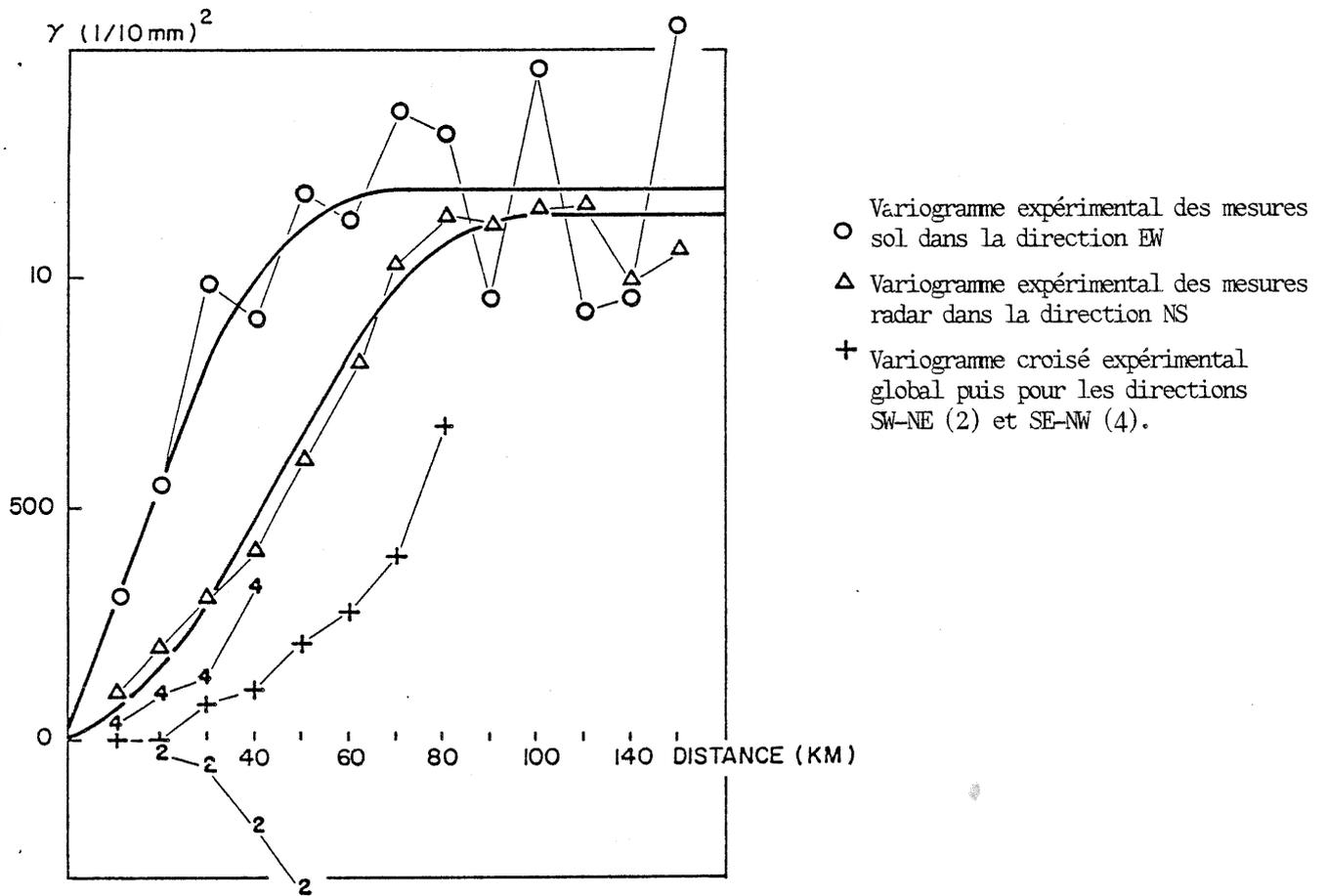
La comparaison des modèles ainsi ajustés avec ceux obtenus avec les mesures sol fait ressortir les points suivants :

* pour six journées (7/10/80, 8/10/80, 15/12/80, 13/7/82 et peut être 23/6/82 et 22/7/82) les deux types de mesures possèdent des modèles de structure identiques ; lorsque ces modèles présentent une portée, celle du sol est toujours légèrement inférieure à celle du radar, ce qui va dans le sens attendu en raison de l'intégration spatiale effectuée par le radar ; il serait imprudent de faire le même type de remarque à propos des portées ou des pentes : la variabilité des mesures radar est toujours moins forte mais l'intégration n'explique cela que de manière certainement très marginale par rapport aux problèmes déjà cités de sous-détection très forte notamment en régime frontal. La réserve émise à propos des deux dernières journées vient de la difficulté d'apprécier la structure très cahotique de ces événements convectifs.

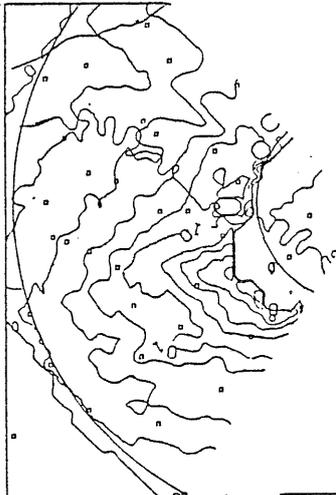
* pour trois autres journées (10/10/80, 14/12/80 et 22/10/82) chaque type de mesure présente une bonne cohérence spatiale mais les modèles ajustés sont distincts. Il semble que pour ces événements les défauts de la mesure radar sont suffisamment forts et spatialement organisés pour perturber sa perception du phénomène. Le 10/10/80 constitue un bon exemple illustré (figure VII.2) ; il est très net, d'après les cartes isohyètes, que le radar a sous-déTECTÉ, de manière croissante en fonction de l'éloignement du radar, un champ pluvieux à la répartition spatiale assez homogène. Ce défaut a tout d'abord modifié la dérive du champ qui présente un gradient grossièrement N-S d'après les mesures sol et E.W d'après le radar : ceci nous a conduit à retenir les variogrammes expérimentaux suivant des azimuts différents pour chaque type de mesure. De plus cette sous-détection n'a pas permis au radar de percevoir d'importantes cellules pluvieuses aux grandes distances, le privant ainsi d'une appréciation fidèle de la variabilité du phénomène.

* enfin pour les deux journées restantes (2/8/81 et 15/7/82), la distribution spatiale cahotique de la pluviométrie liée à des situations

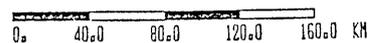
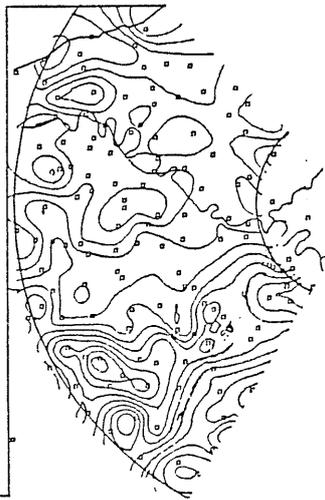
Figures VII.2 : Variogrammes directs et croisé des mesures radar et sol en 98 points pour la journée du 10.10.80 ; cartes isohyètes correspondantes.



MESURES RADAR



MESURES SOL



convectives rend les variogrammes difficilement utilisables (présence de nombreuses valeurs nulles et de forts gradients); tous les variogrammes obtenus présentent un fort effet de pépite, particulièrement sensible pour les mesures sol.

Analyse de la structure croisée des mesures radar et sol.

Nous ne reprendrons pas en détail l'analyse préliminaire de la symétrie des fonctions de covariance croisée. En effet, les décalages géométriques que la disymétrie de ces fonctions indiquent ne trouvent pas d'explication liée à la physique des mesures puisqu'elles varient d'un point à l'autre ; ils n'ont pas non plus d'explication météorologique simple car leur amplitude (souvent supérieure à 5 km) exclut l'hypothèse d'une advection latérale par le vent. En fait, le mode de construction de ces fonctions dans un contexte spatial pur rend ces fonctions très sensibles aux tendances à grande échelle du phénomène qui, même aux faibles interdistances, ne sont pas filtrées ; les décalages géométriques identifiés s'expliquent donc généralement par l'effet d'un maximum local et ne concernent pas l'ensemble du domaine étudié. Nous ferons donc une hypothèse d'absence de décalage en raison de l'impossibilité de les identifier de manière sûre et avec la conviction "physique" de leur faible amplitude.

Les variogrammes croisés présentés (tableau VII.1 et figure VII.1) ont été calculés à l'aide de 98 points de mesure communs aux deux dispositifs et modélisés empiriquement. Lorsque ces variogrammes croisés présentent une évolution nettement positive (valeurs du variogramme global nettement différentes de zéro au-delà de 10 km et variogrammes par azimuth tous positifs) nous avons ajusté des modèles théoriques en prenant les précautions habituelles (voir paragraphe III.2) ; dans le cas contraire (exemple du 10/10/80 donné figure VII.2) nous avons considéré que le variogramme croisé est nul à toute distance, c'est à dire que les signaux radar et sol sont indépendants.

Tableau VII.1 : Détails des résultats de la reconnaissance des structures directes et croisées des mesures sol et radar obtenus pour 11 journées et 98 points de mesures.

Date de l'événement	Variogrammes des mesures sol	Variogrammes des mesures radar brutes	Variogrammes croisés radar/sol	Variogrammes des écarts	Variogrammes croisés écarts*/radar
7.10.80	S 1800 50 0	S 80 60 0	S 150 60 0	S 1600 60 0	P 0
8.10.80	S 8000 80 0	S 380 100 0	S 1000 80 0	S 7000 90 0	S -800(P(2) 0) 80 0
10.10.80	S(1)1200 60 0	S(3) 1200 100 0	P 0	S 3200 100 0	S 1000 100 0
14.12.80	S(1)500 80 0	S(3) 300 150 0	P(1) 0	S 650 100 0	P 0
15.12.80	L 12 (S(1)600 0 70 0)	L 1.8 (S(4) 120 0 90 0)	L 2.5 (S 200 0 20 0)	L 10 (S(2) 400 0 60 0)	P 0
22.10.80	L 50 0	S 1250 100 0	P(3) 0	L 60 0	S 300 30 0
2.8.81	L 3 900	S 400 70 100	L 3 0	P 1100	P 200
23.6.82	L 6 300	L 0.36 (S(1) 50 25 100 0)	L .9 25	P 500	P 0
13.7.82	S 11000 50 0	S 3500 70 500	S 2800 50 0	S 10000 50 0	P 0
15.7.82	L 2.4 600	S 70 60 10	S 100 50 20	P 600	S -40 50 0
22.7.82	L 2.5 450	L .45 20	L .65 30	P 500	P 0

Unités : Pluviométrie en 1/10 mm ; Distances en km

Description des modèles utilisés : S, modèle à portée : palier, portée, pépite
L, modèle linéaire : pente, pépite
P, modèle purement pépitique : pépite

* Différence entre la mesure radar corrigée à l'aide d'un facteur multiplicatif constant et la mesure sol intégrée sur le pixel radar correspondant.

Les modèles de structure croisée ainsi obtenus pour les onze journées disponibles permettent de distinguer trois cas de figure qui rejoignent les résultats obtenus par analyse de la structure directe des deux types de mesure :

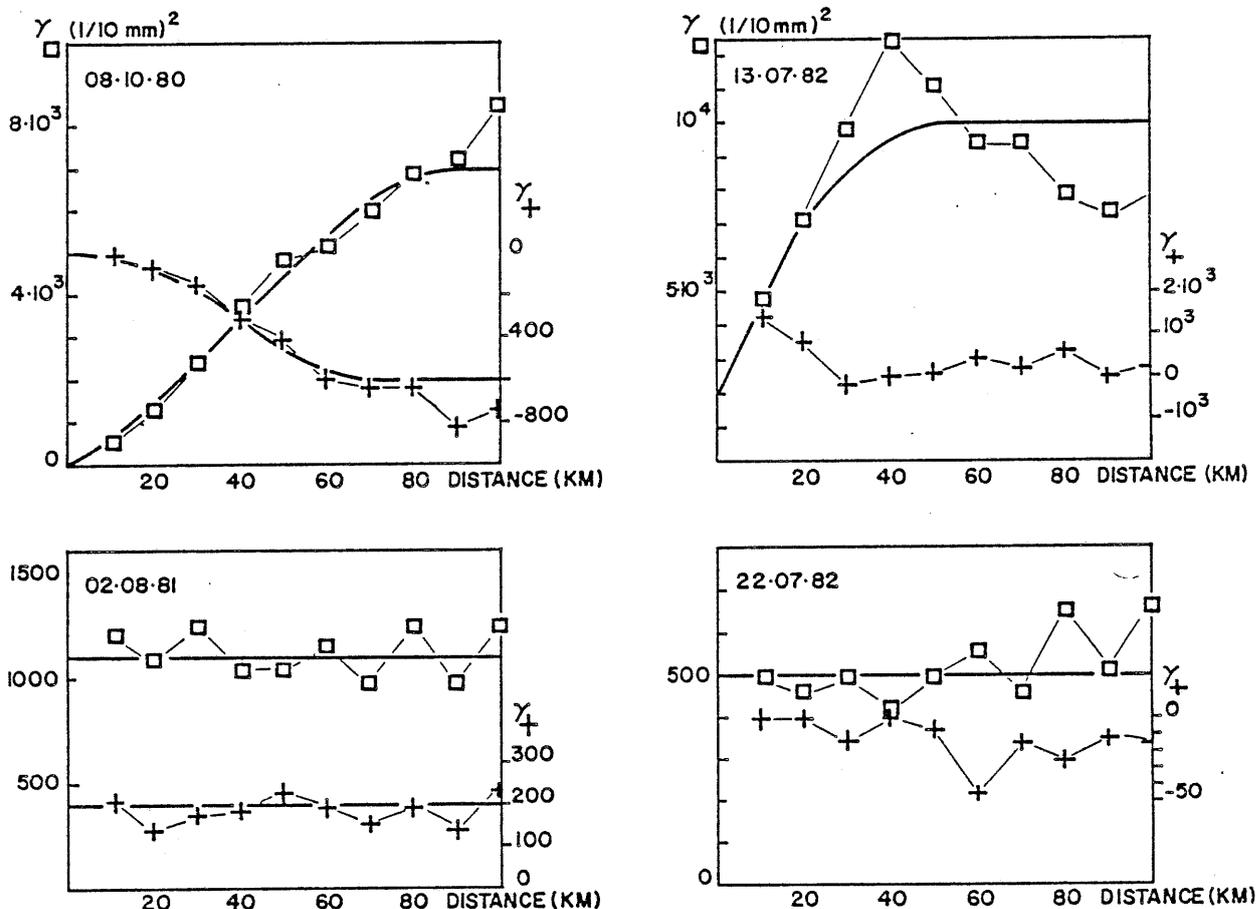
* un groupe de six journées (7/10/80, 8/10/80, 15/12/80, 23/6/82, 13/7/82 et 22/7/82) pour lequel un modèle de corrélation intrinsèque est plausible. A de petites fluctuations près, les différents variogrammes directs et croisés sont liés par une relation d'affinité respectant la condition de positivité évoquée Chapitre IV (formule IV.36). Nous pouvons remarquer que la variabilité croisée est généralement du même ordre de grandeur que la variabilité des mesures sol ; ceci rend plausible une décomposition orthogonale du type recalage d'une pré-estimation (voir Chapitre VI).

* trois journées (10/10/80, 14/12/80 et 22/10/82) pour lesquelles les mesures radar et sol sont indépendantes. Ce sont, nous l'avons vu, des journées où les mesures radar comportent des défauts très sensibles. Il conviendrait donc de ne pas prendre en considération ces mesures.

* deux journées (2/8/81 et 15/7/82) où les différents variogrammes ont été ajustés à l'aide de modèles de types différents (linéaires et sphériques). Il est sans doute nécessaire pour aboutir à une modélisation plus cohérente d'ajuster à chaque variogramme une combinaison de ces modèles ; malheureusement, les variogrammes expérimentaux se prêtent mal à de tels raffinements ; de plus pour ce type de situation météorologique, nous nous écartons très sensiblement des hypothèses de stationnarité.

En résumé, cette analyse de la structure directe et croisée des mesures sol et radar nous permet de justifier pleinement une approche recalage de pré-estimation pour six journées sur onze au sens où les résultats obtenus seront théoriquement équivalents à ceux du cokrigage si la variabilité spatiale des mesures radar égale la variabilité croisée (ce qui est presque vérifié). Pour les journées restantes, cette équivalence n'est pas acquise, néanmoins ce type de correction additive est toujours

Figure VII.3 : Variogrammes des écarts entre les mesures sol et radar en 98 points pour quatre journées représentatives. Variogrammes croisés des écarts et des valeurs radar.



Variogramme expérimental

□ des écarts
 et + croisé

raisonnable si l'on peut s'assurer que les termes correctifs sont spatialement organisés.

Analyse de la structure des écarts entre les mesures sol et radar.

Les résultats de cette analyse sont résumés par le tableau VII.1 et illustrés pour quatre journées figure VII.3. Deux types de remarques peuvent être faites :

- du point de vue de la cohérence globale de la modélisation proposée, cette analyse peut être utilisée comme un moyen complémentaire de vérification. En effet, il est facile de montrer que le variogramme des écarts s'exprime en fonction des variogrammes directs et croisés des variables étudiées :

$$\hat{\gamma}_{Z-ZT}(h) = \hat{\gamma}_{ZZ}(h) + \hat{\gamma}_{ZT,ZT}(h) - 2\hat{\gamma}_{Z,ZT}(h) \quad (\text{VII.1})$$

et la vérification de cette relation est un gage de cohérence de l'ensemble de la modélisation adoptée. Cette vérification est aisée dans les cas de corrélation intrinsèque et montre que pour les six journées concernées, quatre satisfont la relation VII.1 (les 7/10/80, 8/10/80, 15/12/80 et 13/7/82) alors que pour les deux journées restantes (23/6/82 et 22/7/82) la nature déjà fortement pépitiqque des structures directes et croisées devient prépondérante pour la structure des différences. Pour les journées où les deux types de mesures sont indépendants (10/10/80, 14/12/80 et 22/10/82), nous pouvons au moins vérifier que la variabilité des écarts est proche de la somme des variabilités des mesures radar et sol même si les modèles adoptés ne permettent pas une vérification rigoureuse de la relation VII.1. Les deux journées restantes se prêtent mal à ce type de vérification pour les raisons déjà évoquées aux paragraphes précédents.

- d'un point de vue pratique, nous pouvons remarquer que la variabilité des écarts n'est inférieure à celle des mesures sol que pour les quatre journées où le modèle de corrélation intrinsèque paraît sûr (7/10/80, 8/10/80, 15/12/80 et 13/7/82) et pour la journée du 15/7/82. Ce sont donc les seuls cas où les variances théoriques d'estimation par

recalage d'une pré-estimation seront plus faibles que les variances théoriques d'estimation par krigeage des seules mesures sol.

Enfin, nous avons vérifié l'indépendance entre les écarts et les mesures radar (hypothèse formulée par l'approche recalage d'une pré-estimation) en construisant les variogrammes croisés entre ces variables (tableau VII.1 et figure VII.3). Cette indépendance est remise en cause essentiellement pour les journées où les mesures radar sont douteuses (10/10/80 et 22/10/82 notamment).

c - Résultats

Pour deux réseaux d'étalonnage de densités différentes (29 puis 10 stations - voir figure III.6), nous avons combiné les mesures radar et sol des 11 journées disponibles par recalage d'une pré-estimation en utilisant deux modes de calcul différents :

- le premier utilise le krigeage pour interpoler les écarts entre les mesures sol et radar ; il est tout à fait conforme à la simplification décrite au chapitre précédent.

- le second mode de calcul interpole les écarts entre les mesures sol et radar à l'aide de fonctions spline ; nous avons rappelé au Chapitre I que ce système d'interpolation est une forme de krigeage universel utilisant systématiquement une fonction de covariance généralisée donnée afin de garantir certaines propriétés géométriques aux surfaces d'interpolation produites. Le principal défaut de cette technique est de ne pas prendre en compte la variabilité spatiale de la variable interpolée; celle-ci est toujours supposée avoir un comportement très lisse. Ce défaut s'accompagne d'un avantage : la fonction de structure étant choisie a priori, il n'est pas nécessaire d'effectuer une reconnaissance de structure; dans un contexte opérationnel, cette simplicité de mise en oeuvre nous a paru intéressante. Notons tout de même que cette solution constitue un pis-aller par rapport à une approche climatologique qui aurait cet avantage de simplicité en respectant mieux la variabilité propre du phénomène ; malheureusement, faute de série chronologique suffisamment longue une telle approche est interdite ici.

Les critères habituels d'efficacité ont été utilisés pour comparer les résultats obtenus aux valeurs de référence en 69 cibles test; afin de situer les résultats de ces méthodes de combinaison, les performances des dispositifs radar et sol considérés séparément sont rappelés pour chaque critère :

- le diagramme d'efficacité, présenté figure VII.4, a été allégé en raison de la grande similitude des résultats obtenus lorsqu'ils sont regardés à travers ce critère. En effet, quelle que soit la technique d'interpolation employée (krigeage ou spline - non représentée mais pratiquement confondue aux résultats du krigeage), la correction de l'image radar par interpolation des écarts conduit à une réduction importante du biais souvent conséquent des mesures radar seules ; cette réduction conduit à des résultats très comparables à ceux obtenus à l'aide du seul réseau d'étalonnage (figure III.7). Cette correction réduit également l'écart-type des erreurs de façon indépendante du nombre de points d'étalonnage (sauf pour les 7 et 8/10/80). Cette double réduction du biais et de l'écart-type d'erreur permet aux différentes mesures radar corrigées d'égaliser, au sens de ce critère, les valeurs de référence de manière satisfaisante (seuls le 10/10/80 et le 13/7/82 échappent à cette conclusion en raison du maintien respectivement d'un écart-type d'erreur et de biais trop forts).

- les distributions de facteurs de différences présentés figure VII.5 rappellent tout d'abord la distance importante au sens de ce critère entre les performances du radar seul et celles d'un réseau sol seul indépendamment de la densité choisie ici (29 ou 10 stations). Le recalage de la pré-estimation radar conduit à des résultats systématiquement mais faiblement meilleurs que ceux du dispositif sol seul ; cette amélioration est plus nette pour le réseau sol de plus faible densité ; l'interpolation par krigeage est plus performante que les fonctions spline dans le seul cas du réseau de 10 stations.

- les courbes d'efficacité obtenues pour un réseau d'étalonnage de 29 stations (figure VII.6) montrent un alignement systématique des performances du radar corrigé sur celles du meilleur dispositif considéré individuellement à l'exception de la journée du 2/8/81. Pour les journées à

Figure VII.4 : Diagramme d'efficacité comparant les résultats du radar seul et corrigé par krigeage des écarts avec le sol en 29 puis 10 stations d'étalonnage.

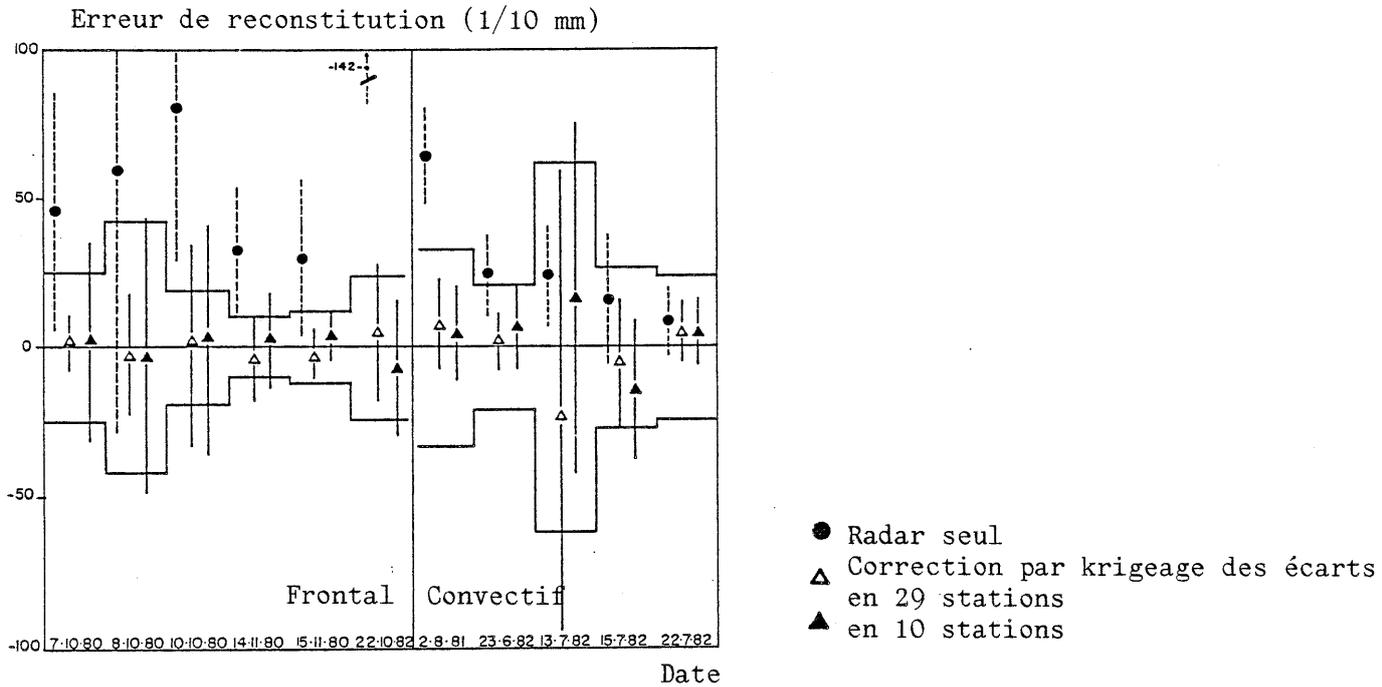


Figure VII.5 : Histogrammes cumulés des facteurs de différence obtenus pour les différentes méthodes d'étalonnage employées avec (a) un réseau de 29 stations puis (b) de 10 stations.

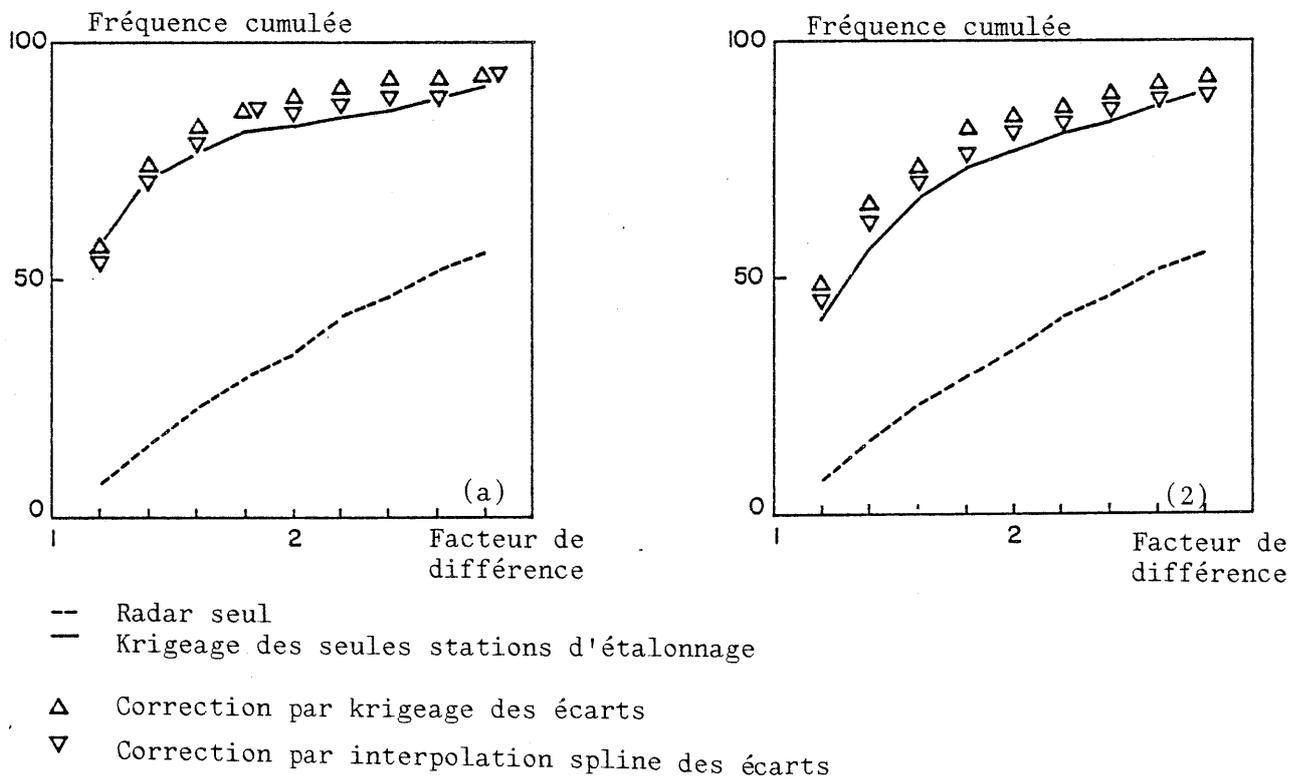


Figure VII.6 : Courbes d'efficacité des différentes méthodes d'étalonnage employées avec un réseau de 29 stations ; en référence sont portés les courbes d'efficacité du radar brut et du seul réseau d'étalonnage.

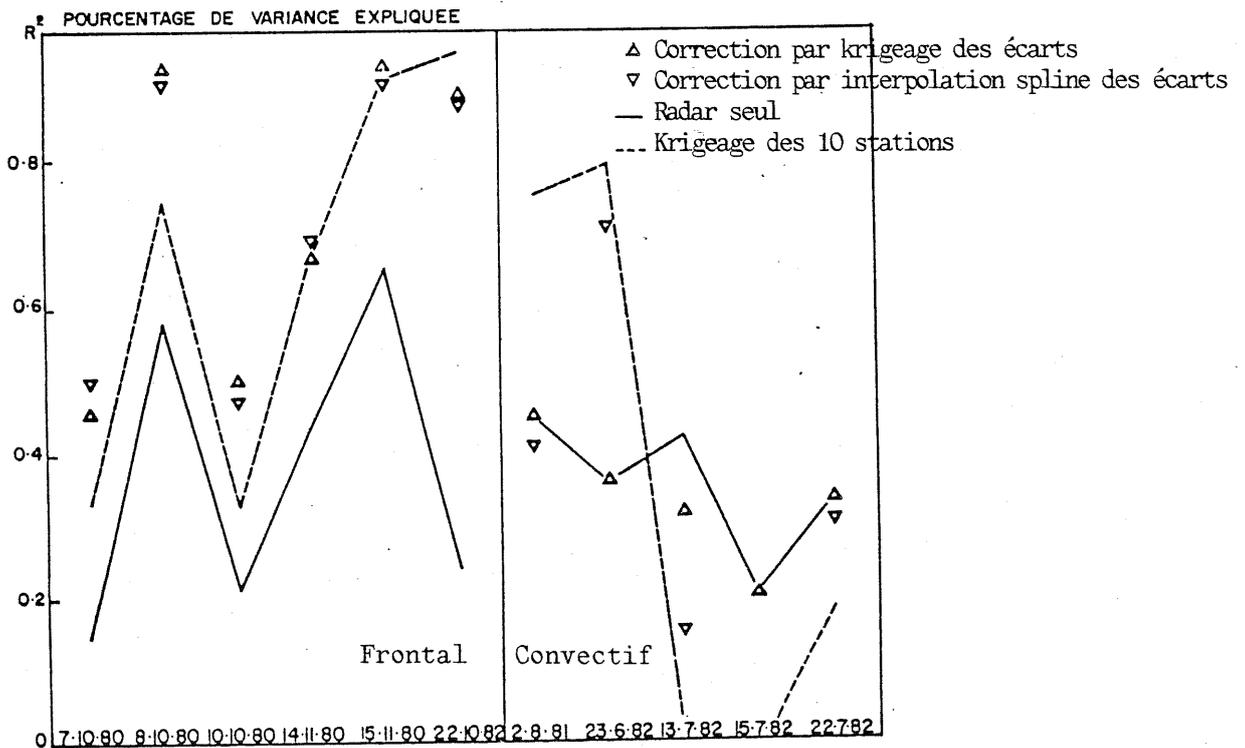
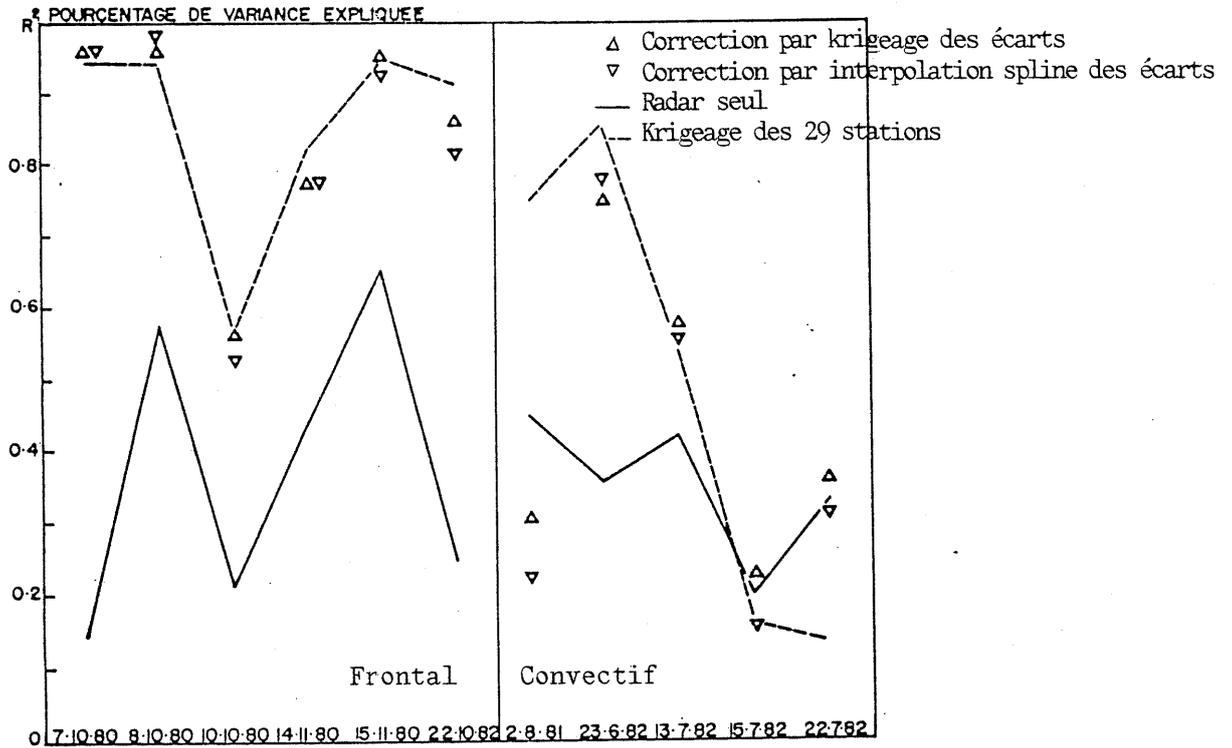


Figure VII.7 : Courbes d'efficacité des différentes méthodes d'étalonnage employées avec un réseau de 10 stations ; en référence sont portés les courbes d'efficacité du radar brut et du seul réseau d'étalonnage.

caractère météorologique frontal, cet alignement s'effectue sur les performances du sol qui sont toujours nettement supérieures à celles du radar ; les plus mauvais résultats de la combinaison sont logiquement obtenus lorsque les mesures combinées sont indépendantes d'après la modélisation de la structure croisée (10/10/80, 14/12/80 et surtout 22/10/82) ; une bonne co-fluctuation des signaux ne permet cependant pas de dépasser nettement les performances du sol déjà fortes ; les fonctions spline sont aussi efficaces que le krigeage. Pour les journées à caractère convectif les performances du sol sont parfois inférieures à celles du radar ; il est alors encourageant de constater que la combinaison de ces mesures s'aligne alors sur le radar ; par contre, la journée du 2/8/81 donne l'exemple d'un mauvais résultat de combinaison qui se place au-dessous des deux dispositifs de base : nous n'avons pas d'explication particulière à donner à propos de cette journée si ce n'est la présence (comme pour le 15/7/82 et 22/7/82) de nombreuses valeurs nulles qui rendent délicate l'application de notre modèle de correction et de notre critère de validité ; il convient donc de rester prudents sur les résultats obtenus à propos de ces journées convectives.

- les courbes d'efficacité obtenues pour un réseau d'étalonnage de 10 stations (figure VII.7) confortent les résultats précédents. En situation frontale l'apport de la combinaison est évident puisqu'il permet trois fois sur six de dépasser sensiblement les performances du sol seul quelle que soit la technique d'interpolation utilisée. En situation convective, il est toujours aussi délicat de tirer un enseignement clair des résultats obtenus, la solution la plus sage consiste sans doute, comme le fait le krigeage des résidus sous hypothèse de structure pépitique (pour les 5 journées à l'exception du 13/7/82), à corriger les mesures radar par un simple terme constant, ce qui n'a pas d'effet au sens de ce critère.

- d'un point de vue plus global, les différents coefficients de corrélation et erreurs de mesure regroupés dans le tableau VII.2 montrent que la correction des images radar (indifféremment par krigeage ou fonctions spline) n'a d'intérêt que lorsque le réseau sol est de faible densité. Le calcul de différents coefficients de corrélation partielle permet d'affiner cette constatation ; ainsi, si l'on dispose déjà d'un réseau sol de 29 stations (resp. 10), l'utilisation d'un dispositif radar

Tableau VII.2 : Coefficients de corrélation entre les valeurs de référence en 69 cibles et les valeurs reconstituées par le radar, le réseau sol seul et leur combinaison par recalage de pré-estimation.

	29 stations d'étalonnage			Radar	10 stations d'étalonnage		
	Sol seul	Interpolation des écarts			Sol seul	Interpolation des écarts	
		krigeage	spline			krigeage	spline
Corrélation	.94	.92	.90	.73	.90	.92	.90
Pourcentage de variance expliquée	.88	.85	.81	.53	.81	.85	.81
Corrélation après centrage par rapport aux moyennes spatiales	.76	.76	.74	.58	.50	.60	.63
Pourcentage de variance expliquée correspondante	.58	.58	.55	.34	.25	.36	.39
Erreur absolue normée par la moyenne	19.8	22.2	25.0	61.9	27.3	25.2	27.5

et sa correction à l'aide de ces stations permet d'expliquer 6 % supplémentaires (resp. 12 %) de variance de la référence ; inversement, si l'on dispose d'un radar, la correction de ses images par un réseau de 29 stations (resp. 10) entrainera un gain de 25 % (resp. 5%) de variance expliquée. Ces pourcentages, traduits en terme de corrélation partielle (c'est à dire normés par la variance résiduelle) montrent que l'apport relatif du radar au réseau sol seul est assez constant (14 % et 16 % pour 29 et 10 stations) ; l'apport relatif de l'étalonnage est plus variable (38 % et 8 % pour 29 et 10 stations).

En résumé, nous rappellerons tout d'abord que l'échantillon de journées disponibles est trop réduit pour permettre une conclusion ferme. Néanmoins, l'étalonnage des images radar apparait comme indispensable en situation météorologique frontale : la sous-estimation systématique du radar doit être corrigée et, de plus, il semble que la prise en compte de l'information sol améliore significativement la cofluctuation spatiale entre estimation et référence ; le recours aux fonctions spline n'entraîne pas de dégradation sensible de la qualité des résultats, il constitue donc une solution opérationnelle intéressante. En situation convective, les enseignements sont moins clairs : le biais des mesures radar est moins net et l'apport de l'information sol ne joue pas de manière constante. Si l'on confond les deux types de situation, les critères présentés ci-dessus montrent que le radar a plus besoin du réseau sol que le réseau sol n'a besoin de lui ; il faut cependant accompagner cette conclusion en remarquant que cette expérience a été menée à un pas de temps favorable aux mesures sol dans une zone géographique défavorable aux mesures radar (éloignement de plus de 70 km du radar).

2 - MESURE DE LA FRACTION JOURNALIERE D'INSOLATION EN FRANCE

a - Choix d'un étalonnage simplifié

Nous avons traité au Chapitre V l'étalonnage des mesures Cactus de la fraction journalière d'insolation dans le cadre général du cokrigeage. Les résultats obtenus peuvent être considérés comme bons, mais la relative lourdeur de la méthode incite à la recherche de simplifications : le but de ce paragraphe est d'identifier une simplification bien adaptée aux caractéristiques de ces mesures et d'apprécier la dégradation qu'elle entraîne au niveau des résultats.

L'analyse de structure présentée au Chapitre V a montré la bonne adaptation du modèle de corrélation intrinsèque aux différentes mesures centrées par rapport à leur moyenne climatologique puis par rapport à leur moyenne spatiale notées X et X_T. L'examen des différentes fonctions de covariance et de corrélation directes et croisées a indiqué que ces variables étaient vraisemblablement liées par une relation du type V.4 :

$$X^W_T(t) = X^W(t) + e^W(t)$$

Une telle décomposition peut permettre des simplifications du système de cokrigeage que nous avons regroupées sous le terme de cokrigeage d'erreurs au Chapitre VI. Malheureusement ces simplifications ne sont effectives qu'à la condition d'avoir des réseaux de mesures confondus, ce qui n'est pas le cas ici.

Une autre source de simplification possible est offerte par l'existence d'une corrélation intrinsèque. En effet, nous rappelons au Chapitre VI que l'étalonnage d'une image par cokrigeage est équivalent à son étalonnage par simple régression à condition que la corrélation intrinsèque soit parfaite, c'est à dire que le coefficient de corrélation:

$$\rho_{X,XT} = C_{X,XT}(h) / \sqrt{C_{XX}(h) \cdot C_{XT,XT}(h)}$$

soit égal à 1. Certes les résultats obtenus indiquent que le coefficient de corrélation intrinsèque des mesures de X et X_T s'écarte assez sensiblement

Tableau VII.3 : Coefficients de régression entre les valeurs centrées mesurées par le réseau sol et le radar pour différentes densités. Calcul effectué pour chacune des 15 journées sélectionnées.

Dates	Nombre de stations d'étalonnage					
	23		13		5	
	a	b	a	b	a	b
2.3.84	.83	6.	.77	3.	.76	4.
4.3.84	.44	16.	.65	8.	.53	17.
8.3.84	.52	18.	.59	14.	.29	21.
10.3.84	.77	1.	.83	- 2.	.89	- 5.
11.3.84	.83	- 5.	.82	- 5.	.86	- 6.
12.3.84	.65	2.	.66	2.	.69	8.
14.3.84	.74	- 7.	.72	- 7.	.72	- 7.
15.3.84	.67	- 9.	.67	- 7.	.65	-14.
16.3.84	.59	-10.	.57	- 8.	.54	- 8.
19.3.84	.65	-12.	.65	-10.	.12	3.
21.3.84	.64	- 9.	.65	-10.	.60	- 9.
22.3.84	.53	- 4.	.55	- 1.	.38	8.
23.3.84	.56	- 3.	.69	- 7.	.48	5.
24.3.84	.80	- 4.	.87	0.	.67	- 4.
25.3.84	.87	6.	.79	4.	.34	-11.

de 1 : sur l'ensemble des 71 journées étudiées ce coefficient vaut environ .92 et si l'on traite séparément les 15 journées de notre sélection, ce coefficient varie entre .61 et .93 (voir tableau V.3). Nous avons cependant pris le "risque" de faire une telle hypothèse en raison des simplifications radicales qu'elle entraîne mais en restant conscients de la sous-optimalité des résultats que nous obtiendrons.

En fait, nous rejoignons par cette démarche les préoccupations pratiques déjà exposées au paragraphe précédent : si l'étalonnage des images est destiné à être une pratique opérationnelle, il doit être aussi simple que possible et les utilisateurs potentiels de ces techniques doivent pouvoir situer en terme de précision les méthodes sophistiquées par rapport à des méthodes rudimentaires.

b - Résultats obtenus

Les résultats présentés ont été obtenus à l'aide de trois densités de réseau sol (23, 13 puis 5 stations héliographiques). Pour chaque journée ont été calculés les coefficients de régression liant la valeur centrée du sol à celle fournie par le système Cactus ; nous avons effectué ces régressions sur les valeurs Y et Y_T centrées uniquement par rapport aux moyennes climatologiques EZ et EZ_T puisque la corrélation filtre la moyenne spatiale de ces valeurs ; le tableau VII.3 regroupe donc les facteurs a et b tels que :

$$\hat{Y}(t) = a Y_T(t) + b + \varepsilon(t)$$

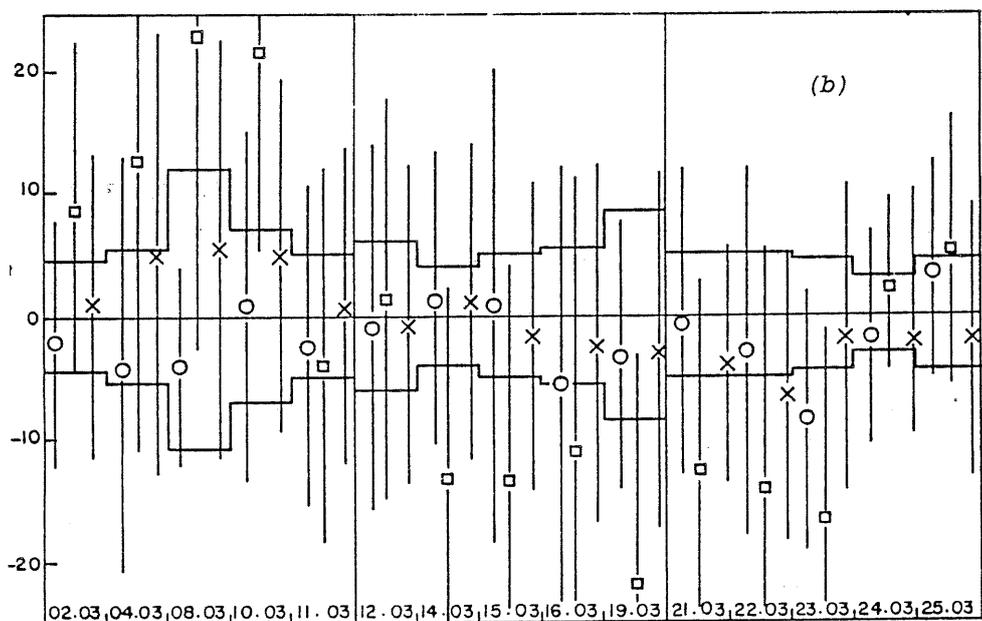
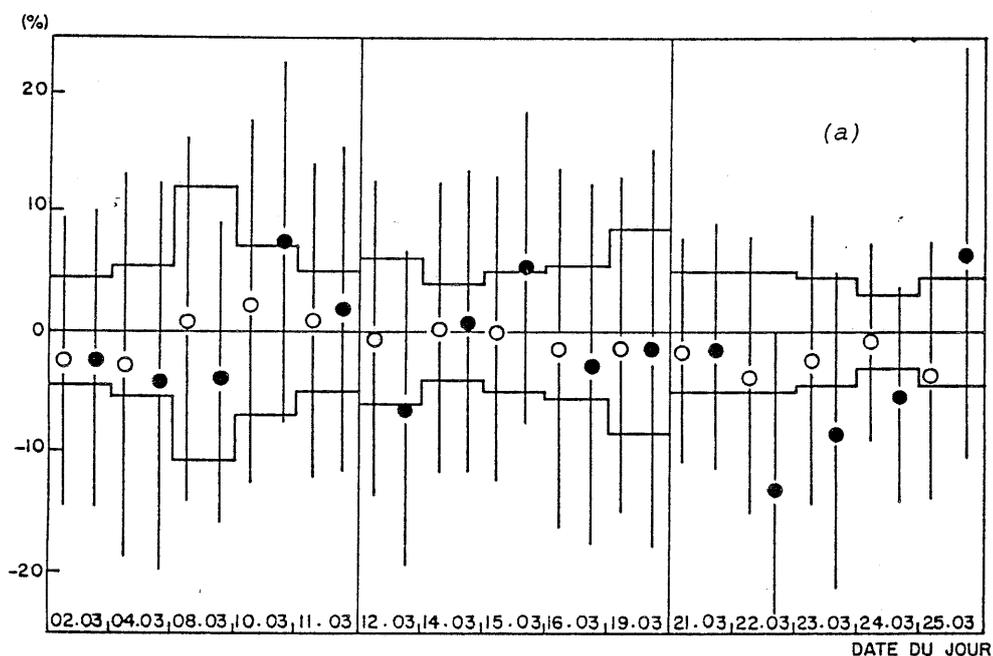
où $\varepsilon(t)$ est un résidu de moyenne nulle et de variance minimale. Nous avons ensuite calculé les écarts $\hat{Y}(t_g)$ aux noeuds de la grille puis fourni une estimation de $Z(t_g)$ suivant :

$$Z^*(t_g) = \hat{Y}(t_g) + EZ_T(t_g)$$

en considérant que la seule estimation de la moyenne climatologique disponible au point t_g est fournie par le système Cactus.

Figure VII.8 : (a) Diagramme d'efficacité de la régression à l'aide de 23 (o) puis 5 (●) stations héliographiques pour 61 cibles test et 15 journées;

(b) Diagramme d'efficacité du système Cactus (□) puis du krigeage (o) et de la régression (x) à l'aide de 13 stations héliographiques.



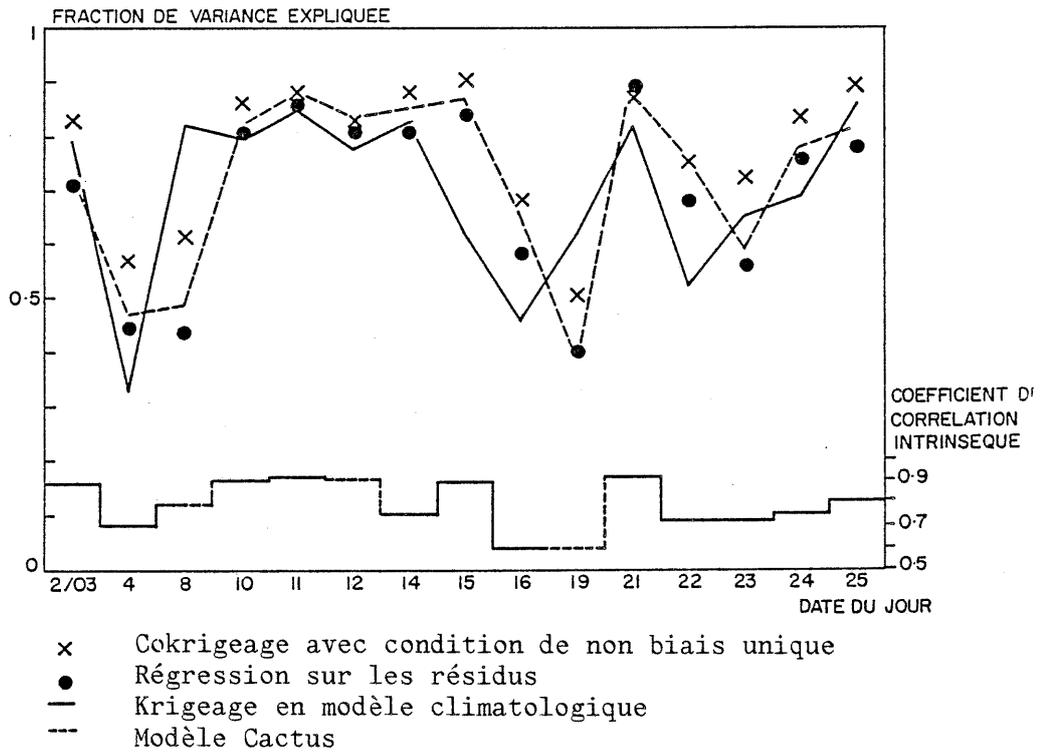
Dans un souci de simplification, nous n'avons pas repris systématiquement l'ensemble des résultats obtenus pour les différentes densités de réseau lors de l'examen des critères de ressemblance entre valeurs estimées et valeurs de référence. Ces critères montrent :

- une bonne efficacité de la méthode de régression proposée, en terme de réduction du biais des mesures Cactus. Les diagrammes d'efficacité de la figure VII.8 montrent que les erreurs moyennes après régression deviennent très faibles notamment lorsque l'on dispose d'une bonne densité de stations sol (23 ou 13 stations). Les erreurs absolues normées par la moyenne regroupée dans le tableau VII.4 confirment ce résultat de manière globale : quelle que soit la densité du réseau héliographique cette erreur après régression est sensiblement inférieure à celle commise par le système Cactus et même légèrement inférieure à celle obtenue après cokrigeage.

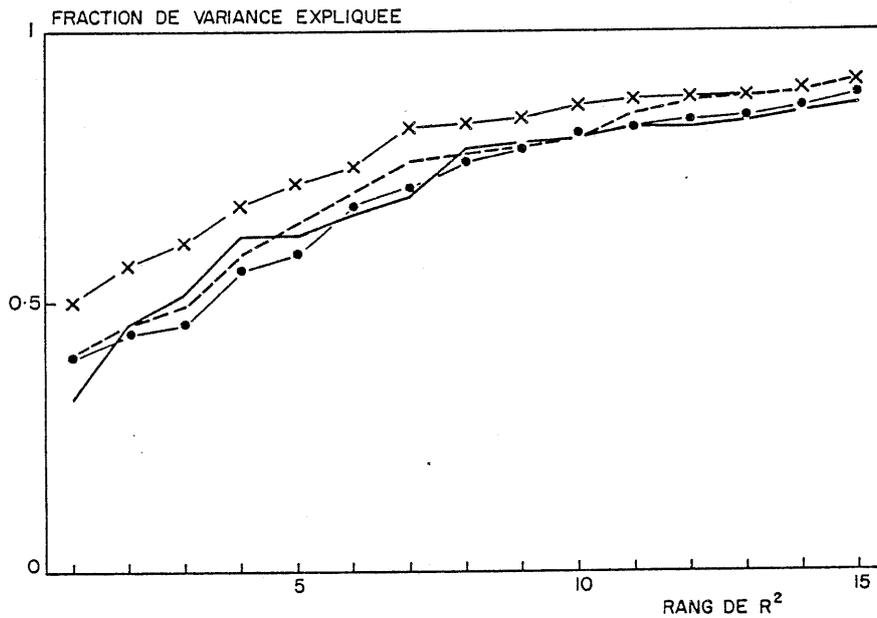
- une efficacité nulle en terme d'amélioration de la cofluctuation des estimations produites avec les valeurs de référence. La figure VII.9 illustre cette incapacité de la régression à améliorer les performances individuelles des systèmes qu'elle combine dans le cas d'un réseau héliographique de 13 stations. Il est à noter que la distance entre les performances du cokrigeage et de la régression, au sens de ce critère, est assez étroitement liée à la valeur du coefficient de corrélation intrinsèque (porté en référence) ; cette liaison est logique puisque nous ne pouvons attendre une égalité des performances de ces méthodes que dans le cas où ce coefficient de corrélation vaut 1, conformément à l'hypothèse formulée. Les résultats obtenus avec 23 ou 5 stations sont de même nature, comme l'indique de manière globale les coefficients de corrélation donnés par le tableau VII.4 : la régression est systématiquement plus mal placée que le cokrigeage, notamment après centrage par rapport aux moyennes spatiales.

En conclusion, cette simplification du système de cokrigeage se révèle incapable de tirer parti des propriétés complémentaires des mesures qu'elle combine ; lorsque le réseau employé est dense, les performances de ce réseau seul sont à tous égards supérieures à celles de la régression ; dans le cas d'un réseau de faible densité, l'utilisation du système Cactus seul est préférable car la légère réduction de biais obtenue par régression

Figure VII.9 : a) Fraction journée par journée de la variance des valeurs de référence en 61 cibles expliquée par les résultats de la méthode de régression à l'aide de 13 stations héliographiques.



b) Courbes d'efficacité correspondantes



est compromise par une détérioration de la cofluctuation spatiale des valeurs estimées avec la référence.

Tableau VII.4 : Coefficients de corrélation entre les valeurs référence en 61 cibles et les valeurs reconstituées par le modèle Cactus et, pour différentes densités de stations héliographiques, par krigeage, cokrigeage et régression des résidus climatologiques (pour 15 journées).

	Krigeage			Cactus	Cokrigeage			Régression		
	5	12	23		5	13	23	5	13	23
Nombre de stations sol	5	12	23		5	13	23	5	13	23
Corrélation	.67	.88	.90	.83	.86	.89	.90	.85	.88	.89
Pourcentage de variance expliquée	.45	.77	.81	.69	.74	.78	.81	.72	.77	.79
Corrélation après centrage par rapport aux moyennes spatiales	.57	.80	.84	.84	.84	.86	.87	.75	.80	.79
Pourcentage de variance expliquée correspondant	.32	.64	.70	.70	.71	.73	.75	.56	.64	.62
Erreur absolue normée par la moyenne	38.	24.	20.	38.	30.	25.	23.	26.	23.	22.

CONCLUSION

APPORTS DE CE TRAVAIL ET PERSPECTIVES OUVERTES

Nous avons tenté, par ce travail, d'apporter une réponse aux problèmes de validation et d'étalonnage de mesures de télédétection, en suivant une approche géostatistique.

Les trois premiers chapitres nous ont servi à définir un protocole de validation, en portant une attention particulière à la qualité des valeurs de référence et à la clarté des critères de ressemblance entre cette référence et les mesures validées. Un certain nombre de précautions ont été prises lors du calcul de la référence pour assurer l'indépendance des sources d'informations utilisées et l'homogénéité des mesures comparées. Les critères de ressemblance retenus présentent des résultats relatifs par rapport à la précision absolue de la référence, ou bien par rapport aux performances d'autres dispositifs de mesure.

Les quatre derniers chapitres décrivent différentes possibilités de combiner les mesures de télédétection à des mesures sol, respectant les caractéristiques statistiques de la co-fluctuation spatiale de ces deux types de mesures. Ces méthodes permettent de substituer à la démarche classique d'étalonnage d'un dispositif de mesure à l'aide d'un autre dispositif, une approche plus générale de combinaison, capable de se plier plus simplement à un critère d'optimalité et susceptible d'être étendue à plus de deux dispositifs.

Nous avons pris soin d'illustrer les détails théoriques de cette approche par des exemples pratiques, afin de donner au lecteur la possibilité de juger de l'intérêt ou des faiblesses des réponses apportées aux problèmes traités. Il est donc vraisemblable que chacun possède d'ores

et déjà sa propre conclusion, selon les buts qu'il poursuit, ou bien les moyens dont il dispose.

Pour notre part, nous nous contenterons de souligner la contradiction, à notre sens importante, entre la lourdeur des mesures de télédétection et la difficulté de compléter efficacement les simples mesures sol. La lourdeur des mesures de télédétection réside à la fois dans la nécessité de les interpréter et dans leur volume ; dans l'objectif d'une utilisation quantitative des images de télédétection, cette interprétation, toujours complexe, doit être suivie d'une critique rigoureuse que le volume des données alourdi considérablement. L'efficacité des mesures de télédétection (dans le contexte particulier des mesures dont nous disposons) est souvent faible en comparaison des mesures sol conventionnelles ; cette faiblesse invite à essayer de tirer parti de la combinaison de ces deux types de mesure a priori complémentaires ; les tentatives de combinaison entreprises se sont révélées souvent décevantes, y compris lorsqu'existait une très bonne co-fluctuation des deux types de mesures.

Cette contradiction peut être comprise comme un échec de l'utilisation quantitative de la télédétection lorsque des moyens conventionnels de mesure existent déjà. Il faut toutefois faire preuve de plus de nuance dans ce jugement, car il porte sur une utilisation particulière d'un ensemble de mesures réduit. Si l'on analyse de manière un peu plus fine nos résultats, il est possible d'identifier différentes sources d'échec que l'on peut ranger par ordre d'importance décroissante :

- la mauvaise qualité de certaines images de télédétection au sens de leur utilisation quantitative (cas du radar en particulier) ;
- le "choix" de pas de temps a priori très favorables aux moyens de mesure sol en raison de la faible variabilité spatiale des phénomènes à ces échelles de temps ;
- la difficulté d'inférer les modèles statistiques choisis pour combiner les deux types de mesure.

Dans le but d'éviter ces écueils, nous évoquons, ci-dessous, quelques perspectives de recherche auxquelles nous commençons à nous intéresser.

Sur le plan général de l'intérêt des méthodes que nous avons proposées, il nous semble important de souligner brièvement leurs avantages par rapport aux méthodes classiques :

- La clarté des critères choisis lors des différentes étapes (validation et étalonnage) grâce à un cadre théorique unique et rigoureux.
- La simplicité d'application obtenue par des hypothèses souvent peu restrictives.
- L'ouverture possible au cas de trois dispositifs de mesures distincts (ou plus).

Améliorer l'interprétation quantitative des mesures de télédétection

Cette proposition peut d'abord être prise dans son sens direct ; il s'agit alors d'imaginer des algorithmes plus performants pour traiter l'information disponible, et si possible, d'accéder à des informations complémentaires. Pour citer l'exemple du radar, une réflexion sur la différence des modes d'échantillonnage du radar et des pluviomètres (réflexion déjà menée par différents auteurs : I. ZAWADSKY, 1975 ou Ch. OBLED, 1985) doit conduire à une meilleure homogénéisation des mesures radar et sol ; la disponibilité nouvelle de mesures de réflectivité différentielle (mesures fournies par les radars à diversité de polarisation) doit donner des indications sur la distribution granulométrique des gouttes de pluie qui pourront être utilisées lors de la conversion réflectivité-intensité.

Mais cette proposition doit surtout être prise a contrario ; connaissant les difficultés d'interprétation quantitative dans certaines conditions d'utilisation des capteurs de télédétection, il est indispensable d'en limiter l'usage dans les limites où l'interprétation est la plus sûre. Dans le cas du radar, il paraît ainsi primordial de ne pas

travailler avec un faisceau à trop haute altitude (ce qui était typiquement le cas dans l'exemple du Bassin Parisien) ou bien en présence d'échos de sol. De telles précautions rendent sans doute possible l'utilisation de ces dispositifs à des pas de temps plus courts où leur compétitivité avec les moyens conventionnels doit s'améliorer sensiblement.

Assouplir les hypothèses des modèles statistiques et échapper au schéma linéaire

De nombreuses difficultés rencontrées lors de l'utilisation des modèles statistiques choisis ont résulté du contexte mono-réalisation auquel nous a ramené l'absence de séries climatologiques assez longues ; dans ce contexte des hypothèses de stationnarité très sophistiquées sont nécessaires et souvent mal vérifiées. Nous avons vu, à propos des mesures d'insolation, qu'un contexte multi-réalisations permet, par un codage approprié, de recourir à des hypothèses plus simples (élimination des problèmes de dérive, validité de fonctions de structure plus élémentaires).

Nous pensons qu'il est possible d'aller plus loin dans le sens d'un assouplissement des hypothèses à condition de se placer en contexte climatologique, ce qui est toujours envisageable en météorologie. Il est par exemple regrettable de maintenir une hypothèse de stationnarité des moments d'ordre 2 des mesures de télédétection sachant, d'une part, que cette hypothèse est rarement vérifiée et, d'autre part, que la forte résolution de ces mesures permet de calculer ces moments pour "tous" les couples de points du domaine d'étude. Une approche telle que l'analyse en composantes principales de processus (Ch. OBLED et J.D. CREUTIN, 1986) devrait permettre, sans cette hypothèse, de caractériser la structure des mesures télédétectées avec plus de nuance.

Enfin, quel que soit le contexte, une transformation préalable de la distribution fréquentielle des mesures vers une distribution Gaussienne est certainement souhaitable. Nous avons, à tort, négligé cette démarche qui permet d'échapper au schéma linéaire.

Travailler sur des données simulées

Cette dernière perspective de recherche ne constitue pas vraiment un remède aux difficultés rencontrées dans ce travail mais présente plutôt une alternative à notre approche.

Par simulation, nous entendons une expérience numérique comportant deux étapes :

i) La simulation de valeurs des variables météorologiques caractérisant le phénomène mesuré. Dans l'exemple des mesures pluviométriques par radar, il sera nécessaire de disposer avec une résolution temporelle fine et dans trois dimensions des distributions granulométriques de gouttes de pluie et, éventuellement, des caractéristiques du vent.

ii) La simulation du fonctionnement des différents dispositifs de mesure utilisés.

Un tel simulateur peut permettre de manière économique d'explorer une gamme de situations beaucoup plus large que celle couverte par l'expérience physique ; il est en effet possible de modifier de manière combinatoire les caractéristiques du phénomène et des appareils de mesure. De plus, la valeur exacte du phénomène est toujours connue. Un travail de validation et d'étalonnage tel que le nôtre devient alors plus robuste puisque deux causes majeures d'incertitudes disparaissent : l'absence de valeur de référence exacte et le nombre limité de situations étudiées.

Malheureusement se pose alors la question de la qualité du simulateur, de son aptitude à reproduire fidèlement l'expérience physique. Cette nouvelle incertitude vient remplacer celles que nous avons rencontrées : c'est en ce sens que cette approche n'est pas un remède à nos difficultés. Peu d'auteurs ont entrepris une telle démarche ; à notre connaissance seuls W. KRAJEWSKY ET K. GEORGEAKAKOS, 1985 ont essayé dans le domaine de la mesure de pluie par radar et pluviomètres ; leur simulateur élude la délicate première étape évoquée ci-dessus ; en prenant pour

variable météorologique de base une longue série d'images radar haute résolution (issue des expériences GATE) ; les mesures radar et sol déduites prennent en compte un nombre limité de sources d'erreurs possibles.

La nature simpliste de cette expérience est bien sûr un indice de la complexité de cette approche. En fait, la construction d'un simulateur très complet nous paraît illusoire ; par contre, ce type de démarche est certainement fructueux pour tester la sensibilité des mesures à une caractéristique précise de l'appareil de mesure dans un contexte météorologique simplifié.

BIBLIOGRAPHIE

- APPERT N., C. LOUVEAU, 1973
Principales méthodes de calcul sur ordinateur. Rapport du Centre de Recherche, Pont à Mousson n° 219.73, 198 pp.
- ATTEIA M., 1966
Existence et détermination de fonctions spline à plusieurs variables C.R. Ac. Sc. de Paris, T. 262, série A, 575-578.
- BATTAN L.J., 1973
Radar observation of the atmosphere. University of Chicago Press, 324 pp.
- BARNES S.L., 1964
A technique for maximizing details in numerical weather map analysis. J. App. Meteor., 3, 326-409.
- BARNSTON A.J., THOMAS J.L., 1983
Rainfall measurement accuracy in FACE : a comparison of gage and radar rainfalls. J. Appl. Meteor., 22, 2038-2052.
- BARRETT E.C., 1977
The assessment of rainfall in North-Eastern Oman through the integration of observations from conventionnal and satellite sources. Consultant report to FAO of the USA, W/K7629, 55 pp.
- BASTIN G., GEVERS M., 1984
Identification and Optimal Optimisation of Random Fields Structure from Scattered Pointwise data. Automatica,
- BRANDES E.A., 1975
Optimizing rainfall estimates with the aid of radar. J. Appl. Meteor., 14, 1339-1345.
- BOX G.E.P., JENKINS G.M., 1976
Time Series Analyses : Forecasting and Control. Holden-Day Ed., 575 pp.
- CHAUVET P., PAILLEUX J., CHILES J.P., 1976
Analyse objective de champs météorologiques par cokrigeage. La Météorologie 6(4), 37-54.
- CHAUVET P., 1977
Exemple d'application de la géostatistique non-stationnaire : le cokrigeage. Note 502, Centre de Géostatistique, ENSMP - Fontainebleau - France, 10 pp.

- CHILES J.P., 1978
L'inférence statistique automatique des FAI-k. Note n° 584, Centre de Géostatistique, ENSMP - Fontainebleau - France, 45 pp.
- CHILES J.P.,
Notice d'utilisation du programme GAMMA. Centre de Géostatistique, ENSMP - Fontainebleau - France, 36 pp.
- CHRISTAKOS G., 1984
On the problem of permissible covariance and variogram models. Water Resources Research, 20, 251-265.
- COLLIER C.G., P.R. LARKE, MAY B.R., 1983
A weather radar correction procedure for real time estimation of surface rainfall. Quart. J.R. Met. Soc. 109, 589-608.
- CREUTIN J.D., 1979
Méthodes d'interpolation optimale de champs hydrométéorologiques ; comparaison et application à une série d'épisodes cévennols. Thèse Docteur-Ingénieur. Institut de Mécanique de Grenoble, France, 174 pp.
- CREUTIN J.D., Ch. OBLED, 1982
Objective analysis and mapping techniques for rainfall fields : an objective comparison. Water Resources Research, 18, 413-431.
- CREUTIN J.D., 1983
Analyse structurale et cartographie de la fraction d'insolation et de quelques paramètres statistiques associés : application au réseau météorologique français. Météorologie et Ressources Renouvelables. Colloque AFME-OMM, Valbonne 1983, 20 pp.
- DARRICAU-BEUCHER H., 1981
Approche géostatistique du passage des données de terrain aux paramètres des modèles en hydrogéologie. Thèse ENSM de Paris, Centre de Géologie Informatique, 119 pp.
- DELFINER P. and DELHOMME J.P., 1973
Application du krigeage à l'optimisation d'une campagne pluviométrique en zone semi-aride. Actes du colloque UNESCO - OMM - AIHS, MADRID 1973, Tome I, 191-210.
- DELFINER P., 1975
Linear estimation of non stationary spatial phenomena. Advanced Geostatistics in the Mining Industry, M. Guarascio et al. Ed., Hingham - Mass. 49-68.
- DELFINER P., DELHOMME J.P., CHILES, 1978
BLUEPACK-3D, Notice d'utilisation. Centre de Géostatistique, ENSMP - Fontainebleau - France, 198 pp.
- DELFINER P., MATHERON G., 1980
Les fonctions aléatoires intrinsèques d'ordre k. Rapport n° C84. Centre de Géostatistique - ENSMP - Fontainebleau - France, 36 pp.

- DELFINER P., DELHOMME J.P., CHILES J.P., GALLI A. and RENARD D., 1984
BLUEPACK-3D, Notice d'utilisation. Centre de Géostatistique, ENSMP -
Fontainebleau - France, 198 pp.
- DELHOMME J.P., 1974
La cartographie d'une grandeur physique à partir de données de
différentes qualités. Symposium de l'Assoc. Int. des Hydrogéologues,
Mémoire n° X(1), 185-194.
- DELHOMME J.P., 1978
Kriging in hydrosociences. Advan. Water Resources, 1(5), 251-266.
- DELHOMME J.P., 1979
Etude de la géométrie du réservoir de Chemery. Contract report,
Centre d'informatique géologique, Fontainebleau - France, pp. 31.
- DELRIEU G., 1986
Evaluation d'un radar météorologique pour la mesure des précipita-
tions : validation et étalonnage par technique géostatistique.
Application au Bassin Parisien. Thèse USM et INP Grenoble, IMG,
200 pp.
- DOVIAK R.J., 1983
A survey of radar rain measurement techniques. J. Appl. Meteor.
22, 832-849.
- DUBRULE O., 1982
Krigage et splines en cartographie automatique. Thèse de l'Ecole
des Mines de Paris, 141 pp.
- DUBRULE O., 1984
Comparing splines and kriging. Computers and Geosciences, vol. 10,
n° 2-3, 327-338.
- DUCHON J., 1975
Fonctions spline du type plaque-mince en dimension 2. Séminaire
d'Analyse Numérique de l'IMAG n° 231, Université de Grenoble, 13 pp.
- DURAND, 1985
Communication personnelle concernant la méthode utilisée par la
Météorologie Nationale (SMM/PREVI) pour l'établissement des cartes
de géopotentiel.
- EDDY A., 1964
The Objective Analysis of Horizontal Wind Divergence Fields.
Quart. J. Roy. Meteor. Soc., 90, 424-440.
- FRANCOIS-BONGARCON D., 1981
Les corégionalisations, le cokrigeage. Lecture notes C86, Centre
de Géostatistique, Fontainebleau - France, 26 pp.
- FROMENT G., 1981
Projet Hydromel. Rapport interne de la Météorologie Nationale,
32 pp.

GALLE S., 1987

Mesures journalières de l'insolation en France à l'aide d'un système de télédétection et d'un réseau sol conventionnel : étude statistique de la cohérence et de la combinaison de ces mesures. Thèse de l'Université de Grenoble, 21 mai 1987.

GALLI A., GERDIL-NEUILLET F., DADOU C., 1985

Factorial Kriging Analysis : a Substitute to Spectral Analysis of Magnetic Data. "Geostatistics for National Resources Characterization" (Proc. 2nd NATO/ASI, Lake Tahoe, California). D. Reidel Ed., Dordrecht.

GANDIN L.S., 1965

Objective analysis of meteorological fields. Israel Program for Scientific Translation, Jerusalem, 242 pp.

GILET M., OLIVIERI J., GAILLARD G., 1980

La chaîne d'acquisition radar SAPHYR. La Météorologie, 23, 25-32.

GILET M., CICCIONE M., GAILLARD G., TARDIEU J., 1982

Projet Aramis : le réseau français de radars météorologiques. Note EERM/CRPA de la Météorologie Nationale, 15 pp.

GOUSSEBAILLE J., 1974

Analyses statistiques des données journalières d'insolation et d'irradiation ; application au Sud-Est de la France. Thèse de l'Université de Grenoble, 213 pp.

GRIFFITH C.G., 1986

Error analyses of satellite rain estimates for the central United States during august 1979. Presented at "Modeling of rainfall fields" Chapman Conference, Caracas, 23 p

HALL A.J., BARCLAY P.A., 1975

Methods of determining areal rainfall from observed data. In "Prediction in catchment hydrology", Australian Academy of Science, Canberra - Australia, 47-57.

HARROLD T.W., ENGLISH E.J., NICHOLASS C.A., 1974

The accuracy of radar - derived rainfall measurements in hilly terrain. Quart. J.R. Met. Soc., 100, 331-350.

HILDEBRAND P.H., TOWERY N., SNELL M.R., 1979

Measurement of convective mean rainfall over small areas using high density raingage and radar. J. Appl. Meteor., 18, 1316-1326.

HUDLOW M.O., 1983

Proposed off-site precipitation processing system for NEXRAD. Preprints 21st Conf. Radar Meteorology, Edmonton Amer. Meteor. Soc., 394-403.

JAMESON A.R., 1985

Deducing the microphysical character of precipitation from multiple parameter radar polarization measurements. J. of Clim. and Applied Meteor., 24, 1037-1047.

- JOURNEL A., HUIJBRECHTS Ch.J., 1978**
Mining geostatistics. Academic Press, pp. 600.
- JOWETT G.H., 1955**
The comparison of means of industrial time series. J. of Appl. Stat., 4, 32-46.
- KOISTINEN J. and PAHUKKA T., 1984**
A comparison between different gauge-radar adjustment techniques of daily rainfall. 22nd Conf. on Radar Meteorology, Zurich (paper not available).
- KRAJEWSKY W.F. and GEORGAKAKOS K.P., 1986**
Synthesis of radar rainfall data. Water Res. Research 21-5, 764-768.
- KRAJEWSKY W., 1986**
Sensitivity analysis of co-kriging of noisy radar and raingage data. Presented at "Modeling of rainfall fields" Chapman Conference, Caracas.
- KRUGER H.B., 1964**
A Statistical Dynamical Objective Analysis Scheme. Canadian Meteorological Memoirs, n° 18, Met. Branch, Dpt. of Transport, Toronto.
- LACOMBA P., 1986**
Evaluation des précipitations par combinaison d'images satellitaires visible et infra-rouge et de réseau de pluviomètres : applications à la péninsule arabique et au sud-ouest de la France. Thèse de Docteur-Ingénieur de l'USTMG et l'INPG, Grenoble, 217 pp.
- LEBEL T., 1984**
Moyenne spatiale de la pluie sur un bassin versant : estimation optimale, génération stochastique et gradex des valeurs extrêmes. Thèse Docteur-Ingénieur, Université et INP de Grenoble, 339 pp.
- MARBEAU J.P., 1978**
Géostatistique forestière : état actuel et développements nouveaux pour l'aménagement en forêt tropicale. Thèse de Docteur-Ingénieur, Ecole des Mines, Fontainebleau - France, 180 pp.
- MARECHAL A., 1970**
Cokrigeage et regression en corrélation intrinsèque. Note 205, Centre de Géostatistique, ENSMP - Fontainebleau, France, 38 pp.
- MARSHALL J.S., LANGILLE R.C. et W.McK. PALMER, 1947**
Measurement of rainfall by radar. J. of Meteor., 4, 186-192.
- MARSHALL J.S. et W.McK. PALMER, 1948**
The distribution of raindrops with size. Journal of Meteorology, 5, 165-166.

- MATHERON G., 1965**
Les variables régionalisées et leur estimation. Masson Ed., Paris, 305 pp.
- MATHERON G., 1970,**
La théorie des variables régionalisées et ses applications.
Fascicule 5, Centre de Géostatistique - ENSMP - Fontainebleau, France, 212 pp.
- MATHERON G., 1971**
The theory of regionalized variables and its applications. Centre de Géostatistique, Fontainebleau - France, 212 pp.
- MATHERON G., 1971**
La théorie des fonctions aléatoires intrinsèques généralisées.
Rapport n° 117. Centre de Géostatistique - ENSMP - Fontainebleau - France, 39 pp.
- MATHERON G., 1973**
The intrinsic random functions and their applications. Adv. Appl. Prob.5, 439-468.
- MATHERON G., 1975**
Le choix des modèles en géostatistique. Advanced Geostatistics in the Mining Industry, M. Guarascio et al Ed., D. Reidel Publ. Comp. Hingham - Mass., 11-27
- MATHERON G., 1979**
Recherche de simplification dans un problème de cokrigage. Internal note n. 698, Centre de Géostatistique, Fontainebleau - France, 19 pp.
- MATHERON G., 1981**
Splines et krigeage : le cas fini. Note n° 698. Centre de Géostatistique. ENSMP, Fontainebleau - France, pp.
- MEDAL D., E. RICHARD, R. ROSSET, OBLED Ch., NICKERSON E.C., 1984**
A comparison between observed and computed precipitations over complex terrain with a three dimensional mesoscale model including parameterized microphysics. Proc. 9th Intern. Cloud Physics Conference, Tallin - USSR.
- MYERS D.E., 1982**
Matrix formulation of co-kriging. Mathematical Geology, 14, 249-257.
- OBLED Ch., LACOMBA P., FREIN Y., 1984**
Predicting rainfall from satellite images on area I. Rapport interne IMG-BRGM, 43 pp.
- OBLED Ch., 1985**
Le spatio-pluviomètre. Séminaire à l'Université de Clermont-Ferrand.
- OBLED Ch., CREUTIN J.D., 1986**
Some developments in the use of EOF's for mapping meteorological fields. J. of Climate and Applied Meteorology, Vol. 295, n° 9, pp. 1189-1204.

- PAIHUA-MONTES L., 1978**
Méthodes numériques pour le calcul de la fonction spline à une ou plusieurs variables. Thèse U.S.M.Grenoble, 171 pp.
- PASTRE C., 1981**
Développement d'une méthode de détermination du rayonnement solaire global à partir des données Météosat. La Météorologie, VI-24, 5-15.
- RODDA J.C., 1967**
The rainfall measurement problem. "Geochemistry, Precipitation, Evaporation, Soil-moisture, Hydrometry", General Assembly of Bern, 215-231.
- RICHARDS W. et CROZIER C., 1983**
Precipitation measurements with a C-band radar in Southern Ontario. Atmos. - Ocean., 21, 125-137.
- RUTHERFORD I.D., 1972**
Data Assimilation by Statistical Interpolation of Forecast Error Fields. J. of Atm. Sc., 29, 809-815.
- SAUVAGEOT H., 1982**
Radarmétéorologie. Eyrolles Editeur, 296 pp.
- SCHLATTER T., 1974**
Some Experiments with a Multivariate Statistical Objective Analysis Scheme. Monthly Weather Review, 103, 246-257.
- SERRA J., 1967**
Echantillonnage et estimation locale d'un phénomène de transition minier. Thèse Université de Nancy (2 tomes).
- SLIMANI M., 1985**
Etude des pluies de fréquence rare à faibles pas de temps sur la région Cévennes-Vivarais : Estimation, relation avec le relief et cartographie synthétique. Thèse Docteur de l'Institut National Polytechnique de Grenoble, 23/9/1985.
- THORPE W.R., ROVE C.W., SILMPSON R.W., 1979**
Areal interpolation of rainfall using a double Fourier series. J. of Hydrol., 42, 171-177.
- TOURASSE P., 1981**
Analyses spatiales et temporelles des précipitations et utilisation opérationnelle dans un système de prévision des crues ; applications aux régions cévennoles. Thèse de l'Université et de l'Institut Polytechnique de Grenoble, 189 pp.
- VAN DIJK A., 1981**
Précipitation assessment from environmental satellite data for Northwest Lybia including the Gefara Rain. Consultant's Report to FAO, 45 pp.

- VAUCLIN M., VIEIRA S.R., VACHAUD G., NIELSEN D.R., 1983
The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J., 47, 175-184.
- VIALAR J., 1956
Calcul des probabilités et statistique (Tome II). Editions de la Météorologie Nationale, 201 pp.
- WACKERNAGEL H., 1985
L'inférence d'un modèle linéaire en géostatistique multi-variable. Thèse ENSM de Paris, Centre de Géostatistique, 100 pp.
- WILSON J.W. and E.A. BRANDES, 1979
Radar measurement of rainfall - a summary. Bull. Amer. Meteor. Soc., 60, 1048-1058.
- WOODLEY L.W., OLSEN A.R., HERNDON A., WIGGERT V., 1975
Comparison of gage and radar methods of convective rain measurement. J. Appl. Meteor., 14, 909-928.
- YUDIN M.I., 1961
Some Regularities in the Structure of Geopotential Field. Trudy GGO, n° 121.
- ZAWADSKY I.I., 1975
On radar-raingage comparison. J. of Appl. Meteor. 14, 8, 1430-1436.
- ZAWADSKY J., 1984
Factors affecting the precision of radar measurement of rain. 22nd Conference on Radar Meteorology, Boston, 251-256.

Titre de l'ouvrage :

**VALIDATION ET ÉTALONNAGE D'IMAGES DE TÉLÉDETECTION
A L'AIDE DE MESURES PONCTUELLES EN SUIVANT
UNE APPROCHE GÉOSTATISTIQUE**

**Applications à la mesure des précipitations
et de l'insolation**

Nom de l'auteur : Jean-Dominique CREUTIN

Établissement : Institut de Mécanique de Grenoble

RÉSUMÉ

Comment utiliser de manière quantitative l'information fournie par un dispositif de télédétection météorologique afin d'améliorer la mesure conventionnelle donnée par les réseaux de stations au sol ? Pour répondre à cette question, importante d'un point de vue pratique, ce mémoire propose une approche géostatistique permettant, d'une part, de vérifier la qualité des mesures obtenues par chaque dispositif, et d'autre part de combiner les mesures sol aux images de télédétection. Des exemples d'application concernant la mesure des précipitations et de l'insolation sont analysés afin d'illustrer la méthodologie proposée en la comparant à des approches plus traditionnelles.

La définition d'un protocole de validation s'appuie sur des rappels de géostatistique classique (Chapitre I). Ce protocole fixe des valeurs de référence permettant de vérifier les résultats obtenus par un dispositif de mesure d'après différents critères statistiques d'erreur (Chapitre II).

Le système de cokrigage est décrit en terme de méthode d'étalonnage capable de combiner des mesures ponctuelles et de télédétection (Chapitre IV). Différentes simplifications de cette méthode sont détaillées et rapprochées des méthodes classiques d'étalonnage (Chapitre VI).

Les données utilisées pour mettre en oeuvre ces méthodes de validation et d'étalonnage (Chapitres III, V et VII) sont des mesures de précipitations en région tempérée (radar et pluviomètres dans le Bassin Parisien) ou semi-aride (satellites et pluviomètres en Arabie) et des mesures d'insolation en région tempérée (satellite et héliographes en France).

MOTS-CLÉS

Télédétection météorologique ; Mesure des précipitations ; Mesure du rayonnement solaire ; Cokrigage ; Géostatistique multivariable ; Analyse objective multivariable ; Analyse d'image ; Validation et étalonnage d'image.