



**HAL**  
open science

**Analyse des champs spatiaux par utilisation de la  
télédétection : estimation de la durée quotidienne  
d'insolation en France à l'aide d'images du satellite  
Météosat et de mesures sol**

Sylvie Galle

► **To cite this version:**

Sylvie Galle. Analyse des champs spatiaux par utilisation de la télédétection : estimation de la durée quotidienne d'insolation en France à l'aide d'images du satellite Météosat et de mesures sol. Traitement du signal et de l'image [eess.SP]. Université Scientifique et Médicale de Grenoble; Institut National Polytechnique de Grenoble - INPG, 1987. Français. NNT : . tel-00694114

**HAL Id: tel-00694114**

**<https://theses.hal.science/tel-00694114>**

Submitted on 3 May 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 par

**Sylvie GALLE**

pour obtenir le titre de

DOCTEUR

de l'INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

(Arrêté ministériel du 5 Juillet 1984)

Spécialité : MÉCANIQUE

---

**ANALYSE DES CHAMPS SPATIAUX  
PAR UTILISATION DE LA TÉLÉDÉTECTION.  
ESTIMATION DE LA DURÉE QUOTIDIENNE D'INSOLATION  
EN FRANCE A L'AIDE D'IMAGES  
DU SATELLITE MÉTÉOSAT ET DE MESURES SOL.**

---

Date de soutenance : 21 Mai 1987

Composition du Jury :

M. Cl. DELORME	Président et Rapporteur
M. D. MYERS	Rapporteur
M. N. BERIOT	
M. J.D. CREUTIN	Examineurs
M. Ph. BOIS	

Thèse préparée au sein de l'INSTITUT DE MÉCANIQUE DE GRENOBLE  
Laboratoire de l'USTMG - INPG, Associé au CNRS (LA 6)

## AVANT-PROPOS

Cette étude a été réalisée au sein de l'équipe "Hydrologie et Ressources Renouvelables" de l'Institut de Mécanique de Grenoble. La qualité de l'environnement scientifique et humain que j'ai rencontré dans cette équipe de recherche, et plus largement dans le groupe Hydrologie que dirige M. G. VACHAUD, a contribué de manière déterminante au succès de cette étude.

Qu'il me soit permis de remercier les membres du jury qui ont accepté d'évaluer ce travail:

M. Cl. DELORME, Professeur à l'Université d'Avignon, qui a bien voulu assurer la présidence de ce jury. Il a ainsi apporté son expérience approfondie sur l'utilisation de la télédétection pour l'estimation de la ressource énergétique solaire.

M. D.E. MYERS, Professeur à l'Université de Tucson, Arizona, qui a pris la peine de lire minutieusement un document écrit en français, pour y apporter ses remarques pertinentes de mathématicien.

M. N. BERIOT, Directeur Adjoint du Centre de Météorologie Spatiale de Lannion a mis à notre disposition les données du programme Cactus, qui ont ainsi pu être évaluées dans un laboratoire de recherche différent. Son intérêt pour mon travail a toujours été enrichissant.

M. Ph. BOIS, Professeur à l'ENSHMG, a suivi le déroulement de ce travail. Sa compétence et sa hauteur de vue trouvent souvent leur expression dans des anecdotes qui donnent à réfléchir et dont j'ai beaucoup apprécié la forme.

Je dois une reconnaissance particulière à M. J.D. CREUTIN, Chargé de recherche au CNRS. Sa compétence et sa facilité à expliquer les problèmes, posés par les applications de la géostatistique, m'ont appris à aborder les problèmes. Il sait le plaisir que j'ai eu à travailler avec lui durant ces quatre années.

Je tiens également à remercier M. C. OBLED, Professeur à l'ENSHMG, pour l'intérêt qu'il a porté à mon travail et la richesse de ses suggestions, particulièrement pour ce qui concerne l'ACP et son interprétation.

De manière plus large, il convient de rappeler la collaboration avec Mme C.VILLIEN, ainsi que E.PESIN et M.P.RAOUL de la Météorologie Nationale qui m'ont toujours bien accueillie.

Je n'oublierai pas l'ambiance chaleureuse et la découverte des difficultés et des enthousiasmes de la recherche avec les autres chercheurs et thésards de l'équipe : Hervé Andrieu, Guy Delrieu, Philippe Lacomba, Jean Nalbandis, Mohamed Slimani, Philippe Bompard. Sans oublier le café avec les "hydrosout", toujours très convivial.

Pour finir, je remercie chaleureusement tous ceux qui m'ont aidé de façon très pratique mais surtout par leur amitié : Claudie, Olivier, Marion, Muriel, Jub, Chantal, Jean-Marc, Marie-Pierre, Jean.

## SOMMAIRE

INTRODUCTION	5
1ère Partie : DESCRIPTION DES DONNEES	10
I - La ressource énergétique solaire	10
1. Rayonnement hors atmosphère	10
2. Transmission à travers l'atmosphère	11
3. Rayonnement solaire atteignant le sol	12
II - Mesures du rayonnement solaire au sol	13
1. Le rayonnement direct	13
2. Le rayonnement global	13
3. Le rayonnement diffus	14
4. La durée d'insolation	14
5. La nébulosité	15
6. Réseaux de mesure en France	15
7. Liaison entre mesures énergétiques et mesures plus simples	16
III - Mesures indirectes par télédétection	19
1. Mesures fournies par les satellites	19
2. Quelques méthodes d'évaluation de l'insolation à partir de mesures de satellites	22
3. Le programme CACTUS	33
IV - Données disponibles pour cette étude	38
Conclusion	42

2ème Partie : ANALYSE DE LA STRUCTURE SPATIALE	43
Introduction	44
Notations utilisées	45
I - Modélisation statistique dans le cadre de l'analyse objective	46
1. Notion de variable régionalisée	46
2. Problème de l'inférence statistique	46
3. Formulation mathématique des fonctions de structure classique	47
4. Inférence statistique dans le cadre de la stationnarité	48
5. Inférence statistique dans le cas non stationnaire	52
6. Calcul et modélisation des fonctions de structure	53
7. Interprétation	57
8. Influence d'un bruit blanc sur la structure de la mesure	58
9. Structure croisée	61
10. Conclusion	63
II - Application à l'analyse du rapport d'insolation	64
1. Mise en forme des données : codage utilisé	64
2. Analyse de la structure spatiale des mesures effectuées au sol et par le satellite	80
III- Analyse en composantes principales	106
1. Analyse en composantes principales classique	106
2. ACP de processus	109
3. Analyse des cosinus directeurs du rapport d'insolation	116
IV - Renforcement de réseau	137
1. Introduction	
2. Comparaison de la structure perçue par les héliographes et la méthode CACTUS	138
3. Critères de sélection utilisés	140
4. Résultats obtenus dans le sud-ouest de la France	142
5. Conclusion	144

3ème Partie : VALIDATION	145
Introduction	146
Critères utilisés	146
Prise en compte des variations saisonnières	148
I - Validation des mesures satellitaires	149
1. Etude de la corrélation des valeurs journalières	149
2. Distribution des erreurs en fonction des valeurs de CACTUS	164
3. Intégration des mesures (temporelle et spatiale)	167
II - Performances des interpolations basées sur le sol seul	177
1. But de cette interpolation	177
2. Description des méthodes utilisées	177
3. Performance des différentes méthodes	185
III- Comparaison des performances des différentes estimations sur un jeu de données indépendant	192
1. Les courbes d'efficacité	192
2. Le tableau de performance	194
3. Remarques sur les courbes d'efficacité	194
4. Illustration pour un jour de printemps	197
Conclusion	202

4ème Partie : ETALONNAGE DU SATELLITE A L'AIDE DE MESURES HELIOGRAPHIQUES	205
Introduction	206
I - L'interpolation optimale multivariable	207
1. Système d'interpolation optimale multivariable	207
2. Simplification possible : recalage de Guess field	211
II - Etalonnage des valeurs satellitaires	212
1. Fonction de structure utilisée	212
2. Choix de la variable	213
3. Estimation de la moyenne climatologique	214
4. Choix du voisinage	216
5. Résultats obtenus	216
Conclusion	221
 CONCLUSION GENERALE	 223
 ANNEXES	 
A1. Détermination des seuils du programme CACTUS	227
A2. La projection conique conforme de Lambert	234
A3. Positionnement de la grille calculée par CACTUS	239
A4. Liste des jours disponibles en 1984	245
A5. Calcul de la durée maximale d'insolation	246
A6. Caractéristiques des cosinus directeurs du rapport d'insolation	250
 REFERENCES	 251

## INTRODUCTION

La connaissance de la structure spatio-temporelle de la ressource énergétique solaire est indispensable aux besoins des utilisateurs du monde technologique (concepteurs de systèmes solaires par exemple) ou scientifique (nécessité d'initialiser ou de forcer des modèles météorologiques ou climatiques par exemple).

Ce travail porte sur l'étude de l'insolation à l'échelle synoptique (la France) à l'aide d'un ensemble de mesures de la durée quotidienne d'insolation réalisées au sol d'une part, et par télédétection satellitaire d'autre part. L'objectif poursuivi est d'analyser une série de données suffisamment longue pour caractériser de manière statistique le comportement climatologique du phénomène dans l'espace. Les applications de cette analyse sont diverses : cartographie ou simulation de champs, optimisation de réseaux de mesures par exemple.

Les mesures du réseau sol sont précises mais ponctuelles, et leur interpolation en un point sans mesure nécessite la formulation d'hypothèses sur la structure du phénomène mesuré, qui sont parfois peu compatibles avec la réalité. C'est pourquoi l'utilisation de la télédétection et de sa forte définition spatiale est très attractive. En effet, le rayonnement solaire au sol est principalement dépendant de la réflexion du flux incident par les nuages, et ce flux réfléchi est bien perçu par les satellites.

Les données satellitaires de haute définition spatiale peuvent-elles, après traitement par un programme, compléter voire même se substituer efficacement au réseau de mesures sol dont elles n'atteignent pas la précision ponctuelle ?

La réponse à cette question nécessite un important jeu de données. Elle est fondamentale pour des continents tels l'Afrique ou l'Amérique du Sud par

exemple, qui ne disposent pas d'un réseau de mesures sol suffisant, mais qui sont bien couverts par les satellites. Par ailleurs, la question reste ouverte à nos latitudes pour lesquelles le test est alors possible grâce à un réseau sol plus fourni.

Depuis une dizaine d'années, de nombreuses méthodes ont été développées pour quantifier l'énergie solaire disponible au sol, à l'aide d'images de satellites. La plupart d'entre elles ont été testées sur une période inférieure au mois. Trois méthodes seulement (TARPLEY, GAUTHIER, PASTRE) ont traité un échantillon représentatif de l'ensemble des situations météorologiques rencontrées sur une année.

Pour notre part, nous avons utilisé le modèle CACTUS mis au point en 1980 par C. PASTRE et N. BERIOT au Centre de Météorologie Spatiale de Lannion. Ce modèle concerne l'évaluation de l'enuagement, qui est l'étape déterminante de toutes les méthodes de détermination du rayonnement solaire. D'un point de vue énergétique, le choix de cette variable se justifie en raison de sa forte liaison avec le rayonnement global. L'intérêt de cette méthode est double. D'une part, la simplicité de l'algorithme proposé (un simple seuillage de l'image visible) en fait une méthode opérationnelle et peu coûteuse, permettant la création d'une banque de données climatologiques (actuellement en routine à la Météorologie Nationale). D'autre part, les mesures de la durée d'insolation sont relativement plus abondantes que celles du rayonnement global et permettent ainsi une évaluation spatiale plus robuste de la méthode.

La démarche adoptée dans cette étude pour évaluer l'apport de la télédétection à la connaissance spatiale et climatologique du rapport d'insolation en France reste valable pour tout autre champ scalaire bidimensionnel.

Dans un premier temps, la variabilité spatiale du rapport d'insolation mesuré par les deux types de capteurs est analysée, ainsi que leur corrégionalisation. Cette étape est fondamentale pour l'orientation des suivantes.

Un protocole de validation des mesures de télédétection est ensuite proposé. Il comprend un test des mesures satellitaires ainsi que leur évaluation par rapport aux méthodes classiques d'interpolation, fondées sur le seul réseau sol. On veut ainsi évaluer l'apport de la télédétection à la connaissance du rapport d'insolation.

Nous proposons, enfin, une technique d'étalonnage des mesures de télédétection par combinaison avec les mesures sol dans le but de tirer parti de leur complémentarité.

Les méthodes d'analyse des champs spatiaux utilisées se rattachent, de manière générale, à la théorie des fonctions aléatoires. De par les hypothèses qu'elles formulent, elles prennent des formes différentes : analyse objective mono- et multi-variables, ou bien analyse de processus au moyen de fonctions orthogonales empiriques.

1ère PARTIE

DESCRIPTION DES DONNEES

## I - LA RESSOURCE ENERGETIQUE SOLAIRE

Le rayonnement solaire hors atmosphère est pratiquement constant, la distance terre-soleil étant son principal facteur de variation. Le rayonnement solaire arrivant au sol est, par contre, très variable : il dépend de la composition de l'atmosphère traversée. Cette ressource énergétique peut être assimilée à un "gisement solaire" dont l'exploitation nécessite la mesure. Nous allons en rappeler brièvement les traits principaux.

### I-1. Rayonnement hors atmosphère

Le soleil, situé à 150 millions de kilomètres de la terre, émet un rayonnement électromagnétique qui peut être assimilé, en première approximation, à celui d'un corps noir à 5800 K (Fig. I-1). Le spectre énergétique montre que 98 % de l'énergie émise se situe dans la bande de longueur d'onde comprise entre 0,2 et 3  $\mu\text{m}$ .

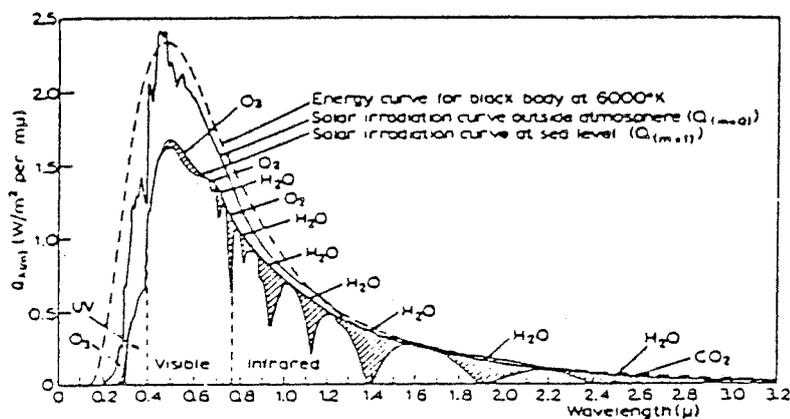


Figure I-1 Spectre solaire hors atmosphère comparé à celui d'un corps noir à 6000°K et à la distribution spectrale au niveau de la mer, pour un soleil au zénith. (Handbook of Geophysics, 1960)

La valeur moyenne de l'éclairement reçu par une surface perpendiculaire aux rayons et se situant à la limite de l'atmosphère est dite "constante solaire" et vaut environ 1370  $\text{w/m}^2$ . Cette valeur varie légèrement au cours de l'année, avec la distance terre-soleil ( $\pm 3.4$  %).

## I-2. Transmission à travers l'atmosphère

L'atmosphère est constituée de composantes fixes comme l'azote ou l'oxygène principalement, et de composants très variables comme le gaz carbonique, l'ozone, l'eau (sous forme de vapeur, gouttelettes ou cristaux), et les aérosols.

Les divers composants de l'atmosphère provoquent l'atténuation du rayonnement solaire par absorption et diffusion, ils donnent naissance à un rayonnement diffus. Le spectre du rayonnement solaire arrivant au sol est modifié à la fois en intensité et composition spectrale. Le sol enfin, réfléchit une partie du rayonnement qui sera à son tour absorbée et diffusée.

### I-2-a. Absorption sélective par les gaz

L'absorption due aux gaz affecte la composition spectrale du rayonnement atteignant le sol (cf. fig. I-1). Elle est due à l'ozone, l'oxygène, l'azote et la vapeur d'eau.

### I-2-b. Diffusion atmosphérique

La diffusion par une particule correspond à la réémission d'une onde de même fréquence, mais de direction différente. La répartition directionnelle du rayonnement diffusé dépend du rapport du diamètre de la particule à la longueur d'onde considérée. On distingue :

- la diffusion de RALEIGH par les particules de diamètre petit devant la longueur d'onde (molécules), qui affecte les radiations de courte longueur d'onde (on lui doit la couleur bleue du ciel),
- la diffusion de MIE par les particules de l'ordre de grandeur de la longueur d'onde incidente (aérosols, poussières, nuages). Elle dépend peu

de la longueur d'onde (couleur blanche des nuages, blanchâtre de la pollution).

### I-2-c. Réflexion

Elle peut être simple, par le sol ou par les nuages, ou multiple : le rayonnement réfléchi par le sol est renvoyé par les nuages vers la terre, etc... L'albédo d'une surface est le rapport de la puissance réfléchie à la puissance incidente. Il est fonction à la fois de la longueur d'onde incidente, de l'angle d'incidence et de la nature de la surface. L'albédo moyen de la terre est de 30 %.

### I-3. Rayonnement solaire atteignant le sol

On le décompose comme la somme du rayonnement direct I qui provient du disque solaire et qui a subi l'atténuation due à la traversée de l'atmosphère, et du rayonnement diffus D qui provient de la voûte céleste (RAYLEIGH), des aérosols, des nuages (MIE) et du sol (réflexion multiple). Le rayonnement global atteignant une surface s'écrit :

$$G = I \sin h + D$$

où h est l'angle entre la normale à la surface et la direction du soleil.

Le rayonnement solaire parvenant au sol par ciel clair est essentiellement fonction de la hauteur du soleil h, connue de façon déterministe et, au second ordre, des contenus en gaz et en aérosols. Environ 80 % du rayonnement extra-atmosphérique parvient au sol lorsque le soleil est au zénith. Les fluctuations autour de cette moyenne, liées surtout aux modifications des contenus en vapeur d'eau et aérosols, restent faibles, de l'ordre de 5 % (DEDIEU, 1984).

Par ciel couvert, la réflexion par les nuages est le phénomène majeur dans l'extinction du rayonnement atteignant le sol (leur albédo moyen est de 50%

, atteignant parfois 90 %). L'absorption par les nuages intervient au second ordre.

## II - MESURES DU RAYONNEMENT SOLAIRE AU SOL

Les mesures effectuées par les Services Météorologiques sont celles du rayonnement global sur un plan horizontal, du direct, du diffus, de la durée d'insolation et de la nébulosité. Ces deux dernières variables ne mesurent pas directement le rayonnement mais lui sont liées comme nous le verrons plus loin.

### II-1. Le rayonnement direct

Il est mesuré par un pyréliomètre : une thermopile est placée au fond d'un tube, qui suit le soleil dans sa course journalière. La position du tube doit être vérifiée tous les jours. Le pyréliomètre doit être périodiquement étalonné.

### II-2. Le rayonnement global

Il est mesuré par un pyranomètre : un thermocouple mesure la différence de température entre un corps noir et une surface réfléchissante blanche (ou la température de l'air). Cette différence est proportionnelle au flux global incident. Le domaine spectral est de 0,3 à 2,8  $\mu\text{m}$ . Sa sensibilité varie avec la hauteur du soleil. Il doit être régulièrement étalonné et entretenu (poussière, condensation).

### II-3. Le rayonnement diffus

Il est mesuré par un pyranomètre associé à une bande pare-soleil qui occulte le rayonnement direct. Cette bande doit être positionnée tous les jours. La présence de cette bande, qui cache une partie du ciel et donc du diffus, implique de corriger la mesure.

### II-4. La durée d'insolation

La durée d'insolation journalière est la durée pendant laquelle le rayonnement solaire direct dépasse  $120 \text{ w/m}^2$  (norme internationale, fixée en 1981, à Mexico). Ce seuil correspond environ à l'éclairement suffisant pour voir les ombres portées au sol.

La durée d'insolation est mesurée à l'aide d'un héliographe. L'appareil le plus couramment utilisé est l'héliographe CAMPBELL-STOKES : une boule de verre focalise le rayonnement solaire sur un diagramme cartonné, gradué en heures, laissant sur ce carton une trace de décoloration ( $\approx 120 \text{ w/m}^2$ ) ou de brûlure ( $\approx 250 \text{ w/m}^2$ ), suivant l'intensité de l'éclairement solaire. Le papier est changé tous les jours, la durée d'insolation est déduite de la longueur  $L$  de la trace. La mesure est donnée en  $1/10^{\text{ème}}$  d'heure.

Cet appareil, fort simple, a l'avantage d'être peu coûteux et de nécessiter peu d'entretien. Par contre, le dépouillement des mesures s'avère délicat. En particulier, lors de passages nuageux rapides, la longueur mesurée  $L$  aura tendance à surestimer la durée d'insolation réelle, l'occultation du soleil étant trop brève pour se distinguer nettement sur la trace. L'expérience de l'opérateur intervient alors. La précision de la mesure est donc liée au type de journée : type de nuages rencontrés et leur vitesse de déplacement. On considère (COUDERT, 1984) que pour une intégration sur une décade (dix jours), les erreurs ne dépassent pas 5 %, mais elles peuvent être beaucoup plus importantes pour la journée.

## II-5. La nébulosité

La nébulosité est la portion de voûte céleste couverte par les nuages. La nébulosité intéresse le rayonnement solaire dans la mesure où les nuages sont le principal facteur de son atténuation. La nébulosité est mesurée visuellement par un météorologue, elle s'exprime en octa : 0 correspond à un ciel dégagé, 8 à un ciel totalement couvert.

## II-6. Réseaux de mesure en France

L'ensemble des mesures de la Météorologie Nationale en France métropolitaine est résumé dans le tableau suivant.

variable	instrument de mesure	(1)	(2)	(3)	(4)
ray <sup>t</sup> . direct	pyrhéliomètre	6 mn	2	1971	± 2%
ray <sup>t</sup> . global	pyranomètre	heure	31	1962-76	± 3%
ray <sup>t</sup> . diffus	pyranomètre + bande pare-soleil	heure	12	1967-78	± 10%
durée d'insol.	héliographe	jour/2	91	1951	
nébulosité	observateur	3 heures	100		

- (1) : Période de mesure  
 (2) : Nombre d'appareils de la météorologie nationale  
 (3) : Années d'implantation  
 (4) : Précision au pas de temps journalier

Tableau I-1 : d'après "Energie solaire : catalogue des données disponibles" du service Météorologique Métropolitain, 6/12/82.

Leur répartition géographique est donnée figure I-2. On note une assez bonne répartition spatiale des stations; cependant, certaines régions sont moins bien instrumentées comme, par exemple, le Massif Central. Ce sous équipement est dû au fait que les stations se trouvent le plus souvent dans les aéroports des grandes villes. A titre de comparaison, il est

intéressant de noter que le réseau climatologique national est constitué de 3500 pluviomètres. La densité spatiale bien supérieure des mesures de pluie est due à sa plus grande variabilité spatiale, mais aussi aux intérêts économiques qu'elle a mis en jeu très tôt. Les mesures de l'énergie solaire sont récentes : 1951 pour les héliographes, 1970 seulement pour les radiomètres.

Le récent intérêt pour la mesure du rayonnement solaire pose le problème de la longueur des séries de données (pour une étude climatologique) et de la faible densité spatiale des mesures (pour l'interpolation à une zone non instrumentée). L'interpolation du rayonnement direct (deux stations de mesure en France) et même du diffus (32 stations) est difficile à partir des seules données sol. Toutes les études traitant ces données fournissent actuellement uniquement des résultats locaux. C'est pourquoi, les mesures indirectes (héliographes ou télédétection) ont un grand intérêt dans l'évaluation de la ressource énergétique, pour la connaissance de sa variabilité spatiale.

#### II-7. Liaison entre mesures énergétiques et mesures plus simples

La ressource énergétique n'est mesurée qu'en un nombre très limité de points. La durée d'insolation et la nébulosité en revanche sont mesurées en de nombreuses stations, et depuis beaucoup plus longtemps. De nombreux auteurs se sont attachés à tirer parti des mesures indirectes, en étudiant leur liaison avec les grandeurs énergétiques. Nous rappelons ici brièvement les relations établies entre la durée d'insolation et le rayonnement global journalier.

##### II-7-a. Liaison entre rayonnement global et fraction d'insolation

La fraction d'insolation  $SS_0$  est le rapport de la durée d'insolation  $S$ , mesurée avec un héliographe, à la durée astronomique du jour  $S_0$  éventuellement corrigée de l'effet de masque dû aux obstacles entourant l'héliographe.

$$SS_0 = \frac{S}{S_0}$$

L'estimation du rayonnement global journalier  $G$  se fait à l'aide de formules de régression linéaire. Nous citerons pour mémoire deux types de relation :

$$\text{- Black} \quad \frac{G}{G_0} = a + b SS_0$$

$G_0$  est l'irradiation globale journalière calculée hors atmosphère. Elle se calcule à partir de la constante solaire  $I_0$ , du lieu et de l'heure. Les coefficients  $a$  et  $b$  varient avec le lieu et la saison.

$$\text{- Angström} \quad \frac{G}{G_1} = a' + b' SS_0$$

$G_1$  est l'irradiation globale journalière estimée par beau temps. Son calcul nécessite l'introduction de données locales concernant la vapeur d'eau et les aérosols notamment. Les coefficients  $a'$  et  $b'$  sont moins variables que  $a$  et  $b$  mais la précision sur  $G$  est liée au bon choix de  $G_1$ .

Il existe de nombreuses autres relations, certaines non linéaires (SCHUEPP, P. de BRICHAMBAUD). Ces équations de régression ont été testées localement (voir, par exemple, sur figure I-3, les résultats à Carpentras). On constate que le global journalier est bien reconstitué à partir de la fraction d'insolation, quelle que soit la relation utilisée. Cependant, pour les valeurs faibles de la fraction d'insolation ( $<0.2$ ), les valeurs de rayonnement global ont tendance à être surestimées.

## II-7-b. Liaison entre rayonnement global et nébulosité

Certaines relations ont été étudiées, nécessitant en outre l'utilisation de la masse atmosphérique (la masse d'air traversée par un rayon solaire).

Fig I-2 (d'après Choissnel et Bedel, 1984)

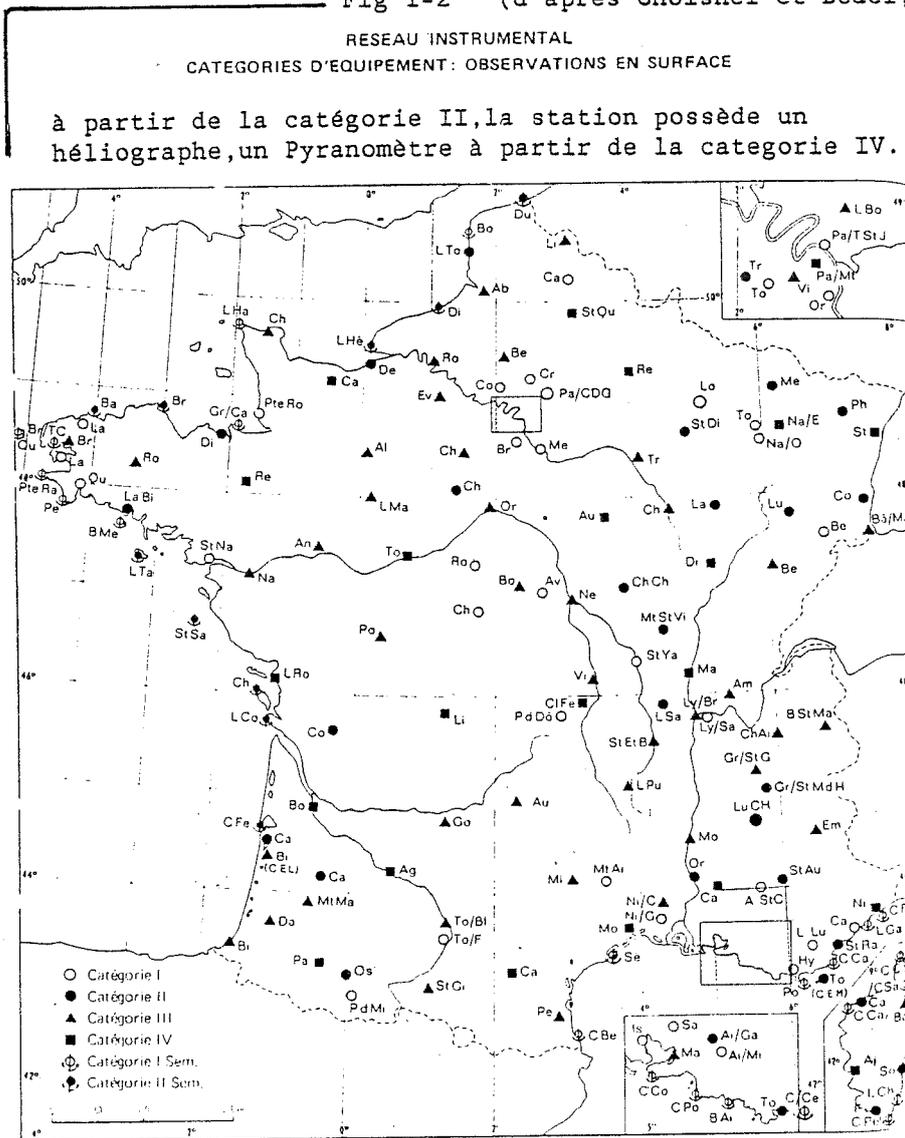
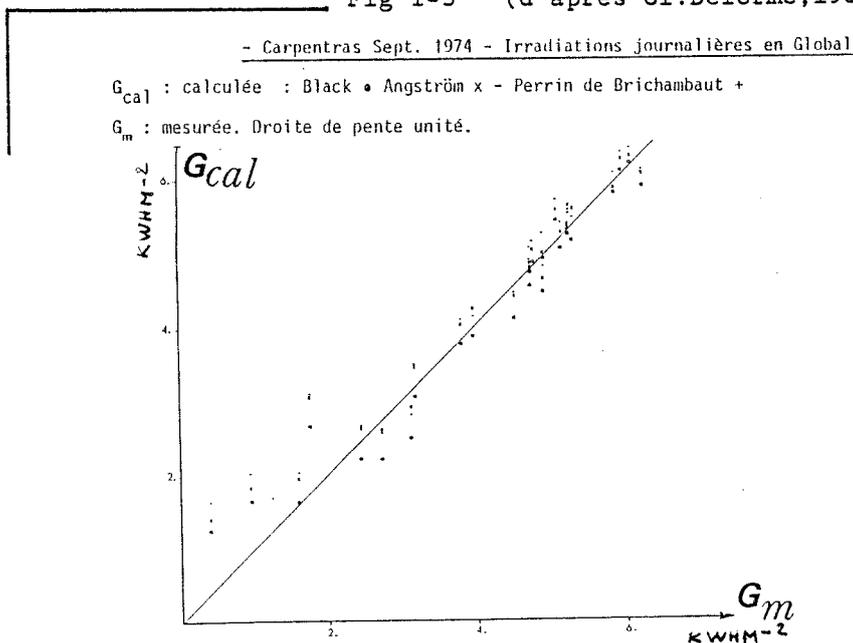


Fig I-3 (d'après Cl. Delorme, 1984)



### III - MESURES INDIRECTES PAR TELEDETECTION

Les satellites météorologiques apportent une information bien répartie et fréquente sur l'état de l'atmosphère. Ils semblent particulièrement aptes à évaluer l'enneuagement, facteur principal du rayonnement solaire au sol.

Après quelques rappels sur les caractéristiques des satellites, les modèles existants pour l'évaluation de l'irradiation solaire, leur principe et les résultats obtenus seront présentés. Les objectifs et le principe de la méthode CACTUS du Centre de Météorologie Spatiale de Lannion, utilisée et évaluée dans cette étude, sont ensuite plus particulièrement détaillés.

#### III-1. Mesures fournies par les satellites

Le principe de l'image satellitaire est le suivant : un radiomètre est embarqué à bord d'un satellite et fournit une image par balayage : chaque élément d'image (pixel) est le résultat d'une mesure du rayonnement émis dans l'angle de visée du capteur. Cette image est fonction de la gamme de longueur d'onde utile du radiomètre. Pour couvrir plusieurs gammes de longueur d'onde, un satellite peut posséder plusieurs radiomètres, chacun étant destiné à un usage particulier.

Il existe de nombreux types de satellite ; on les caractérise par leur orbite et leur équipement radiométrique. Chaque type correspond à un besoin précis.

##### III-1-a. Satellite à défilement

Ils passent périodiquement au-dessus d'une même région. Un satellite à orbite basse a une résolution spatiale élevée (jusqu'à 10 m pour SPOT) mais

leur fréquence de passage au-dessus d'une même région est faible (de deux images par jour à une image tous les 18 jours pour LANDSAT par exemple). Si on veut une bonne résolution temporelle, il faut utiliser plusieurs satellites, ce qui pose le problème de la superposabilité des images (angle de prise de vue, sensibilité des radiomètres). Ces satellites sont utilisés en sciences de la terre plutôt qu'en météorologie.

### III-1-b. Satellites géostationnaires

Ils se déplacent à la vitesse nécessaire pour voir toujours la terre sous le même angle. La physique impose qu'ils soient dans le plan de l'équateur à une distance de 36 000 km. Ils sont donc très éloignés de la terre. Cependant, leur résolution et leur fréquence de prise de vue est utile en météorologie (pour le suivi des perturbations atmosphériques par exemple).

Un réseau de cinq satellites géostationnaire permet d'observer l'ensemble de la terre. C'est sur leurs observations que sont fondées la plupart des méthodes de reconstitution du rayonnement. Nous allons tout d'abord détailler plus particulièrement l'un d'entre eux, Météosat qui couvre l'Afrique et l'Europe, et dont les caractéristiques sont très proches des autres satellites géostationnaires.

### III-1-c. Météosat II

Ce satellite est situé à la verticale de l'équateur et du méridien de Greenwich. Son champ de vision est donné figure I-4. Il a été lancé en juin 1981, et fait partie du programme européen d'exploitation des satellites météorologiques (EUMETSAT), en tant que système préopérationnel. Le lancement du premier satellite opérationnel (MOP1) est prévu en août 1987 et sera suivi de deux autres satellites (MOP2 et MOP3) qui devront assurer la couverture météorologique satellitaire jusqu'en 1995 au minimum.

### Bandes spectrales mesurées par Météosat

Météosat est muni d'un radiomètre à télescope qui permet de mesurer la radiance dans trois bandes spectrales (voir tableau I-1). On remarque que le canal invisible (0.4-1.1  $\mu$ m) correspond à une bande spectrale inférieure à celle du spectre solaire, mais où la densité énergétique solaire est forte (cf fig. I-1).

### Résolution spatiale et temporelle de Météosat

Météosat fournit toutes les 30 mn, soit un jeu de deux images (Vis 5000 lignes et IR thermique 2500 lignes), soit un jeu de trois images (Vis, IR thermique et IR vapeur d'eau) de 2500 lignes chacune. On note ici que deux pixels provenant de la même "image" peuvent avoir 25 mn de décalage dans le temps.

La résolution de l'image satellite dépend du canal et de l'inclinaison de la visée (cf fig I-5). La résolution maximale est celle du point sous satellite :

- 2.5 km pour l'image 5000 lignes
- 5 km pour l'image 2500 lignes.

En France (45° de latitude), la résolution spatiale est d'environ 4.5 km pour l'image 5000 lignes.

### Superposabilité des images

Ces images de radiance sont transmises sous forme numérique (codée de 0 à 255) à la station de Darmstadt (R.F.A.) où elles subissent un traitement destiné à corriger les défauts dus au système de génération de l'image et à la dynamique du satellite. Ces images sont ensuite renvoyées, par l'intermédiaire de Météosat aux différents utilisateurs. Celles-ci ne sont corrigées des défauts dus à l'imperfection de stationnarité du satellite (dites erreurs de navigation), que depuis 1986. Les images de la période antérieure doivent donc subir un autre prétraitement dans le but de superposer des images prises à des instants différents.

Fig I-4 :

Champs de vision du satellite Météosat.

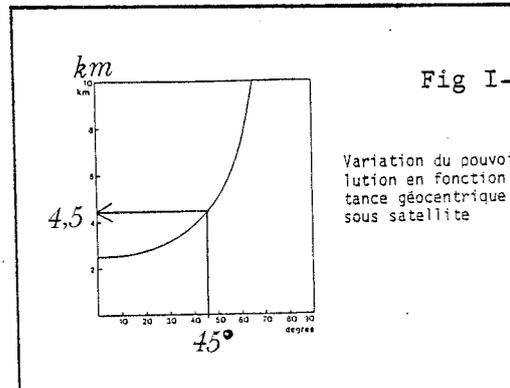
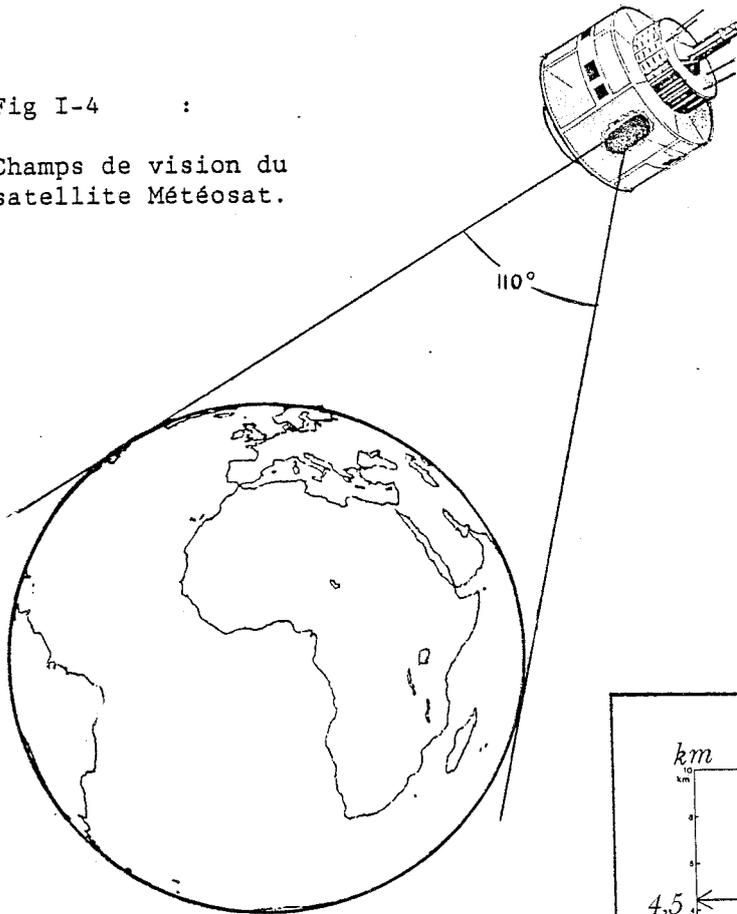


Fig I-5

Variation du pouvoir de résolution en fonction de la distance géocentrique au point sous satellite

Tableau I-1 : Bandes spectrales des radiomètres de Météosat.

Spectral Bands	Visible 0.4-11 $\mu\text{m}$	IR (Water vapour) 5.7-7.1 $\mu\text{m}$	IR (Window) 10.5-12.5 $\mu\text{m}$
Number of channels	2 (Simultaneous)	1 (in time-sharing with 1 vis-channel)	1 (+ 1 redundant)
Number of lines, pict.	5000(2500)*	2500	2500
Number of samples line	5000	2500	2500
Resolution (subsat. point)	2.5 km	5 km	5 km
Line duration		30 ms	
Line recurrence		600 ms	
Image-taking duration		25 min	
Image recurrence		30 min	
Transmission DATTS	Digital 166 kbit s (normal) 2.7 Mbit s (back-up)		

\* Case in which water-vapour channel is also transmitted.

(d'après "Introduction to the Meteosat system")

### III-2. Quelques méthodes d'évaluation de l'insolation à partir de mesures de satellites

Depuis 1971, de nombreux auteurs ont cherché à estimer l'énergie solaire à partir des images de satellite. Les études réalisées s'intéressent essentiellement au rayonnement global, qui intervient dans les calculs de bilan d'énergie. Celui-ci est estimé pour chaque image, puis intégré pour obtenir les moyennes journalières, mensuelles ou annuelles.

Nous avons porté une attention particulière à huit méthodes qui ont été testées sur un volume de données important (supérieur au mois). Nous n'avons pas retenu les publications portant sur le traitement de données photographiques. Celles-ci (M. REZKI, 1982, par exemple) ont démontré la possibilité de l'utilisation des images satellitaires à la détermination de l'insolation sur un grand volume (2500 images) de données (fig.I-6-h). Le traitement manuel qu'elles impliquent ne les rend pas compétitives avec le traitement informatique des images numériques.

Les méthodes décrites dans la littérature sont généralement classées en deux catégories : statistiques (J.S. ELLIS et T.H. VONDER HAAR, 1978 ; J.D. TARPLEY, 1979 ; C. PASTRE, 1981 ; D. CANO et J.M. MONGET, 1982, par exemple) ou reposant sur une modélisation physique (K. HANSON, 1971 ; C. GAUTHIER, G. DIAK et S. MASSE, 1980, W. MOSER et E. RASCHKE, 1981 ; J. AMADO et Cl. DELORME, 1982 ; G. DEDIEU, P.Y. DESCHAMPS et Y.H. KERR, 1983, par exemple). Par la suite, on désignera une méthode par le nom de son premier auteur.

#### Satellites utilisés

Les premières méthodes, mises au point aux Etats-Unis (HANSON, ELLIS) ont utilisé des données provenant du satellite défilant NOAA (deux images par jour). Leurs résultats ont conduit à travailler avec des satellites géostationnaires (METEOSAT ou GOES) pour prendre en compte la variabilité temporelle du phénomène qui est indispensable à la qualité de

estimation. Une des différences essentielles entre METEOSAT et GOES (EST et OUEST), qui couvrent les Etats-Unis, est la largeur de la bande spectrale visible qui est plus fine pour les GOES (0.55 à 0.75  $\mu$ m).

#### Calibration du radiomètre

Afin de relier directement les comptes numériques fournis par le satellite à une luminance (énergie), certains auteurs utilisent des coefficients de calibration obtenus par référence à deux valeurs extrêmes de l'image (REZKI, GAUTHIER, PASTRE), ou par référence à des moyennes (AMADO) ou à des valeurs sol (DEDIEU). Les autres prennent en compte les variations du radiomètre par normalisation des mesures avec une brillance minimum (TARPLEY, MOSER) déterminée sur une série d'images.

#### Détermination d'un seuil de brillance

Un seuil de brillance est généralement utilisé (TARPLEY, PASTRE, GAUTHIER, MOSER, AMADO) pour classer tout d'abord chaque pixel comme clair ou couvert. Cette distinction vaut pour les méthodes "statistiques" ou "physiques". Suivant la valeur de luminance, on détermine ensuite l'équation de régression, ou le modèle physique adapté au pixel. Seuls CANO et DEDIEU s'affranchissent de la détermination d'un seuil. Le choix du seuil est fait de façon absolue, avec prise en compte de la seule hauteur du soleil (PASTRE), ou par référence à l'albédo de surface de chaque heure pour les autres. La connaissance de l'albédo nécessite la connaissance de l'absorption atmosphérique par ciel clair (GAUTHIER, DEDIEU, MOSER, AMADO, CANO). Cette modélisation est assez délicate, notamment en cas de présence d'aérosols (P. RABERANTO, 1984, a étudié ce seul point dans sa thèse).

#### Modélisation physique

Deux cas sont distingués :

#### - Modèle ciel clair

Le principal facteur d'atténuation du rayonnement solaire par ciel clair est l'angle zénithal solaire. L'absorption par les composants fixes, du second ordre, est prise dans la littérature (GAUTHIER, MOSER, DELORME, DEDIEU). L'absorption par la vapeur d'eau (de teneur variable) est estimée par régression à partir du point de rosée (GAUTHIER) ou de la masse atmosphérique (MOSER) ou des pyranomètres (DEDIEU). L'absorption par les aérosols est en général négligée. Elle est cependant prise en compte implicitement à travers l'albédo de surface (teneur en aérosols minimal). DELORME a tenté d'étudier les aérosols avec les mesures IR.

Les résultats obtenus par ciel clair sont généralement très bons (de l'ordre de 5 % d'erreur), mais on a vu que l'ordre de grandeur des fluctuations de la luminance solaire par ciel clair est lui-même de 5 %. L'apport de la télédétection par rapport à la simple utilisation de la valeur moyenne n' est pas significative.

#### - Modèle ciel nuageux

Le cas du ciel nuageux est beaucoup plus délicat. Le principal facteur d'atténuation est alors la réflexion et la diffusion par les nuages (jusqu'à 70 % de réflexion), qui sont bien saisies par les satellites ; l'absorption (qui atténue de 10 % à 20 % du rayonnement incident) est par contre difficile à estimer, car elle dépend de l'épaisseur optique du nuage, sur laquelle on n' a pas d' information. Pour un ciel nuageux, l'absorption atmosphérique se fait au-dessus et au-dessous du nuage dans des proportions variables et le satellite ne saisit pas cette dernière. GAUTHIER fait l'hypothèse que l'épaisseur du nuage est proportionnelle à sa luminance, MOSER teste leur hauteur avec le canal IR, DEDIEU néglige ce problème. L'absorption par la vapeur d'eau est toujours négligée.

#### Modélisation statistique

Les auteurs présentant des méthodes statistiques jugent peu réalisable la modélisation des phénomènes de diffusion et d'absorption du rayonnement solaire par ciel clair et surtout dans les nuages. La relation pratiquement linéaire existant entre la nébulosité (ou la durée d'insolation qui est son

complémentaire) et le rayonnement global est ici mise à profit. Les coefficients d'une (CANO) ou plusieurs (TARPLEY) équations de régressions sont ajustés à l'aide de pyranomètres. Ces équations sont fonction d'un indice de nébulosité continu (CANO) ou discret (TARPLEY). AMADO utilise une modélisation physique pour les ciels clairs et une approche statistique pour les ciels nuageux en séparant les nuages en quatre classes affectées d'un coefficient de transmission.

### Résultats

Les principaux résultats sont présentés dans le tableau I-2. Les nuages de corrélation les plus récents ou portant sur le plus grand nombre d'évènements ont également été reportés (fig. I-6). L'observation de ces nuages et de leur dispersion complète les valeurs de l'erreur RMS\*\*.

Certaines de ces méthodes sont actuellement opérationnelles, comme celle de TARPLEY au NESDIS\*1 de Washington (le nuage de corrélation pour un an de données (1983) et 13 pyranomètres est représenté fig. I-6-b) ; elle est utilisée pour les besoins agronomiques ; la méthode de GAUTHIER est utilisée depuis 1984 au SSEC\*2 du Wisconsin ; la méthode de PASTRE est opérationnelle au CMS\*3 de Lannion depuis 1984, l'analyse de ces données fait l'objet de cette thèse ; la méthode de CANO enfin a été utilisée au CTAMN\*4 de Valbonne en 1983 pour fournir des cartes du rayonnement mensuel. La plupart des méthodes opérationnelles ont fourni des cartes du rayonnement mensuel qui n'ont cependant pas pu être testées, mais seulement comparées à la climatologie connue.

---

\*\* : Relative Mean Square (erreur quadratique moyennne)

\*1 : National Environmental Satellite Data and Information Service

\*2 : Space Science and Engineering Center

\*3 : Centre de Météorologie Spatiale

\*4 : Centre de Télédétection et Analyse des Milieux Naturels

Tableau I-2 : (d'après N.Bériot;1984)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ellis, Vonder Haar (78)	NOAA-4		10-15%	10 x 10	USA	juin 75	1/jour
Tarpley (79)	GOES-E	10-30-50%	5-10%	50 x 50	USA	été 77	7/jour
"	"		10%	"	"	"	2/jour
"	"		20%	"	"	"	1/jour
Justus, Tarpley (83)	GOES-E	16%	10%	40 x 40	USA		7/jour
Gautier, Diak, Masse (80)	GOES-E			1 pixel	Canada	11j78/79	1/heure
"	"		9%	20 x 20	"	"	"
Diak, Gautier, Masse (82)	GOES-E		9%		Océan Tropical	6j GATE	1/heure
Möser, Raschke (83)	Météosat	19%		25 x 25	Europe, Afrique	juin 79	6/jour
Möser (83)	Météosat		7%	25 x 25	Europe, Afrique	juin 79	6/jour
Dedieu, Deschamps, Kerr (83)	Météosat	21%		160x120	Europe	été 79	1/heure
"	"	15%		40 x 30	"	"	"

Résultats de quelques méthodes d'évaluation de l'irradiation solaire globale

- (1) noms des auteurs
- (2) satellite
- (3) erreur RMS sur l'irradiation globale horaire
- (4) erreur RMS sur l'irradiation globale journalière
- (5) résolution spatiale des évaluations de l'irradiation en km x km
- (6) zone géographique
- (7) période
- (8) nombre d'images traitées par jour ou par heure

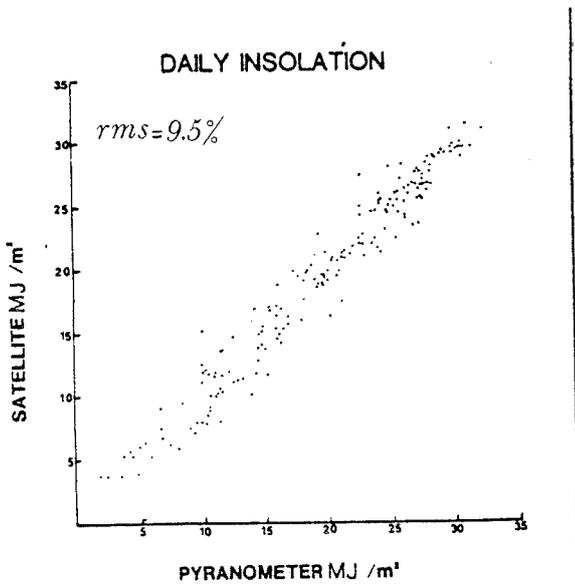
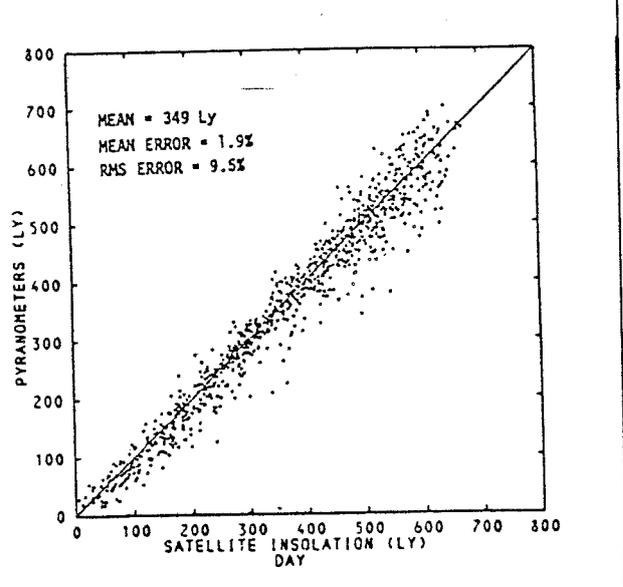


Fig. 1. Daily insolation scatter plot of satellite estimates vs. pyranometer measurements.

90 jours x3 pyranomètres  
(d'après Gauthier;1982)

a



228 jours x13 pyranomètres  
(d'après Tarpley;1984)

b

Fig I-6 :

Différentes estimations du rayonnement global :

- (a) journalier
- (b) et (c) mensuel

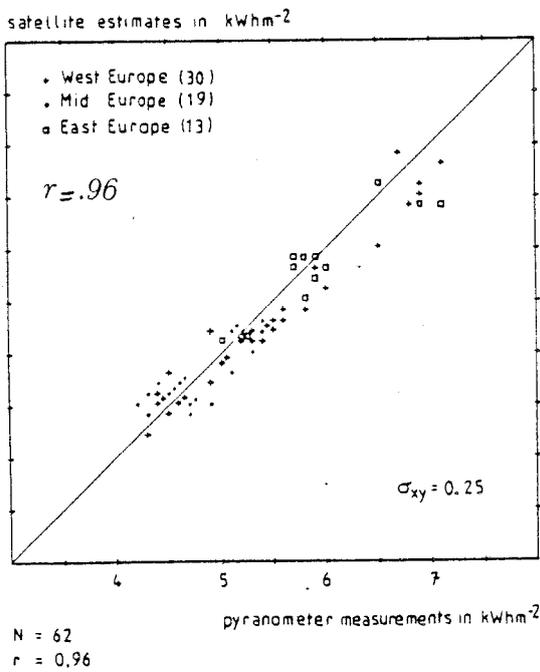


FIG. 6. Comparison of monthly averages of daily sums of global radiation as obtained from pyranometer stations with the estimates derived from METEOSAT I data for June 1979.

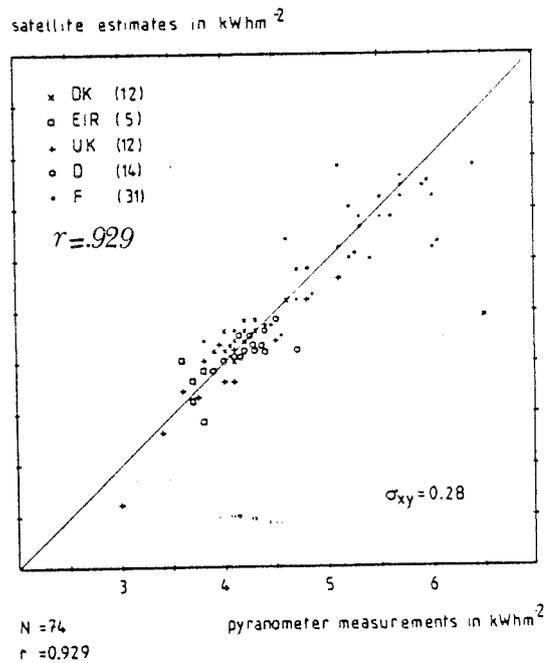


FIG. 7. As in Fig. 6, except that METEOSAT II data for April 1982 have been used. (The symbols for the various countries are identified in Table 1.)

(d'après Möser,1984)

c

été 1977  
(d'après Tarpley;1979)

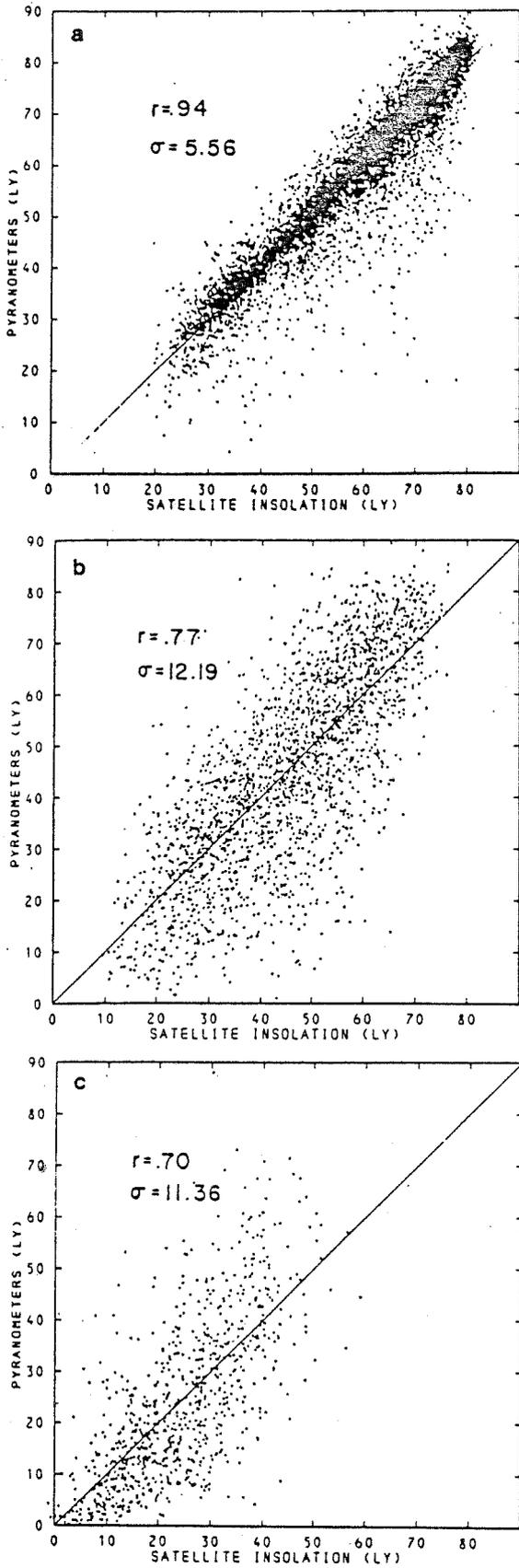


Fig. 4. Hourly insolation measured by the 22 pyranometers plotted against satellite estimates (dependent data) for 5736 clear cases ( $n < 0.4$ ), a, 2217 partly cloudy cases ( $0.4 \leq n < 1.0$ ), b, and 822 totally cloudy cases ( $n = 1$ ). c. Units are langleyes. If there were no errors, all points would lie on the diagonal line.

d

Fig I-6 (suite): valeurs horaires

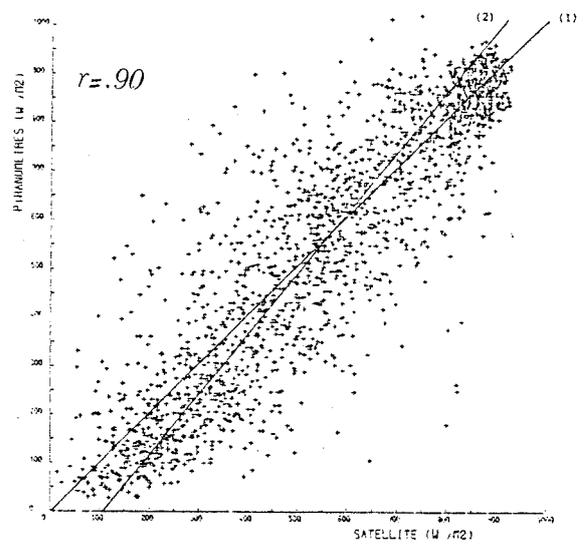


Figure 32. Comparaison des estimations satellitaires du rayonnement global aux mesures des pyranomètres (12h00-13h00 TU) pour les données des mois de mars, mai et juillet 1979. Mailles de 5 x 5 pixels (40 x 30 km<sup>2</sup>). Unité : W.m<sup>-2</sup>.  
corrélation : 0.90  
écart type : 121 W.m<sup>-2</sup> (21% de la moyenne)  
mesure = 1 + 1. estimation (droite 1)  
estimation = 108 + .805 mesure (droite 2)

100 heures (midi) x  
27 pyranomètres  
(d'après Dedieu;1984)

e

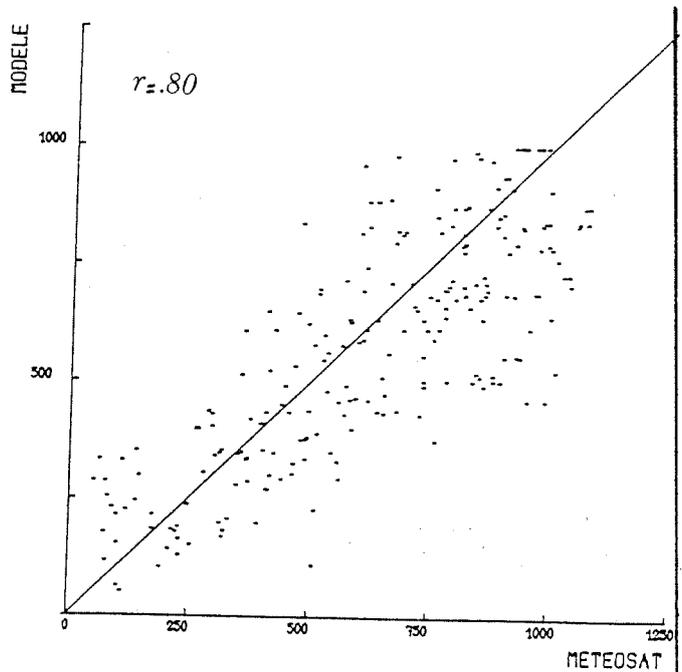


Fig. III.17.- Nuage de points pour les valeurs horaires des 11 stations

21 heures x 11 rayonnements  
horaires reconstitués  
(d'après Amado;1982)

f

Fig I-6 (suite): valeurs horaires

COURBES RAYONNEMENT ESTIME/RAYONNEMENT MESURE

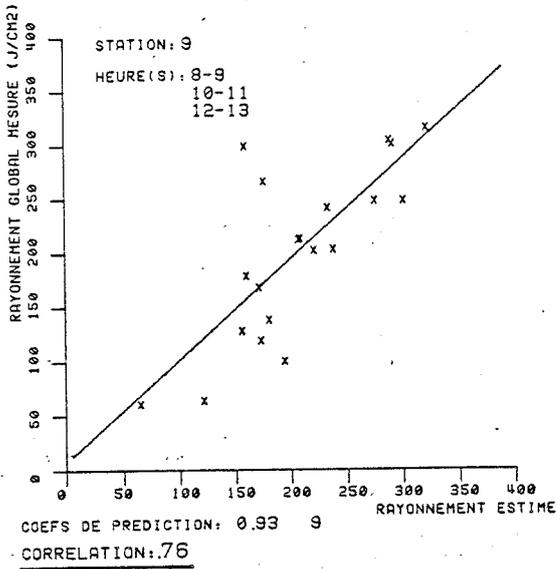


Figure III<sub>15</sub>  
Dijon

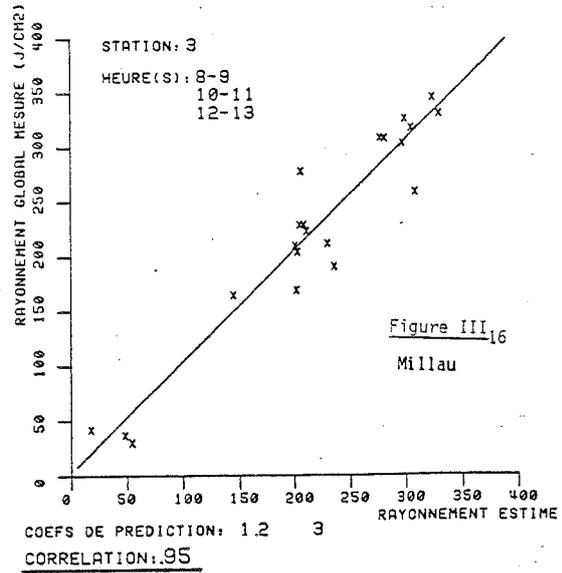
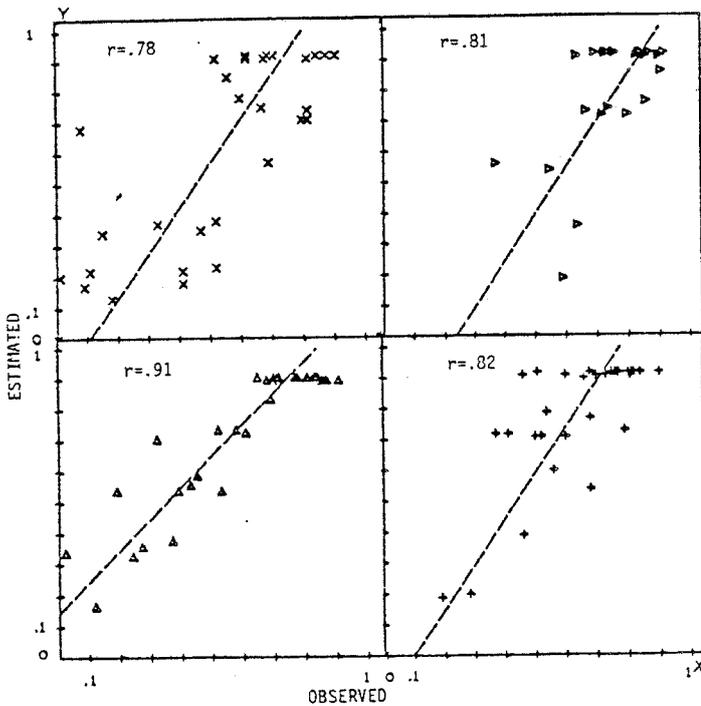


Figure III<sub>16</sub>  
Millau

Résultats pour 2 pyranomètres (Dijon, Millau) x 24 images horaires (mai 79). La corrélation moyenne, pour 27 pyranomètres est de  $r = .87$  (d'après Cano; 1982)

g

DAILY SUNSHINE DURATION



Autres Méthodes :

(d'après Rezki; 1982)

septembre 1979

- x Trappes
- ▷ Carpentras
- ▲ Alger
- + Annaba

4 Héliographes x 1mois

h

En résumé, le rayonnement global est bien estimé par les différentes méthodes puisque l'erreur RMS se situe vers 10 % pour les valeurs journalières, et 15 à 20 % pour les valeurs horaires qui sont plus délicates à estimer. Le satellite a tendance à surestimer le rayonnement global dans le cas nuageux (TARPLEY, GAUTHIER), principalement à cause de l'absorption par les nuages, mal saisie par le canal VIS seul. Les aérosols sont, eux aussi, peu accessibles par la mesure satellitaire. Seule, leur diffusion est négligée.

Le critère d'erreur généralement retenu pour juger de la qualité des méthodes est uniquement l'erreur RMS, qui n'indique pas si il y a co-fluctuation avec les valeurs sol, contrairement au coefficient de corrélation. C'est pourquoi, nous avons reproduit ici les nuages de corrélation (dans la mesure où ils sont fournis). Deux autres remarques peuvent également être ajoutées à propos des résultats. Les tests ont pour la plupart été effectués uniquement en été (mis à part GAUTHIER et TARPLEY), durant lequel les situations de ciel clair, favorables à l'estimation satellitaire sont plus fréquentes. Cet échantillonnage induit une surestimation de la précision. Enfin, il est parfois assez difficile de déterminer si l'échantillon choisi est indépendant de celui sur lequel la méthode a été mise au point.

#### Principales difficultés rencontrées dans l'utilisation d'images satellitaires

Les difficultés liées à l'utilisation d'images satellitaires sont de trois ordres. La première est l'étalonnage du radiomètre dont nous avons discuté plus haut. La deuxième est la localisation géographique précise des mesures. Quel que soit le degré de sophistication apporté à la navigation, tous les auteurs utilisent un contrôle final visuel de la position des images satellitaires. Ce problème n'est donc pas parfaitement résolu. La dernière enfin concerne la validation des modèles. La comparaison des estimations satellitaires avec les mesures des pyranomètres est délicate. Le radiomètre du satellite donne une valeur pratiquement instantanée du

rayonnement et dans une gamme de longueur d'onde réduite. Le pyranomètre, au contraire, fournit une valeur intégrée sur une heure dans tout l'angle solide et tout le spectre du rayonnement solaire. Pour tenir compte de la différence de temps et d'angle, la plupart des auteurs font une somme spatiale des estimations satellitaires sur des mailles allant de 20x20 km<sup>2</sup> (PASTRE) à 50x50 km<sup>2</sup> (TARPLEY). La taille des mailles est fonction de la période temporelle qu'elles sont sensées représenter (plus celle-ci est longue, plus la maille devra être grande).

La sensibilité de la méthode au nombre d'images par jour, a été testée pour une maille constante, (TARPLEY,79, REKI,82) et pour une maille variable (GAUTHIER,84). La correction de la brillance du satellite sur toute la bande du flux solaire par un modèle de transfert radiatif n'est en général pas faite (le canal VIS concerne la majeure part de l'énergie solaire) ; elle est prise en compte dans le coefficient de calibration du canal visible (DEDIEU).

La méthodologie de validation est en général peu détaillée par les auteurs (nombres de journées testées, type de temps correspondant, indépendance des valeurs sol, choix des critères), si on la compare au soin apporté à la description et au développement de la méthode.

### Conclusion

Le rayonnement solaire global atteignant le sol est estimé par différentes méthodes qui se distinguent plus par leur facilité de mise en oeuvre et leur possibilité de généralisation à une zone différente que par la qualité de leurs résultats qui est assez cohérente.

L'application d'une méthode à une zone différente implique le réajustement des coefficients de régression à partir de pyranomètres pour les méthodes statistiques. Pour les méthodes physiques, outre l'utilisation de nouveaux paramètres climatologiques, l'utilisation de certaines relations concernant l'absorption par la vapeur d'eau ou par les nuages nécessite également l'utilisation de pyranomètres. C'est le cas pour GAUTHIER, MOSER, DEDIEU. La différence entre les méthodes statistiques et physiques paraît alors assez floue.

La simplicité et la rapidité de calcul apparaît comme un critère de choix de la méthode. L'estimation de la nébulosité est une étape fondamentale du traitement et a été testée par peu d'auteurs (CANO, PASTRE). C'est pourquoi, la méthode de PASTRE, qui estime la nébulosité et qui est extrêmement simple puisqu'elle possède un seuil de brillance absolu (qui ne nécessite pas l'estimation de l'albédo ni de l'absorption atmosphérique) est très intéressante. Sa mise en place opérationnelle par la Météorologie Nationale permet de plus de disposer d'un important jeu de données et d'envisager de constituer une série climatologique.

Dans un premier temps, la validation du modèle d'estimation de la nébulosité pourra se faire sur l'ensemble des données disponibles, couvrant tous les types de temps rencontrés en France. Dans un deuxième temps, l'étude de la relation entre ces données et le rayonnement global pourra se faire de façon objective, au regard des résultats obtenus. Cette deuxième étape n'entre pas dans le cadre de ce travail.

### III-3. LE PROGRAMME CACTUS

Le programme CACTUS a été développé au Centre de Météorologie Spatiale de Lannion, par C. PASTRE (1981). Ce programme, fondé sur un seuillage de l'image visible, fournit, à partir des données numériques de Météosat, une carte semi-journalière d'estimation de la durée d'insolation.

#### III-3-a. Description technique du programme CACTUS

Pour chaque demi-journée, le programme fournit en pratique deux types de grille :

- une grille de pourcentage d'ennuage ou nébulosité
- une grille de brillance minimale, appelée albédo par abus de langage.

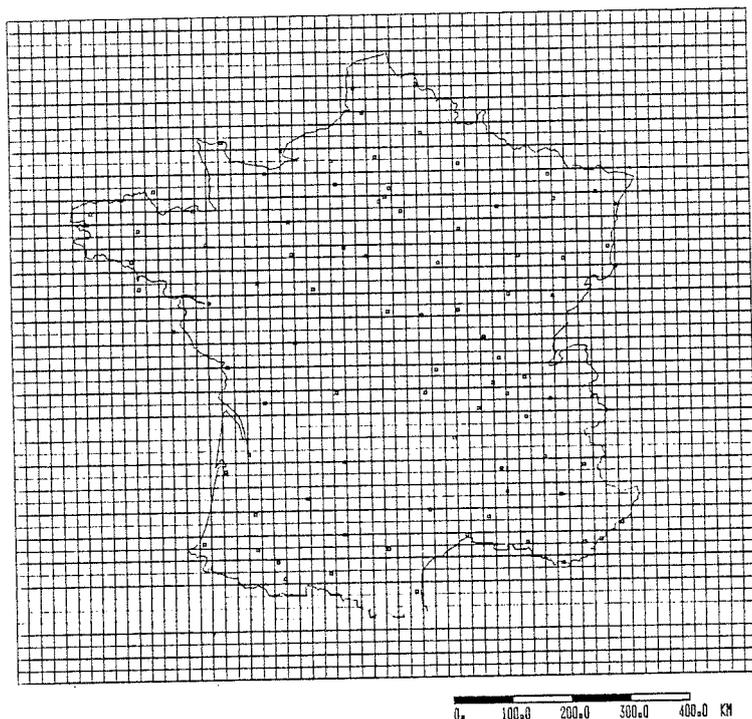
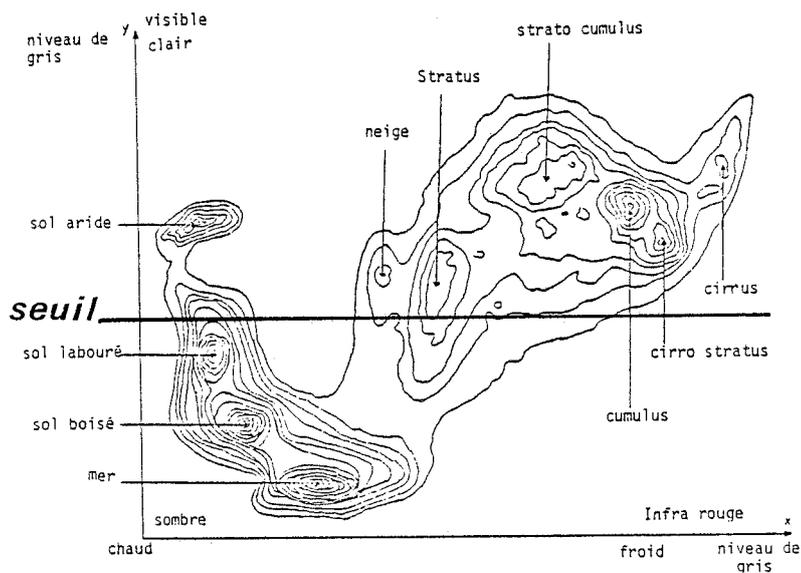


Fig I-7 : Grille 55x60 mailles couvrant la France, fournie par le programme Cactus. Le réseau de 88 Héliographes de la Météorologie Nationale a été superposé.

Fig I-8 : (d'après Rezki;1982)

Un seuil dans le seul canal Visible ne permet pas de distinguer la neige (ou un sol aride) des montagnes.



Histogramme bidimensionnel des radiations visibles (axe y) et infra-rouge (axe x). (Niveau de gris sur photos visible et infra-rouge des différents objets au sol). Cette figure illustre quelques-uns des problèmes posés par l'utilisation d'un canal unique. C'est ainsi par exemple qu'un cirrus au dessus d'un sol aride aura même brillance dans la bande visible mais pas le même niveau de gris que le sol dans la bande infrarouge.

Le programme fonctionne pour la France métropolitaine qui est couverte par une grille 60x55 (3300 mailles) (voir Fig. I-7). La maille élémentaire est de 20x20 km<sup>2</sup>. Chaque maille contient un ensemble de pixel (environ 4x4 pixels) de l'image numérique initiale de Météosat.

#### La carte de nébulosité

La nébulosité d'une maille est définie pour chaque image disponible comme le rapport du nombre de pixels couverts au nombre total de pixels de la maille. Chaque pixel est classé comme clair ou couvert par comparaison de sa réflectivité avec un seuil. Le choix de ce seuil est donc la clé du traitement. Il doit être assez haut pour qu'il n'y ait pas de confusion avec le sol, assez bas pour détecter les nuages de réflectivité faible. Afin de pouvoir raisonner en terme de luminances, la mesure satellitaire est étalonnée : le zéro de l'échelle des comptes numériques correspond au noir de l'espace, le sommet est pris par correspondance avec un réflecteur lambertien d'albédo un. Cet étalonnage grossier (10 % de la pleine échelle) est considéré comme suffisant par C. PASTRE. Le seuil de luminance est une fonction déterministe de l'angle zénithal solaire. Le traitement de l'image n'est pas effectué si le soleil se trouve à moins de 6° de l'horizon, pour des raisons de numérisation. La méthode empirique de réglage des seuils utilisés pour cette étude est décrite dans l'annexe A1.

Le classement entre pixels clair et couvert se fait avec le seul canal visible. Ce choix est justifié par le parti pris de simplicité du programme qui évite ainsi les problèmes de superposition d'images provenant de deux canaux différents.

On peut visualiser les erreurs liées à ce choix en construisant un histogramme bidimensionnel de Météosat (cf fig I-8). La construction d'un tel histogramme est faite à partir d'un grand nombre d'images (VIS et IR), pour lesquelles on a porté tous les couples (VIS,IR) observés. Les isolignes représentent le nombre de couples. Un pic correspond à un grand nombre de couples recensés. Cet histogramme ne donne aucune indication sur la répartition spatiale des luminances mais sur leur spectre.

Dans un deuxième temps, on peut chercher une correspondance entre les pics observés sur l'histogramme et les objets observés par les pixels correspondants : c'est ainsi qu'une valeur forte de réflectivité en IR (corps froid), correspondant avec une valeur forte de réflectivité en VIS (surface réfléchissante) est attribuée aux cirrus : ces nuages très élevés sont très froids et réfléchissent bien le rayonnement solaire. Certaines zones de l'histogramme ne correspondent à aucune réalité physique (luminance faible en IR et forte en Visible par exemple) : on n'y a pas observé de couple.

Si on place un seuil sur le seul canal VIS, comme indiqué sur la figure, on voit qu'on pourra discriminer le sol (faible réflectivité) et les nuages (forte réflectivité) mais il subsiste une ambiguïté pour deux types de sols :

les sols arides et la neige. Le problème des sols arides ne se pose pas en France mais sera important en Afrique. Le cas de la neige, par contre, devra être pris en compte. La solution envisagée est décrite par la suite.

#### Construction de l'image journalière d'ensoleillement

On dispose, pour chaque demi-heure (résolution temporelle de Météosat), d'une image de nébulosité. La moyenne des différentes images disponibles sur la demi-journée puis sur la journée fournit une nébulosité moyenne journalière. Cette nébulosité est ensuite assimilée à une durée d'ennuage, rapportée à la durée potentielle du jour ; cette mesure est le complément à un de la fraction journalière d'insolation.

Il faut faire attention à la définition de cette "nébulosité" qui ne correspond pas à une nébulosité classique, telle qu'elle est observée sur terre par les météorologues pour un instant et un lieu donnés, mais à une moyenne dans le temps d'un pourcentage de couverture nuageuse sur une surface donnée.

Si on admet que la nébulosité change surtout par advection et que la vitesse de déplacement moyenne des systèmes nuageux en France est de 10 m/s (soit environ 40 km/h), la taille de la maille (de 20 km de côté) est alors cohérente avec le régime météorologique français pour des prises de vue espacées d'une demi-heure (20 km par demi-heure). Il existe, en effet, une

relation entre la couverture nuageuse autour d'un point, à un instant donné, et les nuages qui vont passer entre ce point et le soleil pendant l'intervalle de temps séparant deux images.

#### La carte d'albédo de CACTUS

La carte d'"albédo" n'est pas utilisée dans le traitement des données, mais elle sert à la vérification de la position géographique et à la détection de la neige. Cette carte est obtenue en retenant pour chaque maille la plus faible valeur de réflectivité ponctuelle, pour une série mensuelle d'images. La carte d'albédo est donc une image composite, dont la valeur en chaque point peut correspondre à un instant différent de prise de vue.

Disposant de cette carte d'albédo, on peut l'utiliser pour tracer la frontière mer/terre, celle-ci est en effet bien marquée durant toute l'année : l'influence saisonnière de la végétation est négligeable.

#### Le positionnement de la grille

Les images, initialement en projection satellitaire, sont transformées par le programme CACTUS en projection "Lambert 30-60" : la terre est projetée sur un cône sécant aux parallèles 30° et 60°. Cette projection est conforme : l'échelle est la même quelle que soit la direction au voisinage d'un point, mais elle varie dans le plan en fonction de la latitude. Une description plus précise de cette projection est donnée en annexe A2.

La superposition de la carte d'albédo et de la carte de France en projection 30-60 a fait apparaître un problème de navigation, comprenant à la fois une différence d'échelle (homothétie) et une translation (voir annexe A3). Toutes les grilles ont donc subi une correction. Ce décalage résulte en partie de la précision des logiciels de projection mais les problèmes de navigation du satellite sont eux-mêmes complexes. Ils peuvent être résolus en se repérant à des amers de façon automatique (M. FORTIN, 1983) ou manuelle (dans notre cas). Ce décalage évolue assez lentement pour qu'une correction mensuelle soit suffisante.

### La détection de la neige

C'est la carte d'albédo qui va servir à détecter les mailles enneigées. La neige correspond à une zone d'albédo fort et persistant sur plusieurs jours. Si il y a eu une chute de neige, on réinitialise la carte d'albédo et on regarde son évolution. Une fois la zone enneigée déterminée, le programme n'estimera pas le pourcentage de nébulosité sur les mailles considérées.

La détermination exacte de la zone enneigée pose un problème car il est rare d'avoir un temps dégagé après une chute de neige. De plus, la valeur d'albédo de la maille est la valeur de réflectivité du pixel le plus faible. Il suffit donc d'un pixel non enneigé (une vallée par exemple) pour la maille entière soit considérée comme dégagée. La couverture nuageuse est alors surestimée.

### Conclusion sur les données du programme CACTUS

Ce programme simple détermine la nébulosité sur des mailles en fonction d'un seuil. Le choix de ce seuil est délicat. Mais une fois celui-ci déterminé, le traitement ne nécessite aucune donnée extérieure. Pour des raisons d'ordre pratique (temps de calcul), on n'a utilisé qu'une image par heure sur les deux disponibles. Cette contrainte pénalisera davantage l'hiver durant lequel on trouve des situations d'advection à évolution rapide, que l'été où prédominent les situations convectives.

## IV - DONNEES DISPONIBLES POUR CETTE ETUDE

Dans un premier temps, on a disposé de dix jours du programme CACTUS : du 31 mars au 9 avril 1983. C'est sur cette période que les seuils ont été déterminés.

Dans un deuxième temps, grâce aux bonnes performances des premiers résultats, le programme CACTUS a fonctionné de manière opérationnelle pendant l'année 1984. Après avoir éliminé les jours pour lesquels on disposait de moins de six images ou pour lesquels la répartition des images

Fig I-9 : Réseaux d'Héliographes



(a) de la Météorologie Nationale,



(b) autres postes

était trop irrégulière (plus de trois heures sans images), il nous reste 200 journées (voir la liste en Annexe A4). La saisie des données héliographiques des mois correspondants a été faite pour les 88 stations du réseau synoptique de la Météorologie Nationale dont les données sont publiées dans le Résumé Mensuel du Temps (R.M.T.) (Fig.I-9a). Ces données sont disponibles quelques mois après leur relevé dans les stations. On note ici l'avantage de la centralisation des données satellitaires pour la rapidité d'accès à l'information. De plus, le volume de stockage est très réduit. Les données disponibles constituent un ensemble d'environ 11x30x3 grilles de 3300 points, soit environ 3 millions de nombres. L'ensemble des résultats de 1984 tient sur une bande de 1200 pieds (densité 1600 BPI). L'archivage, nécessaire à la constitution de séries climatologiques est donc très facilement envisageable.

La remarque sur la disponibilité des mesures sol doit être modulée dans le cas de fichiers informatisés et accessibles par Météotel par exemple. Ce traitement est actuellement mis en place par la Météorologie Nationale.

On ajoutera enfin que les 88 héliographes de la Météorologie Nationale couvrent régulièrement le territoire français et peuvent éventuellement être complétés (dans le cas d'une étude courte) par les données d'héliographes dépendant d'autres organismes (tel l'INRA par exemple). Malheureusement, ceux-ci fonctionnent de façon irrégulière, suivant les besoins de leurs utilisateurs (nous en avons recensé 100 sur la France (Fig.I-9b)).

Nous avons, pour les besoins de notre étude, été amenés à partager notre réseau en sous réseaux de répartition régulière. La densité de ces différents réseaux est résumée dans le tableau I-3 ci-après :

nombre d'héliographes du réseau		distance interstations (km)		
		moyenne	minimum	maximum
réseau complet	188	35	2	90
Météorologie Nationale	88	60	12	110
sous réseaux	61	80		
	23	150		
	5	500		

Tableau I-3

Variable utilisée

Les variations journalières de la durée d'insolation possèdent une composante astronomique liée au lieu et à la saison, dont on veut s'affranchir. C'est la raison pour laquelle on utilise le rapport d'insolation  $z$

$$z = \frac{S}{S_0}$$

$S$  durée d'insolation mesurée par l'héliographe

$S_0$  durée maximale d'insolation = durée astronomique du jour corrigée de l'effet de masque.

Le calcul de la durée maximale d'insolation  $S_0$  et les masques utilisés pour les différentes stations sont décrits en annexe A5.

## CONCLUSION

La durée d'insolation n'est pas une donnée énergétique, mais sa mesure est très utile à l'étude spatiale et temporelle du gisement énergétique. Ceci pour deux raisons: elle est bien liée au rayonnement global au niveau journalier, et elle est mesurée de façon beaucoup plus dense (88 héliographes et seulement 31 pyranomètres), et depuis bien plus longtemps que celui-ci.

Son lien avec la nébulosité en fait d'autre part une mesure directement liée aux données satellitaires de réflectivité. Cette propriété est exploitée par les différentes méthodes d'estimation du rayonnement solaire au sol, à partir des mesures de satellites. La méthode de C. PASTRE, développée au C.M.S. de Lannion, se propose, dans un premier temps, d'estimer le rapport d'insolation et non le rayonnement solaire. On peut ainsi tester directement les performances de la méthode dans une étape fondamentale pour tous les modèles : les nuages sont le principal facteur d'atténuation du rayonnement solaire.

On souligne enfin que la différence de nature des deux types de mesure (au sol et par satellite) implique de porter une attention particulière sur l'effet de l'intégration spatiale des mesures satellitaires instantanées par rapport aux mesures sol ponctuelles et continues.

2ème PARTIE

**ANALYSE DE LA STRUCTURE SPATIALE**

## INTRODUCTION

Tout phénomène météorologique est variable dans le temps et dans l'espace. De précédentes études se sont attachées à examiner et à quantifier de manière statistique les fluctuations temporelles de l'insolation en un certain nombre de stations au sol (M.J. MEJON, J. GOUSSEBAILE, M. MARQUES). Nous nous intéressons ici à la description statistique des liaisons spatiales ou structure spatiale du rapport d'insolation.

Un phénomène météorologique peut être lisse ou chahuté, de variabilité plus ou moins importante. La structure spatiale de la pluie ou celle de l'insolation ne se ressemblent pas. Utiliser la fonction de structure empirique, c'est prendre en compte une expérience physique dans tout traitement statistique ultérieur.

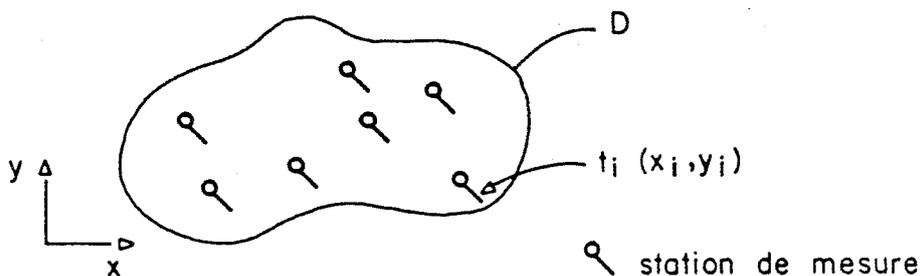
L'étude de la structure peut s'appliquer, non seulement à un phénomène unique ; mais aussi à plusieurs phénomènes conjointement. On parle alors de structure croisée. Le but n'est plus l'étude de la cohérence interne mais d'apprécier la co-fluctuation de leurs distributions spatiales. Cette information est d'autant plus intéressante que les phénomènes sont difficilement comparables de façon directe, en termes physiques. C'est le cas, par exemple, de mesures de nature différentes comme celles des héliographes et de la télédétection.

L'analyse de la structure spatiale est à la base de nombreuses utilisations allant de l'optimisation de réseau à la cartographie optimale et à l'étalonnage d'image, ... C'est pourquoi nous y avons porté un soin particulier, soulignant les hypothèses faites à chaque étape de calculs car elles conditionnent les applications ultérieures.

Deux approches différentes ont été adoptées pour étudier la structure spatiale du rapport d'insolation : la première d'entre elles concerne la modélisation développée dans le cadre de l'analyse objective. Elle fait l'objet des deux paragraphes suivants. La seconde est sa décomposition par des techniques d'analyse factorielle et fait l'objet du troisième chapitre. Le dernier chapitre, enfin, concerne une application de l'analyse de la structure à l'optimisation de réseau.

#### NOTATIONS UTILISEES

Dans l'espace à 2 dimensions, on s'intéresse à un domaine borné  $D$ , constitué de points  $t$ , de coordonnées géographiques  $(x,y)$ . Dans ce domaine sont réparties des stations de mesure  $t_i$ , de coordonnées  $(x_i,y_i)$  ( $i$  indice la station).



Sur ce domaine, se développe un phénomène  $z$  (par exemple la durée d'insolation) qui est fonction du lieu  $t$  et du temps  $k$ , on notera  $z_k(t)$  la valeur du phénomène  $z$ , au lieu  $t$ , pour l'évènement  $k$ . Les mesures faites par le réseau sol seront notées  $z_{sk}(t)$ , celles faites à partir de la télédétection  $z_{Tk}(t)$ .

On dispose de  $p$  stations de mesure sol, et de  $G$  points de grille (mesures de télédétection), ceci pour  $n$  évènements.

## I - MODELISATION STATISTIQUE DANS LE CADRE DE L'ANALYSE OBJECTIVE

### I-1. Notion de variable régionalisée

Les phénomènes naturels s'organisent dans le temps et dans l'espace en y manifestant une certaine structure. En se plaçant en un point  $t$  du domaine considéré  $D$ , on peut considérer les valeurs  $z_k(t)$  comme les réalisations d'une variable aléatoire (VA). Le développement spatial de ce phénomène apparaît alors comme la juxtaposition de VA qui, en général, ne sont pas indépendantes. Une variable caractérisant un tel phénomène est appelée variable régionalisée (VR) (G. MATHERON, 1965) ; cette variable régionalisée est interprétée comme une réalisation de fonction aléatoire (notée en majuscule  $Z$ ). Elle est traitée avec la théorie probabiliste.

### I-2. Problème de l'inférence statistique

Afin de pouvoir utiliser la théorie des FA, il faut pouvoir reconstituer sa loi : c'est l'inférence statistique. Dans notre cas, l'inférence des 2 premiers moments est suffisante. Si on dispose d'un grand nombre de réalisations  $n$  (contexte dit "climatologique"), il sera possible d'estimer les moments de la FA. Il faut pour cela s'assurer que ces  $n$  réalisations peuvent être considérées comme le résultat de  $n$  tirages de la même FA : unicité du processus générateur. Si le nombre de réalisations disponibles n'est pas suffisant ou si une réalisation unique est disponible (contexte dit "spatial"), l'inférence statistique requiert alors l'introduction d'hypothèses supplémentaires.

L'analyse de structure, au sens où l'on étudie la cohérence spatiale de la VR, nécessite l'inférence des moments d'ordre 2. Les principales fonctions de structure utilisées de façon classique sont rappelées ci-dessous.

### I-3. La formulation mathématique des fonctions de structure classiques

Pour quantifier la dépendance statistique dans l'espace, le corrélogramme, la fonction de covariance ou le variogramme sont les fonctions de structure généralement utilisées selon que l'on se situe dans un contexte "climatologique" ou "spatial".

Les différents moments de la FA Z sont notés :

- la moyenne temporelle  $m_{zi}$  du phénomène z au point  $t_i$ , calculée sur n évènements :

$$m_{zi} = E(z_k(t_i))$$

la moyenne est estimée par  $m_{zi}^* = \frac{1}{n} \sum_{k=1}^n z_k(t_i)$

- La covariance  $Cov_{zij}$  entre deux points  $t_i$  et  $t_j$ , pour le phénomène z

$$Cov_{zij} = E((z_k(t_i) - m_{zi})(z_k(t_j) - m_{zj}))$$

- La variance  $Var_{zi}$  au point  $t_i$  est un cas particulier de la covariance

$$Var_{zi} = Cov_{zii}$$

- Le coefficient de corrélation  $r_{zij}$  entre deux points  $t_i$  et  $t_j$ , pour le phénomène z

$$r_{zij} = \frac{Cov_{zij}}{\sqrt{Var_{zi} Var_{zj}}}$$

- Le variogramme  $\gamma_{z_{ij}}$  est défini par

$$\begin{aligned} 2 \gamma_{z_{ij}} &= \text{Var} (z(t_i) - z(t_j)) \\ &= \text{Var } z_i + \text{Var } z_j - 2 \text{Cov}_{z_{ij}} \end{aligned}$$

#### I-4. Inférence statistique dans le cadre de la stationnarité

Le phénomène qui nous intéresse n'est connu qu'en un nombre limité de points  $p$ , répartis sur un domaine  $D$ . La connaissance de la structure ou liaison existant entre 2 points quelconques du domaine n'est alors possible que sous certaines hypothèses. L'hypothèse couramment formulée est celle de la stationnarité de la FA.

##### I-4-a. Hypothèse de stationnarité

Par définition, les moments à tous les ordres d'une FA stationnaire sont invariants par translation. Dans notre cas la stationnarité d'ordre 2 (celle des deux premiers moments) est suffisante. Elle exprime que

- la moyenne est indépendante du lieu  $t$

$$m_z(t) = m_z$$

- la covariance est uniquement fonction de l'interdistance des points  $t$ ,  
et  $t'$

$$\text{Cov}_z(t, t') = \text{Cov}_z(h_{tt'})$$

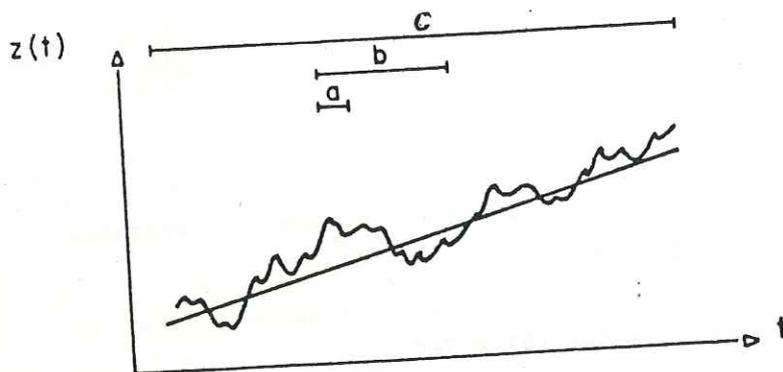
Cette condition implique une variance finie et  
indépendante du lieu :

$$\text{Var}_z(t) = \text{Cov}_z(0) = \text{Var}_z$$

Par abus de langage, on parlera dans la suite du texte de "stationnarité"  
sans préciser l'ordre.

La deuxième condition est l'expression de l'homogénéité du phénomène. Elle  
permet la connaissance de la dépendance statistique entre deux points  
quelconques à partir d'une connaissance discrète. Dans la pratique, cela  
consiste à travailler sur un phénomène auquel on a retiré toute composante  
déterministe. Le mot "déterministe" signifie : lié à la position  
d'observation dans le temps ou dans l'espace. En d'autres termes, si la  
connaissance de la position ( $k$  dans le temps ou  $t$  dans l'espace), d'un  
point de mesure  $z_k(t)$  de la F.A.Z, nous renseigne sur sa valeur, la FA  
n'est pas stationnaire.

Une telle contrainte est rarement compatible avec la réalité du phénomène.  
La stationnarité locale ou quasistationnarité est plus réaliste. C'est  
pourquoi on sera souvent amené à travailler en voisinage restreint. La  
notion de stationnarité est liée à la taille du domaine  $D$  à travers lequel  
on "regarde" le phénomène. Considérons l'exemple suivant à une dimension :



Fig(II)-1

On s'intéresse, dans cet exemple, à la stationnarité de la seule moyenne. Si le phénomène  $z$  est observé à travers la fenêtre (c), il apparaît une nette tendance : le phénomène croît avec le lieu  $t$ . Cette tendance de la moyenne est appelée dérive. Si on utilise une fenêtre de taille (b), par contre, le phénomène peut être considéré comme stationnaire d'ordre 1. A l'échelle (a) enfin, l'hypothèse de stationnarité n'est plus compatible avec les observations. Cet exemple illustre l'importance de l'échelle sur la notion de stationnarité.

#### I-4-b. Inférence du modèle stationnaire

##### - Contexte climatologique

Il est intéressant de travailler sur les valeurs centrées par la moyenne climatologique, car ce changement de variable garantit la stationnarité d'ordre 1 : la moyenne de la variable centrée est uniformément nulle. Par contre, on ne peut garantir, par un simple changement de variable, la stationnarité d'ordre 2. En effet, la réduction par l'écart type implique une variance constante, mais ne contraint pas la covariance à n'être fonction que de l'interdistance. Cette hypothèse doit donc être vérifiée

expérimentalement. On rappelle que le calcul des moments nécessite l'unicité du processus générateur de la série de réalisations disponible : unicité de la FA.

Sous hypothèse de stationnarité d'ordre 2, les fonctions de structure s'expriment simplement :

- covariance :  $\text{Cov}_{z_{ij}} = \text{Cov}_z(h_{ij})$

- corrélation :  $r_{z_{ij}} = \frac{\text{Cov}_z(h_{ij})}{\text{Cov}_z(0)} = r_z(h_{ij})$

la corrélation est une covariance normée par la variance de la variable.

- variogramme :  $\gamma_{z_{ij}} = \text{Cov}_z(0) - \text{Cov}_z(h) = \gamma_z(h)$

le variogramme est le complémentaire, à la variance de la variable, de la covariance.

Il est indifférent de choisir l'une ou l'autre des fonctions de structure puisqu'elles sont toutes liées analytiquement. La fonction de covariance paraît préférable pour apprécier la variabilité spatiale relative de plusieurs phénomènes. Le corrélogramme, au contraire, est normé et standardise les échelles. Le choix de l'une d'entre elles dépend du but poursuivi.

#### - contexte spatial

Ce contexte défavorable n'est rencontré dans cette étude que dans le cadre de l'estimation du champ des moyennes ou des écarts types. Il est cité ici pour mémoire, un développement clair et didactique est proposé par J.P. DELHOMME, 1976 et JOURNEL, 1978. On utilise l'hypothèse de stationnarité d'ordre 2 des accroissements de la variable ou hypothèse intrinsèque :

$$\begin{aligned} & \cdot E(z(t+h)-z(t)) = 0 \\ & \cdot \text{Var}(z(t+h)-z(t)) = 2 \gamma_z(h) \end{aligned}$$

On évite ainsi le problème de l'estimation de la moyenne nécessaire au calcul de la covariance. Cette hypothèse est plus faible que la stationnarité d'ordre 2 : la stationnarité d'ordre 2 implique la stationnarité intrinsèque (le variogramme s'exprime alors simplement en fonction de la covariance. Cf. contexte climatologique) mais la réciproque n'est pas vraie (on n'a pas a priori sur l'expression de la variance dans le cas intrinsèque).

Dans le contexte spatial, afin d'éviter l'estimation des premiers moments (moyenne, variance) le variogramme est la fonction de structure utilisée.

#### I-5. Inférence statistique dans le cas non stationnaire

Si la stationnarité d'ordre 2 n'est pas compatible avec la réalité du phénomène, on peut chercher une combinaison linéaire d'accroissements d'ordre  $k$  qui soit, elle, stationnaire : c'est l'hypothèse intrinsèque d'ordre  $k$  (cf. O. DUBULE, 81). Le fait de travailler sur ces accroissements permet de filtrer la composante déterministe de la moyenne (ou dérive) du phénomène. Il se pose alors le délicat problème de l'identification du degré de la dérive et de l'inférence de la fonction de structure appelée covariance généralisée.

De précédentes études, J.D. CREUTIN 79, G. DELRIEU 86, ont montré la lourdeur d'une telle démarche. C'est pourquoi, dans notre application, si l'hypothèse de stationnarité n'est pas vérifiée, on cherchera à s'y ramener. Par un simple changement de variable ou en prenant en compte toute réalité physique susceptible d'expliquer l'hétérogénéité (altitude, azimut, ...), soit encore en travaillant de façon locale.

### I-6. Calcul et modélisation des fonctions de structure

Nous nous situons dans la suite de cette étude dans l'hypothèse de stationnarité d'ordre 2 de la FA. Le phénomène est homogène, la fonction de structure ne dépend que de l'interdistance entre deux points.

A partir des données expérimentales, on reporte la valeur estimée de la corrélation, la covariance ou du variogramme en fonction de l'interdistance  $h_{ij}$ . Une certaine dispersion des points expérimentaux est alors observée (cf. Figure (II)-2). Celle-ci est due à l'échantillonnage : au nombre limité d'observations disponibles pour l'estimation de la structure.

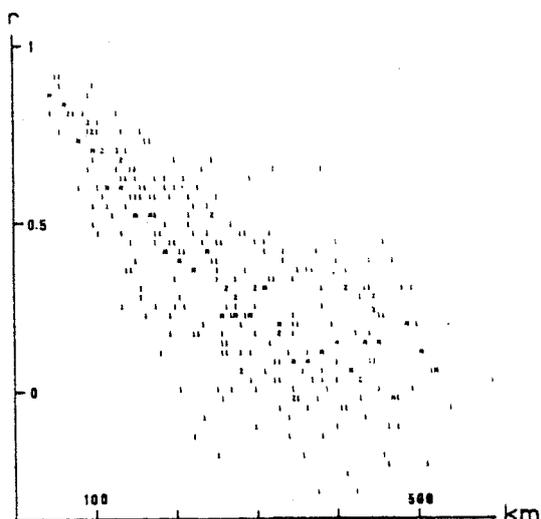


Fig. II-2 :  
corrélogramme empirique du  
rapport d'insolation mesuré  
par le sol.

L'écart type d'échantillonnage est connu de façon théorique. Si la dispersion du nuage est supérieure à sa distribution théorique, il y a hétérogénéité du phénomène. Un moyen de contrôle de l'hypothèse d'homogénéité est ainsi disponible. L'ajustement d'un modèle au nuage de points est décrite ci-dessous pour la corrélation. La démarche est

identique pour les autres fonctions de structure qui lui sont liées de façon analytique.

#### I-6-a. Modèle autorisé

Toute fonction n'est pas autorisée pour modéliser la fonction  $r_z(h)$ . Les modèles décrivant la structure doivent garantir que la variance de toute combinaison linéaire de la variable reste positive : une telle fonction est dite définie positive. Un modèle couramment utilisé en analyse objective est le modèle à 2 paramètres :

$$r_z(h) = e^{-ax} \cos bx$$

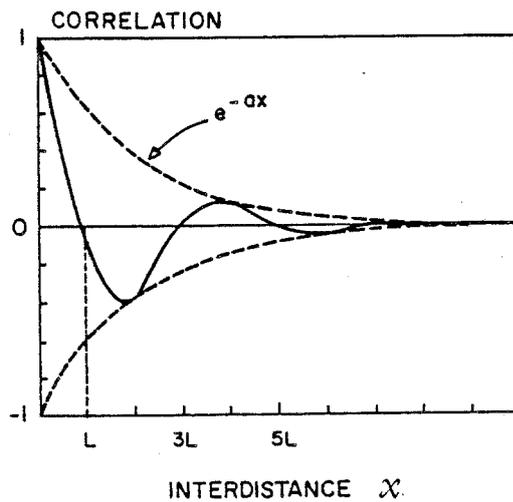


Fig. (II)-3

Sa forme s'adapte aux différents cas rencontrés. Son ajustement est décrit ci-dessous.

## I-6-b. Ajustement du modèle

L'ajustement se fait sur la valeur moyenne de la corrélation calculée pour des classes de distance de largeur  $\Delta L$ . On obtient ainsi une courbe empirique qui met en évidence le comportement moyen du nuage. Par la suite, dans l'utilisation et l'interprétation de ces courbes empiriques, la dispersion initiale du nuage (cf. Figure (II)-2) ne devra pas être perdue de vue. L'estimation de la valeur moyenne de la corrélation sur un intervalle de largeur  $\Delta L$ , centré en  $L$  s'écrit :

$$\bar{r}_z(L) = \frac{1}{N(L)} \sum_{l=1}^{N(L)} r_z(h_l) \quad \forall h_l \in \left[ L - \frac{\Delta L}{2} ; L + \frac{\Delta L}{2} \right]$$

où  $N(L)$  est le nombre de couples de stations dont l'interdistance  $h_1$  est comprise dans l'intervalle  $\Delta L$ .

Dans la pratique, il faut se limiter à un nombre  $N_c$  de classes suffisamment faible pour que  $N(L) > 30$ , sinon la valeur de  $r_z(L)$  est peu significative. Ce problème se pose plus spécialement pour les faibles et fortes distances. De plus, pour les grandes distances, on remarque que les quelques couples de points qui y participent présentent souvent des directions privilégiées dues à la forme particulière du domaine  $D$ .

A partir de ces  $N_c$  points  $r_{zij} = r_{zL}(h)$ , on peut ajuster une fonction mathématique de façon manuelle ou automatique. Pour notre part, nous avons préféré la solution manuelle qui offre l'avantage, outre une programmation simplifiée, de permettre l'introduction de l'interprétation de l'utilisateur.

## Principe du calage manuel

Ce calage s'effectue à l'aide d'un logiciel interactif. On propose des valeurs pour les paramètres  $a$  et  $b$ , et on superpose la courbe théorique

ainsi obtenue aux valeurs expérimentales. L'ajustement direct de a et b étant délicat, on indique en pratique les points caractéristiques de la courbe, tels que sa valeur minimum, et la distance (2l) de ce point, etc. suivant les différents cas de figure que l'on peut rencontrer.

#### I-6-c. Contrôle des hypothèses de stationnarité

Le corrélogramme manifestant une certaine dispersion, il importe de vérifier si celle-ci peut être expliquée par les seules fluctuations d'échantillonnage ou si elle contient une part due à l'hétérogénéité du phénomène.

Si la population est gaussienne, l'écart type de distribution de la corrélation calculée sur un échantillon de n observations s'écrit pour n grand ( $n > 30$ ) (réf. J. VIALAR, 1956)

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}$$

On peut alors superposer au nuage de corrélation, le modèle autorisé ajusté, et ses intervalles de confiance. L'intervalle  $r \pm 1,28 \sigma_r$  qui contient théoriquement 80 % du nuage a été choisi. Si le nuage est compris dans ces bornes, la fluctuation observée est expliquée par les seules fluctuations d'échantillonnage, l'hypothèse de stationnarité est adaptée à la réalité du phénomène. Si ce n'est pas le cas, on essaiera de chercher les causes de l'hétérogénéité observée afin de se situer dans un contexte plus favorable.

La non homogénéité du champ peut être due à l'anisotropie de la variable : la structure est fonction de l'interdistance mais aussi de la direction d'observation. On utilisera alors les corrélogrammes par azimut (avec une certaine ouverture d'angle).

Si l'azimut n'est pas suffisant pour expliquer l'hétérogénéité, il faut se tourner vers d'autres explications : l'altitude, etc. suivant notre connaissance physique du phénomène.

### I-7. Interprétation

Nous ne reviendrons pas sur l'interprétation de la fonction de structure qui est fort bien expliquée dans de nombreux travaux (GANDIN,65; DELHOMME,76). Nous retiendrons simplement que la structure caractérise la façon dont se dégrade l'information apportée par une station à mesure qu'on s'éloigne de celle-ci. On s'intéresse plus particulièrement au comportement de cette fonction à l'origine et à l'infini :

- A l'origine, pour une distance nulle, la corrélation est parfaite  $r_z(0) = 1$ . Dans la pratique, l'extrapolation à la distance nulle de la courbe expérimentale conduit souvent à une valeur inférieure à 1. Cette discontinuité de la fonction de structure souvent appelée pépite\* peut être attribuée à un bruit blanc (dû par exemple à des erreurs de mesures) ou à une microrégionalisation du phénomène. Celle-ci n'est pas saisie par le réseau de mesure, trop lâche, elle traduit cependant la réalité du phénomène et ne doit pas être considérée comme une erreur.

- Deux points très éloignés l'un de l'autre ne sont pas liés : à l'infini, la corrélation est nulle. La distance à laquelle le corrélogramme se stabilise autour de 0 est appelée portée. Elle correspond au rayon de la zone d'influence d'un point expérimental. Si le corrélogramme oscille autour de la valeur nulle, on appellera portée la distance à laquelle elle s'annule pour la première fois.

REMARQUE : Alors que la covariance et le corrélogramme sont nécessairement bornés, le variogramme ne l'est pas. On montre (DELHOMME,76) que la présence d'un palier implique la stationnarité d'ordre 2.

Ajoutons pour conclure que l'examen d'une fonction de structure est subtil et l'expérience de l'utilisateur apporte énormément tant dans l'ajustement

---

\* L'origine de ce terme provient de l'étude des gisements d'or où la teneur en or varie considérablement selon que le point se situe à l'intérieur ou à l'extérieur d'une pépite.

du modèle que dans les renseignements qu'on peut en déduire : le "coup d'oeil" est irremplaçable. Pour notre part, la constante confrontation de notre réalité expérimentale avec l'expérience plus large de l'équipe hydrologie a contribué à atteindre l'interprétation et la mise en forme exposée au chapitre suivant.

#### I-8. Influence d'un bruit blanc sur la structure de la mesure

Une erreur est indissociable de la mesure de toute grandeur physique. Cette erreur n'est pas corrélée dans l'espace puisqu'elle n'est liée qu'à l'instrument de mesure, on peut donc la considérer comme un bruit blanc. Le phénomène mesuré  $z$  se décompose alors en la somme du phénomène "vrai" sous-jacent  $y$  et d'une erreur de mesure aléatoire  $\varepsilon$ , de variance  $V_\varepsilon$

$$z_k(t) = y_k(t) + \varepsilon_k(t)$$

L'intérêt de ce cas particulier réside à la fois dans la simplicité de la décomposition, et dans sa modélisation réaliste des mesures physiques. Les deux phénomènes  $z$  et  $y$  sont très proches, cependant leurs fonctions de structure diffèrent. La connaissance de cet effet est un outil utile pour l'interprétation des fonctions de structure expérimentales.

On suppose le phénomène vrai  $y(t)$  stationnaire d'ordre 2,  $z(t)$  est alors également stationnaire.

#### - Expression de la covariance de $z$

$$\text{Cov}_z(h) = \text{Cov}_{y+\varepsilon}(h) = \text{Cov}_y(h) + \text{Cov}_\varepsilon(h) + \text{Cov}_{y\varepsilon}(h) + \text{Cov}_{\varepsilon y}(h)$$

or  $\text{Cov}_{y\varepsilon}(h) = 0$  car  $\varepsilon(t)$  est un bruit blanc.

$$\begin{aligned} \text{Cov}_\varepsilon(h) &= 0 & \text{si } h &= 0 \\ &= V_\varepsilon & \text{si } h &= 0 \end{aligned}$$

$$\text{d'où } \text{Cov}_z(h) = \text{Cov}_y(h) + \text{Cov}_\varepsilon(h)$$

La covariance de z est identique à celle de y sauf pour la distance nulle, où la variance de z est majorée de V .

- Expression de la corrélation

$$r_z(h) = \frac{\text{Cov}_z(h)}{\text{Var}_z} = \frac{\text{Cov}_y(h) + \text{Cov}_\varepsilon(h)}{\text{Var}_y + V_\varepsilon}$$

$$h = 0 \quad r_z(0) = 1$$

$$h = 0 \quad r_z(h) = \frac{\text{Cov}_y(h)}{\text{Var}_y + V_\varepsilon} = r_y(h) \frac{\text{Var}_y}{\text{Var}_y + V_\varepsilon}$$

L'autocorrélation de r est inférieure à celle de y dans le rapport

$$\frac{\text{Var}_y}{\text{Var}_y + V_\varepsilon}$$

Elles tendent vers la même limite quand h tend vers l'infini car  $r_y(h)$  tend alors vers zéro.

- Expression du variogramme

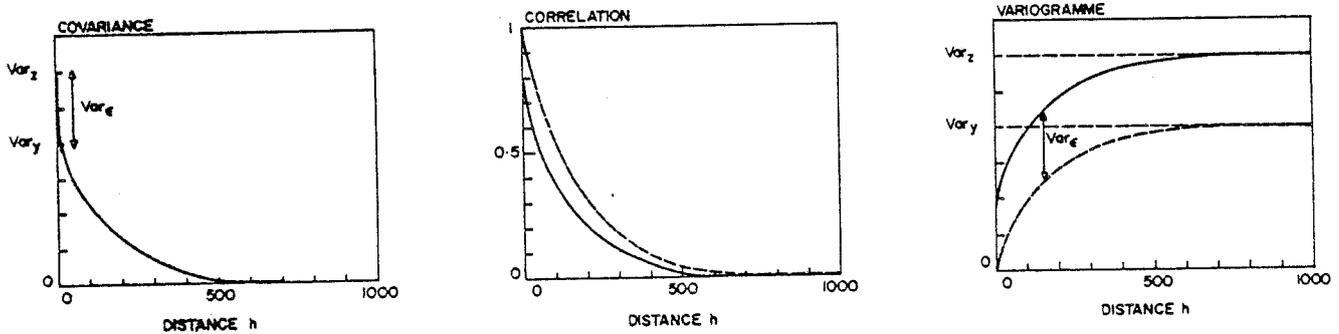
$$\begin{aligned} z(h) &= \text{Var}_z - \text{Cov}_z(h) \\ &= \text{Var}_y - \text{Cov}_y(h) + V_\varepsilon - \text{Cov}_\varepsilon(h) \\ &= \gamma_y(h) + \gamma_\varepsilon(h) \end{aligned}$$

$$\begin{aligned} h = 0 \quad \gamma_z(h) &= \gamma_y(h) \\ h = 0 \quad \gamma_z(h) &= \gamma_y(h) + V_\varepsilon \end{aligned}$$

Le variogramme de  $z$  est le même que celui de  $y$  à la constante  $V_\varepsilon$  près.

Les fonctions de structures des phénomènes  $z$  et  $y$  sont visualisées ci-dessous :

Fig. (II)-4 : effet d'un bruit blanc sur la structure d'un phénomène stationnaire.



Le bruit blanc est à l'origine de l'effet de pépite observé. Le variogramme est très sensible à un bruit blanc alors que seul le premier point de la fonction de covariance est modifié. La portée reste bien entendu inchangée.

## I-9. Structure croisée

### I-9-a. Nécessité de l'analyse croisée

La structure croisée permet d'estimer la cofluctuation spatiale de deux phénomènes. Celle-ci est obtenue de la même manière que la structure directe : on regarde la liaison entre la valeur du premier phénomène au point  $t_i$  et celle du deuxième phénomène au point  $j$ . Cette étape est nécessaire, car la seule connaissance de leur structure directe respective n'est pas suffisante pour estimer leur cofluctuation. Deux phénomènes peuvent avoir la même structure, mais ne pas cofluctuer. Dans ce cas, leur structure croisée sera nulle.

#### Remarques :

1. La valeur du corrélogramme croisé pour la distance nulle est égale au coefficient de corrélation entre les deux phénomènes.
2. Les deux phénomènes peuvent être décalés géographiquement (mauvais positionnement de l'image satellitaire par exemple). La fonction de covariance est alors sensible au sens dans lequel on considère le couple de variables, sinon elle est symétrique. Un décalage éventuel peut être détecté, en testant, pour plusieurs azimuts, la symétrie de la fonction de covariance. On obtient ainsi la direction, le sens et la valeur du décalage (cf. CREUTIN, 1987).
3. La différence de nature éventuelle des phénomènes étudiés n'est pas un obstacle au calcul de leur structure croisée qui sanctionne la cofluctuation.

### I-9-b. Modèles autorisés

Les conditions que doit remplir le modèle sont plus complexes que dans le cas de la structure directe. Pour un développement complet, on se reportera à WACKERNAGEL, 1985.

Rappelons simplement que si on dispose des réalisations de  $N$  fonctions aléatoires indicées en  $M$  :  $Z_M(t)$ , les modèles décrivant la structure doivent garantir que le calcul de la variance de toute combinaison linéaire des variables régionalisées  $Z_M(t)$  donne un résultat positif. Pour résoudre ce problème, on suppose l'existence d'une famille de fonctions aléatoires orthogonales, susceptibles d'engendrer linéairement les variables étudiées. On note  $U_l(t)$  ces fonctions de bases telles que  $C_{ll'}(h) = 0$  pour  $l \neq l'$ .

$$Z_M(t) = \sum_{l=1}^L a_{lM} U_l(t)$$

Dans ces conditions de décomposition de  $Z_M(t)$ , toutes les fonctions de covariance directes et croisées s'expriment comme une combinaison linéaire de  $L$  fonctions covariance de base :

$$C_{MM'}(h) = \sum_{l=1}^L b_{lMM'} C_l(h)$$

Ceci entraîne une condition sur les  $b_{lMM'}$  : pour  $l$  fixé, la matrice  $b_{kk'}$  doit être définie positive.

Ce modèle est dit modèle linéaire de corégionalisation. Il appelle 3 remarques :

1. Les fonctions de covariances autorisées sont symétriques : les variables ne doivent pas être décalées géométriquement.

2. Si on ne dispose que de deux FA, la condition sur les  $b_{lMM'}$ , devient

$$\forall i \quad b_{11} > 0 \quad |b_{12}| = |b_{21}| < \sqrt{b_{11} b_{22}}$$

3. Cas particulier : la corégionalisation intrinsèque :

Si toutes les variables de base  $U_1(t)$  ont la même covariance  $C_0(h)$ , le modèle de corégionalisation devient :

$$C_{MM'}(h) = b_{MM'} C_0(h)$$

Les variogrammes directs et croisés peuvent être déduits l'un de l'autre par simple affinité.

Dans le modèle linéaire de corégionalisation, toutes les FA se décomposent sur une base de  $L$  fonctions aléatoires orthonormées. Toutes les fonctions de covariances directes et croisées sont des combinaisons linéaires des fonctions de covariance de base.

#### I-10. Conclusion

L'utilisation de la théorie des fonctions aléatoires permet une approche statistique de la structure spatiale d'un phénomène observé sur un domaine  $D$  en un nombre limité de points.

L'hypothèse utilisée pour modéliser la structure est la stationnarité d'ordre 2, que l'on devra donc toujours chercher à satisfaire au mieux. Dans la pratique, cela consiste à travailler sur des variables centrées, en voisinage restreint, par azimuth, etc.

La modélisation de la fonction de structure est nécessaire à de nombreuses applications. Le traitement conjoint de plusieurs phénomènes (étalonnage d'image par exemple) nécessite de plus la connaissance de leur co-fluctuation spatiale ou structure croisée. Il faut donc porter à l'inférence de la structure (directe ou croisée) un soin particulier, et surtout, connaître ses faiblesses, dues principalement aux hypothèses formulées et à la densité du réseau de mesure utilisé.

Chaque classe de distance de la fonction de structure contient les particularités de l'échelle spatiale qu'elle représente. Toutes les échelles ne sont pas perceptibles au réseau de mesure : trop lâche à petite échelle et trop localisé à grande.

La fonction de structure, inférée sous hypothèse de stationnarité d'ordre 2 du phénomène, est un outil utile et commode. Son expression mathématique finale, a priori simple, ne peut être maniée de façon efficace par la suite sans la connaissance des limites exposées ci-dessus.

## II- APPLICATION A L'ANALYSE DU RAPPORT D'INSOLATION

### II-1. Mise en forme des données : codage utilisé

Le contexte dit "climatologique" est favorable à l'inférence de la structure spatiale : l'estimation des différents moments d'un phénomène est d'autant plus robuste que le nombre d'observations disponibles indépendantes est grand. Nous avons la chance, pour cette étude, de disposer d'un nombre important d'observations : 200 journées pour lesquelles le rapport d'insolation a été mesuré à la fois par le réseau sol et par le "satellite". Un tel jeu de données de télédétection est rarement disponible en raison du volume de données à manipuler, de leur saisie et de leur coût.

Ces 200 journées couvrent une bonne partie de l'année 1984 (voir Annexe 4) et recouvrent ainsi des situations météorologiques différentes. On se posera donc la question de l'unicité du processus générateur.

L'inférence statistique de la structure d'une fonction aléatoire connue à travers un nombre fini de points de mesure est simple sous hypothèse de

stationnarité d'ordre 2. La réalité physique du phénomène mesuré  $z_k(t)$  n'est pas toujours compatible avec la stationnarité, même locale. Le codage du phénomène mesuré (dit aussi "brut")  $z_k(t)$  a pour but de trouver une grandeur  $x'_k(t)$ , liée à  $z_k(t)$ , compatible avec cette hypothèse. Par extension, nous appellerons également cette grandeur  $x'_k(t)$  "phénomène".

#### II-1-a. Unicité du processus générateur

On veut se placer dans des conditions telles que tous les évènements soient considérés comme des réalisations d'une même FA. Ce n'est pas le cas pour le rapport d'insolation dont certaines caractéristiques varient de façon saisonnière. Cette tendance est observée alors que l'influence purement astronomique a été éliminée dans le calcul du rapport d'insolation (rapport de la durée mesurée à la durée potentielle). Cette variation s'explique par la différence de nature des conditions météorologiques au cours de l'année : les situations météorologiques de type frontal sont plus fréquentes en hiver, alors que l'été regroupe plus de situations convectives. Il s'agit bien de processus générateurs différents. Nous avons donc partagé l'année en périodes, à l'intérieur desquelles les caractéristiques statistiques des évènements sont suffisamment homogènes pour que les évènements la composant puissent être considérés comme les réalisations d'une même fonction aléatoire.

Le choix du découpage de l'année 1984 en périodes a été fait en se basant sur les deux premiers moments des mesures sol et satellite. On a cherché à obtenir des périodes homogènes du point de vue statistique, sous réserve de conserver des effectifs d'évènements suffisants (cf. Figure (II)-5 abc). Nous avons ainsi choisi de diviser l'année en 3 périodes constituées de 70 évènements environ chacune :

- La première période regroupe les mois d'hiver et s'étend de septembre à février ; elle est caractérisée par une moyenne du rapport d'insolation (insolation réelle / insolation potentielle) (33 %) assez faible, et un écart type (32 %) du même ordre de grandeur. Seuls 4 jours sur les 67 dépassent 70 % de rapport d'insolation.

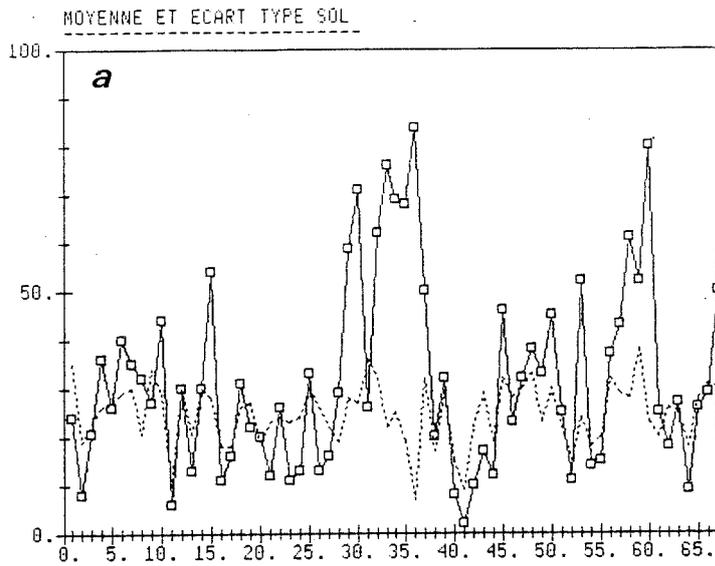
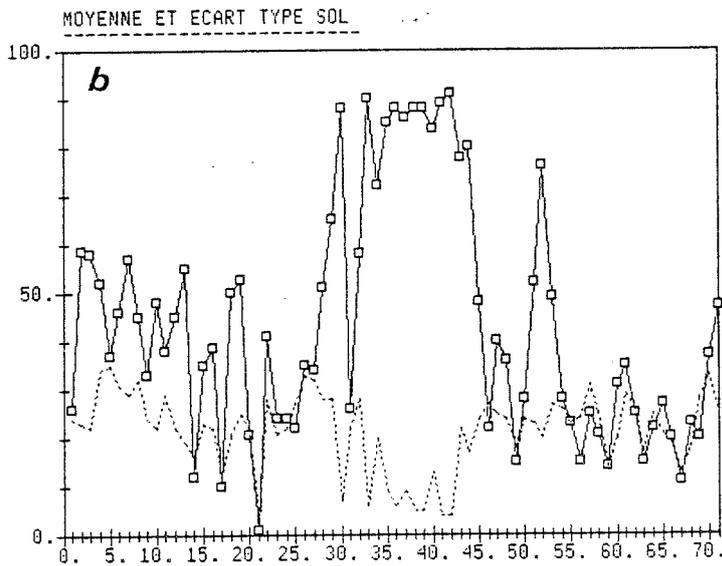


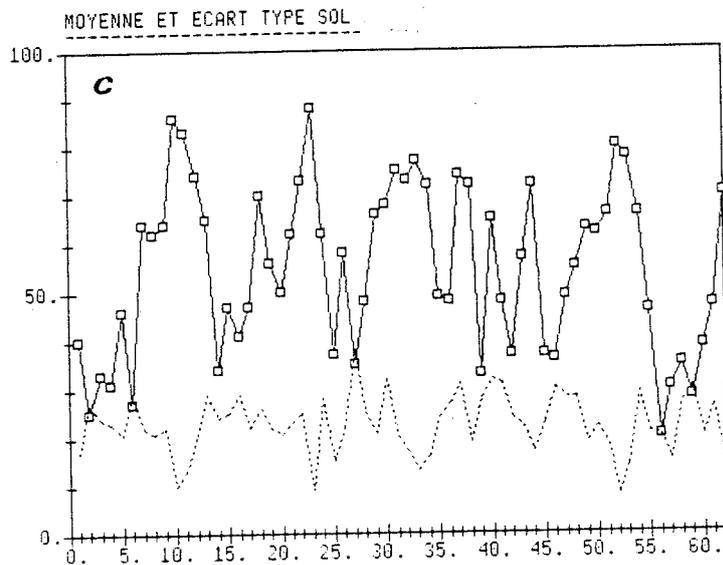
Fig II-5 :

Moyenne ( $\square$ ) et écart type (---) des mesures sol pour les 200 journées disponibles en 1984.

HIVER



PRINTEMPS



ETE

- La deuxième période comprend les mois de mars, avril et mai : nous l'appellerons le "printemps". Sa moyenne est de 44 %, et l'écart type calculé sur cette période de 34 %. On note une série de 12 jours de très beau temps (le rapport d'insolation est supérieur à 85 %) en avril (Figure 1-b). Nous nous intéresserons par la suite plus particulièrement à cette période pour deux raisons : 1) les seuils du programme CACTUS ont été réglés sur le mois d'avril (voir Annexe 1) ; 2) on dispose de données satellitaires sur les années précédentes et suivantes pour cette période uniquement, des calculs ultérieurs seront alors possibles.

- La troisième période, enfin, qui s'étend de juin à août, correspond à l'été. Sa moyenne (55 %) est élevée et son écart type (29 %) est plus faible. Elle est caractérisée par un nombre important de jours de beau temps de très faible variabilité spatiale.

#### II-1-b. Indépendance temporelle des évènements

La robustesse de l'inférence des moments statistiques est liée au nombre d'évènements indépendants considérés. Or, le rapport d'insolation, comme la plupart des phénomènes météorologiques, est autocorrélé dans le temps. La période d'autocorrélation typique pour la durée d'insolation journalière est d'environ 3 jours (cf. GOUSSEBAILE, 74). Pour s'assurer de l'indépendance des évènements utilisés, on pourrait envisager de ne garder les résultats du rapport d'insolation qu'un jour sur trois. Mais cet échantillonnage n'est possible qu'avec un jeu important de données. Pour notre part, nous avons décidé d'utiliser tous les jours disponibles, l'échantillonnage étant décidé par les jours de fonctionnement du programme CACTUS qui ne sont pas toujours consécutifs. Le résultat de ce tirage "aléatoire" est donné dans la liste des jours disponibles (Annexe 4). Mis à part le mois de mai où on trouve une longue série de jours disponibles consécutifs, ce mode d'échantillonnage est assez régulier.

#### II-1-c. Stationnarité de la FA

Le rapport d'insolation n'est pas une variable stationnaire : sa moyenne au moins présente des tendances géographiques systématiques (opposition

nord/sud, vallée, montagne). Notre but est alors d'utiliser l'information disponible pour définir une transformation qui nous permette de nous ramener à un phénomène stationnaire.

Pour étudier la tendance géographique des deux premiers moments du rapport d'insolation, nous pouvons les considérer chacun comme une réalisation unique d'une fonction aléatoire. On étudie alors leur structure spatiale à l'aide du variogramme. Pour étudier l'anisotropie éventuelle des moments, on construit le variogramme dans différentes directions de l'espace.

#### - Structure spatiale de la moyenne climatologique du sol

Les trois variogrammes construits à partir des moyennes estimées pour chacune des trois périodes aux 84 stations du réseau sol sont représentés Figure (II)-6 abc. Ils apportent des renseignements sur le champ des moyennes de chacune des trois périodes :

- Ce champ présente un gradient nord-ouest/sud-est. Cette dérive apparaît dans la direction (3) du variogramme qui ne présente pas de palier : la variance du champ paraît infinie dans cette direction. Cette constatation reste valable pour toutes les périodes. C'est pourquoi on se réfère au variogramme calculé dans la direction perpendiculaire (2) pour laquelle l'effet de la dérive est insensible. Il est considéré comme le variogramme réel (sous-jacent) de la moyenne. Cette démarche nous évite d'avoir à inférer la dérive. Les deux remarques suivantes sont faites à partir du variogramme de la moyenne dans la direction (2) sans dérive ou (\*) moyen.

- Ce champ possède une bonne structure spatiale, avec une portée supérieure à 600 km pour les trois périodes considérées.

- La variance du champ est fonction de la saison. Elle est plus forte en hiver (100) qu'en été (75) ou au printemps (30).

On conclut à une variabilité spatiale significative de la moyenne climatologique du sol. Les cartes des moyennes pour les trois périodes sont présentées par la suite (Figure III-12).

Fig II-6 :

variogrammes de la moyenne du rapport d'insolation.

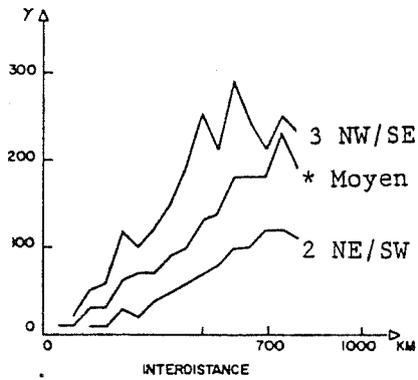
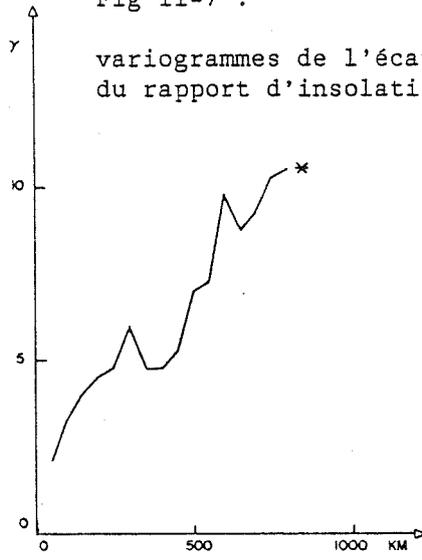
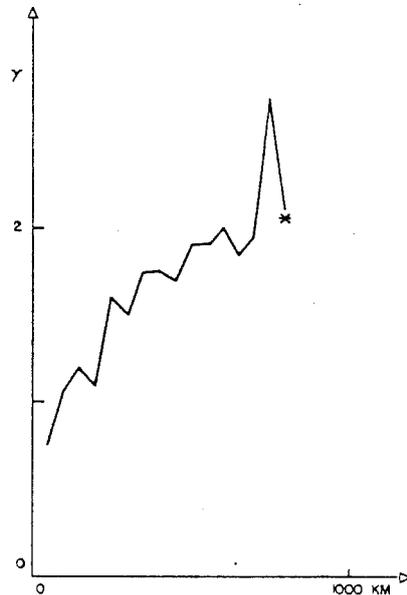
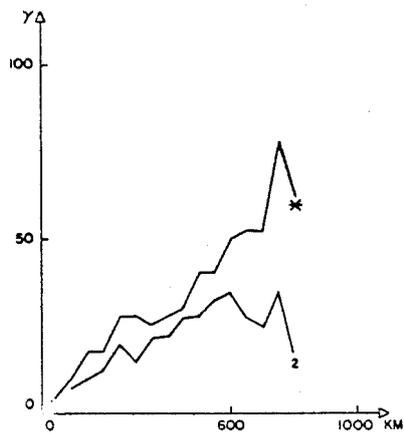


Fig II-7 :

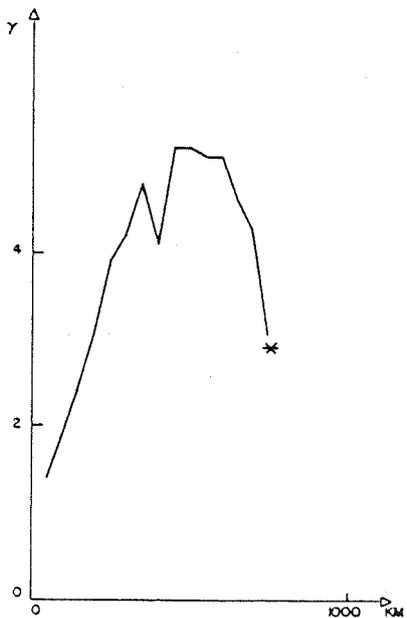
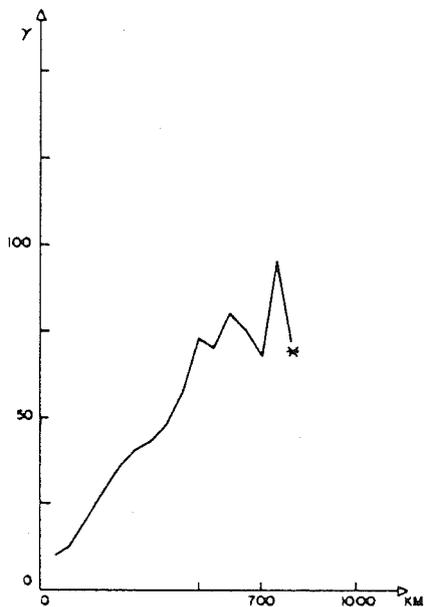
variogrammes de l'écart type du rapport d'insolation



**a**  
HIVER



**b**  
PRINTEMPS



**c**  
ETE

- Structure spatiale de l'écart-type climatologique

Les variogrammes construits à partir de l'écart-type des 84 stations sol (Figure II-7 abc) appellent les trois remarques suivantes :

Il n'apparaît pas de direction privilégiée : le champ des écart-types est isotrope.

On observe une pépite environ égale au cinquième du palier.

La portée, de l'ordre de 300 km, est beaucoup plus courte que celle du champ des moyennes.

Le champ des écart-types apparaît donc comme peu structuré, et de plus, faiblement variable puisque sa variance, rapportée à sa moyenne est respectivement de 5/31, 2/34 et 5/24 pour chacune des 3 périodes (soit environ 10 %). Les cartes qui sont présentées dans la troisième partie I. confirment la faible variabilité, et le peu de cohérence de ce champ. On ne tiendra donc pas compte des variations spatiales des écart-types.

Puisque la moyenne est structurée sans que l'écart-type le soit, la corrélation spatiale de ces deux grandeurs doit être faible. Cette propriété peut être vérifiée en traçant le nuage de corrélation entre la moyenne et l'écart-type calculés en chacune des 84 stations, pour chaque période. Les coefficients de corrélation sont respectivement de 0.5, -0.5 et 0,4 pour l'été, le printemps et l'hiver : le pourcentage de variance expliqué ne dépasse pas 25 %. Nous considérerons que la moyenne et l'écart-type climatologiques ne sont pas liés.

II-1-d. Ergodicité de la F.A.

Une fonction aléatoire est ergodique si chaque réalisation particulière suffit pour obtenir les caractéristiques de cette fonction. En d'autres termes, ses moments spatiaux calculés sur un domaine suffisant seront alors

égaux à ses moments climatologiques. Il faut remarquer qu'une fonction aléatoire stationnaire n'est pas toujours ergodique.

Pour une fonction aléatoire stationnaire non ergodique les moments calculés à partir de chaque réalisation disponible sont différents entre eux et ne permettent pas d'estimer les moments temporels.

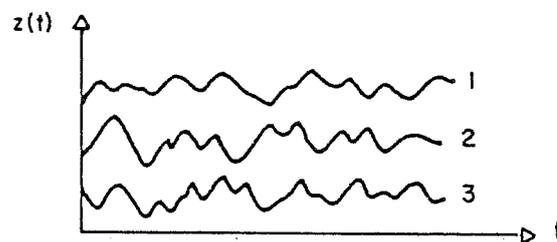


Fig. (II)-8 : 3 réalisations d'une FA stationnaire non ergodique au sens des moments du premier ordre.

On montre (VENTSEL, 1973) qu'un tel processus aléatoire est décomposable en processus aléatoires élémentaires dont chacun a une certaine probabilité de réalisation et des caractéristiques particulières. C'est cette possibilité de décomposition qui est la cause de la non ergodicité.

En particulier, la non ergodicité d'une F.A. peut être liée à la présence, dans sa composition, d'une variable aléatoire  $M_k$  :

$$Y_k(t) = X_k(t) + M_k$$

où  $X_k(t)$  est une fonction aléatoire stationnaire ergodique. Si  $X(t)$  et  $M$  ne sont pas corrélés, alors :

$$m_y(t) = m_x(t) + m_M$$

$$C_{yy}(h) = C_{xx}(h) + V_M$$

où  $m_M$  et  $V_M$  sont les deux premiers moments de la variable  $M$ .

La fonction de covariance de  $Y(t)$  diffère de celle de  $X(t)$  par le terme constant  $V_M$

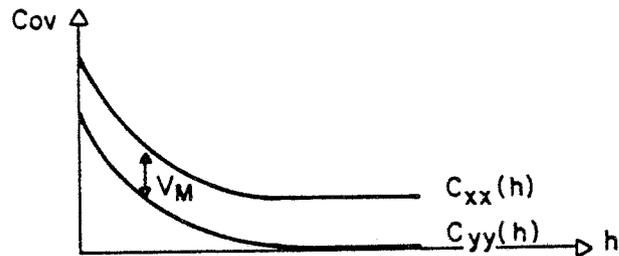


Fig. (II)-9 : Fonction de covariance de deux FA séparées par une variable aléatoire de variance  $V_M$ .

La fonction de covariance  $C_{yy}(h)$  surestime la cohérence strictement spatiale (liée uniquement à la position relative de deux points) de la variable  $Y(t)$ . En effet, une part de la covariance est due alors à la variance de  $M$ . C'est cette part qu'on observe quand  $h$  tend vers l'infini. C'est alors la fonction  $C_{xx}(h)$  qui doit être utilisée pour l'étude de la structure spatiale.

#### - Variabilité de la moyenne spatiale du rapport d'insolation

La moyenne spatiale du rapport d'insolation  $\bar{z}_k$  est estimée, à partir des 84 stations disponibles, pour chaque évènement (cf. Fig. 1). La variabilité de cette moyenne ( $\sqrt{V_{\bar{z}}}=23$ ) est supérieure à l'écart-type d'estimation de celle-ci, calculée sur un nombre fini de stations ( $\sqrt{V}=4$ ) ; elle est donc significative. Il y a alors lieu d'assimiler la moyenne spatiale à la

variable aléatoire  $M_k$  de la décomposition donnée ci-dessus. La moyenne spatiale joue comme un "bouton de réglage" de niveau. Le phénomène s'organise autour de ce niveau qui est significativement variable suivant l'évènement considéré comme nous venons de le voir.

- Variabilité de l'écart-type spatial

L'écart-type est lié de façon non linéaire à la moyenne spatiale, ainsi que le montre le nuage de corrélation (Fig.(II)-10). La forme en croissant du nuage est directement liée au caractère borné du rapport d'insolation : un évènement de moyenne égale à 0 % ou 100 % ne peut correspondre qu'à un écart-type nul. A l'exception de ces valeurs extrêmes, c'est-à-dire pour les valeurs de la moyenne spatiale comprises entre 10 % et 80 % (soit plus de 90 % des cas), la corrélation entre les deux moments spatiaux est inexistante. De plus, la variabilité de l'écart-type  $\nabla$  est faible ( $\sqrt{V_{\nabla}}=7$ ), nous ne prendrons donc pas en compte l'écart type spatial par la suite.

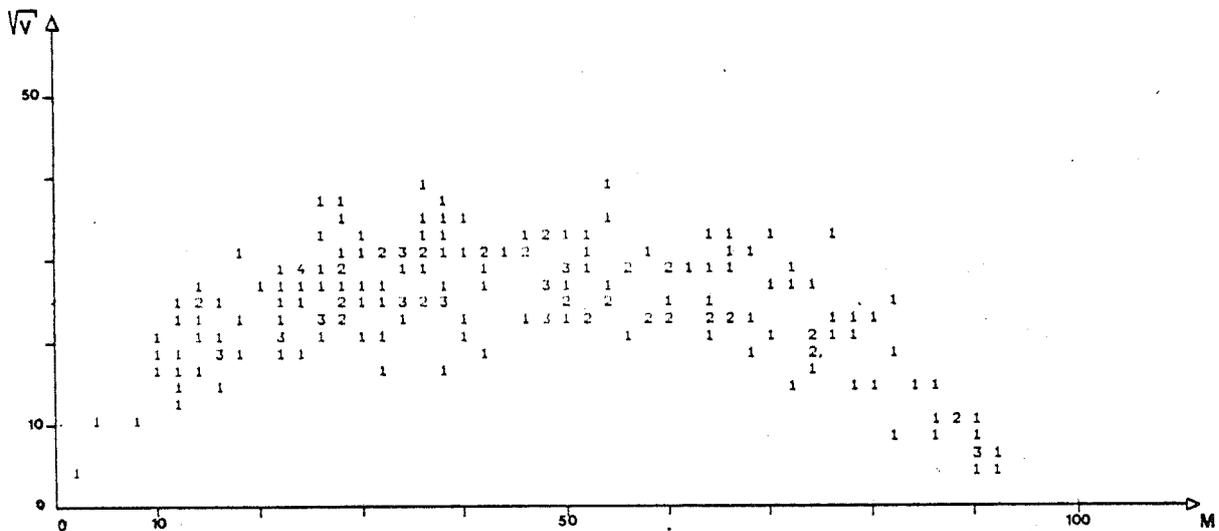


Fig. II-10

Corrélation entre la moyenne et l'écart type spatiaux du rapport d'insolation pour 200 journées.

## II-1-e. Codage proposé

Le rapport d'insolation  $z_k(t)$  est un phénomène dont les caractéristiques statistiques sont complexes. Afin de s'approcher au mieux d'un phénomène stationnaire et ergodique, nous proposons, en fonction des remarques faites dans les paragraphes précédents, un codage qui consiste à faire subir un double centrage (climatologique et spatial) au phénomène brut. En effet, seuls les premiers moments spatial et temporel du phénomène  $Z$  ont une variabilité significative.

- centrage climatologique :

$$y_k(t) = z_k(t) - m_z(t)$$

où  $m_z(t)$  est la moyenne climatologique du phénomène  $Z$  au point  $t$ .

- centrage spatial :

$$x_k(t) = y_k(t) - \bar{y}_k$$

où  $y_k$  est la moyenne spatiale de l'évènement  $k$ .

Il est intéressant de noter que l'ordre dans lequel sont effectués les deux centrages est indifférent. On peut aussi bien faire tout d'abord un centrage spatial ou climatologique du phénomène  $z_k(t)$ .

Notons  $a_k(t)$  la valeur centrée spatialement de  $z_k(t)$  :

$$a_k(t) = z_k(t) - z_k$$

La moyenne temporelle de  $a_k(t)$  peut alors s'écrire :

$$m_a(t) = \frac{1}{n} \sum_{k+1}^n a_k(t)$$

$$\begin{aligned}
&= \frac{1}{n} \sum_k z_k(t) - \frac{1}{n} \sum_k \bar{z}_k \\
&= m_z(t) - \frac{1}{n} \sum_k \frac{1}{D} \int_D z_k(t) dt \\
&= m_z(t) - \frac{1}{D} \int_D \frac{1}{n} \sum_k z_k(t) dt \\
&= m_z(t) - \bar{m}_z
\end{aligned}$$

où  $\bar{m}_z$  est la moyenne spatiale de la moyenne temporelle du phénomène brut  $m_z(t)$ .

Le phénomène  $x_k(t)$ , défini en fonction de  $y_k(t)$  peut également s'exprimer en fonction du phénomène brut  $z_k(t)$ .

$$\begin{aligned}
x_k(t) &= y_k(t) - \bar{y}_k \\
&= z_k(t) - m_z(t) - \overline{z_k(t) - m_z(t)} \\
&= z_k(t) - \bar{z}_k - (m_z(t) - \bar{m}_z)
\end{aligned}$$

Ce qui permet de l'écrire simplement en fonction de la valeur centrée spatialement  $a_k(t)$ .

$$x_k(t) = a_k(t) - m_a(t)$$

La grandeur  $x_k(t)$  est aussi la valeur centrée climatologique de  $a_k(t)$ . L'ordre des centrages est donc bien indifférent. Cette double possibilité pour atteindre  $x_k(t)$  peut se schématiser sur la figure suivante.

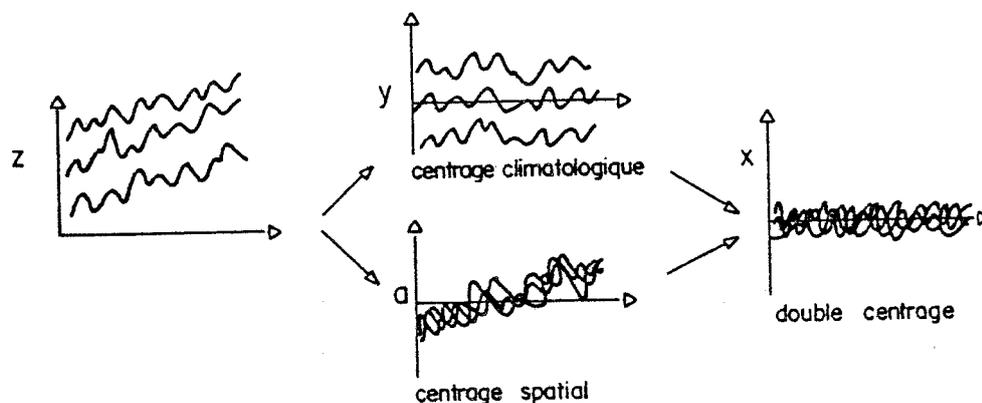


Fig. II-11

#### II-1-e. Distribution des valeurs codées

La distribution des valeurs du rapport d'insolation  $z_k(t)$  (Fig. III-4) n'est pas gaussienne mais bimodale. Il y a une nette prédominance des jours de mauvais temps. Cette dissymétrie de la distribution implique des précautions avant toute manipulation statistique de ces valeurs (la covariance, par exemple, sanctionne une distribution gaussienne). Par contre, la distribution des valeurs  $x_k(t)$  centrées deux fois est symétrique et s'approche d'une distribution normale. Cet effet sur la distribution a été observé pour toutes les périodes. Cette propriété est un argument supplémentaire pour travailler sur les valeurs centrées deux fois.

#### II-1-f. Exemple illustratif

Pour illustrer la nécessité et l'effet du double centrage de nos données, nous avons pris un exemple concret. Cet exemple a été choisi à une dimension afin de simplifier la visualisation : on s'intéresse à une coupe du phénomène. L'axe choisi est orienté Nord-Sud (cf. Fig. II-12). La haute

densité spatiale des mesures satellitaires nous permet une connaissance fine du phénomène suivant cet axe puisque 36 mailles l'échantillonnent.

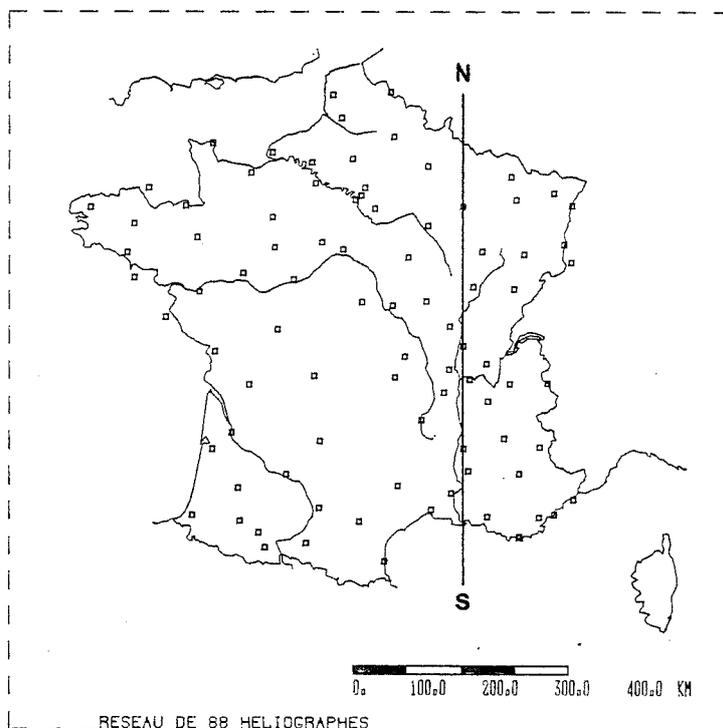
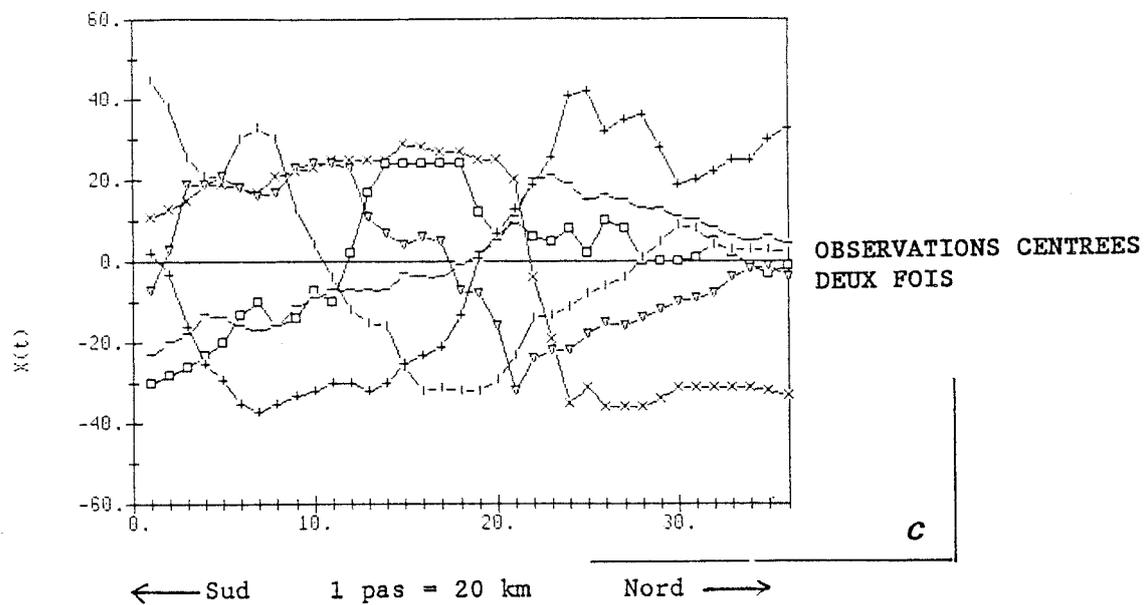
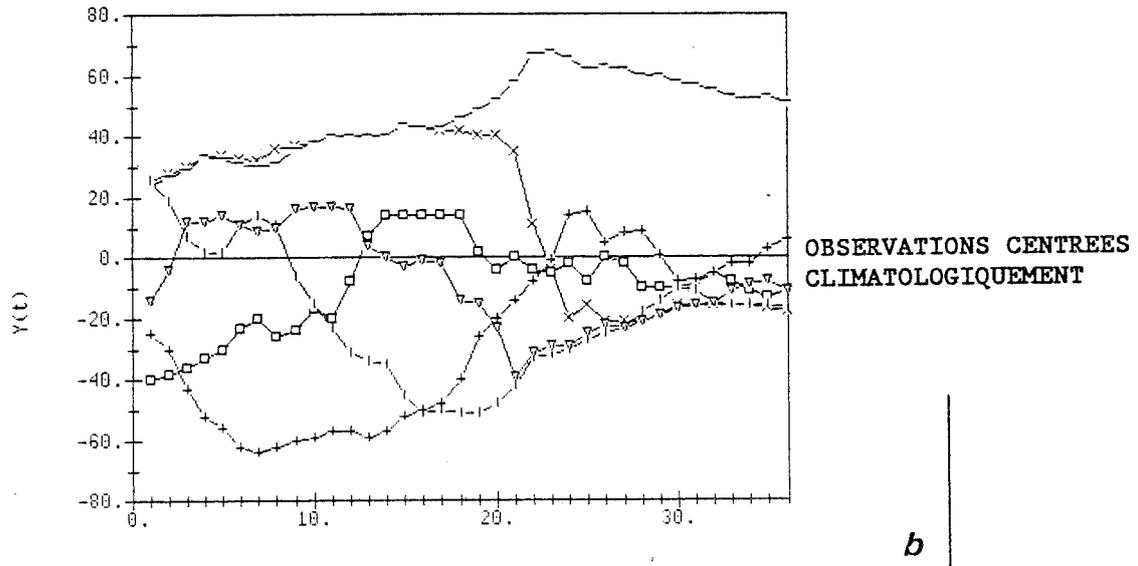
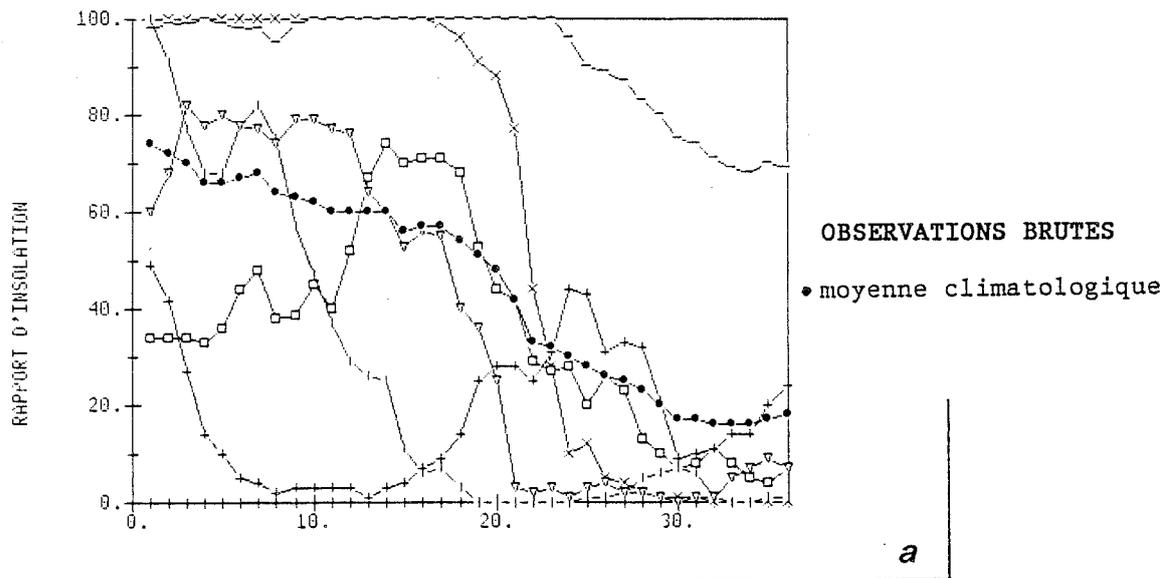


Fig. II-12: axe Nord/Sud sélectionné

Nous avons choisi six journées représentant différents types de temps pour illustrer l'effet successif des deux centrages. Nous avons représenté successivement (Fig. II-13 abc) l'évolution le long d'une ligne nord-sud des valeurs brutes  $z_k(t)$  (et la moyenne climatologique  $m_z(t)$ ), des valeurs centrées par la moyenne climatologique  $y_k(t)$ , et des valeurs centrées deux fois  $x_k(t)$ . (Ceci correspond au chemin du haut de la figure II-11). On observe, pour les différentes étapes, l'évolution du comportement des six réalisations.

- Données brutes : une forte dérive de la moyenne climatologique avec la latitude (elle passe de 75 % au sud à 15 % au nord) est présente. On distingue nettement les journées de beau temps (-) et de mauvais temps (+).

Fig II-12 : Evolution le long d'une ligne N/S de 6 observations



← Sud 1 pas = 20 km Nord →

- Centrage climatologique : la distinction entre les belles (-) et les mauvaises (+) journées est toujours visible, le phénomène n'est pas ergodique.

- Centrage climatologique et spatial : les variations spatiales de  $x(t)$  sont relativement homogènes. Les six réalisations paraissent plus "aléatoires" que précédemment : on ne peut les distinguer les unes des autres. Les six réalisations du phénomène  $X(t)$  présentent toutes une composante périodique de grande amplitude et de période comparable à la dimension du domaine étudié. Les fluctuations de plus faible portée ont une amplitude bien moindre. Ces différentes fluctuations, de portée et d'amplitude différentes, sont caractéristiques de la cohérence spatiale du rapport d'insolation.

#### II-1-g. Conclusion

Le rapport d'insolation  $z_k(t)$  est un phénomène dont les caractéristiques statistiques sont complexes, mais nous avons montré qu'un codage assez simple permet de se ramener à une grandeur  $x_k(t)$  beaucoup plus satisfaisante du point de vue de la stationnarité et de l'ergodicité ainsi que de la distribution. Ce codage consiste en un double centrage (spatial et temporel) des valeurs brutes. Le même codage a été appliqué aux mesures sol et satellite, l'analyse des caractéristiques statistiques des deux types de mesure amenant aux mêmes conclusions.

L'inférence des moments du rapport d'insolation peut alors se faire, à partir de ceux de la grandeur transformée  $x_k(t)$  qui possède les propriétés statistiques adéquates. En ce qui concerne plus particulièrement l'inférence de la structure spatiale du rapport d'insolation, on utilise directement la fonction de structure  $x_k(t)$ . Celle-ci ne caractérise que la cohérence spatiale du rapport d'insolation : on lui a retiré toutes les composantes triviales significatives détectées ci-dessus. C'est cette fonction de structure que l'on considèrera par la suite.

## II-2. Analyse de la structure spatiale des mesures effectuées au sol et par le satellite

Notre but est de quantifier la structure spatiale du rapport d'insolation vu par le sol  $z_{KS}(t)$  et par le satellite  $z_{KT}(t)$ , et de caractériser la co-fluctuation de ces deux types de mesure.

L'analyse des premiers moments du rapport d'insolation a montré la nécessité de travailler sur le phénomène  $x_k(t)$  (qui est une mesure centrée deux fois), afin de s'affranchir des tendances systématiques et ne prendre en compte que les liaisons purement spatiales. Ce chapitre traite donc de l'analyse de la structure du phénomène centré deux fois  $x_k(t)$ .

La diversité des régimes météorologiques rencontrés au cours de l'année nous a amenés à partager celle-ci en trois périodes. Nous allons tout d'abord discuter en détail la structure des différents phénomènes pour une période particulière : le printemps. Puis nous discuterons l'analogie des résultats obtenus avec ceux des deux autres périodes (hiver et été).

### II-2-1. Structure directe des deux types de mesures au printemps

#### II-2-1-1. Les mesures sol

Nous ferons deux remarques préliminaires concernant l'échelle d'observation et l'effet du codage sur la structure spatiale, avant d'aborder l'analyse de la structure des mesures centrées deux fois  $x_k(t)$ .

##### a) Echelle d'observation

Les covariances, comme les corrélations, sont calculées à partir des données des 84 héliographes disponibles. La distance entre deux héliographes est comprise entre 12 et 110 km, avec une moyenne de 60 km. Il ne sera donc pas possible de caractériser d'organisation spatiale se développant à une échelle inférieure à cette distance. Tout ce que l'on pourra détecter de ces microrégionalisations se limitera à leur variabilité

spatiale moyenne. Cette dernière se traduit par une pépite qu'on ne peut distinguer de celle induite par une variabilité strictement ponctuelle comme celle due à des erreurs de mesure par exemple.

b) Effet du codage sur la structure spatiale

Le phénomène  $y_k(t)$  centré de façon climatologique se décompose de la façon suivante :

$$y_k(t) = x_k(t) + \bar{y}_k$$

où  $\bar{y}_k$  est la moyenne spatiale du phénomène  $y$  pour l'évènement  $k$ . En assimilant la moyenne spatiale à une variable aléatoire de variance  $V_y$  et non corrélée à  $X$ , la covariance du phénomène  $y_k(t)$  s'écrit simplement :

$$C_{yy}(h) = C_{xx}(h) + V_y$$

Les deux fonctions de covariance de  $X$  et  $Y$  ne sont séparées que par une constante. Lorsque l'interdistance  $h$  tend vers l'infini, la covariance de  $Y$  ne tend pas vers 0 mais vers  $V_y$ . La covariance de la fonction aléatoire  $X$  est systématiquement plus faible à distance égale, car on lui a retiré la cohérence due à la variable aléatoire  $y$ , indépendante de l'interdistance  $h$ .

Les covariances expérimentales des mesures centrées climatologiquement  $y_k(t)$  et deux fois  $x_k(t)$  ont été tracées sur les mêmes axes (Fig.(II)-14). Les deux courbes se déduisent l'une de l'autre par une translation de  $550(\%)^2$ , très proche de la variance expérimentale de la moyenne spatiale ( $V_y = 590 \%$ ). Le modèle proposé, qui suppose l'orthogonalité des mesures centrées deux fois avec la moyenne spatiale, explique donc les résultats expérimentaux de façon satisfaisante. Les corrélogrammes des phénomènes  $y_k(t)$  et  $x_k(t)$  sont ceux égaux par définition, car le coefficient de corrélation centre les variables.

Dans la suite de ce chapitre, nous ne nous intéresserons plus qu'à la structure du phénomène centré deux fois  $x_k(t)$ .

COVARIANCE SOL AVANT ET APRES CENTRAGE SPATIAL

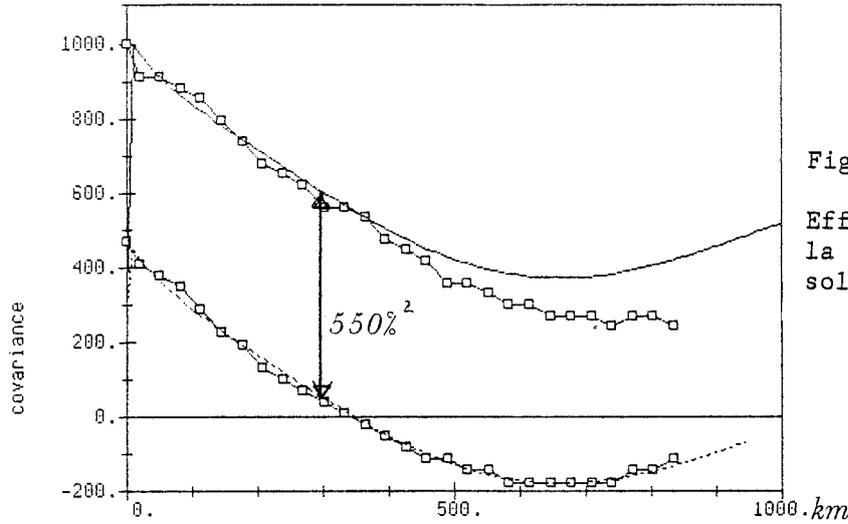


Fig II-14 :

Effet du centrage spatial sur la fonction de covariance du sol (au printemps).

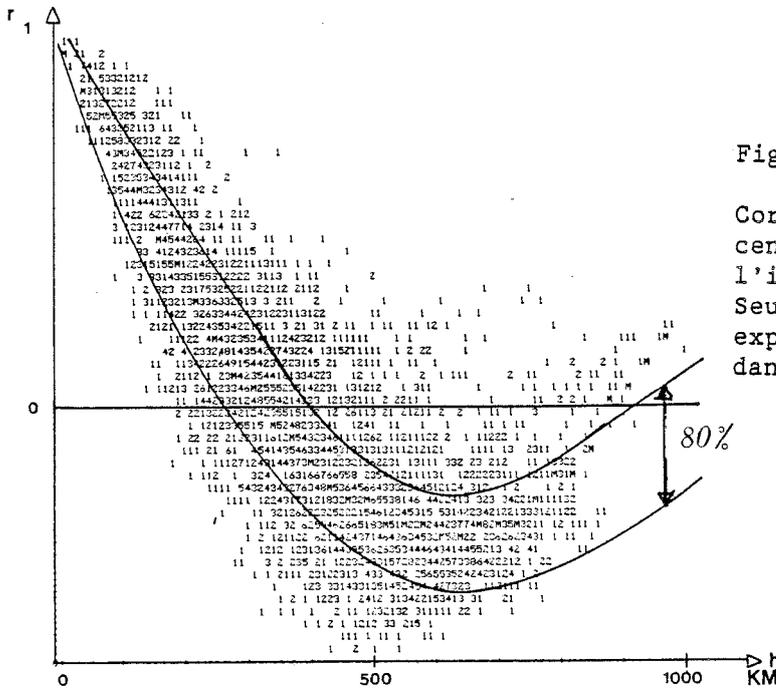


Fig II-16 :

Correlogramme des données sol centrées deux fois avec l'intervalle de confiance 80%. Seules 60% des valeurs expérimentales s'inscrivent dans cet intervalle.

correlogramme sol N0.N45.N90.N135

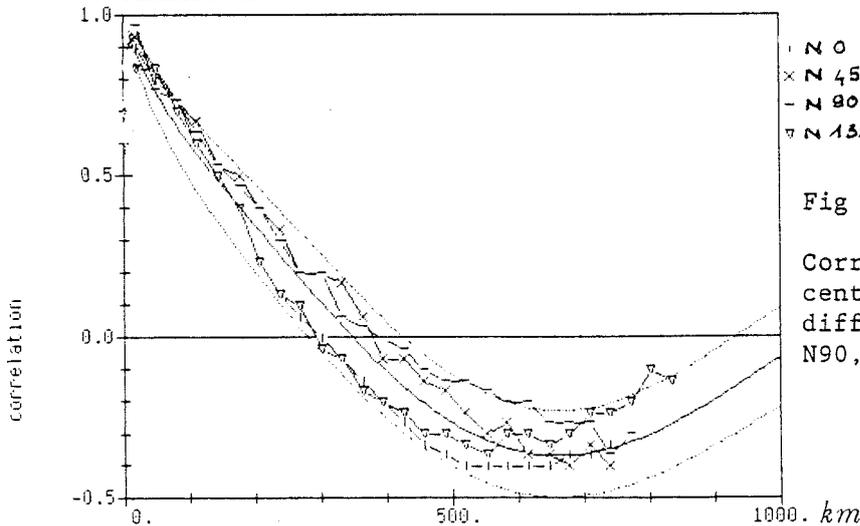


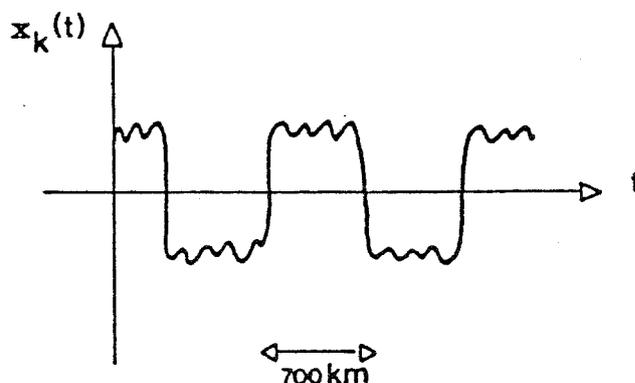
Fig II-17:

Correlogramme des données sol centrées deux fois suivant différents azimuts (N0,N45, N90,N135).

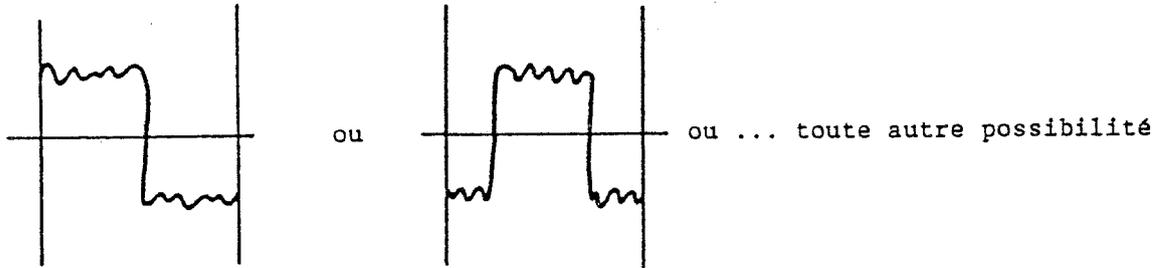
c) Variabilité spatiale des mesures centrées deux fois

A partir de la fonction de covariance des mesures sol centrées deux fois (Fig.(II)-14), on déduit tout d'abord la variance du champ :  $V_x = 480 \text{ } \mu^2$  et la portée ( $p = 350 \text{ km}$ ). En effet, la covariance s'annule pour  $h = 350 \text{ km}$  environ : deux points distants de  $350 \text{ km}$  sont donc indépendants statistiquement. On remarque ensuite des covariances négatives. Celles-ci résultent simplement de l'effet mécanique du centrage par rapport à la moyenne spatiale qui entraîne un nombre approximativement égal de valeurs positives et négatives (car  $X$  est distribué normalement). L'organisation spatiale du phénomène  $X$  a ensuite pour effet le regroupement des valeurs de même signe : la fonction de covariance est positive aux faibles distances et négative ensuite. L'alternance de zones de valeurs positive et négative conditionne la périodicité de la fonction de structure. Les mesures centrées deux fois possèdent une composante de demi-période<sup>de</sup>  $700 \text{ km}$ . Ce qu'on peut illustrer pour une direction par le schéma suivant :

Fig(II)-15



Avec notre fenêtre d'observation d'environ 1000 km, on saisit moins d'une période, soit :



La France est coupée  
en 2 zones de comportement  
opposé.

La zone centrale  
est opposée aux  
bords.

La position de la fenêtre est quelconque par rapport à la période. Sinon, cette tendance apparaîtrait climatologiquement et serait retirée avec la moyenne. Cette conclusion est en accord avec les observations faites sur six champs dans une direction (Fig.II-13 c).

#### d) Modélisation de la fonction de covariance

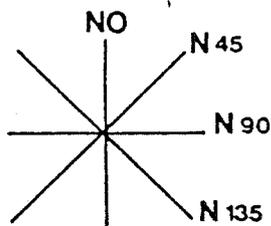
Le modèle utilisé pour représenter la fonction expérimentale est une somme de modèles autorisés. On ajuste tout d'abord un modèle du type  $e^{-ah} \cos bh$  qui décrit la période de 1400 km. La décroissance à l'origine de la courbe expérimentale est plus rapide que celle du modèle en cosinus. On lui ajoute alors une fonction de plus faible portée (200 km) et de faible puissance ( $180 z^2$ ) de la forme  $e^{-ah}$ . Le modèle définitif adopté s'exprime donc (fig.(II)-14) :

$$C_{xx}(h) = 300 e^{-ah} \cos bh + 180 e^{-a'h}$$

e) Homogénéité et isotropie

Dans ce qui précède, nous avons étudié la fonction de structure moyenne, c'est-à-dire la valeur moyenne des coefficients de corrélation obtenus pour des couples de points appartenant à différentes classes de distance. Ces coefficients présentent en fait une certaine dispersion autour de ces valeurs moyennes. Cette dispersion peut être provoquée par l'effet d'échantillonnage, ou par une non stationnarité d'ordre 2. Pour toute valeur de  $r$ , on sait calculer (2ème partie I-4-b) son écart-type d'estimation qui est fonction du nombre d'observations sur lequel il est calculé (71 champs au printemps par exemple). On a superposé au nuage de corrélation les deux courbes correspondant à un intervalle de confiance à 80 % ( $r \pm 1.28 \sigma_r$ ) si la distribution de ces coefficients est gaussienne. Seules, 60 % des valeurs du nuage expérimental (Fig.II-16) s'inscrivent à l'intérieur de ces courbes. Les fluctuations d'échantillonnage ne sont donc pas suffisantes pour expliquer la totalité de cette dispersion. Une certaine non stationnarité est présente.

Afin de voir si cette non stationnarité est due à de l'anisotropie, on a construit les corrélogrammes pour différents azimuts (N0, N45, N90 et N135):



Ainsi que les intervalles de confiance du corrélogramme global (Fig.(II)-17). Il apparaît que les caractéristiques des corrélogrammes sont différentes suivant les directions considérées : la direction N135(V) a une portée plus courte (300 km) que la direction perpendiculaire N45(X) (400 km). Aux grandes distances, seul l'azimut N135 est représenté (grande diagonale de la France). Le corrélogramme "moyen" ne représente plus que cette direction, il est biaisé.

f) Conclusion

Les mesures sol centrées deux fois  $x_k(t)$  ont une portée de 350 km, une variance de  $480 \text{ } \mu^2$ , leur structure est périodique de période 1400 km. Le rapport d'insolation est donc un phénomène bien structuré spatialement, même après une double centrage des mesures brutes. Le fait de se trouver dans le midi par exemple est très important pour  $z_k(t)$  mais n'intervient plus pour  $x_k(t)$ , grandeur centrée climatologiquement. Cependant,  $x_k(t)$  est toujours très liée aux stations voisines. Pour une interdistance de 70 km, ordre de grandeur de l'espacement moyen des stations du réseau, le coefficient de corrélation passe de  $r_{zz}(70) = .87$  pour les valeurs brutes à  $r_{zz}(70) = .80$  pour les valeurs centrées deux fois.

On a détecté d'autre part une anisotropie du phénomène  $x_k(t)$  qui explique en partie la dispersion du corrélogramme. L'azimut N45 possède la portée la plus longue. Pour les faibles distances (inférieures à 200 km) cependant, cette anisotropie n'est pas perceptible. Il est donc raisonnable de considérer la fonction aléatoire X, génératrice des valeurs centrées deux fois, comme stationnaire d'ordre 2, au moins de façon locale : on parlera alors de quasistationnarité de X.

## II-2-1-2. Les mesures satellitaires CACTUS

a) Variabilité spatiale

Le même codage a été adopté pour les mesures satellitaires et pour les mesures sol. La figure (II)-18 montre la fonction de covariance obtenue à partir de 84 mailles satellitaires correspondants aux 84 stations sol. Sa forme globale est la même que celle de la fonction de covariance du sol : cosinus amorti de portée 350 Km. La variance du champ est, elle, plus forte :  $V_{XXT} = 680 \text{ } \mu^2$  contre  $V_{XXS} = 480 \text{ } \mu^2$  pour le sol.

Un raisonnement similaire à celui tenu pour les mesures sol centrées deux fois implique une périodicité à grande échelle (1400 km) des mesures CACTUS

DONNEES CACTUS

Fig II-18 :

Effet du centrage spatial sur la fonction de covariance du satellite (au printemps).

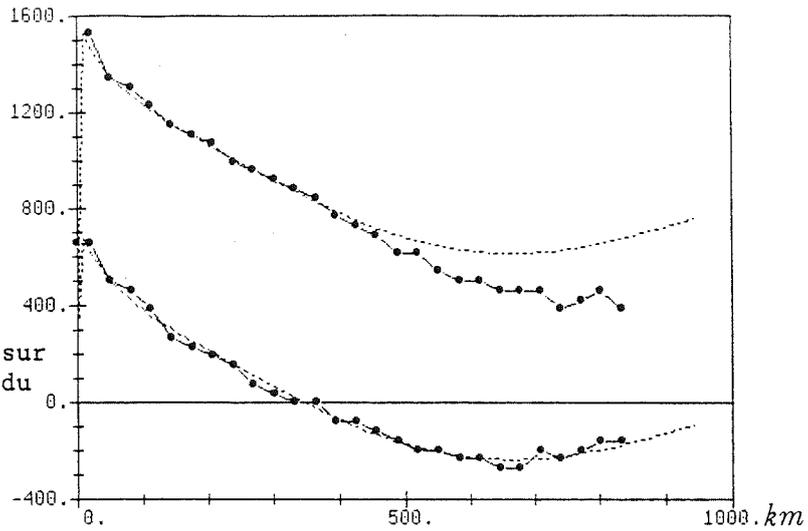


Fig II-19:

Correlogramme des données satellitaires centrées deux fois suivant différents azimuts (N0,N45,N90,N135).

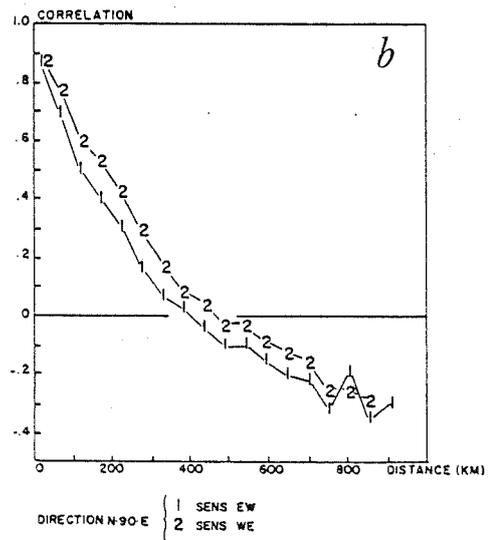
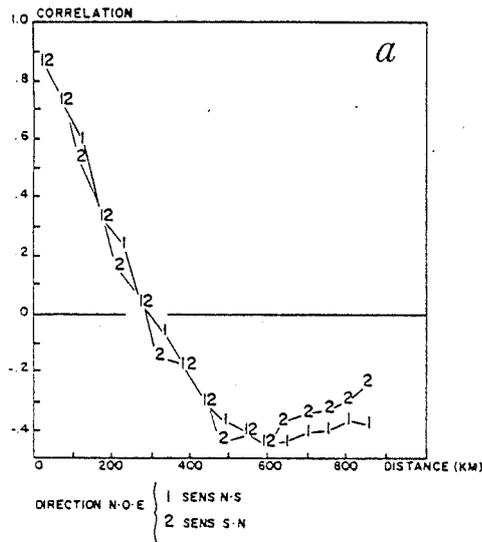
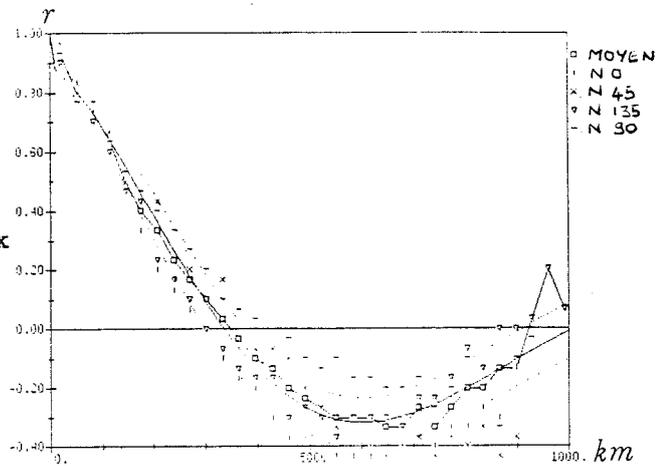


Fig II-20: (d'après Creutin,1987)

Influence du sens sur le correlogramme croisé.

centrées deux fois. Seule l'amplitude de la fonction de structure, c'est-à-dire la variabilité du phénomène, varie.

b) Homogénéité et isotropie

On a tracé le corrélogramme pour les mêmes azimuts que le sol : N0, N45, N90 et N135 (Fig.II-19). L'anisotropie constatée y est comparable : la portée du phénomène est plus grande suivant l'axe N45 que pour l'axe N135.

c) Conclusion

La fonction de covariance des valeurs CACTUS centrées deux fois est de forme identique à celle des valeurs sol, seule l'amplitude est supérieure. Les deux types de mesure ont une perception très cohérente de la structure spatiale de la variable rapport d'insolation centré deux fois, leur différence de nature ne joue que sur leur variabilité.

II-2-2. Structure croisée du sol et du satellite au printemps

La forme identique des fonctions de structure du sol et du satellite ne suffit pas à prouver leur co-fluctuation. C'est pourquoi, il est nécessaire de calculer leur fonction de structure croisée.

Contrairement aux structures directes, la fonction de structure croisée peut être dissymétrique : c'est le cas si les deux phénomènes sont décalés géographiquement. Le corrélogramme est alors sensible au sens dans lequel on considère les couples de valeurs, pour une direction donnée. Aucune dissymétrie significative liée au sens n'est observée pour les deux corrélogrammes croisés suivant les deux azimuts N0 et N90 (Fig.II-20). Il n'y a donc pas de décalage géographique systématique entre les deux types de mesure. Le recalage (cf. Annexe A3) est ainsi validé. Cette constatation étant faite, on ne distinguera plus, par la suite, le sens de la fonction de structure croisée.

Les fonctions de covariances directes et croisées ont été portées sur les mêmes axes afin de les comparer (Fig.II-21). La variance croisée entre les mesures centrées deux fois du sol et du satellite ( $V_{xSxT} = 500 \text{ } \mu^2$ ) a une valeur comprise entre les variances de mesures sol ( $V_{xSxS} = 480 \text{ } \mu^2$ ) et satellite ( $V_{xTxT} = 680 \text{ } \mu^2$ ). La fonction de covariance se situe, pour toute la gamme de distance entre les covariances directes. Celle des valeurs centrées deux fois du satellite étant toujours supérieure en valeur absolue. Cette situation traduit l'excellente co-fluctuation des deux types de mesure puisqu'à l'optimum, on aurait  $C_{xTxS}(h) = \sqrt{C_{xTxT}(h) C_{xSxS}(h)}$ . Le coefficient de corrélation intrinsèque des deux phénomènes :

$$\rho(h) = \frac{C_{xTxS}(h)}{\sqrt{C_{xSxS}(h) C_{xTxT}(h)}}$$

quantifie la qualité de la co-fluctuation. Dans notre cas,  $\rho(h) = .9$  est très proche de 1 et ne varie pas de façon significative avec la distance  $h$ .

En terme de corrélation enfin (Fig.II-22), les trois corrélogrammes sont confondus, sauf pour les faibles distances où la corrélation croisée est légèrement inférieure : on a un effet de pépite sur la structure croisée.

Les deux types de mesure du rapport d'insolation centré deux fois sont donc équivalents à une différence près qui peut résulter d'une microrégionalisation ou d'une erreur.

### II-2-3. Structure des différences

L'excellente co-fluctuation des valeurs sol et télédétection centrées deux fois a été mise en évidence au paragraphe précédent. Les deux types de mesure ne sont séparées que par un bruit, sensible pour les faibles distances. On va chercher ci-dessous à caractériser ce bruit, en étudiant

Fig II-21 :

Covariance des données centrées deux fois. (77 évènements)

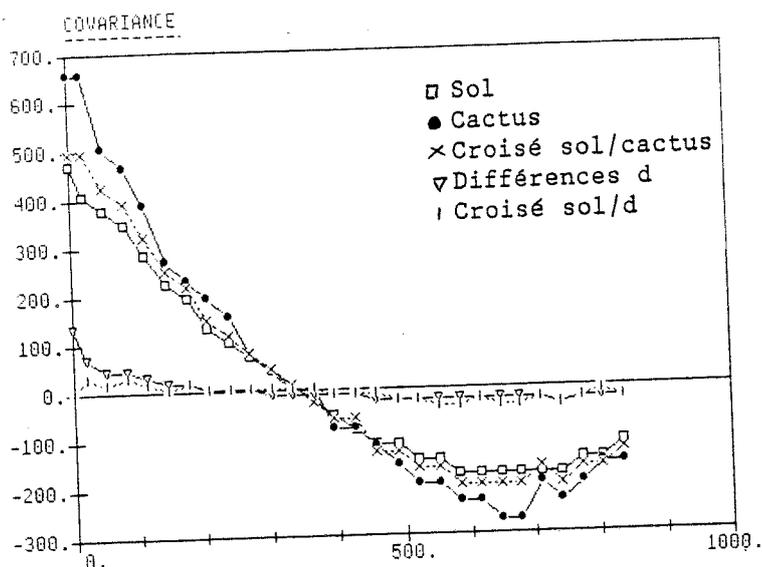


Fig II-22 :

Correlogramme des données centrées deux fois. (77 évènements)

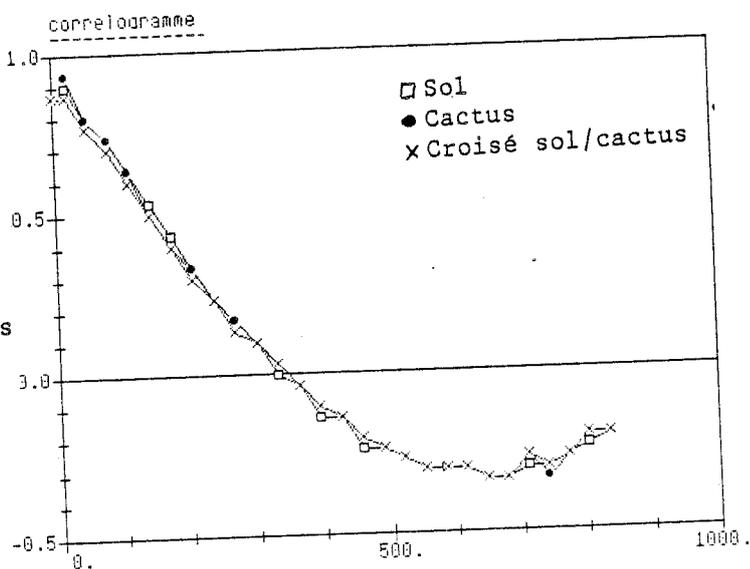
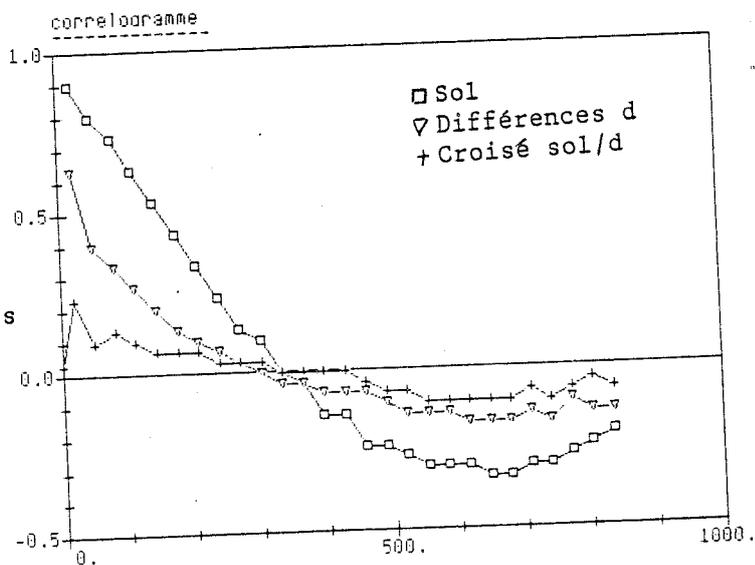


Fig II-23 :

Correlogramme des données centrées deux fois. (77 évènements)



la structure des différences entre les valeurs centrées deux fois de chaque type de mesure.

L'étude de la cohérence spatiale du bruit nous permettra également de juger de l'opportunité d'un recalage de l'image satellite par addition d'une carte des différences. Cette technique, appelée recalage de "guess-field", est couramment utilisée par les météorologues.

La différence entre les valeurs centrées deux fois du sol et de la télédétection se note  $d_k(t)$  :

$$d_k(t) = x_{tk}(t) - x_{sk}(t)$$

On a porté sur le graphe (II)-21 la fonction de covariance des différences (V) et la fonction de covariance croisée entre les différences et le sol (+) afin de tester la co-fluctuation de ces deux phénomènes. La variance des différences est très faible devant celle du sol (25 %). Elle a une décroissance rapide sur 50 km, ce qui est à la limite des possibilités de détection de notre réseau étant donné son espacement moyen pour devenir pratiquement nulle après 200 km. On note cependant de façon très atténuée une périodicité de fréquence identique à celle observée sur les données sol (portée de 350 km). On retrouve cette même périodicité pour la structure croisée entre le sol et les différences. La structure des différences semble encore (faiblement) influencée par la structure propre du phénomène.

Les fonctions de covariance nous ont permis de juger de la variabilité spatiale (ou puissance) relative des différents phénomènes. Les corrélogrammes standardisent les échelles (Fig.II-23). La période 1400 km y apparaît clairement, avec une amplitude supérieure aux fluctuations d'échantillonnage : elle est significative.

Les différences entre les valeurs centrées deux fois du sol et de la télédétection ont une variabilité quatre fois plus faible que celle des valeurs sol. La fonction de structure montre une décroissance rapide pour les 50 premiers kilomètres, mais la corrélation ne s'annule pas avant 350 km. Cette longue portée, caractéristique d'un phénomène lisse, est

surprenante : on s'attendait en effet à un bruit de faible portée. Elle s'explique par la corrélation des valeurs sol avec les différences. En effet, celles-ci ne sont pas strictement orthogonales comme le montre leur corrélogramme croisé.

On s'est également intéressé à la co-fluctuation des différences et des valeurs de télédétection. Leur covariance croisée est toujours supérieure en valeur absolue à celle obtenue avec les valeurs sol. Ces phénomènes ne sont pas indépendants.

La portée relativement longue des différences justifie leur interpolation éventuelle. Cependant, les cartes ainsi obtenues ne semblent pas être en mesure de recaler l'image satellite de façon efficace, puisque l'information qu'elles apportent est déjà comprise en partie dans l'image satellitaire.

#### II-2-4. Relation entre les valeurs centrées mesurées par le sol et le satellite au printemps

##### II-2-4-1. Détermination de la relation

Les cinq fonctions de covariance calculées précédemment ont été portées sur les mêmes axes. Elles s'annulent toutes à la distance  $h = 350$  km. L'analyse de ces cinq fonctions nous a conduit à proposer la décomposition orthogonale suivante pour la mesure de télédétection centrée deux fois : un phénomène sol  $x_{Sk}(t)$ , auquel s'ajoute un bruit  $\mathcal{E}_k(t)$  de plus faible portée.

$$x_{Tk}(t) = a x_{Sk}(t) + \mathcal{E}_k(t)$$

avec la condition d'orthogonalité :

$$C_{xS\mathcal{E}}(h) = 0 \quad \text{pour toute valeur de } h.$$

Les différences calculées précédemment s'écrivent alors :

$$d_k(t) = x_{Tk}(t) - x_{Sk}(t) = (a-1) x_{Sk}(t) + \mathcal{E}_k(t)$$

Elles s'expriment en fonction des valeurs sol  $x_{Sk}(t)$ . Elles n'en sont décorrelées que pour  $a = 1$  ; dans ce cas,  $d_k(t)$  se confond avec  $\mathcal{E}_k(t)$ .

Les cinq fonctions de structure croisées et directes s'expriment alors simplement en fonction de celle de  $x_S$  et  $\mathcal{E}$ .

$$\begin{aligned} C_{xTxT}(h) &= a^2 C_{xSxS}(h) + C_{\mathcal{E}\mathcal{E}}(h) \\ C_{xTxS}(h) &= a C_{xSxS}(h) \\ C_{dd}(h) &= (a-1)^2 C_{xSxS}(h) + C_{\mathcal{E}\mathcal{E}}(h) \\ C_{dxS}(h) &= (a-1) C_{xSxS}(h) + C_{\mathcal{E}\mathcal{E}}(h) \end{aligned}$$

La fonction de covariance croisée filtre le bruit  $\mathcal{E}_k(t)$  : il n'apparaît pas dans son expression.

Remarque : Si le bruit est blanc (de portée nulle), on aura, pour tout  $h$  supérieur à zéro :

$$\left. \begin{aligned} C_{xTxT}(h) &= a^2 C_{xSxS}(h) \\ C_{xTxS}(h) &= a C_{xSxS}(h) \\ C_{dd}(h) &= (a-1)^2 C_{xSxS}(h) \\ C_{dxS}(h) &= (a-1) C_{xSxS}(h) \end{aligned} \right\} \forall h > 0$$

Il y a alors corrélation intrinsèque : les différentes fonctions de covariance se déduisent d'un modèle commun unique, par affinité.

La décomposition orthogonale des mesures de télédétection centrées deux fois, proposée ci-dessus, entraîne le conditionnement des cinq fonctions de structure. Le nombre de paramètres à ajuster s'élève seulement à trois : les deux fonctions de covariance  $C_{xSxS}(h)$  et  $C_{\mathcal{E}\mathcal{E}}(h)$  et le coefficient  $a$ .

Il existe donc deux contraintes redondantes. Elles serviront à valider l'adéquation de cette décomposition à la réalité expérimentale.

Application au rapport d'insolation centré deux fois :

La fonction de covariance  $C_{xSxS}(h)$  est modélisée (cf. Fig.II-24-b). L'ajustement du coefficient  $a$ , et la modélisation de la fonction de structure du bruit  $C_{\mathcal{E}\mathcal{E}}(h)$  sont obtenus à partir des deux courbes  $C_{xxT}$  et  $C_{dd}$  (Fig.II-24 d et c), qui donnent  $a = 1.15$ . La modélisation des deux courbes  $C_{xTxT}$  et  $C_{dx}$  est alors parfaitement définie .

$$x_{Tk}(t) = 1.15 x_{Sk}(t) + \mathcal{E}_k(t) \quad \text{avec } x_S \perp \mathcal{E} .$$

Ayant ainsi déterminé le coefficient  $a = 1.15$ , on peut calculer les résidus  $\mathcal{E}_k(t)$  en tout point :

$$\mathcal{E}_k(t) = x_{Tk}(t) - ax_k(t)$$

et tracer leur corrélogramme expérimental ( $\nabla$ ) (Fig.II-25). La portée des résidus est de 200 km, leur pépite est forte, d'où l'appellation de "bruit". La structure croisée entre ce bruit et les valeurs sol ( $X$ ) est nulle ; le bruit  $\mathcal{E}(t)$  est donc orthogonal aux mesures sol, alors que les différences  $d_k(t)$  ne l'étaient pas (Fig.II-23 b). On vérifie l'hypothèse  $C_{x_S \mathcal{E}}(h) = 0$ .

#### II-2-4-1. Interprétation

La mesure de télédétection centrée deux fois se décompose comme la somme de la mesure sol centrée deux fois et pondérée du coefficient 1.15 et d'un bruit  $\mathcal{E}(t)$  de plus faible portée. Cette relation, fondée sur l'étude des fonctions de structure directe et croisée des deux phénomènes peut être mise en rapport avec leur différence de nature.

Les mesures sol sont ponctuelles mais continues dans le temps. Les mesures CACTUS, au contraire, représentent une moyenne sur une maille ( $400 \text{ km}^2$ ), mais résultent de prise de vue "instantanée" (météosat balaye la France (200 lignes) en une minute, une maille (environ  $4 \times 4$  pixels) en 3 secondes).

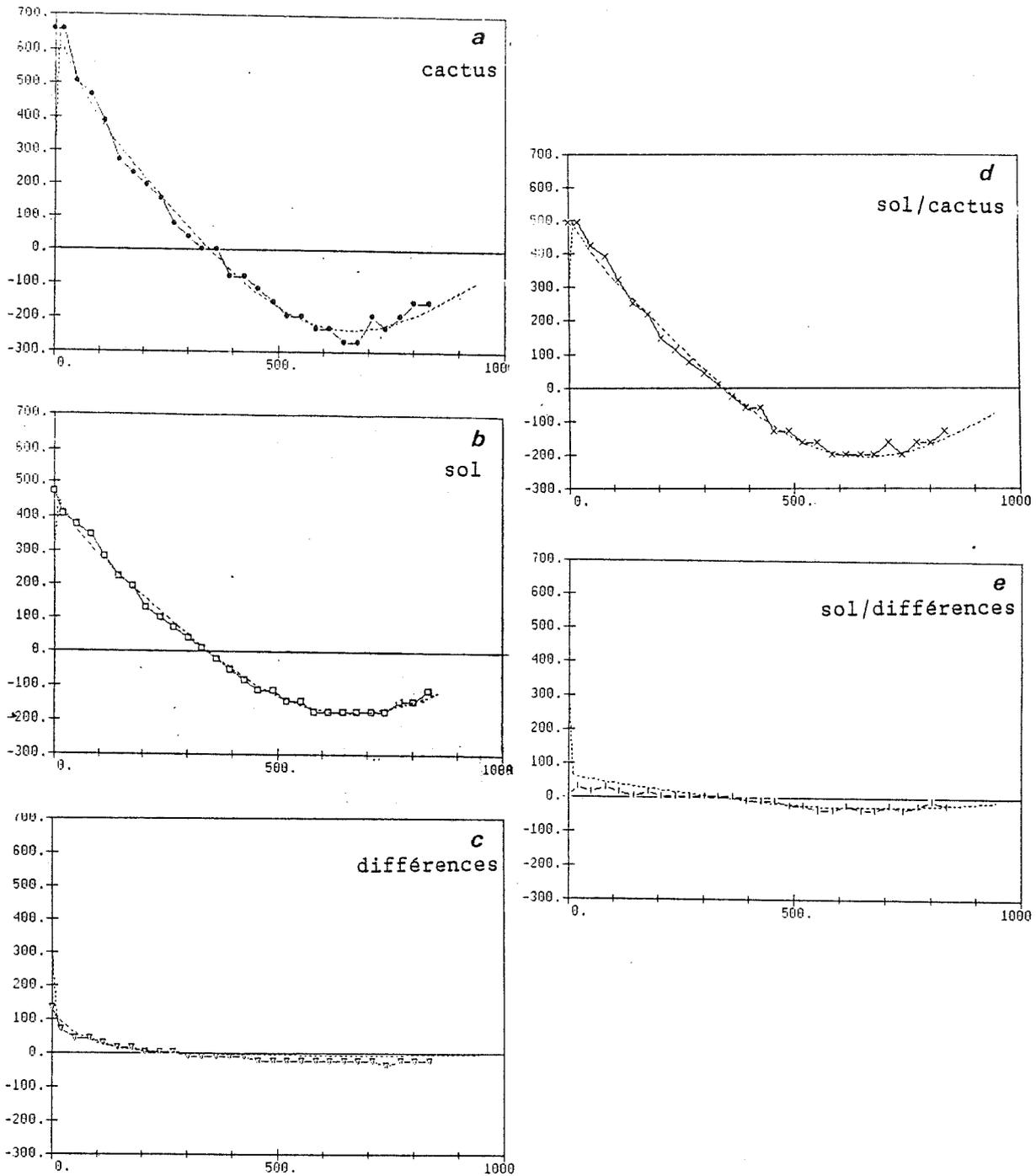


Fig II-24 :

Ajustement d'un modèle à 3 paramètres à l'aide 5 fonctions expérimentales .

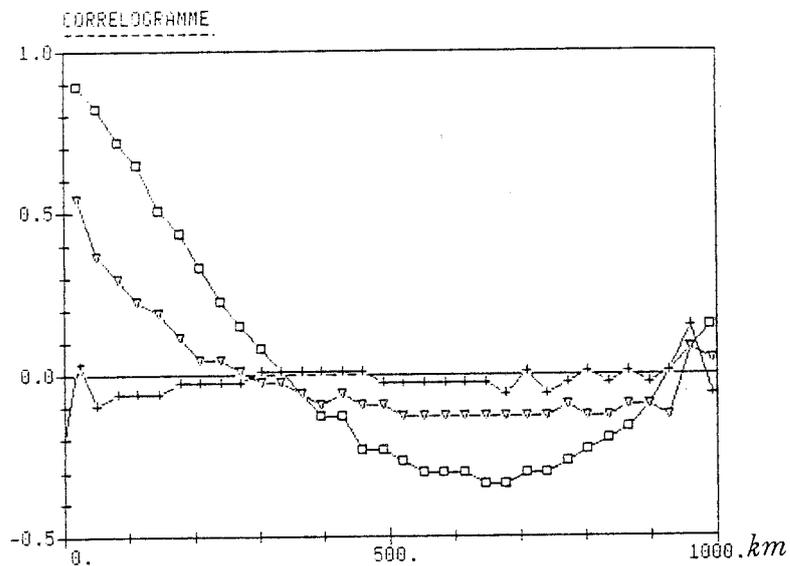


Fig II-25 :

Correlogramme des valeurs centrées deux fois

- Sol
- ▽ Résidus  $\varepsilon$
- + Croise sol/résidus

(71 évènements)

La structure croisée est nulle: le résidu est orthogonal aux mesures sol.

Ces deux sortes d'intégration vont dans le sens d'une compensation de l'une par l'autre : les nuages contenus dans la maille CACTUS sont, pour partie, ceux qui vont passer au-dessus de l'héliographe pendant la demi-heure séparant deux clichés météosat. La taille de la maille, déterminée par C.PASTRE, a été choisie en fonction de la vitesse moyenne des nuages en France et de l'intervalle de temps séparant deux clichés météosat (cf. 1ère partie). Nous allons discuter ci-dessous de l'effet de ces deux types d'intégration appliqués à un même phénomène.

- Effet de l'intégration spatiale sur la fonction de structure

L'intégration sur une surface  $D$  d'un phénomène  $X(t)$  (ou régularisation de  $X(t)$ ) est un autre phénomène  $\bar{X}(t)$ , dont la covariance est liée à celle de  $X(t)$ .

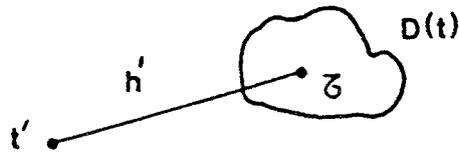
$$\bar{X}(t) = \frac{1}{D} \int_{D(t)} X(z) dz$$

On suppose  $X$  stationnaire d'ordre 2, de moyenne nulle. On montre (JOURNEL and HUIJBRECHTS, 78) que  $X$  est alors également stationnaire d'ordre 2. On peut exprimer la structure croisée entre  $X$  et  $\bar{X}$  en fonction de la structure directe du phénomène ponctuel  $C_{XX}(h)$

$$C_{X\bar{X}}(t, t') = E (X(t) \bar{X}(t'))$$

$$= \frac{1}{D} \int_{D(t')} C_{XX}(t, z) dz \quad (i)$$

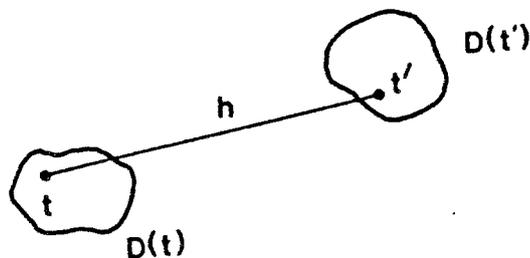
Cette covariance est égale à la valeur moyenne de la covariance ponctuelle  $C_{XX}(h)$  quand le point  $z$  décrit le domaine  $D$ .



La fonction de covariance directe  $C_{XX}(t, t')$  s'exprime également en fonction de  $C_{XX}(h)$  :

$$C_{\bar{X}\bar{X}}(t, t') = \frac{1}{D^2} \int_{D(t)} \int_{D(t')} C_{XX}(z, z') d z d z' \quad (j)$$

$C_{\bar{X}\bar{X}}$  est la valeur moyenne de  $C_{XX}(h)$  quand  $z$  et  $z'$  décrivent indépendamment les domaines  $D(t)$  et  $D(t')$ .



L'expression (j) montre que la covariance  $C_{\bar{X}\bar{X}}(0)$  sera toujours inférieure à celle du phénomène ponctuel  $C_{XX}(0)$ . En effet, la covariance est une fonction décroissante exponentielle de la distance ( $X$  est supposé stationnaire), ce qui implique l'inégalité  $C_{XX}(0) < C_{XX}(h)$  et par suite

$$C_{\bar{X}\bar{X}}(0) = \frac{1}{D^2} \iint C_{XX}(z, z') d z d z' < C_{XX}(0)$$

L'influence de l'intégration (ou régularisation) du phénomène  $X(t)$  sur la covariance est d'autant plus important que la variation de  $C_{XX}(h)$  est grande sur le domaine. A la limite, lorsque le domaine tend vers un point, ou pour des interdistances de l'ordre de la portée,  $C_{XX}(h)$  est pratiquement constant sur le domaine et on a :

$$\begin{aligned} C_{X\bar{X}}(h) &= C_{XX}(h) \\ C_{\bar{X}\bar{X}}(h) &= C_{XX}(h). \end{aligned}$$

Dans notre cas, le domaine  $D$  de 20 km de côté est petit devant la portée de 350 km du rapport d'insolation centré deux fois. L'effet de l'intégration ne sera sensible qu'aux faibles distances, pour atteindre son maximum pour  $h = 0$ .

L'intégration spatiale du phénomène  $X(t)$  a un effet variable en fonction de la distance sur la covariance. A la distance nulle, la variance du phénomène intégré est inférieure à celle du phénomène ponctuel ; à l'infini elles sont égales.

#### - Effet de l'échantillonnage temporel sur la fonction de structure

L'insolation est un phénomène continu. En tout instant, on détermine si il y a eu insolation ou pas (si le ciel a été clair ou couvert), en un lieu donné. La durée d'insolation est l'intégration temporelle de ces résultats. Le satellite ne fait pas une mesure continue. Plus ses prises de vue sont espacées et moins le résultat sera précis. A la limite avec une seule photo par jour, l'estimation sera binaire : clair ou couvert. La diminution de la fréquence des photos a pour effet d'augmenter la variabilité du rapport d'insolation (dont la discrétisation est moins fine qu'avec une mesure continue). Pour un jour dégagé avec un bref passage nuageux au lieu  $t$ , le satellite aura peu de chance d'avoir une prise de vue exactement à cet instant et estimera probablement une journée parfaitement dégagée au lieu  $t$  (ce qui entraîne une surestimation des valeurs fortes, en moyenne). L'effet de l'échantillonnage sera d'autant plus important que le phénomène est peu

autocorréllé dans le temps. L'échantillonnage joue sur l'estimation de la fonction de covariance du phénomène, mais cet effet est semblable en tout lieu.

#### - Conclusion

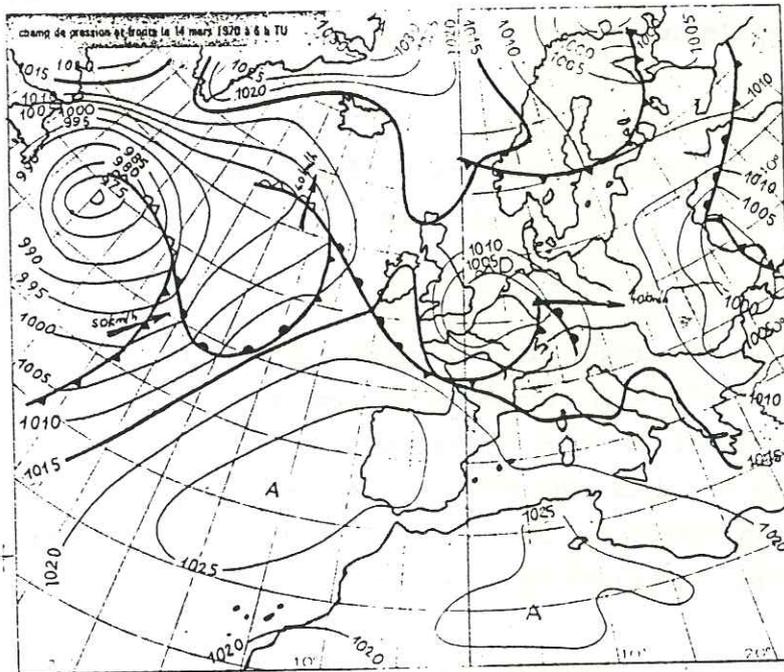
L'intégration spatiale et l'échantillonnage temporel ont des effets opposés sur l'estimation de la fonction de covariance du phénomène ponctuel. L'un tend à la sous-estimer et l'autre à la surévaluer. Les résultats expérimentaux montrent que la fonction de covariance du satellite (intégré et échantillonné) est toujours supérieure à celle du sol (ponctuel et continu). Il est donc probable que c'est l'échantillonnage temporel qui joue le rôle prédominant et qu'il se traduit par le coefficient  $a$ , égal à 1.15. L'échantillonnage temporel, initialement prévu au pas de temps de la demi-heure, a finalement été espacé à l'heure, afin de réduire le temps de calcul du programme CACTUS dans sa version opérationnelle. L'interprétation des fonctions de structure nous conduit à conclure que la maille élémentaire (20x20 km<sup>2</sup>) aurait alors également dû être agrandie. On aurait ainsi une meilleure adaptation des estimations CACTUS aux données sol ponctuelles mais intégrées.

#### II-2-5. Lien entre la fonction de structure et les caractéristiques météorologiques du temps en France

Les caractéristiques de la structure spatiale décrites ci-dessus peuvent, elles, être mises en rapport avec les situations météorologiques classiques se développant sur la France ? Ayant mis en évidence les relations qui lient, après double centrage, toutes les fonctions de covariance à celle du sol, on ne s'attachera qu'à celle-ci, et plus précisément à la période de 1400 km dont on aimerait expliquer l'origine. Bien que cette période ait été observée sur les valeurs centrées deux fois, on peut la lier au phénomène brut pour ce qui concerne sa cohérence purement spatiale.

Le rapport d'insolation caractérise le pourcentage de la durée astronomique du jour pour lequel le ciel est dégagé. Ce rapport est donc fonction des passages nuageux rencontrés au cours de la journée, au point d'observation.

Malgré la grande diversité des situations météorologiques rencontrées en France, on peut donner une image schématique du principal type de temps au printemps : le régime perturbé d'ouest. Au printemps, le temps est régi essentiellement par les grands centres d'action que sont l'anticyclone des Açores et la zone dépressionnaire d'Islande. La position respective de ces deux centres est à l'origine des perturbations, qui sont créées par familles. Elles se déplacent d'ouest en est. A chaque perturbation est associé un système nuageux. C'est le passage des fronts des perturbations qui conditionne principalement l'alternance de ciel nuageux et dégagé. La distance entre deux fronts est de l'ordre du millier de kilomètres - ce qui est également l'ordre de grandeur de la période des valeurs sol centrées deux fois.



Une situation fréquente : passage d'une famille de perturbations à 50° de latitude N

Nous n'avons décrit ci-dessus que des situations instantanées. Le rapport d'insolation représente une valeur intégrée dans le temps. L'ordre de grandeur des vitesses de déplacement des perturbations est de 500 km/24h soit environ 200 km pendant la durée du jour. Ce déplacement est faible en regard de la distance séparant deux fronts (quelques milliers de kilomètres) et ne la modifie donc pas sensiblement.

Le sens général de passage d'Ouest en Est des perturbations induit une cohérence spatiale différente suivant la direction d'observation. On explique ainsi l'anisotropie observée entre les directions E/W et N/S. Cependant, cette explication n'est à l'évidence pas suffisante. En particulier, elle ne justifie pas la période d'environ 1200 km observée également pour la direction Nord/Sud. Cette opposition n'est pas de nature climatologique puisqu'on travaille sur des mesures centrées deux fois : elle ne peut provenir que d'opposition de comportement en mouvement. Il est bien évident que d'autres régimes météorologiques peuvent intervenir mais dans quelle mesure ? Le lien entre la période de la fonction de structure et la réalité météorologique qu'on a souligné ici existe mais ne permet qu'une explication partielle de celle-ci.

#### II-2-6. Structure spatiale du sol et du satellite au cours de l'année

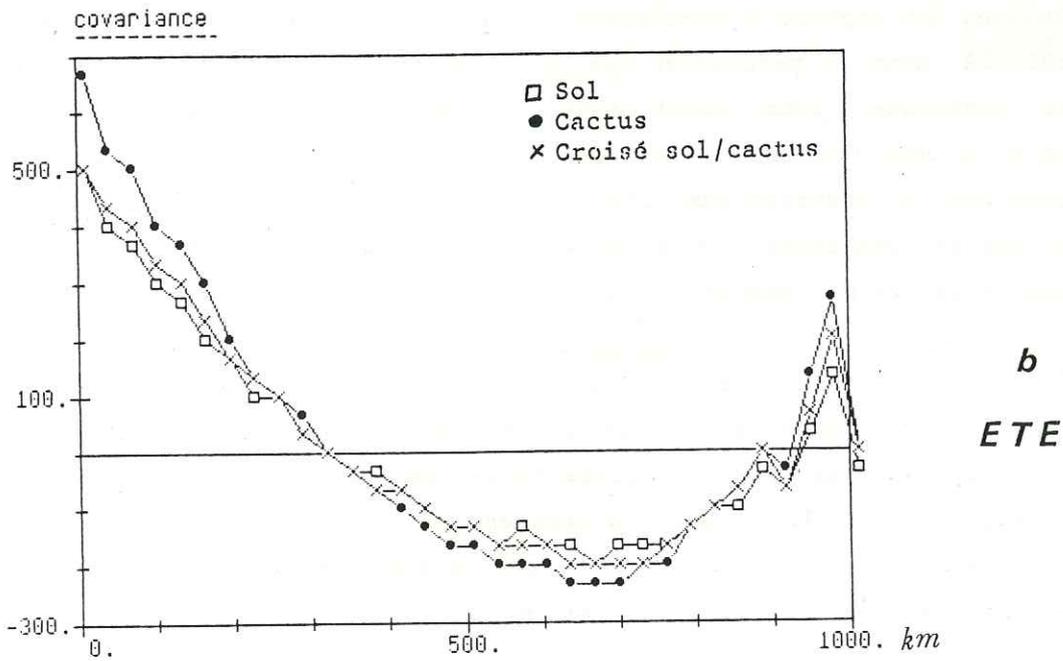
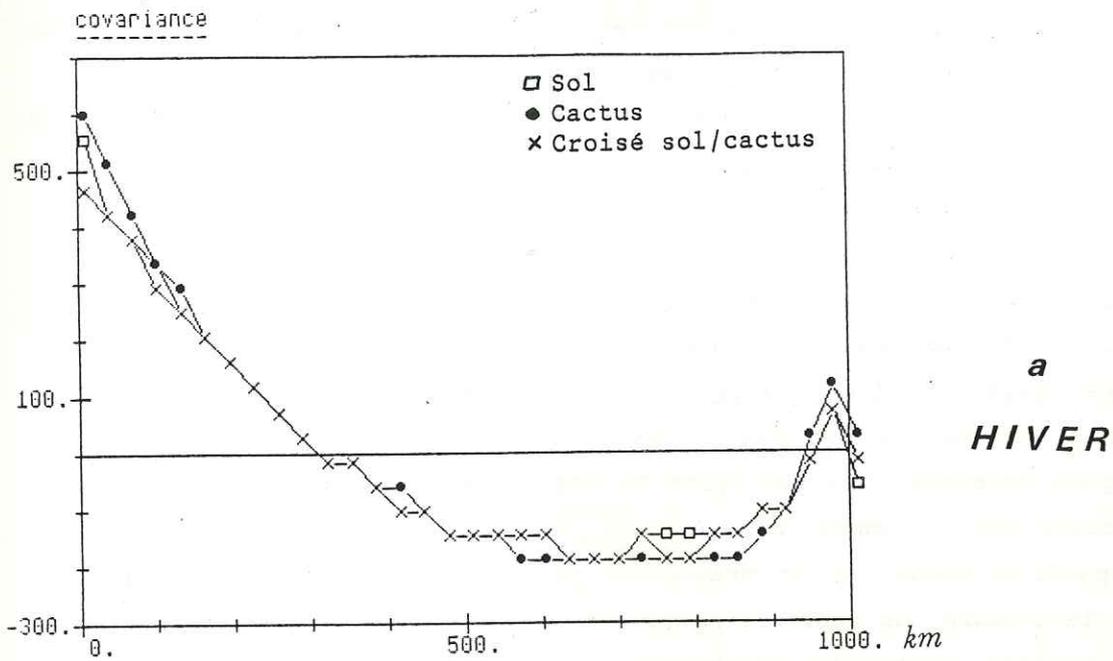
Nous allons maintenant traiter rapidement des deux autres périodes disponibles : l'hiver et l'été, en résumant brièvement les résultats obtenus au printemps.

Nous avons effectué un centrage climatologique et spatial des mesures brutes, de la même manière et pour les mêmes raisons que pour le printemps. Les fonctions de covariance directes et croisées ont été tracées sur les mêmes axes pour chacune des deux périodes considérées (Fig.(II)-26 a et b). Ces figures sont à rapprocher de la figure (II)-21 a. Pour chacune des trois périodes, les trois fonctions de covariance sont très proches.

Les fonctions de structure directes des deux types de mesure sont cohérentes : même allure, même portée de 350 km. Elles ne se distinguent que pour les faibles distances, la covariance du satellite est alors systématiquement plus forte, ce qui traduit une amplitude plus forte des valeurs  $x_{Tk}(t)$ , écarts à la valeur moyenne spatiale. Cette différence est moins sensible en hiver.

Fig II-26 :

Covariance directes et croisées entre les mesures sol-et cactus.



La covariance croisée est pratiquement maximale puisque le coefficient de corrélation intrinsèque  $\rho(h)$  est respectivement de .9, .9 et 1 au printemps, en été et en hiver (à l'exception des courtes distances). Il vaut .87, .86 et .83 pour chacune des trois périodes respectivement, à la distance nulle. La cofluctuation des valeurs centrées obtenues par les deux types de mesures est excellente pour toutes les périodes considérées. En hiver, elle est meilleure aux grandes distances et légèrement moins bonne pour les faibles distances.

Les trois figures suivantes (Fig.(II)-27 abc) illustrent l'évolution au cours de l'année des fonctions de covariance du sol, du satellite et croisée. Celle-ci n'est pratiquement pas perceptible (en particulier pour la structure du sol) et mis à part la distance nulle des fonctions de structure directes. Les fonctions de structure croisées sont parfaitement confondues toute l'année, même à l'origine. Les valeurs centrées deux fois du rapport d'insolation ne présentent pas d'évolution caractéristiques au cours de l'année. On pourrait alors penser à regrouper les périodes, mais on a préféré garder la séparation, car on a vu que ces périodes correspondent à des types de temps et donc à des processus générateurs différents.

En conclusion, le rapport d'insolation centré deux fois est un phénomène bien structuré, dont la perception par les deux types de mesure disponible est très cohérente. Leur coefficient de corrélation intrinsèque est supérieur à .9 pour toutes les périodes, ce qui est excellent. La baisse de la cofluctuation observée aux faibles distances est attribuée à un bruit de faible portée, orthogonal aux valeurs centrées deux fois du sol, dû à la différence de nature des mesures.

La modélisation des fonctions de structure ainsi proposée est fondée sur l'hypothèse de stationnarité d'ordre 2 des valeurs centrées deux fois. Celle-ci n'explique de façon satisfaisante la dispersion des nuages expérimentaux que pour les courtes distances (il y a quasistationnarité). Le modèle ajusté aux données expérimentales n'a donc pas la même fiabilité pour toute la gamme de distances considérées.

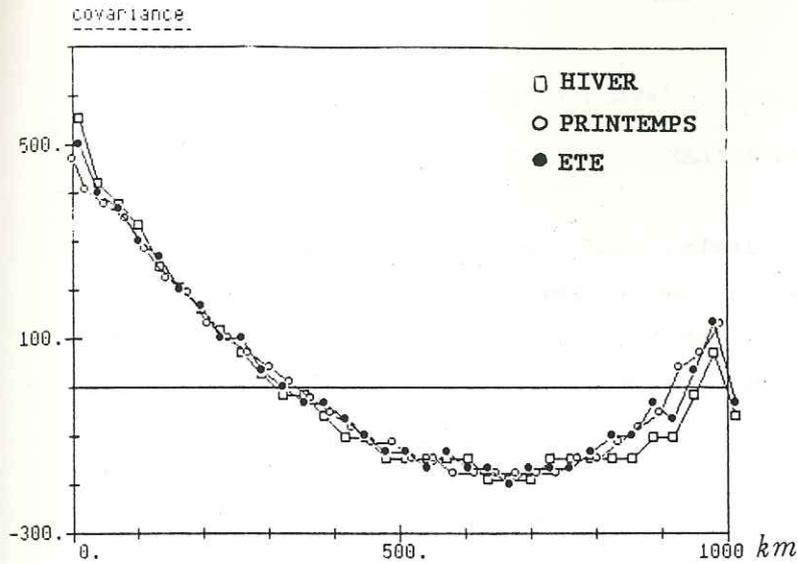
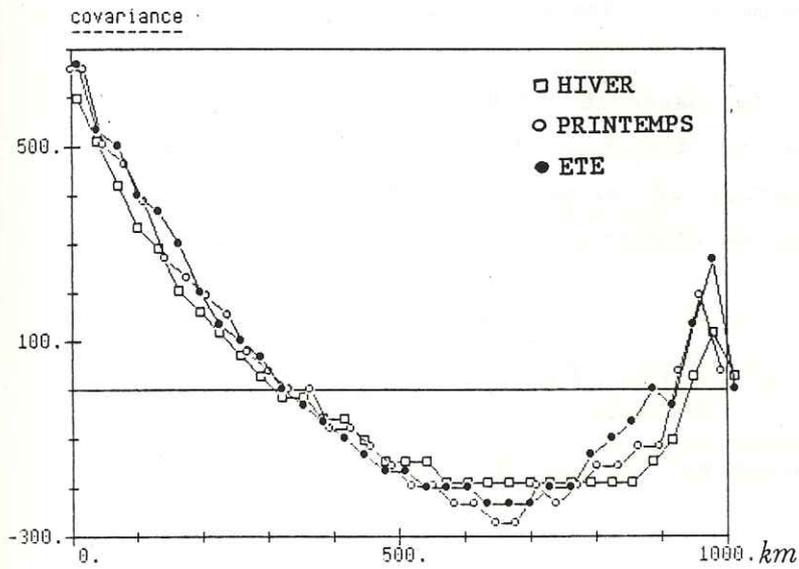


Fig II-27 :  
Evolution de la fonction  
de covariance avec la  
saison considérée.

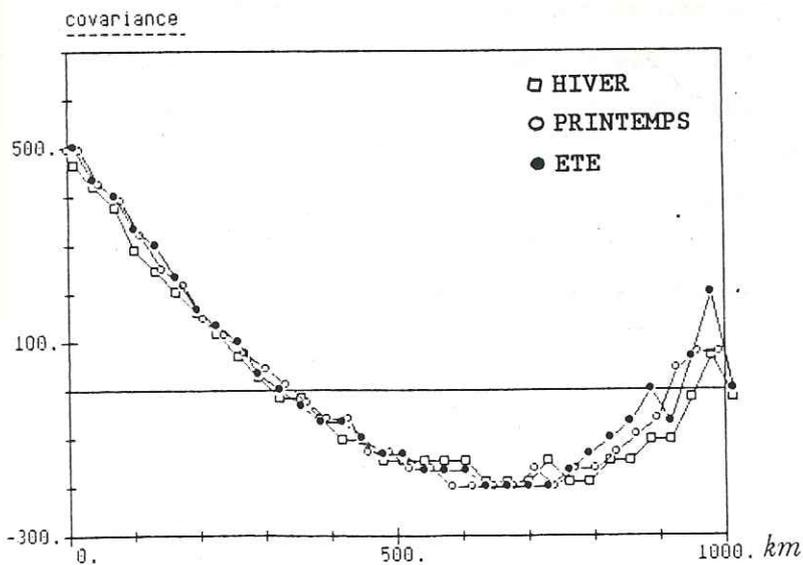
SOL

a



CACTUS

b



CROISE  
SOL/CACTUS

c

### III-ANALYSE EN COMPOSANTES PRINCIPALES

L'analyse en composantes principales (ACP) est une technique classique d'analyse de données, fondée sur un critère de variance expliquée. Sa principale utilisation est de réduire le volume initial de données, tout en conservant cependant la majeure partie de l'information initiale.

Lorsque les données analysées résultent d'un processus qui se développe dans le temps, mais aussi dans l'espace comme par exemple la durée d'insolation, l'ACP, dite alors de processus, nous renseigne également sur la structure spatiale. Cette connaissance nous permettra d'utiliser par la suite l'ACP de processus comme méthode d'interpolation.

Après avoir brièvement rappelé les bases de l'ACP classique, on développera l'ACP de processus ; ensuite, on discutera de la structure spatiale du rapport d'insolation mesuré par le sol ou par le satellite et de l'apport de ce dernier à la connaissance de l'hétérogénéité spatiale du processus.

#### III-1. Analyse en composantes principales classique

Cette méthode d'analyse classique est rappelée brièvement afin de préciser les notations.

On dispose d'un tableau T, de n observations de p variables centrées réduites. Chaque élément de cette matrice est noté  $y_{ki}$ , k<sup>ème</sup> observation de la i<sup>ème</sup> variable. La lettre y représente dans ce chapitre la variable centrée et réduite. On a vu précédemment que l'écart-type climatologique est peu variable, la réduction se ramène alors à une affinité. Cet ensemble de données peut être considéré comme un nuage de n points (les observations) repérés sur p axes (les variables). Les p variables étant souvent liées entre elles, on recherche une autre base, de variables non corrélées, qui engendre l'espace défini par ces p variables. Les

observations s'écrivent dans la nouvelle base :

$$y_{ki} = \sum_{l=1}^p v_{kl} \varphi_{il} \quad (1)$$

$v_{kl}$  est le score de l'observation  $k$  sur la  $l^{\text{ème}}$  composante.  
 $\varphi_{il}$  est le cosinus directeur de la  $i^{\text{ème}}$  variable sur la  $l^{\text{ème}}$  composante.

La nouvelle base étant orthonormée :

$$\vec{\varphi}_p \cdot \vec{\varphi}_m = \sum_{i=1}^p \varphi_{il} \varphi_{im} = \delta_{lm} \quad (2)$$

où  $\delta_{lm}$  est le symbole de kroneker.

Parmi les bases orthonormées ainsi définies, on choisit celle qui minimise l'erreur quadratique moyenne de reconstitution de tout élément du tableau  $T$ , à partir d'un nombre  $L$  de composantes (dites principales), inférieur à  $p$ , ceci quel que soit  $L$ . On note  $y_{Lki}$  la valeur ainsi reconstituée :

$$y_{Lki} = \sum_{l=1}^L v_{kl} \varphi_{il} \quad (3)$$

On cherche les cosinus directeurs  $\varphi_{il}$  qui vérifient la relation suivante:

$$\sum_k (y_{ki} - y_{Lki})^2 \text{ minimum}$$

On montre (LEBART et FENELON, 1973) qu'ils sont solution des  $p$  équations :

$$\sum_{j=1}^p r_{ij} \varphi_{j1} = \varphi_{i1} \lambda_1 \quad \text{pour } i = 1 \text{ à } p \quad (4)$$

Les vecteurs ainsi définis sont donc les vecteurs propres de la matrice R des corrélations intervariables  $r_{ij}$  ; les

$\lambda_1$  sont les valeurs propres associées. Nous en rappelons les principales propriétés :

- Les vecteurs propres sont non corrélés, leur variance est égale à la valeur propre associée.

$$\sum_{k=1}^n v_{k1} v_{km} = \delta_{lm} \lambda_1 \quad (5)$$

- La somme des variances des p variables est égale à p

$$\sum_{l=1}^p \lambda_l = p \quad (6)$$

- L'erreur quadratique s'exprime

$$\sum (y_{ki} - y_{Lki})^2 = \sum_{i=L+1}^p \lambda_i \quad (7)$$

- Le coefficient de corrélation total entre la l<sup>ème</sup> composante et la i<sup>ème</sup> variable se calcule par :

$$\varphi_{y_l z_i} = \sqrt{\lambda_1} \varphi_{1i}$$

L'ACP, fondée uniquement sur l'analyse des corrélations, ne prend pas en compte d'autres caractéristiques des variables. Dans notre cas, les variables sont des stations mesurant la durée d'insolation, c'est-à-dire un phénomène continu et peut être organisé spatialement. Il est évident que leur organisation géographique joue un rôle dans leur corrélation. Nous avons, par exemple, modélisé la fonction de corrélation en fonction de la seule distance interstations au chapitre précédent. Ce modèle décrit la structure d'un processus stationnaire d'ordre 2 ; il ne suffit pas à expliquer toutes les fluctuations expérimentales observées mais il en décrit bien la majeure partie. L'ACP de processus, comme l'ACP classique, résume un tableau avec des composantes principales, mais celles-ci prennent en compte la disposition géographique des variables. La puissance et l'intérêt de cette technique est de décrire la structure en une somme de composantes d'importance décroissante, et d'éviter ainsi sa modélisation qui nécessite des hypothèses de stationnarité. Les hétérogénéités ou même les ruptures locales climatologiques peuvent ainsi être prises en compte, dans la mesure où l'échantillonnage spatial les a saisies.

### III-2. ACP de processus

#### III-2-a. Décomposition orthogonale de processus

On montre (DAVENPORT, 1958, OBLED et CREUTIN, 1986) que tout processus aléatoire stationnaire  $Y(t)$  peut se décomposer en une série doublement orthogonale sur tout intervalle fini  $[ab]$  :

$$y(k)(t) = \sum_1 U_{k1} F_1(t) \quad (a \leq t \leq b) \quad (8)$$

Les deux conditions se notent :

. orthogonalité des  $F_l(t)$

$$\int_a^b F_m(t) F_l(t) dt = \delta_{ml} \quad (9)$$

. décorrélation des  $U_{kl}$

$$E(U_{kl} U_{km}) = \delta_{lm} \lambda_l \quad (10)$$

Ces équations sont à rapprocher de (1), (2), (5). Les fonctions  $F_l(t)$  dépendent ici de la taille et de la forme du domaine  $ab$ . On montre que ce sont les fonctions caractéristiques de l'équation intégrale (11) et qu'il en existe une infinité dénombrable.

$$\int_a^b R(t, t') F_l(t') dt' = F_l(t) \lambda_l \quad (11)$$

Par analogie entre la valeur ponctuelle  $F_l(t_i)$  et le cosinus directeur  $\psi_{li}$ , on appellera également la fonction  $F_l(t)$  "cosinus directeur" de la  $l^{\text{ème}}$  composante. Le coefficient  $U_{lk}$ , ou score de l'évènement sur la composante  $l$ , est donné par :

$$U_{kl} = \int_a^b z_k(t) F_l(t) dt \quad (12)$$

Pour toute décomposition limitée à  $L$  termes,

$$y_{Lk}(t) = \sum_{l=1}^L U_{kl} F_l(t)$$

l'erreur quadratique moyenne est minimum et égale à

$$E \left( \int_a^b (y_k(t) - y_{Lk}(t))^2 \right) = \sum_{l=L+1}^{\infty} \lambda_l \quad (13)$$

Cette décomposition est dite de Karheinen-Loeve pour le cas à 1 dimension. Pour le cas d'un phénomène se développant dans l'espace, on parle de décomposition en fonctions orthogonales empiriques (E.O.F. en anglais). On remplace alors la somme sur l'intervalle [ab] par une somme double sur le domaine D.

III-2-b. Lien avec l'analyse harmonique

Dans le cas d'un processus aléatoire périodique, l'équation (8) est la décomposition en série de Fourier du processus  $y_k(t)$ . L'énergie moyenne de chaque fréquence  $F_l(t)$  est égale à la valeur propre associée  $\lambda_l$ . Dans le cas général, la décomposition peut être interprétée comme une décomposition spectrale en puissance sur les fonctions  $F_l(t)$ , du processus aléatoire  $y_k(t)$  (DAVENPORT, 1958). Les valeurs propres  $\lambda_l$  tendent vers la densité spectrale en puissance du processus, et ceci d'autant mieux que le domaine est grand devant la portée. Par abus de langage, on a parfois utilisé par la suite le mot "harmonique" pour désigner le cosinus directeur  $F_l(t)$ .

### III-2-c. Rôle de l'échantillonnage dans l'estimation des cosinus directeurs

Le processus aléatoire continu  $z_k(t)$  n'est connu qu'à travers un nombre fini de stations  $p$ , réparties de façon irrégulière sur le domaine  $D$ . Il est alors nécessaire de rechercher des estimateurs calculés à partir des données discrètes disponibles, et qui convergent vers les fonctions  $R(t,t')$  et  $F_1(t)$ .

Ne disposant que de  $p$  stations, on ne pourra estimer que  $p$  fonctions propres et non une infinité : les petites longueurs d'ondes ne pourront donc être perçues par un réseau trop lâche.

### III-2-d. Approche générale

Elle a été proposée par DEVILLE (1974) et employée par OBLED et CREUTIN (1986) dans le cas d'un processus bidimensionnel.

On se donne une base de fonctions canoniques  $e_i(t)$  engendrant l'espace de dimension  $p$ , telles que le phénomène  $y_k(t)$  puisse être approché par la formule :

$$y_k^*(t) = \sum_{i=1}^p y_k(t_i) e_i(t) \quad (14)$$

La valeur estimée  $z_k^*(t)$  du phénomène en tout point  $t$  s'exprime comme une combinaison linéaire des  $p$  valeurs connues  $y_k(t_i)$ . On montre (DEVILLE, 1974) que la fonction d'autocorrélation  $R^{**}(t,t')$  des valeurs estimées converge vers celle du processus réel  $R(t,t')$ , quand  $p$  tend vers l'infini. Les cosinus directeurs  $F_1^*(t)$  et les scores  $U_{1k}^*$  associés tendent respectivement, vers  $F_1(t)$  et  $U_{1k}$

C'est donc la fonction d'autocorrélation  $R^{**}(t, t')$ , qui s'exprime en fonction des coefficients interstations,  $R(t_i, t_j)$  (éq.15), qui va être analysée ci-dessous.

$$R^{**}(t, t') = \sum_i \sum_j R(t_i, t_j) e_i(t) e_j(t) \quad (15)$$

Les fonctions caractéristiques recherchées  $F^*_1(t)$  sont solutions de l'intégrale

$$\int_D R^{**}(t, t') F^*_1(t') dt' = F^*_1(t) \lambda_1 \quad (16)$$

Elles peuvent, elles aussi, se décomposer sur la base des fonctions canoniques  $e_i(t)$  :

$$F^*_1(t) = \sum_{i=1}^p f_{i1} e_i(t) \quad (17)$$

L'équation intégrale (16) s'écrit alors :

$$\int_D \sum_i \sum_k R(t_i, t_k) e_i(t') e_k(t') \sum_j f_{1j} e_j(t') dt' = \sum_i f_{1i} e_i(t) \cdot \lambda_1$$

$$\Rightarrow \sum_i e_i(t) \sum_k \sum_j R_{ik} f_{1j} \int_D e_j(t') e_k(t') dt' = \sum_i e_i(t) f_{1i} \lambda_1$$

Ce qui donne le système de  $p$  équations :

$$\sum_j \left( \sum_k R_{ik} E_{jk} \right) f_{1j} = f_{1i} \quad i = 1 \text{ à } p \quad (18)$$

$$\text{avec } E_{jk} = \int_D e_j(t') e_k(t') dt' \quad (19)$$

Le coefficient  $E_{jk}$ , élément de la matrice  $[E]$ , est le coefficient caractérisant la ressemblance entre les fonctions  $e_i(t)$  et  $e_j(t)$ , sur le domaine  $D$ .

Les coefficients  $f_{1i}$  sont les cosinus directeurs de la matrice  $[R] \cdot [E]$  (éq.18).

On peut calculer (cf. éq.12) les scores  $U^*_{1k}$  associés aux  $F^*_{1}(t)$  :

$$\begin{aligned} U^*_{1k} &= \int_D y^*_k(t) F^*_{1}(t) dt \\ &= \sum_i \sum_j y_k(t_i) f_{1j} E_{ij} \end{aligned} \quad (20)$$

De façon concrète, l'estimation des cosinus directeurs  $F^*_{1}(t)$  et des scores  $U^*_{1k}$  consiste tout d'abord à choisir une base  $e_i(t)$  ( $i=1$  à  $p$ ) vérifiant (14), puis à diagonaliser la matrice  $[R \ E]$  pour déterminer ses cosinus directeurs  $f_{1i}$  ; enfin, les cosinus directeurs (éq.17) et les scores (éq.20) sont reconstitués à partir des cosinus directeurs ponctuels  $f_{1i}$  et des fonctions de base  $e_i(t)$  et éventuellement des mesures  $y_k(t_i)$ .

La matrice de corrélation interstation  $R$  que l'on diagonalise ici est pondérée par la matrice  $E$ . C'est par elle qu'on prend en compte la répartition géographique du réseau. Sa valeur dépend de la forme des fonctions  $e_i(t)$  employées.

Considérons des fonctions simples, valant un dans la surface d'influence de  $t_i$  et zéro partout ailleurs (les surfaces d'influence sont par exemple celles des polygones de Thiessen). Les  $E_{ij}$  (éq.19) pondèrent alors les résultats par la surface associée à chaque station, ce qui compense les effets de l'irrégularité du réseau. L'équation intégrale (15) est discrétisée :

$$\sum_i R(t_i, t_j) F_1(t_j) \Delta t_i = F_1(t_j) \Delta t_j \quad j = 1 \text{ à } p$$

où  $\Delta t_i$  représente l'élément de surface associée au point  $t_i$ .

Dans le cas d'un réseau régulier, toutes les surfaces  $\Delta t_i$  sont égales. Les résultats coïncident avec ceux d'une ACP classique (éq.4). D'autres formes, plus sophistiquées, ont été proposées pour modéliser les  $e_i(t)$  (cf. OBLED et CREUTIN, 1986) qui prennent en compte la décroissance de l'influence d'une station.

On dispose ainsi avec l'ACP de processus d'un outil qui permet d'atteindre les cosinus directeurs  $F_1(t)$  du phénomène réel continu  $Z(t)$ , connu de façon discrète par un réseau de disposition irrégulière.

### III-2-e. Conclusion

L'ACP classique fait l'analyse factorielle d'un tableau de données, sans tenir compte explicitement de l'organisation géographique éventuelle des variables.

Par contre, cette organisation est prise en compte dans l'ACP de processus, à travers les coefficients d'une matrice  $E$ . Celle-ci pondère la matrice des corrélations dont on fait ensuite l'ACP. A partir des cosinus directeurs ainsi calculés, on peut reconstituer les premiers cosinus directeurs et les scores associés du phénomène à concurrence du nombre de stations de mesure disponibles. Les cosinus directeurs  $F_1(t)$  dépendent à la fois du processus  $Y(t)$  et du domaine  $D$ . Ils peuvent être interprétés comme les "harmoniques" génératrices du processus  $Y(t)$ , dont la puissance est donnée par les valeurs propres  $\lambda_1$ . Si on pondère les cosinus directeurs par leur puissance, on obtient la valeur du coefficient de corrélation entre la composante 1 et le point considéré. En ce sens, ils caractérisent la structure du phénomène.

### III-3. Analyse des cosinus directeurs du rapport d'insolation

Nous avons vu, au chapitre précédent, que la structure spatiale du rapport d'insolation centré deux fois pouvait être expliquée pour une large part, par la seule interdistance entre les points considérés : c'est l'hypothèse de stationnarité d'ordre 2. Une certaine hétérogénéité est cependant présente et entraîne une dispersion du corrélogramme ; essayons de la mettre en évidence, et de la distinguer du phénomène.

Nous disposons d'un réseau de 84 stations mesurant l'insolation sur la France. Ces stations sont relativement bien réparties : environ une station par département. Une ACP simple a donc tout d'abord été effectuée.

#### III-3-a. Analyse d'un processus homogène pris comme référence

Par définition, la liaison statistique entre deux points mesurant un processus homogène ne dépend pas de leur point d'appui, mais seulement de leur interdistance. Un tel processus ne possède aucune tendance liée à la géographie. Pour un modèle  $r(h)$  donné, les cosinus directeurs sont parfaitement déterminés par la seule forme du domaine  $D$  (éq. 11).

Fig II-28 :

Cosinus directeurs obtenus à partir des matrices de corrélation  
 (a) du sol (b) du satellite (c) du modèle homogène

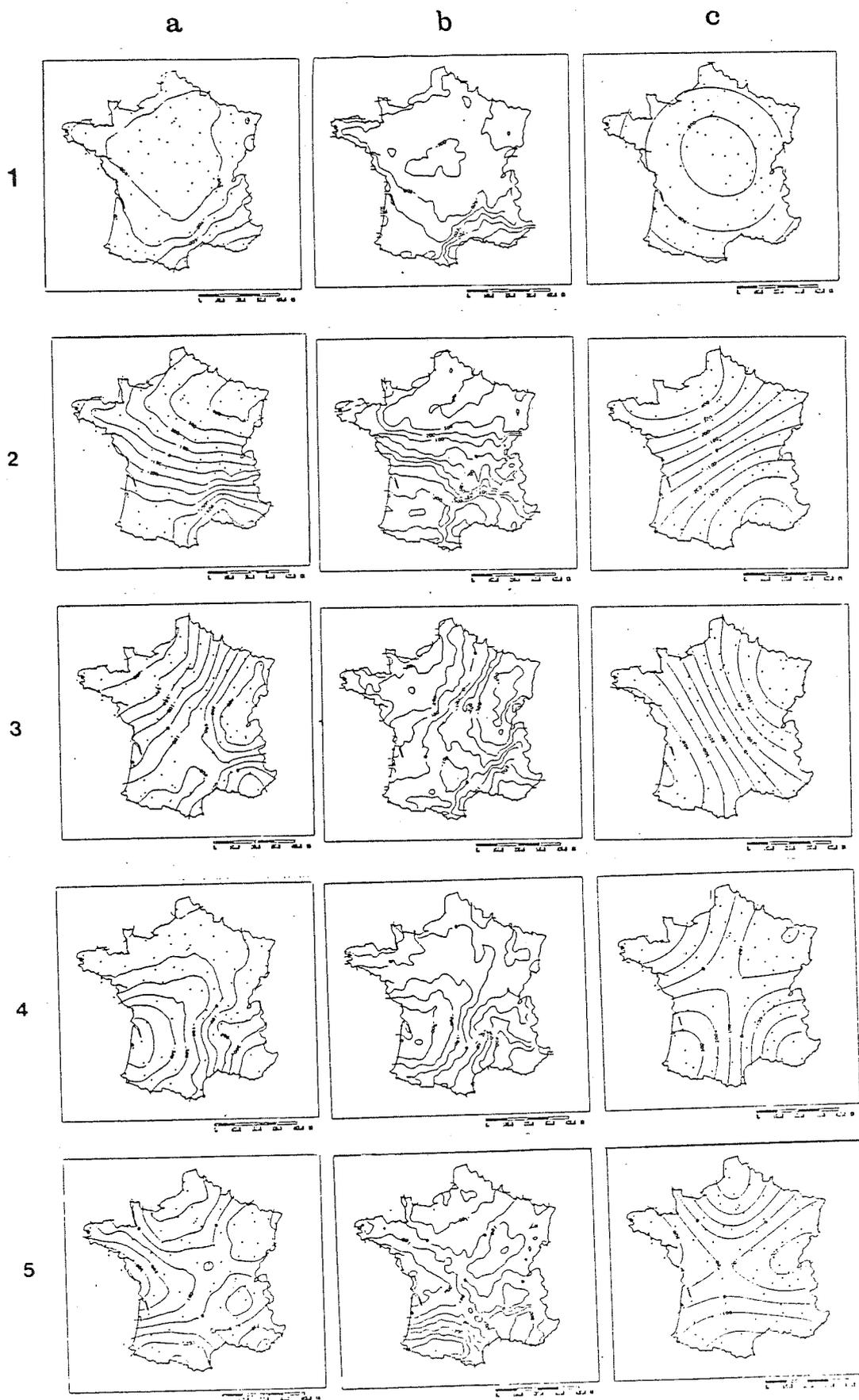
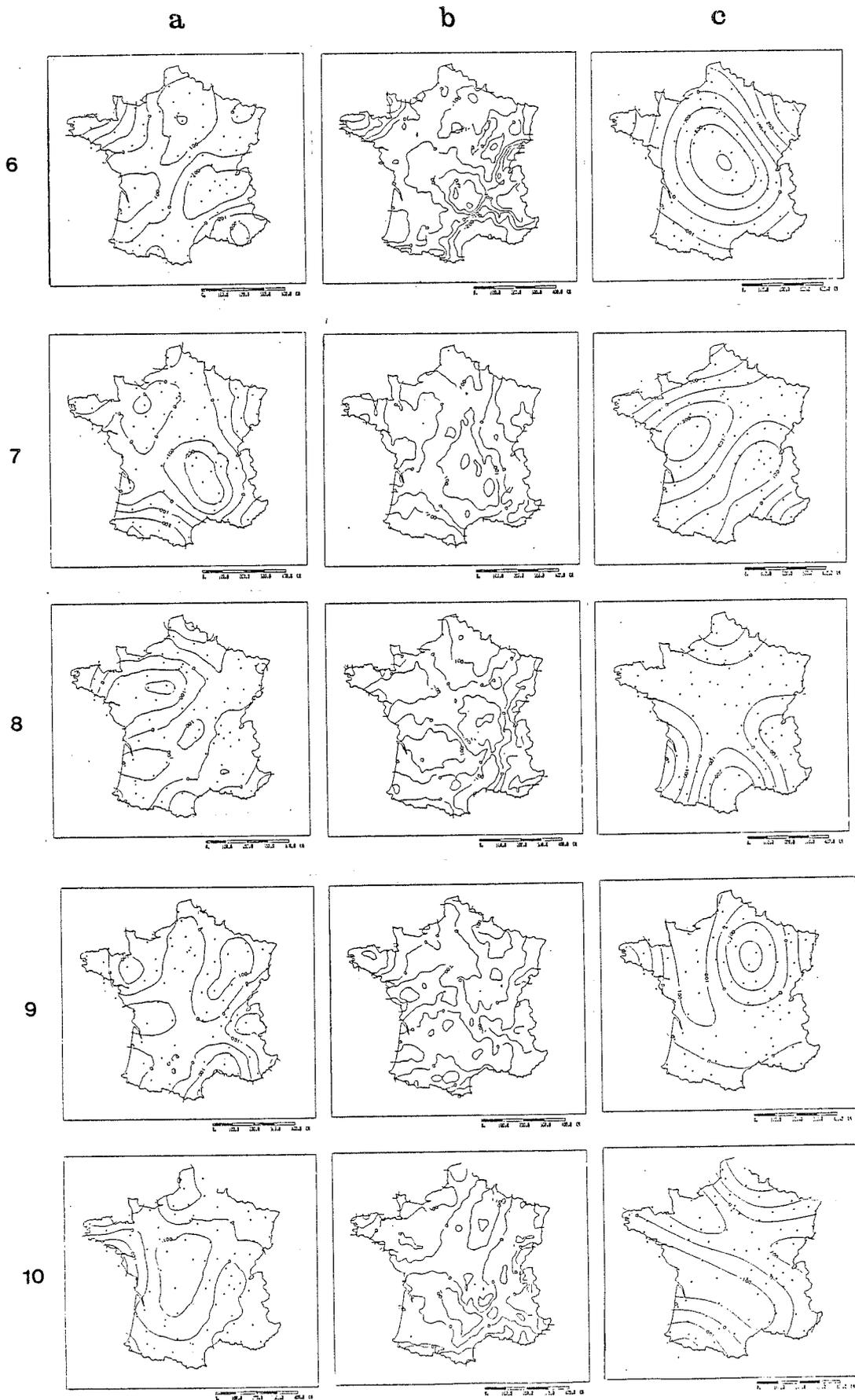


Fig II-28 (suite):



Nous avons donc utilisé le modèle de corrélation ajusté sur les données sol centrées et réduites de façon climatologique pour calculer une matrice de corrélation parfaitement homogène mais proche de la réalité du phénomène. L'ACP réalisée sur cette matrice modélisée nous a ensuite servi de référence pour juger de l'influence de la forme de la France, et de la géométrie du réseau sur les cosinus directeurs du phénomène. Les dix premiers d'entre eux (Fig.(II)-28 colonne c), de valeur propre donnée (Fig.(II)-32-a), sont détaillés ci-dessous.

- Première composante

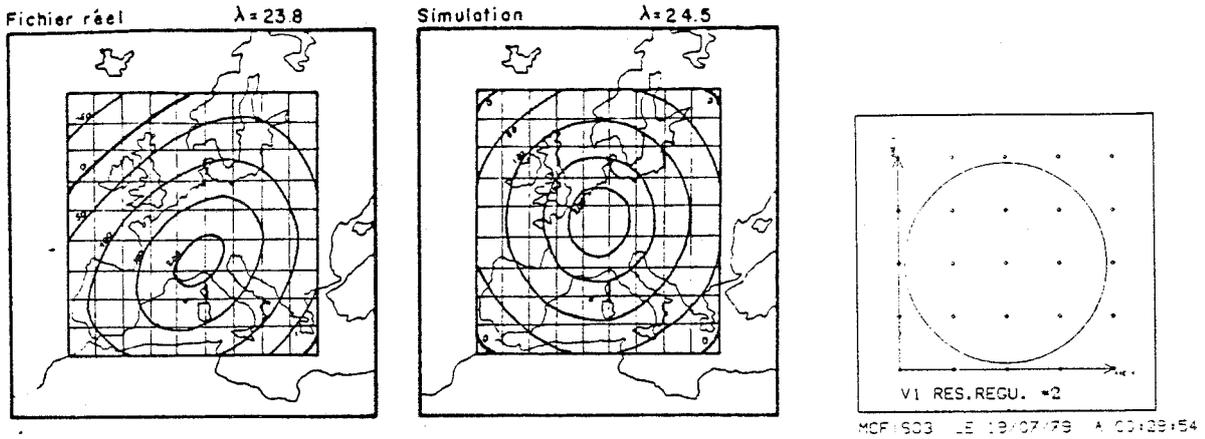
Elle possède une très forte puissance  $\lambda = 44$  soit plus de la moitié (52%) de la variance totale du phénomène. Cette composante est très proche de la moyenne spatiale du champ et se traduit par des cosinus directeurs pratiquement constants sur le domaine. Le fait de travailler sur un domaine limité induit la légère forme en dôme observée : les bords du domaine sont moins corrélés que le centre, à la moyenne spatiale. Cette interprétation se trouve confirmée si l'on fait maintenant l'ACP de la matrice  $R_x$ , modélisée par la fonction de corrélation ajustée sur les valeurs de l'insolation centrées et réduites, puis centrées spatialement. Cette première composante disparaît alors et la première composante obtenue correspond à la deuxième composante de la figure 1.

La première composante, de très forte puissance, traduit simplement un effet de taille (c'est le rôle du "bouton de réglage" vu précédemment). Le phénomène se décompose en premier lieu comme une valeur moyenne, puis des fluctuations autour de celle-ci, ce qui justifie le centrage spatial effectué précédemment. La structure de ces résidus est analysée dans les composantes suivantes.

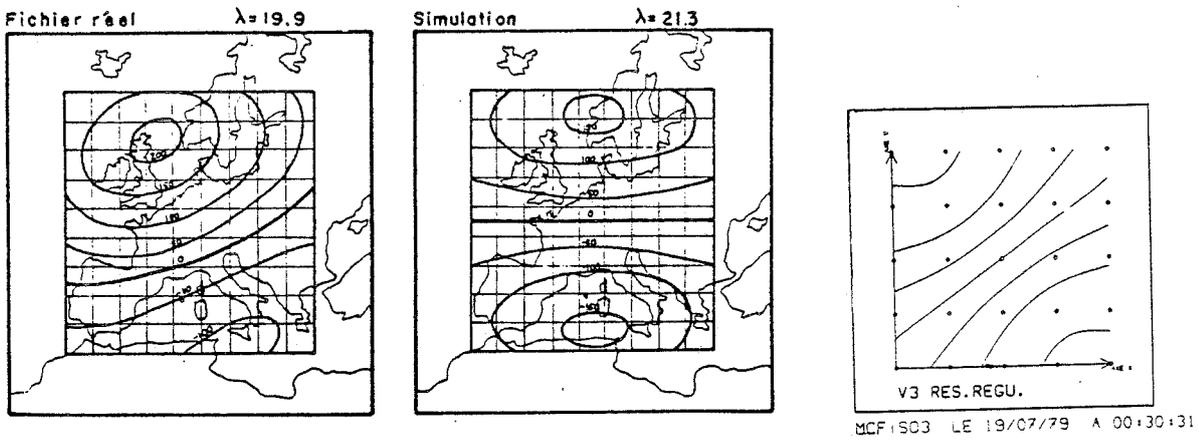
- Deuxième et troisième composantes

Elles sont liées de façon évidente par la similitude de leur forme (seule une rotation de  $90^\circ$  les sépare). Elles correspondent à une valeur propre multiple d'ordre 2 ( $\lambda = 9$ ). Elles traduisent une opposition de

PREMIERE COMPOSANTE PRINCIPALE CSRI



2<sup>e</sup> COMPOSANTE PRINCIPALE CSR2



3<sup>e</sup> COMPOSANTE PRINCIPALE CSR3

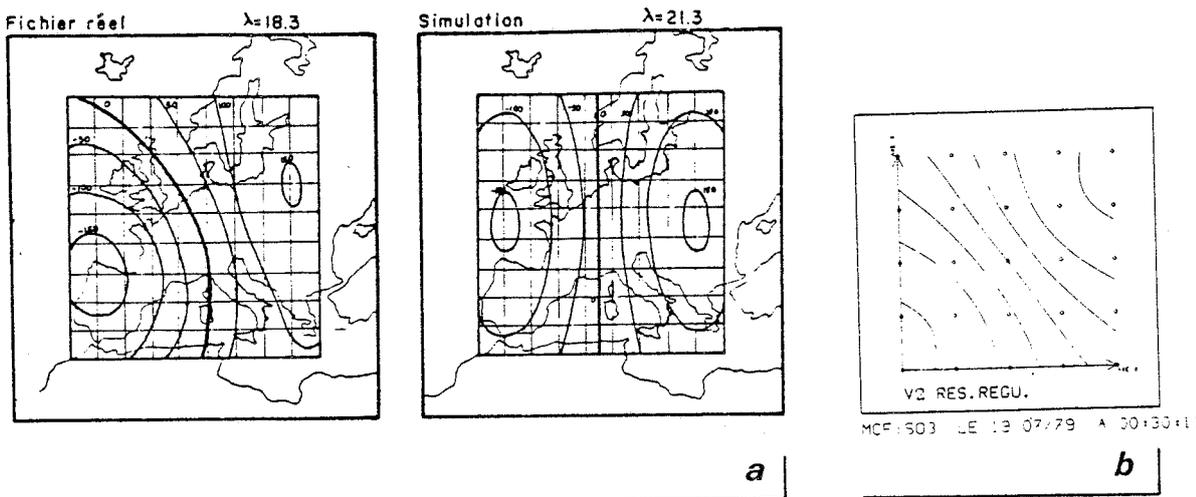


Fig II-29 : Cosinus directeurs (a) des altitudes géopotentielle (d'après Kueny,1977); (b) d'un processus markovien bidimensionnel (d'après Obléd,1979) .Des phénomènes très divers ont des cosinus directeurs semblables.

comportement entre le NW et le SE du domaine, ainsi qu'entre le SW et le NE. Il faut ici se rappeler que ces fonctions ne dépendent que du domaine  $D$  et de  $R_z(h)$  (qui dépend de l'interdistance et non de la position absolue des deux points de mesure). Il faut se garder d'interpréter de façon physique ces directions qui ne sont déterminées que par la forme du domaine (la France). Si on fait subir une rotation au domaine, les cosinus tourneront de même et ne seront plus alors NW/SE.

Des observations identiques sur la forme des cosinus ont été faites par J.L. KUENY (1977), à propos du champ des altitudes géopotentielle 500 mbar désaisonnalisées, sur l'Europe (Fig.(II)-29-a). Ainsi que par C. OBLED (1979), pour un processus markovien bidimensionnel (Fig.(II)-29-b). La similitude des cosinus directeurs pour des phénomènes aussi divers et des domaines d'étendue différentes est troublante au premier abord. Elle résulte en fait de la forme du domaine, proche d'un carré dans tous les cas, et de la forte cohérence spatiale des données pour chacun de ces cas. La forme du domaine, conduit à des formes de cosinus plus "élaborées" dans le cas de la France, dont le contour est, lui aussi, plus compliqué. Les formes des cosinus obtenus sont cependant comparables jusqu'à la sixième composante, après elles divergent. Pour illustrer maintenant la sensibilité des cosinus à la portée de la fonction de corrélation, prenons un cas extrême.

Considérons un tableau de données dont on a mélangé, de façon aléatoire, les coordonnées des stations. La matrice de corrélation calculée à partir de ces données sera identique à celle des données initiales, les valeurs, et vecteurs propres également. Toute l'information spatiale réside dans le corrélogramme, qui sera parfaitement incohérent puisque la distance n'a plus de sens par rapport à la corrélation. Par suite, les fonctions  $F_1(t)$  sont elles aussi incohérentes: la valeur des cosinus directeurs aux points de mesure est identique, mais ceux-ci sont répartis de façon aléatoire sur le domaine ; ils n'ont aucune cohérence spatiale. On ne retrouve pas les formes caractéristiques des cosinus directeurs décrites ci-dessus. Les cosinus directeurs d'un processus sans cohérence spatiale sont eux aussi chaotiques. Mais revenons à la description des cosinus directeurs.

- Quatrième, cinquième et sixième composantes

La valeur propre multiple associée,  $\lambda = 2.5$ , est plus faible. Les cosinus directeurs  $F_1(t)$  ont une portée plus courte. La France est divisée en quatre parties, opposées en signe deux à deux, ou encore, pour la sixième composante, le bord est opposé au centre. Cette dernière composante n'est pas comparable à la première pour laquelle on ne notait pas d'opposition entre le centre et les bords mais seulement une légère décroissance de la corrélation sur les bords.

- Septième composante et suivantes

Les valeurs propres sont de plus en plus faibles et les longueurs d'onde associées également. Ceci donne un aspect de plus en plus chaotique aux cartes des cosinus directeurs  $F_1(t)$ . Les portées associées diminuent de même.

- Conclusion

Nous avons fait l'analyse en composantes principales d'un phénomène parfaitement homogène : modélisé à partir de la fonction de corrélation du rapport d'insolation. Les cosinus directeurs ainsi calculés nous serviront de référence par la suite, afin de distinguer l'effet dû à l'homogénéité et à la forme du domaine, de celui dû à l'hétérogénéité ou à la physique du phénomène qui est celui qui nous intéresse.

Les formes observées pour les différents cosinus directeurs sont générales et ont déjà été observées par d'autres auteurs, pour des phénomènes différents. La première composante est proche de la moyenne spatiale du champ, les suivantes caractérisent la structure spatiale des résidus à cette moyenne. On peut comparer les cosinus directeurs  $F_1(t)$  à des harmoniques génératrices du phénomène de longueur d'onde décroissante avec la composante  $l$ . Dans notre cas, comme souvent en météorologie, la densité spectrale de puissance est monotone décroissante : ce sont les harmoniques de plus basse fréquence qui transportent le plus de puissance.

### III-3-b. Analyse en composantes principales du rapport d'insolation

Nous nous intéressons maintenant à la matrice de corrélation calculée à partir des données réelle centrées réduites  $y_{sk}(t)$  du sol. Les cosinus de cette matrice sont cartographiés (Fig.(II)-28- colonne a). Ils ont un aspect moins lisse que ceux de la matrice modélisée, qui ont été reportés (Fig.(II)-28 - colonne c). Ces différences sont dues à l'hétérogénéité (spatiale) du rapport d'insolation, ainsi qu'à l'échantillonnage. Si le sol et le satellite ont même structure, cette hétérogénéité doit se retrouver de façon identique dans les fonctions  $F_{1S}(t)$  et  $F_{1T}(t)$ . Afin d'éliminer l'influence de l'échantillonnage spatial, on a comparé les fonctions estimées par un réseau identique pour le sol et le satellite, soit 84 stations.

Les principales caractéristiques des cosinus directeurs du sol et du satellite, pour les trois périodes considérées sont résumées dans l'annexe 6(p250). Plutôt que de comparer visuellement, et donc subjectivement, les cosinus directeurs obtenus, nous avons calculé les coefficients de corrélation spatiaux entre les cosinus du sol et du satellite, aux 84 stations. Ils traduisent de façon synthétique leur cofluctuation spatiale.

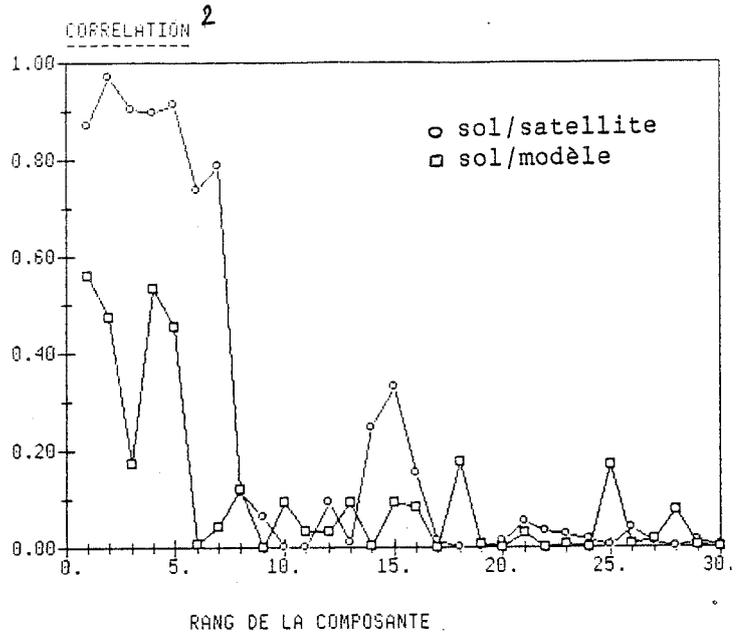
La cofluctuation des premiers cosinus directeurs, en terme de variance expliquée ( $r^2$ ), est excellente pour toutes les périodes (Fig.(II)-30 abc). Il intervient ensuite une chute brutale et les corrélations ne sont plus significatives. On peut donc interpréter les premières composantes comme la structure commune aux deux types de mesure. Les suivantes sont chaotiques et de faible puissance, elles n'expliquent que des microfluctuations locales ou les erreurs perçues apparemment de manière différente par le sol et le satellite, ou des inversions de fonctions propres (qui sont décorréliées à cause de l'orthogonalité déterministe). Cette dernière hypothèse sera détectée si on prend soin de comparer les cartes correspondantes (Fig II-29).

Dans le cas du printemps, les cosinus directeurs du sol ont également été corrélés à ceux du modèle homogène précédent (Fig.(II)-30-a). Le pourcentage de variance expliqué par le modèle est nettement plus faible

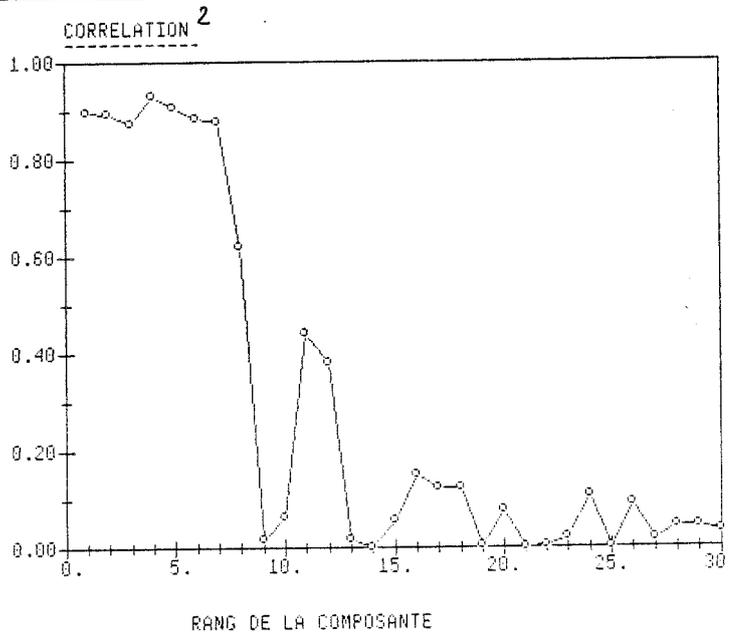
Fig II-30 :

Diminution de la  
correlation des cosinus  
directeurs du sol avec  
ceux du satellite en  
fonction du rang de la  
composante.

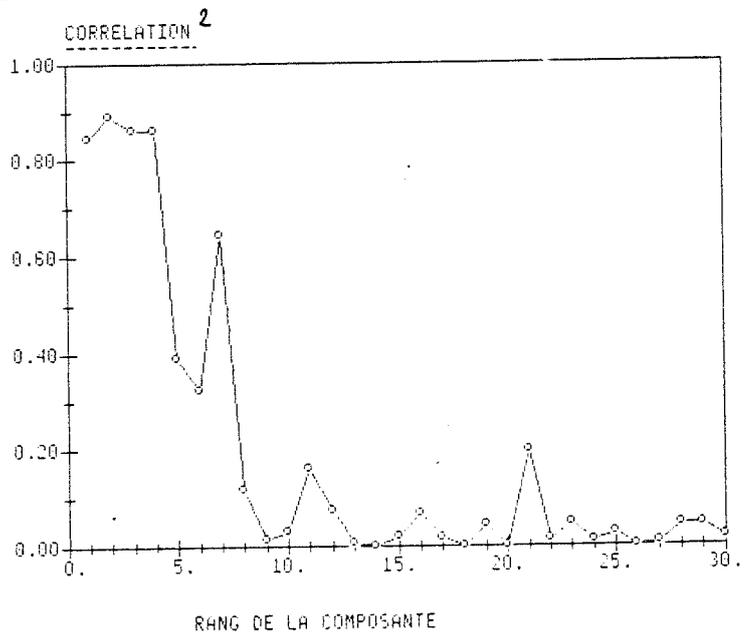
PRINTEMPS



ETE



HIVER



que celui expliqué par le satellite. Pour le premier cosinus directeur, par exemple, le modèle homogène explique 58 % de la variance du sol mais le satellite 88 %. La faible performance des deuxième et surtout troisième cosinus directeurs s'explique par une inversion des deux composantes associées à une valeur propre multiple (nous y reviendrons plus loin). De plus, la chute de la corrélation intervient plus tôt : la sixième composante du modèle est déjà totalement décorrelée de celle du sol alors que c'est la huitième pour le satellite. C'est l'hétérogénéité du phénomène sol qui explique la différence de ses cosinus avec ceux du modèle homogène. La corrélation est meilleure avec les cosinus du satellite, c'est donc que leur perception de l'hétérogénéité est, dans une certaine mesure, semblable.

Les sept premiers cosinus directeurs du sol et du satellite confluent de façon remarquable. Ils expliquent à eux seuls environ 85 % de la variance totale (voir tableau II-1), ce qui est déjà très correct. Une meilleure appréhension de la confluctuation nécessiterait également le calcul de la corrélation croisée entre fonctions propres. Nous nous sommes contentés de comparer les cartes pour détecter les inversions.

### III-3-c. Apport de la résolution du satellite

Nous avons vérifié sur 84 stations que la structure spatiale du rapport d'insolation est perçue de façon remarquablement semblable par le sol et la télédétection. La haute résolution spatiale du satellite va maintenant être utilisée afin d'affiner l'estimation des cosinus directeurs  $F_1(t)$  du processus générateur, à partir des données expérimentales. Plus le nombre de points d'échantillonnage est important, meilleure est l'estimation des cosinus directeurs du processus. De plus, la régularité de la grille autorise alors une parfaite équivalence entre l'ACP classique et l'ACP de processus.

L'ACP d'une matrice de corrélation aussi volumineuse (1222x1222) est possible, car le nombre d'évènements étant de l'ordre de 70, il suffit de se ramener à une matrice 70x70 (cf LEBART et FENELON, 1973 p202-204), ce qui est tout-à-fait raisonnable.

Pour chaque composante (de la première à la huitième), on a tracé les trois cartes de cosinus directeurs estimés respectivement à partir : des 84 stations sol (a), du satellite complet (b), et de la matrice de corrélation modélisée (c) (Fig.(II)-28). Les cosinus directeurs obtenus à l'aide du satellite présentent une variabilité plus forte, qui recouvre souvent une réalité physique identifiable. On note, par exemple, pour le premier cosinus directeur, le comportement singulier de la vallée du Rhône ou encore de la côte bretonne par rapport à l'intérieur de la Bretagne : détails que le seul réseau sol ne parvient pas à restituer. Comme nous l'avons vu les deuxième et troisième cosinus directeurs sont associés à la même valeur propre dégénérée ( $\lambda \approx 10$ ). L'ordre de ces deux composantes a été inversé pour les deux matrices expérimentales par rapport à la matrice modélisée, ce qui explique leur faible cofluctuation (Fig.(II)-30-a). De plus, le gradient du champ des cosinus a subi une légère rotation l'amenant à opposer Nord/Sud plutôt que NE/SW. Cette rotation et cette inversion des composantes n'est pas un fait de la géométrie car il apparaîtrait alors dans le cosinus modélisé mais bien du processus lui-même. L'opposition Nord/Sud est plus importante que NW/SE pour le rapport d'insolation. On finira en remarquant le comportement singulier de la bordure méditerranéenne, sur la deuxième composante, qui est un élément important dans la description du rapport d'insolation.

On pourrait ainsi continuer à détailler de façon approfondie les cosinus directeurs des premières composantes, pour la région et l'application qui nous intéresse. Ces comparaisons ont un sens jusqu'au rang sept, pour la deuxième période considérée ici (Fig.(II)-28). Les cosinus directeurs suivants, outre leur faible puissance qui les rend peu explicatifs, ne cofluctuent pas pour les deux types de mesure.

Les cartes de cosinus directeurs ont également été tracées (Fig.(II)-31) pour les deux autres périodes de l'année. Leurs grandes lignes sont très comparables, les cosinus directeurs relatifs à l'hiver apparaissant toutefois plus "chahutés".

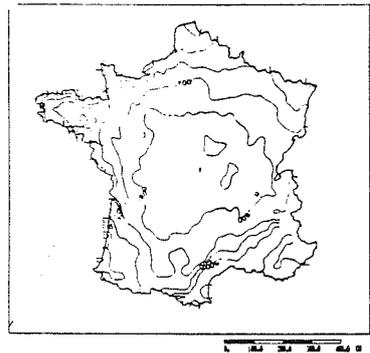
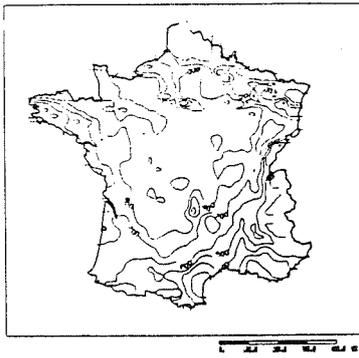
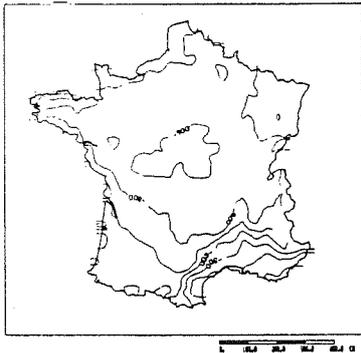
Fig II-31 : cosinus directeurs du satellite

PRINTEMPS

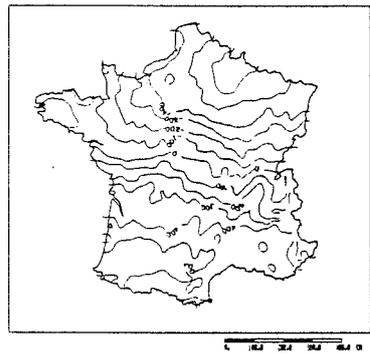
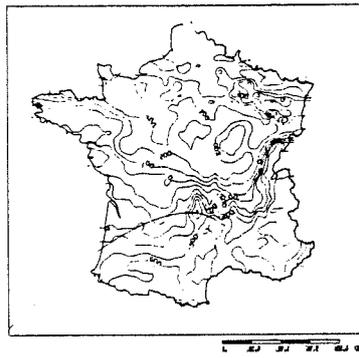
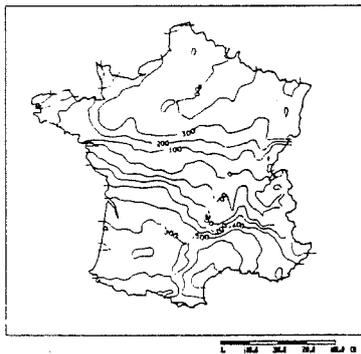
ETE

HIVER

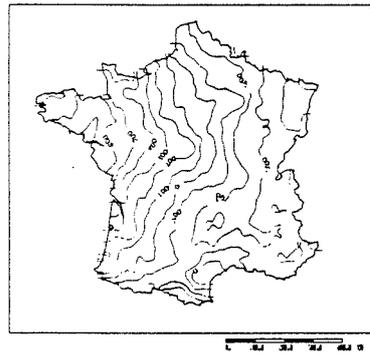
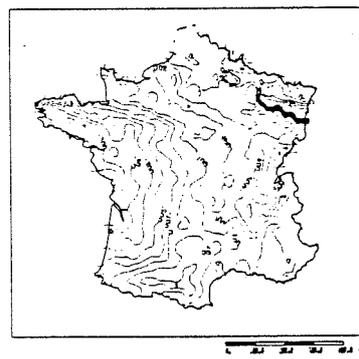
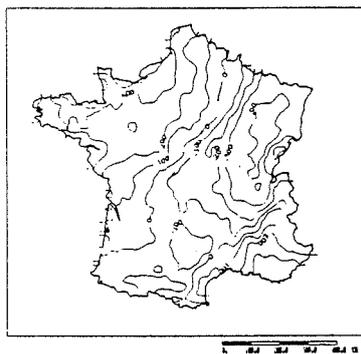
1



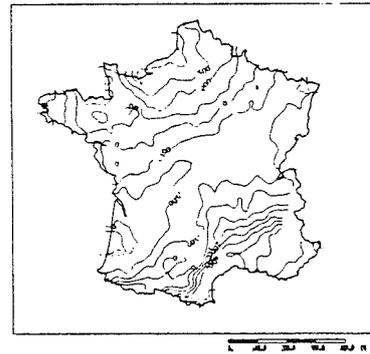
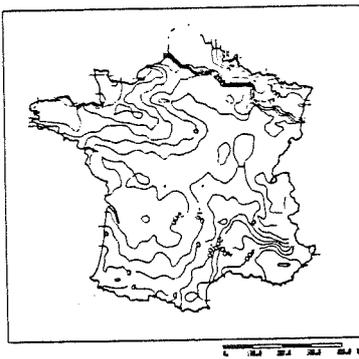
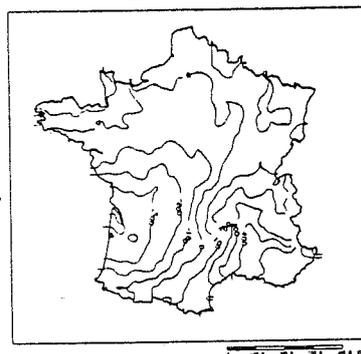
2



3



4



5

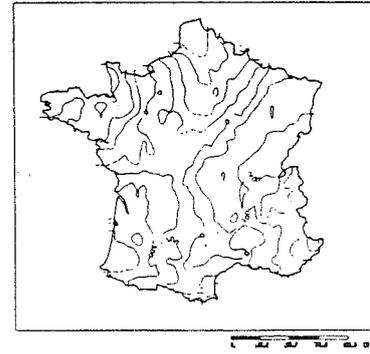
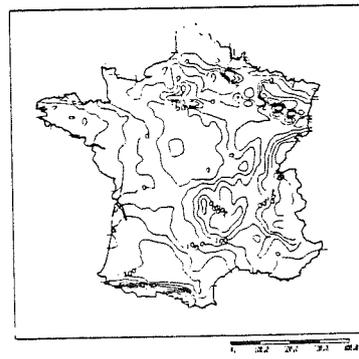
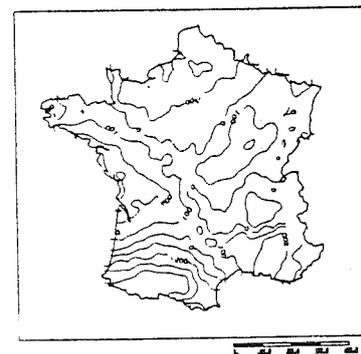


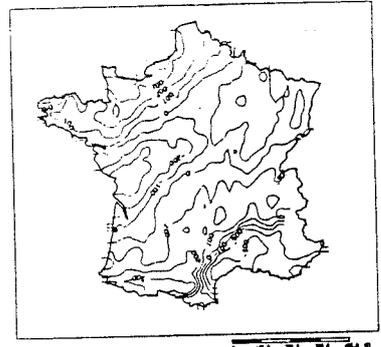
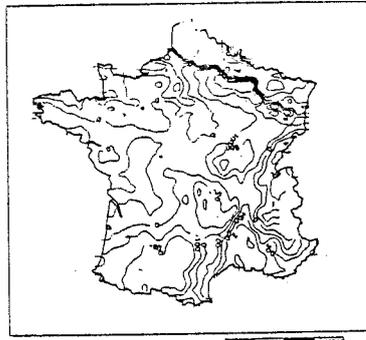
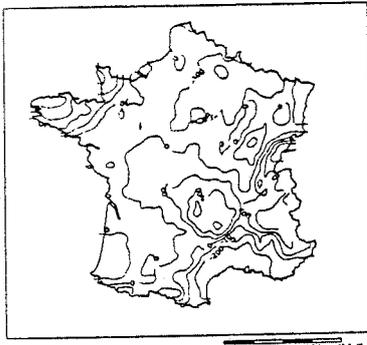
Fig II-31 (suite) :

PRINTEMPS

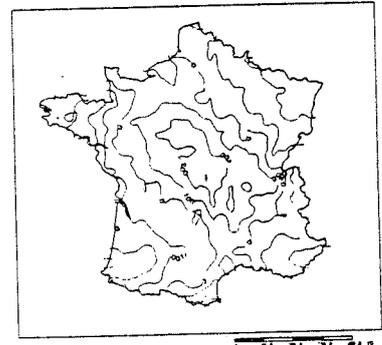
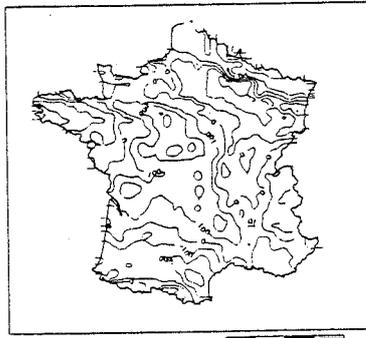
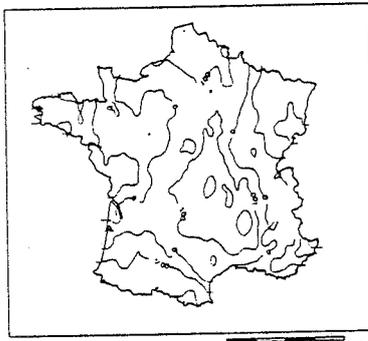
ETE

HIVER

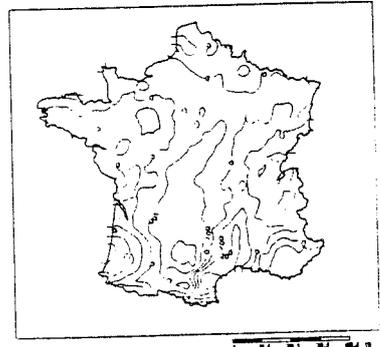
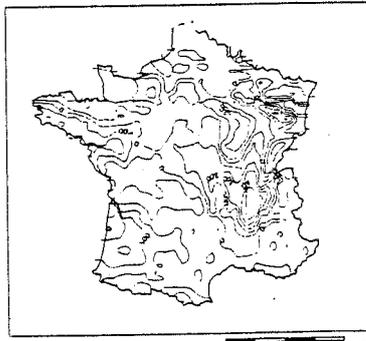
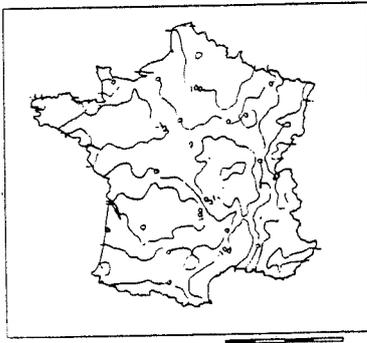
6



7



8



### III-3-d. Détermination du nombre de composantes significatives

La fonction de corrélation associée au processus aléatoire se décompose sur les cosinus directeurs :

$$R(t, t') = \sum_{l=1}^P \lambda_l F_l(t) F_l(t')$$

Cette décomposition ne nécessite pas d'hypothèse d'homogénéité sur le phénomène. Si on se contente des L premières composantes, on lisse la fonction de corrélation. Ce lissage correspond à une perte de variance que l'on peut quantifier, ce qui n'est pas le cas lorsqu'on ajuste un modèle homogène sur le nuage de points  $R(t, t') = f(d)$ .

Le choix du nombre de composantes "significatives" L est délicat. Un certain nombre de règles empiriques est couramment utilisé. Une description détaillée en est faite par C. OBLED, 1979.

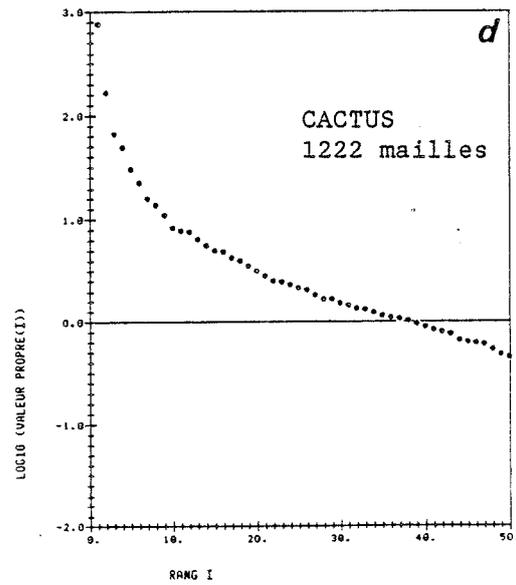
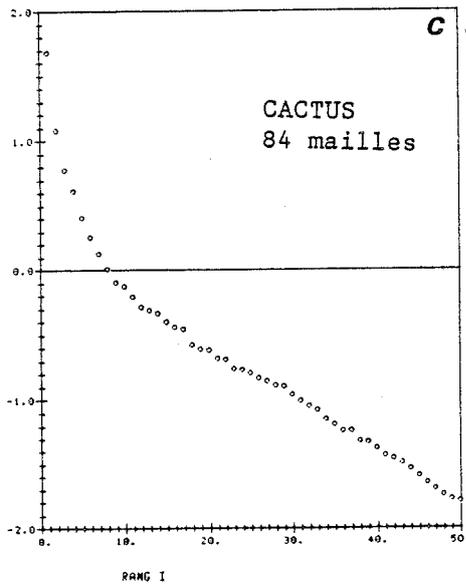
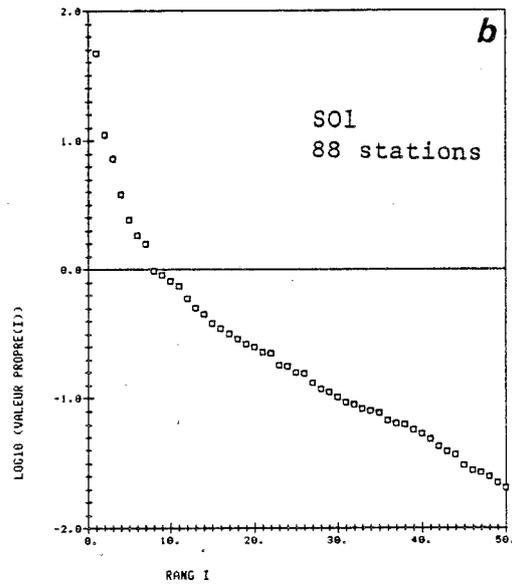
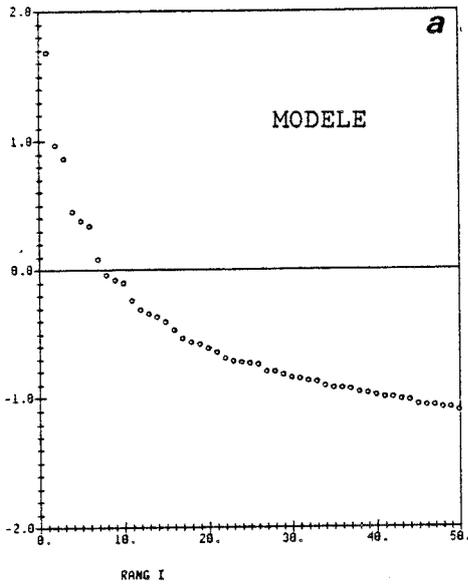
#### . Règle de Kaiser

C'est la plus simple. Elle consiste à retenir comme facteurs significatifs ceux dont les valeurs propres associées sont supérieures ou égales à un. Ce seuil de un correspond au souci de ne pas ignorer une variable qui serait totalement indépendante des autres.

#### .Méthode de LEV (Log Eigen Value)

Cette méthode a été proposée par les météorologues anglais, dans le cadre de l'ACP de champs spatiaux (JM.CRADDOCK, 1973). On porte sur un graphique le logarithme des valeurs propres expérimentales en fonction du rang i de la composante. Ce diagramme présente en général une partie rectiligne correspondant à une valeur propre multiple de faible valeur. Seules les valeurs propres précédentes sont conservées.

Fig II-32 : Evolution du logarithme de la valeur propre  $\lambda_i$ , en fonction de son rang  $i$



Ces constatations empiriques peuvent s'expliquer ainsi : pour un phénomène régionalisé, les premiers facteurs soulignent une structure d'ensemble forte, puis chaque variable introduit une variabilité propre, du même ordre de grandeur. La valeur propre associée est multiple. Une valeur propre multiple du processus réel présente une petite variabilité du fait de l'échantillonnage. Ces valeurs propres ne sont donc pas strictement égales, mais alignées.

Si on observe plus particulièrement le graphe  $\log_{10}(\lambda_i) = f(i)$  (Fig.(II)-32-a), obtenu pour les valeurs théoriques, on remarque que les valeurs propres sont groupées par paquets : la première, puis la deuxième et la troisième, puis (4,5,6), puis (7,8,9,10), puis enfin la onzième et les suivantes. Ces dernières sont celles que l'on supprime par la méthode de LEV. Ces différents groupes s'interprètent chacun comme une portée caractéristique, génératrice du processus, associée à une puissance. Plus celle-ci est courte, plus la multiplicité de la valeur propre correspondante est d'ordre élevé .. Il faut donc prendre garde à ne pas séparer des composantes correspondant à la même valeur propre, ce qui n'aurait pas de sens physique. Dans notre cas (Fig.(II)-32-a), on gardera une, trois, six ou dix composantes pour résumer la matrice modélisée, de préférence dix.

On a reproduit (Fig.(II)-32) les courbes  $\log_{10}(\lambda_i) = f(i)$  pour les matrices (b) du sol, (c) du satellite calculé avec 84 mailles, (d) du satellite calculé avec toutes les mailles, soit 1222. La séparation des valeurs propres multiples n'est pas aussi évidente que dans le cas des valeurs modélisées.. Une cassure est tout de même assez apparente à la septième (huitième) composante, pour le sol (le satellite). Ce sera le nombre de composantes à conserver. Le nombre de composantes significatives ne varie pas avec le nombre de mailles considérées dans la matrice de corrélation.

. Portée des cosinus directeurs

Par analogie avec l'analyse harmonique, nous avons parfois appelé les cosinus directeurs "harmonique". Le processus n'étant pas périodique, ceux-ci ne le sont pas non plus, mais ils manifestent dans l'espace une certaine structure. Cette structure est caractérisée par la portée du cosinus directeur considéré. Seuls les cosinus directeurs qui manifestent une structure sont considérés comme significatifs au sens de l'analyse de la structure du phénomène puisque le phénomène  $y_k(t)$  s'exprime encore comme leur combinaison linéaire :

$$y_{Lk}(t) = \sum_{l=1}^L U_{lk} F_l(t)$$

Chaque cosinus directeur peut être considéré comme la réalisation unique d'une variable régionalisée. Le variogramme construit pour chacun d'entre eux nous donne les caractéristiques de sa structure spatiale. Les variogrammes des cosinus directeurs du sol en hiver (Fig.(II)-34) ont été tracés pour différents rangs de composantes, allant de la première à la vingt-quatrième. La portée, supérieure à 500 km pour la première, passe à 350 km pour la huitième et à 100 km pour la seizième. La portée diminue avec le rang de la composante (voir le tableau II-1 pour des résultats complets). Cette diminution de la portée s'accompagne d'une diminution du palier, c'est-à-dire de la variance du champ ; mis à part pour le premier cosinus directeur qui a un comportement singulier (on a vu qu'il était peu dépendant du lieu). L'interpolation du champ des cosinus directeurs de rang élevé n'a pas de sens.

Une chute importante de la portée a lieu entre la cinquième et la sixième composante (on passe de 500 à environ 300 km) pour toutes les périodes considérées (cf. tableau II-1). Ces cinq premiers cosinus directeurs ont une portée supérieure à celle du rapport d'insolation centré deux fois (350 km). Les deux ou trois cosinus directeurs suivants ont une portée du même ordre de grandeur (300 km) puis celle-ci baisse encore (200 km), jusqu'à la seizième, après quoi les variogrammes sont pépitiqes.

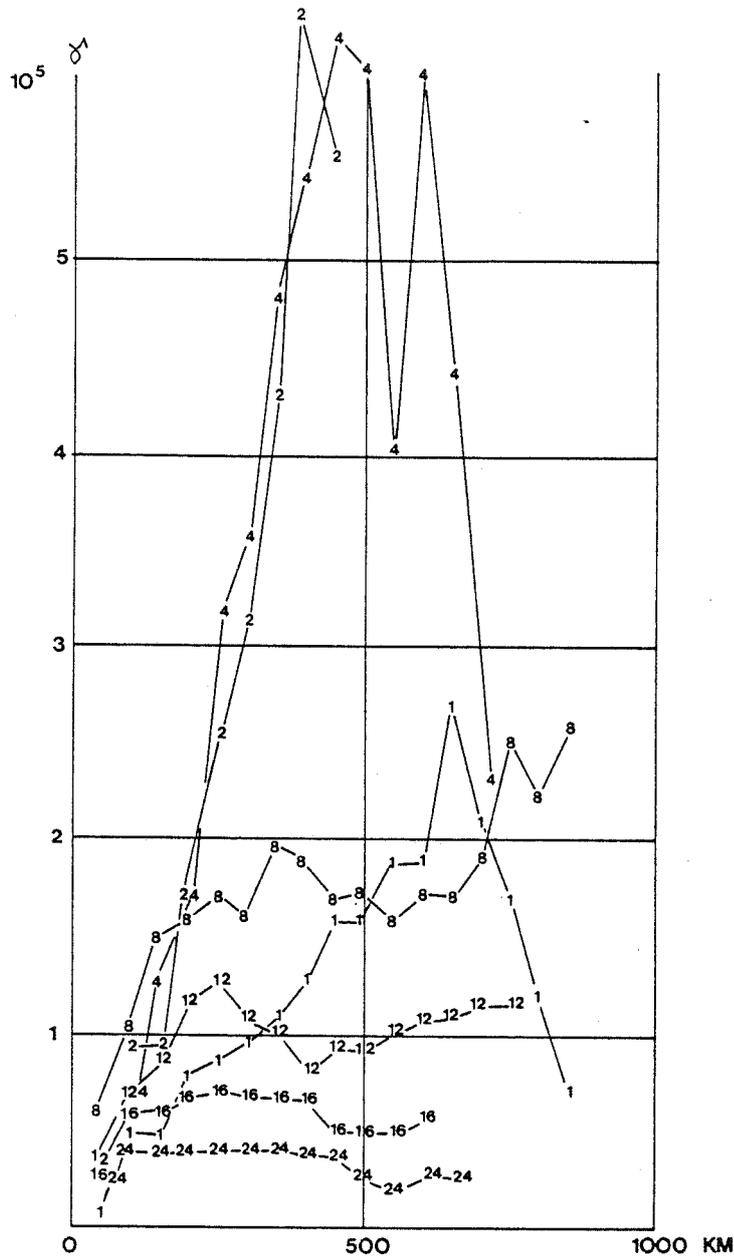


Fig II-33 : Variogramme des cosinus directeurs pour différentes composantes ( de la 2<sup>e</sup> a la 24<sup>e</sup> ).  
On note une diminution de la variance du champ des cosinus, et de sa portée, avec le rang de la composante.

. Conclusion

Chacune de ces méthodes est basée sur des critères différents. Le nombre de composantes à retenir suivant la méthode utilisée est résumé dans le tableau ci-dessous : nombre de composantes retenues (et pourcentage de variance expliquée correspondant).

méthode	hiver	printemps	été
KAISER	10(87 %)	7(88 %)	10(88 %)
LEV	9(86 %)	7(88 %)	8(86 %)
PORTEE	5(77 %)	5(85 %)	5(78 %)

Tableau II-2

Le nombre de composantes à retenir varie entre cinq et dix suivant la méthode employée. Cet écart est faible en regard du nombre de composantes total de la matrice de corrélation (environ 70). Le pourcentage de variance expliquée correspondant est en moyenne de 85 %. Le critère de la portée est le plus sélectif et celui de Kaiser le moins.

Le critère sur la portée choisi ici est assez sévère, puisqu'on ne garde que les composantes dont le cosinus directeur associé a une portée supérieure à celle du rapport d'insolation. Si on prend en compte tous les cosinus directeurs qui ont une structure spatiale significative, on peut aller jusqu'à quinze. Huit cosinus ont une portée équivalente à celle du rapport d'insolation (on rejoint alors le nombre de composantes retenues par les autres méthodes).

La méthode de Kaiser conserve le plus grand nombre de composantes : elle est basée sur la sécurité. Son seul but est de ne pas ignorer une variable.

La méthode de LEV, enfin, fournit des résultats semblables pour les trois périodes en terme de pourcentage de variance expliquée. Le nombre de composantes retenues pour la matrice calculée sur la grille satellite complète (1222 points) est identique à celui choisi pour la sous matrice (84 mailles), malgré le nombre élevé de variables. La méthode de Kaiser, par contre, ne saisit pas l'interdépendance des mailles, et en choisit trente huit. La méthode de LEV paraît plus performante.

L'ensemble des méthodes utilisées semble converger vers le même résultat, pour les deux types de mesure, et pour les trois périodes. On choisit de conserver huit, sept et huit composantes (pour l'hiver, le printemps, l'été) dites "significatives" au sens des critères cités ci-dessus. Ces nombres sont à mettre en rapport avec celui du nombre de cosinus directeurs semblables pour les deux types de mesure (sept). Seuls les cosinus directeurs significatifs sont corrélés de façon significative ; ce sont ceux présentant une portée de l'ordre de grandeur de celle du phénomène.

### III-3-e. Conclusion sur l'ACP du rapport d'insolation

L'analyse en composantes principale du rapport d'insolation centré et réduit  $y_k(t)$  mesuré par le sol et par le satellite présente les caractéristiques suivantes :

La première composante, de valeur propre (ou puissance) très élevée, explique à elle seule pratiquement la moitié de la variance totale. Cette composante est très proche de la moyenne spatiale du champ : si on fait l'ACP des valeurs centrées deux fois, elle disparaît, et on obtient directement les composantes suivantes. Elle traduit la prépondérance de la variabilité des événements sur la variabilité spatiale.

Les sept composantes suivantes, de puissance importante ( $\lambda_l > 1$ ), contiennent et résument la structure spatiale, propre au rapport d'insolation. Elles expliquent environ 85 % de la variance totale. Les cosinus directeurs correspondants ont une portée de l'ordre de grandeur de celle du rapport d'insolation centré deux fois. Ils confluent de façon très satisfaisante ( $r^2 = 90 \%$ ) pour les deux types de mesure.

Les dernières composantes, enfin, ont une puissance faible ; elles n'expliquent que les fluctuations ponctuelles d'échantillonnages ou les erreurs de mesure. Elles n'ont pas de signification climatologique pour le rapport d'insolation.

Le satellite est un bon capteur de la structure spatiale du rapport d'insolation. Sa haute définition spatiale permet d'estimer de façon détaillée les cosinus directeurs du rapport d'insolation. Ceux-ci sont d'autant plus proches des cosinus directeurs de la variable régionalisée que le nombre de points de mesure augmente. De plus, la répartition spatiale régulière évite le choix arbitraire des fonctions  $e_i(t)$ , puisque toutes les surfaces d'influence des stations sont égales. L'ACP de processus revient alors à une ACP classique.

L'hétérogénéité du rapport d'insolation est mise en évidence par comparaison des cartes de cosinus directeurs obtenues par l'information satellite fine, avec celles obtenues pour une matrice modélisée. L'ACP de l'image de télédétection permet de quantifier dans les cosinus directeurs l'hétérogénéité spatiale du rapport d'insolation, contrairement à un modèle inféré sous hypothèse de stationnarité.

## IV - RENFORCEMENT DE RESEAU

### IV-1. Introduction

Le renforcement d'un réseau de mesure, et l'élimination de stations de mesure existantes, sont des problèmes fréquemment rencontrés dans les sciences de l'environnement ; ils sont connus sous la désignation d'"optimisation de réseau de mesure".

L'élimination de stations existantes se résoud simplement par les techniques statistiques classiques d'élimination de variables. La sélection est fondée sur des critères de distance ou d'information, calculés à partir de la matrice des corrélations empiriques interstations. Le réseau sélectionné préserve alors le maximum d'informations contenues dans le réseau initial.

Le renforcement de réseau, à l'inverse, nécessite une information inconnue a priori sur le comportement du phénomène aux points où un renforcement est souhaitable. La meilleure solution pour résoudre ce problème est de considérer le phénomène mesuré comme un processus aléatoire. Alors, sous hypothèse d'homogénéité, la variance de l'erreur d'estimation de ce processus peut être calculée en tout lieu du domaine (DELHOMME et DELFINER, 1973). Les stations seront implantées aux lieux où cette variance est maximale c'est à dire là où les estimations sont les moins précises avec l'ancien réseau.

L'hypothèse fondamentale qui sous-tend cette démarche est l'homogénéité du phénomène étudié. Connaissant sa structure en quelques stations, on peut l'inférer et l'étendre à tout le domaine. Cette hypothèse n'est malheureusement pas vérifiée de façon satisfaisante par la plupart des phénomènes naturels. Nous avons vu, ainsi, que la dispersion du corrélogramme du rapport d'insolation ne peut être imputée aux seules fluctuations d'échantillonnage. Nous avons plus précisément constaté, à l'aide de l'ACP, le comportement particulier de certaines régions comme la vallée du Rhône ou la côte bretonne.

Si, en plus des mesures sol, on dispose de mesures de télédétection, la grille de télédétection peut alors être considérée comme un réseau redondant. Le problème du renforcement se ramène alors à celui de la sélection de variables. Toute hypothèse d'homogénéité est ainsi évitée et on peut prendre en compte, de façon quantitative, les comportements locaux particuliers. Mais il faut pour cela démontrer que les deux types de mesure ont une appréciation identique de la variabilité du phénomène on peut alors considérer la grille satellitaire comme un deuxième réseau.

#### IV-2. Comparaison de la structure perçue par les héliographes et la méthode CACTUS

Les fonctions de corrélation directes et croisées des deux types de mesure reflètent le comportement et la co-fluctuation moyenne des coefficients de corrélation expérimentaux, en fonction de la distance interstation. Leur similarité (cf. 2ème partie II) ne suffit cependant pas pour l'application qui nous intéresse : leur perception de l'hétérogénéité doit également être semblable. Cette deuxième contrainte a été vérifiée par ACP (cf. 2ème partie III). En effet, la décomposition du rapport d'insolation sur ses composantes principales donne des cosinus directeurs similaires pour les deux types de mesure (au moins les sept premiers). Ces cosinus directeurs traduisent la structure et l'hétérogénéité du rapport d'insolation, qui est donc perçue de façon similaire par le sol et le satellite.

Cette analyse peut être complétée par l'examen de la dispersion expérimentale des corrélogrammes. Nous avons porté la corrélation du sol en fonction de celle du satellite pour les couples de points communs aux deux réseaux. La figure (II)-34 regroupe à la fois ce nuage et les deux corrélogrammes expérimentaux du sol et du satellite. La dispersion du premier nuage de points est remarquablement plus réduite que celle des deux corrélogrammes. La dispersion du corrélogramme s'explique donc en partie par la localisation géographique.

Les deux types de mesures ont une appréciation similaire de la variabilité moyenne du rapport d'insolation, ainsi que de son hétérogénéité. A titre d'illustration de la dispersion due à l'hétérogénéité et à

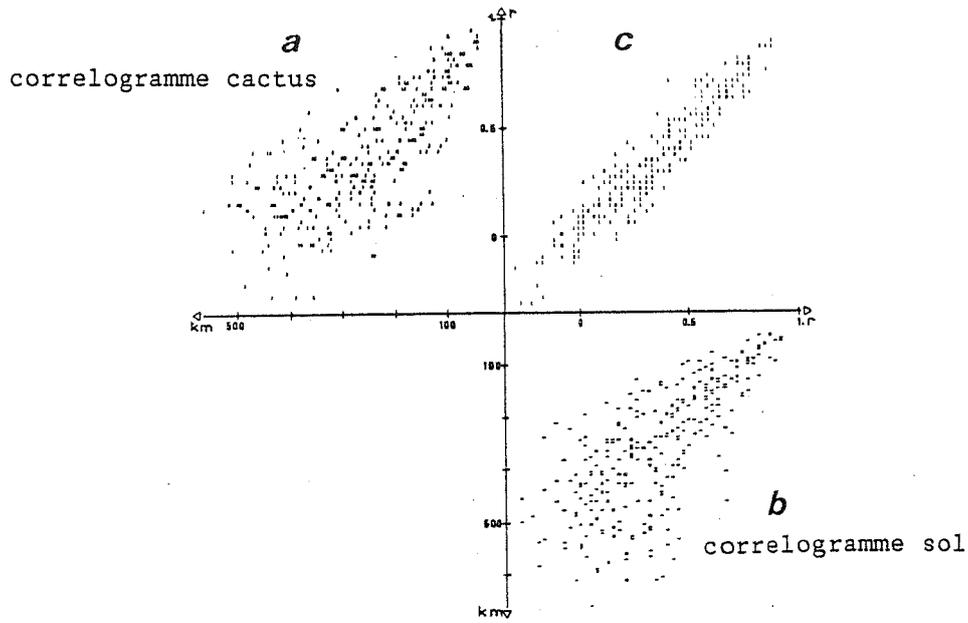


Fig II-34 : Le nuage de correlation (c) qualifie la liaison existant entre les coefficients de correlation du sol et du satellite pour un couple de stations donné.

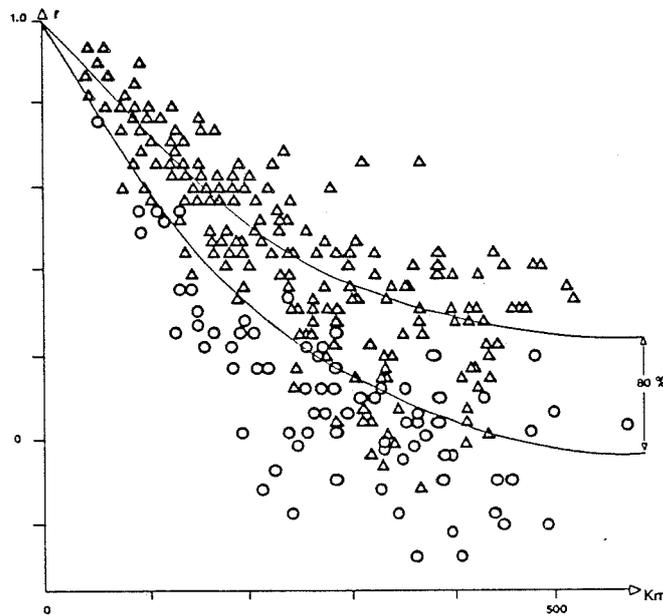


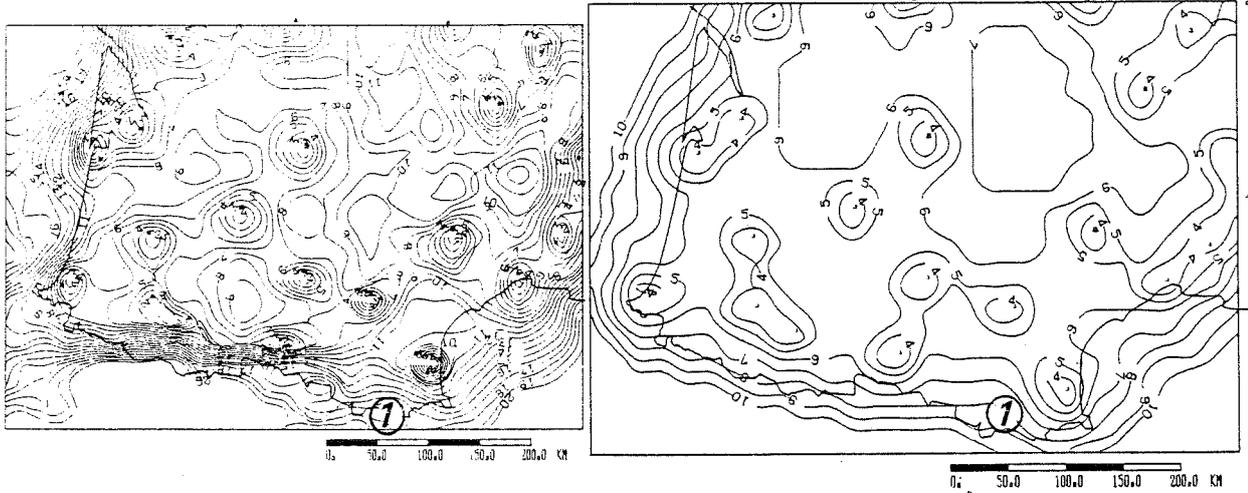
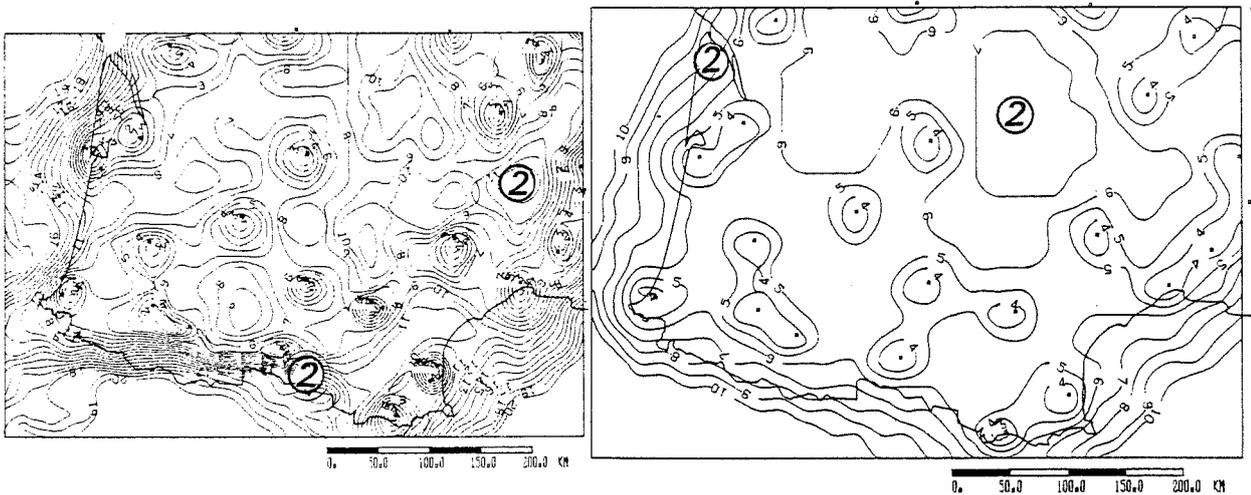
Fig II-35 : Répartition des couples de points correspondant aux quatres stations : Clermont, St Etienne, Le Puy, Lyon (repérées ○), dans le correlogramme des 23 stations du sud ouest (Δ). Leur correlation est systématiquement plus faible pour une distance donnée.

l'échantillonnage, on a pointé (Fig.(II)-35) le corrélogramme expérimental des 23 stations sol du sud-ouest (Fig.(II)-37), au printemps. Sur ce nuage, on a tracé les intervalles de confiance à 80 % sur la distribution de la corrélation autour du modèle homogène, pour l'échantillon considéré. Le nuage y est compris à 70 % seulement. Dix pour cent de la dispersion du nuage reste inexpliquée. On a, de plus, distingué les points correspondant aux intercorrélations avec les quatre stations suivantes : Clermont, Lyon, St Etienne et Le Puy. Ces quatre stations sont voisines géographiquement et ont un comportement nettement distinct des autres : leurs corrélations (o) sont significativement inférieures, pour une interdistance donnée à celle des autres stations entre elles ( $\Delta$ ). Le phénomène n'est manifestement pas homogène.

#### IV-3. Critères de sélection utilisés

Notre but est de sélectionner parmi les G mailles du satellite le lieu où il semble judicieux d'ajouter une station sol compte tenu des p stations sol déjà implantées. Ces p stations sol ne correspondent, en général, pas aux centres des mailles : on les affectera à la maille la plus proche. De même, c'est une maille qui sera sélectionnée (20x20 km<sup>2</sup>) et non un site ponctuel. L'approximation ainsi faite permet une certaine latitude quant au lieu exact de l'implantation. D'autre part, la précision définie par une maille est très fine devant l'interdistance moyenne initiale des stations.

La littérature offre un vaste éventail de possibilités pour sélectionner des variables (C. OBLED, J.D. CREUTIN, 1980). Les premières, fondées sur la notion de distance, sont des méthodes d'agrégation. Elles partent de variables isolées que l'on regroupe. Une variable représentative est ensuite choisie dans chaque groupe. Les autres méthodes utilisent des critères d'information. Elles sélectionnent individuellement des variables. Ce deuxième type de méthode peut fonctionner dans le sens de l'élimination (méthode descendante) ou bien de la sélection (méthode ascendante).

1<sup>er</sup> pas2<sup>ème</sup> pas

structure du satellite

structure homogène

Fig II-36 : Cartes de variance d'estimation, et lieu d'implantation de la nouvelle station (sur un maximum), avec et sans hypothèse d'homogénéité.

On a présenté les deux premiers pas de sélection. On note l'effet d'implantation d'une nouvelle station sur la variance d'estimation.

Dans notre cas, l'existence d'un réseau initial nous conduit à utiliser une méthode ascendante dont les premiers pas sont forcés sur les stations existantes qui sont des variables choisies a priori (S. GALLE, J.D. CREUTIN, 1986). Nous avons utilisé deux critères différents : I et II.

I. A chaque pas, la maille qui minimise la variance résiduelle moyenne est sélectionnée.

II. Sélection de la maille présentant la corrélation partielle la plus faible avec le réseau existant (Fig.(II)-36). Notons que ce critère correspond à la variance d'estimation utilisée en géostatistique. Notre avantage, tiré de la télédétection est de connaître la structure en tout point et d'éviter ainsi l'hypothèse d'homogénéité habituelle.

Ces deux critères sont différents par deux aspects : i) l'un se réfère aux mailles non encore expliquées et l'autre à celles déjà choisies ; ii) l'un considère une moyenne et l'autre une valeur extrême.

#### IV-4. Résultats obtenus dans le sud-ouest de la France

Afin de juger l'apport de la télédétection qui évite l'hypothèse d'homogénéité, les corrélations utilisées pour chaque critère ont été estimées de deux manières :

i) en utilisant la fonction de corrélation  $r_{ZZ}(h)$ , c'est-à-dire en faisant une hypothèse d'homogénéité,

ii) à partir des corrélations expérimentales du satellite.

Les résultats sont présentés (Fig.(II)-38). Les sites sélectionnés sont numérotés en fonction de leur rang de sélection.

#### Sensibilité des résultats à l'hypothèse d'homogénéité

Sous l'hypothèse d'homogénéité, les sites sélectionnés reflètent les lacunes du réseau, de façon géométrique. L'utilisation de la télédétection, au contraire, délaisse de grandes zones pourtant sous

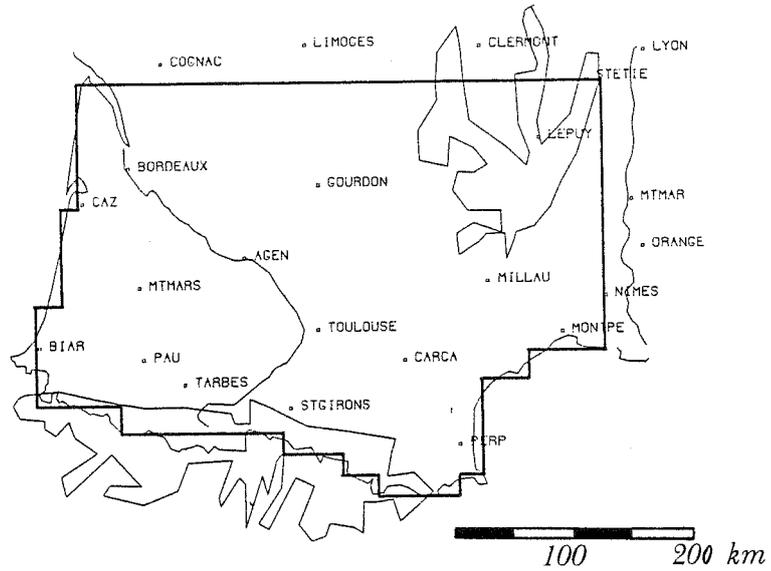


Fig II-37 : Localisation du réseau d'héliographes du sud-ouest, et détermination de la zone d'étude.

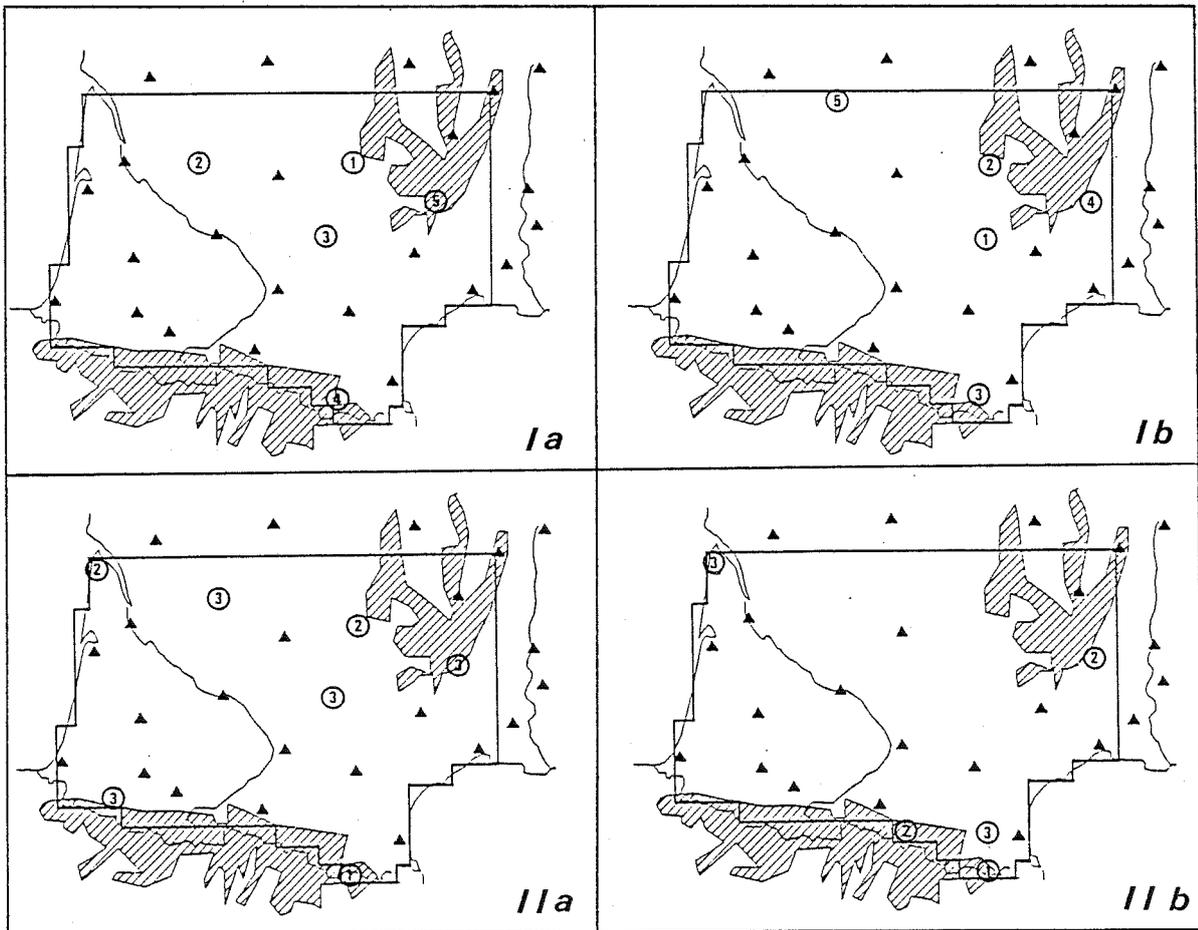


Fig II-38 : Lieu d'implantation des nouvelles stations numérotées suivant leur rang de sélection.

I : premier critère  
a : structure modélisée

II : deuxième critère  
b : structure calculée sur les données cactus

équipées (Bassin Aquitain) pour leur préférer des régions montagneuses (Pyrénées, Massif Central).

#### Sensibilité des résultats aux critères choisis

Les deux critères conduisent à des choix assez différents. Alors que le premier choisit les zones peu instrumentées, le deuxième sélectionne les sites situés en lisière du domaine (ils ne sont pas entourés). L'un travaille sur des moyennes, l'autre sur des valeurs extrêmes.

#### IV-5. Conclusion

Les résultats obtenus montrent clairement l'intérêt de prendre en compte l'hétérogénéité du phénomène.

Les sites ainsi sélectionnés sont cohérents avec notre connaissance a priori de la durée d'insolation. Toutefois, une telle méthode n'est pas apte à séparer les fluctuations d'échantillonnage de l'hétérogénéité. Il faut donc rester vigilant. Deux solutions peuvent être envisagées pour pallier à cette difficulté. La première serait de considérer une période plus longue, la deuxième d'utiliser une matrice de corrélation lissée : reconstituée avec  $L$  composantes "significatives". Le bruit serait ainsi éliminé.

**3ème PARTIE**

**VALIDATION**

## INTRODUCTION

Dans cette troisième partie, les mesures du programme CACTUS sont évaluées par rapport aux données sol.

Pour valider les mesures fournies par la méthode CACTUS, nous disposons des héliographes du réseau synoptique de la Météorologie Nationale. Ces mesures nous fournissent la "vérité sol", ou référence, à laquelle on va comparer les mesures satellite.

A l'aide de différents critères, on quantifie d'abord la co-fluctuation des données fournies par les deux types de mesure. Cette étude est faite pour différents pas de temps, allant du plus fin (journalier) au plus grossier (trimestre).

Afin de mieux situer les performances du satellite ainsi quantifiées, nous les avons ensuite comparées à celles de techniques d'interpolation classiques, utilisant les seules données sol. Cette deuxième étape est possible grâce à la densité du réseau d'héliographes couvrant la France. Ce réseau a été séparé en deux : une partie utilisée uniquement pour établir la référence et l'autre utilisée pour les interpolations à partir du sol seul. Les tests réalisés se réfèrent toujours au même réseau de référence indépendant.

### Les critères utilisés

La haute résolution spatiale du satellite nous paraît a priori très attractive. L'observation d'une image, par les détails qu'elle fournit, nous conduit spontanément à attacher une grande valeur de réalisme à ces données.

Il faut cependant se méfier de tels raisonnements car peu importants les détails si ils n'ont pas un sens quantitatif : un bruit blanc n'apporte aucune information, il est pourtant variable. Dans un premier temps, la

cofluctuation des phénomènes sol  $Z_S$  et satellite  $Z_T$  sera estimée par leur coefficient de corrélation, calculé sur les points de mesure commun.

$$r_{Z_S Z_T} = \frac{\text{Cov}_{Z_S Z_T}}{\sqrt{\text{Var}_{Z_S} \text{Var}_{Z_T}}}$$

Le coefficient de corrélation quantifie la relation linéaire existant entre  $Z_S$  et  $Z_T$  (le coefficient est maximum lorsque la relation  $Z_T = aZ_S + b$  est vérifiée). Ce coefficient possède l'avantage de n'être sensible qu'à la cofluctuation des deux phénomènes : il est insensible à toute transformation linéaire ( $r_{Z_S, Z_T} = r_{aZ_S + b, Z_T}$ ).

On rappelle que son carré  $r_{Z_S Z_T}^2$ , appelé coefficient de détermination

exprime le pourcentage de variance de  $Z_S$  expliqué par  $Z_T$  à travers la relation linéaire existant entre eux. L'interprétation du coefficient de

détermination est immédiate : si  $r = .9$ , alors  $r^2 = 81$ , c'est à dire 81 % de la dispersion de  $Z_S$  est liée à la variabilité de  $Z_T$ . Le coefficient de corrélation peut être calculé sur des séries temporelles (caractérisant la cofluctuation des mesures en une station), et peut également être défini sur un ensemble de points de mesure (caractérisant la cofluctuation des mesures pour une observation donnée) ; les moments sont estimés sur les  $n$  stations disponibles pour l'observation considérée.

Le nuage de corrélation est la représentation graphique des  $n$  couples de points  $(Z_S(t_i), Z_T(t_i))$ . Il est indispensable de le représenter pour chaque coefficient de corrélation calculé. On évite ainsi de tomber dans les pièges classiques de la corrélation (un point isolé peut augmenter ou diminuer artificiellement la corrélation, une relation non linéaire entre deux phénomènes n'est pas prise en considération par la corrélation

classique). Si nous n'avons pas pu tous les présenter ici pour une raison de place, nous les avons tous visualisés au cours du travail. Nous avons également calculé les erreurs de reconstitution de  $Z_S$  par  $Z_T$  pour une observation donnée  $k$ . Différents types d'erreurs peuvent être définis, nous avons choisi de travailler sur les valeurs absolues des erreurs simples normées par la moyenne du champ  $Z_{Sk}$

$$E_k = \frac{1}{p} \frac{\sum_i |Z_{Si} - Z_{Ti}|}{\bar{Z}_{Sk}} \quad \text{et} \quad E_a = E(E_k)$$

Ainsi une erreur forte sera moins importante si elle intervient sur un champ de moyenne élevée que sur un champ de moyenne faible. Cette erreur est à rapprocher des "erreurs" habituellement citées dans la littérature, bien que leur mode de calcul exact soit rarement précisé. Nous avons enfin étudié l'histogramme des erreurs simples. Cette information est contenue dans le nuage de corrélation, mais l'histogramme la visualise plus facilement.

#### Prise en compte de la variation saisonnière

Les 200 journées disponibles de l'année 1984 ont été regroupées en trois périodes homogènes du point de vue des caractéristiques statistiques du rapport d'insolation : le printemps, l'été et l'hiver. La détermination de ces périodes a été discutée dans la partie précédente. Leur composition est donnée en annexe 4.

## I - VALIDATION DES MESURES SATELLITAIRES

### I-1. Etude de la corrélation des valeurs journalières

Une image Météosat traitée par le programme CACTUS ou "image CACTUS" est constituée de 3300 valeurs. Le réseau d'héliographes est bien moins abondant. Seules les mailles correspondant aux stations sol de référence sont retenues pour l'évaluation. Certaines régions de l'image satellite qui correspondent à une zone géographique sans héliographe sont ainsi ignorées par le test. Le rapport d'insolation étant un phénomène de grande portée (350 km) et bien instrumenté au sol (interdistance moyenne des héliographes = 60 km environ), l'image de télédétection sera globalement bien validé.

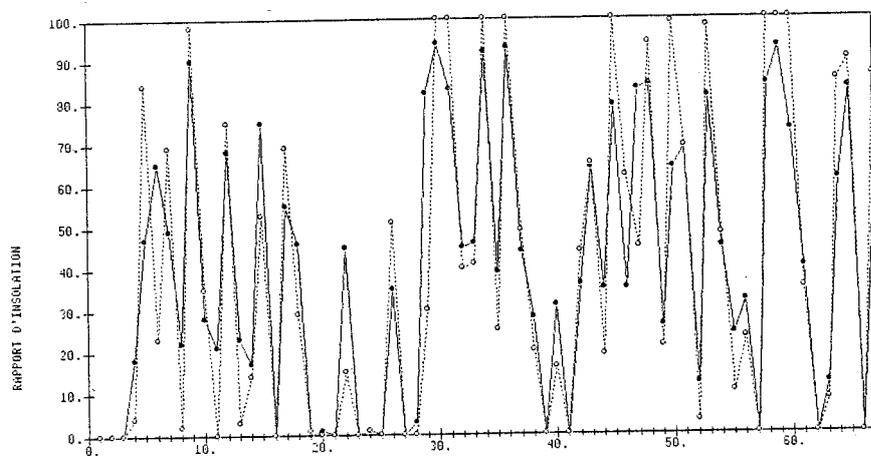
D'autre part, les mesures satellite intégrées sur des mailles sont comparées à des valeurs sol ponctuelles. L'étude de la structure spatiale des deux types de mesure a montré que l'effet d'intégration de la mesure CACTUS est compensé par l'effet de l'échantillonnage temporel (cf. 2ème partie, II-2) : les deux types de mesure sont donc directement comparables.

#### I-1-a. Illustration de la co-fluctuation temporelle

Nous avons voulu tout d'abord illustrer la co-fluctuation temporelle des deux types de mesure. Deux stations ont été choisies : Toulouse et Bourg Saint Maurice. Pour chacune d'entre elles, on a porté sur le même graphique les mesures sol et satellite, pour tous les jours disponibles de 1984.

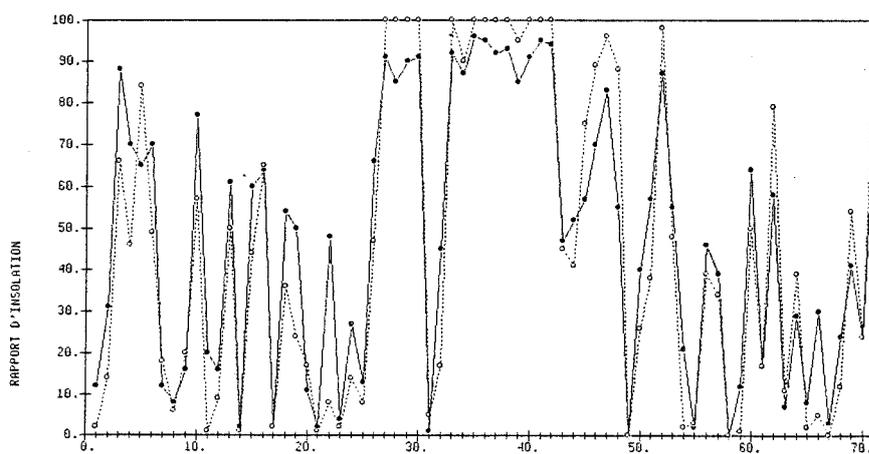
A Toulouse (Fig.(III)-1), on observe une bonne correspondance entre les mesures sol et CACTUS ; cette concordance reste valable pour des valeurs moyennes du rapport d'insolation qui sont, a priori, plus difficiles à quantifier (en effet, il est facile de dire à l'aide de l'image satellite visible qu'une journée a été totalement dégagée ou totalement couverte, il

Fig III-1 : cofluctuation temporelles des mesures (○) cactus, et (●) sol, pour les trois périodes.

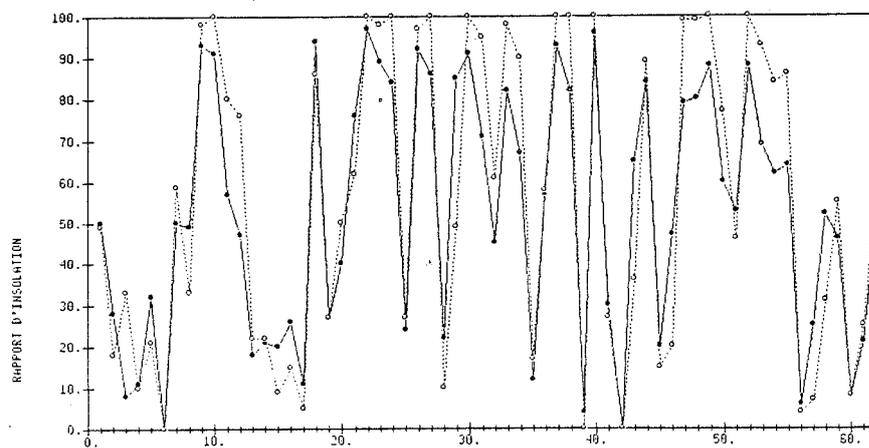


TOULOUSE

HIVER

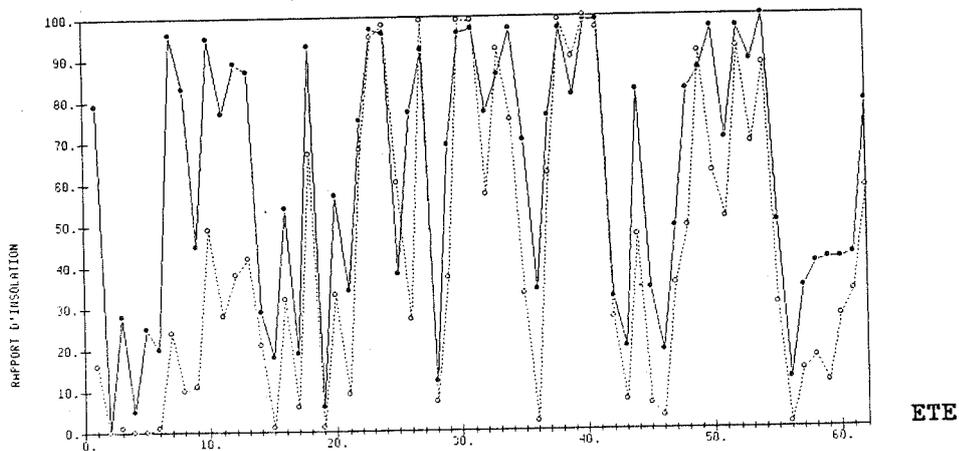
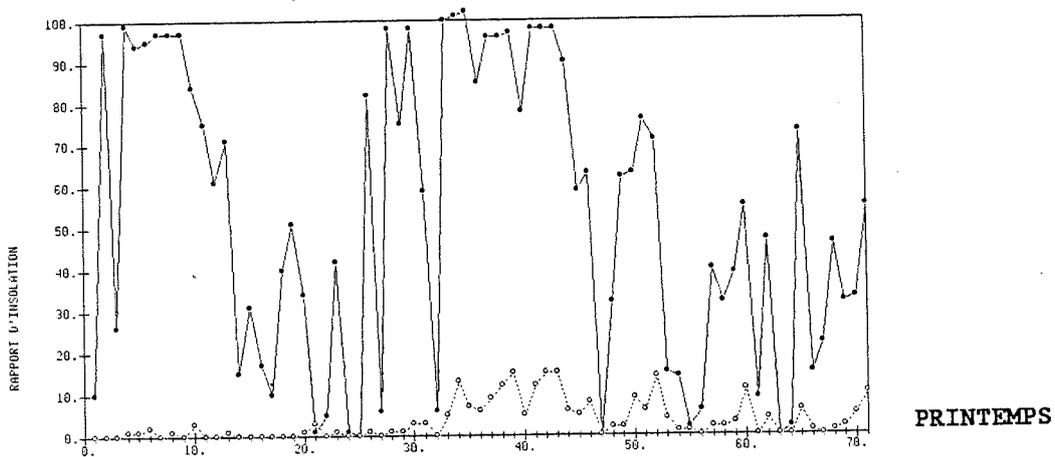
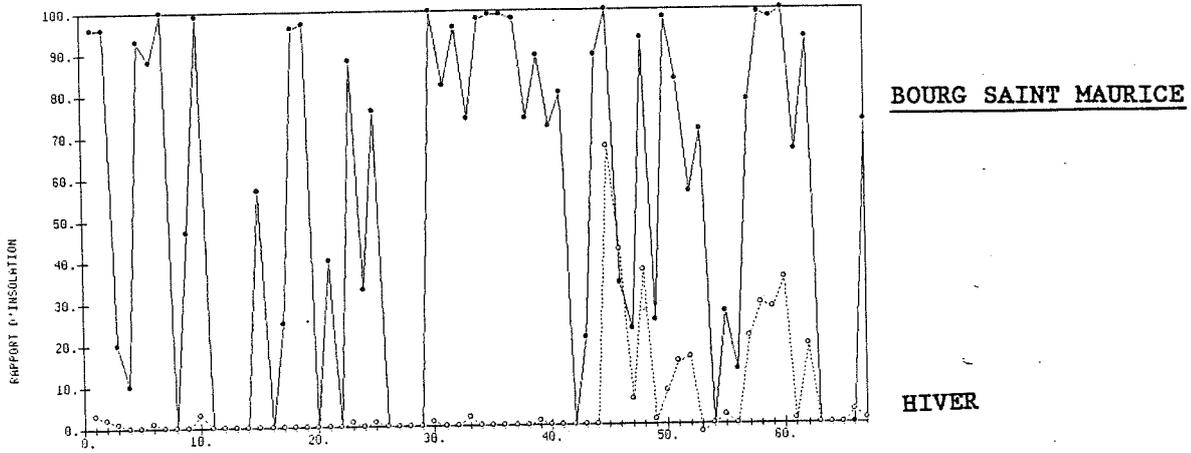


PRINTEMPS



ETE

Fig III-2 : confluctuation temporelles des mesures (0) cactus, et (●) sol, pour les trois périodes.



est plus difficile de quantifier des passages nuageux). Il apparaît également que le satellite a tendance à saturer plus vite à 0 % et 100 % que le sol. Cette observation est à rapprocher de la remarque relative à l'échantillonnage temporel (cf. 2ème partie II-2). Cet échantillonnage provoque une perte d'information et appauvrit les nuances. La variance du satellite est ainsi plus forte que celle du sol.

Les résultats observés pour Bourg-St-Maurice sont très différents de ceux de Toulouse : le satellite s'éloigne très peu de 0 % (couverture nuageuse totale) alors que le sol fluctue normalement. Cette disparité entre les deux types de mesure est expliquée par la présence d'un couvert neigeux ; la station de Bourg-St-Maurice est entourée de sommets élevés et ses environs sont souvent enneigés. Le satellite ne réussit pas à distinguer la neige des nuages à l'aide d'un seul canal visible, ce qui explique sa mauvaise performance. Cette constatation est très nette pour la période d'hiver. En été, en revanche, la co-fluctuation entre les mesures sol et CACTUS est beaucoup plus affirmée, mais s'accompagne toujours d'une sous-estimation considérable du rapport d'insolation par le programme CACTUS. Seule une partie des pixels composant la maille du satellite correspond à une zone dégagée, les autres restant enneigés (les sommets montagneux). La co-fluctuation observée est due aux seuls pixels recouvrant les vallées.

Ce comportement a été observé pour toutes les stations de montagne, aussi avons nous décidé de ne pas étudier les résultats du programme pour les mailles dont un pixel est situé à une altitude dont la probabilité d'enneigement est forte (supérieure à 2000 m) - ceci quelle que soit la période de l'année.

#### I-1-b. Corrélation temporelle

##### - Corrélation globale

Pour chacune des trois périodes, on a calculé le coefficient de corrélation entre les données des 84 héliographes du réseau synoptique de

la Météorologie (situés à moins de 2000 m d'altitude) et les valeurs des mailles correspondantes du satellite.

Ces résultats sont résumés dans le tableau suivant :

période	r	r <sup>2</sup> %	moyenne (et écart-type)		erreur E <sub>a</sub> %
			sol %	CACTUS %	
- hiver	.88	77	33 (32)	29 (36)	35
- printemps	.93	86	44 (34)	45 (39)	25
- été	.91	83	55 (29)	58 (36)	22

Tableau III-1

Remarques générales sur les nuages de corrélation correspondants (cf. Fig.(III-3 abc) :

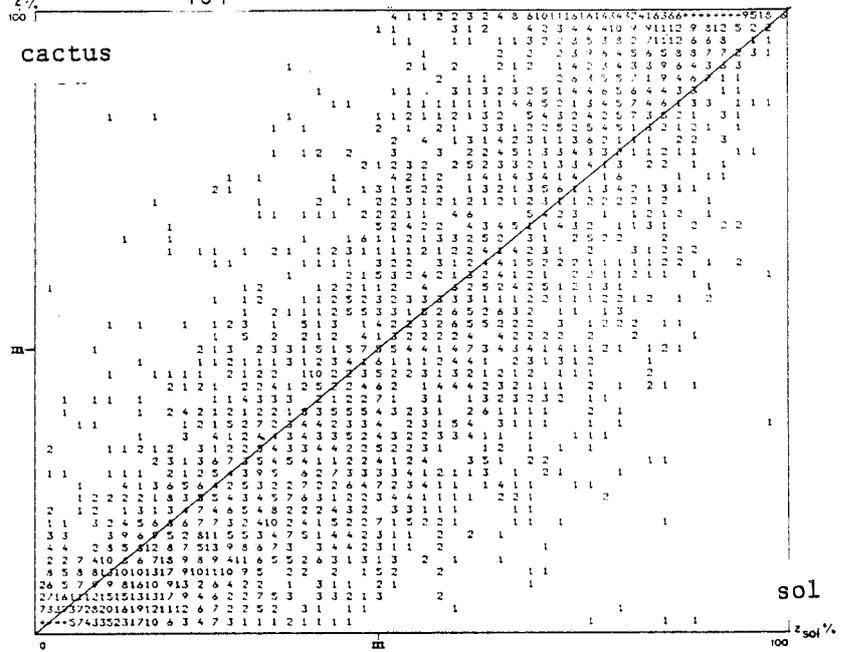
- Pour les trois périodes, il n'y a pas de sur ou sous-estimation systématique du rapport d'insolation par le satellite. Une modification du seuil ne semble donc pas utile.
- Le satellite sature plus vite que le sol à la fois à 0 % (valeurs faibles) et 100 % (valeurs fortes), ainsi que le montre l'exemple de la station de Toulouse.
- La distribution du rapport d'insolation, nettement bimodale, (cf. Fig.(III)-4), conduit à des nuages en "haltères" qui augmentent artificiellement la valeur des coefficients de corrélation. Chaque période possède des caractéristiques propres :

Fig III-3 :

Nuage de corrélation des mesures sol avec les données cactus

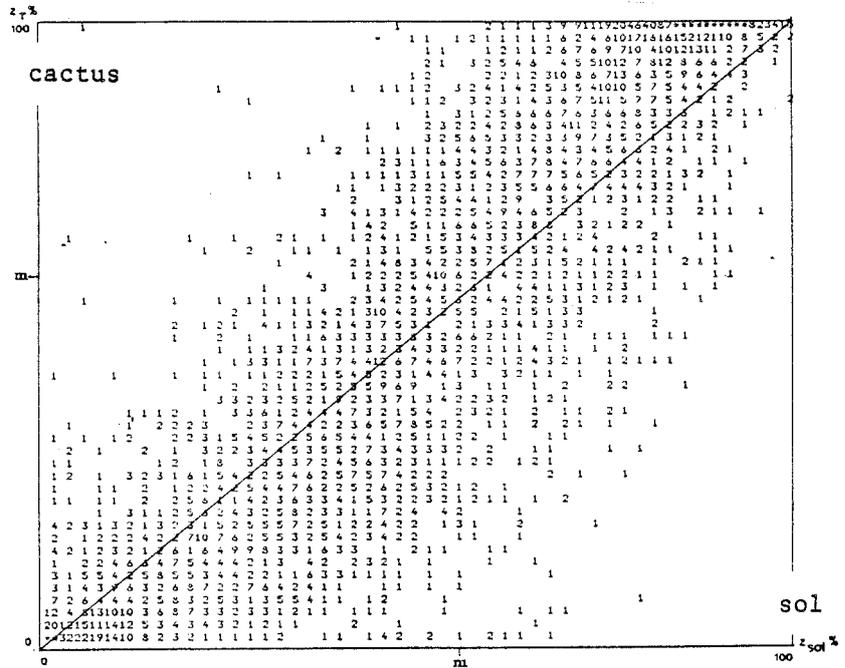
PRINTEMPS

r = .93



ETE

r = .91



HIVER

r = .88

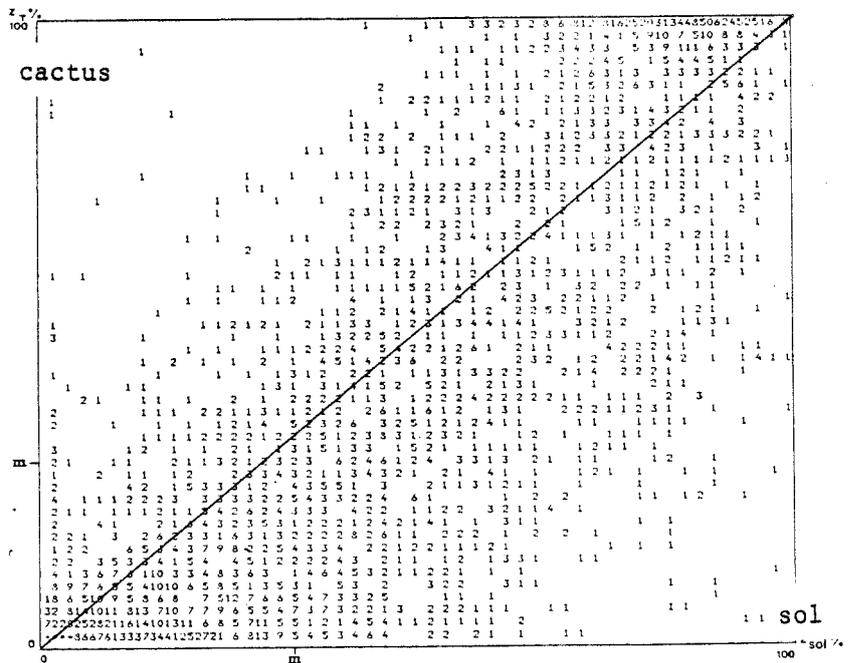
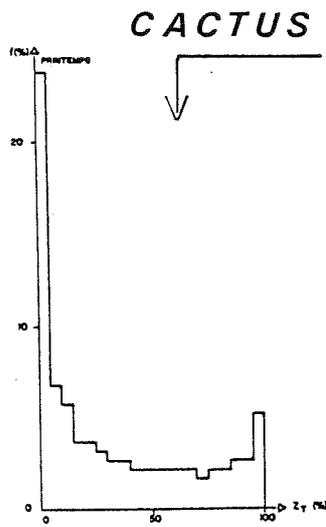
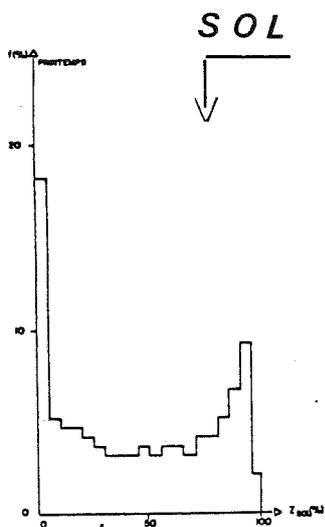
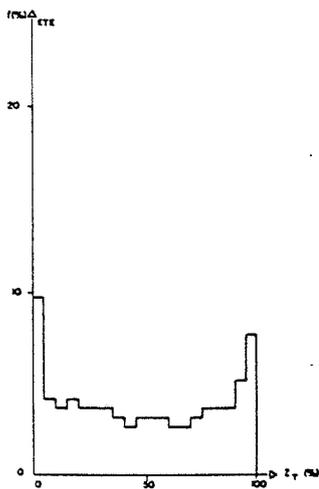
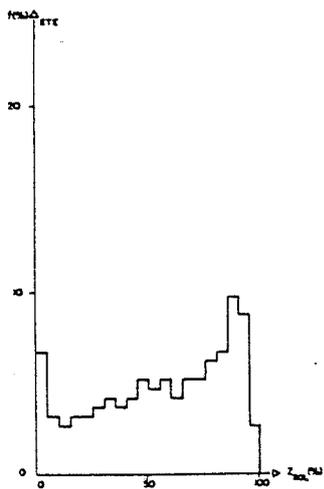


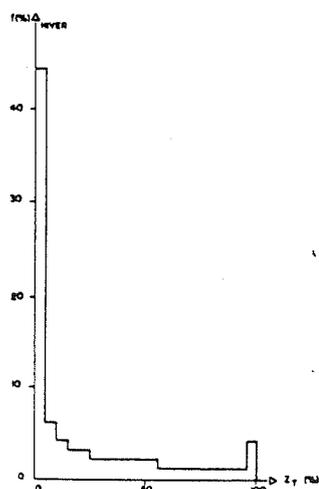
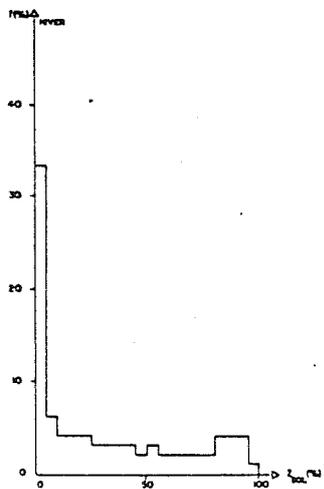
Fig III-4 : Histogramme bimodal des données sol et Cactus



PRINTEMPS



ETE



HIVER

a) Hiver

C'est en hiver que les performances du satellite sont les moins bonnes. Le nuage de corrélation est dispersé. Ceci peut être lié au choix du seuil (qui a été réglé pour avril), mais plus probablement au temps d'hiver, souvent de type frontal, avec des passages rapides de nuages. Le pas de temps choisi : une image par heure est alors trop lâche ; un nuage (ou une éclaircie) a le temps de traverser complètement une maille entre deux prises de vue : il ne sera alors pas pris en compte par le satellite mais, par contre, l'héliographe aura enregistré son passage.

De plus, l'élimination des stations situées en haute montagne ne supprime pas totalement, en hiver, les erreurs provoquées par les zones enneigées, qui peuvent s'étendre aux altitudes inférieures à 2000 m. Les points présents dans le quart bas droit du nuage de corrélation traduisent ce problème d'enneigement.

L'estimation du rapport d'insolation à partir d'images instantanées horaires du satellite permet cependant d'expliquer plus de 75 % de la variance des données sol.

b) Printemps

C'est pour le mois d'avril, situé au milieu de cette période qu'ont été réglés les seuils du programme CACTUS et c'est également là qu'il donne les meilleurs résultats ( $r = .93$ ). Le nuage de corrélation est beaucoup moins étalé qu'en hiver. On n'observe pratiquement plus de points dans les zones extrêmes (bas droit, haut gauche).

c) Eté

Il y a une très importante proportion de valeurs très fortes du rapport d'insolation qui sont très bien vues par le satellite. Les résultats obtenus en été sont pratiquement semblables à ceux du printemps.

#### d) Conclusion

Le rapport d'insolation est bien estimé globalement par le satellite puisque dans le cas le plus défavorable en hiver, on a 77 % de variance expliquée, pour atteindre 86 % en été. Cependant, l'observation des nuages de corrélation montre une dispersion assez forte, ainsi qu'une sous-estimation des valeurs faibles et une surestimation des valeurs fortes.

#### I-1-c. Corrélation des valeurs centrées par la moyenne du champ

Le coefficient de corrélation est un outil destiné à l'étude de phénomènes de distribution gaussienne. Le rapport d'insolation, comme nous l'avons vu plus haut, a une distribution bimodale. Afin de se rapprocher d'une distribution gaussienne, on centre les valeurs par la moyenne du champ. Cette transformation est nécessaire (cf. 2ème partie, II-1) pour retirer la part due à la variabilité inter-réalisation contenue dans le rapport d'insolation. C'est ce codage qui a été utilisé pour déterminer la structure spatiale du phénomène. La corrélation ainsi calculée ne sanctionne plus que la cofluctuation spatiale des deux phénomènes.

Les nuages de corrélation obtenus pour les valeurs centrées (Fig.(III)-5 abc) sont toujours meilleurs en été qu'en hiver, mais, par rapport au tableau III-1, les valeurs des coefficients de corrélation diminuent sensiblement.

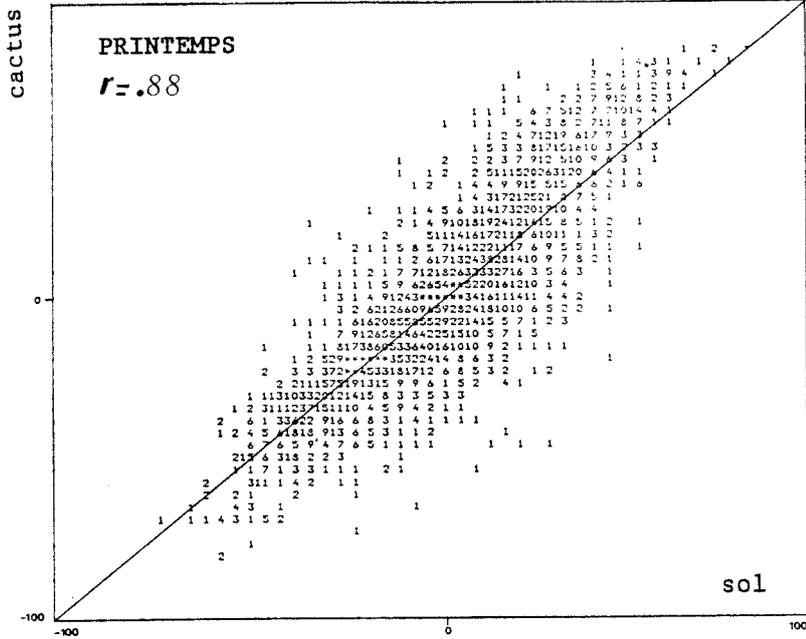


Fig III-5 :

Nuage de corrélation des mesures sol avec les données cactus :

valeurs centrées deux fois

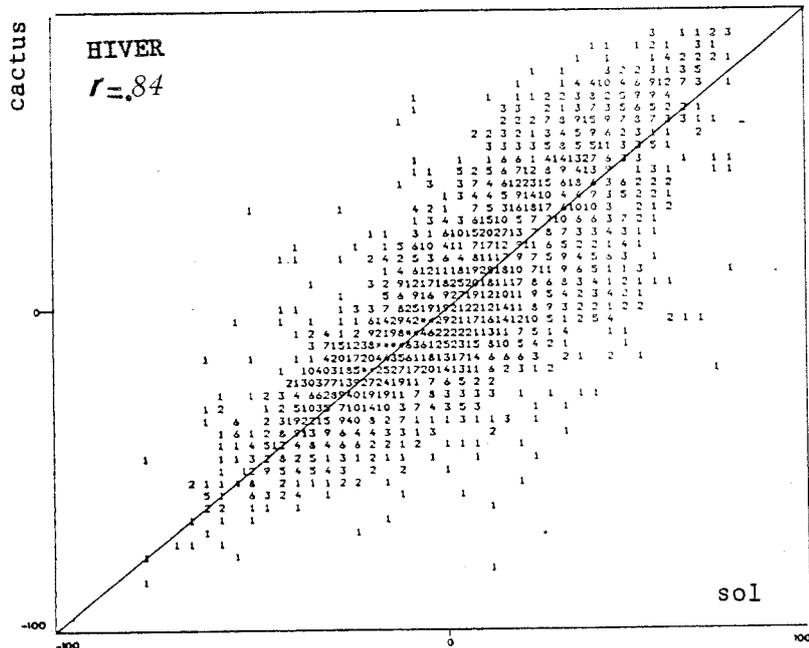
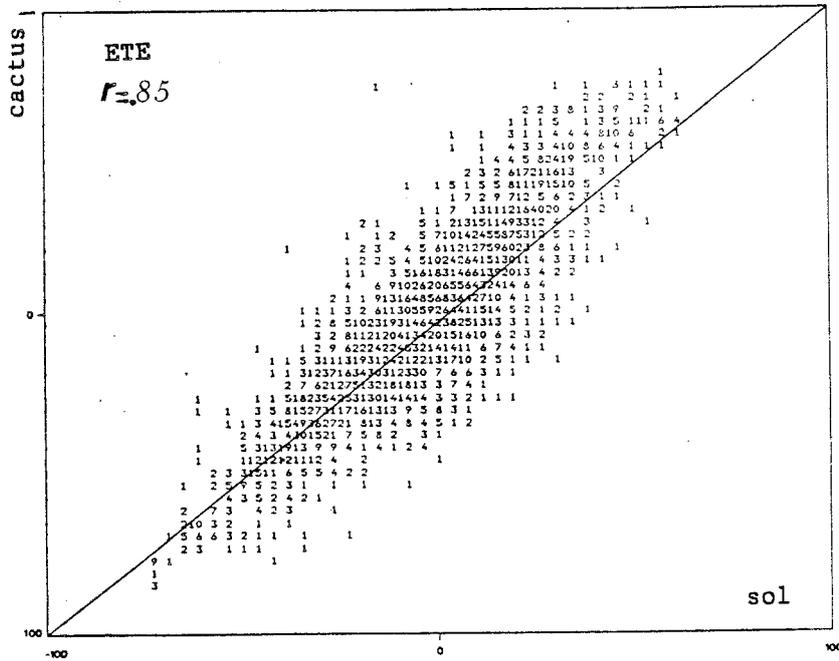


Fig III-4 : Histogramme des données sol et Cactus centrées deux fois

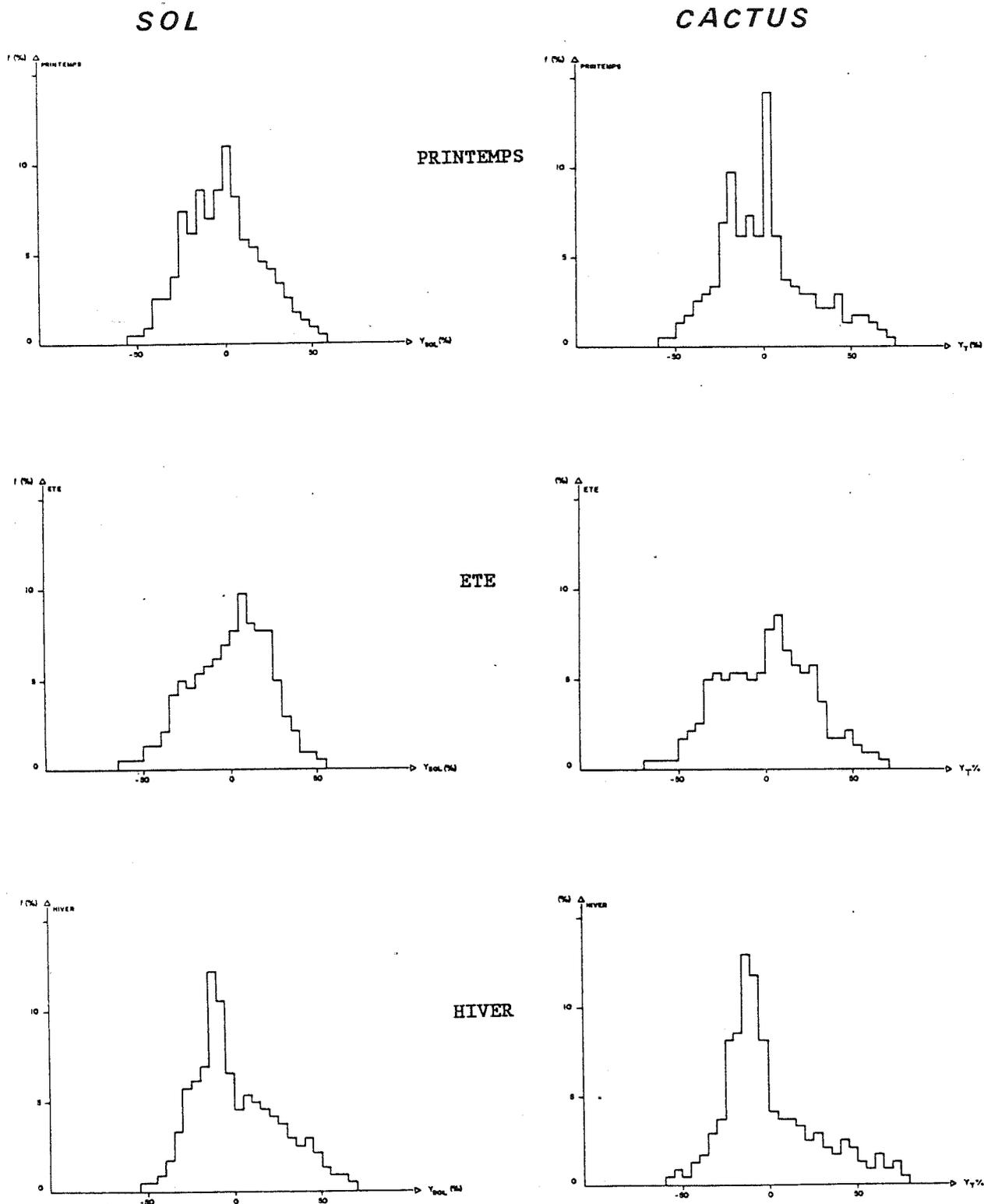
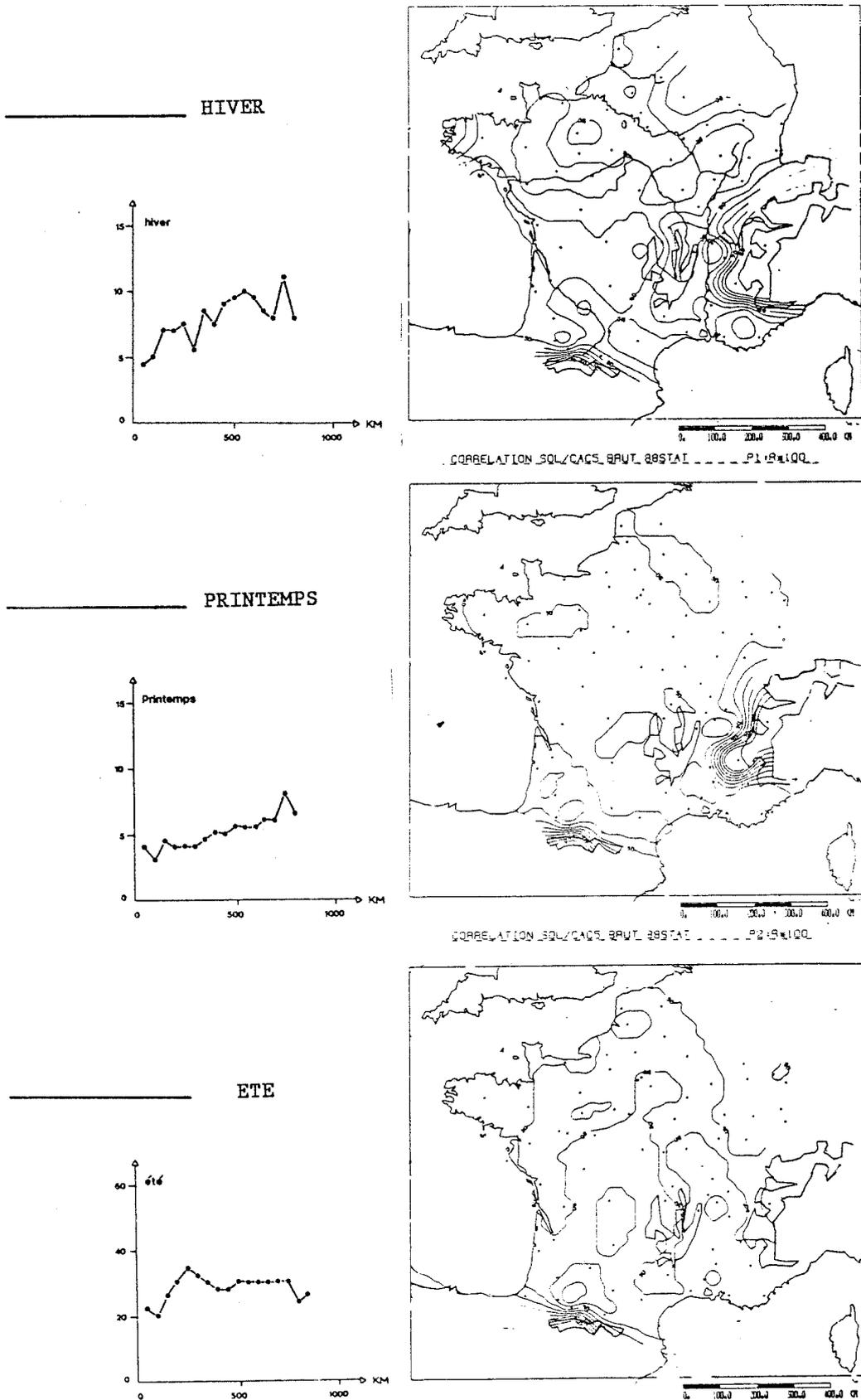


Fig III-7 : Cartes (et variogrammes associés) de la corrélation entre les données sol et cactus, (obtenues à partir de 88 stations).



période	r	r <sup>2</sup>	moyenne		erreur E <sub>a</sub>
			sol	CACTUS	
- hiver	.84	71	0.7	0.5	-
- printemps	.88	77	0.5	0.5	-
- été	.89	79	0.5	0.5	-

Tableau III-2

La valeur de l'erreur absolue normée par la moyenne E<sub>a</sub> n'a plus de sens puisque, par définition, la moyenne spatiale est nulle. La bonne co-fluctuation spatiale du sol et du satellite est confirmée. Ces coefficients de corrélation sont par définition les mêmes que ceux des grandeurs centrées deux fois ; seule la pente de la droite de régression est différente. Elle est de 1.1 pour les valeurs centrées deux fois, ce qui est à mettre en rapport avec le coefficient multiplicateur  $\alpha = 1.15$  déterminé par l'analyse de la structure (cf. 2ème partie, II).

#### I-1-d. Cartographie de la corrélation temporelle

Nous avons jusqu'alors étudié les résultats globaux sans tenir compte d'une quelconque répartition géographique. Il s'agit maintenant de voir s'il existe des zones privilégiées ou défavorisées de façon systématique par rapport à la corrélation.

Pour ceci, nous avons tracé la carte décrivant les corrélations temporelles, obtenue à partir des valeurs aux 88 stations test (on a inclus les zones de montagne (Fig.(III)-7)). Les variogrammes correspondants ont

été calculés à partir de 84 stations pour éliminer la dérive due aux zones enneigées. Ces variogrammes sont très pépitiqes : il ne se dégage pas de structure bien définie. Ceci est dû au fait que toutes les valeurs de corrélation sont assez voisines. Le champ des corrélations ne présente pas d'organisation spatiale nette.

Le tracé des cartes montre effectivement qu'il ne se dégage pas de tendance générale, ou de zone particulière pour les régions situées à basse altitude. Pour les zones de haute montagne en revanche (Alpes, Pyrénées), la présence de la neige entraîne une chute brutale des performances du satellite ; ce gradient disparaît toutefois en été sur les Alpes. Il est intéressant de noter également que des massifs d'altitude moyenne, comme le Massif Central, présentent en hiver un gradient en  $r$  (moins marqué cependant) qui disparaît aux autres périodes.

#### I-1-e. Corrélation spatiale

Après avoir regardé comment varient deux séries de mesures en un point donné, nous allons examiner, pour chaque observation, la co-fluctuation spatiale des mesures du sol et du satellite. Pour ceci, on a utilisé les courbes d'efficacité : le coefficient de détermination spatial est porté en fonction des différents événements disponibles. Rappelons que les coefficients de détermination ( $r^2$ ) expriment le pourcentage de variance expliquée. Sur le même graphique, a été portée la valeur de l'écart-type spatial du sol comme indicateur de la dispersion des observations. Il faut, en effet, faire attention aux pièges de la corrélation. Pour les deux journées du 8/4 et 19/4 ( Fig.(III)-8 a,b), les observations sont réparties de 0 à 100 %, la valeur de  $r$  est élevée (.94 et .92) et reflète la qualité de l'estimation du sol par le satellite. Pour le 23/4 (Fig.(III)-8 c), au contraire, les mesures sol sont toutes proches de 100 % d'insolation. C'est un jour parfaitement découvert (l'écart type est égal à 6) ; le satellite l'estime bien mais avec peu de nuances (il "sature" à 100 %). Ce sont ces seules nuances que sanctionne le coefficient de corrélation qui apparaît donc comme faible ( $r=0.4$ ). Il faut

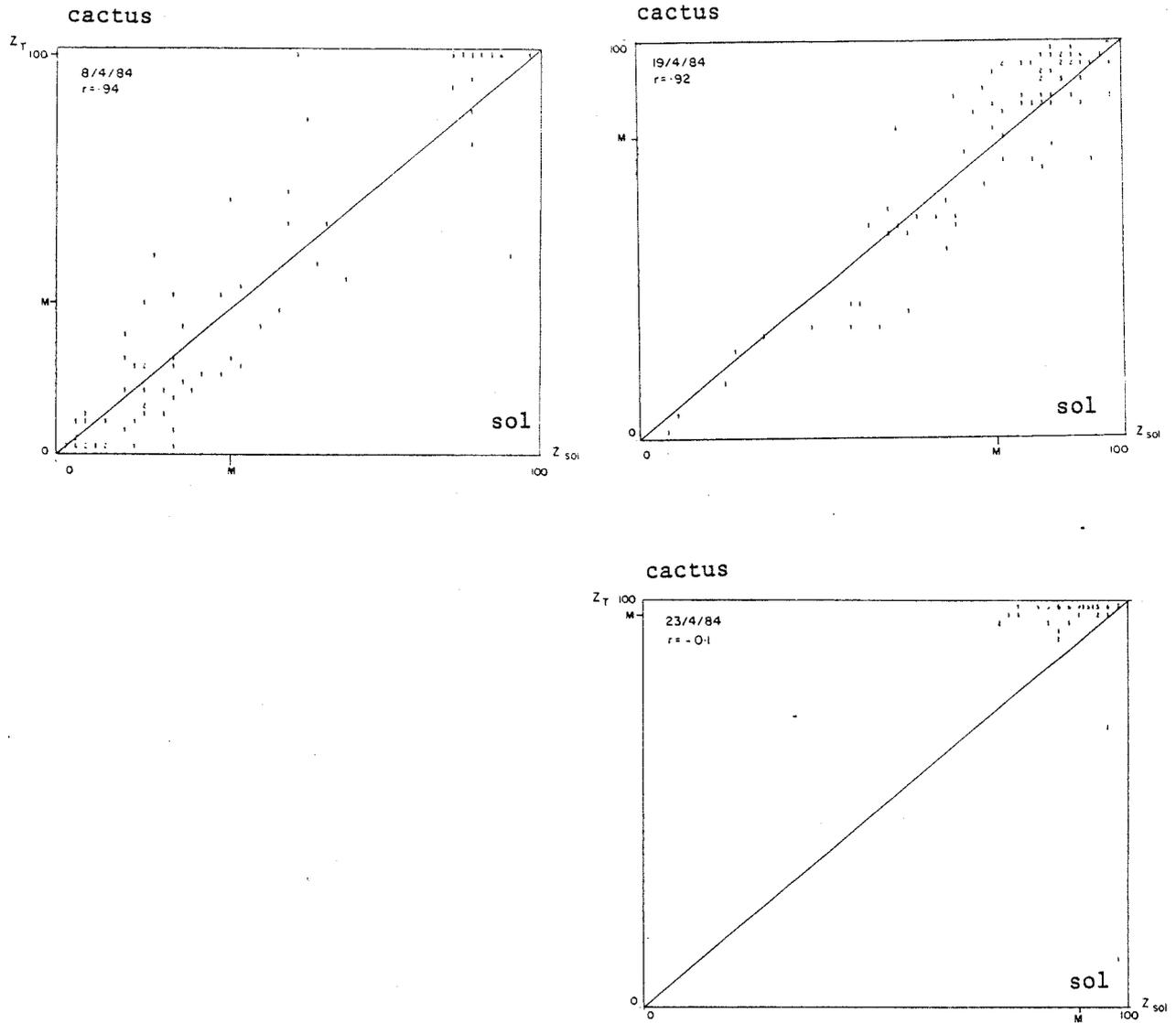


Fig III-8 : Nuages de corrélation entre les données sol et cactus pour trois journées d'avril 1984, illustrant un piège de la corrélation.

donc faire attention lorsqu'on manipule des coefficients de corrélation et garder à l'esprit l'échelle des fluctuations que l'on désire quantifier.

Les courbes d'efficacité des trois périodes (Fig.(III)-9 abc) montrent des coefficients de détermination nettement supérieurs à 0.6, en dehors des journées dont l'écart type est inférieur à 5. Ces journées correspondent toutes à des jours parfaitement dégagés ou complètement couverts, sur toute la France.

Le satellite permet donc d'expliquer environ 80 % de la variance spatiale du rapport d'insolation. Ces performances, en terme de corrélation, sont d'autant meilleures que le champ est plus contrasté.

On peut également remarquer que les performances des mesures satellitaires sont globalement meilleures en été (pour des types de temps convectifs) qu'en hiver, où les passages nuageux sont plus rapides et donc moins bien saisis par l'échantillonnage temporel du satellite.

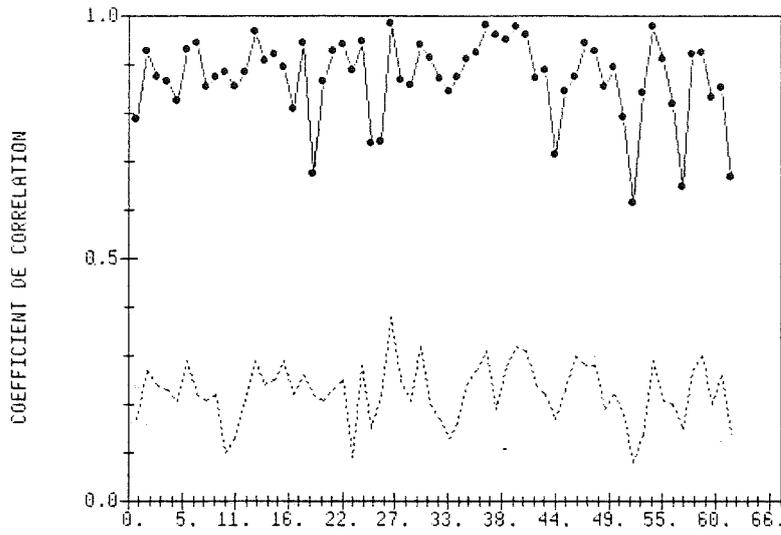
## I-2. Distribution des erreurs en fonction des valeurs de CACTUS

Nous avons déjà observé que les erreurs ne sont pas les mêmes suivant la valeur du rapport d'insolation. Afin de préciser cette liaison, nous avons calculé la corrélation entre le rapport d'insolation estimé par CACTUS  $z_{kT}(t)$  et l'erreur  $d_k(t)$  qui lui est associée en tout point  $t$   $d_k(t) = z_{kT}(t) - z_{kS}(t)$ .

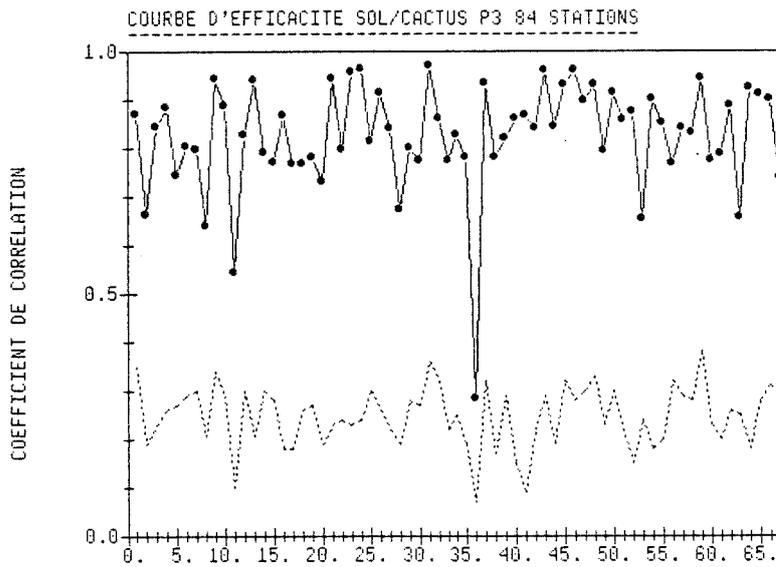
Les nuages de corrélation sont présentés sur la figure (II)-10. Les erreurs sont nécessairement bornées car la variable l'est ; les valeurs de ces bornes sont reportées dans le diagramme. Il faut donc considérer la répartition du nuage à l'intérieur de la zone ainsi définie.

En hiver, le nuage est étalé, la corrélation n'est pas significative ( $r=.32$ ) soit 10 % de variance expliquée. Au printemps et en été, les

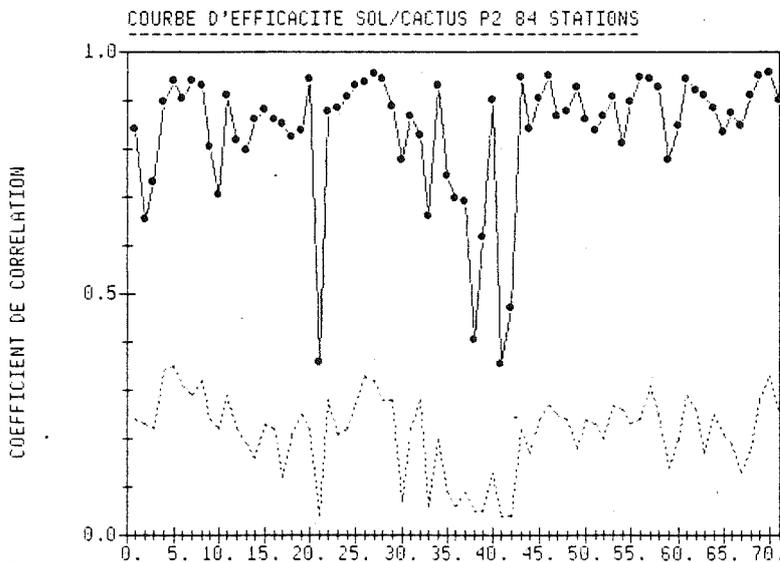
Fig III-9 : Coefficient de corrélation entre les données sol et cactus. Le coefficient de corrélation chute pour les champs d'écart type faible.



HIVER



ETE



PRINTEMPS

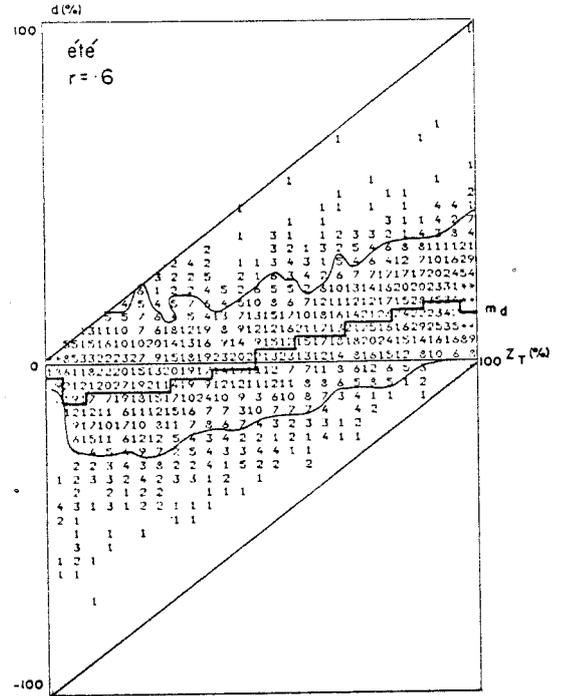
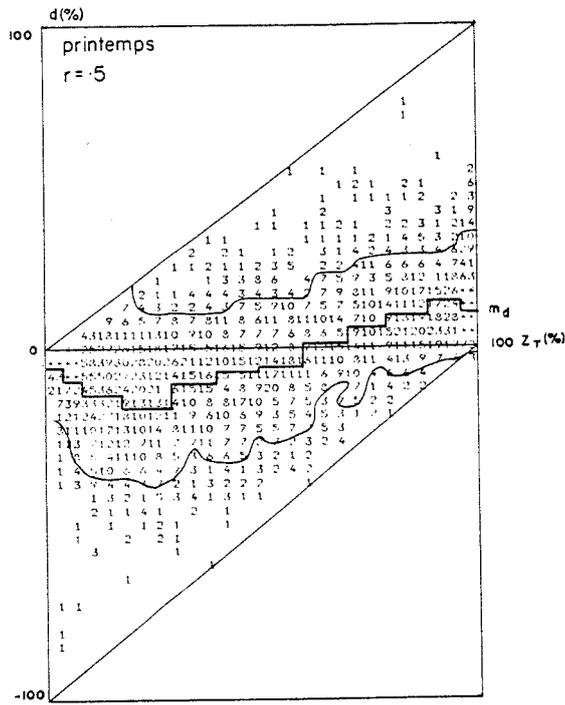
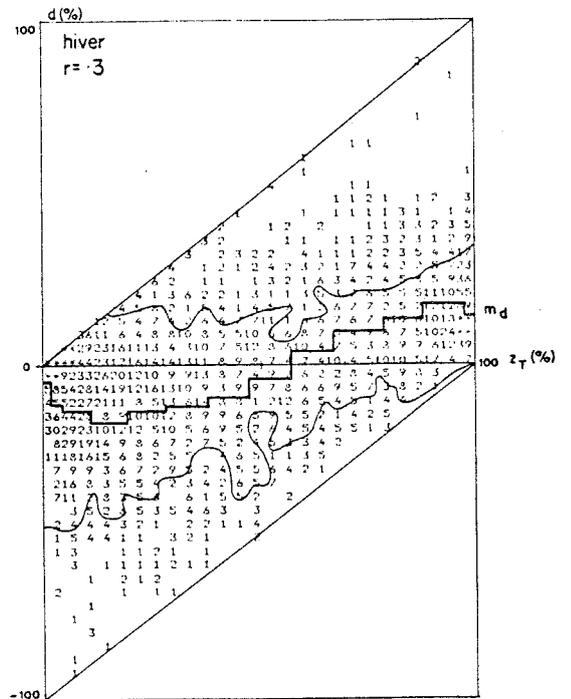


Fig III-10 :

Nuage de corrélation entre les différences  $d = z_t - z_s$ , et les valeurs cactus  $z_t$ . Le domaine autorisé est limité par les deux traits.

$m_d$  : moyenne des différences pour une valeur de télé-détection  $z_t$



erreurs se rapprochent plus de zéro et les coefficients de corrélation atteignent .46 et .58 (ce qui est toujours faible). La forme du domaine autorisée est en partie responsable de ces valeurs non nulles de corrélation. Si le satellite estime le rapport d'insolation à 0 %, son erreur ne pourra être qu'une sous-estimation, et inversement pour 100 %. Les erreurs sont ainsi "liées" aux valeurs estimées.

### I-3. Intégrations des valeurs dans le temps

#### I-3-a. Moyennes temporelles

Une intégration dans le temps des observations permet de travailler sur un phénomène plus lisse et donc plus facile à estimer. Deux pas de temps d'intégration ont été choisis : le mois (en moyenne 20 journées disponibles) et la totalité de la période étudiée (environ 70 jours).

#### - Corrélation des valeurs mensuelles

L'année a été découpée en neuf mois, le neuvième d'entre eux étant composé des mois de septembre, octobre et novembre (7, 8 et 8 journées disponibles respectivement). Ce regroupement, qui totalise 23 jours, est homogène avec le nombre de données des autres mois.

Pour chaque période, on a calculé le coefficient de corrélation défini par p couples (moyenne calculée sol, moyenne satellite). Les valeurs obtenues sont regroupées ci-dessous :

période	r	moyenne (et écart-type)			erreur $E_a$ %
		$r^2$ %	sol %	CACTUS %	
- hiver	.93	86	33 (12)	29 (15)	18
- printemps	.95	91	44 (15)	45 (19)	11
- été	.89	79	55 (10)	58 (13)	9

Tableau III-3

Le fait de moyenner le rapport d'insolation journalier sur une vingtaine de jours permet d'obtenir des coefficients de corrélation allant jusqu'à .95 (91 % de variance expliquée) avec une erreur  $E_a$  normée par la moyenne du champ, de 11 % (alors qu'il était de 25 % pour la même période pour les valeurs journalières du rapport d'insolation). Le satellite permet donc une très bonne estimation du rapport d'insolation mensuel, quelle que soit la période de l'année (cela reste vrai en hiver qui donnait les plus mauvais résultats au pas de temps journalier). Les moyennes ayant une structure plus lisse sont plus faciles à estimer. En outre, il faut noter que les valeurs héliographiques sont beaucoup plus fiables sur une période de temps supérieure à la décade (cf. première partie), ce qui doit également contribuer à expliquer la meilleure relation sol/satellite.

#### - corrélation des valeurs trimestrielles

Nous avons enfin calculé la moyenne trimestrielle du rapport d'insolation en chaque station. Le nuage de corrélation (Fig.(III)-11) regroupe les résultats des trois périodes (p x 3 couples). Le nuage est resserré mais

biaisé (pente 0.71). Les résultats pour chaque période considérée séparément sont détaillés dans le tableau suivant :

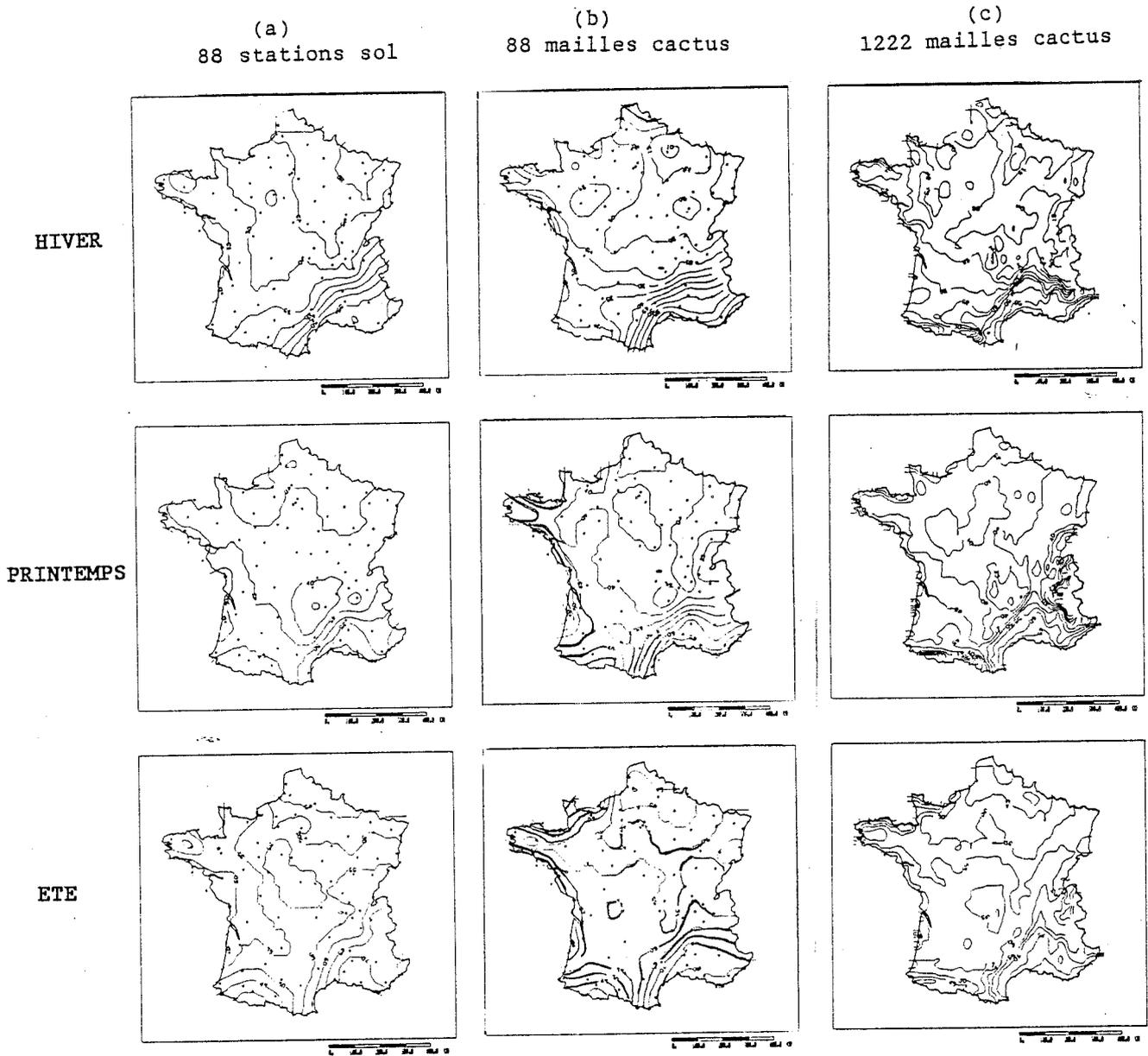
période	r	r <sup>2</sup> %	moyenne		erreur E <sub>a</sub> %
			sol %	CACTUS %	
- hiver	.94	88	32	29	16
- printemps	.87	75	44	45	9
- été	.88	80	55	58	8
Global	.96	92	44	44	10

Tableau III-4

Pour les valeurs trimestrielles du rapport d'insolation, il y a une très bonne corrélation entre les deux types de mesure aux stations test.

Intéressons-nous maintenant aux cartes de la moyenne du rapport d'insolation journalier pour chaque période étudiée. Observons tout d'abord Fig.(III)-12 (a) les cartes tracées à partir des 88 héliographes pour le sol et (b) celles tracées à partir des 88 mailles correspondantes du satellite (on a volontairement limité notre connaissance des mesures satellite à ces seules mailles). Pour chaque période, les cartes des moyennes mesurées par le sol et par le satellite sont tout-à-fait semblables ; les gradients se situent aux mêmes endroits. La structure de la moyenne sol a déjà été discutée (2ème partie II-1).

Fig III-12 : Moyenne climatologique estimée par :



JUILLET

Fig III-13 :

(d'après "Le gisement solaire en France", Météorologie Nationale, 1980).

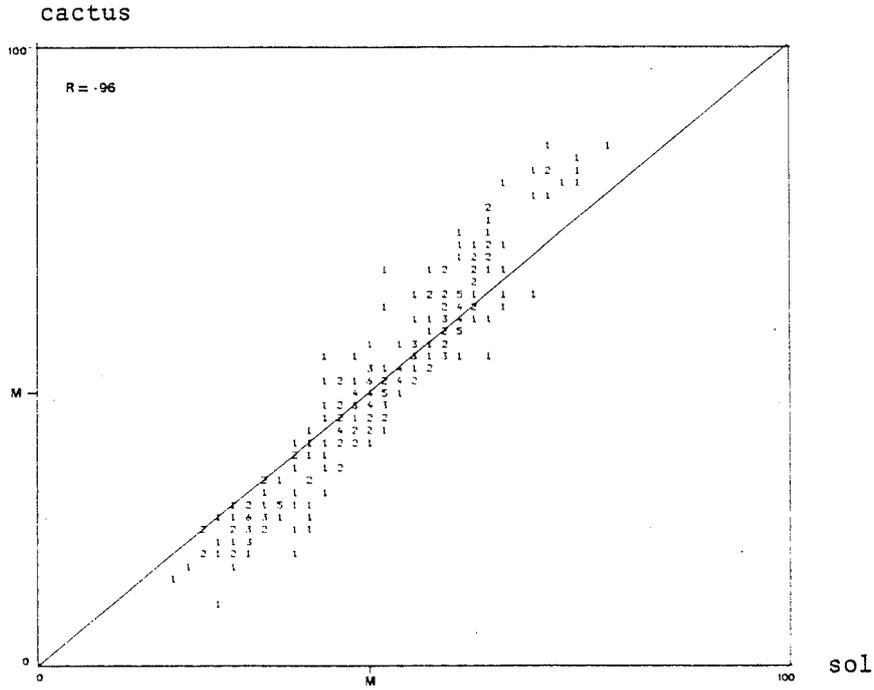
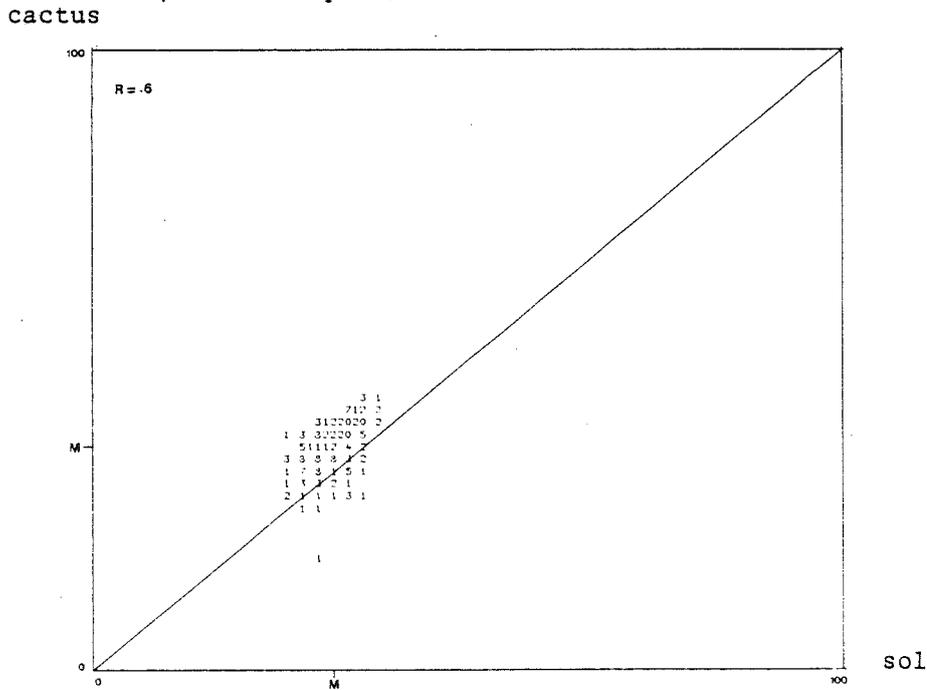


Fig III-11 : Correlation entre les moyennes trimestrielles calculées à partir des mesures sol ou cactus (84 x 3 couples)

Fig III-14 : Correlation entre les écarts types temporels des 3 périodes calculées à partir des mesures sol ou cactus (84 x 3 couples)



Il est intéressant de regarder l'information complémentaire apportée par la très haute résolution spatiale du satellite (1200 mailles couvrent la France) sur les cartes de la valeur moyenne journalière. On a donc calculé la moyenne temporelle en chaque maille et tracé la carte résultante qui ne nécessite aucune hypothèse sur la structure du phénomène (Fig.(III)-12 c).

La comparaison de la moyenne du sol et de celle estimée par le satellite à pleine résolution est très instructive. L'aspect globalement plus variable de la carte du satellite permet de préciser des comportements locaux. Il apparaît ainsi nettement l'influence de la vallée du Rhône, bien établie en météorologie ; cette particularité n'était pas saisie par le réseau sol, trop lâche à l'est du Rhône. De même, le comportement côtier est pris en compte par le satellite (fort gradient du rapport d'insolation moyen sur une distance faible : la côte est beaucoup plus ensoleillée que l'arrière-pays). Le réseau sol, pour sa part, a mesuré une différence climatologique entre Rennes et Brest, mais il ne dispose d'aucune appréciation de la limite de ces deux zones.

Cette appréciation complémentaire est introduite de manière "subjective" par les météorologues ; ils obtiennent ainsi des cartes de moyennes semblables à celles obtenues à l'aide des mesures satellite (Fig.(III)-13), à partir des seules données sol. Cette interpolation subjective n'est possible que pour une zone connue et observée depuis longtemps. Le satellite semble arriver, de manière "objective", aux mêmes résultats.

### I-3-b. Ecart-type temporel

L'écart-type du rapport d'insolation a été calculé, en chaque station et pour chaque période, pour les deux types de mesures : sol et satellite. Le nuage de corrélation résultant (Fig.(III)-14) possède un coefficient de corrélation très faible.

La comparaison de la variabilité du champ des écart-types avec celle du champ des moyennes montre que ce dernier est très peu variable. Tous les points sont regroupés en une très petite zone du graphique. Le coefficient de corrélation ( $r = 0.6$ ) ne quantifie que la co-fluctuation et est donc

faible. Cependant, les deux types de mesure fournissent le même ordre de grandeur pour l'écart-type temporel (de moyenne 31 % pour le sol et 35 % pour le satellite).

L'écart-type ne possède pas de cohérence spatiale (cf. 2ème partie II-1) et les cartes tracées pour les différentes périodes (Fig.(III)-14) sont pépitiqes. En hiver, la présence de la neige sur les zones de montagnes (Alpes, Pyrénées et aussi Massif Central) abaisse localement l'écart-type pour le satellite : il confond la neige et les nuages, toutes les journées lui semblent uniformément "couvertes". Ce gradient local disparaît en été.

### I-3-c. Moyenne et écart-type spatiaux

La moyenne spatiale de chaque évènement k a été estimée à partir des mesures sol  $z_{Sk}$  ou satellitaires  $z_{Tk}$ , en p stations.

$$z_{Sk} = \frac{1}{p} \sum_{i=1}^p z_{kS}(t_i)$$

$$z_{ST} = \frac{1}{p} \sum_{i=1}^p z_{kT}(t_i)$$

Le nuage de corrélation des moyennes estimées par les mesures sol ou satellite (Fig.(III)-15) montre une très bonne corrélation ( $r = .96$ ) et peu de dispersion. Les valeurs faibles sont légèrement sous-estimées alors que les valeurs fortes sont plutôt surestimées par CACTUS. Cet effet avait déjà été remarqué sur les corrélations ponctuelles et s'explique par le phénomène de saturation à 0 et 100 %.

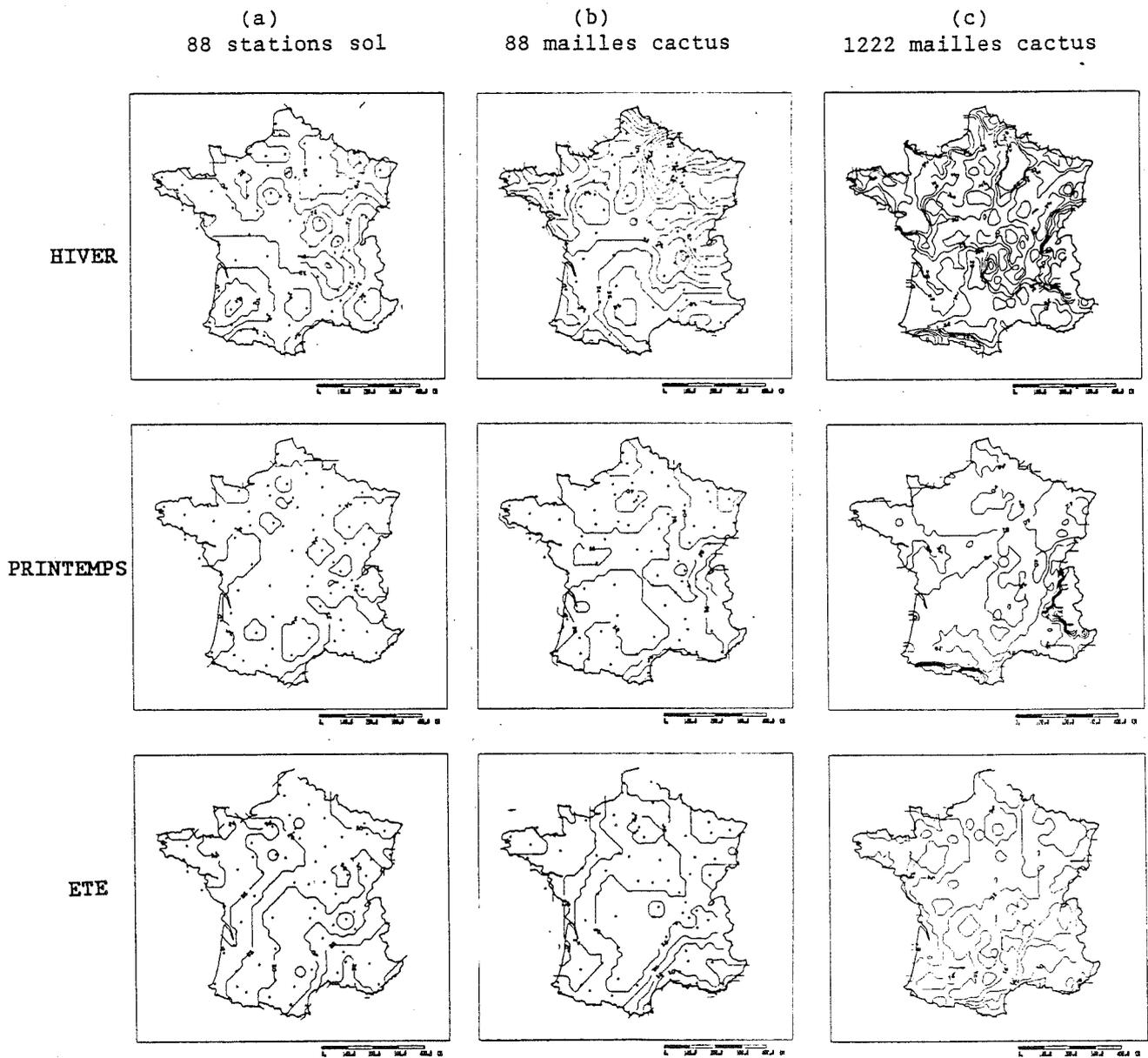
L'écart-type spatial de chaque évènement k a également été estimé

$$T_k = \frac{1}{p} \sum_{i=1}^p (z_k(t_i) - z_k)$$

à partir des mesures sol et satellite. Leur corrélation est de .93 et le nuage (Fig.(III)-16) montre que la variabilité spatiale (calculée sur 84 mailles) de l'image CACTUS est plus forte que celle calculée à partir du sol. Cet effet est à nouveau expliqué par le phénomène de saturation.

Les deux premiers moments spatiaux sont donc très bien estimés par le programme CACTUS.

Fig III-14 : champ des écarts type estimés par :



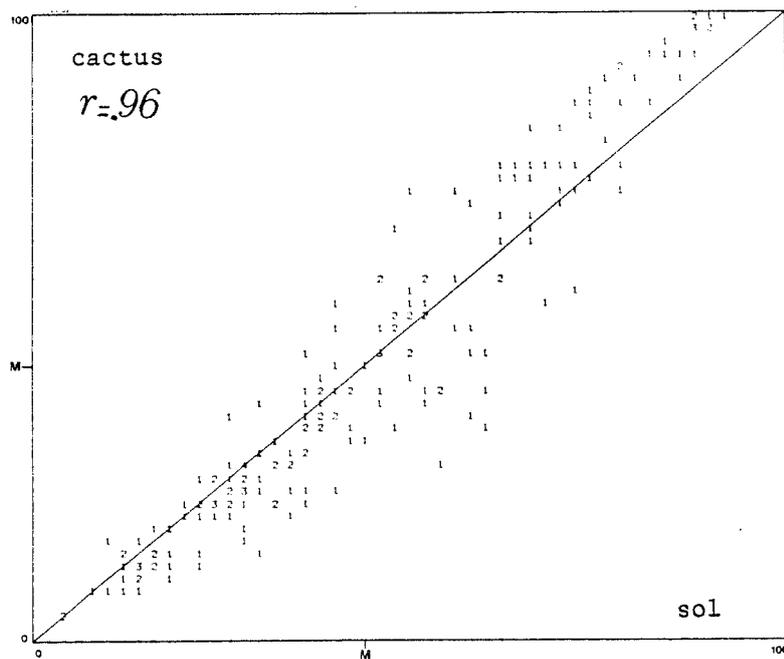
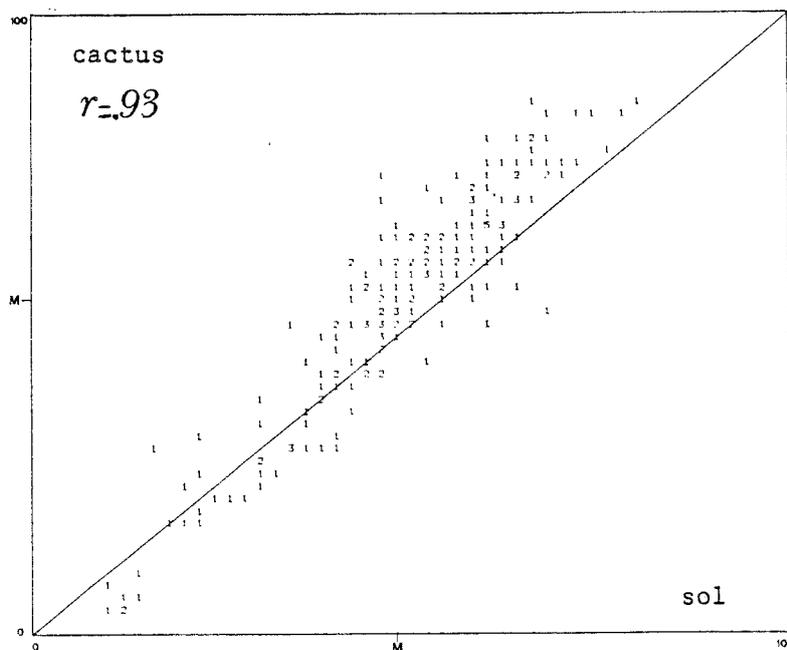


Fig III-15 : Correlation entre les moyennes spatiales estimées par les données sol et cactus pour les trois périodes (200 couples).

Fig III-16 : Correlation entre les écarts type spatiaux estimées par les données sol et cactus pour les trois périodes (200 couples).



## II- PERFORMANCES DES INTERPOLATIONS BASEES SUR LE SOL SEUL

### II-1. But de cette interpolation

Il nous a paru indispensable de situer les performances du satellite par rapport à d'autres méthodes d'estimation de l'insolation. Si cela était réalisable, il serait bien sûr intéressant de comparer la méthode CACTUS à d'autres méthodes utilisant l'information satellitaire. Mais il est aussi nécessaire d'évaluer les performances des techniques d'interpolation qui utilisent la seule information sol, afin de déterminer l'apport de la télédétection par rapport aux moyens classiques de mesure.

Etant donné que l'information sol est utilisée pour établir les valeurs de référence, des précautions doivent être prises pour en assurer l'indépendance avec la source d'information sol destinée à l'interpolation. Pour cela, le réseau sol a été scindé en deux :

- un réseau utilisé pour l'interpolation des données sol
- un réseau destiné à établir les valeurs de référence;

### II-2. Description des méthodes d'interpolation utilisées

De nombreux schémas d'interpolation sont utilisés pour estimer la valeur d'un phénomène naturel tel que l'insolation en tous points d'un domaine d'étude D. Nous en avons sélectionné six, de complexité variable, et qui sont toutes capables de s'appuyer sur un réseau de mesures réparties dans le plan de façon irrégulière.

#### II-2-a. Le plus proche voisin

A tout point  $t$  du domaine D, on attribue la valeur de la station de mesure la plus proche. Cette méthode a pour avantage son extrême simplicité.

### II-2-b. Moyenne arithmétique

On prend pour valeur au point  $t$  non plus une seule mesure, mais la moyenne des  $m$  plus proches stations (ici  $m = 3,5$  et  $10$ ) : on effectue une moyenne glissante.

### II-2-c. Interpolation optimale de GANDIN

L'interpolation décrite par GANDIN, 1965, est optimale au sens où elle minimise la variance de l'erreur d'interpolation. Soit un phénomène  $z$  qui se développe dans l'espace. Ce phénomène est connu à travers une série d'observations (considérées comme les  $n$  réalisations d'une FA.Z), en  $p$  points de mesure  $t_i$ .

L'interpolation de GANDIN reconstitue les écarts à la moyenne climatologique. On calcule en tout point de mesure  $t_i$  :

$$y_k(t_i) = z_k(t_i) - E_z(t_i)$$

Ce changement de variable permet de se ramener à une FA.Y, stationnaire d'ordre 1.

L'estimateur  $y^*$  de la FA.Y en tout point  $t_0$  est choisi sous forme de combinaison linéaire des valeurs mesurées.

$$y_k^*(t_0) = \sum_{i=1}^p \lambda_i y_k(t_i)$$

Les poids  $\lambda_i$  sont calculés de manière à satisfaire la condition d'optimalité :

$$\text{Var}(y_k^*(t_0) - y_k(t_0)) \text{ minimum.}$$

On notera l'erreur  $\epsilon_k(t_0) = y_k^*(t_0) - y_k(t_0) = \sum_{i=0}^p \lambda_i y_k(t_i)$   
 (où  $\lambda_0 = -1$ ).

La variance de l'erreur s'écrit en fonction de la covariance de  $y$  :

$$\text{Var} (\epsilon_k(t_0)) = \sum_{\alpha=0}^p \sum_{\beta=0}^p \lambda_\alpha \lambda_\beta \text{cov}(y(t_\alpha), y(t_\beta))$$

La covariance peut être calculée entre tous les couples de stations ( $\alpha = 1$  à  $p$  et  $\beta = 1$  à  $p$ ), mais n'est pas connu pour les couples de points incluant le point à estimer ( $\alpha = 0$ ), lorsque celui-ci n'est pas un point de mesure. Il est donc nécessaire, pour calculer la variance des erreurs

$\epsilon_k$ , de modéliser les covariances expérimentales. Celle-ci a pour but de lier la covariance à un paramètre simple, caractérisant tout couple de points du domaine d'étude (leur interdistance par exemple).

L'hypothèse de stationnarité d'ordre 2 de la FA.Y est généralement utilisée ; elle permet de modéliser simplement le comportement de la covariance :

$$\text{Cov} (y(t_\alpha), y(t_\beta)) = \text{Cov} (h_{\alpha\beta}) = C_{\alpha\beta}$$

La condition d'optimalité conduit alors à un système de  $p$  équations, où  $C_{ij}$  ne dépend que de la distance.

$$(1) \quad \sum_{j=1}^p \lambda_j C_{ij} = C_{i0} \quad i = 1 \text{ à } p$$

Les poids  $\lambda_i$ , solutions de ce système sont fonction de la forme de la fonction de covariance et de la répartition géographique du réseau autour du point à estimer  $t_0$ . Ils sont les mêmes pour toutes les observations  $k$  (dans la mesure où les stations ont fonctionné pour la série d'observations). Les poids des stations éloignées sont faibles, on opère

généralement en voisinage glissant afin de réduire les temps de calcul. Le fait de travailler en voisinage restreint allège la contrainte de stationnarité d'ordre 2 : il lui suffit alors d'être vérifiée localement. L'hypothèse de stationnarité d'ordre 2, a priori lourde, est alors réaliste. Le rapport d'insolation la vérifie (cf. 2ème partie II).

Une propriété très puissante de l'interpolation de GANDIN est la connaissance en tout point  $t_0$  de la variance de l'erreur. En effet, lorsque les poids  $\lambda_i$  vérifient le système (1), il vient :

$$\text{Var}(\hat{\varepsilon}_k(t_0)) = - \sum_{i=0}^p \lambda_i c_{i0}$$

La variance ne dépend que du lieu  $t_0$  et de la structure spatiale, non de la valeur observée. Cette propriété peut être utilisée pour renforcer un réseau de mesure (cf. 2ème partie IV) : on choisira d'implanter une nouvelle station au lieu où la variance de l'erreur est maximale. La carte de variance de l'erreur d'estimation complète toute carte calculée par estimation optimale en indiquant son degré de précision.

Une fois  $y_k^*(t_0)$  calculée, la reconstitution de la valeur brute du phénomène  $z_k(t_0)$  nécessite la connaissance de la moyenne climatologique au point  $t_0$ . Si celle-ci ne peut être connue de façon exogène, il faudra l'interpoler à partir des moyennes climatologiques estimées aux  $p$  points de mesure  $t_i$ . Cette étape complémentaire va augmenter l'erreur de reconstitution. Toutefois, l'erreur sur l'estimation de la moyenne reste a priori faible devant celui des valeurs centrées  $y$  : sa portée est plus grande, l'amplitude des fluctuations moindre (voir Fig.(III)- 7 ).

On pourrait penser à former directement un estimateur de  $z$ , mais il faudrait alors lui appliquer une contrainte de non biais :

$$E(z_k^*(t) - z_k(t)) = 0$$

qui augmente la variance d'estimation car celle-ci est alors minimisée sous contrainte (de non biais). C'est pourquoi, on a travaillé sur les

valeurs centrées auxquelles on ajoute la valeur de la moyenne estimée par krigeage.

D'un point de vue formel, on peut montrer (O. DUBRULE, 1981) que l'interpolation spline est équivalente à l'interpolation optimale de GANDIN, sous certaine condition de non biais et pour une fonction de structure choisie à priori ( $C(h) = h^2 \text{Log } h$ ). Cette fonction de structure a pour avantage l'esthétique des cartes produites mais n'est pas toujours optimale au sens de la minimisation de la variance de l'erreur.

On remarquera pour finir que le jeu de poids  $\lambda_i$ , calculé pour tout point  $t_0$  ne dépend pas de l'évènement  $k$  considéré, et peut ainsi être réutilisé (si les mêmes stations ont fonctionné pour tous les événements). L'interpolation d'un évènement  $k'$  est donc assez rapide.

#### II-2-d. Krigeage

Cette méthode, comme l'interpolation de GANDIN, est optimale au sens où elle minimise la variance de l'erreur d'estimation. Développée dans le cadre des sciences de la terre, elle s'applique aux phénomènes connus à travers une réalisation unique. Ce contexte, que l'on appelle "spatial" par opposition au cas "climatologique" traité par GANDIN, pose le problème de l'inférence de la structure. Son développement est décrit par DELHOMME, 1976. La démarche suivie est similaire à celle de GANDIN : l'estimateur est une combinaison linéaire des données disponibles

$$z^*(t_0) = \sum_{i=1}^p \lambda_i z(t_i)$$

Les poids  $\lambda_i$  sont déterminés par les deux conditions

- optimalité :  $\text{Var} (z^*(t_0) - z(t_0))$  minimum
- non biais :  $E (z^*(t_0) - z(t_0)) = 0$

La condition de non biais n'était pas nécessaire dans le contexte climatologique puisqu'on travaillait sur les valeurs centrées. La connaissance du champs des moyennes étant impossible dans le contexte spatial, la recherche de l'optimalité se fait sous contrainte de non biais.

Le calcul de la variance de l'erreur implique la connaissance de la structure spatiale. La variabilité spatiale de la moyenne (ou dérive) biaise l'estimation de la fonction de covariance. Le krigeage permet d'élaborer, suivant l'hypothèse faite sur la dérive, une fonction de structure capable de filtrer son influence (le variogramme dans le cas d'une dérive constante).

#### II-2-f. ACP de processus

La reconstitution par ACP procède d'une approche différente de l'analyse objective. Elle ne nécessite pas d'hypothèse d'homogénéité de la structure, car elle ne modélise pas la corrélation mais la décompose en harmoniques  $F_1(t)$  appelés cosinus directeurs (cf. 2ème partie III), de puissance associée  $\lambda_1$ .

$$R(t, t') = \sum_1 \lambda_1 F_1(t) F_1(t')$$

Tout processus centré et réduit  $y_k(t)$  se décompose sur les cosinus directeurs :

$$y_k(t) = \sum_1 U_{1k} F_1(t)$$

Le score  $U_{1k}$  ne dépend que de l'évènement  $k$  et de la composante  $1$  ; il est indépendant du lieu.

Le cosinus directeur  $F_1(t)$  est indépendant de l'évènement. Il est approché aux stations de mesures  $t_i$  par  $f_{i1}$ , le cosinus directeur de la matrice  $[R]$  .  $[E]$  . Ces valeurs  $f_1(t_i)$  sont interpolées par krigeage, pour obtenir l'estimateur  $y^*$  ( à partir des  $M$  premières composantes).

$$y^*_{Mk}(t_0) = \sum_{l=1}^M U_{lk} f^*_l(t_0)$$

Si les cosinus directeurs  $f_1(t)$  sont plus lisses que le phénomène (c'est le cas pour les cinq premiers), alors l'erreur d'interpolation sur les cosinus directeurs est plus faible que celle sur le phénomène  $y$ .

On montre (OBLED et CREUTIN, 1986) qu'un tel estimateur peut s'exprimer comme une combinaison linéaire des données disponibles  $y_k(t_i)$ . Le système ainsi défini est similaire à celui de l'interpolation de GANDIN. Seule la façon de prendre en compte la structure diffère: l'hypothèse d'homogénéité pour GANDIN est remplacée par une décomposition orthogonale pour l'ACP. Le fait de n'utiliser qu'un nombre restreint  $M$  de composantes, lisse les corrélations. Le nombre de composantes à retenir se pose à nouveau.

L'estimateur ainsi déterminé nous fournit des valeurs centrées réduites que l'on transforme pour obtenir un estimateur des valeurs brutes  $z^*_{kM}(t_0)$ .

$$z^*_{kM}(t_0) = m^*(t_0) + \sigma^*(t_0) \left( \sum_{l=1}^M U_{lk} f^*_l(t_0) \right)$$

Comme pour l'interpolation optimale, l'erreur d'estimation de la moyenne reste faible devant celle des valeurs centrées  $y^*_{kM}(t_0)$  : sa portée est plus grande (600 km). L'écart-type pour sa part est peu structuré mais pratiquement constant ; on l'estime par sa valeur moyenne sur les  $p$  stations. Les différentes étapes de la reconstitution du rapport d'insolation se résume ainsi :

- interpolation de la moyenne et de l'écart-type
- interpolation des  $M$  premiers cosinus directeurs  $f^*_l(t)$

- calcul des M scores  $U_{1k}$

$$U_{1k} = \sum_i y_k(t_i) f_1(t_i)$$

- addition des différentes surfaces de cosinus directeurs pondérées par les scores.

Par conséquent, la reconstitution de tout nouvel évènement  $k'$  ne nécessite que les deux dernières étapes, si les cosinus directeurs ont déjà été interpolés. Ces deux étapes sont très rapides. La reconstitution par ACP présente l'avantage d'éviter l'hypothèse d'homogénéité de la structure du phénomène  $y$ . Cette hypothèse est, par contre, faite sur les cosinus directeurs  $f_1(t)$ . Les cosinus directeurs du rapport d'insolation vérifient cette condition (cf. 2ème partie III), du moins les premiers.

#### II-2-g. Conclusion

La première méthode d'interpolation par moyenne glissante est d'une extrême simplicité et sera notre référence "plancher": toute méthode moins performante est à éliminer.

Le krigeage et l'interpolation de GANDIN sont des méthodes d'interpolation optimale dans le sens où elles utilisent la structure spatiale du champ pour l'interpoler.

L'ACP ne fait pas d'hypothèse d'homogénéité sur la structure du phénomène mais sur celle de ses cosinus directeurs. Le fait de tronquer la décomposition, en ne gardant que M composantes, en fait une méthode de lissage.

Regardons maintenant les résultats obtenus pour le rapport d'insolation en France.

### II-3. Performances des différentes méthodes d'estimation sol

Nous avons évalué les performances de chaque méthode d'estimation du rapport d'insolation à partir des seules données sol. Les performances sont évaluées aux noeuds d'un réseau test de 61 héliographes (Fig.(III)-17a) qui n'ont pas servi pour les interpolations. Les interpolations sol ont été faites à partir d'un autre réseau, de 23 héliographes (Fig.(III)-17b) ou un sous réseau de celui-ci réduit à 5 stations.

Les performances sont appréciées sur les courbes d'efficacité, en terme de coefficient de détermination (coefficient de corrélation au carré) qui exprime le pourcentage de variance expliqué. L'échelle est donc directement interprétable (ces coefficients  $r^2$  sont toujours plus faibles que  $r$ ). Nous avons utilisé les courbes d'efficacité "classées", c'est-à-dire que les évènements sont rangés par ordre de coefficient de détermination croissant. Ces courbes sont une estimation de la loi de répartition de  $r^2$ . Le classement permet une lecture plus aisée des courbes et une comparaison immédiate des méthodes mais présente l'inconvénient de ne pouvoir retrouver les performances des différentes méthodes pour un jour donné. C'est pourquoi, on a également tracé les courbes d'efficacité par ordre chronologique. Ces courbes, très chargées, et peu lisibles car entrecroisées, n'ont cependant pas été toutes reproduites ici.

Nous allons d'abord évaluer chaque méthode d'estimation à l'aide des seules données sol pour ensuite comparer les meilleures d'entre elles au programme CACTUS (au chapitre suivant).

NB : De nombreuses méthodes fournissent des estimations négatives en un nombre limité de points. Le rapport d'insolation est par définition compris entre 0 et 100 %. Devant cette apparente contradiction, nous avons décidé de ne pas borner artificiellement à 0 les estimations (ce qui aurait des effets différents sur chaque méthode, suivant la façon dont elles peuvent s'en éloigner), ni de négliger ces valeurs négatives (-1 n'est pas une valeur fautive), mais de les considérer telles quelles dans la corrélation : les valeurs corrélées vont de -10 à 100 % pour les estimations (0 à 100 % pour les valeurs de référence). Nous n'avons pas rencontré de valeurs estimées inférieures à -10%.

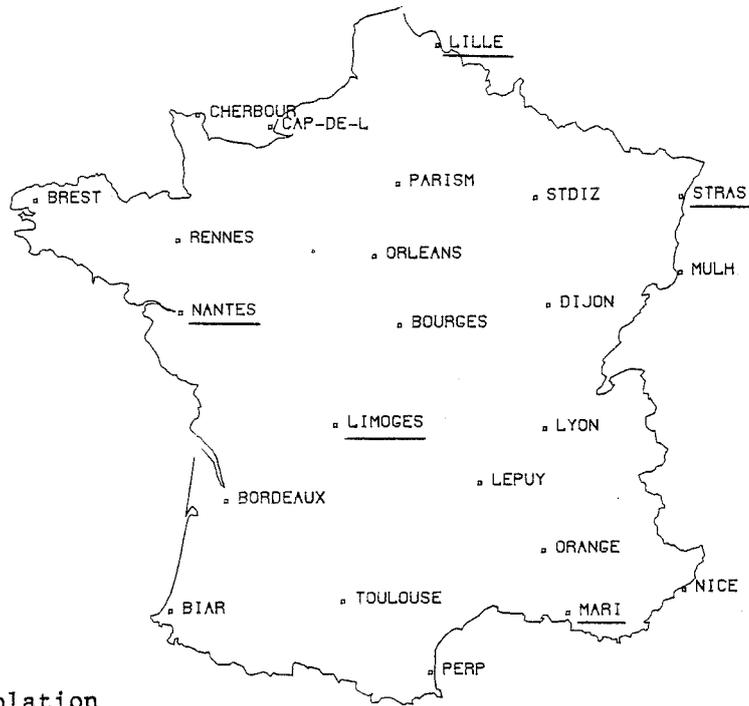
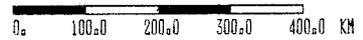


Fig III-17 :

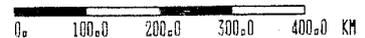
(a)

Réseau d'interpolation  
de 23 et 5 (souligné)  
héliographes.



(b)

Réseau test comp osé  
de 63 héliographes.



### III-3-a. Moyenne arithmétique

Nous avons estimé la valeur aux 61 points test à l'aide d'un nombre de voisins variant de 1 à 23 (totalité des stations disponibles pour l'interpolation) - ce dernier cas revenant à estimer tout point du domaine par la moyenne spatiale (estimée par 23 héliographes).

#### Estimation par la moyenne spatiale

L'estimation ainsi faite étant constante sur le champ, le coefficient de détermination qui qualifie sa co-fluctuation avec les 61 stations sol de référence est nul. Le coefficient de détermination a été calculé sur l'ensemble des 71 évènements ; il indique 52 % de variance expliquée. La variabilité inter-épisode explique donc plus de la moitié de la variance du rapport d'insolation des 71 évènements considérés. Toute autre méthode doit au moins faire mieux. Nous avons déjà constaté cette importance de la moyenne dans l'analyse de la structure, où la première composante qui correspond à la moyenne spatiale avait une valeur propre associée de 55 %.

#### Estimation avec N voisins

Lorsqu'on augmente le nombre de voisins, la corrélation croît dans un premier temps, puis diminue à mesure qu'on se rapproche de la moyenne spatiale qui est uniforme pour l'évènement. Pour le rapport d'insolation et un réseau de 23 héliographes, le nombre optimal se situe aux alentours de trois voisins (Fig.(III)-18). Il est directement lié à la portée du phénomène considéré ; les stations situées au-delà de la portée n'apportent plus d'information en terme de variabilité, mais tendent à lisser (en la moyennant) l'estimation.

Cette méthode fournit des résultats corrects en regard de son extrême simplicité.

### II-3-b. Interpolation optimale

La fonction de structure utilisée ici est la fonction de covariance, elle a été estimée à l'aide des données centrées spatialement. L'interpolation se fait en deux étapes :

- estimation des valeurs centrées par la moyenne climatologique
- addition de la moyenne climatologique, interpolée par krigeage.

Nous avons effectué, à partir du réseau complet de 23 héliographes, une interpolation en voisinage glissant avec 5 voisins (les autres ont des poids négligeables).

Nous avons également effectué une interpolation à l'aide du sous réseau de 5 héliographes, en voisinage unique. Les moyennes climatologiques en tout point du domaine sont alors estimées à partir de ces seules 5 stations.

On a quantifié (Fig.(III)-19) la sensibilité de l'interpolation de GANDIN à la densité du réseau disponible. Lorsqu'on passe de 5 à 23 stations, on explique globalement 30 % de variance en plus. Le premier réseau est nettement insuffisant pour obtenir une estimation correcte du rapport d'insolation : il est trop lâche par rapport à la portée du phénomène. Le deuxième, au contraire, est satisfaisant : mis à part 10 évènements (qui correspondent, après vérification, à des situations peu contrastées), il explique nettement plus de 50 % de la variance. De façon globale, pour les 71 évènements, le coefficient de détermination vaut .86.

### II-3-c. ACP de processus

Nous avons reconstitué la valeur au point test à l'aide d'un nombre croissant de composantes. On a pour cela utilisé le réseau de 23 héliographes. A partir duquel on a calculé les scores (indépendants du lieu), et les cosinus directeurs (interpolés par krigeage) puis on a empilé M surfaces de cosinus directeurs (pondérées par les scores correspondants), pour reconstituer les valeurs centrées réduites, et enfin on a multiplié par l'écart-type et ajouté la moyenne estimée par krigeage pour atteindre le rapport d'insolation.

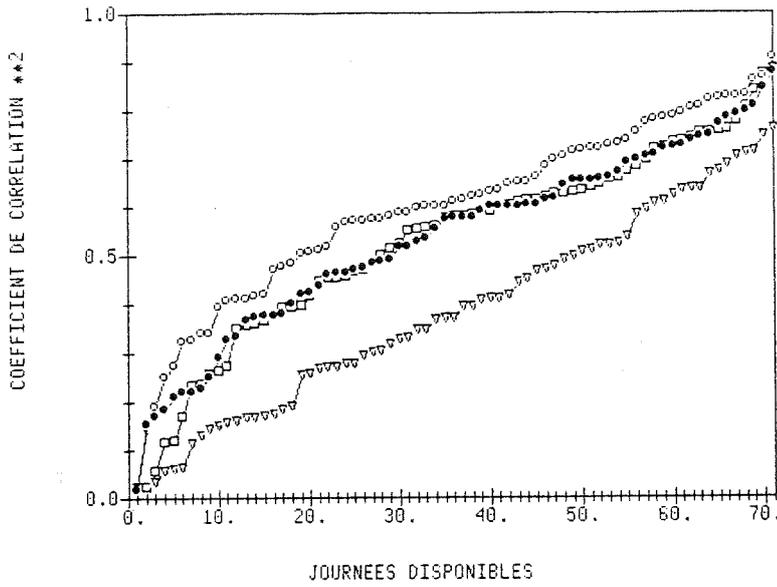


Fig III-18 :

MOYENNE GLISSANTE  
nombre de voisins  
variable

- 1
- 3
- 5
- ▽ 10

(réseau de 23  
héliographes)

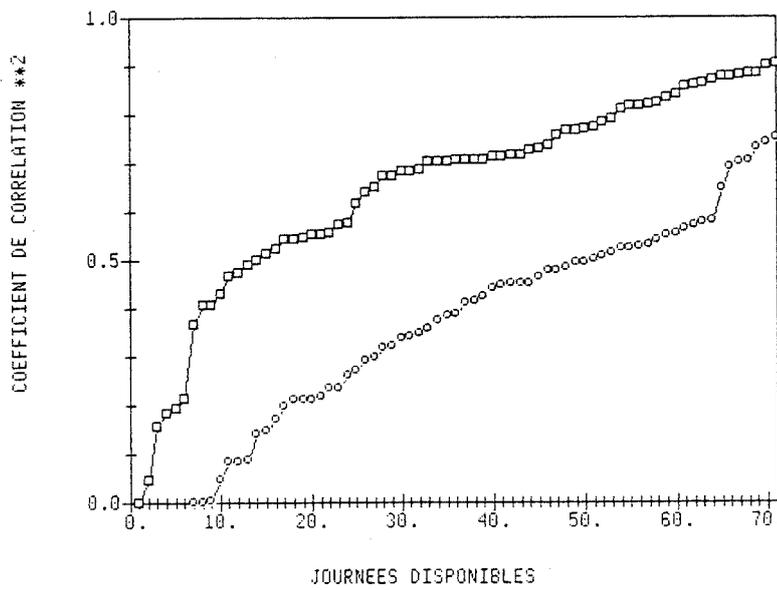


Fig III-19 :

INTERPOLATION OPTIMALE  
avec un réseau de :

- 5 héliographes
- 23 héliographes

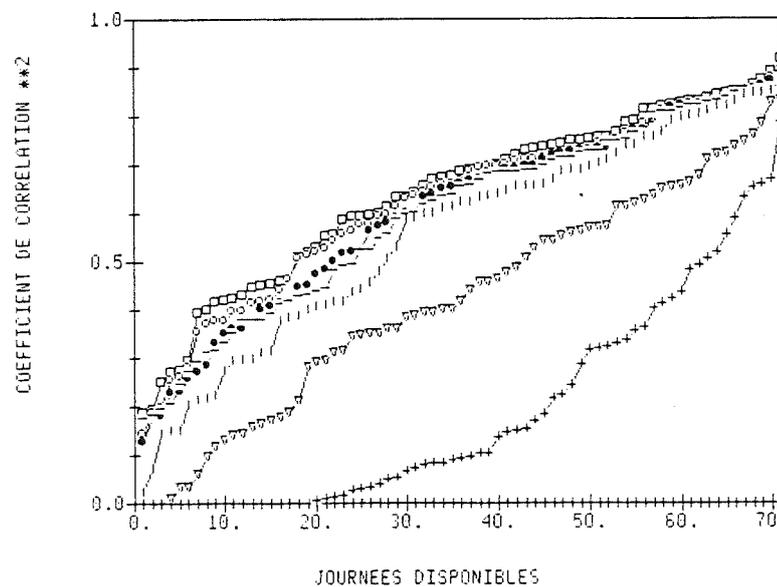


Fig III-20 :

INTERPOLATION PAR ACP  
nombre de composantes  
retenues :

- + 1
- ▽ 3
- | 7
- 8
- 9
- 10
- 14

(réseau de 23  
héliographes)

Le coefficient de détermination (calculé sur les 61 stations test), pour la reconstitution avec M composantes, tend vers la somme des valeurs propres correspondantes :

$$\frac{1}{p} \sum_{i=1}^M \lambda_i$$

Cette somme est cependant toujours supérieure à la valeur empirique du coefficient de détermination, car la valeur du cosinus au point test est interpolée (et non plus exacte) et de plus la moyenne et l'écart-type sont eux aussi interpolés.

On remarque la grande stabilité des valeurs des  $\sum \lambda_i$  (puissance des cosinus directeurs) quel que soit le nombre de stations utilisé pour les déterminer.

Lorsque M augmente, l'écart entre le coefficient de détermination et les  $\sum \lambda_i$  croît. Les cosinus directeurs d'ordre élevé n'apportent pas toute l'information qu'on attend d'eux. Non seulement, ils sont plus difficiles à interpoler car leur portée est plus courte mais, en outre, ils reflètent les fluctuations particulières des mesures des 23 héliographes pendant la période utilisée pour les calculer, plutôt qu'une réalité propre du phénomène. Ajouter une surface n'ajoute plus une information mais un bruit. Le pourcentage de variance expliquée n'augmente plus.

nombre de composantes retenues = M	1	3	7	8	9	10	14
coefficient de détermination global *	56	75	82	84	84	85	86
$\frac{1}{23} \sum_{i=1}^M \lambda_i$ (avec 23 stations) *	53	76	89	91	92	93	97
$\frac{1}{88} \sum_{i=1}^M \lambda_i$ (avec 88 stations) *	55	77	88	90	91	92	95

\* résultats en pourcentage.

L'apport d'une composante supplémentaire dans la reconstitution est quantifié (Fig.(III)-20). Au-delà de huit composantes, l'apport est faible en terme de coefficient de performance journalier. Ce nombre correspond à celui choisi comme le nombre de composantes comportant une information spatiale significative (sept au printemps). On retiendra qu'avec huit composantes, on capte globalement 85 % (au mieux 90 %) de la variance du rapport d'insolation. L'interpolation des huit cosinus directeurs et des deux moments expliquent cette différence de 5 % qui reste faible.

### III - COMPARAISON DES PERFORMANCES DES DIFFERENTES ESTIMATIONS SUR UN JEU DE DONNEES INDEPENDANTES.

Il s'agit, dans ce dernier chapitre, de comparer les performances des différentes estimations étudiées précédemment, aussi bien celles basées sur les mesures sol que satellitaires. Après la présentation des résultats, on les illustrera par un exemple avant de conclure.

#### III-1. Les courbes d'efficacité .

On a porté sur le même graphe les courbes d'efficacité de chacune des méthodes d'estimation utilisées, à partir du sol et du satellite (Fig.(III)-21). Les performances de la méthode CACTUS sont supérieures à celles des méthodes d'estimation basées sur un réseau d'héliographes, de 5 comme de 23 héliographes, quelle que soit la période. Parmi les méthodes utilisant seulement les données sol, il vient, par ordre décroissant, le krigeage et l'ACP, de performance comparable, puis la moyenne arithmétique.

Le programme CACTUS est donc plus performant en terme de cofluctuation, malgré le fait qu'il ne travaille que sur une mesure indirecte ; ceci est dû à sa bonne appréhension spatiale du phénomène, ce qui lui permet de saisir les fluctuations locales. Un facteur décisif dans la qualité de l'estimation d'une méthode est donc la densité de mesure disponible en regard de la portée de la variable analysée.

Si l'interdistance entre stations est supérieure à la distance de décorrélation du phénomène, alors il existe des zones où le réseau n'explique plus la variabilité du phénomène. Le réseau test de 61 héliographes est précisément placé dans ces zones. Les méthodes basées sur les interpolations des mesures sol, sont donc désavantagées par rapport à la méthode satellitaire, par notre réseau test. D'autre part, le critère de corrélation ne sanctionne que la cofluctuation spatiale. Il est donc logique que les performances du réseau de 5 héliographes (interdistance moyenne de 500 km, et donc supérieure a la portée) soient médiocres. Le réseau de 23 héliographes est déjà beaucoup plus performant.

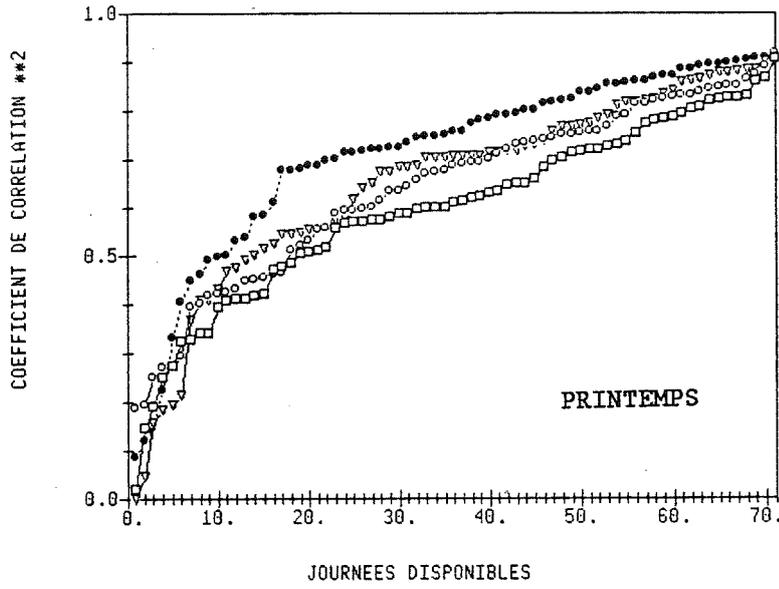
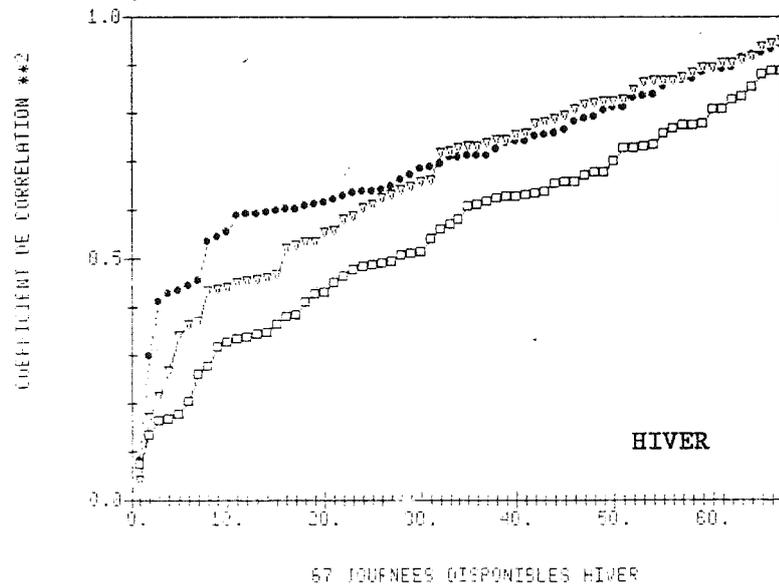
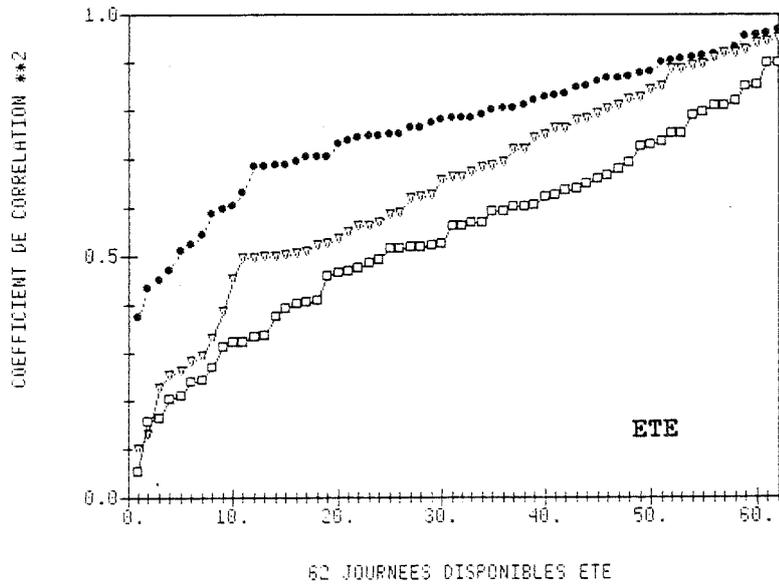


Fig III-21 :

Comparaison de différentes méthodes d'interpolation

- cactus
- ▽ gandin
- acp
- 1 voisin



Etant donné la fonction de structure, un réseau plus dense amènerait des résultats encore meilleurs. Mais il devient alors difficile de réaliser un test, car augmenter le nombre de stations interpolées revient à diminuer le nombre de stations test et donc à dégrader la robustesse de notre test. D'autre part, une augmentation du réseau d'interpolation augmente le temps de calcul, même en voisinage glissant. Cependant, il existe certainement une densité de réseau sol qui fournisse une estimation équivalente à celle de l'image CACTUS.

### III-2. Tableau de performance

Les courbes d'efficacité caractérisent uniquement la cofluctuation. Dans le tableau ci-contre, nous avons noté différents paramètres afin de compléter la comparaison des différentes méthodes d'estimation. On note que la pente et l'origine de la droite de régression sont proches respectivement de un et de zéro sauf pour la méthode CACTUS (respectivement .8 et 9) et la méthode du plus proche voisin (respectivement .9 et 4). Au contraire, pour l'ACP, les coefficients de la droite de régression sont parfaitement égaux à un et zéro. Il n'y a pas de correction à apporter à ces données. L'ACP reconstitue des valeurs non biaisées. L'erreur absolue normée par la moyenne  $E_a$  n'apporte pas une grande information par rapport à  $r$ . On a enfin porté l'écart-type d'estimation moyen sur le champ, connu pour l'interpolation optimale de GANDIN.

### III-3. Remarques sur les courbes d'efficacité

1 - Le coefficient de corrélation calculé sur un champ chute lorsque ce champ a une faible variabilité spatiale. Par exemple, s'il s'agit d'un jour uniforme de beau temps, même une reconstitution à 5 % d'erreur aléatoire (ce qui serait un exploit car c'est de l'ordre de l'erreur de mesure) donnerait une corrélation très faible, si la variabilité spatiale est du même ordre. Pour la période considérée nous avons une série de 10

## TABLEAU DE PERFORMANCE

Estimation pour 71 évènements (printemps) testés en 61 stations

Valeurs de référence : 61 héliographes

moyenne du rapport d'insolation = 44

variance = 1127

$$E_a = \frac{1}{n} \sum_{k=1}^n \frac{1}{p} \sum_{i=1}^p \frac{|z_k(t_i) - \bar{z}_k(t_i)|}{\bar{z}_k} \quad \begin{array}{l} n = 71 \\ p = 61 \end{array}$$

méthode utilisée	r		pente origine		valeurs reconstituées		E <sub>a</sub>	écart-type d'estimation
	r	r <sup>2</sup>	régré.	régré.	moyenne	variance		
	r	r <sup>2</sup>	r	r	r	r <sup>2</sup>	r	r
CACTUS	.92	85	.79	9	44	1545	25	
1 VOISIN	.89	79	.89	4	45	1119	24	35
3 "	.91	83	1.00	-1	44	930	22	28
5 "	.89	80	1.04	-2	44	835	25	27
10 "	.84	70	1.05	-3	44	726	32	26
23 "	.72	52	1.02	-2	45	564	42	25
Gandin 23	.93	86	1.00	1	43	977	19	9
Gandin 5	.81	65	.91	2	46	889	34	6
ACP 1	.75	56	.99	0	44	639	40	
3	.87	75	1.00	0	44	840	29	
7	.91	82	1.00	0	44	923	23	
8	.92	84	1.00	0	44	937	22	
9	.92	85	1.00	0	44	944	22	
10	.92	85	1.00	0	44	948	21	
14	.93	86	1.00	0	44	961	21	

journées très ensoleillées d'écart-type inférieur à 10. Après avoir éliminé ces 10 journées de notre jeu de données, nous avons calculé à nouveau les courbes d'efficacité classées. On n'observe qu'un simple décalage d'origine. Ces 10 journées, de faible variabilité spatiale, correspondent aux 10 moins bonnes performances, pour toutes les méthodes. Le classement de celles-ci n'est pas modifié.

2 - Les différentes méthodes que nous avons employées ont toutes été classées d'après leurs courbes d'efficacité classées. Elles peuvent être considérées comme les fonctions de répartition de la variable "coefficient de détermination". Il vient alors deux questions sur la robustesse statistique du classement entre les méthodes ainsi départagées.

•quel est l'écart-type d'estimation des coefficients de corrélation ?

•quel est l'écart-type sur les quantiles de la fonction de répartition ?

Le premier point est lié au nombre de stations du réseau test. La densité de notre réseau test (interdistance moyenne de 80 km) assure une robustesse statistique considérée comme suffisante, au regard de la portée du phénomène (350 km).

Le deuxième est lié au nombre d'évènements sur lesquels on a estimé la fonction de répartition. Plutôt que de calculer l'écart-type sur les quantiles, qui suppose connue la loi de répartition de  $r^2$ , on a préféré tester la similarité des distributions. Ce qui nous intéresse, c'est de déterminer si deux distributions de  $r^2$  sont significativement différentes. Pour cela, on a utilisé le test du  $\chi^2$ .

L'histogramme sur dix classes d'effectifs environ constants, pour les méthodes CACTUS et GANDIN par exemple donne un  $\chi^2 = 14.98$ , ce qui nous permet d'affirmer que ces deux méthodes ont des performances différentes avec un risque d'erreur de 10 %. Les courbes d'efficacité indiquent que la méthode CACTUS est supérieure à la méthode GANDIN, avec 90% de chances.

Le risque est identique entre l'ACP et CACTUS. Il est relativement élevé, et la simple observation de la figure (III)-21 ne nous l'aurait pas laissé supposer. Toutes les autres estimations à partir du sol seul ont

des performances inférieures et sont, elles, significativement différentes.

Si deux distributions ne sont pas significativement différentes, la position relative des courbes d'efficacité correspondantes peut induire en erreur, car elle ne permet pas de conclure sur les performances relatives des différentes méthodes testées. Par exemple, sur un échantillon de 10 journées en avril 1983, le classement entre les méthodes CACTUS et GANDIN semblait inversé: L'échantillon, trop réduit, ne nous permettait pas de tirer de conclusion : Les distributions n'étaient pas significativement différentes. Pour soixante-dix événements, les conclusions sur le classement des courbes d'efficacité deviennent possibles .

#### III-4. Illustration pour un jour de printemps

Afin d'apprécier de façon plus concrète les particularités des différentes méthodes d'interpolation, on a tracé les cartes calculées par chacune d'entre elles pour une même journée de printemps : le 8 mai 1984. C'est une journée contrastée : le Nord-Ouest, le Centre et les Pyrénées sont couverts, alors que le Bassin Aquitain ainsi que la côte méditerranéenne sont dégagés. La limite entre ces deux types de temps est à déterminer.

##### III-4-a. Carte de référence

La carte de référence a été établie à partir des données sol. On a utilisé pour la tracer une méthode utilisant les données sol parmi les plus performantes : l'interpolation de GANDIN. La totalité des données disponibles (88 héliographes) a été utilisée afin de diminuer la variance d'estimation (Fig.(III)-22 a).

#### III-4-b. Estimation CACTUS

La carte satellite (Fig.(III)-22 f) est riche en détails, d'où l'aspect contourné des isolignes. On note que globalement les valeurs CACTUS sont plus faibles que les valeurs sol. On perçoit un gradient côtier en Bretagne et sur le Nord du Cotentin. Les Pyrénées sont couvertes, le Massif Central aussi, mais entre les deux s'étend une zone de beau temps (de rapport d'insolation supérieur à 75 %), qui est également vue par la carte de référence. La zone de beau temps sur le bassin aquitain est moins nuancée, (saturation à 100%), que celle de la carte de référence.

A la visualisation avec des isolignes, on préfère une mosaïque dont l'élément est une maille de la grille. Ce procédé ne permet pas de reproduction noir et blanc, mais a été couramment utilisé au cours de cette étude. La visualisation sur console couleur (DN 660 APOLLO) d'une image 55 x60 mailles est très rapide (de l'ordre de quelques secondes avec le programme VISU, mis au point par P. LACOMBA et S.GALLE). Toutes les journées ont été ainsi contrôlées et prises en photo.

#### III-4-c. Interpolation optimale de GANDIN

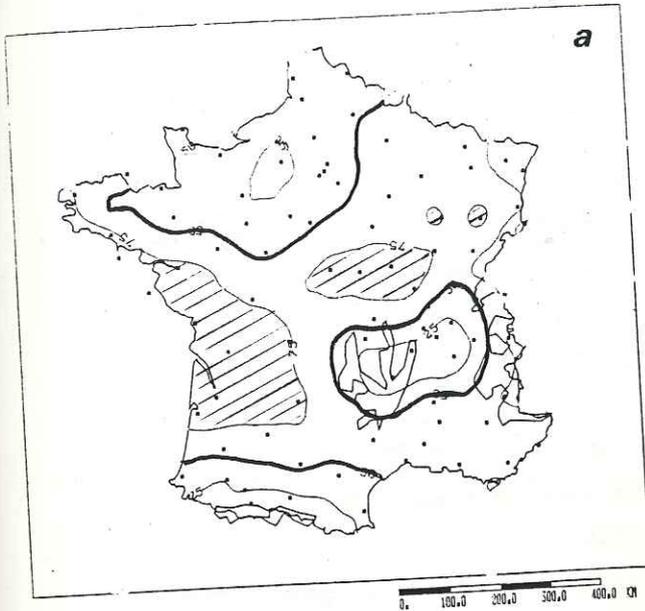
##### Réseau de 23 héliographes (Fig.(III)-22 g)

On obtient globalement la même forme que par la méthode CACTUS, mais l'éclaircie au sud du Massif Central n'est pas perçue (il n'y a pas d'héliographe dans cette zone), et l'éclaircie du nord du Cotentin n'est plus estimée comme un phénomène côtier, mais généralisée à tout le Cotentin, aucune information ne permettant de situer une frontière sur la côte plutôt qu'ailleurs. Les Pyrénées sont couvertes.

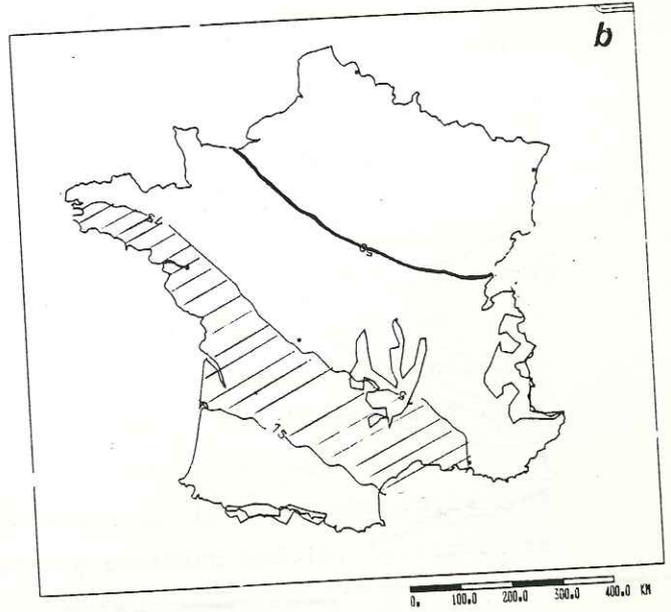
##### Réseau de 5 héliographes (Fig.(III)-22 b)

Etant donné le faible nombre de stations, on ne saisit qu'une tendance générale : temps plutôt couvert dans le Nord.

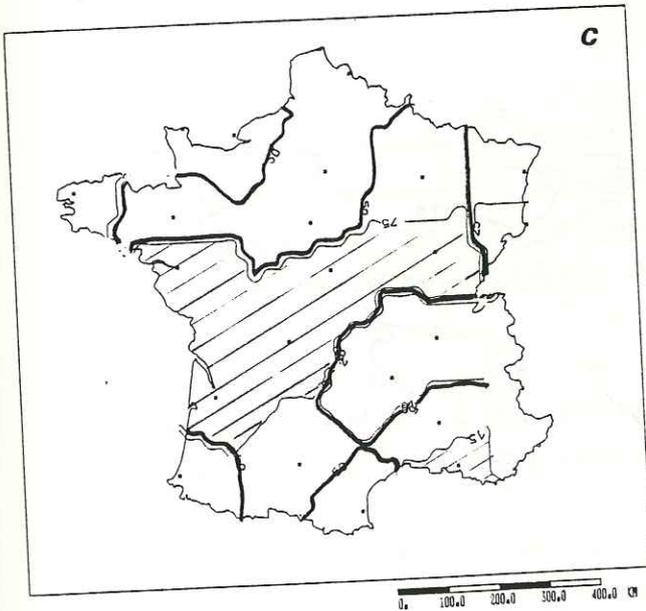
Fig III-22 : rapport d'insolation le 8 mai 1984 :  
interpolation par différentes méthodes.



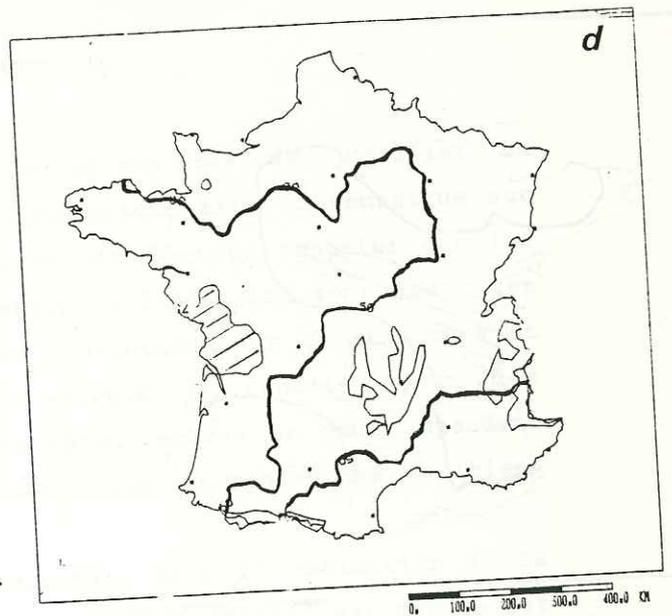
interpolation optimale à  
partir de 88 héliographes



idem avec 5 héliographes  
seulement

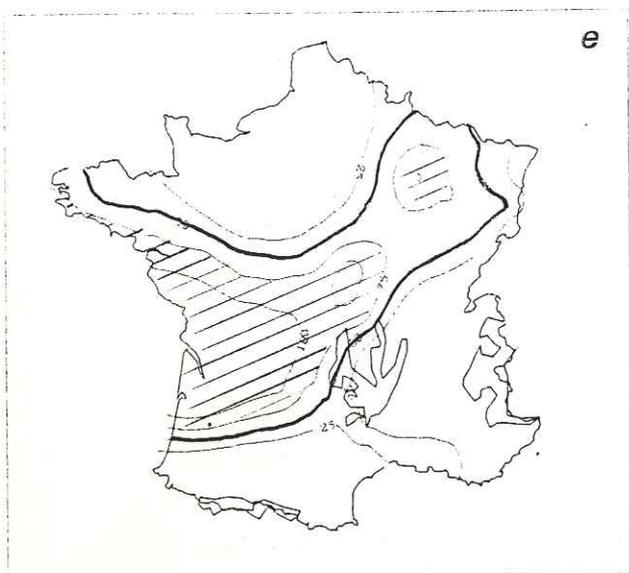


plus proche voisin



moyenne des trois plus  
proches voisins

Fig III-22 (suite):



ACP avec 8 composantes retenues

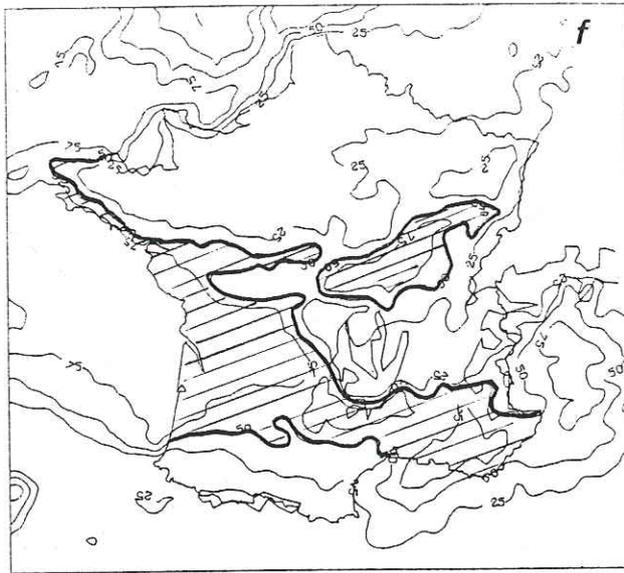
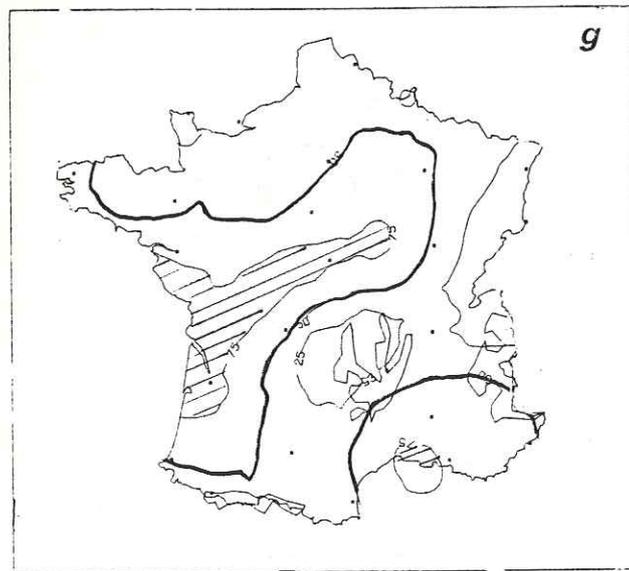
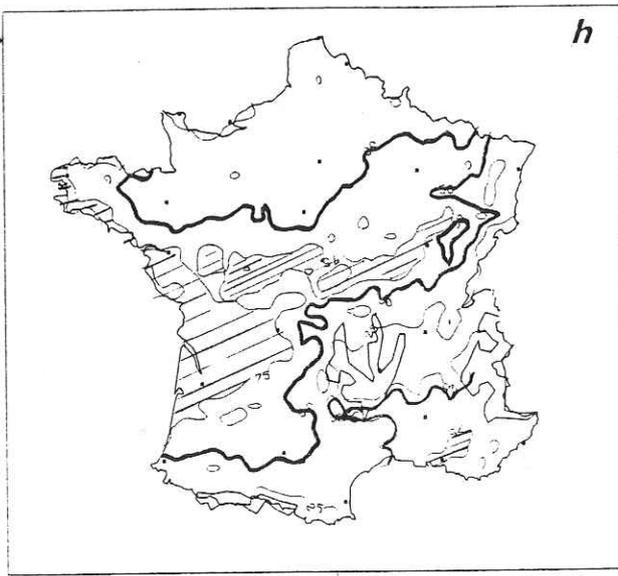


image cactus initiale



interpolation optimale à partir de 23 héliographes



étalonnage de l'image cactus par 23 héliographes

#### III-4-d. Moyenne arithmétique

##### Un voisin (Fig.(III)-22 c)

La carte se compose d'un ensemble de polygones (qui représentent la zone d'influence d'une station), que l'on affecte à une valeur ; d'où l'aspect géométrique des isolignes.

##### Trois voisins (Fig.(III)-22 d)

En augmentant le nombre de voisins à prendre en compte, on augmente le nombre de polygones (qui correspondent maintenant à des zones de trois voisins communs), et en diminuant ainsi leur taille, on obtient une carte plus nuancée. Le résultat obtenu est globalement semblable à celui de l'interpolation de GANDIN avec 23 héliographes. Une observation plus attentive permet toutefois de noter des différences importantes : la couverture des Pyrénées n'a pas été vue, ni le gradient côtier en Bretagne par exemple. Contrairement à l'interpolation GANDIN, la structure spatiale du phénomène n'est pas prise en compte.

#### III-4-e. Conclusion

Le choix d'une journée particulière ne permet pas de justifier un classement entre différentes méthodes d'estimation, mais il renseigne sur l'aspect des cartes obtenues. Il nous a permis de nous rappeler que les zones de plus forte hétérogénéité du rapport d'insolation sont les côtes et les reliefs. Le satellite ne fait pas d'hypothèse sur la structure et cerne bien ces difficultés. Le sol, suivant sa densité, peut les "manquer". Cette supériorité des données satellitaires ne peut cependant être saisie que par un réseau test adéquat : suffisamment dense et "bien" situé.

L'implantation du réseau test est fondamentale dans l'estimation de la qualité d'une méthode. Le satellite, dont la qualité de l'estimation est spatialement constante, (cf fig. III-7), sera privilégié par un réseau test situé dans des zones hétérogènes, que ne perçoivent pas les méthodes sol. Les méthodes sol, de variance d'estimation variables dans l'espace,

seront privilégiées par des stations test proches des stations d'interpolation.

Pour notre cas, le réseau test disponible est un réseau régulier, systématiquement éloigné du réseau d'interpolation, mais pas non plus adapté pour saisir l'hétérogénéité ; (l'ouest de la Bretagne n'est pas testé par exemple)

#### IV - CONCLUSION

Le programme CACTUS est une méthode d'estimation du rapport d'insolation journalier très simple : fondée sur un seuil de brillance, c'est une méthode binaire. Sa mise en pratique n'est cependant pas triviale : problèmes du réglage du seuil, du calage géographique par exemple. Nous avons pu établir, en se référant à la vérité sol (mesurée par des héliographes), que cette méthode est performante puisque de façon globale pour chaque période de l'année considérée (d'environ 70 jours chacune) le pourcentage de variance expliquée est de 86 % au printemps, 83 % en été et 77 % en hiver (ce qui correspond à des coefficients de corrélation de .93, .91 et .98). Lorsque l'on cumule ces résultats, aussi bien dans l'espace que dans le temps, on travaille sur une variable plus lisse et les estimations tirées du programme CACTUS sont encore plus performantes (comme celles des méthodes d'interpolation classiques).

La validation du programme CACTUS a été complétée par comparaison de ses performances avec celles obtenues par des méthodes d'interpolation basées sur les seules données sol (réseau de 23 héliographes). Ces dernières sont performantes, puisque la méthode "plancher" du plus proche voisin explique (au printemps) déjà 80 % de la variance ! Les méthodes les plus élaborées comme l'ACP ou l'interpolation de GANDIN expliquent respectivement 84 et 86 % de la variance globale. C'est par rapport à ces résultats que l'apport de la télédétection doit être évalué.

La distribution des coefficients de détermination ( $r^2$ ) calculés pour chaque journée entre les 61 stations test et les mesures CACTUS a été montrée significativement supérieure à celles obtenues à partir d'un réseau de 23 héliographes, quelle que soit la méthode utilisée. La haute densité spatiale des données CACTUS (réparties tous les 20 kilomètres) compense le fait qu'on utilise des mesures indirectes. Le programme CACTUS saisit l'hétérogénéité spatiale du rapport d'insolation. Son information est particulièrement utile dans les zones de fort gradient comme les côtes et les reliefs.

Toutefois, on gardera en mémoire les coefficients de la droite de régression (pente 0.8, ordonnée à l'origine 9). Ces coefficients traduisent un biais des mesures CACTUS. Ce biais n'existe pas pour les méthodes d'interpolation GANDIN ou l'ACP. Cet avantage en précision des méthodes basées sur les données sol n'est pas sanctionné par le coefficient de corrélation qui ne quantifie que la fluctuation de deux séries de mesures. C'est pourquoi nous avons également calculé l'erreur absolue normée par la moyenne  $E_a$ . Nous proposons dans le chapitre suivant un étalonnage de l'image CACTUS par les mesures sol.

4ème PARTIE

ETALONNAGE DU SATELLITE  
A L'AIDE DE MESURES HELIOGRAPHIQUES

## - INTRODUCTION

La valeur que l'on souhaite estimer est la durée quotidienne d'insolation. Les héliographes nous en fournissent une mesure directe mais ponctuelle. L'interprétation des images de télédétection au contraire nous fournit des données indirectes mais de haute densité spatiale. L'étalonnage d'image a pour but de tirer parti au mieux des avantages de ces deux types de mesure, de nature différente. Une telle démarche ne peut cependant apporter une amélioration de l'estimation finale que si les deux signaux confluent. A la limite, s'ils sont indépendants, la connaissance de l'un est inutile à l'appréhension du deuxième.

Dans notre cas, la fonction de covariance croisée entre les valeurs sol et télédétection centrées deux fois est pratiquement confondue avec les deux fonctions de covariance directes : la confluctuation est donc excellente (cf. 2ème partie II-2-2). Elle ne s'en démarque qu'aux faibles distances à cause de la présence d'un bruit blanc de courte portée (la confluctuation étudiée au chapitre précédent illustre le cas particulier où l'interdistance est nulle). C'est pourquoi, les images CACTUS se prêtent à l'étalonnage par des données sol, dans le but d'estimer le rapport d'insolation de façon encore plus précise.

L'interpolation optimale multivariable (appelé cokrigeage dans le cas mono-réalisation) est une généralisation du système d'interpolation optimale. Elle prend en compte les structures directes et la corrégionalisation des différentes sources de données disponibles.

L'intérêt et les limites de l'étalonnage seront évalués par le gain qu'il est susceptible d'apporter par rapport aux performances de chaque type d'estimation pris isolément.

## I - L'INTERPOLATION OPTIMALE MULTIVARIABLE

L'interpolation optimale multivariable permet d'estimer de façon "optimale" une variable régionalisée principale  $Z_0(t)$  en prenant en compte d'autres variables  $Z_N(t)$ . Dans le cadre de l'étalonnage d'image, ce nombre  $N$  se réduit à un. Nous ne considérerons ici que ce cas particulier, pour lequel on rappellera les résultats principaux. Pour une étude complète, on se référera par exemple à JOURNEL; 78 ou CREUTIN; 87, pour les applications à l'étalonnage d'image.

D'un point de vue théorique, l'interpolation optimale multivariable est une généralisation de l'interpolation optimale. Le système d'interpolation doit prendre en compte les fonctions de structure directes, mais aussi croisées, comme nous allons le voir.

## I-1. Système d'interpolation optimale multivariable

L'estimateur  $z_k^*(t_0)$  de la variable  $Z$ , pour le  $k^e$  évènement, au point  $t_0$  sans mesure, s'exprime comme une combinaison linéaire des données disponibles de chacune des variables. Les données disponibles sont notées:

nature de la mesure	nombre de points de mesure	lieu de mesure	nombre d'évènements	données disponibles
sol	$p$	$t_i$	$n$	$z_k(t_i)$
télé-détection	$G$	$t_g$	$n$	$z_{kT}(t_g)$

$$\text{d'où } z_k^*(t_0) = \sum_{i=1}^p \lambda_i z_k(t_i) + \sum_{g=1}^G \lambda_g z_{kT}(t_g).$$

On impose à cet estimateur les deux contraintes suivantes :

1- optimalité  $\text{Var}(z_k^*(t_0) - z_k(t_0))$  minimum

2- non biais  $E(z_k^*(t_0) - z_k(t_0)) = 0$

Cette deuxième condition ne s'impose plus lorsque l'on travaille sur des valeurs centrées, non biaisées par définition.

Sous hypothèse de stationnarité d'ordre 2 de la corrégionalisation (moments directs et croisés des deux variables), ces deux conditions donnent le système d'interpolation suivant :

$$\sum_i \lambda_i C_{zz}(h_{ij}) + \sum_g \lambda_g C_{zTz}(h_{gj}) + \mu = C_{zz}(h_{0j})$$

$$\sum_i \lambda_i C_{zzT}(h_{ij}) + \sum_g \lambda_g C_{zTzT}(h_{gj}) + \mu_T = C_{zzT}(0_j)$$

$$\sum_i \lambda_i = 1$$

$$\sum_i \lambda_g = 0$$

où  $C_{zzT}(t_i, t_j) = E(z(t_i) - m_z)(z_T(t_j) - m_{zT}) = C_{zzT}(h_{ij})$

NB : La relation  $C_{zzT}(h_{ij}) = C_{zTz}(h_{ij})$  n'est vraie que si les deux variables ne sont pas décalées spatialement, ce qui est notre cas (cf 2ème partie II).

Ce système peut encore se noter sous forme matricielle

$$\begin{pmatrix} C_{zz} & C_{zTz} & 1 & 0 \\ C_{zzT} & C_{zTzT} & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_i \\ \lambda_g \\ \mu \\ \mu_T \end{pmatrix} = \begin{pmatrix} C_{zz} \\ C_{zzT} \\ 1 \\ 0 \end{pmatrix}$$

A l'optimum, la variance de l'erreur s'écrit en fonction des poids  $\lambda_i$  et  $\lambda_g$ , et des fonctions de covariance directe du sol, et croisée entre le sol et le satellite.

La taille d'un tel système ( $p+G+2$ ) est énorme, le nombre de mesures de télédétection disponible  $G$  étant très grand. Il est exclu de le proposer à des fins opérationnelles. On devra donc le simplifier ou travailler en voisinage glissant. Cette deuxième possibilité permet en outre une hypothèse plus faible car la stationnarité locale est alors suffisante.

Les remarques suivantes ont pour but de mieux saisir le fonctionnement de ce système d'interpolation multivariable, en prenant des cas extrêmes :

- Si les deux variables sont indépendantes, les poids  $\lambda_g$  seront nuls : la deuxième variable n'apporte rien à l'estimation de la première.
- Si, au contraire, les deux variables coïncident parfaitement le système s'écrit comme celui du krigeage simple (alors appliqué à un grand nombre de points de mesures)

$$\begin{bmatrix} C_{zz} & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_i \\ \mu \end{bmatrix} = \begin{bmatrix} C_{zz} \\ 1 \end{bmatrix}$$

Seules les conditions de non biais distinguent les deux types de mesure. Les deux conditions de non biais traduisent le fait que l'on privilégie la variable  $Z$  pour l'estimation de la moyenne. Si la mesure de télédétection n'est pas biaisée ( $m_z = m_{zT}$ ), on se ramène à la condition unique :

$$\sum_i \lambda_i + \sum_g \lambda_g = 1$$

Par contre, si les fonctions de structure des deux variables sont différentes (par exemple par une pépite due à des erreurs de mesure de

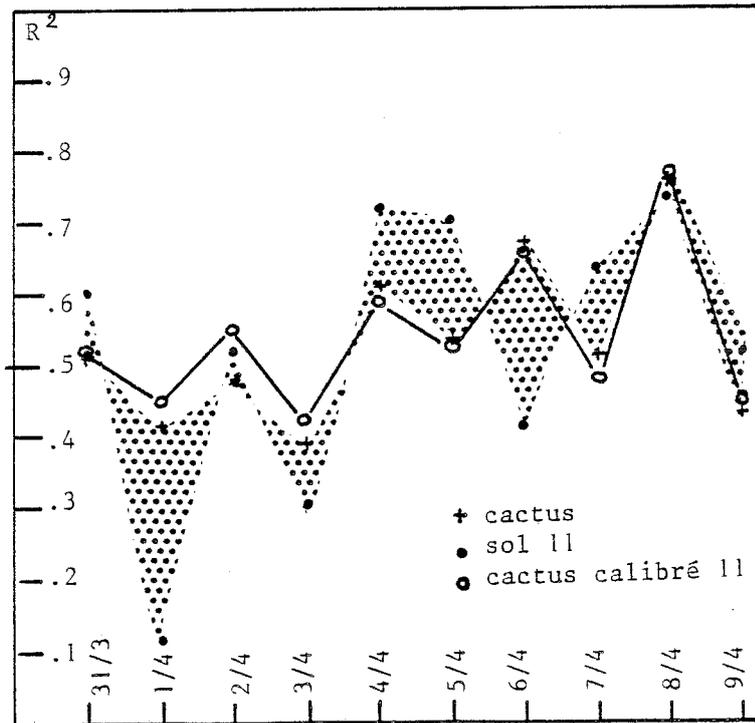


Fig IV-1 : Courbe d'efficacit e des m ethodes : (+) cactus, (●) interpolation optimale a partir de 23 h eliographes, (○) image cactus recal ee par 23 h eliographes.

10 journ ees d'avril 1983

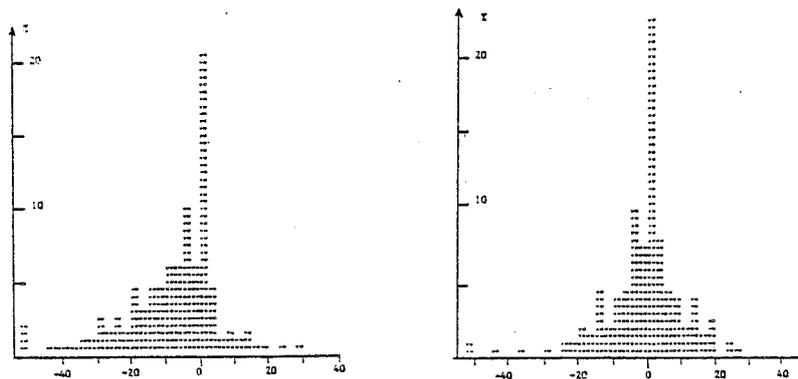


Fig IV-2 : HISTOGRAMME DE FREQUENCE D'ERREURS

avant et apr es recalage de l'image satellitaire par les donn ees sol

natures différentes), l'interpolation optimale multivariable est utile, même si les mesures ne sont pas biaisées.

- La résolution d'un tel système n'autorise que certains modèles pour les fonctions de covariance directes et croisées. Nous ne discuterons pas ce point dans ce chapitre (voir partie 2-1).

#### I-2. Simplification possible : recalage de guess-field

Cette simplification, envisagée par DELHOMME,1979, et utilisée par DELRIEU,86, CREUTIN,87, repose sur la décomposition orthogonale suivante :

$$z_k(t) = z_{kT}(t) + e_k(t)$$

où  $e_k(t)$  est un terme correctif régionalisé, statistiquement indépendant de  $z_{kT}(t)$ . Cette orthogonalité permet de décomposer le système de cokrigage en deux sous-systèmes indépendants. La mesure de télédétection étant connue avec une définition spatiale suffisante en tout point  $t_0$ , on écrit alors simplement :

$$z^*_k(t_0) = z_{kT}(t_0) + \sum_i \lambda_i C_k(t_i)$$

La valeur de référence s'exprime comme la somme de la valeur de télédétection et d'une erreur (égale à la différence entre les mesures sol et satellite) estimée en tout point par interpolation optimale.

Cette méthode d'étalonnage a été appliquée sur dix journées d'avril 83 que l'on a traitées dans un contexte mono réalisation (on a utilisé le cokrigage). Les résultats en terme de cofluctuation (fig.IV-1) montrent une amélioration pour quatre journées (les 1,2,3 et 8 avril 1983) obtenue par l'étalonnage de l'image CACTUS à l'aide de 11 héliographes et une dégradation les 4,5 et 7 avril. Cette amélioration est appréciée en

fonction des performances de chaque méthode d'interpolation (CACTUS ou krigeage) prise individuellement.

L'observation des histogrammes de fréquence d'erreur (fig.IV-2) montre la correction du biais par l'étalonnage. Des résultats plus complets, pour différentes densités de réseau d'étalonnage, ainsi que les variogrammes directs et croisés expérimentaux, sont décrits dans CREUTIN et GALLE, 1985.

La conclusion sur l'apport de l'étalonnage simplifié pour l'évaluation de la durée d'insolation, si elle est encourageante, reste cependant limitée par le faible nombre de journées disponibles.

Une approche climatologique a, depuis, été rendue possible, grâce aux 200 journées disponibles en 1984. L'analyse de la covariance directe et croisée (2ème partie II) nous a conduit à la relation suivante :

$$z_{TK}(t) = \alpha z_k(t) + \mathcal{E}_k(t)$$

où  $\mathcal{E}_k(t)$  est un terme correctif régionalisé de faible portée, orthogonal à la mesure sol  $z_k(t)$ , et non pas à la mesure de télédétection. La simplification du système d'interpolation multivariable n'est alors plus "optimale" ; c'est pourquoi, nous avons préféré conserver le système complet.

## II- Etalonnage des valeurs satellitaires par interpolation optimale multivariable

### II-1. Fonction de structure utilisée

Les fonctions de structure utilisées pour caractériser la dépendance spatiale des mesures sol, des mesures satellitaires et leur interdépendance, sont les fonctions de covariances directes et croisées. Celles-ci ont été calculées à partir des variables centrées deux fois

( $X_S(t)$  et  $X_T(t)$ ), afin de soustraire les influences déterministes connues (cf. 2ème partie II) et se rapprocher de la structure purement spatiale du rapport d'insolation.

## II-2. Choix de la variable

On cherche un estimateur du rapport d'insolation  $Z(t)$ . Cette variable n'est pas stationnaire. Un double centrage climatologique et spatial permet de s'approcher d'une variable à la fois normale et stationnaire d'ordre 2.  $X(t)$ . Deux problèmes pratiques se posent alors pour fournir une estimation de  $Z(t)$  : la reconstitution de la moyenne climatologique de  $Z$  et l'estimation des moyennes spatiales  $Y_k$  et  $Y_{Tk}$  des résidus climatologiques  $Y(t)$  et  $Y_T(t)$ .

Le problème de l'estimation des moyennes spatiales pose la question de savoir sur quelle variable ( $X$  ou  $Y$ ) va porter l'interpolation optimale multivariable. Le choix de  $X$  (valeur centrée deux fois) évite la condition de non biais du système d'interpolation mais nécessite l'estimation des moyennes spatiales. Cette estimation devient délicate lorsque le nombre de points d'étalonnage décroît. Le choix de  $Y$  (écart à la moyenne climatologique), au contraire, permet de s'affranchir des estimations de  $Y(t)$  et  $Y_T(t)$ , qui seront filtrées par les conditions de non biais alors imposées au système.

Nous avons choisi d'interpoler la variable  $Y$  avec une condition de non biais unique pour les poids  $\lambda$  :

$$\sum_{i=1}^p \lambda_i + \sum_{g=1}^G \lambda_g = 1$$

Car les valeurs centrées  $Y_S(t)$  et  $Y_T(t)$  ont même moyenne climatologique. La valeur du rapport d'insolation au point  $t_0$  est ensuite reconstituée par addition de la moyenne climatologique en ce point.

$$z^*_k(t_0) = y^*_k(t_0) + m^*_z(t_0)$$

### II-3 Estimation de la moyenne climatologique

Nous disposons, pour estimer la moyenne climatologique, en tout point  $t_0$ , à la fois des moyennes aux  $p$  points de mesure sol et aux  $G$  noeuds de la grille CACTUS. Les moyennes calculées à partir des données CACTUS sont biaisées (cf. 3ème partie I) mais leur co-fluctuation avec les moyennes des valeurs sol est bonne : le pourcentage de variance expliqué est de 88 % en hiver, 75 % au printemps et 80 % en été. C'est pourquoi, nous avons choisi d'estimer la moyenne climatologique par cokrigage.

$$m_z^*(t_0) = \sum_i^p \lambda_i m_z(t_i) + \sum_g^G \lambda_g m_{zT}(t_g)$$

Le biais de la moyenne climatologique de télédétection nous a amené à imposer deux conditions de non biais à ce système. Les performances du cokrigage pour l'estimation de la moyenne climatologique sont évaluées par rapport à celles des deux méthodes simples : CACTUS et krigeage des moyennes sol. C'est à dire que l'on veut apporter un gain par rapport aux possibilités des deux systèmes de mesure (sol et CACTUS) pris isolément. Le test s'effectue sur 61 héliographes qui n'ont pas servi pour l'éta lonnage (tableau IV-1).

Tableau IV-1 : Performances des différents estimateurs de la moyenne

période : printemps

nombre de jours : 71

nombre de station-test : 61

estimateur	r	r <sup>2</sup> %	penne	ordonnée origine	E <sub>a</sub> %
CACTUS	.86	75	.51	2	9
krigeage					
23 stations	.85	73	.9	3	5.2
5 stations	.73	54	1.1	-6	7.8
cokrigeage					
23 stations	.89	79	.9	3	4.7
5 stations	.75	55	1.1	-3	7.7
gain CACTUS					
23 stations	+.03	+4	+	-	+4.3
5 stations	-.11	-20	+	-	+2.3
gain sol					
23 stations	+.04	+6	0	0	+0.5
5 stations	+.02	+1	0	+	0

E<sub>a</sub> = erreur absolue normée par la moyenne

Le gain sur l'estimation de la moyenne climatologique, apporté par le cokrigeage des deux types de mesure, dépend de la densité du réseau sol utilisé. Si la densité du réseau sol est trop faible (5 stations), l'étalonnage ne se justifie pas : ses performances égalent celles du sol seul, dont la précision (en terme de cofluctuation) est très inférieure à celle de CACTUS. Pour le réseau de 23 héliographes par contre, l'étalonnage apporte une amélioration de la précision à la fois en terme de cofluctuation et de biais.

#### II-4. Choix du voisinage

L'interpolation des valeurs centrées  $Y(t)$ , par interpolation optimale multivariable conduit à un système énorme (de taille  $G+p$ ) si on considère l'ensemble de la grille. C'est pourquoi, nous avons effectué une interpolation en voisinage glissant.

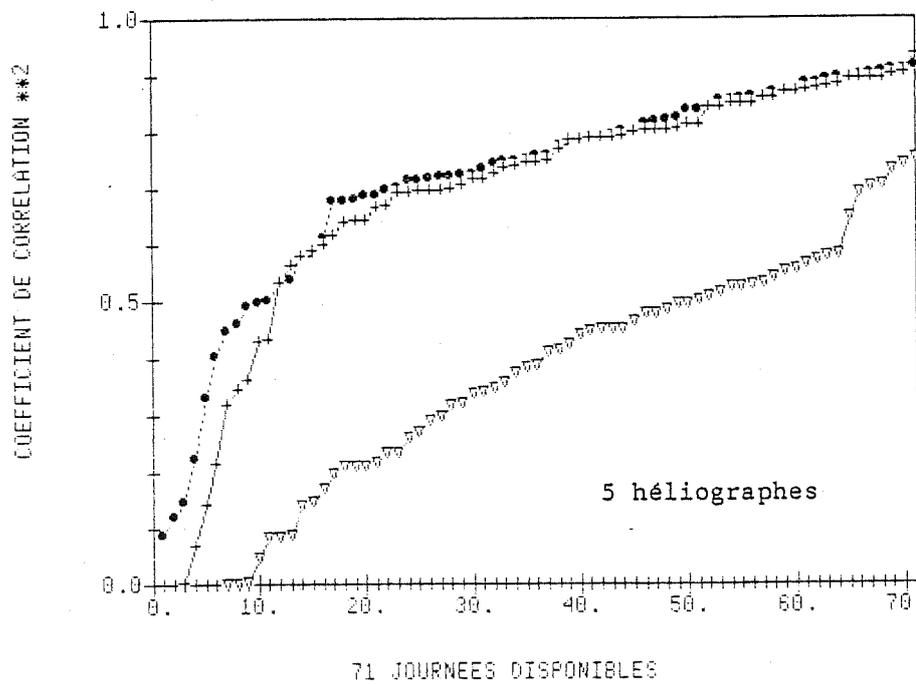
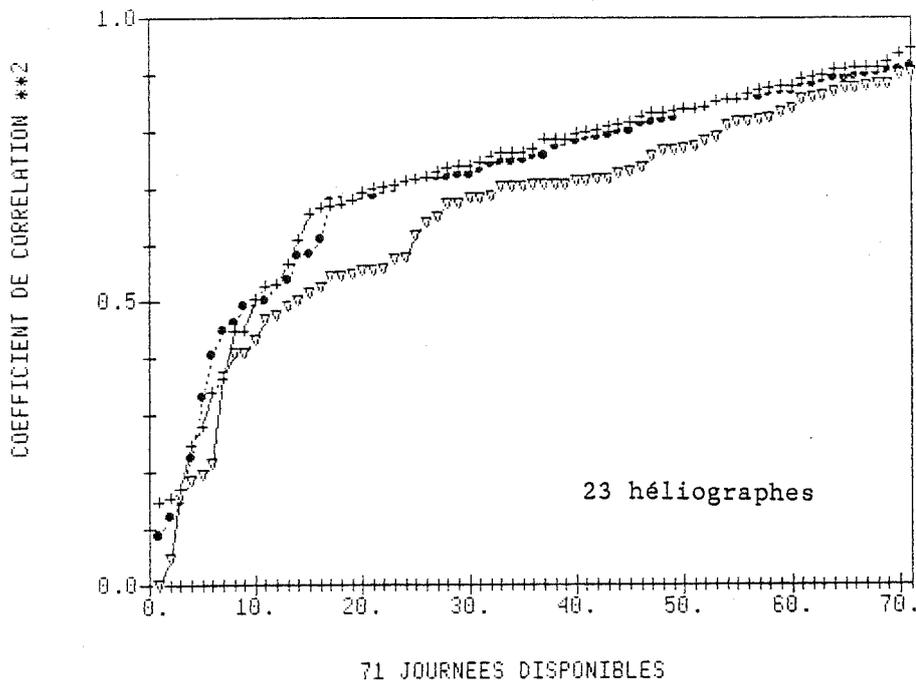
Le choix de la taille du voisinage est plus délicat que lorsque l'on ne dispose que d'une variable unique. En effet, les densités spatiales des deux réseaux de mesure sont très différentes, pouvant entraîner un choix différent pour chacune d'elle. Un point de l'image satellitaire sera affecté d'un poids  $g$  faible, s'il est masqué par un autre point de grille : il est inutile de prendre en compte une grande surface autour du point considéré pour les mesures satellitaires, très denses. D'autre part, on veut s'assurer d'un nombre minimum de valeurs sol pour l'étalonnage. Nous avons donc choisi un nombre de voisins unique pour les deux réseaux, ce qui entraîne des surfaces de voisinage différentes.

#### II-5. Résultats obtenus

Nous avons testé les estimations du rapport d'insolation obtenues par interpolation optimale multivariable, sur le réseau de 61 héliographes n'ayant pas servi à l'étalonnage. Les critères utilisés sont les mêmes qu'au chapitre précédent.

L'examen des courbes d'efficacité obtenues au printemps pour différentes densités de réseau sol, (fig. IV-3a et b), montre que les performances du cokrigeage (+) en terme de cofluctuation sont égales à celle du meilleur dispositif de mesure (CACTUS dans les deux cas) mais sans toutefois les dépasser. L'intérêt de l'étalonnage peut alors être mis en cause. Cependant, le but fixé à l'étalonnage des images du satellite ne se limite pas à la précision en terme de cofluctuation, mais cherche également la correction du biais des images satellitaires. La corrélation est, par

Fig IV-3 : Courbe d'efficacité des méthodes : (●) cactus, (▽) interpolation optimale a partir des héliographes, (+) image cactus étalonnée par les héliographes.



PRINTEMPS

(71 évènements)

définition, insensible à ce biais. Nous utiliserons pour le caractériser l'erreur absolue normée par la moyenne  $E_a$ .

L'amélioration apportée par l'étalonnage de l'image CACTUS est alors claire (cf. tableau IV-2) : on passe de 25 % d'erreur absolue sur l'image brute à 18 %, pour un étalonnage avec 23 stations (qui égale alors l'erreur absolue de 19 %, faite avec le dispositif sol seul).

NB : cette correction du biais se traduit également sur les coefficients de la droite de régression qui passent de (.8;9) à (.9;2) et s'approchent ainsi de (1;0).

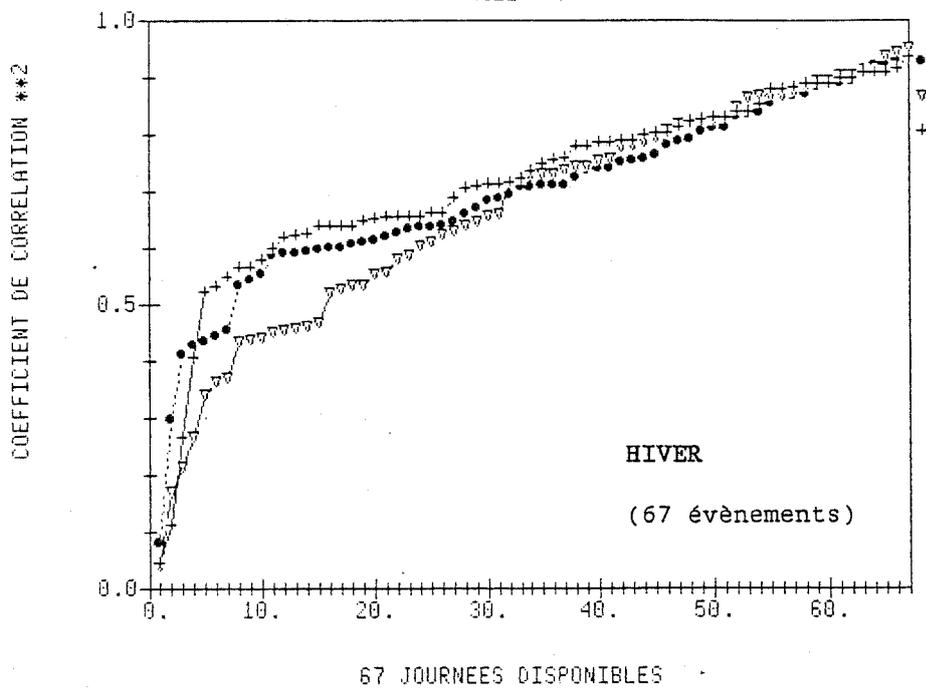
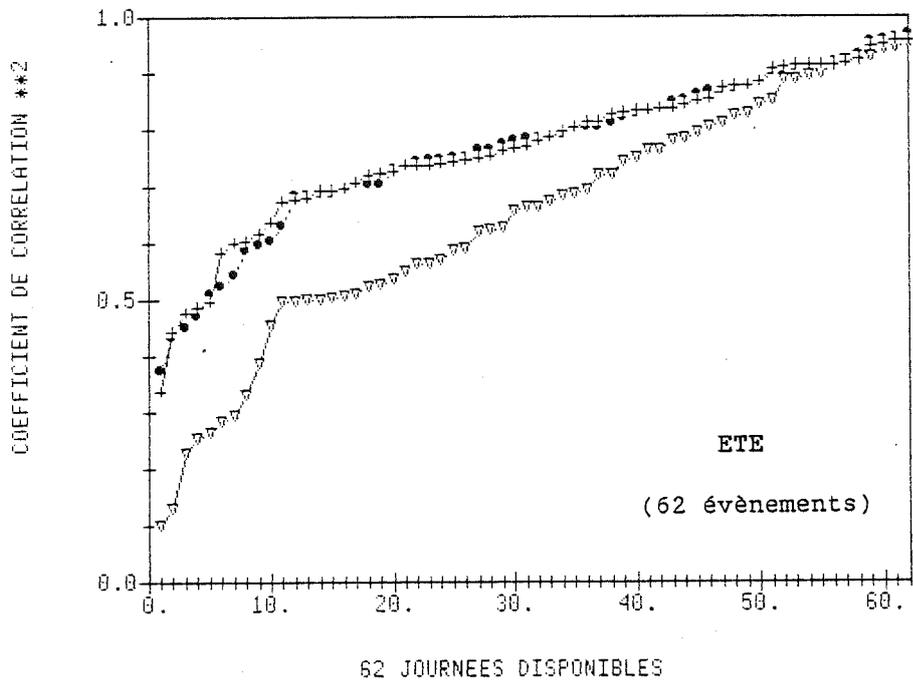
Tableau IV-2 : Comparaison des performances globales des différentes estimations du rapport d'insolation au printemps 1984 (71 jours).

Nombre de stations-test : 61

	r	r <sup>2</sup> %	pente	ordonnée origine %	E <sub>a</sub> %
CACTUS	.92	85	.8	9.	25
Interpolation de GANDIN					
23 stations	.93	86	1.0	1	19
5 stations	.81	65	.9	2	34
Etalonnage					
23 stations	.95	89	.9	2	18
5 stations	.92	85	.9	3	24

Les courbes d'efficacité ont également été calculées pour les deux autres périodes (Fig.IV-4 (a) hiver et (b) été). Elles montrent que les performances de l'image étalonnées sont comparables à celles de l'image brute en terme de co-fluctuation (ce que nous avons déjà remarqué au

Fig IV-1 : Courbe d'efficacité des méthodes : (●) cactus, (▽) interpolation optimale a partir de 23 héliographes, (+) image cactus étalonnée par 23 héliographes.



printemps). Elles semblent même légèrement supérieures, mais de façon non significative.

Tableau IV-4 : Comparaison des performances globales des estimations du satellite sol et du satellite étalonné, pour l'hiver (67 jours) et l'été (62 jours).

Test réalisé sur 61 stations.

méthode	HIVER					ETE				
	r	r <sup>2</sup> %	pen- te	0.0 %	E <sub>a</sub> %	r	r <sup>2</sup> %	pen- te	0.0 %	E <sub>a</sub> %
CACTUS	.88	77	.80	9	35	.91	82	.74	12	22
GANDIN	.89	80	.95	0.5	30	.90	80	.98	1	17
ETALONNAGE	.91	83	.95	1	30	.93	86	.91	4	15

Le tableau IV-4 ci-dessus montre que l'étalonnage corrige le biais de l'image brute : l'erreur absolue normée par la moyenne passe de 35 % à 30 % au printemps et de 22 % à 15 % en été, ce qui est remarquable.

Les estimations du rapport d'insolation sont globalement meilleures en été et moins performantes en hiver, ce que nous avons déjà remarqué sur les images brutes, et attribué à la nature des situations météorologiques. Ce classement n'est pas modifié par l'étalonnage, les estimations sol étant, elles aussi, moins performantes en hiver. C'est en été que l'estimation du rapport d'insolation est la plus aisée.

C'est souvent sur cette seule période que sont évaluées les méthodes décrites dans la littérature (ELLIS 78, TARPLEY 79, GAUTHIER 80 et 82, MOSER 83, DEDIEU 84). Les résultats obtenus et testés sur une période plus longue (TARPLEY 84, GAUTHIER 82) sont rares car ils nécessitent une mise en place opérationnelle, mais eux seuls permettent une évaluation robuste de la méthode.

- Conclusion

L'étalonnage de l'image CACTUS à l'aide de données sol a été effectué par interpolation optimale multivariable. Une année d'images satellitaires a ainsi été étalonnée. De façon générale, les performances de l'image satellitaire étalonnée sont supérieures à celle de l'image brute en terme de biais, et à celle du sol en terme de co-fluctuation. L'étalonnage a donc su utiliser les atouts de chacune des deux méthodes prises séparément pour fournir une estimation du rapport d'insolation qui combine les deux avantages. Ceci, quelle que soit la période de l'année. De façon plus précise, les résultats appellent deux remarques :

1- Si aucun des deux dispositifs n'est performant, l'image étalonnée ne le sera pas non plus. En particulier, un réseau sol trop lâche ne sera pas apte à corriger le biais de façon satisfaisante en tout lieu, de même qu'il ne pouvait fournir une estimation précise du rapport d'insolation dans les zones situées au-delà de la portée d'une station. Il faut ici rappeler que les résultats sont jugés en référence à un réseau test qui se situe justement dans les zones où l'influence de ces stations est le plus faible. L'apport de l'étalonnage à l'image brute ne peut donc être que sous estimé.

Etalonnage par différents réseaux :

Le réseau de 5 héliographes, n'apporte une amélioration que sur les coefficients de la droite de régression, et non sur l'erreur absolue normée par la moyenne. Une telle correction peut se faire, de manière beaucoup moins lourde par correction linéaire des données CACTUS.

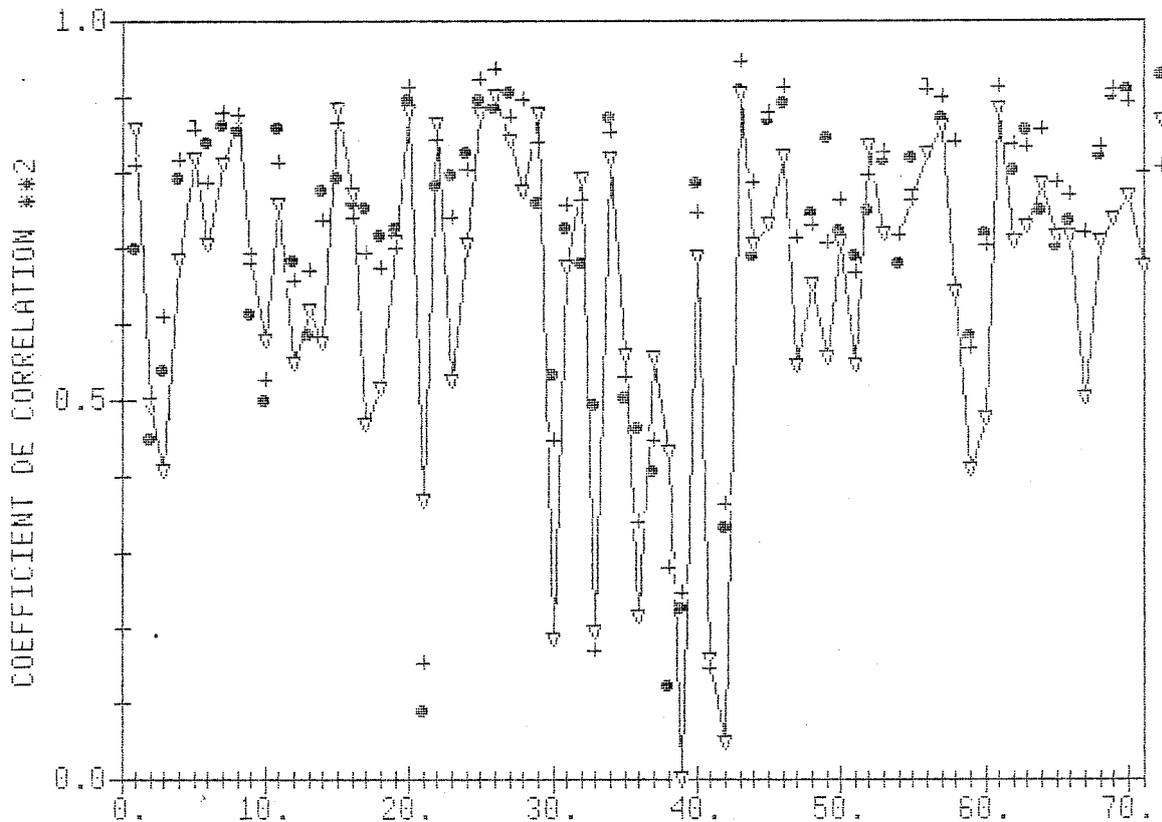
Pour un réseau d'étalonnage plus dense (23 héliographes), les coefficients de la droite de régression sont meilleurs, mais surtout, l'erreur absolue normée par la moyenne est considérablement réduite. Cette diminution d'erreur n'est pas évidente de prime abord.

Il existe donc une densité minimale de réseau d'étalonnage, au-dessous de laquelle l'interpolation optimale multivariable est inutile.

- L'étalonnage de l'image de télédétection n'a pas permis d'obtenir une estimation significativement meilleure que celle du meilleur des deux types de mesure. Cependant, les deux dispositifs étant déjà très performants sur des critères différents, la conjugaison de leurs avantages confère, à elle seule, à l'image étalonnée une qualité supérieure.

Cette remarque globale doit être complétée par l'examen de la figure IV-5 qui montre les performances des deux types de mesure et de l'image étalonnée pour chaque jour du printemps pris individuellement. Il existe des journées, plus spécialement celles où l'image brute était performante, pour lesquelles l'étalonnage est effectivement meilleur que l'interpolation sol et que l'image brute en termes de cofluctuation (et inversement). Le lissage des courbes d'efficacité classées ne doit pas nous faire oublier cette réalité individuelle.

Fig IV-5 : Courbe d'efficacité des méthodes : (●) cactus, (▽) interpolation optimale a partir de 23 héliographes, (+) image cactus étalonnée par 23 héliographes.



71 JOURNEES DISPONIBLES AU PRINTEMPS

### CONCLUSION GENERALE

Tout au long de ce travail, nous avons présenté un certain nombre de conclusions sur la structure spatiale, la corréionalisation et l'évaluation du rapport d'insolation par un réseau d'héliographes et par des images satellitaires, ainsi que les conséquences que l'on pouvait en tirer quant à la complémentarité de ces dispositifs de mesure pour l'étude du rapport d'insolation.

Il faut tout d'abord rappeler les problèmes rencontrés concernant la navigation de l'image satellitaire. Nous les avons résolus de façon manuelle, par visualisation des cartes d'albédo. Nous avons constaté que ce contrôle visuel était systématiquement effectué, y compris après traitement par des systèmes de recalage automatiques. Les progrès récents, réalisés par le centre de réception des images satellitaires de Darmstadt (RFA), permettent d'espérer une simplification de ce problème, qui est à la base de tout traitement des images de télédétection.

L'analyse de la structure spatiale des processus mesurés par le sol et par le satellite a été étudiée à l'aide de la théorie des fonctions aléatoires. Afin d'approcher les hypothèses nécessaires à cette théorie, nous avons codé les données par un double centrage climatologique et spatial. Les structures des deux types de mesure du rapport d'insolation, étudiées par analyse objective d'une part, et par analyse en composantes principales d'autre part, présentent un haut degré de similarité. La perception très fine de la variabilité spatiale par le satellite a bien sûr permis de retrouver les grands traits climatologiques connus, mais a également conduit à identifier les particularités mal perçues par les moyens classiques de mesure (phénomènes côtiers notamment).

Cette analyse de la structure spatiale a été utilisée pour proposer une optimisation du réseau de mesures sol dans le sud-ouest de la France. L'évaluation des mesures satellitaires a été faite par comparaison avec les mesures sol. L'analyse de la différence des modes d'intégration entre les deux types de mesures a montré que la taille des mailles choisie (20x20 km<sup>2</sup>) compense de manière satisfaisante la nature instantanée de la mesure de télédétection pour le rapport d'insolation en France. On a donc directement comparé les valeurs des héliographes à celles des mailles correspondantes. L'étude en co-fluctuation des deux types de mesure a été complétée par une comparaison avec les performances de différents modèles d'interpolation des données sol seules. Le programme CACTUS s'est révélé plus performant en terme de co-fluctuation qu'un réseau sol de 23 héliographes. En terme de biais, par contre, les interpolations sol sont meilleures. Le programme CACTUS a tendance à exagérer les valeurs extrêmes.

L'étalonnage des images de télédétection a pour but d'utiliser les avantages complémentaires des deux types de mesures du rapport d'insolation. Les structures spatiales des deux processus étant semblables, une procédure de combinaison des mesures de télédétection avec les mesures sol par interpolation optimale multivariable a été proposée et testée. Les performances des images satellitaires sont tout-à-fait intéressantes car elles rendent bien compte de la variabilité spatiale du phénomène. Notamment, la co-fluctuation avec les données sol est remarquable. Après étalonnage de ces images, et à condition d'avoir un réseau sol assez dense, les estimations du rapport d'insolation sont en outre débiaisées et l'erreur absolue normée par la moyenne est réduite. La méthode d'étalonnage proposée est enrichissante par rapport aux méthodes d'estimation fondées sur les mesures sol et satellite prises individuellement.

L'étalonnage de l'image satellitaire pourrait également être envisagé par d'autres techniques comme l'analyse en composantes principales de processus par exemple. Cette étude reste à réaliser et à tester.

Une banque de données de télédétection aussi importante (elle couvre pratiquement une année) avait rarement été calculée et testée dans sa totalité (TARPLEY, 1981) et sur des données sol parfaitement indépendantes du modèle. De façon globale, les types de temps d'hiver, de type frontal à

passages nuageux rapides, sont moins bien perçus par le satellite ( $r^2 = 77\%$ ) qu'au printemps ( $r^2 = 85\%$ ) ou en été ( $r^2 = 82\%$ ). Ces résultats soulignent la nécessité de faire des tests sur plusieurs mois différents pour valider une méthode de façon robuste et objective. Ces différences se retrouvent après étalonnage, le sol étant lui aussi moins performant en hiver.

#### Perspectives pour le programme CACTUS

L'utilisation du programme CACTUS pour l'évaluation de l'insolation sur d'autres régions géographiques, en Afrique ou en mer par exemple, nous semblait très prometteuse. Elle nécessite au moins deux réflexions approfondies : la première sur le choix des nouveaux seuils (l'albédo du sol pouvant atteindre des valeurs très élevées en cas de brumes sèches en Afrique par exemple), la deuxième sur la validité du modèle, pour des types de temps très différents de ceux testés dans cette étude.

Ce travail ne porte que sur la durée d'insolation, mais son extension à une variable plus directement intéressante dans l'étude des phénomènes radiatifs ne pose pas de problème théorique. La méthode d'étalonnage de l'image de télédétection proposée est en effet capable de prendre en compte une troisième variable, le rayonnement global par exemple, et de fournir ainsi des cartes de rayonnement, moyennant une étude préalable de sa structure spatiale. On aurait ainsi l'avantage de prendre en compte la régionalisation, contrairement aux méthodes de régression linéaire habituellement proposées. On aurait également une carte des écart-types d'estimation, quantifiant en tout lieu la précision de l'estimation. En conclusion, la constitution d'une banque de données climatologiques du programme CACTUS est aisée, peu volumineuse et riche de possibilités même pour une zone bien instrumentée au sol comme la FRANCE.

## Annexe 1

## DETERMINATION DES SEUILS DU PROGRAMME CACTUS

## A1-1. INTRODUCTION

Chaque pixel de l'image de Météosat est classé comme clair ou couvert, par comparaison de sa réflectivité avec un seuil. Ce seuil est fixé de façon absolue, sans prendre en compte l'albédo du sol situé au-dessous ; il n'est fonction que de l'angle zénithal solaire.

Le choix du seuil est délicat ; celui-ci doit permettre à la fois d'éviter la confusion avec le sol et de détecter les nuages minces, d'albédo faible et proche de celui du sol. Cependant, un voile de cirrus, s'il est mince, n'empêche pas l'impression d'un héliographe et devra donc être classé comme "clair".

D'un point de vue pratique, le seuil dépend de façon discrète de l'angle zénithal solaire  $z_s$ . Quatre classes d'angle zénithal ont été choisies. La quatrième qui correspond à un soleil bas ( $z_s < 6^\circ$ ) n'est pas traitée, pour des raisons de numérisation à bord du satellite. Ce choix a pour conséquence une légère diminution de la durée du jour mesurée par le satellite. Il doit être mis en rapport avec les masques rencontrés sur terre, souvent inférieurs à 5 % mais présents, et qui diminuent également la durée du jour. L'erreur ainsi introduite est très faible sur l'estimation de la durée du jour et joue encore moins lorsqu'on veut estimer le rayonnement, qui est très faible pour un soleil bas.

## A1-2. MOYENS DE VERIFICATION DISPONIBLES

L'estimation du pourcentage de nébulosité journalier fourni par CACTUS est la moyenne des nébulosités obtenues pour les différentes images disponibles de la journée considérée. On peut donc observer l'effet du choix d'un seuil à deux niveaux :

### - journalier

Par corrélation entre la durée d'insolation des héliographes et le pourcentage de nébulosité fourni par CACTUS aux mailles correspondantes. Cette étude a pour avantage d'être automatisée et objective, mais, si les résultats sont mauvais, il faut remonter au niveau de l'image pour déterminer quels sont les seuils à modifier.

### - instantané (une image)

On ne dispose pas, pour ce pas de temps, de "vérité sol". Il va donc falloir utiliser d'autres données. Les clichés pris par les satellites (Météosat, NOAA7) seront notre référence. Le CMS archive une photo toutes les trois heures environ (3h, 6h, 9h, etc.). Sur ces photos, une néphanalyse visuelle (étude de la couverture nuageuse) permet de distinguer les zones couvertes des zones claires. La comparaison de ces photographies aux résultats de nébulosité du programme CACTUS est faite sur une image simplifiée (la nébulosité est exprimée en octa : de 0 pour clair à 8 pour couvert), plus accessible à un examen visuel.

Pour une image donnée, l'angle zénithal solaire n'est pas le même sur toute la France et celle-ci peut être découpée en deux (voire trois zones), correspondant à deux (trois) seuils de réflectivité différents.

## A1-3. PREMIERS RESULTATS JOURNALIERS

Nous avons, tout d'abord, disposé des résultats du programme CACTUS, pour sept journées en avril 1983. L'étude de la corrélation entre les valeurs mesurées au sol et les valeurs de CACTUS correspondantes montre (Fig.A1-1) une très nette sous-estimation du rapport d'insolation par le programme CACTUS. Un grand nombre de points se situent dans la partie bas droite du

diagramme et même sur l'axe des 0 % pour CACTUS, alors que le sol varie de 0 % à 100 %. La corrélation est faible ( $r = .62$ ), soit 38% de variance expliquée.

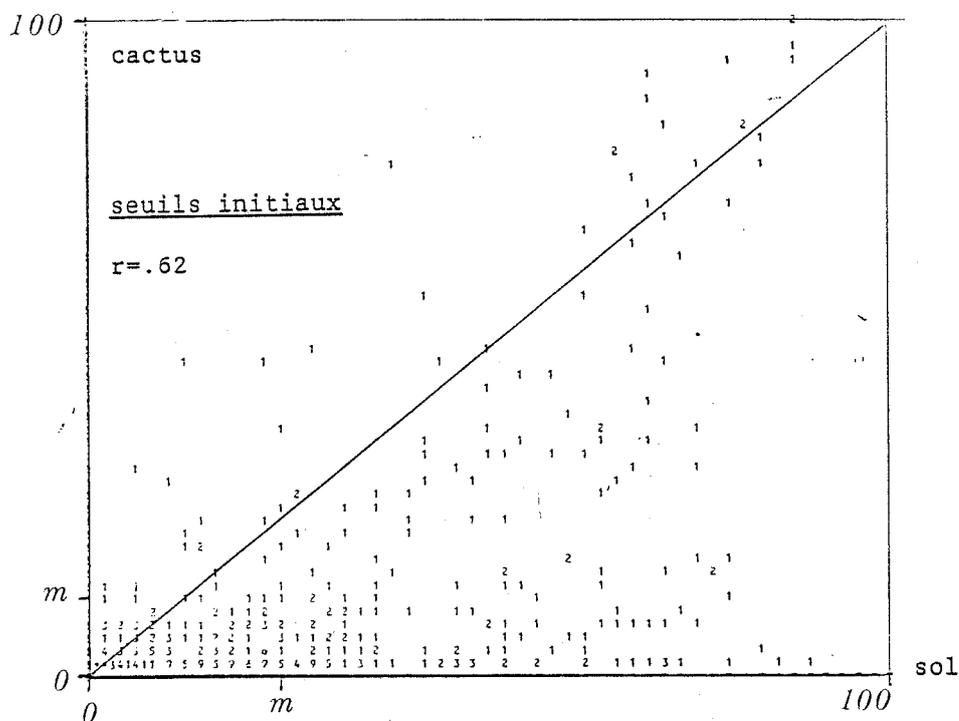


Fig A1-1 : Corrélation entre les valeurs sol et cactus avant le réglage des seuils.

Les seuils choisis ne permettent pas de distinguer le sol d'un nuage, du moins dans certains cas que nous allons devoir déterminer. Cette étude, qui nécessite l'accès aux résultats image par image, a été menée au CMS de Lannion.

#### A1-4-a. Dépouillement des premières observations

Le CMS disposait de certaines sorties horaires simplifiées du programme CACTUS sur une période s'étendant du 15 décembre 1983 au 15 janvier 1984. La comparaison de ces résultats avec les photos NOAA a montré que les résultats étaient assez bons (à bons) à 9h ou 15h TU, et mauvais à 12h TU\*. On peut de plus observer que sur la photo de 12h, les résultats sont assez

(\*) dans la suite du texte, toutes les heures sont des heures TU.

bons sur la mer et mauvais sur la terre, où un ciel totalement dégagé est estimé à 8 (100 % de nébulosité) par le programme CACTUS.

Ces observations montrent que le seuil semble adapté lorsque le soleil est bas (9h et 15h en hiver), mais trop faible pour un soleil haut (12h). La modification apportée a été faite sur la valeur du seuil, et non sur les classes d'angles.

Remarque : Les seuils initiaux (pour soleil haut) étaient effectivement supérieurs à l'albédo de la mer (où ils donnent des résultats corrects), mais inférieurs à celui de la terre, ce que l'on va s'appliquer à corriger en les réévaluant.

#### A1-4-b. Choix de journées intéressantes

##### - Type de journée

Pour trouver le seuil, suffisamment haut pour qu'il n'y ait pas de confusion avec le sol, une journée claire était idéale. Pour que cela soit vrai pour toutes les classes d'angle solaire, il fallait que la zone dégagée s'étende sur toute la zone étudiée (France métropolitaine) et toute la journée si possible. D'autre part, pour étudier la précision du traitement, il fallait une journée ayant une limite nette nuage/clair, et du point de vue photographique une bonne résolution.

Remarque: Nous n'avons pas voulu tester le programme sur des journées de nébulosité difficile à apprécier "à l'oeil" : nuages type cumulus, limite floue, ou résolution inférieure à la maille, dans un premier temps.

##### - Période de l'année

N'ayant pas d'idée sur la saisonnalisation des seuils, ceux-ci ont été déterminés sur une période précise : le mois d'avril. En effet, sept journées d'avril avaient déjà été traitées avec les anciens seuils et l'amélioration pouvaient ainsi être mieux perçue. L'adaptation de ces seuils aux autres saisons sera testée dans un deuxième temps, à l'aide

d'une seule image horaire de décembre, puis plus systématiquement, lorsque les valeurs journalières seront disponibles.

- Journées retenues

Ce sont les journées suivantes :

13 et 17 avril 1984 : deux journées totalement dégagées

7 et 8 avril 1984 : deux journées partiellement couvertes.

28 décembre 1984 : limite nette entre les zones claire et couverte

**A1-4-c. Résultats observés après modification des seuils**

Les seuils ont été choisis de plus en plus haut, jusqu'à atteindre une correspondance jugée bonne entre l'image simplifiée CACTUS et la photo NOAA. Les différentes valeurs de seuils testées, ainsi que les améliorations apportées sur les chaque image, sont décrites dans la note S.GALLE, 1984.

Nous avons ensuite observé les performances de ces seuils sur la journée de décembre. Les résultats montrent encore une légère surestimation de la couverture nuageuse. Cette surestimation est cependant trop délicate à juger pour justifier la modification des seuils, elle devra être vérifiée sur un échantillon plus large.

En conclusion, il semble, d'après l'examen des images instantanées, que les nouveaux seuils choisis soient valables pour l'ensemble de l'année. En effet, une partie de l'influence astronomique de la saison est prise en compte avec l'angle solaire : pour l'image de 12h par exemple, un seul seuil intervient en avril (soleil haut) mais deux seuils entrent en jeu en décembre - ce qui a pour effet de relever automatiquement le seuil, sur une partie de l'image de midi, en décembre.

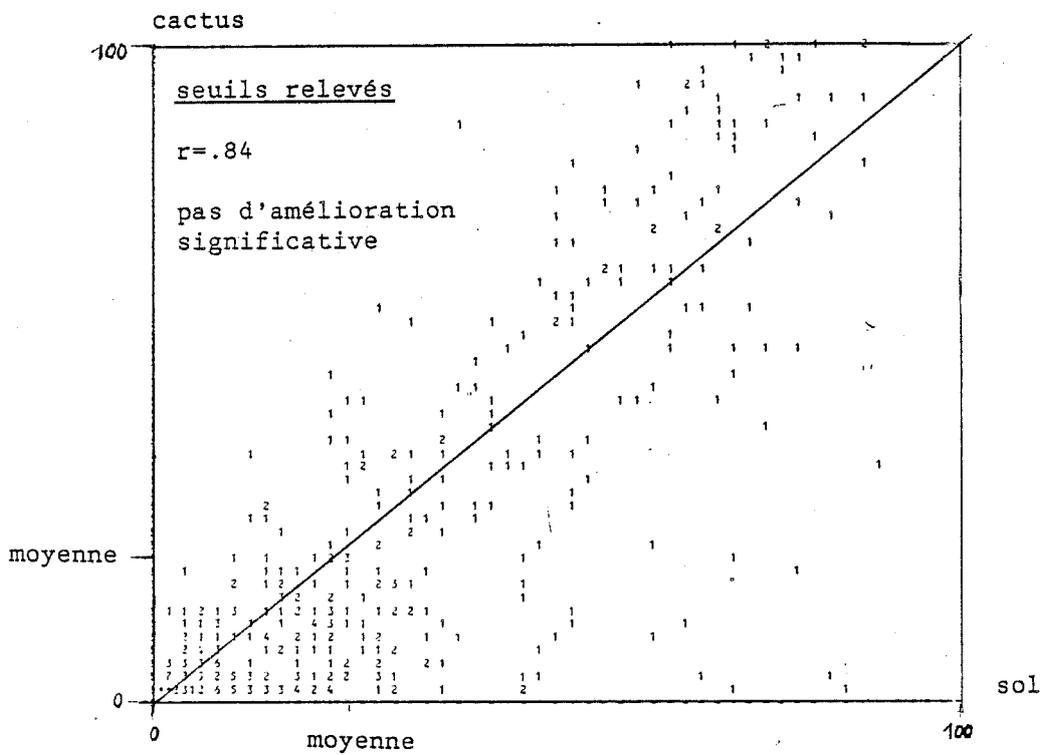
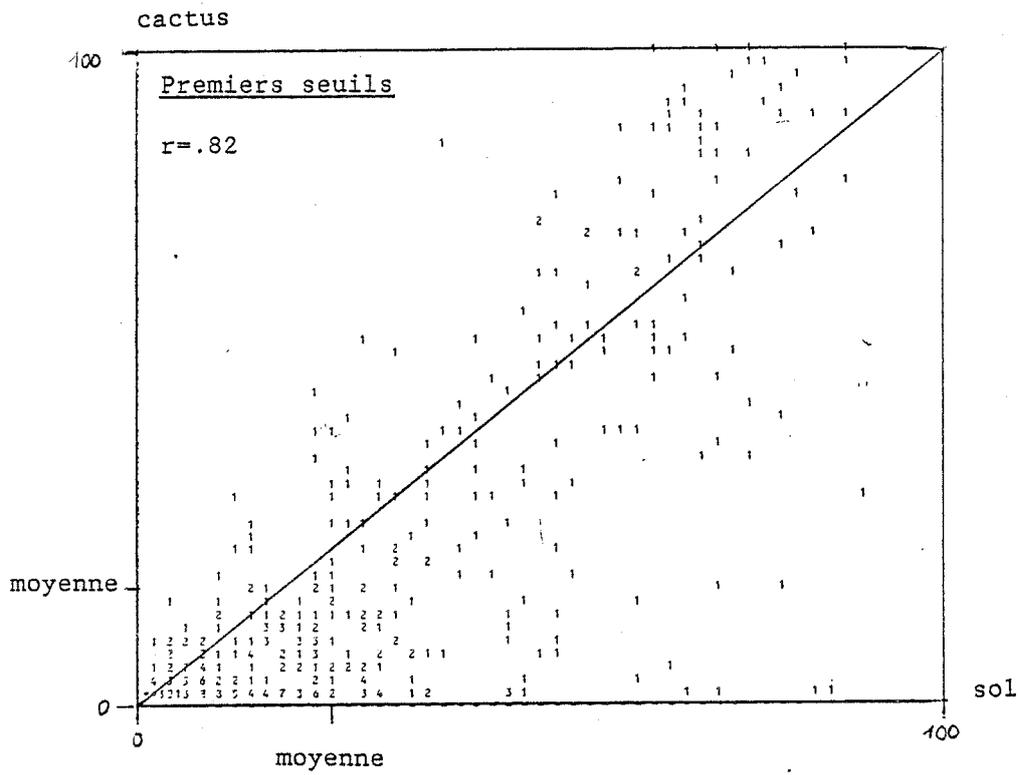


Fig A1-2 : Nuage de corrélation entre les données sol et cactus pour différents valeurs des seuils de celui ci.

#### A1-5. ESTIMATION DES VALEURS JOURNALIERES APRES MODIFICATION

Le réglage des seuils effectué image par image, a ensuite été testé sur les valeurs quotidiennes du rapport d'insolation. La corrélation entre les mesures sol et les estimations CACTUS a été calculée pour les sept journées d'avril 1983 précédemment disponibles (Fig.A1-2). La modification des seuils a amélioré le coefficient de corrélation de .1 pour chaque journée en moyenne. Le coefficient de corrélation global sur les sept journées passe de .62 à .82, soit 67% de variance expliquée au lieu de 38%. La comparaison des figures 1 et 2 montre un regroupement du nuage et surtout la quasi-disparition des points du bas du diagramme. Les quelques points persistants dans cette région correspondent, après vérification, aux stations situées en altitude et enneigées (ou entourées de sommets enneigés). Il ne s'agit plus d'un problème de seuil mais de détection de la neige. Un essai effectué avec des seuils plus élevés (Fig.A1-3) n'a pas apporté de modifications significatives (le coefficient de corrélation passe de .82 à .84).

#### A1-6. CONCLUSION

Le réglage des seuils du programme CACTUS était indispensable à l'amélioration de la qualité de ses résultats : celui-ci surestimait nettement la couverture nuageuse. La détermination des nouveaux seuils a été faite par comparaison des images CACTUS instantanées aux clichés NOAA correspondants. Les seuils ont été réglés sur le mois d'avril et ont donné des résultats satisfaisants pour une journée du mois de décembre. Ces résultats devront par la suite être validés de façon robuste, sur un échantillon significatif.

Un premier test sur les estimations journalières de sept journées d'avril 1983 a montré une amélioration significative de l'estimation CACTUS : il n'y a plus de sous-estimation systématique. Le fait de relever encore les seuils n'a pas apporté de modification sensible des estimations. Ce sont ces seuils qui ont été utilisés pour calculer les données CACTUS analysées dans cette étude.

## ANNEXE 2

## LA PROJECTION CONIQUE CONFORME DE LAMBERT

## A2-1. DEFINITION D'UNE PROJECTION CONFORME

La terre est une géoïde que l'on approche par un ellipsoïde ; chaque pays a défini un ellipsoïde de référence. La triangulation française a été calculée sur l'ellipsoïde de CLARKE 1880. Tout point de l'ellipsoïde est repéré par ses coordonnées géographiques : latitude  $\varphi$  et longitude  $\lambda$ . Une projection permet une correspondance analytique biunivoque entre chaque point de l'ellipsoïde et le point homologue du plan (x,y). Une projection introduit des déformations qui altèrent une partie de la surface à représenter. On classe les projections en fonction des altérations qu'elles entraînent :

projection conforme : elle conserve les angles élémentaires. L'échelle est la même dans toutes les directions au voisinage d'un point, mais celle-ci varie dans le plan.

projection équivalente : elle conserve les surfaces. L'échelle est variable autour d'un point suivant la direction considérée.

Le choix d'un système de projection s'effectue en fonction de l'échelle et du contenu de la carte à produire. Les cartes modernes sont généralement établies dans des systèmes de projection conformes, car les mesures angulaires sont les plus courantes (topographes, géomètres, météorologues).

## A2-2. LA PROJECTION CONIQUE CONFORME DE LAMBERT

La surface du globe terrestre est projetée sur un cône, admettant pour axe de symétrie l'axe des pôles (Fig.A2-1). La projection conique conforme de Lambert possède les propriétés géométriques suivantes :

- les méridiens sont des droites concourantes
- les parallèles sont des arcs de cercle.

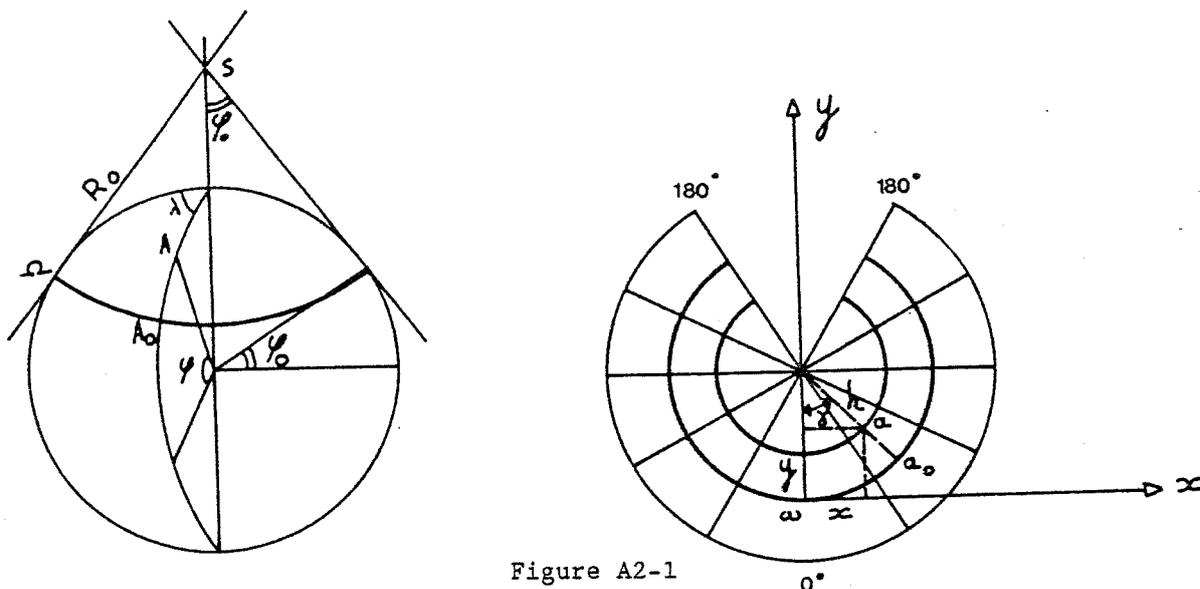


Figure A2-1

Le cône peut être choisi tangent (Fig.A2-1) ou sécant à l'ellipsoïde. Suivant le cas, l'échelle sera identique entre l'ellipsoïde et la projection, pour un ou deux parallèles dits d'"échelle conservée". Pour les autres parallèles, l'altération  $e$  s'exprime en fonction de la seule latitude  $\varphi$  et des parallèles d'échelle conservée (voir plus loin). C'est pourquoi une projection à cône sécant permet de cartographier des surfaces étendues avec une altération moindre qu'un cône tangent.

Calcul des coordonnées en projection Lambert

Tout point a du plan est défini par ses coordonnées rectangulaires (x,y) ou polaires ( $\gamma$ , R). Pour une projection à cône tangent au parallèle  $\varphi_0$  il vient :

$$\begin{cases} \gamma = (\lambda - \lambda_0) \sin \varphi_0 \\ R = f(\varphi) \end{cases}$$

La fonction f est définie par la condition de conformité qui impose :

$$f(\varphi) = R_T \cotg \varphi_0 \left( \frac{\Phi}{\Phi_0} \right)^{\sin \varphi_0}$$

où 
$$\Phi = \frac{\cos \varphi}{1 + \sin \varphi} = \operatorname{tg} \left( \frac{\pi}{4} - \frac{\varphi}{2} \right)$$

Dans le cas d'une projection conique conforme, sécante aux parallèles  $\varphi_1$  et  $\varphi_2$ , ces formules deviennent :

coordonnées polaires 
$$\begin{cases} \gamma = (\lambda - \lambda_0) C \\ R = R_T \frac{\cos \varphi_1}{C} \left( \frac{\Phi}{\Phi_1} \right)^C = R_T \frac{\cos \varphi_2}{C} \left( \frac{\Phi}{\Phi_2} \right)^C \end{cases}$$

coordonnées cartésiennes 
$$\begin{cases} X = R \sin \gamma \\ Y = R_T \cotg \varphi_0 - R \cos \gamma \end{cases}$$

Le paramètre C s'exprime en fonction des parallèles  $\varphi_1$  et  $\varphi_2$  :

$$C = \frac{\text{Log} \left( \frac{\cos \varphi_1}{\cos \varphi_2} \right)}{\text{Log} \left( \frac{\cos \varphi_1 + 1 + \sin \varphi_2}{\cos \varphi_2 + 1 + \sin \varphi_1} \right)}$$

On peut montrer que lorsque  $\varphi_1$  tend vers  $\varphi_2$ , C tend vers  $\sin \varphi_2$  ; on retrouve alors les formules de la projection sur un cône tangent à un parallèle ( $\varphi_0$ ).

#### Calcul des altérations d'une projection Lambert

L'altération  $e$  est définie par  $e = \frac{ab-AB}{AB}$ . Elle caractérise la

différence de longueur entre une distance réelle AB mesurée sur l'ellipsoïde et la distance correspondante ab sur la projection. On la calcule en tout point de latitude :

$$e = 1 - \left( \frac{\phi}{\phi_1} \right)^C \left( \frac{\cos \varphi_1}{\cos \varphi} \right)$$

#### A2-3. LA PROJECTION LAMBERT 30-60

Il s'agit d'une projection sur un cône sécant aux parallèles  $\varphi_1 = 30^\circ\text{N}$  et  $\varphi_2 = 60^\circ\text{N}$ . Le méridien de référence  $\lambda_0$  est celui de Paris ( $\lambda_0 = 2^\circ 20' 14''$ ). La constance C vaut 0.717.

Remarque : La projection habituellement utilisée par l'IGN pour cartographier la France est la projection "Lambert II étendue", c'est-à-dire une projection sur un cône tangent au parallèle  $52 \text{ gr} = 46.8^\circ$  (pratiquement au centre de la France). Le méridien de référence est celui de Paris et C vaut 0.729. La différence entre les projections Lambert 30-60 et Lambert II étendue ne dépasse pas 5 km pour des points situés en France.

La projection Lambert 30-60 utilisée dans cette étude est bien adaptée à la cartographie de vastes zones comme l'Europe.

## ANNEXE 3

## Positionnement de la grille calculée par CACTUS

La grille de valeurs, calculée par le programme CACTUS, doit être située de façon précise, afin de justifier sa haute résolution spatiale. La navigation des images de Météosat effectuée par CACTUS est vérifiée à l'aide de la carte d'albédo mise à jour régulièrement : le contour des côtes ainsi visualisé est superposé à une carte de France dans la même projection (Lambert 30-60). Notre but est de situer non pas la carte d'albédo mais de nébulosité ; l'algorithme d'attribution des pixels Météosat à des mailles étant identique pour les cartes d'albédo et d'ennuagement, il suffit de vérifier la position de l'une d'entre elles.

## Caractéristiques de la grille CACTUS

Le nombre de pas en x et en y est de :  $NX = 60$  ;  $NY = 50$   
 La longueur du pas est de :  $DX = 20$  km ;  $DY = 20$  km  
 Le point bas droit se situe en :  $XO = 10^\circ E$  ;  $YO = 41^\circ N$   
 soit en coordonnées Lambert 30-60 :  $XO = 1245$  km ;  $YO = - 115$  km

## Premiers essais de superposition

En utilisant les caractéristiques ci-dessus, on a superposé la carte d'albédo de juin 84 et le fond de carte Lambert 30-60 (Fig.A3-1). L'examen de cette figure montre un double problème de translation, et d'échelle. Ce problème, vu son importance, doit être attribué à l'algorithme utilisé

Fig A3-1 : Carte d'albédo du 29/6/84

Image initiale  
qui reflète  
un problème de  
translation et  
d'échelle

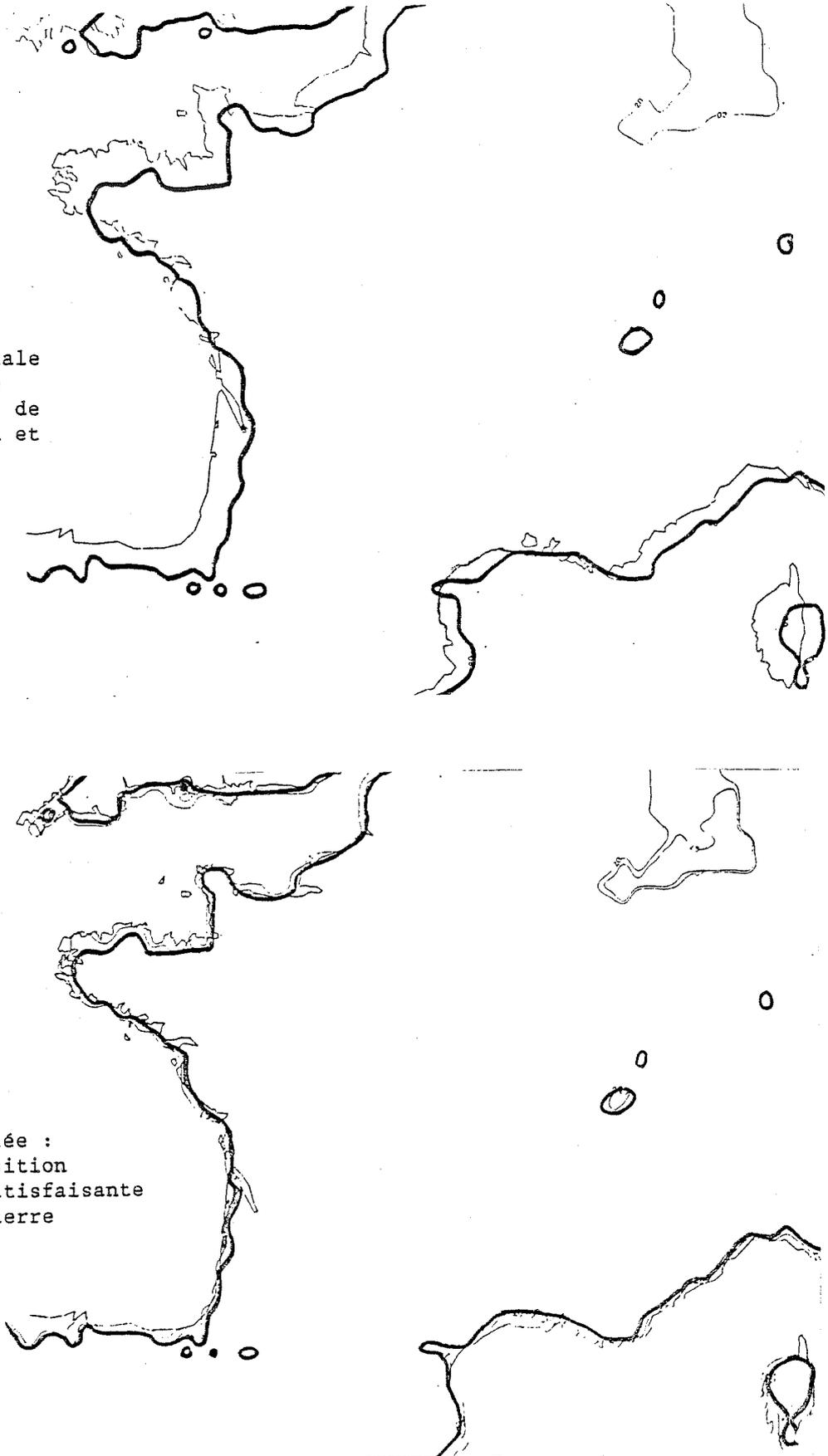


Image recalée :  
la superposition  
est très satisfaisante  
de l'Angleterre  
à la Corse.

dans le programme CACTUS et, au deuxième ordre, à la navigation du satellite.

#### Démarche proposée

Devant le problème rencontré, nous avons tout d'abord essayé de corriger nous-même, de façon simple, la translation géographique des images CACTUS, qui semblait assez systématique. Le premier problème était celui de l'échelle et le deuxième, celui de la translation.

#### Détermination de l'échelle

Une façon simple de modifier l'échelle de la carte obtenue est de faire varier la valeur du pas de la maille (20 km). C'est ce que nous avons fait pour cinq valeurs du pas variant de 20 à 21.2 km. Les cartes d'albédo ainsi tracées ont été superposées au fond de carte de France, disponible sur un transparent. La meilleure superposition est obtenue pour le pas  $DX=DY=20.8$  km. Ce résultat reste vrai pour les 11 cartes d'albédo des 11 mois disponibles. Il faut ici remarquer que la présence de la Corse est un facteur déterminant dans la précision de l'estimation. En effet, une fois la Bretagne ajustée, l'adéquation sur la Côte d'Azur semble toujours assez correcte, mais on remarque aisément la Corse "dans l'eau". C'est pourquoi notre fond de carte comprenait la Corse et l'Angleterre.

#### Détermination de la translation

Une fois l'échelle déterminée, grâce à un calque, on a cherché l'origine de la grille, telle que la carte d'albédo et le fond de carte se superposent sur le même support. Différents essais nous ont amené à choisir la nouvelle origine du point bas droit :  $XO = 1212$  km,  $YO = -90$  km, soit une translation de (33,25) km, par rapport à la position théorique. La précision est estimée à  $\pm 5$  km (soit environ un pixel de Météosat).

Cette origine et ce pas ont été testés sur les dix autres mois et ont donné des résultats satisfaisants . Ils semblent donc assez sûrs et on les gardera pour toutes nos grilles.

Carte d'albédo cumulée : 2 exemples qui nécessitent une vérification des cartes journalières.

Fig A3-2 :

Le décalage d'une image vers le nord a entraîné la diminution de la Corse et de la côte d'azur.

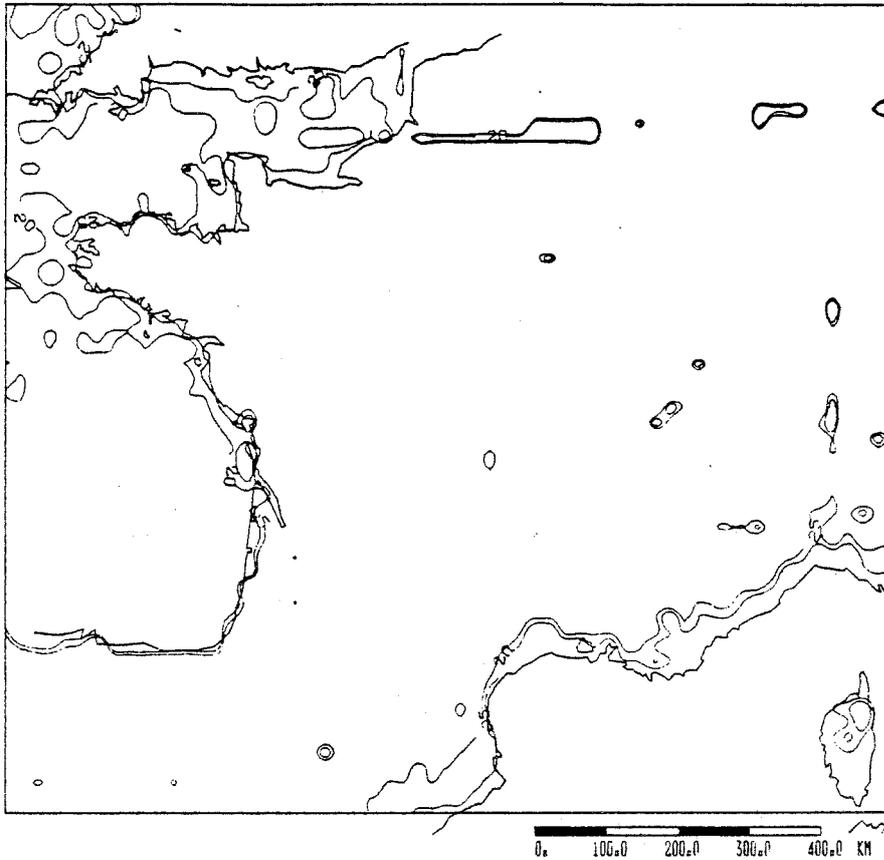
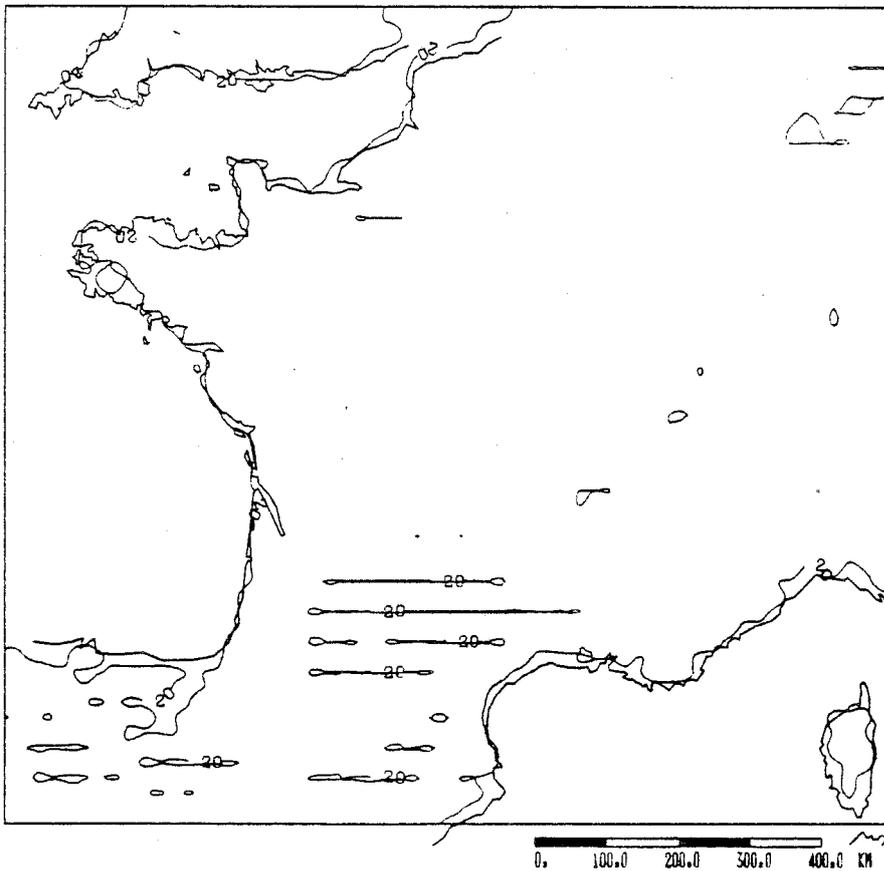


Fig A3-3 :

Une image a présenté des parasites dans le balayage E/W.



#### Remarque sur les cartes d'albédo :

Le fait de cumuler les images d'un jour sur l'autre améliore la carte finale, en sélectionnant les valeurs les plus basses de l'albédo. Mais si pour une raison quelconque, il est un jour apparu une valeur anormalement basse en une maille ou sur une région, celle-ci restera sur la carte jusqu'à la fin du mois. Ainsi, par exemple, si une image ou une partie d'image se trouve décalée vers le nord, la mer, d'albédo plus bas que la terre, va "mordre" sur la terre et les valeurs d'albédo seront modifiées partout où a eu lieu cette superposition. On aura donc une côte "entamée", et ceci jusqu'à la fin du mois considéré.

Ceci expliquerait, par exemple, la diminution de la Corse et de la côte d'azur que l'on peut remarquer sur certaines images comme celle du 29.03.84, isovaleurs 20 et 30 (Fig. A3-2 )

D'autre part, grâce à ce phénomène, en regardant l'albédo du dernier jour seulement, on peut détecter si il y a eu des parasites (Fig.A3-3 ) ou un incident dans le positionnement de l'image satellite. Si on détecte une erreur, on remonte alors les jours pour trouver celui qui a provoqué l'anomalie et on peut alors l'éliminer. L'absence d'erreur sur la carte d'albédo du dernier jour ne veut par contre rien dire.

#### Conclusion

Après une translation et un changement d'échelle, les cartes d'albédo sont superposables avec le fond de carte de France en Lambert 30-60. Ce nouveau positionnement de la carte est estimé à  $\pm 5$  km, une meilleure qualité de localisation se traduit par une amélioration de la coïncidence avec les valeurs sol (Fig.A3-4) : on compare aux stations sol des mailles qui leur sont plus proches.

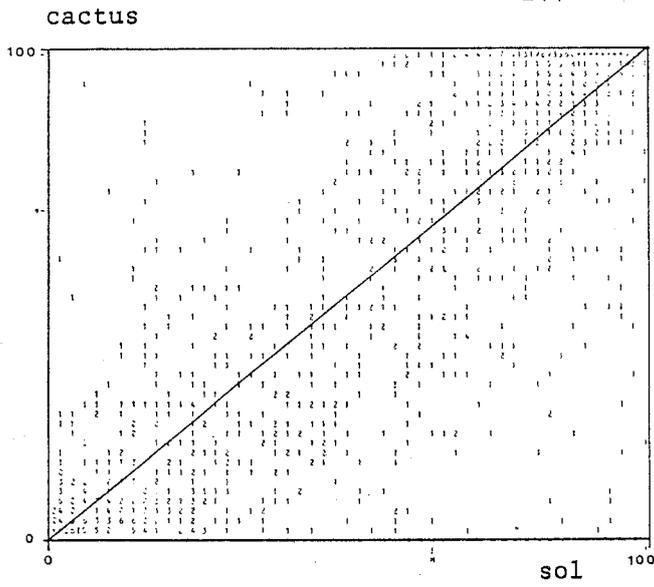
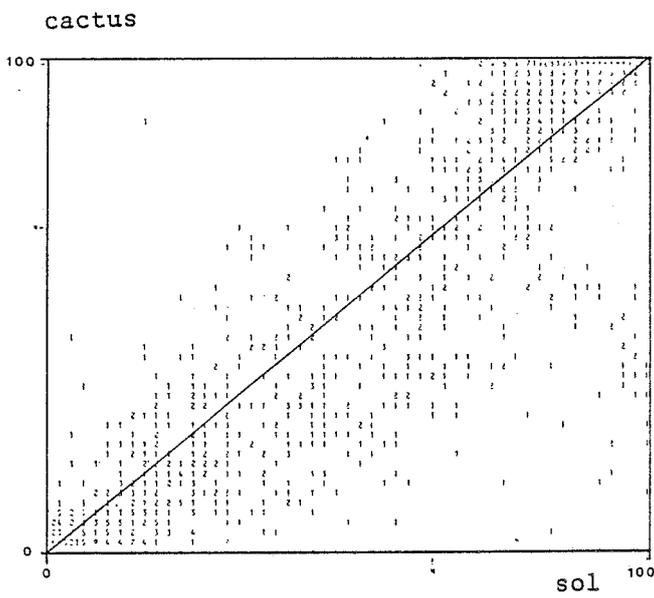
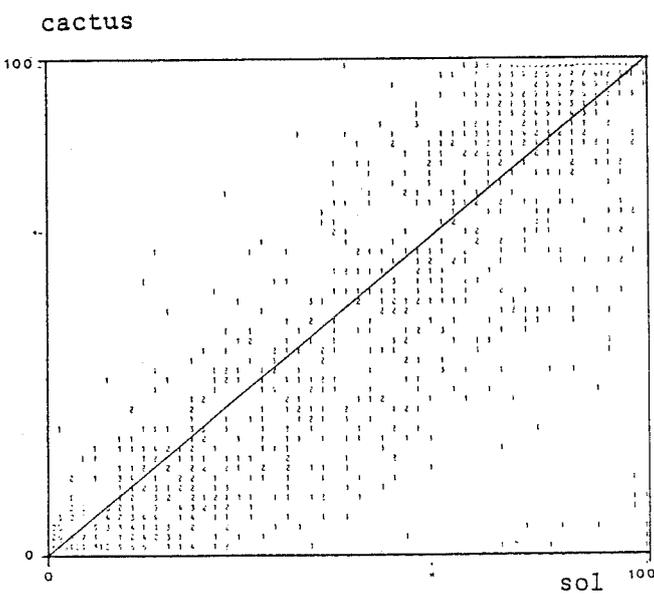


Fig A3-4 :  
Correlation entre les  
rapports d'insolation  
mesurés par le sol et  
estimés par cactus  
(avril 84, 85 stations,  
27 jours)

Avant recalage  
 $r = .91$



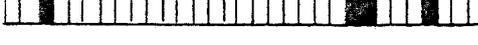
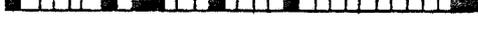
Translation de (10,10 km)  
 $r = .93$



Position définitive  
 $r = .95$

## ANNEXE 4

## LISTE DES 200 JOURNEES DISPONIBLES EN 1984

MOIS	Tableau des jours disponibles	Total disponible	PERIODE	effectif
septembre		7	"hiver"	67
octobre		8		
novembre		8		
décembre		0		
janvier		20	"printemps"	71
février		24		
mars		20		
avril		24	"été"	62
mai		27		
juin		19		
juillet		20		
août		23		
				200

□ jour disponible

■ jour absent ou incomplet

## ANNEXE A5

CALCUL DE LA DUREE MAXIMALE D'INSOLATION  $S_0$ 

La durée maximale  $S_0$  est égale à la durée astronomique du jour  $S_J$  (fonction uniquement de la latitude  $\varphi$  et du jour J), diminuée du nombre d'heures où le soleil est masqué par un obstacle (bâtiment ou montagne par exemple).

## A5-1. DUREE ASTRONOMIQUE DU JOUR.

La durée du jour est calculée à partir de la formule classique d'astronomie donnant la hauteur angulaire du soleil h.

$$\sin h = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos w$$

$\varphi$  : latitude du lieu

$\delta$  : déclinaison

w : angle horaire du soleil

(a) Montpellier :  
station sans masque

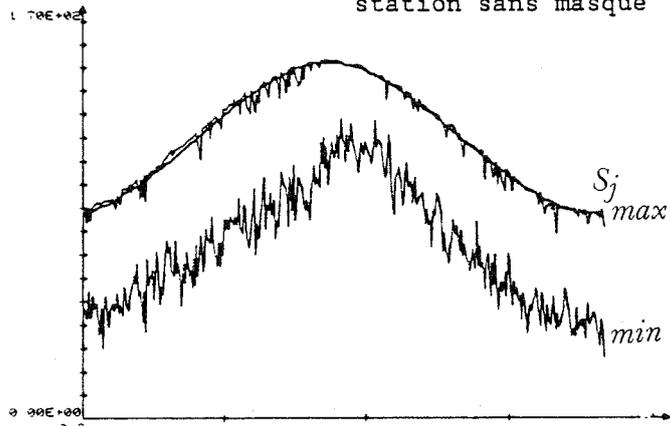
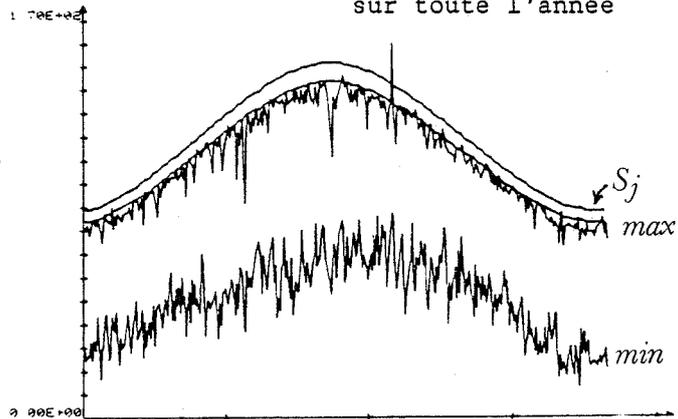


Fig A5-1 :

Estimation du masque

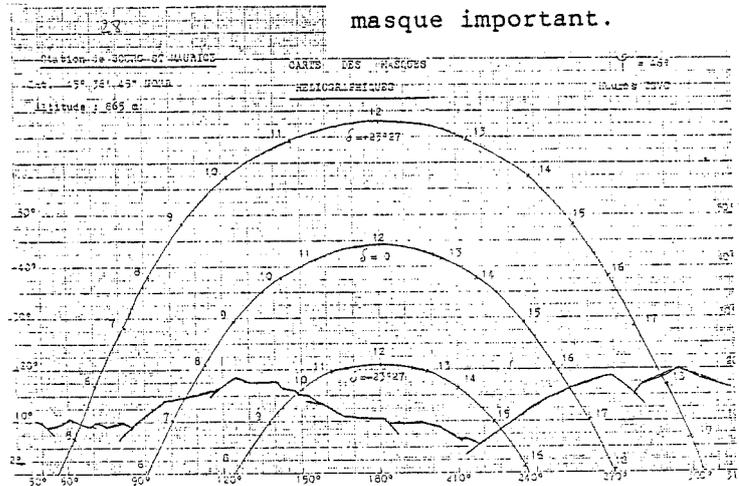
(a) et (b) :  
comparaison de la durée du  
jour théorique  $S_j$ , et de  
la valeur max obtenue  
sur une série de 20 ans.

(b) toulon :  
masque de 5%, régulier  
sur toute l'année



(c) :  
tour d'horizon, indiquant  
la position du soleil pour  
différentes déclinaisons .

(c) Bourg Saint Maurice :  
masque important.



Le soleil se lève ou se couche lorsque  $h = 0$  (soleil à l'horizon) d'où la durée du jour  $S_J$  en heures

$$S_J(\varphi, J) = 2 \times \frac{12}{\pi} \text{Arcos} (-\text{tg } \varphi \text{ tg } \delta)$$

$\delta$  est calculé par la formule approchée :

$$\delta = 23,45^\circ \sin \left( (n-81) \frac{360}{365} \right)$$

$J$  = numéro du jour dans l'année, compté à partir du 1er janvier.

#### A5-2. MASQUE

Le masque dépend de la station. On l'a déterminé pour chacune d'elles et pour chaque décade, en portant plus particulièrement notre attention sur les zones de montagne où il peut être très important.

Principe de mesure du masque : si on dispose d'une série climatologique longue à la station étudiée (supérieure à 20 ans), on cherchera la valeur maximum obtenue sur la série pour chaque décade. On calcule ensuite le rapport à la durée du jour théorique en ce lieu ( $S_J$ ), et on en déduit le masque (en pourcentage), pour chaque décade (Fig.A5-1 a et b). Ces données ont été fournies par S. MARTIN et J.D. CREUTIN.

Si on ne dispose pas d'une série suffisamment longue à la station considérée, il faut alors se référer au tour d'horizon. Celui-ci donne la hauteur en degré des obstacles entourant la station (Fig.A5-1-c: tour d'horizon de la station de bourg Saint maurice).

Le tableau A5-1 donne les valeurs (en pourcentage) des masques utilisés dans cette étude. Le parcours du soleil dans le ciel étant symétrique au cours de l'année, le relevé n'est nécessaire que pour une demi-année. Les masques inférieurs à 5 % n'ont pas été pris en compte dans les calculs.





## REFERENCES

BOIS Ph.

Contribution à l'analyse et à la prévision de variables hydrométéorologiques. Application à la prévision des débits du Niger et des avalanches à Daves. Thèse de Docteur d'Etat, INP Grenoble, 1976.

BERIOT N., 1984

L'évaluation de la ressource énergétique solaire à l'aide de satellites météorologiques. Rapport World Meteorological Organisation. WCP 80.

CHOISNEL et BEDEL, 1984

Disponibilité des données climatiques de référence et des données du temps passé récent. Coll. Météorologie et Energies Renouvelables, Valbonne, 1984.

COUDERT R.

Evaluation de la précision de mesure de la durée d'insolation et du rayonnement solaire global dans un réseau radiométrique. Centre radiométrique de Carpentras. Colloque Météorologie et Energies Renouvelables, Valbonne, mars 1984.

CRADDOCK J.M., 1973

Problems and prospects for eigen vector analysis in meteorology. The Statistician, Vol 20, N°4, p63-72.

CREUTIN J.D., Ch. OBLED, 1980

Elimination de variables et optimisation de réseaux de mesures. Data analysis and informatics E.Diday et al. Ed., 759,775

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.

CREUTIN J.D., Y.FREIN, S.GALLE, P.LACOMBA, C.OBLED, 1984

A simple approach to estimate climatological variables at ungauged points from satellite images and scarce ground data : two case studies about rainfall and sunshine duration. Proceedings of 18th Int. Symp. on Remote Sensing of Environment. Paris, France, oct 1-5 1984

- CREUTIN J.D., GALLE S., 1985  
Estimation de la durée quotidienne d'insolation par utilisation conjointe d'images METEOSAT et de mesures sol : application à la France. 2ème Journées Internationales sur les Transferts Héliothermiques, Rabat, avril 1985.
- CREUTIN J.D., 1987  
Validation et étalonnage d'images de télédétection à l'aide de mesures ponctuelles selon une approche géostatistique. Applications à la mesure des précipitations et de l'insolation. Thèse d'Etat, INP Grenoble.
- CUENIN R., 1972  
Cartographie générale. Tome 1, Collection Scientifique de l'IGN, Ed. Eyrolles.
- DAVENPORT B., ROOT W.L.  
Random signals and noise. Mac Graw Hill, 1958
- 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.
- DELHOMME J.P., 1976  
Application de la théorie des variables régionalisées dans les sciences de l'eau. Thèse D.I. Université de Paris VI.
- DELORME Cl., 1984  
Le rayonnement solaire utilisable. Tome I (Université de Compiègne).
- DELRIEU G., 1986  
Evaluation d'un radar météorologique pour la mesure des précipitations : validation et étalonnage par technique géostatistique. Application au Bassin Parisien. Thèse USM et INP Grenoble, IMG, 200 pp.
- DUBRULE O., 1982  
Krigeage et splines en cartographie automatique. Thèse de l'Ecole des Mines de Paris, 141 pp.
- FORTIN M.  
Recalage d'image multi-sources : application aux images de télédétection . Thèse de 3ème cycle, Paris VI, sept. 1983.
- GALLE S., 1986  
Spatial structure of meteosat and ground measurements of daily sunshine ratio : some aspects of their good spatial agreement. 6th Meteosat user's Meeting. Amsterdam, The Netherlands, nov 25-27 1986

- GALLE S., CREUTIN J.D., 1986  
Ground network optimisation using satellite information :  
application to the French heliographic network.  
Int. Geoscience and Remote Sensing Symposium IGARSS'86  
Zurich, Switzerland, sept 8-11 1986  
Special issue of IEEE transaction on GRS.
- GALLE S., 1984  
Etude expérimentale des seuils dans le programme CACTUS.  
Rapport de la mission effectuée au Centre de Météorologie  
Spatiale de Lannion. Avril 1983
- GALLE S., 1983  
Recalage géographique des images CACTUS. Note pour la Météorologie  
Nationale
- GAMBIER G., 1984  
Notions sur les représentations planes de la terre
- GANDIN L.S., 1965  
Objective analysis of meteorological fields. Israel Program for  
Scientific Translation, Jerusalem, 242 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.
- Introduction to the METEOSAT system  
ESA SP.1041, november 1981.
- JOURNEL A., HULJBRECHTS Ch.J., 1978  
Mining geostatistics. Academic Press, pp. 600.
- KUENY J.L.  
Contribution au traitement statistique de données météorologiques.  
Thèse de D.I., INP Grenoble, 1977, p.70-80.
- LACOMBA P., 1986  
Evaluation des précipitations par combinaison d'images satellitaires  
visible et infra-rouge et de réseau de pluviomètres : applications à  
la péninsule arabe et au sud-ouest de la France. Thèse de  
Docteur-Ingénieur de l'USTMG et l'INPG, Grenoble, 217 pp.
- LEBART et FENELON  
Statistique et informatique appliquées. 2ème Edition Dunod, 1973.
- MARQUES M., 1982  
Conception d'un modèle stochastique de simulation des rayonnements  
solaires direct et global à un pas de temps fin. Application aux  
données de Grenoble. Thèse de 3ème cycle USM et INP Grenoble.

- MATHERON G., 1965**  
Les variables régionalisées et leur estimation. Masson Ed., Paris, 305 pp.
- MEJON M.J., 1979**  
Etude des liaisons entre l'irradiation journalière et les variables météorologiques courantes. Contribution à l'analyse et à la simulation de données d'éclairement direct à pas de temps fin. Thèse de 3ème cycle USM et INP Grenoble.
- MYERS D.E., 1982**  
Matrix formulation of co-kriging. Mathematical Geology, 14, 249-257.
- OBLED C., CREUTIN J.D., 1986**  
Some developments in the use of Empirical Orthogonal Functions for mapping Meteorological fields. JCAM, sept. 1986.
- OBLED C., 1979**  
Contribution à l'analyse des données en Hydrométéorologie. La prévision des phénomènes accidentels et l'analyse des champs spatiaux. Thèse de Docteur d'Etat, INP Grenoble, 1979.
- OBLED C.**  
Recherche de facteurs significatifs dans une ACP. Recette pratiques et résultats théoriques utiles. (à paraître)
- PERRIN DE BRICHAMBAUD Ch.**  
Le gisement solaire. Edition Lavoisier Tec et Doc, 1982.
- VENTSEL H., 1973**  
Théorie des Probabilités. Edition de Moscou, p.439-446.
- VIALART J., 1956**  
Calcul des probabilités et statistiques.  
Direction de la Météorologie Nationale, tome III, p. 49-52.
- WACKERNAGEL H., 1985**  
L'inférence d'un modèle linéaire en géostatistique multi-variable. Thèse ENSM de Paris, Centre de Géostatistique, 100 pp.

## EVALUATION DE L'INSOLATION A PARTIR DE DONNEES SATELLITAIRES

## \_\_\_\_\_ Méthode ELLIS et VONDER HAAR (1978)

ELLIS J.S., VONDER HAAR T.H., 1978

Solar radiation reaching the ground determined from meteorological satellite data. Proceedings of the Third Conference on Atmospheric Radiation, Davis (Cal.), June 28-30 1978.

VONDER HAAR T.H., ELLIS J.S., 1978

Determination of solar energy microclimate of the United States using satellite data. University of Colorado. Final Report NASA Grant NAS5 22372.

## \_\_\_\_\_ Méthode de TARPLEY (1979)

TARPLEY J.D., 1979

Estimating incident solar radiation at the surface from geostationary satellite data. J.A.M. Vol 18, N°9.

TARPLEY J.D., 1981

Satellite derived insolation for agriculture : an up date on the NESS program. Proceedings of the First Workshop on Terrestrial Solar Resource Forecasting and on Use of Satellite for Terrestrial Solar Resource Assessment, Washington D.C., February 2-5 1981.

LEDUC S., 1984

Evaluation and use of satellite products for agricultural assessments. Int. Symp. on remote sensing of environment, octobre 1984.

HANSON K., 1971

Studies of cloud and satellite parameterization of solar irradiance at the earth's surface. Proceedings of the Miami Workshop on Remote Sensing, Miami (FL).

## \_\_\_\_\_ Méthode de GAUTHIER, DIAK, MASSE (1980)

GAUTHIER C., DIAK G., MASSE S., 1980

A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. J.A.M., Vol 19, N°8.

DIAK G., GAUTHIER C., MASSE S., 1982

An operational system for mapping insolation from GOES satellite data". Solar Energy, Vol 28, N°5.

DIKAK G., GAUTHIER C., 1983

Improvement to a simple physical model for estimating insolation from GOES data. JCAM, Vol 22, N°3, March 83.

GAUTHIER C., DIKAK G., MASSE S., 1984

An investigation of the effects of spatially averaging satellite brightness measurements on the calculation of insolation. JCAM, Vol 23, N°9, September 1984.

DIKAK G., WEICKMANN A., 1985

Solar energy in Wisconsin. One year survey from Satellite data for Wisconsin power and light. Report from the Space Science and Engineering Center. The University of Madison.

\_\_\_\_\_ Méthode de MOSER et RASCHKE (1981)

MOSER W., RASCHKE E., 1983

Mapping of global radiation and of cloudiness from METEOSAT image data. Theory and ground truth comparisons. Meteorol. Rdsch. 36, 33-41.

TUZET A., MOSER W., RASCHKE E., 1984

Estimating global solar radiation at the surface from METEOSAT data in the Sahel region. J. Rech. Atmos. 18 (1984), N°1, p.31-39.

MOSER W., RASCHKE E., 1983

Incident solar radiation over Europe estimated from METEOSAT data. JCAM, Vol 23, 1984, p.166-170.

\_\_\_\_\_ Méthode de PASTRE (1981)

PASTRE C., 1981

Développement d'une méthode de détermination du rayonnement solaire global à partir des données METEOSAT. La Météorologie, VIème série, N°24.

\_\_\_\_\_ Méthode de REZKI (1982)

REZKI, 1982

Reconstitutions de l'insolation journalière à l'aide d'un an d'information du satellite METEOSAT 1. Thèse de 3ème cycle USM et INP Grenoble, juin 1982.

---

Méthode DEDIEU, DECHAMPS, KERR (1983)

DEDIEU, 1984

Estimation du rayonnement solaire global au sol et de l'albédo de surface à l'aide de METEOSAT. Thèse de 3ème cycle. Université de Toulouse, mai 1984.

---

Méthode DELORME, AMADO, RABERANTO (1982)

AMADO J., 1982

Cartographie de l'albédo et du gisement solaire à partir des données du satellite METEOSAT. Thèse D.I. Université de Toulouse, mars 1982.

DELORME Cl., AMADO J., RABERANTO P., 1983

The use of METEOSAT for solar radiation mapping. Meteorol. Rdsch - 36.

DELORME Cl., RABERANTO P., 1984

Quelques causes d'incertitude sur le rayonnement solaire à partir de METEOSAT. Col. Météorologie et Energies Renouvelables, Valbonne, m1984, p.465-475.

RABERANTO P., 1984

Etude critique de l'albédo minimal déterminé à partir des images numériques de METEOSAT. Thèse Université de Toulouse, 1984.

---

Méthode MONGET, CANO (1982)

CANO D., 1982

Etude de l'ennuagement par analyse de séquences d'images du satellite. Application à l'évaluation du rayonnement solaire global au sol. Thèse D.I. ENST, novembre 1982.

MONGET J.M., 1984

Global solar radiation mapping from satellite an European Research Program. Col. Météorologie et Energies Renouvelables, Valbonne, Mars 84.

THESE de DOCTORAT de  
l'Institut National Polytechnique de Grenoble  
Spécialité : Mécanique

*Titre de l'ouvrage :*

**ANALYSE DES CHAMPS SPATIAUX  
PAR UTILISATION DE LA TÉLÉDÉTECTION**  
Estimation de la durée quotidienne d'insolation en France  
à l'aide d'images du satellite Météosat  
et de mesures sol

*Nom de l'auteur :* Sylvie GALLE

*Établissement :* Institut de Mécanique de Grenoble

### RÉSUMÉ

L'insolation est étudiée, à l'échelle synoptique (la France) à l'aide d'un ensemble de mesures de la durée quotidienne d'insolation réalisées au sol et par télédétection satellitaire. L'objectif poursuivi est d'analyser une série de données suffisamment longue pour caractériser de manière statistique le comportement climatologique de l'insolation dans l'espace.

Les données sont fournies par 88 héliographes du réseau synoptique français ainsi que par le programme « Cactus » de la Météorologie Nationale, qui utilise les images visibles du satellite Météosat. L'étude porte sur une année de données.

La similarité de la structure spatiale des mesures du sol et du satellite est montrée par analyse objective, et par analyse en composantes principales de processus. La fine perception de la variabilité spatiale du satellite est alors utilisée pour optimiser un réseau de mesures au sol.

Les performances du programme Cactus sont comparées avec celles de méthodes d'interpolation classiques (basées sur 23 héliographes), pour l'estimation du rapport d'insolation. Elles se sont révélées plus performantes en terme de cofluctuation. Il apparaît cependant une tendance du satellite à exagérer les valeurs extrêmes.

L'étalonnage des mesures satellitaires par des mesures sol a pour but de conjuguer les qualités de précision des héliographes, et la définition spatiale du satellite. Une approche géostatistique est proposée. L'effet recherché est observé à partir d'une certaine densité du réseau d'étalonnage.

### MOTS-CLÉS

Champs spatiaux ; Télédétection ; Météosat ; Insolation ; Étalonnage ; Géostatistique ; Analyse objective ; Analyse en composantes principales de processus.