

REMERCIEMENTS

Je tiens à remercier tout particulièrement Denis Allard, qui m'a encadré avec patience et m'a forcé à (un peu) de rigueur et d'organisation.

Je remercie beaucoup Frédéric Baret pour m'avoir fourni un sujet intéressant et pour la semaine de vacances qu'il m'a fait passer à Romilly-sur-Seine, Gérard Douaire et Hervé Nicolas pour nous avoir supporté et avoir maintenu ouverte notre spécialisation.

Je suis très reconnaissant à tous les membres des laboratoires de biométrie, de bioclimatologie et de l'UIC de l'accueil qu'ils m'ont fait (et de la gentillesse des réponses qu'ils ont toujours fait à mes questions). Enfin, c'est évident, merci aux développeurs de Solaris, Emacs, L^AT_EX, ... qui m'ont permis de travailler dans de bonnes conditions.

Je souhaite beaucoup de courage à Marie Weiss pour la suite de ce travail, et je la remercie beaucoup de l'aide qu'elle m'a apporté en télédétection.

Table des matières

1	Introduction	5
1.1	Présentation du cadre	5
1.2	Problématique	6
1.2.1	Stratégie d'échantillonnage	6
1.2.2	Évaluation empirique des lois	8
1.2.3	Définition du modèle géostatistique	9
1.3	Nature des données	10
1.3.1	Les mesures au sol	10
1.3.2	Les données satellites	10
1.4	Zone d'échantillonnage	11
2	Campagne d'échantillonnage sur la zone de Romilly	13
2.1	Echantillonnage mis en œuvre	13
2.1.1	Choix des emplacements de mesure	13
2.1.2	Dispositif expérimental sur une parcelle	15
2.2	Description des données mesurées	15
2.2.1	Localisation des mesures	16
2.2.2	Gestion des croix incomplètes	17
2.3	Présentation de l'image satellite	18
2.3.1	Image SPOT de la zone	18
2.3.2	Classification supervisée	19
3	Analyse statistique et géostatistique des données	21
3.1	Les données LAI	21
3.1.1	Répartition des LAI	21
3.1.2	Quel support considérer?	23
3.2	Les données SPOT	25
3.2.1	Etude des corrélations entre bandes spectrales	25
3.2.2	Etude géostatistique de l'image	26
3.3	Régressions LAI / réflectances	32
3.3.1	Présentation des données	32

3.3.2	Estimation de la régression	33
3.3.3	Régression linéaire	34
3.3.4	Ajustement de modèles non linéaires	36
4	Modèles Géostatistiques	39
4.1	Présentation des différents modèles envisagés	40
4.2	Krigeage ordinaire	42
4.2.1	Modèle de covariance spatiale	42
4.2.2	Cartographie du LAI et de la variance de krigeage	43
4.2.3	Agrégation	44
4.3	Cokrigeage	45
4.3.1	Modèles de covariance nécessaires au cokrigeage	45
4.3.2	Cartographie du LAI et de la variance de krigeage	48
4.3.3	Agrégation	49
4.4	Krigeage des résidus	52
4.4.1	Modèle de covariance spatiale	52
4.4.2	Cartographie du LAI estimé et de la variance de krigeage	59
4.4.3	Agrégation	61
4.5	Discussion sur les techniques d'interpolation pour la cartogra- phie de LAI	63
4.6	Discussion sur les variances de krigeage	65
4.6.1	Erreur d'estimation du LAI	65
4.6.2	Sensibilité des variances de krigeage	67
A	Régressions linéaires	71
A.1	Régression linéaire simple	71
A.2	Régression linéaire généralisée	72
B	Outils géostatistiques	73
B.1	Le variogramme	73
B.1.1	Bases théoriques	73
B.1.2	Calcul des variogrammes expérimentaux	74
B.1.3	Ajustement d'une fonction de covariance	75
B.2	Variogrammes croisés ou covariogrammes	76
C	Présentation des techniques de krigeage	79
C.1	Krigeage de la moyenne	79
C.2	Krigeage ordinaire et krigeage de bloc	80
C.3	Cokrigeage	81

Chapitre 1

Introduction

1.1 Présentation du cadre

Les recherches effectuées lors de ce stage s'inscrivent dans un programme de validation des algorithmes des capteurs à large champ utilisés pour caractériser la biosphère continentale : VALERI, qui réunit des partenaires tels que l'INRA, le CESBIO (CNES), le CIRAD ou l'IRD.

Le lancement de capteurs à large champ tels que VEGETATION, POLDER, MERIS, MODIS... permet d'acquérir des données utilisées pour caractériser la biosphère continentale. Les thématiques concernées sont associées à la connaissance des climats, des écosystèmes et leur évolution, l'influence de l'homme... Les caractéristiques des couverts généralement recherchées à partir de ce type de capteurs sont les suivantes :

1. l'albédo, c'est à dire la fraction du rayonnement d'origine solaire (rayonnement direct et rayonnement diffus) qui est réfléchi par une surface, dans toute l'hémisphère. Pour une surface, il représente la valeur moyenne de sa réflectance spectrale hémisphérique, pour l'ensemble du spectre solaire (de $0,3\mu\text{m}$ à $3,0\mu\text{m}$).
2. La fraction de couverture du sol par la végétation (notée par la suite fCover),
3. l'indice foliaire (noté par la suite LAI pour Leaf Area Index),
4. la fraction du rayonnement photo-synthétique actif absorbé par le couvert (fAPAR),
5. le contenu en chlorophylle du couvert végétal.

L'amélioration de la qualité des capteurs (étalonnage, finesse des bandes, géométrie, prise en compte des paramètres orbitaux...) et les progrès réalisés en algorithmie d'inversion des signaux satellites rendent possible la production

depuis l'espace de champs de variables biophysiques de surface, tels que ceux évoqués précédemment. Ces produits doivent être vérifiés et validés avant de pouvoir être utiles aux études climatiques et environnementales. En raison des multiples erreurs associées à chaque étape de l'algorithme (étalonnage du satellite, élimination des zones de nuages, corrections atmosphériques, normalisation des données en fonction de la direction d'observation...), seule une méthode de validation directe permet de qualifier correctement les produits considérés. L'objectif est donc de comparer des données échantillonnées au sol avec ces produits. Mais comment comparer de rares données sol, au support quasiment ponctuel, avec de très nombreuses données intégrées sur des surfaces de l'ordre du kilomètre carré ?

1.2 Problématique

L'objectif initial du stage était de proposer une méthode de cartographie de LAI à partir d'un échantillonnage au sol et les cartes d'erreurs d'estimation associées. Comme il est difficile d'échantillonner de façon dense une surface proche d'un kilomètre carré (seul un petit nombre d'échantillons est réalisable compte tenu des contraintes temporelles : les mesures doivent être réalisées peu de temps avant ou après la prise des images satellites, surtout si l'on est dans un contexte de croissance rapide des couverts végétaux), on a décidé d'utiliser l'information radiométrique de capteurs tels que SPOT HRV comme source d'information complémentaire. Le passage à la cartographie se fera alors grâce à des méthodes géostatistiques d'interpolation qui s'appuieront sur les zones échantillonnées et, en dehors de celles-ci, sur les informations portées par des images de réflectances à haute résolution. Ainsi la problématique du stage s'articule autour de quatre points :

1. Définir la stratégie d'échantillonnage au sol,
2. Modéliser les liens entre les paramètres biophysiques et les réflectances,
3. Définir le modèle géostatistique adapté pour interpoler,
4. Réaliser un changement de support. En effet, une fois la cartographie réalisée à la résolution SPOT, on veut repasser à une résolution moindre ; celle des capteurs dont on veut valider les produits.

1.2.1 Stratégie d'échantillonnage

Nos échantillons doivent répondre à deux contraintes : permettre de bien estimer le où les liens entre réflectances et LAI (on considère classiquement que cette relation est dépendante du type couvert), et bien décrire l'espace

géographique d'étude, tant au niveau spatial qu'au niveau culturel. Définir cette stratégie revient à répondre à trois questions :

1. comment répartir les échantillons entre les différentes cultures ? Si on considère que l'on a k classes de probabilité (occurrence) P_k et n échantillons à répartir, deux stratégies s'offre à nous : affecter $\frac{n}{k}$ échantillons à chaque classe ou alors en affecter $n \cdot P_k$. Ces stratégies auront un impact différents sur les erreurs d'estimation. Quelle erreur veut-on minimiser ?
2. Comment répartir les échantillons dans l'espace (dans la mesure ou la contrainte précédente aura été satisfaite) ? Quel critère d'erreur prendre en compte, et est-il réaliste de le mettre en œuvre ? En effet on dispose d'une cinquantaine d'échantillons, à l'échelle du pixel, alors qu'un carré de 10 kilomètres de coté représente 250 000 pixels SPOT.
3. Où placer des échantillons sur une courbe pour l'estimer correctement ? Traçons une courbe hypothétique (figure 1.1) et représentons sur le même graphique la fonction de densité de notre abscisse.

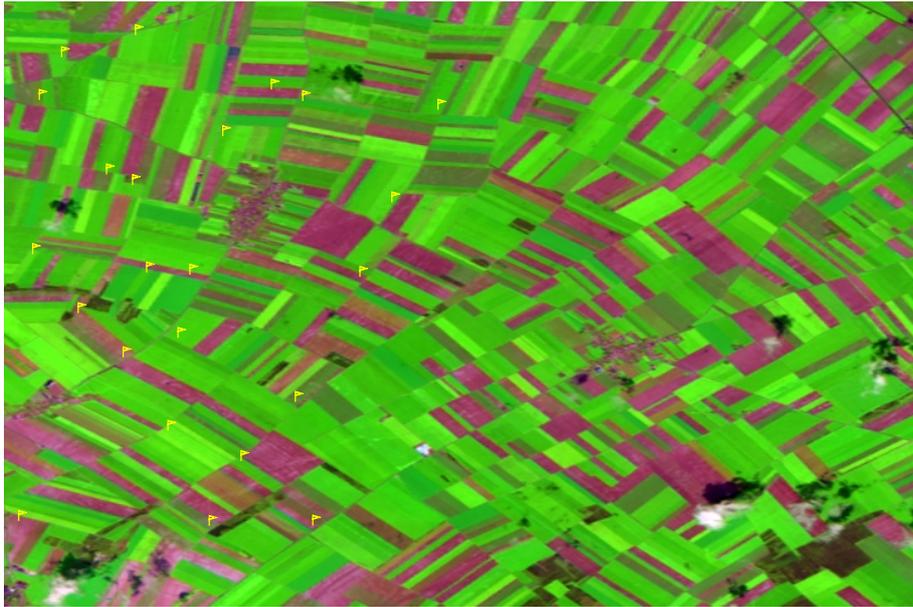


FIG. 1.1 – *Représentation hypothétique de la relation $LAI = f(NDVI)$.*

Pour minimiser l'erreur d'estimation des ordonnées à partir des abscisses, il convient de décrire le mieux possible la courbe dans la région centrale, à probabilité élevée. Mais cette région est-elle pour autant celle ou il faille concentrer l'effort d'échantillonnage pour estimer au mieux notre courbe ?

1.2.2 Évaluation empirique des lois

On veut relier nos paramètres biophysiques à un indice dérivé des réflectances. Les études sur ce sujet montrent généralement deux choses ([23], [10] et [19]) :

1. il est peu probable que ces lois soient les mêmes pour différents types de végétation,
2. ces lois sont classiquement approchées par un modèle exponentiel. En notant $NDVI$ (ou Normalized Differential Vegetation Index) la différence normalisée entre les réflectances dans l'infrarouge, ρ_{PIR} et dans le rouge, ρ_R :

$$NDVI = (\rho_{PIR} - \rho_R) / (\rho_{PIR} + \rho_R)$$

La relation entre LAI et NDVI peut être modélisée comme suit:

$$LAI = -\frac{1}{k_{LAI}} \cdot \ln \frac{NDVI - NDVI_{\infty}}{NDVI_{soil} - NDVI_{\infty}}$$

Où :

- $NDVI_{\infty}$ est la valeur asymptotique du NDVI au nadir, pour un LAI tendant vers l'infini (en pratique pour un LAI de 8),
- $NDVI_{soil}$ est la valeur du NDVI du sol nu,
- K_{LAI} est un coefficient d'extinction.

Ainsi, une relation non linéaire est sans doute la plus à même de relier de façon satisfaisante LAI et NDVI. Cependant, une telle relation n'est observée que si une grande diversité dans la végétation (ou dans la classe de végétation) existe ; or nous allons échantillonner des données pendant une semaine, dans une région d'agriculture intensive. Il est donc possible que nous observions une grande homogénéité au sein d'une même culture. Dans ces conditions précises, peut-être la relation que nous établirons pourra-t-elle être considérée comme linéaire.

Sur la figure 1.2, pour illustrer ce propos, nous avons représenté les mesures (associées à leur écart-type), réalisées lors d'une campagne d'échantillonnage sur la région des Alpilles, sur six parcelles de blés différenciables par leur figuré.

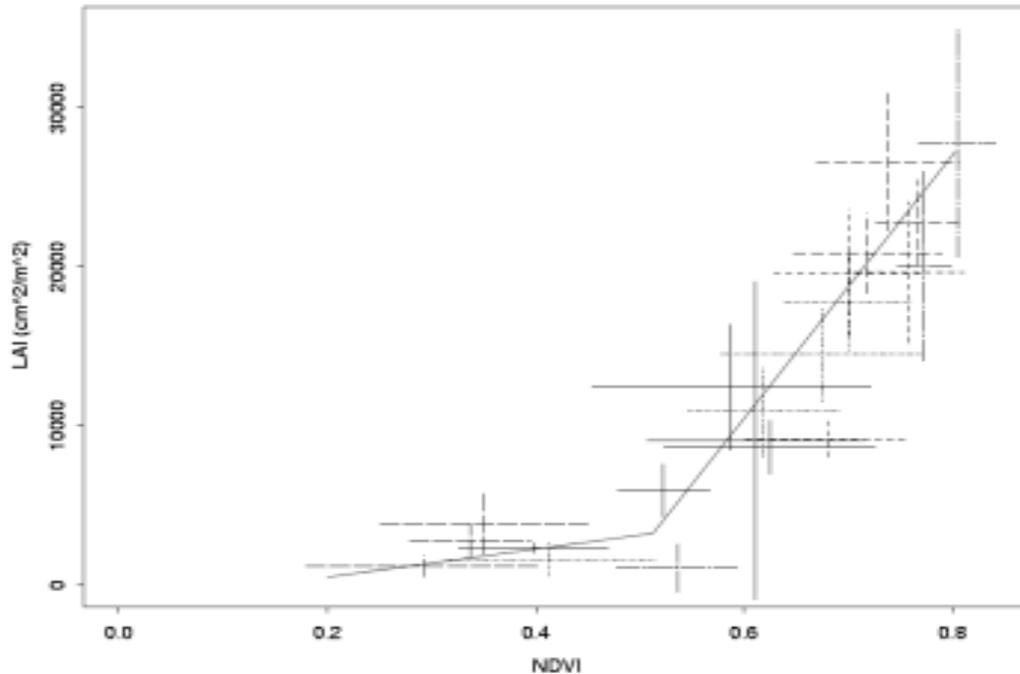


FIG. 1.2 – Relation empirique $LAI = f(NDVI)$ sur du blé.

La large fenêtre temporelle introduit une importante variabilité dans les mesures, cependant, sur les intervalles de NDVI allant de 0,2 à 0,5 et de 0,5 à 0,8 il ne paraît pas extravagant de considérer des relations linéaires.

1.2.3 Définition du modèle géostatistique

La méthode d'interpolation doit s'envisager ici avec l'aide de données extérieures : les réflectances SPOT ou un indice dérivé. Un certain nombre de questions se posent alors :

1. les liens entre réflectances et paramètres seront connus, à ce stade, pour chaque type de culture. D'autre part, le relevé partiel de l'occupation des sols sur la zone, réalisé lors de l'échantillonnage, devrait nous permettre d'obtenir une segmentation de bonne qualité. Faut-il décider, en ces points, que nous sommes dans la classe de probabilité maximale (démarche déterministe) ou plutôt adopter une démarche probabiliste ?
2. Comment prendre en compte les variations de support de mesure ? Les lois d'estimation de nos paramètres biophysiques dépendent du support et les mesures ne sont accessibles que sur une partie seulement d'entre

eux.

D'autre part on pourra se poser la question de l'intérêt d'une telle démarche si l'erreur d'estimation du LAI n'est pas spatialement structurée, c'est à dire si elle est due à autre chose qu'à la variation du type de sol, des réserves hydriques...

1.3 Nature des données

1.3.1 Les mesures au sol

Il a été décidé de répéter une cinquantaine de croix de douze mesures finement géoréférencées. La moyenne sur chaque croix de mesures permet d'obtenir des mesures surfaciques avec un écart type associé, comparables à celles fournies par SPOT : la réflectance intégrée sur 400 mètres carrés. Ces mesures sol sont géoréférencées grâce à un GPS différentiel, de façon à pouvoir les relier aux réflectances SPOT. Le nombre de mesures maximal nous est imposé par des contraintes temporelles : nous ne voulons pas introduire d'écart temporel significatif entre nos mesures sol et les mesures satellite, et nos techniques de mesures ne peuvent être réalisées qu'en condition de rayonnement diffu, soit peu de temps par jour quand il fait beau.

Les mesures de Leaf Area Index sont réalisées au LAI2000 (Li-Cor), qui évalue la fraction de trous du couvert végétal, et de là, calcule une valeur d'indice foliaire. L'évaluation du fCover se fait par comparaison des rayonnements lumineux sur et sous le couvert, dans des bandes spectrales où les feuilles sont opaques (le bleu), cela pour plusieurs secteurs angulaires centrés sur le nadir. Pour les cultures très peu développées, ou difficiles à mesurer au LAI 2000, telles qu'une culture de pomme de terre (relief lié au rang), nous avons pris des photos hémisphériques aux points de mesure. De ces photos on déduit de la même manière LAI et fCover. Une fois binarisées pour discriminer la couverture végétale du sol, il est possible d'évaluer la fraction de trous.

1.3.2 Les données satellites

On utilise les images issues du capteur HRV embarqué sur SPOT. Celles-ci sont à pas de 20 mètres, dans 3 bandes : rouge, vert, proche infra rouge (SPOT4 dispose d'une bande supplémentaire dans le moyen infrarouge). On utilisera des images géoréférencées et si possible corrigées des effets atmosphériques qui perturbent généralement les algorithmes que l'on veut valider.

1.4 Zone d'échantillonnage

Dans un premier temps, il était prévu d'aller faire cette campagne de mesures en Camargue. Nous aurions alors eu accès à un jeu de données assez similaire pour la préparer : les résultats du programme Réséda, sur les Alpilles, issus des mêmes types de culture sous un climat proche. Cependant deux contraintes techniques nous ont incités à changer de site très peu de temps avant la campagne :

- les rizières inondées risquaient de gravement perturber notre interprétation des données satellitales, compte tenu de la lame d'eau couvrant le sol,
- les cultures d'orge et de blé étaient arrivées à maturité. Or nos techniques estiment le LAI à partir de la fraction de trous, mesurée quasiment indépendamment de la couleur du couvert, ce qui n'est pas le cas des indices de végétation issus des mesures radiométriques satellitales.

Nous nous sommes donc reportés sur une région étudiée dans le cadre d'un programme d'agriculture de précision : la vallée de la Seine, entre Troyes et Provins, représentée sur la figure 1.3. En effet, en juin, les cultures n'y sont pas trop avancées. Pour préparer la campagne, nous avons eu accès à une image SPOT4 (sans occupation des sols) sur laquelle nous avons choisi un carré de 10 kilomètres de côté. Le choix s'est porté sur la zone sud-est de Romilly-sur-Seine, zone accessible, au parcellaire régulier et de grande taille, dotée d'un minimum de bois (ceux-ci sont difficiles à échantillonner car souvent inaccessibles et très variables) et de zones urbanisées.

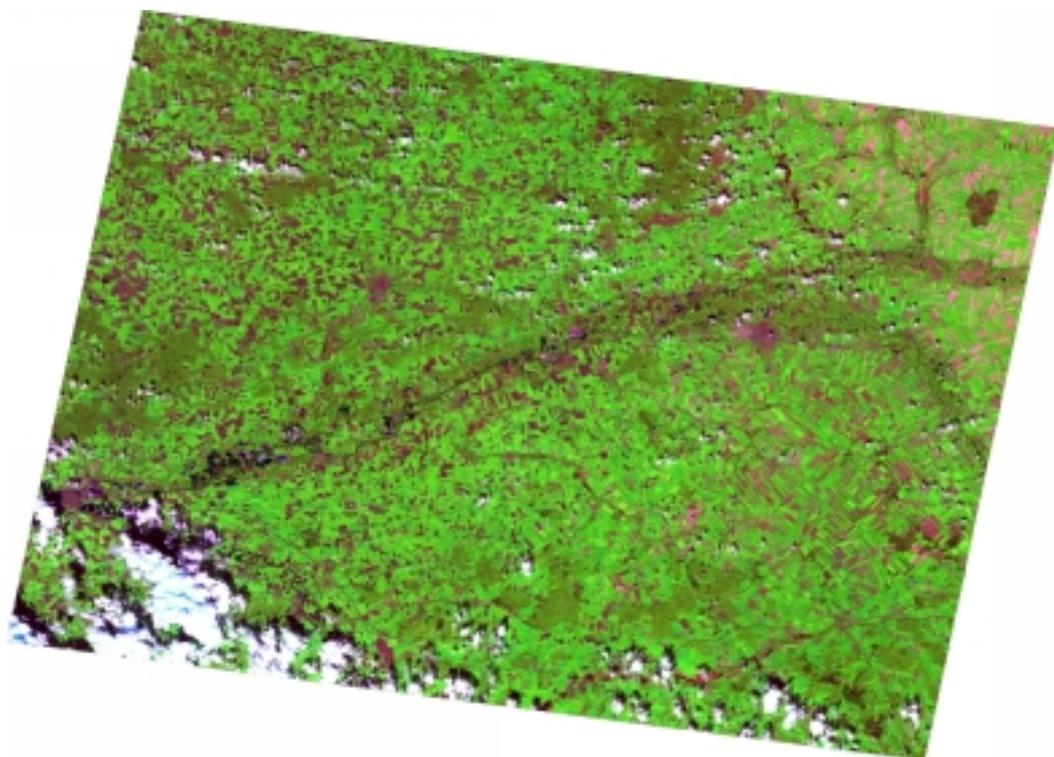


FIG. 1.3 – *Image de la vallée de la Seine, couleurs vraies.*

Chapitre 2

Campagne d'échantillonnage sur la zone de Romilly

Avant d'aller sur le terrain, nous ne disposions que d'une image satellite de la zone. Or notre échantillonnage devait être optimisé pour décrire le mieux possible la ou les lois liant les paramètres à la réflectance et pour permettre de réaliser une cartographie de ces paramètres par des méthodes dérivées du krigeage. Dans cette optique, il était nécessaire de segmenter l'image, afin de modéliser les éventuelles corrélations spatiales, entre les informations spectrales ou biophysiques.

2.1 Echantillonnage mis en œuvre

2.1.1 Choix des emplacements de mesure

Absence de plan d'occupation

Sur la zone d'échantillonnage nous ne disposions pas du plan d'occupation des sols. Or, nous souhaitions récupérer des classes homogènes aux cultures présentes. Les techniques de classifications non supervisées ne répondent que très partiellement à ces exigences. De plus, pour estimer nos LAI à partir des réflectances, il est souhaitable de pouvoir identifier précisément la ou les cultures présentes dans les classes considérées. En d'autres termes il fallait que cette segmentation :

- ne mélange pas les cultures aux réflectances proches mais dont les paramètres biophysiques sont d'un comportement différents,
- permette de mettre un nom de culture ou de type de culture sur une classe.

Or tout ceci n'est pas possible à partir d'une seule image satellite, il paraît donc utopique d'avoir une classification utile avant l'échantillonnage

Implication sur l'échantillonnage

Nous avons soulevé trois points lors de l'exposé du problème. Compte tenu des maigres données préalables auxquelles nous avons eu accès, voici comment nous y avons répondu :

1. Le premier problème consiste dans la répartition des échantillons entre les cultures : on se posait la question de l'erreur à minimiser dans l'échantillonnage des courbes. Comme nos paramètres vont être intégrés sur de grandes surfaces, la stratégie consistant à minimiser l'erreur globale d'estimation paraît plus satisfaisante. Cependant celle ci suppose que l'on connaisse la proportion de chaque culture pour que l'on puisse lui affecter $n \cdot p_k$ échantillons. Sans plan d'occupation des sols et avec une image ne nous permettant pas d'effectuer de classification satisfaisante (prise fin mai, elle ne permettait pas de discriminer les cultures tardives), l'accès à ces proportions était impossible. Nous avons alors décidé de l'estimer sur le terrain, et de définir alors la répartition culturelle des échantillons.
2. Comment placer les échantillons dans l'espace ? Dans le cas idéal, on aurait pu écrire un modèle de krigeage et rechercher la configuration pour laquelle la variance de krigeage est la plus faible. Cependant, nous ne pouvions réaliser plus d'une cinquantaine d'échantillons sur une zone de dix kilomètres de côtés. Si l'on suppose que la covariance spatiale entre les mesures de LAI a une portée du même ordre de grandeur que celle des réflectances, cette covariance spatiale entre deux points sera non nulle pour une distance inférieure à 400 mètres environ, la demi largeur moyenne des parcelles. Dans ces conditions, si on veut bien décrire spatialement la scène, on risque fort de placer 50 échantillons complètement indépendants... Il convient donc de veiller à rapprocher quelques croix de mesures, de façon à avoir une information sur les covariances spatiales aux distances intermédiaires.
3. Quant à la sélection précise des parcelles, permettant de bien décrire les relations $LAI = f(\rho)$, il était manifestement utopique de vouloir la préparer à l'avance : sans classification fiable de la zone nous ne pouvions examiner la dispersion des mesures de réflectances pour une classe de végétation. Quand bien même, cette sélection des points de mesures aurait du être faite en fonction de la dispersion des indices foliaires ; or nous n'avions qu'une idée imprécise des relations entre réflectances et indice foliaire.

Stratégie d'échantillonnage

Compte tenu des remarques précédentes, nous avons opté pour la stratégie la plus sûre : la scène fut divisée en 49 carrés d'égale surface ou nous nous sommes imposé de faire au moins une mesure, si possible la mesure de la culture la plus représentative de ce carré, tout en respectant la représentativité globale de cette culture. Les objectifs sont ainsi correctement respectés :

1. la scène est bien couverte spatialement,
2. les cultures sont échantillonnées en fonction de leur représentativité dans la scène et dans le pixel,
3. les distances de mesure sont assez variables.

D'autre part, pour disposer de données sur les faibles distances nous avons réalisé plusieurs mesures dans certains carrés, et nous avons essayé de choisir des cultures arrivées à des stades végétatifs différents afin de pouvoir estimer plus facilement nos lois. On est très loin, cependant, d'un échantillonnage optimal !

2.1.2 Dispositif expérimental sur une parcelle

Chaque emplacement de mesure de l'indice foliaire est constitué normalement de 12 mesures ponctuelles disposées en croix. Ces mesures ponctuelles sont elles même composées, dans le cas de mesures au LAI200, d'une mesure de rayonnement au dessus du couvert, et de quatre mesures sous le couvert (technique préconisée par les fabricants) ou d'une photo hémisphérique prise au dessus du couvert. Les mesures ponctuelles sont écartées chacune de 4 mètres, les bras de la croix faisant alors 20 mètres de long. Le positionnement précis de ces mesures est assuré par un guide : la croix est matérialisée par quatre piquets reliés par des cordes ou les emplacement de mesures sont précisés.

La position du centre de cette croix est repérée grâce à un GPS, avec une incertitude de positionnement inférieure ou égale au mètre dans 95% des cas.

2.2 Description des données mesurées

Nous avons identifié, sur place, onze cultures différentes :

- des cultures précoces : blé, orge, pois, luzerne, colza, chanvre, pavot ;
- des cultures tardives : maïs, betterave, tournesol, pommes de terre.

A ces cultures, il faut ajouter les bois et les zones urbaines. Nous n'avons pas échantillonné les bois en raison de la difficulté d'accès et du manque de temps. Nous affecterons une valeur plausible de LAI à ces zones. Quant aux

zones urbaines, nous les avons identifiées manuellement.
Nous avons échantillonné une zone de dix kilomètres de cotés, représentée sur la figure 2.1.

2.2.1 Localisation des mesures

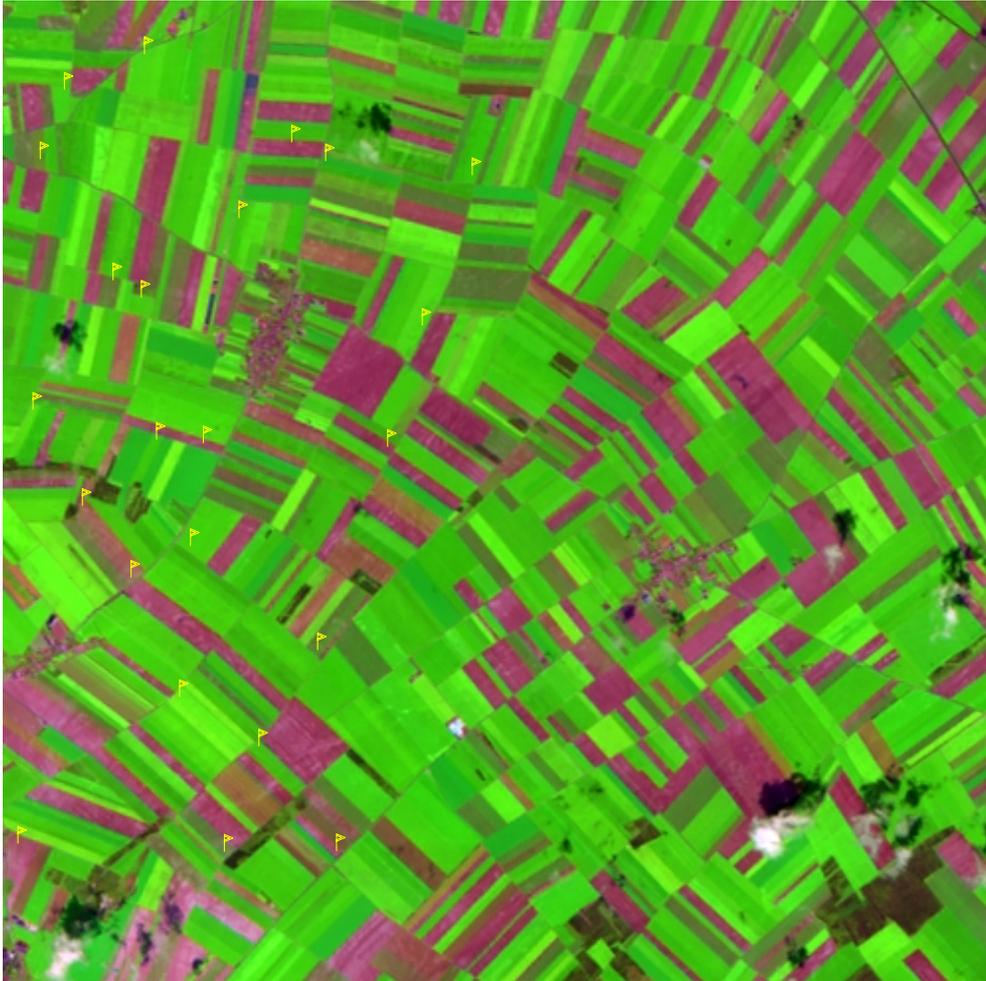


FIG. 2.1 – Image SPOT4 HRV, mai 2000, zone sud est de Romilly-sur-Seine.
Rouge : proche infrarouge, vert : vert, bleu : rouge

Sur cette image, les sites des mesures LAI2000 que nous avons pu exploiter par la suite (une vingtaine parmi les 55 réalisées) sont représentées . La zone d'échantillonnage avait été choisie sur une première image SPOT4 (figure 2.1), et l'image de réflectances qui devait nous servir par la suite

(image SPOT1 de la figure 2.2) ne recouvrait pas totalement la zone échantillonnée... Seuls 40% de la scène furent couverts et nous n'avons eu accès qu'à 23 couples de mesures. D'autre part l'évaluation des LAI à partir des photos hémisphériques n'étant pas encore réalisée, nous n'avons pas eu de résultats sur les pommes de terre et le maïs.

Ces 23 mesures "utilisables" se répartissent comme suit :

Culture	Nombre de croix	Points par croix
blé	5	12
	2	11
orge	2	12
	1	11
	1	9
pois	3	12
luzerne	1	12
	1	11
colza	1	12
chanvre	1	12
pavot	1	12
betterave	3	12
tournesol	1	12

2.2.2 Gestion des croix incomplètes

Dans certains cas nous n'avons pu réaliser les douze mesures ponctuelles prévues: la réapparition du soleil nous a empêché de finir une série, une erreur de manipulation fut détectée lors de l'extraction des mesures... Comment gérer ce type d'incidents?

Il est difficile de rajouter des points de mesures: pour ne pas introduire de biais, il faudrait modéliser la variable aléatoire LAI (s'il s'agit d'une gaussienne, estimer les paramètres espérance et covariance entre les points) et réaliser un tirage aléatoire des points manquants, suivant cette loi, en imposant la covariance spatiale. En quoi l'absence de points peut elle contrarier les calculs? elle n'a pas beaucoup d'influence sur l'estimation de l'espérance d'une croix. En revanche cette absence d'information a un impact sur les variances des mesures de chaque croix, et nous gêne pour modéliser les covariances spatiales entre les points.

Ainsi, lors de l'estimation d'une covariance à grande échelle, on utilisera les valeurs moyennes, équivalentes, mais on mélangera des variances différentes. Dans le cas de la covariance spatiale à petite échelle, il est osé d'estimer une fonction de covariance sur une douzaine de points. On considérera donc que

les covariances spatiales sont les mêmes pour tous les groupes de cultures homogènes (tels que les céréales par exemple). La modélisation de la covariance se fera alors sur l'ensemble de ces mesures ; l'impact d'un point de mesure manquant est ainsi négligeable.

2.3 Présentation de l'image satellite

2.3.1 Image SPOT de la zone

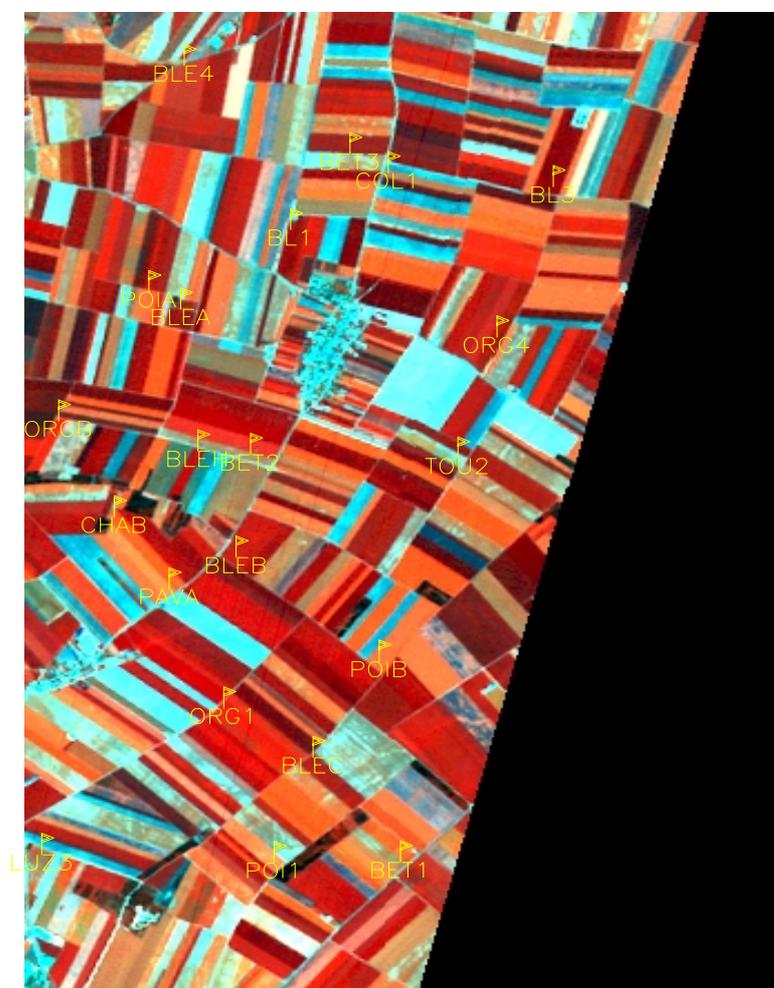


FIG. 2.2 – Image SPOT1 HRV, juin 2000, zone sud est de Romilly-sur-Seine.
Rouge : proche infrarouge, vert : vert, bleu : rouge

2.3.2 Classification supervisée

Nous avons relevé, lors de notre campagne d'échantillonnage, l'occupation d'un certain nombre de parcelles, ce qui nous a permis de construire une classification satisfaisante de la zone. Nous allons présenter ici la méthode du maximum de vraisemblance qui nous a servi à classer chaque pixel de l'image :

- on définit un ensemble d'apprentissage pour chaque classe, sur l'image d'origine,
- à partir de ces ensembles on construit les histogrammes des données radiatives par bande,
- ces histogrammes permettent de construire des cartes de probabilité d'appartenance de chaque pixel de l'image à la classe considérée. Plus exactement, ces histogrammes nous permettent d'évaluer pour chaque pixel x la probabilité d'appartenance à la classe $k, k \in [1, K]$ conditionnellement à x . Pour calculer ces probabilités $p(C_k|x)$, on utilise le théorème de Bayes :

$$p(C_k|x) = p(x|C_k) \cdot p_k / p(x)$$

où :

$$p(x) = \sum_j p(x|C_j) \cdot p_j$$

et p_k est la probabilité à priori d'appartenance à la classe k . Le calcul des $p(x|C_k)$ est aisé à partir des histogrammes.

- Comme p_k est la même pour toutes les classes considérées, on classe le pixel x dans la classe c pour laquelle la probabilité $p(x|C_c)$ est la plus grande. Dans un premier temps on ne peut avoir accès aux p_k , on compare donc les probabilités $p(x|C_k)$. Une méthode itérative permet éventuellement de calculer les p_k grâce à la classification précédente, et ainsi d'affiner la segmentation.

Nous présentons à la figure 2.3 la segmentation de l'image de juin que nous avons retenu :

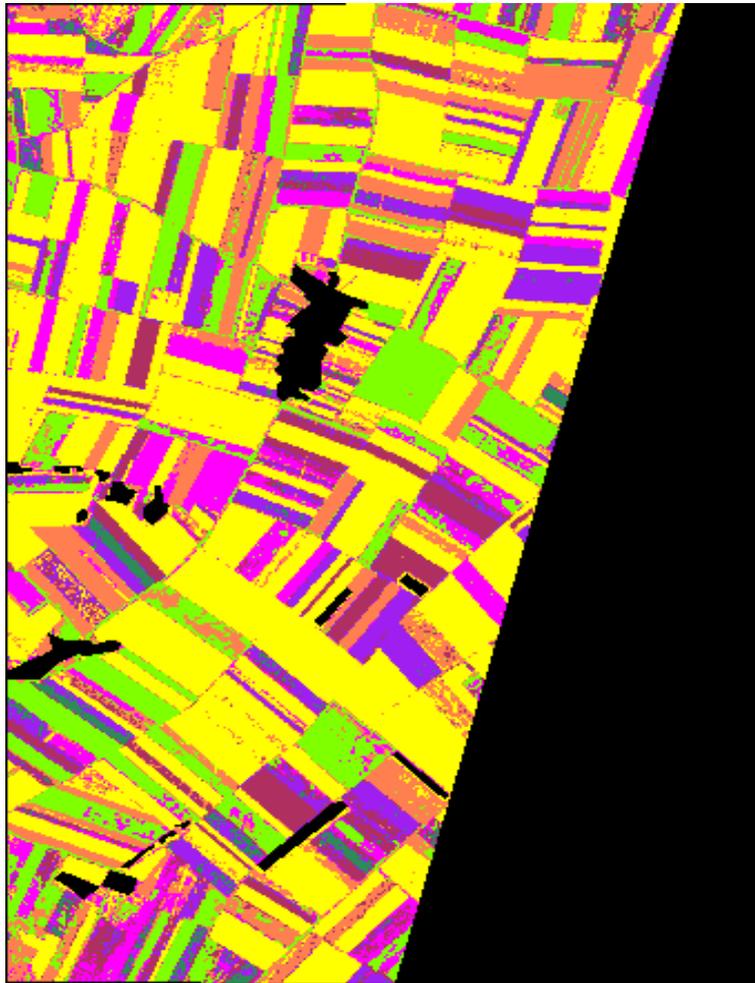


FIG. 2.3 – *Image de classification de la zone, à partir de l'image SPOT1 HRV de juin 2000.*

Les classes de cultures représentées sont les suivantes :

- les céréales en jaune,
- les cultures tardives en vert pâle,
- la luzerne en corail,
- les pois en violet,
- le pavot en vert foncé,
- le chanvre en marron,
- le colza en rose,
- en noir, les bois, que nous avons masqués.

Chapitre 3

Analyse statistique et géostatistique des données

3.1 Les données LAI

3.1.1 Répartition des LAI

On a représenté les fréquences des LAI mesurés sur notre zone d'échantillonnage (histogramme de la figure 3.1). Deux populations se dégagent assez nettement, ce qui n'est pas surprenant : d'un côté les cultures peu développées (betteraves...) de l'autre les céréales, chanvre... Seule la luzerne pose quelque problèmes : en fonction de son âge (la luzerne est en moyenne réimplantée tous les 3 ans) et la date de sa dernière fauche, l'indice foliaire moyen de la culture varie beaucoup. Cependant, en moyenne, on peut classer la luzerne dans le premier groupe.

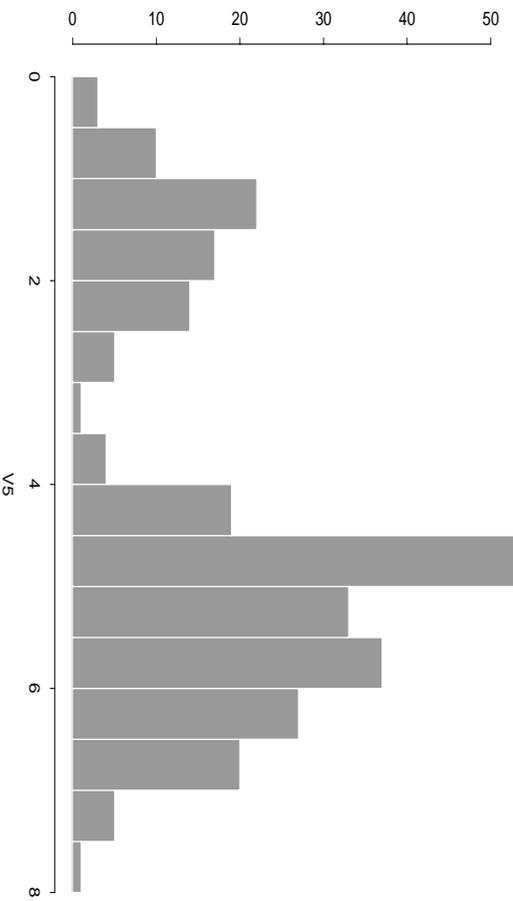


FIG. 3.1 – *histogramme des LAI sur la scène (ensemble des 12 mesures faites sur chacune des 49 croix LAI2000).*

Par cultures on a les histogrammes de la figure 3.2.

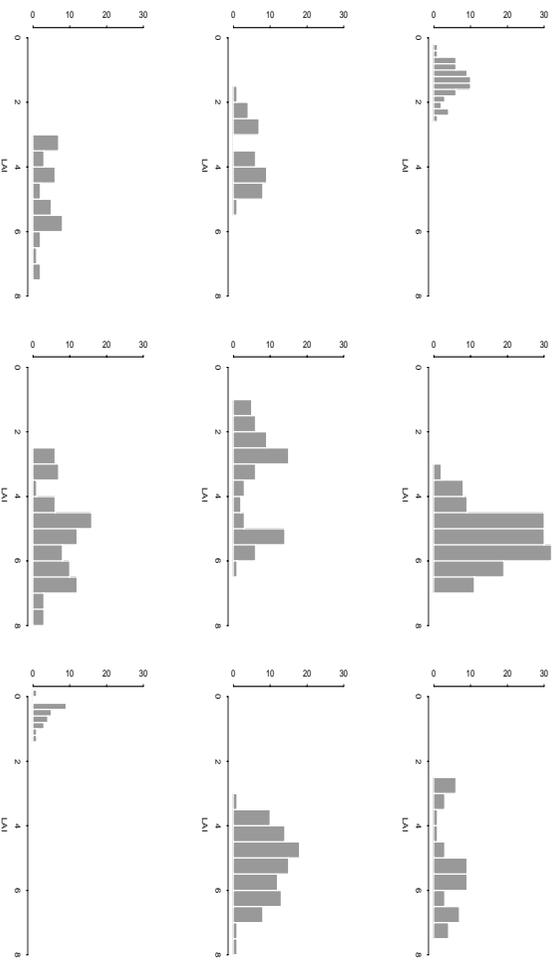


FIG. 3.2 – *Histogrammes du LAI par culture. De gauche à droite et de haut en bas : betterave, blé, chanvre, colza, luzerne, orge, pavot, pois, tournesol (ensemble des 12 mesures faites sur chacune des 49 croix LAI2000).*

3.1.2 Quel support considérer ?

Nous pouvons choisir de travailler sur l'ensemble de nos mesures : n croix de p mesures, ou alors travailler sur une valeur estimée pour chacune des croix.

Impact sur une régression linéaire Si nous considérons le cas d'une régression linéaire classique $Z = \beta X + \varepsilon$, l'estimateur $\hat{\beta}$ est le même que l'on travaille sur les données brutes ou sur la moyenne arithmétique de ces données, à partir du moment où le nombre de mesures est le même par croix. Nous pouvons considérer que nous sommes dans ce cas là (nous devrions introduire une pondération lors de la régression, pondérateurs liés au nombre de mesures par croix. Cependant les différences observées sont si faibles que nous les négligeons). Travailler sur les moyennes présente un avantage important : cela permet de réduire la variabilité des résidus et ainsi de rendre les résultats plus lisibles.

Prise en compte des corrélations Dans le cas d'une régression linéaire généralisée, l'estimateur $\hat{\beta}$ obtenu est le même que l'estimateur fourni par krigeage de la moyenne (Le développement de la technique de krigeage de la moyenne est en annexe). En effet, si nous avons un nombre constant de points, les pondérateurs seraient les mêmes pour les n croix, puisque seule la covariance spatiale intervient dans leur calcul. Or celle-ci est identique pour chacune d'entre elles, nous retrouverions donc bien le même estimateur $\hat{\beta}$ obtenu de façon globale en imposant la matrice de variance-covariance.

Comparaison des deux approches Nous avons modélisé la covariance spatiale à partir du variogramme construits sur les données relatives aux cultures avancées, cela pour pouvoir faire l'hypothèse de l'homogénéité des covariances. Ces variogrammes sont présentés sur la figure 3.3 suivante :

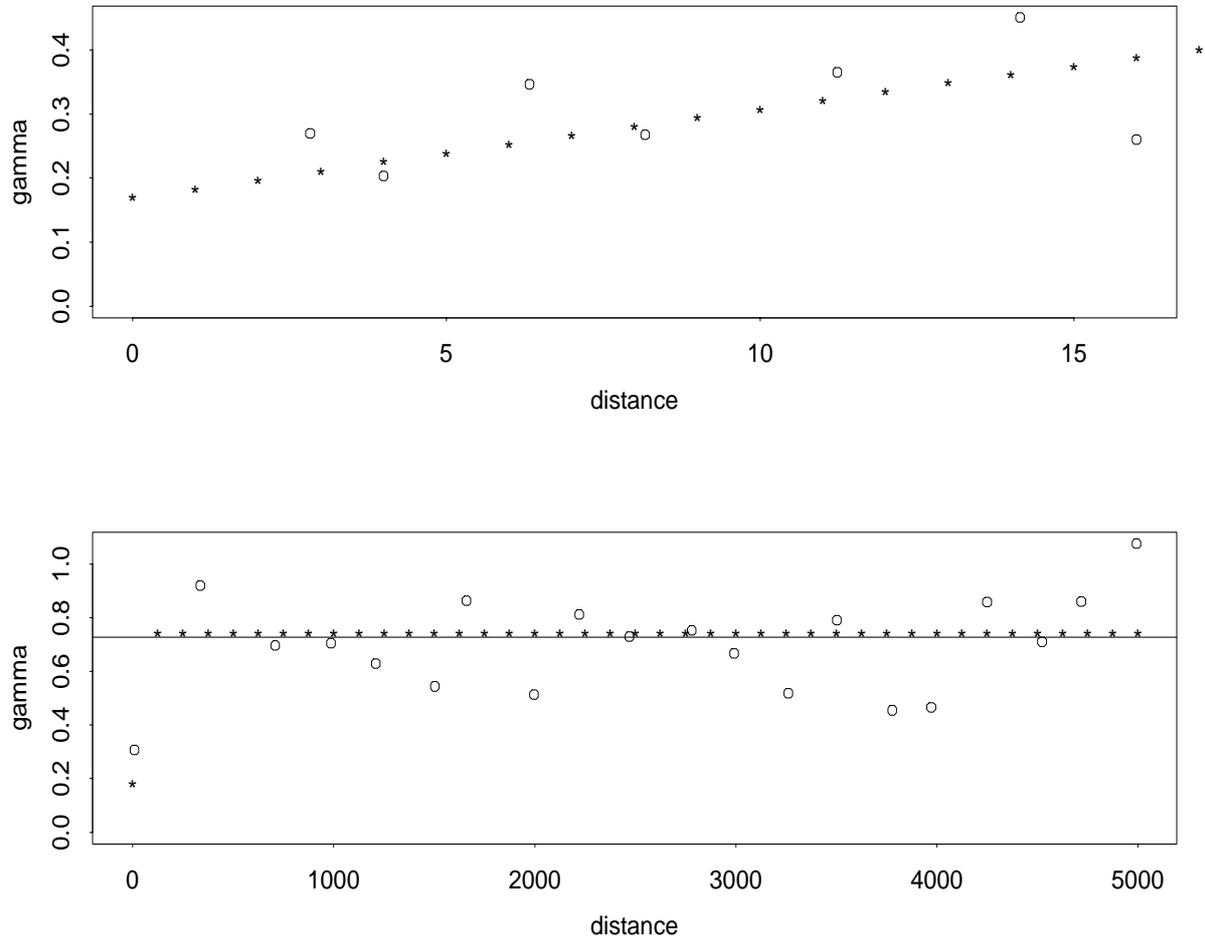


FIG. 3.3 – Variogramme expérimental du LAI (points) et modèle ajusté (*) sur les courtes et longues distances (en mètres). Le variogramme expérimental a été calculé sur les $12 * 49$ mesures.

Nous pouvons ainsi estimer les paramètres du modèle de covariance sphérique. Ici, nous avons estimé la portée à 80 mètres et la valeur de la variance expérimentale à 0,72. Nous avons d'autre part un effet pépité non négligeable d'environ 0,16, qui traduit la discontinuité des mesures ponctuelles (la présentation des modèles de covariances spatiales est en annexe A). Ces informations nous permettent de construire la matrice de variances-covariances Σ liant nos mesures ponctuelles.

Nous avons comparé les moyennes et les estimations par krigeage de la moyenne : le coefficient de corrélation entre les deux vecteurs est supérieur

à 0,98 (pour une moyenne des écarts au carré de 0,016). La différence est non significative, c'est pourquoi nous abandonnerons, dans la partie traitant des régressions, la notion de corrélation spatiale entre les points d'une croix. De plus, dans un souci de lisibilité des résultats, mais aussi pour garantir l'indépendance des différentes mesures, nous travaillerons sur les moyennes arithmétiques.

3.2 Les données SPOT

Nous avons travaillé ici sur l'image SPOT acquise en juin, celle qui nous a permis de former nos couples (mesures au sol, réflectances).

3.2.1 Etude des corrélations entre bandes spectrales

Coefficients de corrélation Soit $\rho(i,j)$ le coefficient de corrélation entre deux variables aléatoires i et j , calculé comme suit : $\rho(i,j) = \frac{Cov(i,j)}{\sqrt{Var(i) \cdot Var(j)}}$.

Le calcul de ρ sur nos différentes bandes spectrales nous donne les résultats suivants :

- $\rho(pir,rouge) = -0.43$
- $\rho(pir,vert) = -0.20$
- $\rho(rouge,vert) = 0.95$

Les bandes rouge et verte apparaissent très corrélées entre elles, mais assez peu avec le proche infrarouge.

Analyse en composantes principales Nous recherchons une base orthogonale de notre espace de réflectances, que nous aurons au préalable centré et réduit, et projetons nos données sur celle ci. Voici les coefficients des combinaisons linéaires des vecteurs réflectances, qui définissent les nouveaux axes, associés à leur pourcentage d'inertie (pourcentage de la variance totale associée à chaque axe).

	axe 1	axe 2	axe 3
inertie	70%	29%	1%
pir	0.37	0.91	0.19
vert	-0.64	0.39	-0.66
rouge	-0.68	0.12	0.72

Comme on le voit, le premier axe, prépondérant, est composé des trois bandes spectrales avec un antagonisme pir / visible, ou les deux bandes du visibles

sont équivalente. Le second représente le proche infra rouge. Deux axes suffisent à décrire 99% de la variabilité ; cela traduit bien l'idée de la redondance de l'information. Ainsi, l'axe trois, formé par les seules bandes du visibles, n'explique presque aucune variabilité.

3.2.2 Etude géostatistique de l'image

Les bases théoriques des outils géostatistiques employés sont en annexes.

Présentation des variogrammes et covariogrammes

Une fois extraites les données radiatives propres à chaque classe, on peut envisager un certain nombre de calculs de statistiques spatiales. Nous avons réalisé ceux-ci de façon à explorer nos données et éventuellement dégager des caractéristiques statistiques intéressantes pour la suite (ils ont, par exemple, aidé à la modélisation de l'auto-covariance du LAI).

Nous avons dans un premier temps calculé les variogrammes isotropes et directionnels sur toutes les cultures confondues :

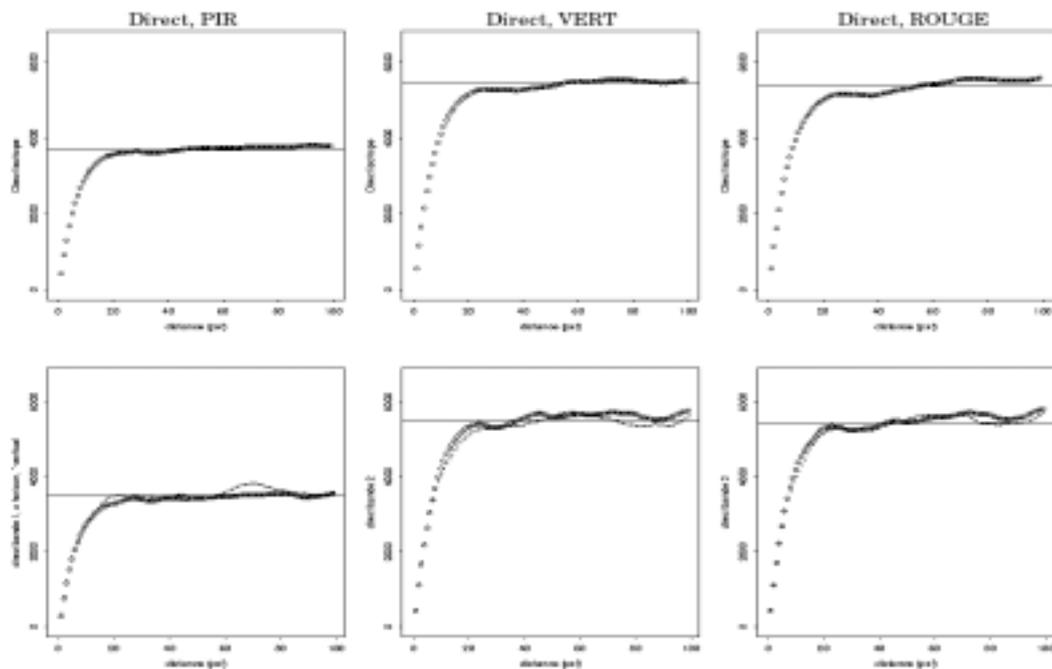


FIG. 3.4 – Variogrammes isotropes en haut, directionnels en bas, construits sur les trois bandes spectrales, dans l'ordre : proche infra rouge, vert, rouge

Les structures (porté, plateau...) des variogrammes expérimentaux calculés dans les cas isotrope et directionnel étant identiques (figure 3.4 en haut et en bas), et cela pour les trois bandes spectrales, nous avons dans la suite réalisé les calculs de façon directionnelle. En effet, on peut alors coder le calcul du variogramme expérimental de façon efficace (voire l'annexe B-1.2). Par ailleurs, les variogrammes présentés sur les figures 3.8 à 3.10 sont calculés sur des données issues d'une classification, qui est entachée d'erreur. Sur les 3 figures suivantes (3.5, 3.6 et 3.7), on a représenté par une ligne horizontale la variance expérimentale des réflectances, dans le cas des variogrammes directs et +/- leur coefficient de corrélation dans le cas des variogrammes croisés.

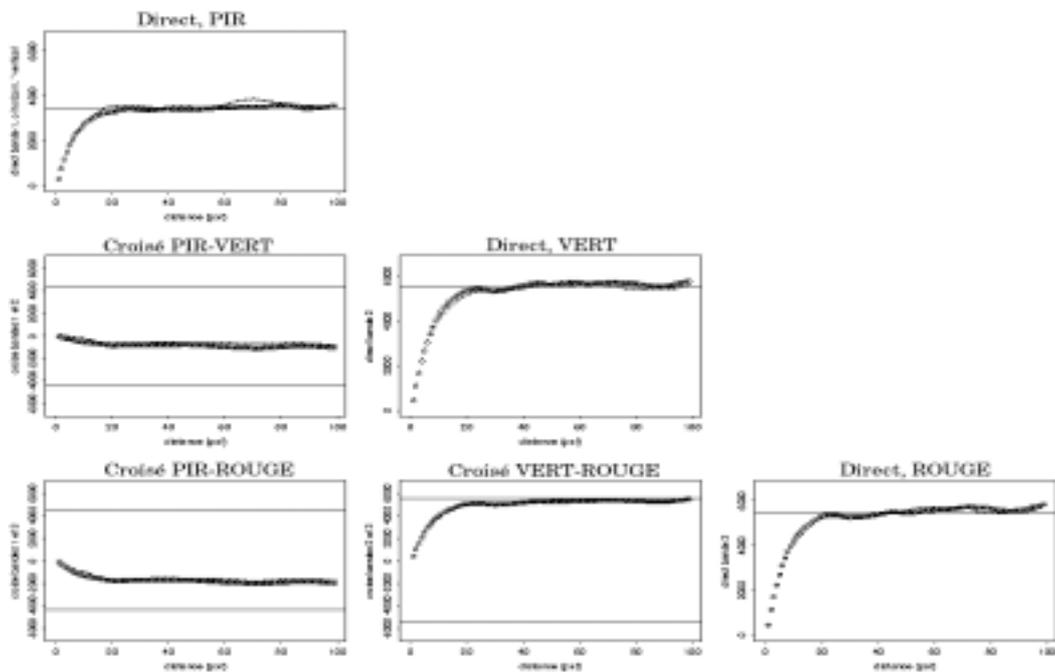


FIG. 3.5 – Variogrammes construits sur les lignes (o) et les colonnes (*) sur toute l'image. Les variogrammes directs sont sur la diagonale.

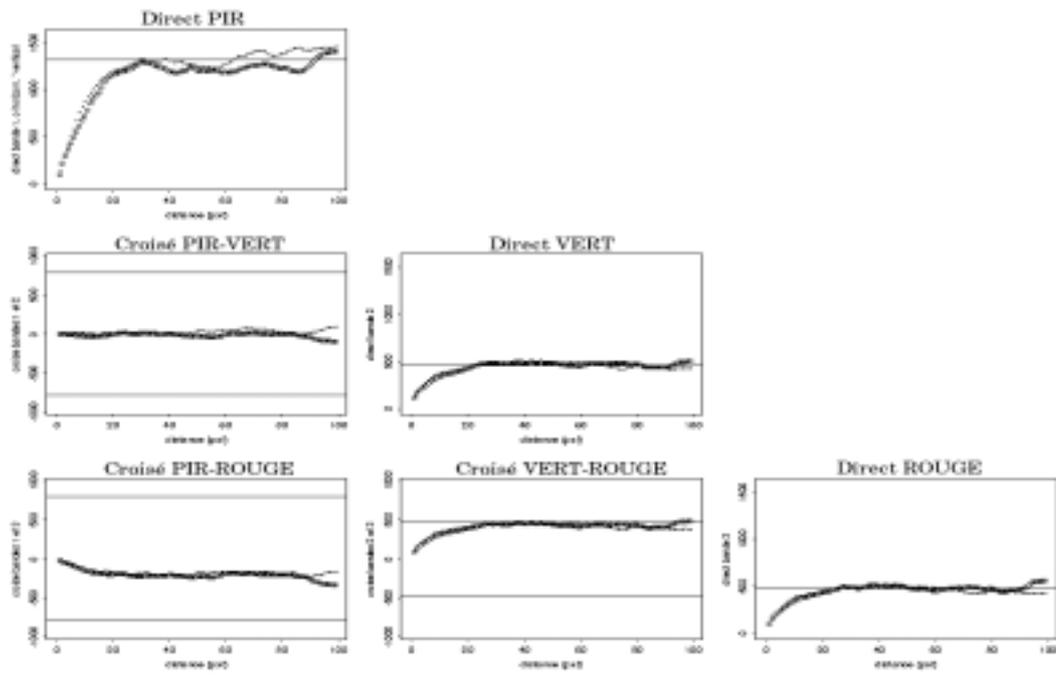


FIG. 3.6 – Variogrammes construits sur les lignes (o) et les colonnes (*) pour la classe céréales. Les variogrammes directs sont sur la diagonale.

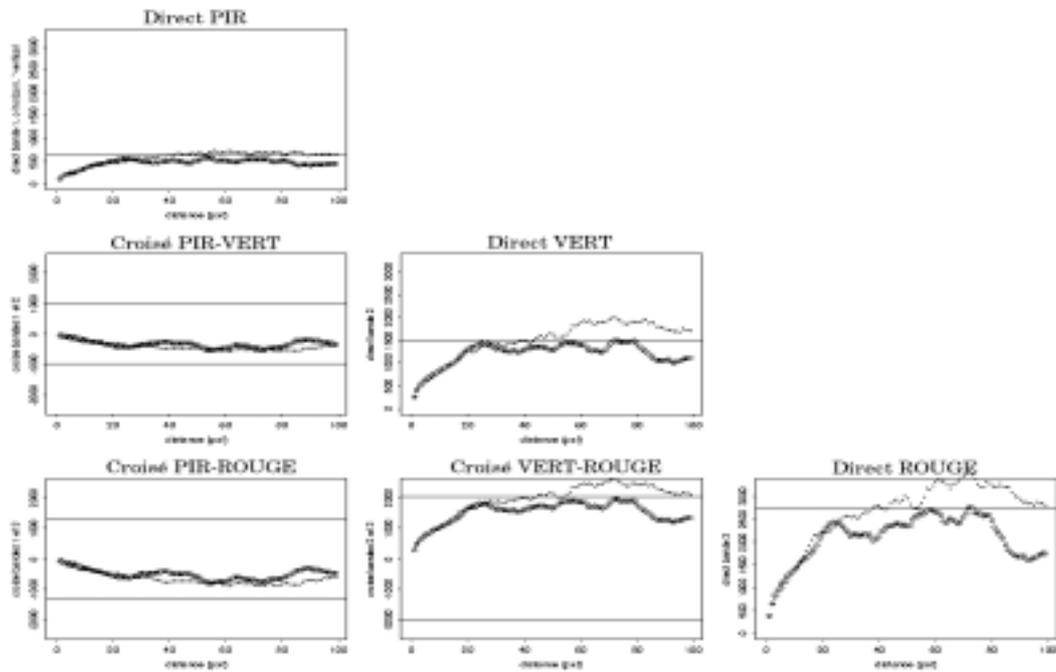


FIG. 3.7 – Variogrammes construits sur les lignes (o) et les colonnes (*) pour les cultures tardives. Les variogrammes directs sont sur la diagonale.

Les covariogrammes et les coefficients de co-dispersion

Pour avoir une meilleure idée des corrélations entre bandes spectrales, nous avons calculé les coefficients de co-dispersion (voir annexes B) pour un certain nombre de classes ; ils sont représentés sur la figure 3.8. Ces coefficients correspondent à des covariances croisées normalisées, qui permettent, entre autre, d’apprécier la stabilité géographique des structures spatiales.

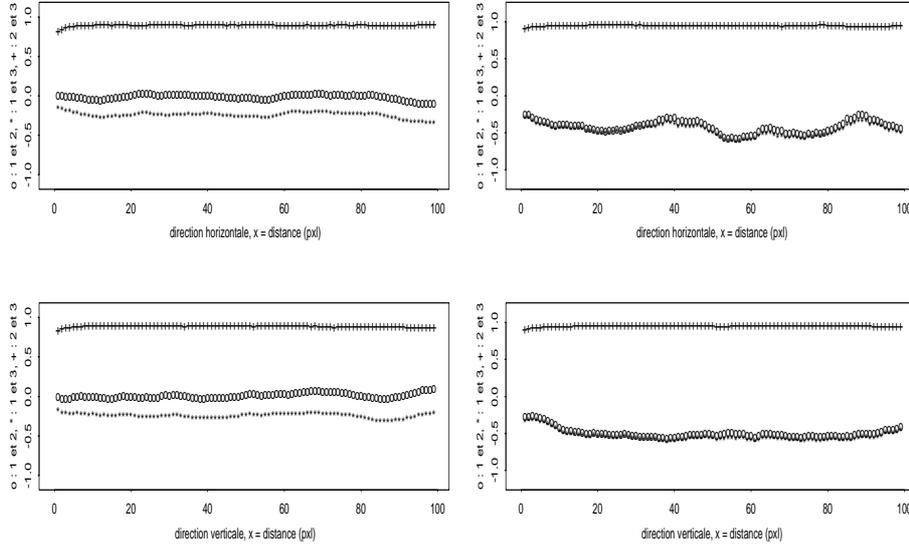


FIG. 3.8 – $\eta(h)$ calculé sur la classe céréales (à gauche) et sur les cultures tardives (à droite). Figuré: (o) entre les bandes 1 et 2, (*) entre les bandes 1 et 3, (+) entre les bandes 2 et 3.

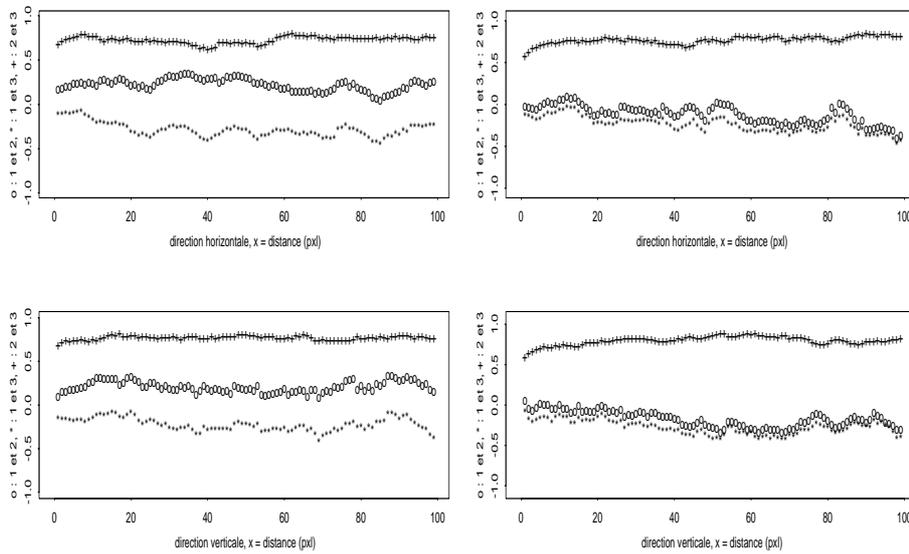


FIG. 3.9 – $\eta(h)$ sur le colza (à gauche) et les pois (à droite). Figuré: (o) entre les bandes 1 et 2, (*) entre les bandes 1 et 3, (+) entre les bandes 2 et 3).

Commentaires sur les figures présentées

Les variogrammes (figures 3.5, 3.6 et 3.7) Plusieurs propriétés peuvent être commentées :

- Les variogrammes obtenus sont bornés et monotones, ce qui est cohérent avec l’hypothèse de stationnarité d’ordre deux (la covariance spatiale décroît avec la distance).
- L’observation de la valeur des paliers atteints par les variogrammes nous montre que la variabilité au sein d’une classe chute beaucoup par rapport à celle observée sur toute l’image : le rapport entre la variance expérimentale calculée sur l’image complète et celle du blé va de 6 à 10 suivant la bande spectrale considérée.
- La variabilité au sein d’une parcelle peut être approchée par la valeur prise par le variogramme aux courtes distances (inférieures à la largeur typique d’une parcelle), autrement dit le variogramme intra parcellaire. Cette variabilité est bien plus faible que la variabilité inter-parcellaire.
- le palier des variogrammes (la portée expérimentale) est atteint, lors des calculs sur l’image complète, pour une vingtaine de pixels. Sur une classe particulière, cette portée est atteinte pour des distances légèrement supérieures, soit 25 à 30 pixels. Ce résultat peut s’expliquer simplement : les hétérogénéités inter-parcellaires sont bien plus élevées que les intraparcélaires. Dans ces conditions on peut s’attendre à ce qu’elles expliquent la plus grande partie de la variabilité spatiale des réflectances, variabilité que traduit le variogramme. Cependant, les discontinuités inter parcellaires sont beaucoup plus faibles lorsqu’on s’intéresse à une classe donnée ; il n’est donc pas surprenant que les portées dépassent la largeur typique des parcelles.

Les covariogrammes Ceux-ci peuvent nous renseigner sur les corrélations existant entre les bandes spectrales. On a déjà vu, lors de l’analyse statistique élémentaire, que les bandes rouge et verte étaient très corrélées. Qu’en est-il d’un point de vue spatial ? Si on note b_{ij} la valeur du palier atteint par les covariogramme, $Cov_{ij}(h) = b_{ij} - \gamma_{ij}(h)$. Ainsi, la covariance croisée entre la bande verte et la bande rouge apparaît effectivement importante sur des distances inférieures à la portée (de 20 à 30 pixels), alors qu’elle l’est beaucoup moins entre ces deux bandes et l’infrarouge.

Les coefficients de co-dispersion (figure 3.8) sont quasiment invariables avec la distance : la structure spatiale de cette information est stable. De plus ils traduisent là encore la redondance de l’information portée par les bandes du visible.

3.3 Régressions LAI / réflectances

3.3.1 Présentation des données

La première étape avant la réalisation des régressions a été de rapprocher nos mesures des valeurs de réflectances associées. Les croix de mesures peuvent très bien se trouver à cheval sur plusieurs pixels. De plus, il y a des incertitudes sur les mesures GPS comme sur le géoréférencement de la scène. Nous avons donc envisagé de rapprocher les valeurs de LAI de la moyenne des 9 réflectances voisines (carré de 9 pixels). Cependant, la fonction de transfert des capteurs ccd de SPOT introduit déjà un effet de moyenne pondéré des pixels : les mesures “bavent” les unes sur les autres. D’autre part, nous avons comparé les résultats de régressions linéaires réalisées sur les données brutes d’une part et les moyennes de l’autre. La différence n’était alors pas significative (les coefficients de régression sont de toute façon très faibles). Nous avons donc choisi la simplicité en conservant les seules données du pixel. Sur la figure 3.10 on a représenté les LAI en fonction des NDVI. Des valeurs extrêmes sont observables (NDVI négatif pour les sols nus, par exemple) pour deux raisons : à ce stade, la correction atmosphérique n’a pas été réalisée, et le sol est rouge et calcaire.

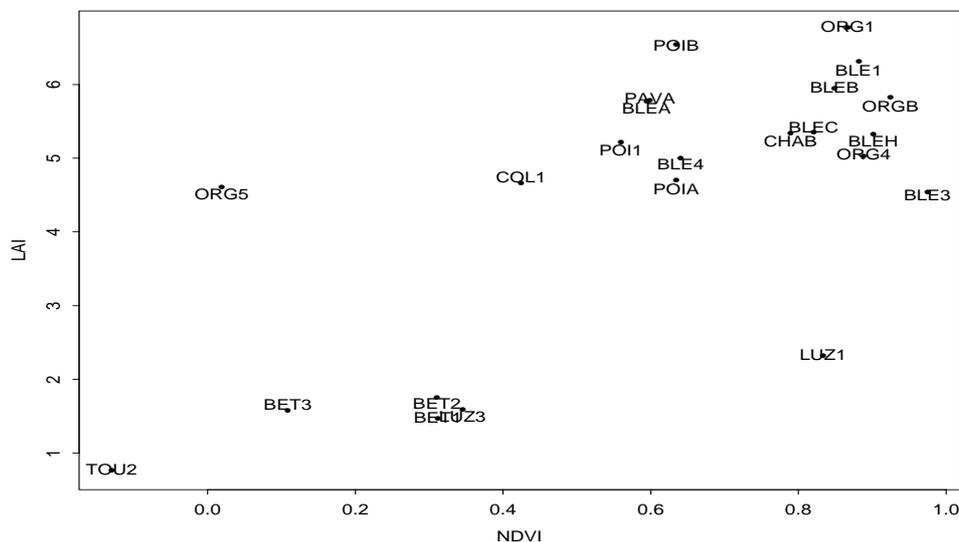


FIG. 3.10 – Rapprochement des données mesurées et satellites. Le nom des mesures correspond aux trois premières lettres de la culture associé à un numéro d'ordre

La figure 3.11 montre les données ponctuelles réparties par culture.

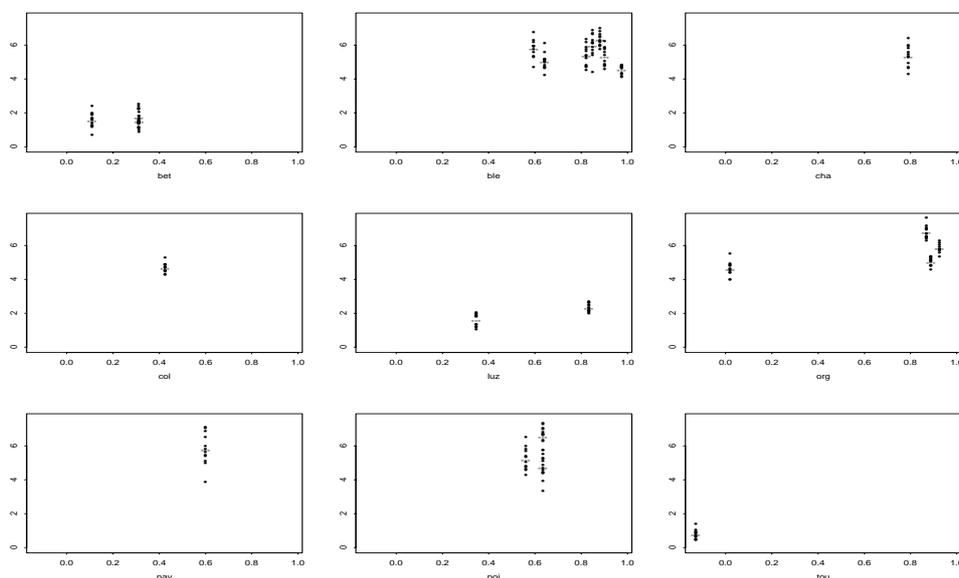


FIG. 3.11 – *Représentation des LAI mesurés en fonction des NDVI, par culture*

Deux mesures posent problèmes : ORG5 a été réalisée sur une culture sénescente, la culture présentait une fraction de trous faible, mais une réponse spectrale très différente d'un couvert vert : la réponse, très forte dans le rouge (culture jaune) et très faible en infrarouge (souvent indicatrice de l'état de croissance des végétaux), explique le NDVI nul. Pour ce qui est de la mesure LUZ1, la luzerne a été très certainement fauchée entre les deux mesures. En conséquence, nos LAI sont très faibles pour un NDVI élevé.

3.3.2 Estimation de la régression

Quel type de régression ?

Le manque de données rend hasardeux l'ajustement d'une fonction paramétrique, d'autant que l'on n'exploite qu'une toute petite plage de données pour chaque culture (une fois les deux mesures aberrantes écartées). Dans ces conditions, l'hypothèse d'une relation linéaire est recevable. En reprenant les notations établies précédemment, nous allons donc chercher à ajuster une relation du type $Z = \beta_0 + Y\beta$, sur le critère des moindres carrés, où β est le vecteur des coefficients de la régression. Puis nous aborderons les relations paramétriques.

Nous avons estimé les LAI associés au support "pixel SPOT" par la moyenne arithmétique des mesures ponctuelles, et réalisé nos régressions dessus. Les rappels sur la régression linéaire sont en annexes.

Sélection des jeux de données

Quelles prédictrices utiliser ? Nous pouvons utiliser comme prédictrices (la matrice Y) différentes variables mesurées : les réflectances tout d'abord, une partie de celles-ci (on a vu que les bandes verte et rouge était très corrélées) ou, comme nous l'envisagions au départ, un indice dérivé tel que le NDVI.

Nous comparerons les résultats obtenus dans les différents cas de figure.

Quel groupes de LAI former ? Compte tenu du faible nombre de couples dont nous disposons, il n'est pas réaliste de vouloir modéliser nos lois pour chaque culture. Cependant, on peut vouloir séparer notre population en deux groupes au comportement spectral très différent : séparer les cultures tardives (betteraves, pommes de terre, tournesol...) des autres. De plus, on peut vouloir écarter les données surprenantes telles les mesures ORG5 et LUZ1.

Nous allons considérer successivement plusieurs régressions :

- régression linéaire,
- régression exponentielle,
- régression logarithmique.

3.3.3 Régression linéaire

Nous avons envisagé quatre groupes de données différents : tous les points de mesure, tous les points sauf ORG5 et LUZ1, et enfin deux groupes formés par la végétation précoce et par la végétation tardive, discriminée à partir des valeurs de LAI (seuil pour LAI=3). Pour chacun de ces groupes nous avons examiné les corrélations entre LAI d'une part et NDVI, proche infrarouge, rouge et vert de l'autre. Nous avons alors testé cette corrélation.

Test sur la corrélation Nous avons pris comme hypothèse :

$$H_0 : \rho = 0$$

$$H_1 : |\rho| > 0$$

Dans le cas de N couples d'observations (x_i, y_i) , on utilisera la corrélation de Pearson :

$$r = \frac{1}{N} \sum_{i=1}^N \left(\frac{z_i - \bar{z}}{s_z} \right) \cdot \left(\frac{y_i - \bar{y}}{s_y} \right)$$

En notant s les estimateurs de la variance et “barre” ceux de l'espérance. Dans le cas d'une distribution binormale, r est l'estimateur du maximum de vraisemblance de ρ . Il a été montré [21] que si $\rho = 0$, la statistique $(N - 2)^{1/2} r / (1 - r^2)^{1/2}$ a une distribution de Student à $N - 2$ degrés de liberté.

La statistique de test n'est vraie que s'il y a indépendance des couples : on teste à la fois cette indépendance et l'existence d'une corrélation. Or ce n'est pas le cas des mesures ponctuelles, ce qui justifie que l'on travaille sur les moyennes.

Données	NDVI		PIR		Vert		Rouge	
	r	$p(H_0 H_0)$	r	$p(H_0 H_0)$	r	$p(H_0 H_0)$	r	$p(H_0 H_0)$
Jeu complet	0,68	0,00032	0,61	0,0018	-0,6	0,002	-0,6	0,002
Jeu incomplet	0,83	0	0,63	0,002	-0,67	0,0008	-0,77	0
Groupe LAI>3	0,18	0,49	0,25	0,34	-0,15	0,48	-0,16	0,54
Groupe LAI<3	0,87	0,054	0,78	0,12	-0,93	0,02	-0,87	0,06

Quel jeu de données conserver ? l'examen de ces résultats nous incite à ne travailler que sur un seul jeu de données, auquel on aura soustrait les deux croix suspectes. En effet :

- Si le LAI est peu corrélé avec chacune des réflectances prises séparément, la qualité d'un modèle dont les prédictrices sont les réflectances ne sera pas satisfaisante. Dans ces conditions, que ce soit pour une régression sur le NDVI ou sur les réflectances, travailler sur deux populations distinctes ne peut s'envisager car le risque de première espèce de 5% est dépassé.
- Si l'on travaille sur une seule population, l'introduction des données ORG5 et LUZ1 fait chuter le coefficient de corrélation de 15%.

On peut choisir de travailler sur deux populations. Cependant le test précédent nous montre l'absence de corrélation dans ces conditions. On travaillerait alors de façon binaire, en utilisant les moyennes des LAI par population. Par exemple, pour tout les pixel ou le NDVI est inférieur à 0,4 (correspondant au seuil de LAI égal à trois), le LAI vaudrait la moyenne arithmétique des LAI mesurés dans cette population.

Régression sur le NDVI ou régression sur les réflectances? Nous avons décidé de continuer à explorer les différentes régressions possibles. Nous cherchons à de terminer ici les prédictrices qui permettent d'expliquer au mieux la variabilité de nos mesures. On comparera donc la racine carrée de la moyenne des résidus au carré (RMSE). Cependant, on observe en général une diminution de cette moyenne lorsque le nombre de prédictrices utilisées augmente. On utilisera donc aussi un indicateur proche du précédent, mais où le nombre d'observations sera remplacé par le nombre de degrés de liberté :

$$RSE = \sqrt{\frac{\sum(Z - \hat{Z})^2}{ddl}}$$

où *ddl* représente le nombre de degrés de liberté.

Finalement, on a recherché les combinaisons de réflectances pertinentes. A cette fin, nous avons réalisé une analyse de variance pas à pas. Si l'on considère le jeu de données complet, l'analyse de variance nous conduit à garder les trois bandes spectrales dans l'ordre *PIR,vert,rouge*. Si l'on supprime les deux valeurs aberrantes, les deux premières sont *rouge,PIR*. La probabilité que le *vert* ne soit alors que du bruit est de 0,053, ce qui est supérieur au risque de première espèce que l'on s'était donné. L'introduction du *vert* réduit cependant le *RSE* de 9%.

Considérons que nous avons rejeté la bande verte comme prédictrice. Dans ces conditions, les coefficients estimés pour la régression linéaire sur le rouge et le proche infra rouge sont -0,027 et +0,02. On se ramène alors quasiment à un indice de végétation.

3.3.4 Ajustement de modèles non linéaires

Le jeu de données Il n'est pas pertinent de vouloir ajuster des modèles paramétriques sur le jeu de données complet ou sur des fractions de mesures correspondant à des types défini de végétations. Nous avons restreint nos ajustements de modèle au cas d'une population sans les 2 données surprenantes.

Modèles dérivés d'une exponentielle Nous avons successivement réalisé des ajustements, avec un nombre croissant de paramètres.

– Modèles pouvant se ramener à un cas linéaire :

Les deux cas suivants peuvent se ramener à une régression linéaire, en transformant les variables (passage au logarithme) :

$$Z = b \exp(Y) \text{ et } Z = b \exp(cY)$$

De plus, on a voulu ajuster le modèle complet, à trois paramètres :

– Modèle non linéaire

$$Z = a + b \exp(cY)$$

Pour estimer les paramètres qui minimisent l'écart-type des résidus, nous avons suivi la démarche suivante : pour un a fixé, on se ramène à un cas linéaire et estimons les paramètres b et c dans l'expression $\ln(Z - a) = \ln(b) + cY$ sur un critère des moindres carrés. Nous recherchons alors la valeur de a qui minimise le *RMSE*.

Pertinence d'une exponentielle? L'ajustement d'une exponentielle pose problème dans notre cas : la majorité des points se situent dans la zone des NDVI élevés, donc dans la zone où la pente de l'exponentielle est la plus forte. Or les résidus sont d'autant plus importants que la pente est forte. Dans le cas de notre modèle à trois paramètres, le *RMSE* est minimisé asymptotiquement lorsque :

$$\begin{aligned} a &\rightarrow -\infty \\ b &\rightarrow +\infty, (a - b) \rightarrow 1,3 \\ c &\rightarrow 0, b \cdot c \rightarrow 5,3 \end{aligned}$$

En développant l'exponentielle ($cNDVI \rightarrow 0$) notre modèle tend vers le modèle suivant : $LAI = 1.3 + 5.3NDVI$, avec les mêmes coefficients que ceux obtenus par régression linéaire. Ce modèle paramétrique ne semble donc pas pertinent s'il n'a pas de justification physique. Enfin nous avons finalement tenté d'ajuster le modèle présenté lors de la définition du NDVI.

Ajustement logarithmique

Modèle Nous reprenons ici la formule empirique présentée en introduction. Nous pouvons estimer le NDVI au niveau du sol à $NDVI_{soil} = -0,5$ et les NDVI pour un LAI infini à $NDVI_{\infty} = 1$. Dans ces conditions, on cherche à estimer b dans l'expression :

$$Z = b \ln\left(\frac{1 - Y}{1,2}\right)$$

Limites Ce type de fonction est en général défini pour un type particulier de végétation, et non pas pour toutes les composantes d'un scène. D'autre part, l'estimation des valeurs extrêmes de NDVI est rendu délicat par la variation de la nature du sol, l'absence de corrections atmosphériques à ce stade, etc. Il n'est donc pas surprenant d'obtenir un RMSE de 2,03. Pour

confirmer la faiblesse de ce modèle dans le cas présent, nous avons laissé libre le paramètre $NDVI_{soil}$ et observé la valeur qu'il prend lors d'un ajustement sur le critère des moindres carrés. L'estimateur de ce paramètre est alors : $ND\hat{V}I_{soil} = 10$ (sic) pour un RSE de 1,52.

Nous avons regroupé ici les résultats précédents :

Modèle	Données	RMSE	RSE	a	b	c	d
$LAI \sim a + bNDVI$	1	1,32	1,38	2.0	4.05		
$LAI \sim a + bNDVI$	2	1,02	1,07	1.26	5.27		
$LAI \sim a + bPIR + cVert + dRouge$	1	0,84	0,97	-0.89	0.035	-0.045	0.038
$LAI \sim a + bRouge + cPIR$	2	0,87	0,94	2.13	-0.028	0.019	
$LAI \sim a + bRouge + cPIR + dVert$	2	0,77	0,86	0.59	0.009	0.028	-0.026
$LAI \sim bexp(NDVI)$	2	1.34			2.38		
$LAI \sim a + bexp(NDVI)$	2	1.26		-1.15	2.94		
$LAI \sim bexp(cNDVI)$	2	1.31			2.15	1.14	
$LAI \sim bexp((1 - NDVI)/1,2)$	2	2.77		2.69			

RMSE: Root Mean Square Error, RSE: Root Square Error, 1: tout le jeu de données, 2: jeux de données sans les mesures ORG5 et LUZ1

La régression linéaire sur les trois bandes spectrales apparaît comme la plus satisfaisante. Cependant nos données ne sont pas forcément très représentatives de la zone, et elle n'ont pas été corrigées des effets atmosphériques. En conséquence, pour des réflectances correspondant à un sol nu, cette régression nous conduit à un LAI négatif (d'environ -2) ce qui d'un point de vue physique est abhérent. Pour éviter ce problème nous considérerons deux relations :

- $LAI = 0.59 + 0.009Rouge + 0.028PIR - 0.026Vert$ pour les réflectances conduisant par cette relation à des LAI positifs.
- $LAI = 0$ sinon.

Cela revient à créer deux populations distinctes. Si nous avions voulu n'avoir qu'une population, il aurait fallu rajouter une condition lors de nos régressions :

On aurait recherché les paramètres a, b, c et d dans la relation $LAI^* = a + bRouge + cPIR + dVert$, tels que la somme des différences $LAI - LAI^*$ au carré soit minimale, et que pour tous les triplets $(pir, vert, rouge)$ de réflectances de la zone d'étude, LAI^* soit positif.

Chapitre 4

Modèles Géostatistiques

Notre objectif final est de produire une cartographie du LAI et de l'erreur associée à son estimation, à une échelle comparables à celle des capteurs à large champs. Le capteur associé à notre image SPOT HRV de juin 2000 est VEGETATION embarqué sur la plate forme SPOT4. Ce capteur produit des images de 1000 mètres de résolution spatiale.

Dans cette partie, afin de limiter les temps de calcul, nous n'avons travaillé que sur une zone de 3 kilomètres de côtés, soit 9 pixels VEGETATION. Cette zone ne contient pas de mesures par photos hémisphériques. Pour les parcelles de sol nu ou de végétation peu développée, le LAI prend une valeur nulle.

La mise en place de modèles spatialisés n'est pertinente que si les variations des variables étudiées répondent effectivement à une structure ! Voici quelques exemples de facteurs structurants :

- le parcellaire, autrement dit les changement de cultures, qui introduisent des variations discontinues, dont l'échelle typique est quelque centaines de mètres,
- les techniques culturales, qui peuvent varier entre deux parcelles (dates de semis, variétés différentes...), mais aussi au sein même d'une parcelle. Ainsi l'échelle typique est plus variable, mais du même ordre de grandeur.
- des facteurs continus sur la scène : la qualité des sols, les facteurs hydriques... Dont on s'attend à ce que l'échelle soit supérieure à celle de la parcelle.

Il serait souhaitable de mettre en place des modèles qui soient adaptés aux grosses discontinuités. Si l'on travaille sur les LAI, il conviendrait donc de faire une étude par classe de végétation, ce que nous n'avons pu mettre en place en raison du nombre insuffisant d'échantillons. Une autre méthode est de s'intéresser aux résidus des régressions du LAI par les réflectances. En ef-

fet, dans le cas idéal, ces résidus devraient être indépendants du parcellaire. Un autre type de variations affecte les données que nous utilisons et perturbe l'appréciation de leurs structures spatiales : les erreurs qui les entachent. Citons par exemple :

- les erreurs de mesures,
- les erreurs dues à l'algorithme d'estimation du LAI à partir des mesures de transmittance au sol,
- l'erreur liée à l'estimation du LAI associé à l'échelle SPOT, à partir de données ponctuelles en croix...

Dans les cas où les données permettaient directement l'ajustement d'un modèle de covariance, ces erreurs peuvent être approchées par l'effet dit de pépité. Lors de l'examen des résidus, nous avons essayé d'approcher plus précisément leur valeur.

Les modèles de krigeage utilisés tout au long de cette partie, sont présentés dans l'annexe C.

4.1 Présentation des différents modèles envisagés

Nous nous sommes intéressé à trois techniques différentes :

1. la plus simple est le krigeage ordinaire des LAI, qui n'utilise pas les réflectances SPOT,
2. si la relation entre LAI et réflectance est linéaire sur notre domaine, on peut mettre en oeuvre un cokrigeage,
3. enfin, on peut s'affranchir de cette condition de linéarité entre LAI et réflectances en krigeant non pas notre variable d'intérêt, mais l'erreur d'estimation qui lui est associée.

Krigeage ordinaire des LAI On construit l'estimateur $Z^*(v)$ comme un estimateur de krigeage ordinaire avec :

- en tout point $Z^*(v) = \sum_{j=1}^n \lambda_j Z(v_j)$,
- la condition de non biais : $\sum_j \lambda_j = 1$,
- le modèle de covariance des LAI, ajusté à partir du variogramme expérimental à l'échelle des pixels SPOT.

Cokrigeage du LAI, avec les réflectances en variables secondaires
Comme on l'a vu, nous avons peu de données pour kriger directement les LAI. L'utilisation des réflectances (ou d'un indice dérivé des réflectances), permet

d'améliorer les estimations. Le cokrigeage, cependant, ne peut être mis en oeuvre qui si nous sommes dans un domaine de relation linéaire entre LAI et réflectances (ou indice dérivé). Dans le cas où nous avons plusieurs cultures aux comportements très différents, la technique du cokrigeage permet de contourner les difficultés liées aux estimations des lois f_x^* , liant la variable principale Z à la variable secondaire Y , estimations critiques pour les cultures peu échantillonnées.

On construit l'estimateur $Z^*(v)$ comme un estimateur de cokrigeage avec :

- en tout point $Z^*(v) = \sum_j \lambda_j Z(v_j) + \sum_i \delta_i Y(v_i)$,
- les conditions de non biais : $\sum_j \lambda_j = 1$ et $\sum_i \delta_i = 0$,
- tous les modèles de variogrammes directes et croisés.

Krigeage de l'erreur Dans les faits, l'erreur $\varepsilon(v)$ d'estimation de la loi f_v^* en v est plutôt multiplicative. Cependant, dans un souci de simplicité, nous avons noté

$$\varepsilon(v) = Z(v) - f_v^*(v)$$

Cette mesure est connue en nos points de mesure j . Ailleurs on peut écrire $\varepsilon^*(v_i) = \sum_j \lambda_j \varepsilon(v_j)$. En considérant alternativement nos k classes (ou occupations du sol) on peut étendre l'erreur d'estimation $\varepsilon_k(v)$ sur la scène, puis composer une cartographie finale en utilisant les probabilités p_k d'appartenance de chaque pixel à chaque classe (fournies par la classification préalable). Pour un pixel v_i on a le modèle suivant :

$$Z^*(v_i) = \sum_k Z_k^* P(v_i \in C_k)$$

avec $Z_k^*(v_i) = f_v^*(Y(v_i)) + \varepsilon^*(v_i)$. On construit l'estimateur $\varepsilon^*(v_i)$ selon l'estimateur du Krigeage Simple avec :

- $\varepsilon^*(v_i) = \sum_j \lambda_j \varepsilon(v_j)$,
- la condition de non biais est automatiquement vérifiée car ε est un résidu donc par définition d'espérance nulle,
- le modèle de variogramme $\gamma\varepsilon$ des résidus.

Le problème de cette méthode, en dehors du fait qu'il faille très bien connaître les lois f_v , réside dans l'interprétation de la carte d'erreur produite. Par exemple, comment intervient l'erreur de krigeage (carte d'erreur de la carte d'erreur...)? En revanche, si l'erreur ne dépend plus de la culture, mais de facteurs sols, stress hydrique... alors elle contourne une des difficultés de l'exercice : on passe à une cartographie continue, indépendante du parcellaire.

4.2 Krigeage ordinaire

4.2.1 Modèle de covariance spatiale

Nous avons essayé de construire un modèle de covariance spatiale sur les 49 LAI moyens mesurés à notre disposition. Cependant, vu les incertitudes auxquelles nous sommes confrontés avec si peu de points, nous avons, au préalable, construit un variogramme sur notre petite zone géographique de 3 Km de côtés. Cette scène est constituée par les moyennes de nos mesures par classes, distribués suivant la segmentation réalisée sur l'image SPOT HRV de juin 2000.

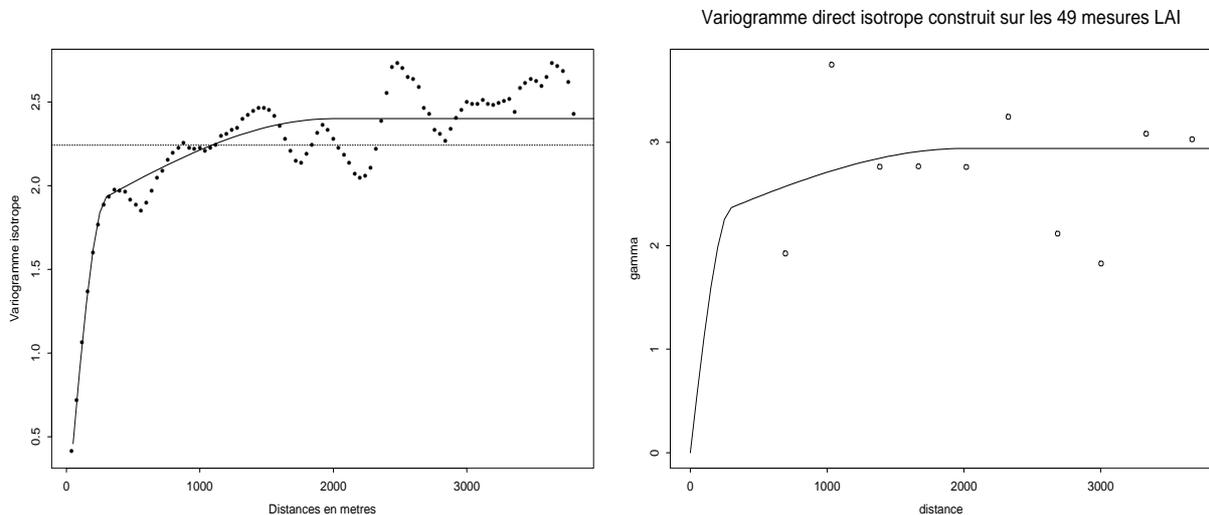


FIG. 4.1 – Variogrammes expérimentaux et modèles ajustés sur les LAI mesurés moyens (à gauche) et les 49 mesures (à droite).

Sur notre étude préalable (figure 4.1 à gauche), nous avons ajusté la somme de deux modèles sphériques. Les LAI étant constants par parcelle, on n'observe pas de pépite. La première structure (portée de 300 mètres) traduit la discontinuité inter-parcellaire, la seconde à beaucoup plus grande portée (2000 mètres) est sans doute liée aux variations des conditions hydriques ou pédologiques.

Nous avons choisi de modéliser le variogramme des LAI (figure 4.1 à droite) de la même façon :

1. une structure de courte portée (égale à la largeur moyenne des parcelles), liée à l'hétérogénéité des LAI intraparcellaires et surtout aux fortes discontinuités inter-parcellaire,

2. une structure de plus longue portée traduisant des phénomènes plus continus (influence du sol...)

Les LAI intra parcellaires sont très homogènes, et se distribuent de façon continue, nous n'ajouterons donc pas de terme pépétique.

Ainsi nous avons le modèle de covariance spatiale suivant :

Pour $h < 300$,

$$Cov(h) = 2.2 \left(1 - \frac{3}{2} \cdot \frac{h}{300} + \frac{1}{2} \cdot \left(\frac{h}{300} \right)^3 \right) + 0.74 \left(1 - \frac{3}{2} \cdot \frac{h}{2000} + \frac{1}{2} \cdot \left(\frac{h}{2000} \right)^3 \right)$$

Pour $300 < h < 2000$

$$Cov(h) = 0.74 \left(1 - \frac{3}{2} \cdot \frac{h}{2000} + \frac{1}{2} \cdot \left(\frac{h}{2000} \right)^3 \right)$$

Pour $h > 2000$ $Cov(h) = 0$ Nous avons ajusté le modèle de façon à ce que la covariance en zéro soit égale à la variance expérimentale des 49 mesures, soit 2.94.

4.2.2 Cartographie du LAI et de la variance de krigeage

En mettant en œuvre un krigeage ordinaire, nous obtenons les cartographies de la figure 4.2.

FIG. 4.2 – Carte de LAI estimé (à gauche) et carte de l'écart type de krigeage (à droite).

L'influence respective de chacune des structures est bien visible : la structure de courte portée provoque l'apparition "d'yeux", car chacune des mesure de LAI marque fortement les estimations locales. D'autre part on observe une opposition générale nord-est / sud-ouest entre les mesures, que la structure de grande portée traduit en un gradient de LAI.

4.2.3 Agrégation

De part sa nature physique le LAI est une variable additive. De plus tous nos pixels SPOT sont de la même nature, la valeur du LAI d'un pixel VEGETATION peut donc être estimée par la moyenne arithmétique des LAI des pixels SPOT qui y sont inclus. Cependant, nous n'avons pas réalisé à proprement parlé d'agrégation ici : nous avons mis en œuvre directement un krigeage de bloc, qui nous fourni directement une estimation du LAI et la variance de krigeage à l'échelle du bloc, représentés sur la figure 4.3

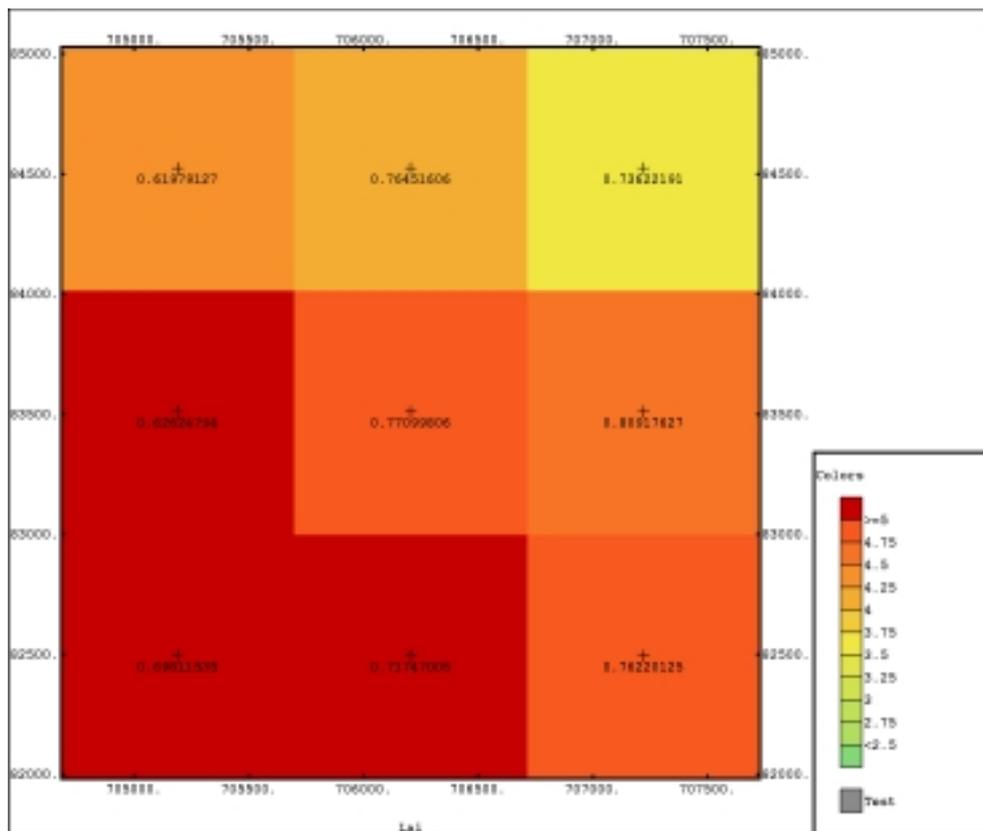


FIG. 4.3 – Carte des LAI estimés à l'échelle VEGETATION. Les écarts types de Krigeage sont précisés sous le centre de pixels .

4.3 Cokrigage

4.3.1 Modèles de covariance nécessaires au cokrigage

Envisager une relation linéaire entre LAI et réflectances n'est pas abusif, on l'a vu. Dans ces conditions, La technique du cokrigage, décrite en annexe, nous permet d'effectuer directement l'interpolation, en tenant compte de toute l'information disponible.

Avec le peu de données dont nous disposons la modélisation des covariance croisées n'est pas aisée. Pour illustrer cette difficulté, nous avons représenté à la figure 4.4 les variogrammes expérimentaux directs et croisés, construits sur les LAI et les réflectances associées.

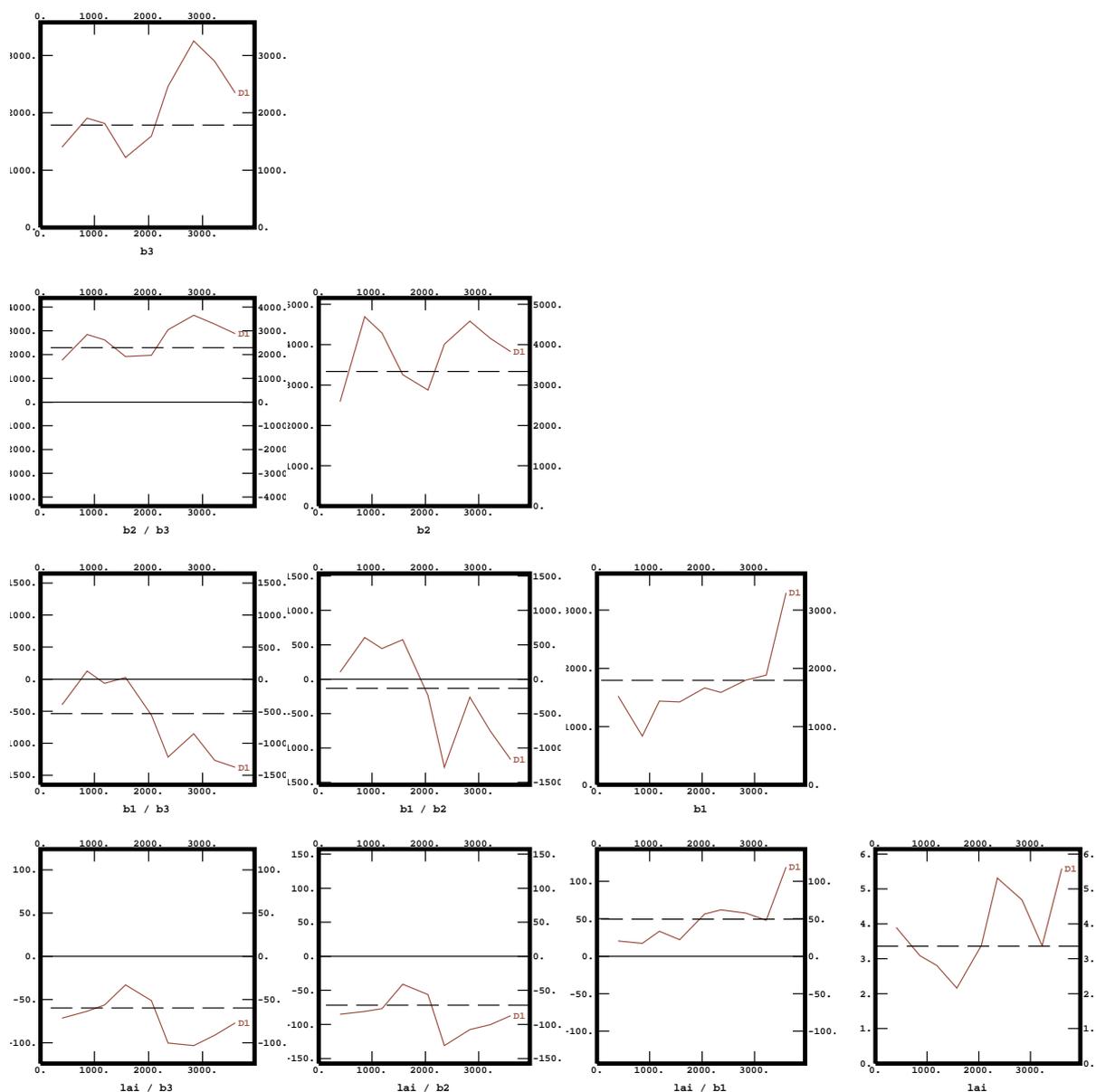


FIG. 4.4 – Variogrammes isotropes, directs et croisés, construits sur le LAI et les trois bandes spectrales.

Le manque de données se traduit par une grande instabilité des figures. De plus, nous ne voulions pas écraser la variable Z par la surabondance des réalisations de Y . Nous avons donc mis en oeuvre un "colocated" cokrigage (cette variante du cokrigage est présentée en annexe C.3). Le modèle d'autocovariance des LAI a déjà été décrit précédemment. La

technique de "colocated" cokrigeage rend inutile la modélisation de l'auto-covariance des variables secondaires, et l'hypothèse de type markovienne que l'on peut faire, nous dispense de modéliser les covariances croisées : $Cov_{ij}(h) = \frac{Cov_{ij}(0)}{Cov_{ii}(0)} \cdot Cov_{ii}(h) = \rho_{ij} \cdot Cov_{ii}(h)$. En revanche, nous ne pouvons faire cette hypothèse si nous considérons plusieurs variables secondaires : la matrice de krigeage deviendrait alors singulière. Nous n'avons donc pas utilisé les réflectances comme variables secondaires, mais les LAI prédits par régression. Cela présente, par ailleurs, un avantage : les deux variables ont la même espérance.

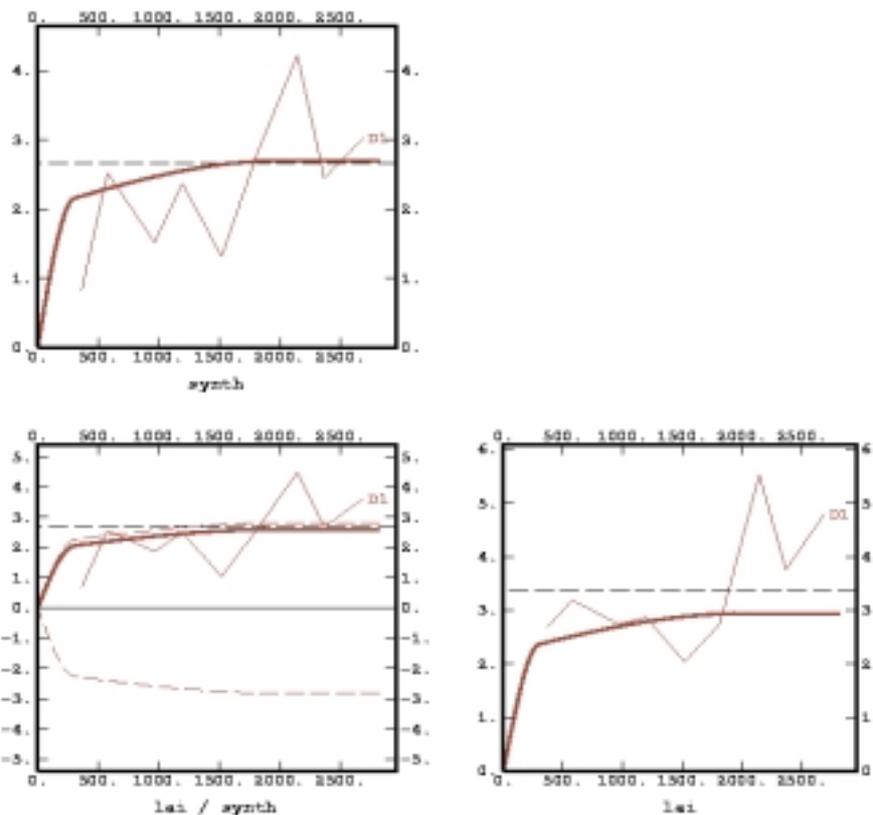


FIG. 4.5 – Variogrammes directs et croisés construits sur les LAI mesurés et les LAI prédits par régression.

Sur les variogrammes de la figure 4.5, nous avons construit un modèle de corégionalisation linéaire (D. Allard, *Introduction au multivariable*, Cours du

DAA Traitement de l'Information Spatiale et [13]). Cette méthode consiste à créer un jeu de fonctions aléatoires d'espérance nulle, indépendantes deux à deux et d'autocovariance connue. On décompose alors les fonctions $Z_k(h) - m_k$, avec $E[Z_k(h)] = m_k$, comme combinaisons linéaires des fonctions $F_i, i \in [1, I]$.

Si Z est le vecteur des fonctions Z_k , M le vecteur des espérances et A_i les I matrices des coefficients des combinaisons linéaires, alors :

$$Z(h) = \sum_{i=1}^I A_i F_i(h)$$

Soit $C(h)$ la matrice de fonctions de covariance correspondante :

$$C(h) = E \left(\left(\sum_i A_i F_i(h) \right) \cdot \left(\sum_{i'} A_{i'} F_{i'} \right) \right) = \sum_i A_i Cov_F^i(h) A_i^t$$

A la condition que les produits matriciels $A_i A_i^t$ soient définis positifs. Dans notre cas, nous avons conservé le variogramme construit sur les 49 mesures de LAI, ce qui nous impose les deux fonctions F_1 et F_2 présentant chacune un structure de covariance, de portée respective 300 et 2000 mètres. Sur les variogrammes expérimentaux (Fig. 4.5), nous avons ajusté le modèle suivant :

$$\begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) \end{pmatrix} = \begin{pmatrix} 2.2 & 1.9 \\ 1.9 & 2 \end{pmatrix} (1 - Cov_{spher_1}(h)) + \begin{pmatrix} 0.74 & 0.7 \\ 0.7 & 0.7 \end{pmatrix} (1 - Cov_{spher_2}(h))$$

en notant Cov_{spher_1} et Cov_{spher_2} deux structures sphériques, de portée respective 300 et 2000, et de variance unitaire en 0. Le coefficient de corrélation ρ utilisé pour l'hypothèse de type markovienne est issu de ce modèle :

$$\rho = \frac{C_{12}(0)}{C_{11}(0)} = \frac{1.9 + 0.7}{2.2 + 0.74} = 0.88$$

4.3.2 Cartographie du LAI et de la variance de krigeage

L'utilisation du modèle décrit précédemment nous permet de dresser les cartes de la figure 4.6.

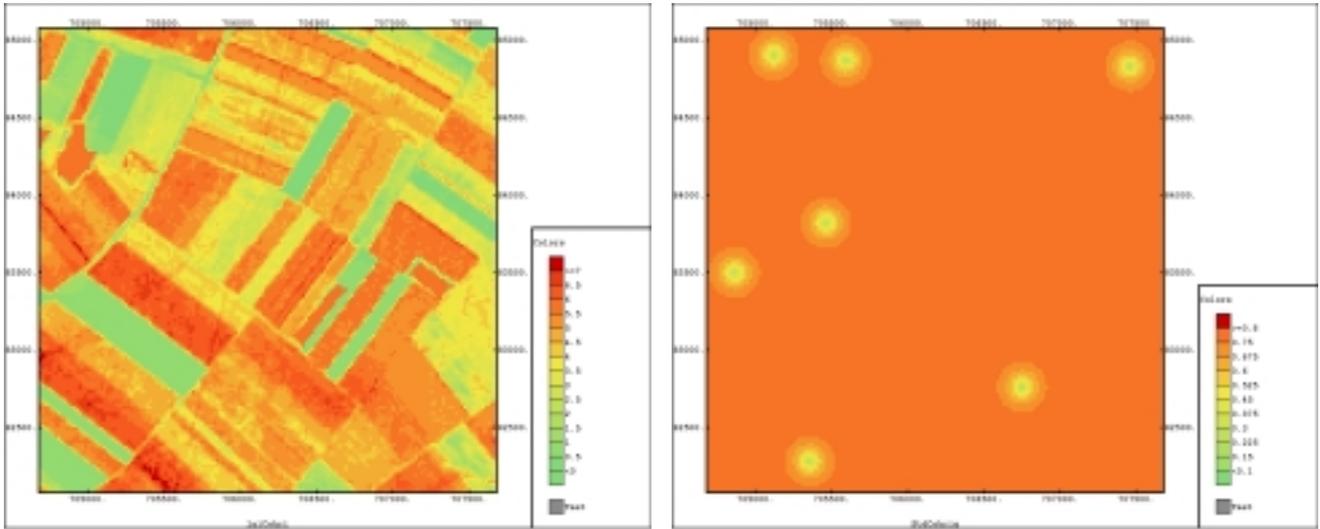


FIG. 4.6 – Cartographie des estimations de LAI par "colocated" cokrigeage (à gauche), et des écarts types (à droite).

L'écart type de krigeage est très stable sur la scène : il prend une valeur comprise entre 0.90 et 0.95, sauf, très localement, aux alentours des points de mesures.

4.3.3 Agrégation

Comme dans le cas précédent, les LAI que nous avons représenté sur la figure 4.7 sont estimé par moyenne arithmétique.

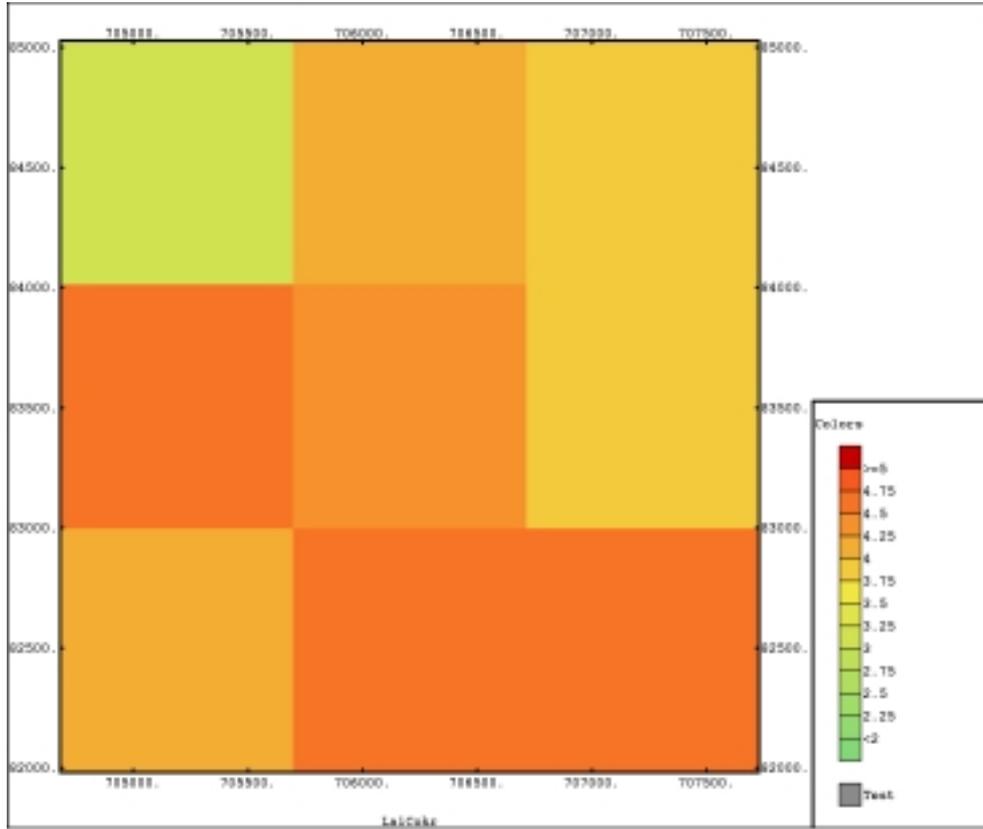


FIG. 4.7 – Cartographie du LAI à l'échelle du capteur VEGETATION.

En revanche produire une carte de variance de krigeage à l'échelle du capteur VEGETATION est plus délicat :

Nous avons calculé les écarts types à l'échelle des n pixels SPOT, les blocs v inclus dans un pixel VEGETATION V . De ceux ci on peut déduire la variance de krigeage à l'échelle du bloc V . En effet :

$$Var(Z(V) - Z^*(V)) = Var(Z(V)) - 2 \cdot Cov(Z(V), Z^*(V)) + Var(Z^*(V))$$

Or :

$$Z^*(V) = \frac{1}{n} \sum_{i=1}^n Z_{v_i}$$

donc

$$Var(Z^*(V)) = \frac{1}{n^2} \sum_i \sum_j Cov(Z^*(v_i), Z^*(v_j))$$

avec :

$$Z^*(v_i) = \sum_{\alpha} \lambda_{\alpha}^i Z_{\alpha} + \lambda_0^i Y_i$$

$$Z^*(v_j) = \sum_{\beta} \lambda_{\beta}^j Z_{\beta} + \lambda_0^j Y_j$$

Donc :

$$Cov(Z^*(v_i), Z^*(v_j)) = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha}^i \lambda_{\beta}^j \cdot Cov_{\alpha\beta} + \sum_{\alpha} \lambda_{\alpha}^i \lambda_0^j \cdot Cov_{\alpha i} + \sum_{\beta} \lambda_{\beta}^j \lambda_0^i \cdot Cov_{\beta j} + \lambda_0^i \lambda_0^j \cdot Cov_{i,j}$$

Avec les notations suivantes :

$$Cov_{\alpha\beta} = Cov(Z_{\alpha}, Z_{\beta})$$

$$Cov_{ij} = Cov(Y_i, Y_j)$$

$$Cov_{\alpha i} = Cov(Z_{\alpha}, Y_i)$$

Les coefficients λ pour tous les pixels SPOT d'un pixel VEGETATION sont accessibles par le calcul. De plus, les termes de covariances croisées se déduisent tous de l'autocovariance de la variables principale :

$$Cov_{\alpha i} = \rho Cov_{Z_{\alpha}, Z_i}$$

D'autre part on peut calculer simplement $Var(Z(V))$:

$$Var(Z(V)) = \overline{Cov}(V, V) = \frac{1}{n^2} \sum_i \sum_j Cov(Z(v_i), Z(v_j))$$

Enfin,

$$Cov(Z(V), Z^*(V)) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n Cov(Z(v_j), Z^*(v_i))$$

donc :

$$Cov(Z(V), Z^*(V)) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{\alpha} \lambda_{\alpha}^i Cov(Z(v_{\alpha}), Z(v_j)) + \rho \lambda_0^i \cdot Cov(Z(v_i), Z(v_j)) \right)$$

Il ne reste qu'à sommer les trois termes pour avoir accès à ces variances. Ce calcul pose des problèmes de mémoire : il y a 2500 pixels SPOT dans un pixel VEGETATION, ce qui nous fait travailler avec des matrices de 6.250.000 éléments... Nous n'avons eu d'autre solution que de discrétiser un peu moins les pixels VEGETATION. On a calculé les écarts types de krigeage pour des discrétisations de $15 * 15$, $20 * 20$ et $25 * 25$. Les résultats sont très stables puisque seule la troisième décimale varie, et dans un interval de 3 unités. Nous avons donc conservé la discrétisation en 400 éléments. Ce code nous a permis de calculer les écarts types de cokrigeage pour les neuf pixels ($\rho = 0.88$) :

	705200	706200	707200
84500	0.448	0.531	0.504
83500	0.444	0.537	0.552
82500	0.488	0.509	0.517

Dans ce tableau, les écarts types de krigeage sont représentés en fonction des coordonnées géographiques des centres de pixel.

On remarquera que les écarts types de krigeage sont plus faibles ici que lors du krigeage ordinaire, résultat que l'on attendait dans la mesure où l'on apporte ici beaucoup plus d'information.

4.4 Krigeage des résidus

4.4.1 Modèle de covariance spatiale

On avait noté : $Z(v)$ les LAI et $y(v)$ les réflectances sur un pixel v et $\varepsilon(v)$, le terme d'erreur qui modélise de façon additive ce qui échappe au modèle $Z(v) = f(y(v))$. On notera v les pixels, c les croix constituées de n mesures ponctuelles.

Présentation du modèle

On peut supposer que l'erreur se décompose en plusieurs termes. Nous avons fait la modélisation suivante : $\varepsilon(v) = \varepsilon_0(v) + \varepsilon_m + \varepsilon_p(v)$; avec :

1. un terme $\varepsilon_0(v)$ spatialisé, lié à un effet de sol par exemple,
2. un terme ε_m inhérent à la mesure elle-même : défaut de l'appareil de mesure, cultures introduisant un effet de rang... Cette erreur est supposée non spatialisée au sein d'une même classe de végétation.
3. Enfin, un terme $\varepsilon_p(v)$ lié à l'erreur d'estimation de $Z(v)$ par \bar{Z} , moyenne arithmétique sur une croix. Cette erreur est spatialisée en théorie.

De plus, On suppose ces erreurs indépendantes, pour avoir :

$$\gamma_\varepsilon = \gamma_{\varepsilon_0} + \gamma_{\varepsilon_m} + \gamma_{\varepsilon_p}$$

Comment approcher ces termes ?

- ε_m est non spatialisé : γ_{ε_m} est un effet de pépité, et $\sigma_{\varepsilon_m}^2 = \frac{1}{n} \cdot pep_Z$ où pep_Z est l'effet de pépité des LAI au sein des croix.
- ε_p est spatialisée en théorie : notons $Z(x_i)$ les douze mesures ponctuelles formant une croix.

$$Cov(\bar{Z}_v - Z_v, \bar{Z}_{v'} - Z_{v'}) = Cov\left(\frac{1}{n} \sum_i Z(x_i) - Z_v, \frac{1}{n} \sum_j Z(x_j) - Z_{v'}\right)$$

$$= \frac{1}{n^2} \sum_i \sum_j Cov_{ij} + \overline{Cov}_{v,v'} - \frac{2}{n} \sum_i \overline{Cov}_{i,v'} - \frac{1}{n} \sum_j \overline{Cov}_{j,v}$$

Où $i \in v$ et $j \in v'$.

Si la distance l entre v et v' est assez grande, la croix est une bonne discrétisation du pixel, et le terme ε_p est presque nul, car alors

$$Cov_{ij} \simeq Cov_{ix'} \simeq Cov_{xj} \simeq Cov_{xx'}$$

avec $x \in v$ et $x' \in v'$. Cependant, pour les pixels très proches, l'approximation n'est plus valable.

Si l'on considère que cette erreur n'est pas spatialisée, on prendra ε_p comme une pépite. $Var(\varepsilon_p) = Var(Z_v - \bar{Z}_v) = Cov(\bar{Z}_v - Z_v, \bar{Z}_v - Z_v)$. A partir de la formule définie précédemment, on obtient la relation :

$$\sigma_{\varepsilon_p}^2 = \frac{1}{n^2} \sum_i \sum_j Cov_{ij} - \frac{2}{n} \overline{Cov}_{i,v} + \overline{Cov}_{v,v}$$

Cette grandeur dépend de la géométrie de la croix dans le pixel v .

- $\varepsilon_0(v)$ est notre véritable variable d'intérêt, spatialisée. On suppose que les phénomènes qu'elle représente sont suffisamment continus pour être sans effet pépite.

Dans le Modèle retenu, $\gamma_\varepsilon(v)$ possède :

1. un effet pépite du uniquement à $\sigma_{\varepsilon_m}^2 = \frac{p\varepsilon p z}{n}$,
2. une structure courte, liée à l'estimation de Z_v par \bar{Z} , que l'on simplifiera éventuellement par une pépite,
3. une structure plus longue, liée à ε_0 .

Estimation des termes du modèle

Les erreurs de mesure Le tracé du variogramme construit sur les mesures de LAI ponctuelles nous permet d'estimer correctement l'effet de pépite lié aux erreurs de mesures :

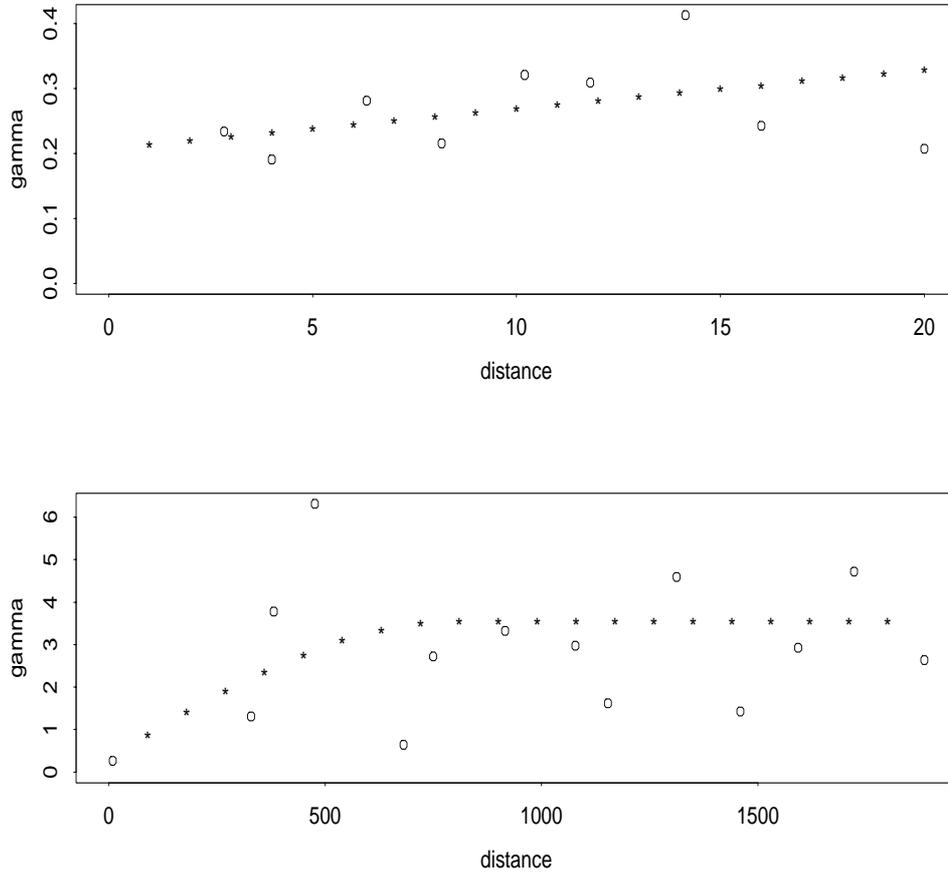


FIG. 4.8 – Variogramme isotrope construit sur les LAI. "o" : variogramme expérimental, "*" : modèle.

La figure 4.8 en haut montre le variogramme calculé à l'intérieur des croix (moyenne des variogrammes intra-croix), celle du bas montre le variogramme calculé entre les croix de mesures. On a alors l'estimateur suivant :

$$\hat{\gamma}_{\varepsilon_m}(h) = \frac{1}{n} \cdot 0.20 \simeq 0.02 \text{ pour } h = 0 \text{ et } \hat{\gamma}_{\varepsilon_m}(h) = 0 \text{ pour } h \neq 0$$

Les erreurs d'estimation des LAI_v par les moyennes arithmétiques

On a décidé d'évaluer $\hat{\gamma}_{\varepsilon_p}(h)$ à partir de $\gamma_Z(h)$, variogramme des LAI, en estimant les fonctions $\tilde{\gamma}_{c,cl}$, $\tilde{\gamma}_{v,vl}$ et $\tilde{\gamma}_{v,i}$, $i \in cl$. Ces fonctions sont sensibles à

la géométrie des mesures : dans le cas du calcul de $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$, la position de la croix de mesure par rapport au pixel est susceptible d'influencer les résultats. Pour essayer de décrire cet impact, nous avons calculé la valeur de $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$ pour des distances croissantes entre c , le centre du pixel, et la croix. Cela, dans quatre cas différents représentés sur la figure 4.9.

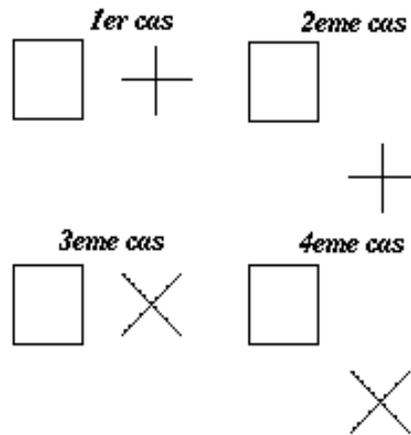


FIG. 4.9 – *Dispositions relatives pixel / croix lors des quatre situations envisagées.*

De plus, on a calculé les valeurs prises par les expressions $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$, i appartenant à une croix de n mesures ponctuelles, pour un pixel v et une croix superposée ; les résultats présentés à la figure 4.10 nous ont permis de choisir une discrétisation satisfaisante d'un pixel :

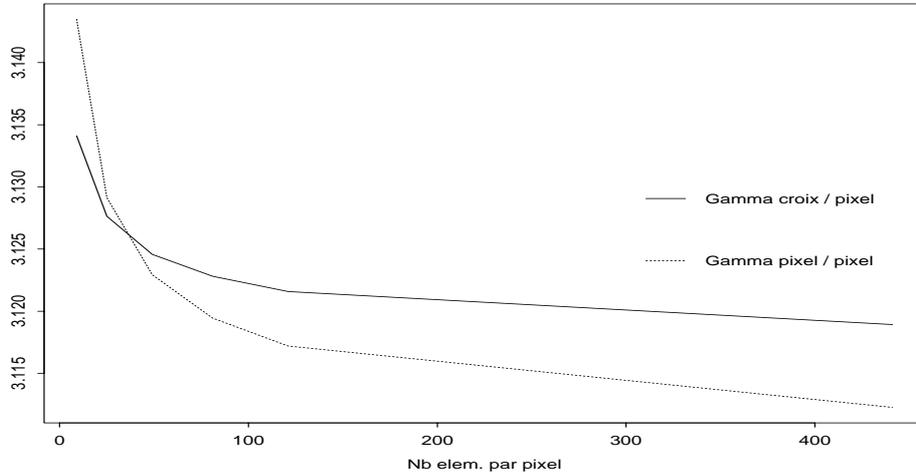


FIG. 4.10 – Effets de la discrétisation sur \overline{Cov} , pour une distance nulle entre le centre du pixel et le centre de la croix.

A partir de la figure 4.10, on s’est donné une discrétisation du pixel en 121 éléments. Pour celle-ci, les expressions $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$ en fonction de la distance sont toutes de la même forme ; nous en avons représenté une à la figure 4.11.

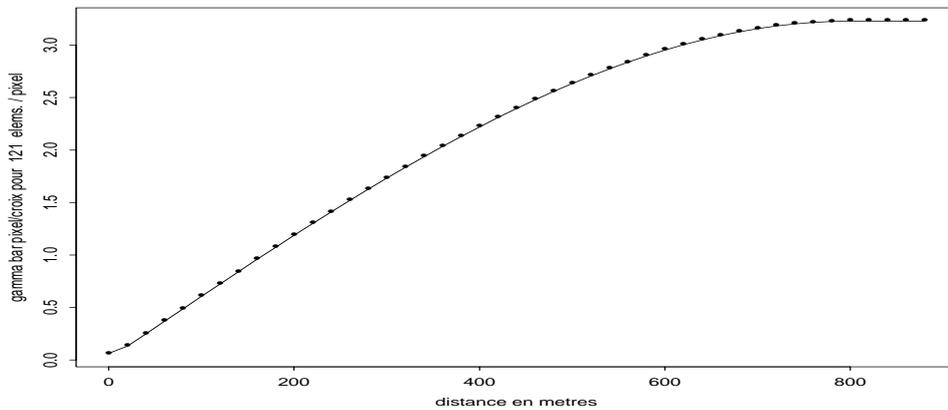


FIG. 4.11 – $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$, pour une distance variable entre le centre du pixel et le centre de la croix.

Finalement, l’effet de la géométrie de la mesure est négligeable, puisque les différences entre les expressions $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$ calculées pour les quatre configurations précédentes sont très petites, comme on le voit sur les graphiques de la figure 4.12

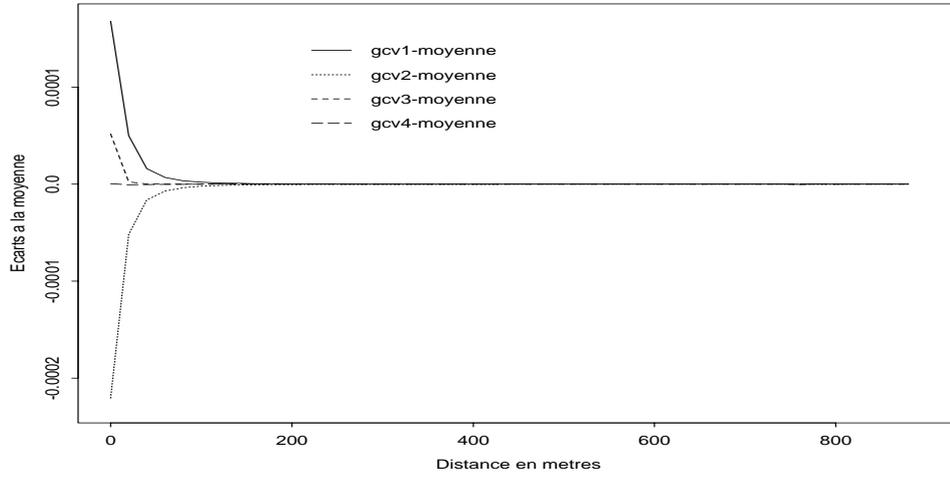


FIG. 4.12 – Représentations des différences obtenues sur les calculs de $\frac{1}{n} \sum_i \bar{\gamma}_{v,i}$ suivant les dispositions relatives croix / pixel

A ce stade, il ne nous reste plus qu'à évaluer $\bar{\gamma}_{v,v}$ (pour cela nous avons conservé la même discrétisation du pixel que précédemment) et $\bar{\gamma}_{c,c}$. Les deux courbes sont très proches des $\bar{\gamma}_{c,v}$, comme on le voit sur la figure 4.13. En effet la discrétisation d'un pixel par une croix est assez bonne.

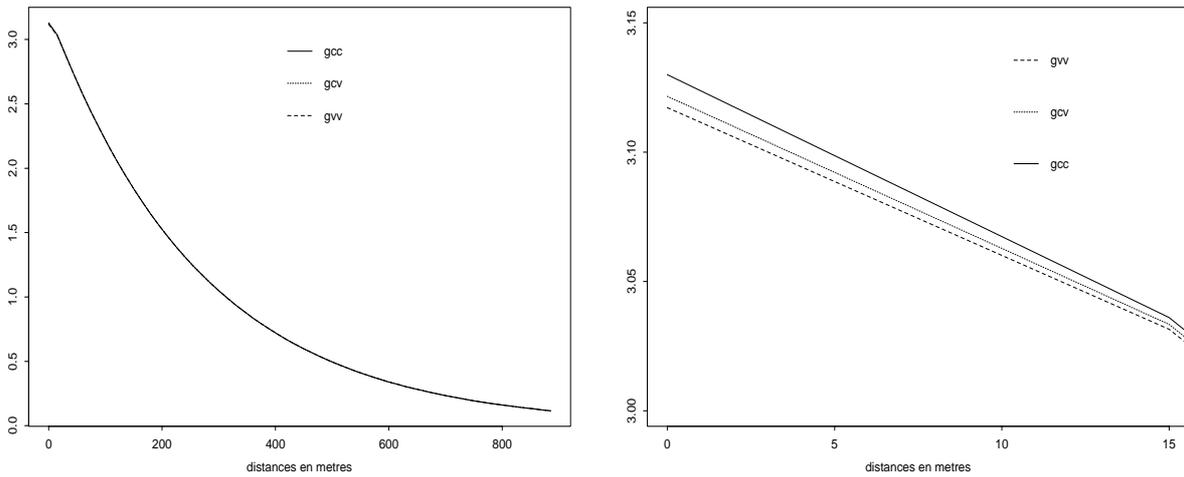


FIG. 4.13 – Représentations graphiques de $Cov_{v,v}$, $Cov_{v,c}$ et $Cov_{c,c}$.

La seule différence perceptible est à l'origine.

Finalement, nous obtenons :

$$Cov_{\varepsilon_p}(h) = \bar{\gamma}_{v,vt} + \bar{\gamma}_{c,ct} - 2\bar{\gamma}_{c,v}$$

que nous avons représenté sur la figure 4.14.

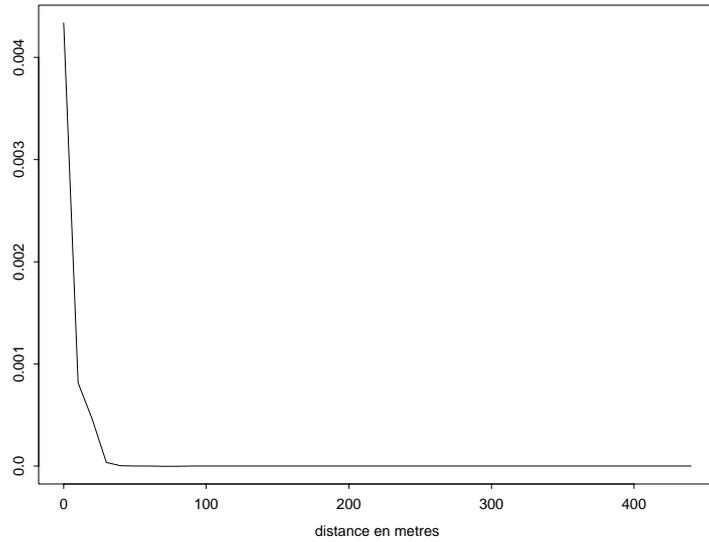


FIG. 4.14 – Covariance spatiale de l'erreur d'estimation du LAI d'un pixel par la moyenne des mesures ponctuelles.

On remarquera que les valeurs des écarts dus à la géométrie de la mesure sont bien négligeables devant les valeurs prises par cette fonction de covariance.

Estimation de $\hat{\gamma}_0(h)$ On modélise finalement $Cov_{\varepsilon_0}(h)$ par une covariance exponentielle :

$$Cov_{\varepsilon_0}(h) = a \exp\left(-\frac{|h|}{b/3}\right)$$

On détermine a et b de façon à ce que l'ajustement de $Cov_{\varepsilon} = Cov_{\varepsilon_0} + Cov_{\varepsilon_p} + Cov_{\varepsilon_m}$ soit correcte. Cependant, le terme Cov_{ε_m} , et le terme Cov_{ε_p} sont négligeables devant Cov_{ε} . Nous avons donc finalement $Cov_{\varepsilon_0}(h) \simeq Cov_{\varepsilon}(h)$. D'autre part, les résidus sont indépendants, nous pouvons donc estimer la valeur du plateau du variogramme par $Var(\varepsilon) = 0.604$. En revanche, la portée du variogramme ne peut être évaluée aisément en raison du manque de données relatives aux courtes portées. Nous avons représenté sur la figure

4.15 le variogramme expérimental et un modèle sphérique de portée 800 mètres.

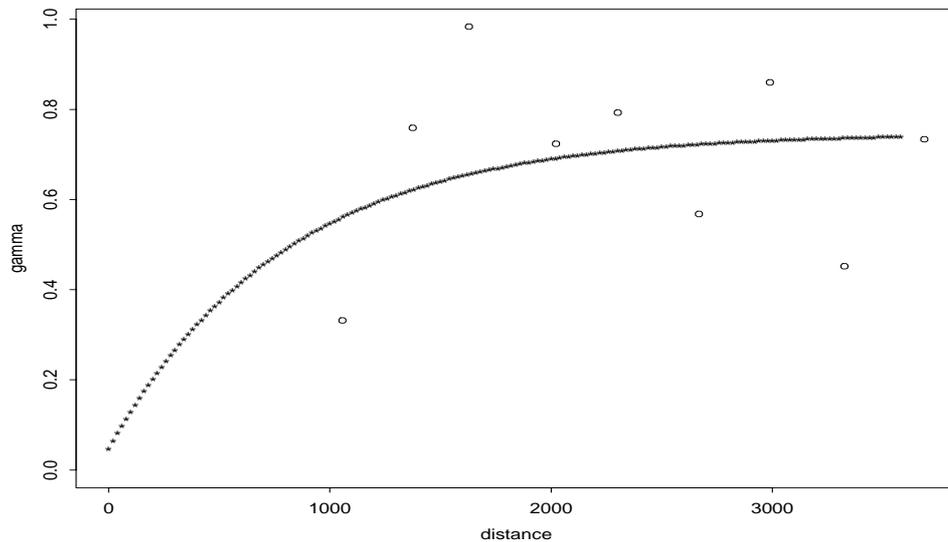


FIG. 4.15 – Variogramme des résidus, "o" expérimental, "*" modèle exponentiel pour une portée de 800 mètres.

Le paramètre de portée est cependant borné : il peut varier entre 0 et 2000, nous envisagerons dans la suite 5 cas de figure, pour des portées de 200, 500, 1000, 1500 et 2000 mètres (une portée nulle ne nous intéresse pas!).

4.4.2 Cartographie du LAI estimé et de la variance de krigeage

Nous avons repris la régression linéaire qui nous permettait d'obtenir une carte la plus précise possible : nous avons choisi de conserver la relation pour laquelle le *RMSE* était minimal :

$$LAI^* = 0.59 + 0.009\rho_{rouge} + 0.028\rho_{pir} - 0.026\rho_{vert}$$

pour les réflectances telles que LAI^* soit positif, et $LAI^* = 0$ sinon. de plus, Nous avons ici fixé arbitrairement le LAI des bois à 6. Appliquée à toute la zone cela nous donne la carte de la figure 4.16.

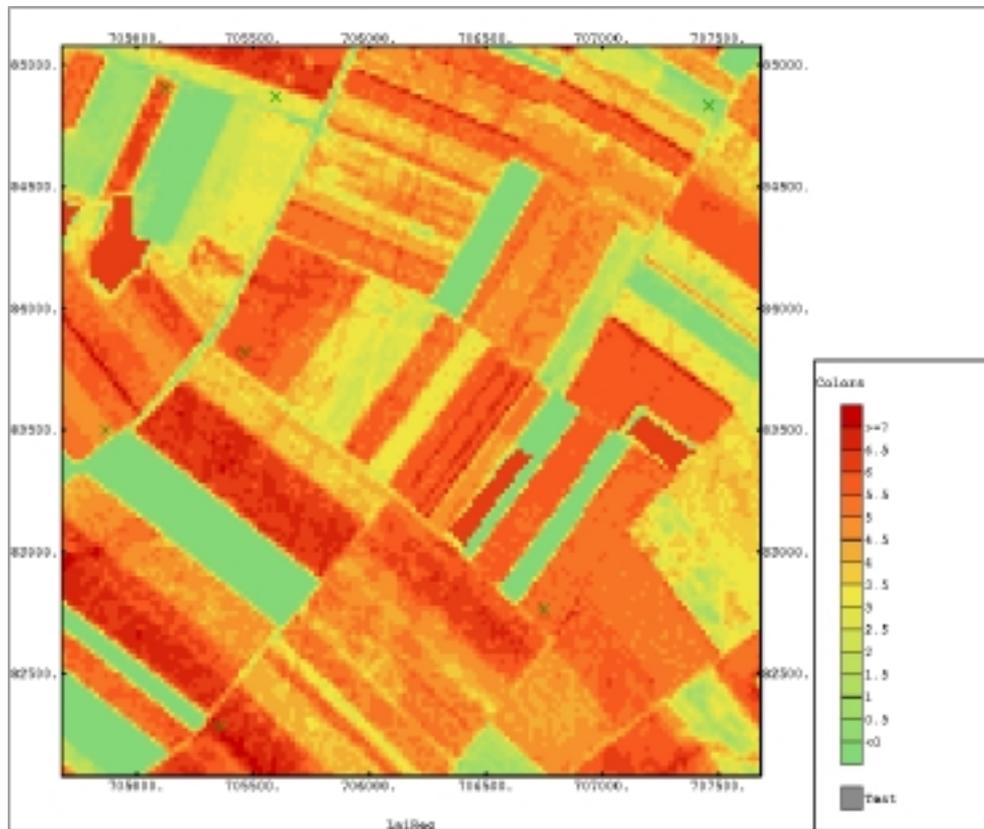


FIG. 4.16 – *Carte des LAI prédits par régression.*

En ajoutant à ces LAI prédits les termes d'erreur, nous obtenons la carte de la figure 4.17.

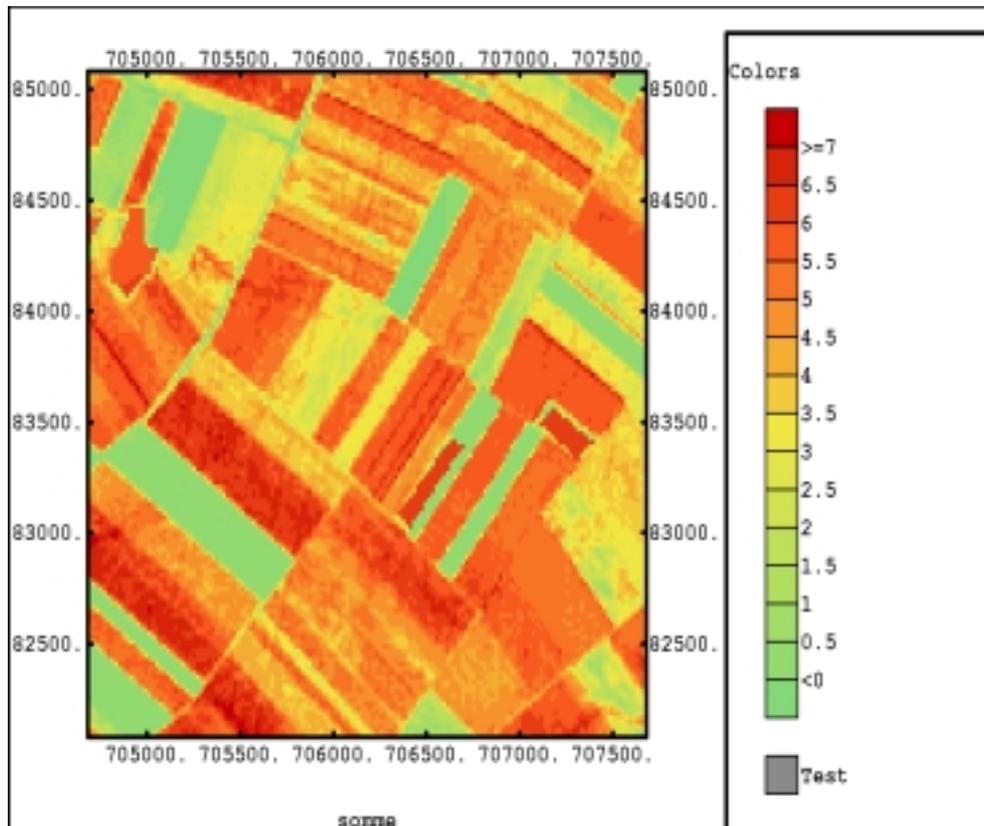


FIG. 4.17 – Carte de la somme des LAI et des résidus.

4.4.3 Agrégation

Comme nous l'avons fait précédemment, nous avons réalisé la moyenne arithmétique des LAI associés aux pixels SPOT pour estimer le LAI d'un pixel VEGETATION. En revanche nous ne pouvons composer si simplement les variances ; on a donc mis en place un krigeage de bloc des résidus afin d'obtenir les variances à l'échelle voulue. La figure 4.18 fait la synthèse de tous ces résultats.

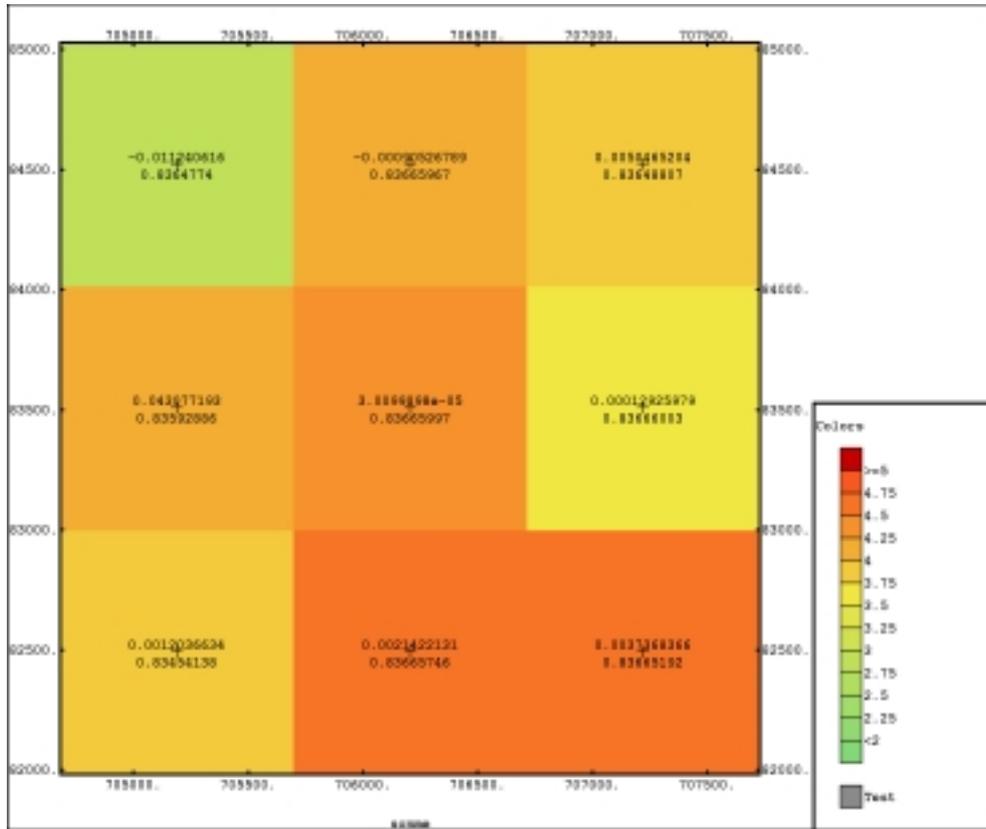


FIG. 4.18 – Cartographie des résidus + LAI prédits par régression. Les écarts types de krigeage sont précisés sous les centres de pixels, la valeur des résidus au dessus.

On a travaillé ici sur des résidus de régression, donc par définition d'espérance nulle. L'intégration sur un grand pixel tend vers l'espérance, qui est ici de 0. Cela explique les très faibles valeurs des résidus à grande échelle.

4.5 Discussion sur les techniques d'interpolation pour la cartographie de LAI

Krigeage ordinaire Nous avons calculé les estimations par krigeage ordinaire afin d'obtenir les variances maximales de cokrigeage. Le krigeage ordinaire correspond, en effet, au cas où les variables principales et secondaires ne seraient pas corrélées. Nous n'avons pas exploité les cartographies de LAI qu'il fournit : cette technique est risquée car elle repose uniquement sur les mesures réalisées, mesures qui sont très peu nombreuses sur cette petite zone (7 mesures), et probablement peu représentatives de l'ensemble des valeurs prises par les LAI sur cette zone.

Cokrigeage Nous introduisons ici une information supplémentaire : les valeurs de réflectances dans les trois bandes spectrales. Cela devrait permettre d'affiner nos estimations. Cette technique demande un effort supplémentaire de modélisation, puisqu'elle repose sur une corégionalisation. En raison du manque de données accessibles, ce qui rendait ces modélisations hasardeuses, nous avons choisi de réaliser un "colocated cokriging". Cette technique permet de faire une hypothèse de type markovienne sur les covariances croisées ; ce qui explique la restriction du voisinage des variables secondaires (les réflectances) au point d'estimation, ce qui est plutôt souhaitable : les voisins, au mieux, n'apportent pas d'information, en raison de la continuité intra parcellaire des réflectances, au pire sont dans une autre parcelle ce qui fausse l'estimation. Dans les cas où la variable principale est échantillonnée de façon clairsemée, cette information devient prédominante (l'examen des poids de krigeage le montre : en moyenne seul 10% du poids total va à la variable principale). Enfin, la conséquence de cette restriction est l'absence de modélisation spatiale des liens entre LAI et réflectances. Toutefois, le logiciel utilisé ne permettait pas l'utilisation de plusieurs variables secondaires, ce qui nous a contraint à utiliser une variable unique qui résume les réflectances. Pour cela, on a utilisé la régression du LAI par les réflectances donnée à la partie 3.3.4

Pour finir, cette méthode diminue les biais locaux du krigeage ordinaire : les cultures non échantillonnées sur la petite zone sont prises en compte au travers des LAI synthétisés à partir des réflectances.

Krigeage des résidus Tout comme la technique précédente, on corrige en partie les biais locaux par la régression, établie sur toutes les mesures. Mais surtout, cette technique permet l'utilisation de relations non linéaires, et ainsi d'ajuster éventuellement des courbes plus pertinentes qu'une simple

régression linéaire (même si l'on a vu que, dans notre cas, cela ne se justifiait pas).

En revanche, pour l'estimation du LAI au niveau de très grands pixels (comme c'est le cas ici), cette méthode n'apporte pas grand chose à la simple régression : comme on l'a vu, l'estimation des résidus sur une grande surface tend vers l'espérance, c'est à dire 0. Cela pourrait nous permettre d'évaluer la taille critique d'un pixel, au delà de laquelle l'estimation par la moyenne arithmétique des prédictions par régression serait suffisante.

Enfin, On peut donc se demander si nos deux modèles (cokrigeage et krigeage des résidus) donnent un résultat très différent, c'est pourquoi nous avons cartographié les différences entre leurs estimations. Ces cartes sont représentées à la figure 4.19.

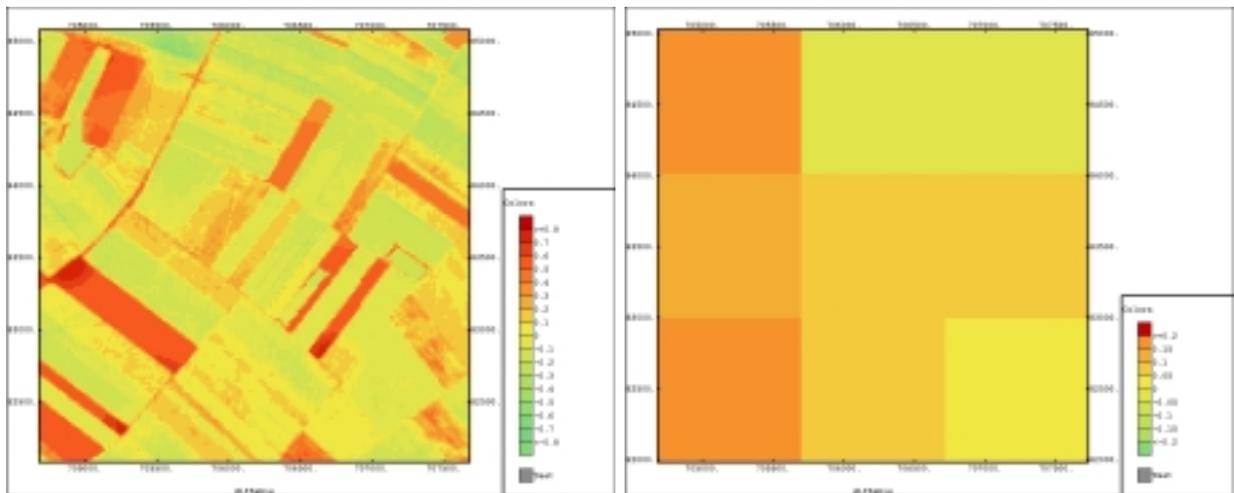


FIG. 4.19 – Différences entre les estimations obtenues par les deux modèles, à l'échelle SPOT (à gauche), intégrés à l'échelle VEGETATION (à droite).

A l'échelle d'un pixel VEGETATION, les différences maximales observées sur l'estimation du LAI sont de l'ordre de $20 \text{ cm}^2/\text{m}^2$. Ce qui n'est pas beaucoup : dans les deux cas, la même information est utilisée.

Quelles cartographies garder ? Notre régression linéaire n'étant pas très satisfaisante, les cartographies qui en sont issues sont contestables.

Avec le jeu de données dont nous disposons nous pouvons difficilement fournir une carte de LAI de référence. Mais est-ce parce que notre échantillonnage est incomplet ? ou est-ce parce que les techniques employées, classiques et éprouvées, n'y sont pas adaptées ?

Un moyen de répondre à cette question serait de mettre en place une validation croisée. L'autre réponse que nous pouvons apporter est que, même si il n'y a pas beaucoup de différences entre les cartographies par régression linéaire, par krigeage des résidus et par cokrigeage, ce qui pousserait à abandonner toute spatialisation des données, ces techniques d'estimation nous fournissent aussi un moyen de quantifier l'erreur d'estimation : les écarts types de krigeage.

4.6 Discussion sur les variances de krigeage

Si l'on fournit une cartographie de LAI estimé, il convient de fournir aussi la carte de tolérance liée aux erreurs d'estimation. Dans le cas d'une cartographie de LAI prédits par régression linéaire, cette tolérance n'est pas spatialisée, et vaut sur toute la scène le RMSE de la régression. Qu'en est-il si nous spatialisons notre estimation?

4.6.1 Erreur d'estimation du LAI

On s'intéresse aux valeurs prises par l'expression $\sqrt{\text{Var}(Z^* - Z)}$ en tout point, où Z^* désigne l'estimateur de la grandeur Z pour les différents modèles considérés, soit l'écart type de krigeage.

Variance de cokrigeage et de krigeage des résidus Dans le krigeage des résidus, la variance d'estimation $\text{Var}(Z_p - Z^*)$ au point p se ramène à la variance de krigeage de l'erreur. En effet, $Z_p = \hat{Z}_p + \varepsilon_p$. Comme d'autre part, $Z_p^* = \hat{Z}_p + \varepsilon_p^*$, on a la relation suivante :

$$\text{Var}(\varepsilon_p - \varepsilon_p^*) = \text{Var}(Z_p - \hat{Z}_p - Z_p^* + \hat{Z}_p) = \text{Var}(Z_p - Z_p^*)$$

Nous avons reproduit les cartes d'écarts types de krigeage des résidus (figure 4.20) et de cokrigeage à l'échelle SPOT (4.21) :

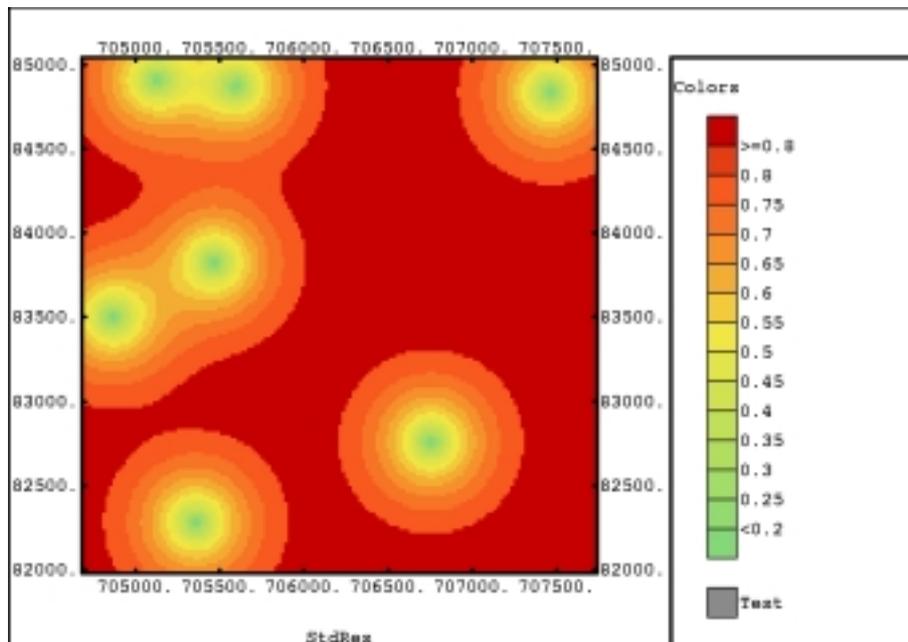


FIG. 4.20 – *Ecart types de krigeage des résidus.*

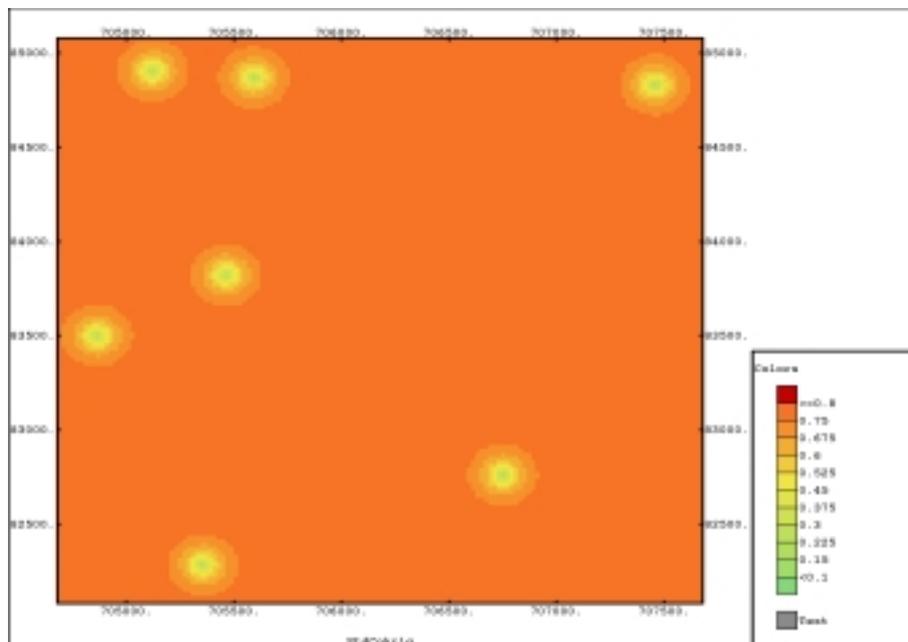


FIG. 4.21 – *Ecart types de cokrigeage.*

L'influence, sur les écarts types, de la position géographique des points

de mesures est bien plus grande dans le cas du krigeage des résidus. Cela n'est pas surprenant, puisque seuls ceux-ci interviennent dans le krigeage. Dans le cas du cokrigeage, l'introduction de la variable secondaire stabilise la variance à une valeur plus faible. L'effet des portées des structures de covariances associées à chacun des modèles est ainsi prépondérant.

Pour avoir une idée plus précise de l'influence de l'échantillonnage et des structures spatiales sur les variances de krigeage, nous avons réalisé une petite simulation dans le cas du krigeage des résidus.

4.6.2 Sensibilité des variances de krigeage

Nous avons calculé les écarts types de krigeage sur les 9 pixels de la scène, avec nos 21 échantillons, et dans un deuxième temps, avec 7 échantillons supplémentaires dans la petite zone. Ces échantillons supplémentaires ont été disposés de façon à améliorer la couverture de la zone, comme on le voit sur la figure 4.22.

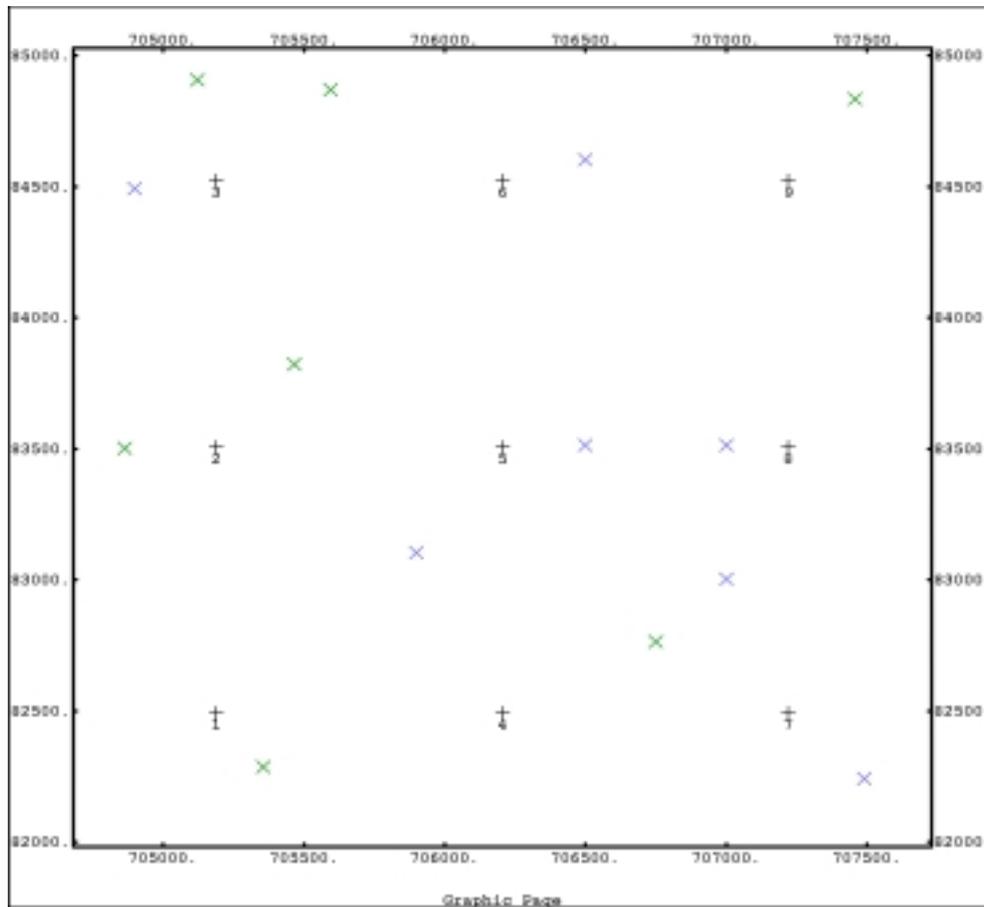


FIG. 4.22 – *Disposition des échantillons : en vert les mesures effectives, en bleu les simulations. Les 9 centres des pixels VEGETATION sont figurés en noir.*

Écart types de cokrigage

On a calculé les écart types de krigage par la méthode présentée dans la partie précédente (section 4.3.3). Dans un premier temps nous avons regardé comment ces écart variaient avec ρ , puis avec l'échantillonnage.

En faisant varier ρ , en particulier pour $\rho = 0$ on a vérifié que la valeur des écart types était identique à celle obtenue par krigage ordinaire. Pour le pixel inférieur gauche les écart types de krigage prennent les valeurs suivantes :

ρ	0	0.1	0.2	0.5	0.8	0.88	0.999
σ	0.69	0.68	0.68	0.68	0.58	0.48	0.50

Il est intéressant de noter que la variance remonte un petit peu quand ρ passe de 0.88 à 0.999. Cela est lié à une petite incohérence entre le modèle de

corégionalisation utilisé et l'hypothèse de type markovienne : si $Cov_{12}(h) = \frac{Cov_{12}(0)}{Cov_{11}(0)}Cov_{11}(h)$, alors les rapports entre les covariances en zéro devraient être égaux, et égaux à ρ . Ce n'est pas exactement le cas ici, ce qui explique ce comportement un peu étrange.

Nous avons calculé ensuite les écarts types de krigeage avec sept échantillons supplémentaires simulés (ρ vaut toujours 0.88) :

7 échantillons			14 échantillons		
0.448	0.531	0.504	0.407	0.470	0.482
0.444	0.537	0.552	0.436	0.442	0.456
0.488	0.509	0.517	0.483	0.485	0.435

Quand on rajoute des points de mesures, la part des écarts types due à l'autocovariance de la variable principale diminue, ce qui explique cette évolution.

Ecarts types de krigeage des résidus

Dans ce modèle, nous n'avons pu fixer tous les paramètres des covariances spatiales grâce à nos mesures. C'est pourquoi nous avons réalisé ces calculs pour une large gamme de portées : 200, 500, 1000, 1500 et 2000 mètres. La figure 4.23 fait la synthèse des résultats de ces calculs.

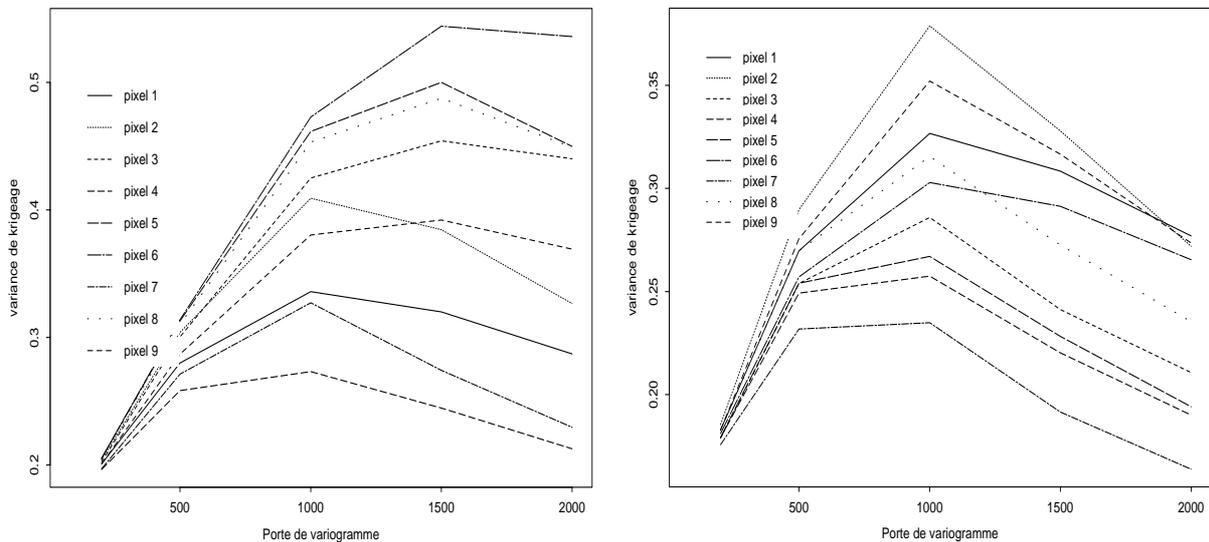


FIG. 4.23 – *Ecarts types calculés, à gauche, pour 7 échantillons, à droite pour 14.*

Résultat prévisible, la variance de krigeage diminue quand le nombre d'échantillons augmente. En revanche, l'évolution de la variance en fonction de la portée du variogramme semble répondre à deux phénomènes contradictoires.

1. Tout d'abord la hausse de la variance expérimentale avec l'augmentation de la portée de l'auto covariance: quand la portée augmente, l'information apportée par les différents voisins peut être simplement redondante, mais aussi contradictoire. Ainsi le terme $\overline{Cov}_{v_0, v_0}$ augmente avec la portée :

portée	200	500	1000	1500	2000
$\overline{Cov}_{v_0, v_0}$	0.013	0.069	0.20	0.31	0.37

2. Dans un deuxième temps, ces variances décroissent : quand la portée augmente, l'estimation en chaque point peut être faite en fonction d'un nombre croissant de voisins et avec une covariance spatiale plus forte. La qualité de notre estimation s'accroît en conséquence.

Conclusion

L'objectif était de proposer une méthode de cartographie d'indices foliaires permettant la validation de produits de capteurs à large champs. Nous avons présenté deux méthodes statistiques, intégrant le caractère spatial de l'information, qui permettent de cartographier le LAI à l'échelle voulue, et de quantifier l'erreur associée ; méthodes qui répondent à la question posée. Cependant, sur les cartographies que nous avons produit, l'erreur d'estimation est de 0.5 pour des valeurs comprises entre 3.5 et 4.5. L'erreur relative est donc assez importante.

Il faut considérer cette erreur avec prudence. Elle peut, en effet, être sous estimée : elle provient d'un modèle donné, estimé à l'aide d'un échantillonnage non exhaustif. Or on a négligé l'erreur d'estimation des paramètres du modèles. Mais elle peut, aussi bien, être sur estimée. Un échantillonnage plus dense permettrait d'affiner les relations paramétriques liant le LAI aux réflectances, et ainsi améliorer la qualité de l'estimation.

Pour avoir une idée de la qualité des estimations, une "cross-validation" permettrait de confronter (de manière partielle) les données à la réalité.

Recommencer l'exercice avec le même effort d'échantillonnage, mais sur une zone plus petite, et en ayant les moyens de le préparer mieux, permettrait d'améliorer sensiblement la qualité des cartographies produites.

Annexe A

Régressions linéaires

A.1 Régression linéaire simple

Soit Y la matrice des prédictrices et Z le vecteur des réponses, β le vecteur des coefficients et ε celui des erreurs.

Moindres carrés On considère $Z = \beta_0 + Y\beta$. On cherche $\hat{\beta}_0$ et $\hat{\beta}$ pour lesquels la somme des carrés des écarts est minimale, soit :

$$\min \sum_i (Z_i - (\beta_0 + Y\beta))^2$$

somme qui s'écrit de la façon suivante :

$$S(\hat{\beta}) = (\mu - Y\hat{\beta})^t(\mu - Y\hat{\beta})$$

où μ est le minimiseur de Lagrange, introduit de façon à respecter la contrainte de non biais. La minimisation de cette fonction sur β conduit à l'équation normale dont la solution est $\hat{\beta}$:

$$[\hat{\beta}_0 \ \hat{\beta}] = (Z^t Z)^{-1} Z^t [Y \ 1]$$

Maximum de vraisemblance Considérons maintenant le modèle suivant :

$$Z = Y\beta + \varepsilon$$

Où Y est la matrice des prédictrices, déterministe, et où ε est une variable aléatoire gaussienne. On recherche le meilleur estimateur de β (Minimum Dispersion Linear Unbiased Estimator), noté $\hat{\beta}$, qui maximise la vraisemblance

$L(\beta, Y(i), Z(i))$. Dans le cas d'une régression linéaire simple, nous faisons l'hypothèse que $\varepsilon \rightsquigarrow \mathcal{N}(0, \sigma^2 I)$ (l'espérance des résidus est nulle par définition). Cette méthode conduit au même estimateur $\hat{\beta}$, pour lequel on a bien :

$$E(\hat{\beta}) = (Z^t Z)^{-1} Z^t (Z\beta + E(\varepsilon)) = \beta$$

et

$$V(\hat{\beta}) = \sigma^2 (Z^t Z)^{-1}$$

ce qui la plus petite matrice de covariance possible, d'un estimateur de β (théorème de Gauss-Markov).

A.2 Régression linéaire généralisée

Ici, nous supposons que les ε ne sont plus indépendants. ε suit une loi gaussienne de covariance :

$$Cov(\varepsilon_i, \varepsilon_j) = C(Z_i, Z_j)$$

avec C la covariance spatiale. On a toujours $E(\varepsilon_i) = 0, \forall i$. Nous allons chercher à nous ramener à un modèle de résidus indépendants en posant : $Y' = \Sigma^{-1/2} Y$ et $\epsilon' = \Sigma^{-1/2} \epsilon$. En reprenant les résultats obtenus précédemment, nous arrivons à l'estimateur suivant :

$$\hat{\beta} = (Z^t \Sigma^{-1} Z)^{-1} Z^t \Sigma^{-1} Y$$

Annexe B

Outils géostatistiques

B.1 Le variogramme

B.1.1 Bases théoriques

Notons $Z(X)$ notre fonction aléatoire d'intérêt, h la distance entre deux sites de mesures.

- Hypothèse de stationnarité intrinsèque

Elle impose deux conditions sur les différences $Z(x+h) - Z(x)$: l'espérance de cette différence doit être nulle, pour une translation d'un vecteur h quelconque. D'autre part, la variance de cette différence a une valeur finie égale à $2\gamma(h)$. Cette variance dépend de la direction et de la norme de h , mais pas de la position de x dans le domaine. Soit :

$$E[Z(x+h) - Z(x)] = 0$$

$$Var[Z(x+h) - Z(x)] = 2\gamma(h)$$

Où $\gamma(h)$ est par définition le variogramme de $Z(x)$

- Le variogramme théorique

A partir des équations précédentes on peut écrire le variogramme γ de la façon suivante :

$$\gamma(h) = \frac{1}{2}E[(Z(x+h) - Z(x))^2]$$

C'est une fonction paire, nulle en zéro, à valeur positive.

- La fonction de covariance

Celle ci est définie sur les bases de la stationnarité des deux premiers moments :

$$\begin{cases} E[Z(x)] = m \\ E[Z(x) \cdot Z(x+h)] - m^2 = C(h) \end{cases}$$

$$\forall x, x + h \in \mathcal{D}$$

On peut toujours déduire un variogramme théorique de la fonction de covariance :

$$\gamma(h) = C(0) - C(h)$$

Cependant, l'inverse n'est pas vrai : l'hypothèse de stationnarité d'ordre deux est plus forte que l'hypothèse intrinsèque et certains variogrammes ne permettent pas d'extraire une fonction de covariance (c'est le cas, par exemple, des variogrammes non bornés).

Estimateur de $\gamma(h)$

Nous cherchons à modéliser des covariances spatiales, à partir d'un nuage de point. Comme on l'a vu, variogramme et fonction de covariance sont liés. La première étape consiste donc en la recherche d'un estimateur de γ , que nous noterons $\hat{\gamma}$. Plusieurs estimateurs ont été proposés [6] :

- Le plus courant est l'estimateur par la méthodes des moments (Matheron, 1962) :

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \cdot \sum_{N(h)} (Z(x_i) - Z(x_j))^2$$

où :

$$N(h) = \{(x_i, x_j) / x_i - x_j = h; i, j = 1, \dots, n\}$$

et $|N(h)|$ est le nombre de paires distinctes dans $N(h)$. Cet estimateur est connu comme l'estimateur "classique".

- Quand les données sont irrégulièrement réparties dans l'espace, l'estimateur précédent est instable. On introduit donc une tolérance sur h en réalisant une partition de l'espace des h en p secteurs tronqués I_p . On a alors :

$$2\gamma^*(I_p) = E[(Z(x_i) - Z(x_j))^2 : h = (x_i - x_j) \in I_p]$$

D'autres estimateurs ont été proposés, notamment pour les cas où les données sont contaminées, on parle alors d'estimateurs "résistants" ou "robustes".

B.1.2 Calcul des variogrammes expérimentaux

Nous utiliserons comme estimateur de variogramme celui que nous avons défini en second, en utilisant des intervalles réguliers. Pour l'estimation du variogramme construit sur une bande spectrale, nous calculons la moitié de

la différence au carré entre deux mesures de réflectance. Comme il est long d'étudier une par une toutes les paires possibles de points, nous avons fait l'hypothèse que la construction des variogrammes dans deux directions, selon les lignes et les colonnes de l'image, permettrait de donner une représentation satisfaisante des variations de réflectance.

Pour vérifier que cette simplification n'est pas abusive, nous avons tracé les variogrammes construits sur les trois bandes spectrales de nos images SPOT. Dans le premier cas nous les avons calculé de façon isotrope avec Splus (plusieurs heures de calcul), dans le second de façon directionnelle grâce à une routine en C (moins d'une minute de calcul).

Pour chacune des bandes spectrales les variogrammes isotropes et les variogrammes directionnels présentent les mêmes structures (portées et variances expérimentales), que ce soit pour l'image de mai ou celle de juin.

Pour obtenir ces résultats, nous avons utilisé l'algorithme suivant : soient **elem** un élément (vecteur de pixels, qu'il soit une ligne ou une colonne de l'image) de taille **nbr**, **dist** la distance entre deux pixels (distance maximale notée **max**), et que l'on note **somme** et **cpt** deux vecteurs de longueur **max**, initialisés à zéro

```

pour tout elem
  pour toute dist ≤ max
    pour tout i ≤ nbr - dist
      on ajoute (elem[i+dist] - elem[i])2 a somme[i]
      on incrémente cpt[i]
    finpour
  finpour
finpour

```

On a alors $\text{variogramme}[i] = \text{somme}[i] / \text{cpt}[i] / 2$

B.1.3 Ajustement d'une fonction de covariance

Nous avons à ce stade un estimateur γ^* du variogramme. Comment ajuster alors une fonction de covariance ?

Si le variogramme est borné (on notera alors σ^2 la valeur de son maximum), on peut toujours écrire la covariance expérimentale comme :

$$C(h) = \sigma^2 - \gamma(\hat{h})$$

Cependant, on préfère souvent ajuster une fonction de covariance théorique, aux propriétés mathématiques connues en tout point. Citons trois exemples

de fonction de covariance, qui seront ajustées graphiquement par rapport au tracé du variogramme ([25]) :

- le modèle “effet pépite”

Cette fonction de covariance modélise une discontinuité à l’origine :

$$C_{pep}(h) = \sigma^2 \text{ pour } |h| = 0$$

$$C_{pep}(h) = 0 \text{ pour } |h| > 0$$

Ce type de modèle est souvent interprété comme le bruit qui accompagne le signal étudié.

- le modèle exponentiel

Ce modèle traduit une décroissance exponentielle de la covariance avec la distance :

$$C_{exp}(h) = b \exp -\frac{|h|}{a}$$

le paramètre a étant alors un coefficient d’extinction. A la distance $|h| = 3a$, la covariance a chuté de 95%. On parle donc de cette distance comme de la “porté pratique”.

- le modèle sphérique

Ce modèle introduit aussi deux paramètres : comme précédemment σ^2 représente la valeur maximale de la covariance, a représente la portée réelle. Le modèle s’écrit alors :

$$C_{sph}(h) = \sigma^2 \cdot \left(1 - \frac{3}{2} \frac{|h|}{a} + \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right) \text{ pour } 0 \leq |h| \leq a$$

$$C_{sph} = 0 \text{ pour } |h| > a$$

Il arrive souvent que l’on combine différents modèles, par exemple sphérique et pépite ou l’on retrouve l’analogie signal / bruit.

B.2 Variogrammes croisés ou covariogrammes

Définition du covariogramme On a développé la démarche conduisant à modéliser la covariance spatiale entre des mesures de même nature, par exemple, les covariances entre les réflectances mesurées dans la bande verte. Cependant, on peut souhaiter modéliser les corrélation entre des mesures dans le vert et le rouge. Si $Z(x) = (Z_i(x), \dots, Z_k(x))$ est un processus spatial multivariable, on peut exprimer la notion de dépendance croisée à l’aide du variogramme croisé, parfois appelé covariogramme ([6]) :

$$\gamma_{ij}^*(h) = \frac{1}{2} \cdot Cov[(Z_i(x+h) - Z_i(x)), (Z_j(x+h) - Z_j(x))] \text{ avec } i, j \in [1, \dots, k]$$

On estimera alors le variogramme croisée par ::

$$\hat{\gamma}_{ij}^*(h) = \frac{1}{2} \frac{1}{|N(h)|} \sum_{x_i - x_j \sim h} [(Z_i(x+h) - Z_i(x)) \cdot (Z_j(x+h) - Z_j(x)) : h \in I_p]$$

Si les hypothèses de stationarité sont respectées, la covariance croisée pour les bandes i et j s'écrit de la manière suivante :

$$Cov_{ij}(h) = E[(Z_i(x) - m_i), (Z_j(x+h) - m_j)]$$

en notant $m_i = E[Z_i]$ et $m_j = E[Z_j]$. Avec la définition précédente du variogramme croisé, on peut écrire la relation ([13]) :

$$\gamma_{ij}(h) = Cov_{ij}(0) - \frac{1}{2} \cdot [Cov_{ij}(h) + Cov_{ij}(-h)]$$

Propriétés Comme toute fonction, la fonction de covariance croisée peut être écrite comme la somme d'une fonction paire et d'une fonction impaire :

$$Cov_{ij}(h) = \frac{1}{2} \cdot [Cov_{ij}(h) + Cov_{ij}(-h)] + \frac{1}{2} \cdot [Cov_{ij}(h) - Cov_{ij}(-h)]$$

γ_{ij} ne traduit que la partie paire ; ainsi la covariance croisée est une fonction plus riche que le variogramme croisé, mais qui nécessite des hypothèses plus fortes. Dans le cas isotrope :

$$Cov_{ij}(h) = Cov_{ij}(-h) = Cov_{ij}(|h|)$$

Dans ces conditions $Cov_{ij}(h) = Cov_{ji}(h)$ et

$$\gamma_{ij}(h) = Cov_{ij}(0) - Cov_{ij}(h)$$

Enfin, la corrélation entre deux vecteurs aléatoires $Z_i(x)$ et $Z_j(x+h)$ tend vers 0 quand h tend vers l'infini. Ainsi, pour les covariogrammes bornés, on a :

$$\gamma_{ij}(h) \rightarrow Cov_{ij}(0) \quad \text{quand } |h| \rightarrow \infty$$

Co-dispersion Par analogie avec le coefficient de corrélation ρ on définit le corrélogramme $\zeta(h)$:

$$\rho = \frac{Cov_{ij}}{\sqrt{Var_i \cdot Var_j}}$$

$$\zeta(h) = \frac{Cov_{ij}(h)}{\sqrt{Cov_{ii}(0) \cdot Cov_{jj}(0)}}$$

Il est cependant souvent plus intéressant de calculer le coefficient de co-dispersion $\eta(h)$, défini de la façon suivante:

$$\eta(h) = \frac{\gamma_{ij}(h)}{\sqrt{\gamma_{ii}(h) \cdot \gamma_{jj}(h)}}$$

Ce coefficient est identiquement nul en l'absence de corrélation spatiale entre $Z_i(h)$ et $Z_j(h)$.

Annexe C

Présentation des techniques de krigage

D'une manière générale, le krigage est une technique d'estimation linéaire, non biaisée et de variance minimale, pour des données spatialisées dont la covariance spatiale ou le variogramme est connu.

C.1 Krigage de la moyenne

Soit \hat{m} l'estimateur linéaire de la moyenne du LAI.

$$\hat{m} = \sum_{\alpha=1}^n w_{\alpha} Z(x_{\alpha})$$

avec $Z(x_{\alpha})$ les mesures ponctuelles de LAI aux points α et w_{α} les pondérateurs de $Z(x_{\alpha})$. La condition de non biais nous impose d'avoir $\sum_{\alpha=1}^n w_{\alpha} = 1$. La variance de l'erreur d'estimation s'écrit alors :

$$Var(\hat{m} - m) = \sum_{\alpha=1}^n \sum_{\beta=1}^n w_{\alpha} w_{\beta} Cov(x_{\alpha} - x_{\beta})$$

La minimisation de cette variance équivaut à résoudre le système formé par l'annulation des $n + 1$ premières dérivées partielles de l'équation suivante :

$$\phi(w_{\alpha}, \mu) = Var(\hat{m} - m) - 2\mu \left(\sum_{\alpha=1}^n w_{\alpha} - 1 \right)$$

en introduisant le multiplicateur de Lagrange μ de façon à satisfaire à la contrainte de non biais. Cela conduit au système suivant :

$$\forall \alpha \in [1, n] \quad \sum_{\beta=1}^n w_{\beta} Cov(x_{\alpha} - x_{\beta}) = \mu$$

et la condition de non biais : $\sum_{\beta=1}^n w_\beta = 1$

C.2 Krigeage ordinaire et krigeage de bloc

On estime ici les valeurs prises par Z sur un support en fonction des mesures alentours. Si l'on veut estimer Z sur un support de même nature que celui des mesures, on parlera de krigeage ordinaire. Si on estime la valeur de Z sur un bloc v plus grand que le support des mesures, on calculera la moyenne de Z sur ce bloc :

$$Z_v = \frac{1}{|v|} \int_v Z(x) dx$$

on réalisera alors un krigeage de bloc. Cependant, dans les deux cas, la méthode d'estimation est la même. La différence réside dans l'emploi, par la technique de bloc de variogrammes moyens :

Les variogrammes moyens

On note $\bar{\gamma}_{x,v}$ la moyenne des variogrammes entre la mesure x et le bloc v d'intérêt. On aura les relation suivantes :

$$\bar{\gamma}_{x,v'} = \frac{1}{|v'|} \int_{v'} \gamma_{x,x'} dx'$$

$$\bar{\gamma}_{v,v'} = \frac{1}{|v|} \int_v \int_{v'} \gamma_{x,x'} dx dx'$$

où $\gamma_{x,x'}$ est une notation abrégée pour $\gamma(x - x')$

Matrice de krigeage

On construit un estimateur linéaire $\bar{Z}_v^* = \sum_\alpha \lambda_\alpha Z(x_\alpha)$. Les coefficients λ_α sont déterminés en v_0 par la résolution du système suivant où μ est le minimisateur de Lagrange introduit de façon à satisfaire à la contrainte de non biais $\sum_\alpha \lambda_\alpha = 1$:

$$\begin{pmatrix} \gamma_{x_1,x_1} & \cdots & \gamma_{x_1,x_n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{x_n,x_1} & \cdots & \gamma_{x_n,x_n} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \bar{\gamma}_{x_1,v_0} \\ \vdots \\ \bar{\gamma}_{x_n,v_0} \\ 1 \end{pmatrix}$$

Nous avons ici présenté la méthode pour le krigeage de bloc. Pour réaliser un krigeage ordinaire il suffit de remplacer $\bar{\gamma}_{x,v}$ par γ_{xx} , dans la matrice de krigeage.

Variance de krigage

La variance de krigage en v_0 se calcul de la façon suivante :

$$\sigma^2 = -\mu - \bar{\gamma}_{v_0, v_0} + \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}_{x_{\alpha}, v_0}$$

Quand v_0 tend vers l'infini, σ^2 tend vers la variance de krigage de la moyenne. Cela est aussi vrai pour les valeurs estimées.

C.3 Cokrigage

Lorsqu'on a accès à une variable secondaire, corrélée à la variable que l'on cherche à estimer, il peut être intéressant de l'utiliser pour améliorer l'estimation. Cela est d'autant plus vrai si cette variable secondaire est plus densément échantillonnée que notre variable principale.

Modèle de cokrigage

On construit l'estimateur de cokrigage $Z^*(x)$, pour Z la variable principale et Y la variable secondaire, de la façon suivante :

$$Z^*(x) = \sum_{i=1}^n \lambda_i Z(x_i) + \sum_{j=1}^m \delta_j Y(x_j)$$

La condition de non biais ($E[Z^*(x)] = Z(x)$) impose deux contraintes : $\sum_{i=1}^n \lambda_i = 1$ et $\sum_{j=1}^m \delta_j = 0$. La minimisation de la variance d'estimation conduit à établir le système suivant :

Matrice de cokrigage

On introduit μ_Z et μ_Y minimiseurs de Lagrange, de façon à respecter les deux conditions de non biais. Les $n + m$ premières dérivées partielles annulées, ainsi que les deux conditions de non biais conduisent au système

suivant :

$$\begin{pmatrix} \gamma_{Z_1, Z_1} & \cdots & \gamma_{Z_1, Z_n} & \gamma_{Z_1, Y_1} & \cdots & \gamma_{Z_1, Y_m} & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{Z_n, Z_1} & \cdots & \gamma_{Z_n, Z_n} & \gamma_{Z_n, Y_1} & \cdots & \gamma_{Z_n, Y_m} & 1 & 0 \\ \gamma_{Y_1, Z_1} & \cdots & \gamma_{Y_1, Z_n} & \gamma_{Y_1, Y_1} & \cdots & \gamma_{Y_1, Y_m} & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{Y_m, Z_1} & \cdots & \gamma_{Y_m, Z_n} & \gamma_{Y_m, Y_1} & \cdots & \gamma_{Y_m, Y_m} & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \delta_1 \\ \vdots \\ \delta_m \\ \mu_Z \\ \mu_Y \end{pmatrix} = \begin{pmatrix} \gamma_{Z_1, Z_0} \\ \vdots \\ \gamma_{Z_n, Z_0} \\ \gamma_{Y_1, Z_0} \\ \vdots \\ \gamma_{Y_n, Z_0} \\ 1 \\ 0 \end{pmatrix}$$

La résolution de ce système nous fournit ainsi, pour tout les V_0 , les λ_i et δ_j qui définissent $Z^*(v_0)$. Si l'on s'intéresse à une zone étendue (un grand nombre de points), la matrice de cokrigage enflera rapidement au point de dépasser les capacités de calcul des machines. D'autre part, la variable secondaire introduit souvent une information redondante, qui peut introduire une instabilité. Il peut donc être intéressant de s'intéresser à un voisinage restreint pour la variable secondaire.

"Colocated cokriging"

On s'intéresse ici au voisinage restreint au maximum, puisque l'on ne prend en compte la variable secondaire qu'au point d'estimation. Cela permet d'éviter que celle-ci prenne trop d'importance par rapport à la variable principale. On ne peut plus imposer qu'une seule condition de non biais, car imposer une somme nulle aux poids de la variable secondaire n'a plus beaucoup de sens quand il n'y en a qu'un. Ainsi dans le cas où les deux variables ont la même espérance, l'estimateur devient le suivant :

$$Z^*(x) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) + \delta Y(x)$$

Avec une seule contrainte de non biais :

$$\sum_{\alpha=1}^n \lambda_{\alpha} + \delta = 1$$

Pour simplifier le modèle, on peut faire une hypothèse de type markovienne : La technique du "colocated cokriging" n'impose la connaissance que de l'autocovariance de la variable principale et de la covariance croisée, et on peut supposer que :

$$Cov_{12}(h) = \frac{Cov_{12}(0)}{Cov_{11}(0)} \cdot Cov_{11}(h)$$

(Une des conditions nécessaire à cette approximation est que seule la variable secondaire au point d'estimation est corrélée à la variable principale.)
Si cette hypothèse de type markovienne peut être faite, on n'a alors à modéliser que l'autocovariance.

Bibliographie

- [1] F. BARET. Un modèle simplifié de réflectance et d'absorptance d'un couvert végétal. *4ème Colloque International Signatures Spectrales d'Objets en Télédétection*, pages 113–120, 1988.
- [2] F. BARET and G. GUYOT. Potentials and limits of vegetation indices for lai and apar assessment. *Remote Sensing of Environment*, 35:161–173, 1991.
- [3] J.P. CHILÈS and P. DELFINER. *Geostatistics, Modeling Spatial Uncertainty*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, 1999. Chapter 2.4. Regularization and Nugget Effect ; 3.5. Estimation of Spatial Average.
- [4] J. G. CLEVERS. The derivation of a simple reflectance model for the estimation of leaf area index. *Remote Sensing of Environment*, 25:53–69, 1988.
- [5] N. A. C. CRESSIE. *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, 1991. Chapter 5.1 : nonlinear geostatistics et 5.2 : change of support.
- [6] N. A. C. CRESSIE. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, 1991. Chapter 2.4 : Estimation of the Variogram, chapter 2.3.2 : Covariogram and Correlogram.
- [7] P.J. CURRAN. Selecting a spatial resolution for estimation of per field green leaf area index. *International Journal of Remote Sensing*, 9:1243–1250, 1988.
- [8] P.J. CURRAN and A.M. HAY. The importance of measurement error for certain procedures in remote sensing at optical wavelengths. *Photogrammetric Engineering & Remote Sensing*, 52:229–241, 1986.
- [9] P.J. CURRAN and H.D. WILLIAMSON. Sample size for ground and remotely sensed data. *Remote Sensing of Environment*, 20:31–34, 1986.

- [10] P.J. CURRAN and H.D. WILLIAMSON. Glai estimation using measurements of red, near infrared and middle infrared radiance. *Photogrammetric Engineering & Remote Sensing*, 53:181–186, 1988.
- [11] M. YERBRUGGHE D. WU, X. GU and G. GUYOT. Radiometric characterisation of dunhuang satellite calibration test site. *7th International Symposium on Physical Measurements and Signatures in Remote Sensing, Courchevel, Balkema publ. Rotterdam*, pages 151–160, 1997.
- [12] J. L. DEUZÉ E. VERMOTE, D. TANRÉ and M. HERMAN. *Second Simulation of the Satellite Signal in the Solar Spectrum (6S)*. 6s User Guide Version 2, 1997.
- [13] P. GOOVAERTS. *Geostatistics for Natural Ressources Evaluation*. Applied Geostatistics Series. Oxford University Press, 1997. Chapter 3.3.3 Multivariate random function ; 6.2.9 Multivariate factorial kriging.
- [14] J. WANG H. GAO and P. ZHAO. The update kriging variance and optimal sample design. *Mathematical Geology*, 28:295–314, 1996.
- [15] B. ANDRIEU J. C. PRICE, M. D. STEVEN and K. JAGGARD. Visible near-infrared radiance parameters for sugar beets. *International Journal of Remote Sensing*, 17:3411–3418, 1996.
- [16] M. Th. van GENUCHTEN K. C. ABBASPOUR, R. SCHULIN and E. SCHLAPPI. An alternative to cokriging for situations with small sample sizes. *Mathematical Geology*, 30:259–274, 1997.
- [17] Y. J. KAUFMAN and C. SENDRA. Algorithm for automatic atmospheric corrections to visible and near infra red satellite imagery. *International Journal of Remote Sensing*, 9:1357–1381, 1988.
- [18] Y. J. KAUFMAN and D. TANRÉ. Direct and inderect methods for correcting the aerosol effect on remote sensing. *6th International Symposium on Physical Measurements and Signatures in Remote Sensing, Val d'Isère, CNES édition*, pages 7–19, 1994.
- [19] M.LEROY M. WEISS, F.BARET. Validation of neural net techniques to estimate canopy biophysical variables from remote sensing data. *submitted to Physics and Chemistry of the Earth*, 2000.
- [20] Remote Sensing of Environment. *Special Issue: Validating MODIS Terrestrial Ecology Products*. Elsevier, 1999. volume 70, number 1, pp 1-128.
- [21] I OLKIN and J. W. PRATT. *Annals of Mathematics and Statistics*, volume 29. 1958.
- [22] G. PAN. Restricted kriging: a link between sample value and sample configuration. *Mathematical Geology*, 26:135–155, 1996.

- [23] J. C. PRICE and W. C. BAUSCH. Leaf area index estimation from visible and near-infrared reflectance data. *Remote Sensing of Environment*, 52:55–65, 1995.
- [24] B. D. RIPLEY. *Spatial Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley Interscience, 1981. Chapter 10, 11 et 12: change of support.
- [25] Hans WACKERNAGEL. *Multivariate Geostatistics*. Springer, 1995.