

Comparaison de procédures d'estimation dans le cadre des modèles non linéaires à paramètres aléatoires : application à la modélisation de l'évolution temporelle de l'indice de surface foliaire de cultures observées par télédétection spatiale

THÈSE

présentée et soutenue publiquement le 4 juillet 2002

pour l'obtention du

Doctorat de l'Université Toulouse III - Paul Sabatier

Discipline : Mathématiques appliquées (spécialité Statistique)

par

Sébastien DEJEAN

Composition du jury

Président : Jean-Marc AZAÏS, Université Toulouse 3

Rapporteurs : Christian LAVERGNE, Université Montpellier 3 - INRIA Rhône-Alpes
Hervé MONOD, INRA Jouy-en-Josas (non présent)

Examineurs : Didier CONCORDET, École Nationale Vétérinaire de Toulouse
Gérard DEDIEU, Centre d'Études Spatiales de la BIOSphère, Toulouse
Robert FAIVRE, INRA Toulouse

Directeur de thèse : Jean-Marc AZAÏS, Université Toulouse 3
Michel MESTE, Université de Perpignan (non présent)

Remerciements

Je tiens tout d'abord à remercier Robert Faivre et Michel Goulard qui m'ont intégré à leur équipe et ont accepté de m'encadrer tout au long de mes travaux de thèse. Je souhaite associer à ces remerciements mes deux directeurs de thèse Jean-Marc Azaïs et Michel Meste qui, comme convenu au départ, m'ont apporté une aide ponctuelle mais efficace.

Ensuite, c'est à Christian Lavergne et Hervé Monod, que j'adresse mes sincères remerciements pour avoir accepté d'être les rapporteurs de mon travail et à Gérard Dedieu et Didier Concordet pour leur participation au jury.

Je tiens également à associer Pascal Sarda, Alain Trubuil et Richard Delécolle dont les conseils prodigués lors des réunions du comité de pilotage m'ont à chaque fois redonné de l'élan pour poursuivre mes travaux.

Je souhaite aussi saluer tous les membres de l'unité de Biométrie et Intelligence Artificielle de l'INRA de Toulouse qui m'ont accueilli et fourni des conditions de travail idéales durant les presque quatre années que j'y ai passées. Je pense notamment aux Wonderwomen de l'équipe administrative qui m'ont épargné la plupart des galères dues aux diverses paperasseries ; alors merci à Monique, Maïté, Jacky, Pascale, Martine et Nelly. Je pense ensuite aux Supermen du réseau, les jumeaux Martin et Abde, qui parviennent à résoudre à longueur de journée toutes les tracasseries informatiques par la seule imposition des mains sur le clavier de la bête. Je n'oublie pas tous les autres, que je ne cite pas pour n'oublier personne, mais je les remercie pour l'ambiance amicale et chaleureuse qu'ils ont su créer autour de moi. À travers Laurence (à qui j'ai fait faire de gros progrès en \LaTeX) et Christophe, c'est tous ceux avec qui j'ai partagé mon bureau que je souhaite remercier pour leur bonne humeur et le soutien qu'ils m'ont apporté.

C'est à mes parents que j'adresse toute mon affection et ma reconnaissance pour m'avoir permis d'en arriver là. Et enfin, à celle qui partage ma vie depuis plusieurs années, c'est bien plus qu'un merci que j'adresse...

*À ma famille
avec une pensée particulière pour ceux qui nous ont quittés durant ces dernières années*

Résumé

Dans le cadre de la télédétection spatiale au service de l'agriculture, nous nous intéressons à l'exploitation d'images à haute résolution temporelle pour estimer les paramètres d'un modèle d'évolution temporelle de l'indice de surface foliaire (LAI) d'une culture. Ces informations utilisées en complément d'un modèle de développement sont susceptibles d'améliorer la prévision de rendement sur une région agricole. Les acquisitions effectuées au cours du temps sur une même zone, assimilables à des données longitudinales sur une population de pixels, fournissent des valeurs de LAI ajustables par une fonction non linéaire caractérisant les phases de croissance et de sénescence de la plante. Actuellement, cette fonction est ajustée en considérant séparément chaque pixel. Pour traiter l'ensemble des pixels simultanément et exprimer explicitement les variabilités intra-pixel et inter-pixels, nous proposons une modélisation par paramètres aléatoires.

Nous présentons une étude bibliographique des différentes spécifications d'un modèle à paramètres aléatoires. La variabilité intra-pixel s'exprime à un premier niveau par une structure de covariance sur le terme résiduel d'une régression non linéaire classique. À un second niveau, une loi de probabilités sur les paramètres du modèle non linéaire traduit la variabilité inter-pixels. Selon les hypothèses que l'on admet sur cette distribution, on a recours à des approches non paramétriques (sans restriction sur la forme de la distribution), paramétriques classiques (la famille des distributions est imposée), voire bayésiennes (approche paramétrique complétée par une information *a priori* sur les paramètres de la distribution). Dans notre cas, nous justifions l'utilisation d'un modèle paramétrique classique comme une première étape vers l'emploi de méthodes non paramétriques moins restrictives.

Dans le cadre du modèle paramétrique, nous passons en revue diverses méthodes d'estimation rencontrées dans la littérature. Elles sont fondées soit sur le traitement séparé des individus soit sur la maximisation d'une vraisemblance globale. Dans ce dernier cas, diverses stratégies ont été proposées : mise en œuvre de l'algorithme EM, approximation numérique (Monte Carlo, quadrature gaussienne), modification du modèle initial en un modèle où la vraisemblance est plus facile à maximiser (linéarisation, pseudo-vraisemblance). La convergence, la normalité asymptotique et l'efficacité de ces divers estimateurs dépendent de conditions différentes selon les méthodes. Certaines, dont la linéarisation, nécessitent un nombre d'individus et un nombre de dates par individu qui tendent vers l'infini. Pour d'autres (pseudo-vraisemblance), le nombre de dates peut être uniformément borné. Nous établissons une synthèse des propriétés des estimateurs issus de ces diverses méthodes ; ceci comble un vide dans la littérature qui se consacre généralement aux aspects numériques des méthodes.

Les exemples traités dans la littérature sont généralement de petite taille par rapport à notre problème. Pour juger du caractère opérationnel des méthodes, nous mettons en œuvre des comparaisons sur données simulées. La première série de tests nous permet de valider dans notre cas l'emploi de la méthode par linéarisation. Ensuite, une deuxième série montre que l'estimation par linéarisation dans le modèle à paramètres aléatoires présente une plus grande robustesse que l'approche par pixel face à des conditions défavorables : faible nombre de dates, données fortement bruitées. Enfin, nous étudions l'influence de structures spatiales dans la simulation de données. Nous mettons d'abord en évidence que la présence d'une structure spatiale sur le bruit n'influence pas la qualité de l'estimation des paramètres du modèle. Ensuite, la simulation de paramètres avec des corrélations spatiales inter-pixels montre que l'estimation par linéarisation parvient à retrouver les zones homogènes ainsi simulées à condition que le bruit résiduel ne soit pas trop important.

Pour illustrer les travaux et conforter les précédents résultats, nous procédons à une étude comparative des estimations par pixel et par linéarisation sur un jeu de données réelles.

Mots-clés : modèle non linéaire mixte, données longitudinales, image satellite, SPOT/Végétation, suivi de cultures.

Abstract

In the spatial remote sensing for agriculture framework, we deal with the use of high temporal resolution images to estimate crop leaf area index (LAI) nonlinear time evolution model parameters. These information used with a crop simulation model can improve yield prediction over a large area. Regular measurements over the same area, assimilable to longitudinal data, give LAI values that can be fitted by a nonlinear function characterizing plant growth and senescence. At present, the nonlinear model is fitted with a per-pixel based approach. To deal with the whole pixels simultaneously and to express explicitly within-pixel and between-pixels variability, we propose to use random parameter nonlinear models.

We present a bibliographical study of different specifications for the random parameter nonlinear models. Within-pixel variability is stated at a first stage through a covariance structure on the residual term of a classical nonlinear regression. At a second stage, a probability distribution on the nonlinear model parameters represents the between-pixels variability. According to the assumptions made on this distribution, we use non parametric approaches (no restriction on the distribution form), classical parametric (the distribution is assumed to belong to a particular family), or bayesian (parametric approach completed by an *a priori* information on parameters distribution). In our case, we justify the use of a classical parametric model as a first step to less restrictive methods as non parametric ones.

In the parametric model framework, we study several estimation methods. They are based either on individuals processes or on the maximum likelihood. In the latter case, several strategies were proposed: EM algorithm implementation, numerical approximations (Monte Carlo, Gaussian quadrature), transformation of the initial model where the likelihood is easier to be maximised (linearisation, pseudo-likelihood). Estimators consistency, asymptotic normality and efficiency rely on different conditions according to the methods. As the linearisation, some methods need both an infinite number of individuals and an infinite number of dates per individual. For others (pseudo-likelihood), the number of dates can be uniformly bounded. We establish a synthesis of estimators properties; this fills a lack in the bibliography which is mainly concerned with numerical aspects.

Examples in the articles are usually small-sized in relation with our problem. To evaluate the computational efficiency of the methods, we perform comparisons on simulated data. The first test series enables us to assess, in our case, the use of the linearisation method which the main advantage is its computational efficiency facing numerical approximations based methods. Then, a second series shows that the linearisation estimation in a random parameter model is more robust than per-pixel based approach towards unfavourable conditions: low number of dates, strong-noisy data. Finally, we study the influence of spatial structures in the simulation process. The estimation *via* a random parameter model is not more disturbed when the noise is spatially structured. Moreover, it gets different areas if the parameters are spatially correlated provided that the noise remains relatively low. Simulating data according to spatial structures show that the estimation through random parameters gets this structure without specifying it in the model.

To illustrate these works, we implement a comparative study of per-pixel and random parameter based approaches on a real dataset.

Key-words : nonlinear mixed effects model, longitudinal data, satellite image, SPOT/Végétation, crop monitoring.

Table des matières

Introduction	1
1 La télédétection spatiale au service de l'agriculture	5
1.1 Généralités	5
1.1.1 Définitions liées à la télédétection spatiale	5
1.1.1.1 Caractéristiques techniques des capteurs de télédétection	6
1.1.1.2 Contraintes liées à l'utilisation d'un satellite	7
1.1.2 Problématiques liées à l'agriculture	10
1.1.2.1 Échelle parcellaire	10
1.1.2.2 Bassin de collecte	10
1.1.2.3 Échelles nationale, continentale et mondiale	11
1.2 Contexte de l'étude	12
1.2.1 SPOT 4	13
1.2.1.1 Caractéristiques des capteurs HRVIR et VEGETATION	13
1.2.1.2 Mesures de réflectances	14
1.2.2 Indices de végétation	15
1.2.2.1 Calcul et choix d'un indice de végétation	15
1.2.2.2 Du NDVI au LAI	16
1.2.3 Modèle d'évolution temporelle du LAI	17
1.2.3.1 Modèle de Baret	17
1.2.3.2 Utilisation d'un modèle temporel du LAI	19
1.3 Analyse de l'existant	20
1.3.1 Les données	21
1.3.2 Traitement des données à haute résolution spatiale	22
1.3.3 Traitement des données à haute résolution temporelle	23
1.3.3.1 Hypothèse de base sur les pixels mixtes	23
1.3.3.2 Résolution du problème de désagrégation	24

1.3.4	Ajustement d'un modèle d'évolution temporelle du LAI	25
1.4	Amorce des travaux de recherche	27
1.4.1	Pistes de recherche	27
1.4.2	Question de recherche	28
1.4.3	Proposition	29
2	Le modèle non linéaire à paramètres aléatoires	31
2.1	Généralités	31
2.2	Différentes spécifications pour l'expression de la variabilité inter-individus .	32
2.2.1	Approche non paramétrique	32
2.2.1.1	Écriture du modèle	32
2.2.1.2	Estimation	33
2.2.2	Approche semi-paramétrique	34
2.2.2.1	Écriture du modèle	34
2.2.2.2	Estimation	35
2.2.3	Approche paramétrique classique	36
2.2.3.1	Écriture du modèle	36
2.2.3.2	Estimation	37
2.2.4	Approche paramétrique bayésienne	38
2.2.4.1	Écriture du modèle	38
2.2.4.2	Estimation	38
2.2.5	Discussion	40
3	Estimation dans le modèle paramétrique gaussien	43
3.1	Généralités sur le modèle paramétrique gaussien	43
3.1.1	Écriture du modèle	43
3.1.2	Hypothèses du modèle	45
3.1.3	Calcul de la vraisemblance	46
3.2	Procédures d'estimation	46
3.2.1	Classification des méthodes	46
3.2.2	Méthodes basées sur le traitement séparé des individus	47
3.2.2.1	Estimation des paramètres individuels	47
3.2.2.2	Estimation des paramètres globaux	48
3.2.3	Vraisemblance directe	49
3.2.4	Vraisemblance directe approchée	53

3.2.4.1	Approximation laplacienne	53
3.2.4.2	Monte Carlo par échantillonnage d'importance	55
3.2.4.3	Quadrature gaussienne	57
3.2.5	Vraisemblance indirecte	59
3.2.5.1	Pseudo-vraisemblance simulée	59
3.2.5.2	Linéarisation	60
3.2.6	Implémentation des méthodes	63
3.3	Propriétés des estimateurs	63
3.3.1	Au sujet de l'asymptotique	63
3.3.2	Revue bibliographique	64
3.3.2.1	Vraisemblance directe	64
3.3.2.2	Vraisemblance directe approchée	64
3.3.2.3	Vraisemblance indirecte	66
3.3.3	Contribution à l'étude des propriétés des estimateurs	69
3.3.3.1	Convergence de l'estimateur du maximum de vraisemblance	69
3.3.3.2	Adaptation pour les méthodes d'estimation par approxi- mation	70
3.3.4	Synthèse sur les propriétés des estimateurs	71
3.4	Conclusion	72
4	Étude par simulation	73
4.1	Simulation de valeurs de LAI	73
4.2	Comparaison linéarisation / approximation numérique	75
4.2.1	Plan d'expérience	76
4.2.1.1	Motivations	76
4.2.1.2	Mise en œuvre	77
4.2.2	Résultats des méthodes d'approximation	78
4.2.2.1	Aspect opérationnel	78
4.2.2.2	Estimation des paramètres	79
4.2.3	Conclusion des tests	82
4.3	Comparaison d'estimations par pixel et par linéarisation	87
4.3.1	Plan d'expérience	87
4.3.1.1	Motivations	87
4.3.1.2	Mise en œuvre	87
4.3.2	Résultats	89

4.3.2.1	Convergence	89
4.3.2.2	Paramètres de moyenne	90
4.3.2.3	Paramètres de variance	91
4.3.2.4	Paramètres individuels	91
4.3.3	Conclusion	92
4.4	Effet d'une structure spatiale sur l'estimation par linéarisation	94
4.4.1	Simulation d'une structure spatiale	94
4.4.2	Structure spatiale sur le bruit	96
4.4.2.1	Plan d'expérience	97
4.4.2.2	Résultats	99
4.4.3	Structure spatiale sur les paramètres aléatoires	100
4.4.3.1	Plan d'expérience	100
4.4.3.2	Résultats	101
4.4.4	Conclusion	104
5	Application à des données réelles	107
5.1	Les données	107
5.2	Traitement des données	109
5.3	Résultats	110
5.3.1	Paramètres globaux	110
5.3.1.1	Paramètres de moyenne β	111
5.3.1.2	Écart-type des paramètres σ_β	113
5.3.1.3	Écart-type résiduel σ_ε	114
5.3.2	Paramètres individuels	114
5.4	Modification du modèle	115
5.4.1	Résultats	115
5.4.1.1	Paramètres globaux	115
5.4.1.2	Paramètres individuels	116
5.5	Conclusion	117
	Conclusion et perspectives	121
1	Conclusion	121
2	Perspectives	122
	Bibliographie	125

Annexe A : Glossaire de télédétection spatiale	133
Annexe B : Mise en œuvre des programmes	135
1 Simulation de valeurs de LAI	135
2 La librairie nlme	135
2.1 Arguments de la fonction nlme	135
2.2 Sorties de la fonction nlme	137
3 Comparaison des estimations par linéarisation et approximation numérique	137
4 Comparaison des estimations par pixel et par linéarisation	139
5 Simulation de structures spatiales	140

Introduction

L'intérêt majeur du suivi de cultures réside principalement dans la connaissance, voire la prédiction des récoltes. Cette information revêt une importance primordiale pour l'ensemble des acteurs de la chaîne agro-alimentaire, des agriculteurs aux décideurs politiques.

La complexité du problème de prédiction de production implique que l'on essaie de tirer profit de toutes les sources d'information disponibles. Au côté des données de terrain et climatiques, dont l'utilisation est classique, les données issues de la télédétection spatiale prennent une place de plus en plus importante. L'exhaustivité spatiale et la répétitivité des mesures en sont les principaux avantages. Les données radiométriques permettent d'identifier les couverts végétaux, de caractériser leur état et de suivre leur évolution au cours du cycle végétatif.

Pour pouvoir tirer profit des données satellites, celles-ci doivent être en accord avec le phénomène étudié. Lorsqu'il s'agit du suivi de cultures, la croissance parfois rapide de la plante nécessite des informations fréquentes (Bouman, [9]). De même, l'identification des cultures nécessite une résolution spatiale relativement haute pour s'adapter à la taille des parcelles agricoles. Or, une contrainte liée aux caractéristiques techniques des satellites empêche de disposer d'images à haute résolution spatiale à une fréquence temporelle d'acquisition élevée. Afin de pouvoir fournir des informations pertinentes, il convient de combiner les deux types de données : haute résolution spatiale/faible résolution temporelle et haute résolution temporelle/faible résolution spatiale.

Le traitement des données à haute résolution temporelle par des techniques de désagrégation (Faivre et Fischer [34], Cherchali *et al.* [13]) offrent la possibilité de séparer les caractéristiques propres à chaque thème observé en mélange ; on peut ainsi disposer d'une mesure ciblée sur la culture d'intérêt avec un pas de temps relativement court. Ces techniques sont généralement basées sur une connaissance *a priori* du plan d'occupation des sols, susceptible d'être établi à partir des données à haute résolution spatiale.

Toutefois, les données de télédétection, aussi pertinentes soient-elles en termes de résolutions spatiale et temporelle, ne suffisent pas à mener à bien une prédiction de production. Ces informations doivent être utilisées soit dans un modèle statistique, soit en complément d'un modèle de simulation de culture (Delécolle *et al.* [25]). Plus précisément, les données issues de la télédétection peuvent servir de paramètres d'entrée du modèle de simulation au même titre que les données terrain, météorologiques, hygrométriques *etc...* Une autre procédure, potentiellement complémentaire à la précédente, consiste à effectuer un forçage du modèle de simulation par l'évolution temporelle d'une variable d'état de la culture, comme l'indice de surface foliaire. De cette façon, on impose des contraintes à la simulation de croissance sur la base de données recueillies par télédétection. Dans ce

cas, les informations fournies au modèle de simulation correspondent à l'ajustement d'un modèle d'évolution de l'indice de surface foliaire sur des données relatives à chaque pixel et à chaque date d'acquisition.

L'étude présentée ici se focalise sur un modèle d'évolution temporelle de l'indice de surface foliaire avec comme objectif de mieux représenter le développement de la culture tout au long du cycle végétatif. Cette étape qui survient après la mise en œuvre éventuelle de techniques de désagrégation dans le traitement des données de télédétection est essentielle pour la prise en compte d'une dynamique temporelle ; elle permet de lier les informations acquises à des pas de temps discrets.

Actuellement, ce type d'étude est mené en considérant les pixels des images indépendamment les uns des autres [32, 33]. C'est partant de ce constat que nous avons choisi d'orienter les travaux vers l'étude d'une méthodologie permettant de traiter l'ensemble des pixels simultanément.

Les modèles non linéaires à paramètres aléatoires permettent de répondre favorablement à cette attente. Ces techniques, apparues dans les années 80 essentiellement dans le domaine de la pharmacologie, sont destinées à permettre la modélisation d'un ensemble de courbes non linéaires en prenant en compte les variabilités intra-individu et inter-individus (Davidian et Giltinan [22]). Le cas typique d'application de ces modèles est l'étude clinique sur plusieurs individus de la réponse à un produit injecté dans l'organisme. L'objectif de ce type d'étude est de modéliser par exemple la diminution de la concentration du produit au cours du temps. Les mesures sont logiquement affectées par les variabilités intra- et inter-individus précédemment évoquées.

La complexité du modèle ne permet pas en général une résolution simple de l'estimation par maximum de vraisemblance. Aussi, diverses stratégies ont été développées pour remédier à ces difficultés d'estimation. Elles prennent généralement en compte l'estimation des paramètres de population et les prédictions individuelles. Les différences entre les méthodes apparaissent aussi bien dans la spécification des modèles (approches paramétriques ou non paramétriques) que dans les procédures d'estimation (linéarisation, approximation de la vraisemblance *etc.*). Ces méthodes sont généralement présentées et éventuellement comparées d'un point de vue quasi-exclusivement numérique sur des jeux de données réels ou simulés de petite taille par rapport au problème de suivi de cultures que nous souhaitons aborder. La taille des jeux de données traités est de l'ordre d'une dizaine de mesures pour quelques dizaines d'individus alors que les images que nous avons à traiter comportent un nombre de pixels supérieur au millier et sont acquises en quantité variant selon la durée d'étude et pouvant se compter par dizaines sur une période de plusieurs mois.

Cette étude bibliographique nous conduit à orienter les travaux de thèse vers l'étude des modèles non linéaires à paramètres aléatoires. L'aspect théorique est abordé *via* les propriétés des estimateurs issus des méthodes paramétriques. L'accent est cependant mis sur la mise en œuvre des modèles non linéaires à paramètres aléatoires sur des jeux de données simulés de la taille des problèmes liés au suivi de cultures par télédétection.

Le premier chapitre de ce mémoire est destiné à situer le cadre de l'étude dans le contexte général du suivi de cultures par télédétection spatiale. Les enjeux du suivi de cultures pour différentes tailles de la surface d'étude sont évoqués et face à ces enjeux, nous évoquons les apports de la télédétection spatiale pour améliorer les procédures mises

en œuvre pour l'estimation des productions, principal objectif du suivi de cultures. Néanmoins, il est reconnu que les données satellites seules ont une utilité restreinte. Leur intérêt réside dans une utilisation conjointe avec des modèles de développement. Pour cela, il convient de fournir à ces modèles des informations pertinentes sur l'évolution de la plante. Les transformations des données brutes vers l'utilisation d'informations utiles par l'agronome sont détaillées dans la suite. La fin du chapitre est consacrée au traitement statistique des données de télédétection. Le constat de certaines insuffisances dans ce traitement nous conduit à formuler la question de recherche basée sur l'utilisation des modèles non linéaires à paramètres aléatoires.

Le deuxième chapitre présente quelques généralités sur le modèle non linéaire à paramètres aléatoires. La prise en compte des variabilités intra-individu et inter-individus est détaillée. Nous proposons une revue des différentes spécifications suggérées dans le cadre de ces modèles. Selon les hypothèses que l'on suppose sur la variabilité inter-individus, on choisira parmi les approches non paramétrique, semi-(non)paramétrique, paramétrique classique et paramétrique bayésienne.

Le troisième chapitre est consacré à l'estimation dans les modèles non linéaires à paramètres aléatoires dans le cadre de l'approche paramétrique classique. Devant les difficultés de calcul de la vraisemblance, diverses stratégies basées sur des approximations ont été développées. Nous proposons une classification des méthodes d'approximation en distinguant d'abord celles basées sur le traitement séparé des individus et celles considérant comme critère le maximum de vraisemblance. Une synthèse des propriétés théoriques des estimateurs issus de ces diverses méthodes clôt ce chapitre.

Le quatrième chapitre repose sur une étude par simulation du modèle non linéaire à paramètres aléatoires. Une première série de simulation compare diverses méthodes d'estimation par approximation dans les modèles non linéaires à paramètres aléatoires sur des jeux de données de taille comparable au problème du suivi de cultures. L'objectif de la seconde série de tests est de comparer l'approche actuelle par pixel à l'estimation par linéarisation dans un modèle non linéaire à paramètres aléatoires. Enfin, une troisième série teste la méthode d'estimation par linéarisation sur des données simulées selon une structure spatiale.

Le cinquième et dernier chapitre présente l'étude d'un jeu de données réelles axée également sur la comparaison des estimations pixel par pixel et par paramètres aléatoires. Nous disposons sur une zone de $36 \times 36 \text{ km}^2$ d'images à haute résolution spatiale grâce auxquelles nous pouvons considérer diverses configurations pour l'analyse des données. La présentation des résultats sous forme cartographique est une manière d'illustrer une des utilisations potentielles des estimations des paramètres d'un modèle d'évolution temporelle de l'indice de surface foliaire sans passer par un modèle de développement de cultures.

Chapitre 1

La télédétection spatiale au service de l'agriculture

Ce chapitre situe les travaux de thèse dans le contexte général de la télédétection spatiale au service de l'agriculture. Dans une première partie, après avoir défini des termes techniques liés à la télédétection, nous évoquons conjointement des problématiques d'ordre agronomique et les apports potentiels de la télédétection spatiale à celles-ci. Par la suite, nous nous plaçons dans le cadre de l'utilisation des données recueillies par les capteurs embarqués sur le satellite SPOT 4; leur exploitation constitue le cadre applicatif des travaux de thèse. Nous présentons alors la chaîne de traitement des données qui, à partir de mesures de réflectances, fournit des données interprétables en termes d'évolution de culture. Le cadre de travail étant précisé, un état de l'art sur le traitement statistique des données de télédétection est proposé. C'est l'analyse de cet existant qui conduit à la définition de la question de recherche. Un glossaire des termes techniques liés à la télédétection spatiale est fourni en annexe.

1.1 Généralités

La définition de termes techniques liés à la télédétection spatiale nous permet de mieux discuter ensuite les apports potentiels des données satellitaires pour les besoins de l'agriculture. Selon la taille de la zone d'étude, les informations utiles sont de natures différentes bien que généralement directement reliées à la rentabilité économique.

1.1.1 Définitions liées à la télédétection spatiale

En préambule à cette section, nous citons la définition de la télédétection donnée dans le glossaire des termes officiels de la télédétection aérospatiale [99]. La télédétection est « l'ensemble des connaissances et techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci ». Une définition un peu plus détaillée est fournie par Bonn et Rochon [7]: « La télédétection est la discipline scientifique qui regroupe l'ensemble des connaissances et des techniques utilisées pour l'observation, l'analyse, l'interprétation et

la gestion de l'environnement à partir de mesures et d'images obtenues à l'aide de plateformes aéroportées, spatiales, terrestres ou maritimes. Comme son nom l'indique, elle suppose l'acquisition d'information à distance, sans contact direct avec l'objet détecté ». La télédétection est dite spatiale quand les mesures sont effectuées à partir d'un satellite en orbite autour de la Terre.

1.1.1.1 Caractéristiques techniques des capteurs de télédétection

La nature des données recueillies par télédétection spatiale dépend des caractéristiques techniques des capteurs embarqués sur le satellite. Nous nous restreignons à quelques définitions concernant la précision des images recueillies par les capteurs.

Résolution spatiale

La résolution spatiale d'un capteur est définie à partir de la surface minimale observable au sol. Plus la résolution spatiale est grande, plus la surface minimale observable est petite. Dans ce cas, le capteur « voit » des détails plus petits ; l'image est plus précise. On parle de haute résolution spatiale lorsque la taille du pixel, unité élémentaire de l'image, représente au sol une zone de côté de l'ordre du mètre à la dizaine de mètres. Une faible résolution spatiale correspond grossièrement à des pixels kilométriques ou plus vaste encore.

On aurait tendance à préférer systématiquement des images à haute résolution spatiale aptes à fournir des détails, mais selon l'objectif poursuivi des images de résolution kilométriques sont tout à fait satisfaisantes. Le choix du type d'images dépend de l'objectif poursuivi. Par exemple, si l'on souhaite observer le développement urbain dans une agglomération, on aura intérêt à obtenir des images précises pour pouvoir distinguer les nouveaux bâtiments et les nouvelles routes. En revanche, si l'on s'intéresse à des parcelles agricoles aux États-Unis, des images de résolution kilométrique seront suffisantes.

Résolution temporelle

En termes de traitement du signal, la résolution temporelle est le plus petit intervalle de temps séparant deux réalisations successives du signal. Pour un capteur satellitaire, cela équivaut à la fréquence d'acquisition des images d'une même zone et donc à la période de révolution du satellite. Nous retiendrons cette définition même si certains satellites sont capables de viser latéralement et d'augmenter ainsi la capacité de re-visite.

Comme pour la résolution spatiale, la résolution temporelle doit être adaptée à l'événement dynamique à surveiller. Pour étudier des phénomènes relativement lents comme la fonte de glaciers ou la croissance d'une forêt, des images à quelques années d'intervalles suffisent. En revanche, lors de la phase de croissance d'une culture annuelle, phénomène beaucoup plus rapide, il est important de disposer d'informations fréquentes de l'ordre de quelques jours.

Résolution spectrale

La résolution spectrale correspond à la largeur des bandes spectrales d'un capteur. Plus la résolution spectrale est fine, plus la largeur des bandes spectrales des différents canaux du capteur sont étroites. Celle-ci peut varier très fortement selon les capteurs : de l'ordre

du nanomètre, voire moins, pour des capteurs hyperspectraux, à plusieurs centaines de nanomètres pour des capteurs panchromatiques.

Comme pour les deux autres caractéristiques, le choix de la résolution spectrale dépend de l'objectif visé. Un capteur à large bande spectrale suffit pour différencier des classes distinctes comme l'eau et la végétation. En revanche, des classes plus spécifiques comme par exemple différents types de cultures ne sont pas aussi faciles à différencier et nécessitent l'utilisation de capteurs à bande spectrale plus fine.

1.1.1.2 Contraintes liées à l'utilisation d'un satellite

Dans un cas idéal, nous disposerions quotidiennement d'images à haute résolution spatiale acquises dans une bande spectrale relativement étroite. Dans la pratique, ce type de données n'est pas accessible car les capteurs sont soumis à des contraintes.

Compromis spatial-temporel

La relation entre la fréquence d'acquisition et la résolution spatiale s'exprime à travers la fauchée du satellite.

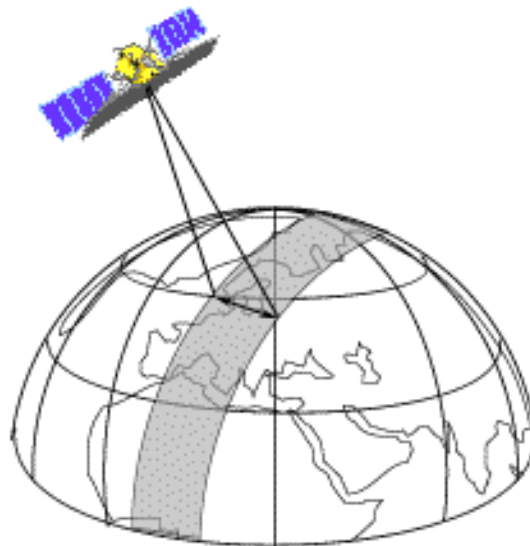


FIG. 1.1 – Représentation de la fauchée d'un satellite d'observation de la Terre [11].

La fauchée d'un satellite, représentée sur la figure 1.1, peut être définie comme la largeur du couloir observé par le capteur. Si la fauchée de l'instrument est large, une très grande surface au sol est « vue ». Dans ces conditions, une même zone est observée même si le satellite ne se situe pas exactement à sa verticale. Ceci a pour conséquence d'augmenter la fréquence d'acquisition. Or, bénéficier d'une large fauchée est équivalent pour un observateur humain à l'oeil nu, à disposer d'une vue d'ensemble sur la cible, ce qui se fait au détriment de la distinction de détails. Dans le cas du capteur satellitaire, cela implique une faible résolution spatiale.

A l'inverse, si la fauchée est étroite cela équivaut à un observateur humain regardant à travers des jumelles. Il voit les détails mais sur une surface moins étendue. Pour le

satellite, la surface vue étant étroite, il faut attendre un nouveau passage exactement à la verticale de la zone d'étude pour disposer d'une nouvelle image. La résolution temporelle est donc plus faible que pour un capteur à large fauchée.

Ce compromis entre les résolutions spatiale et spectrale d'un capteur satellitaire est illustré sur la figure 1.2.

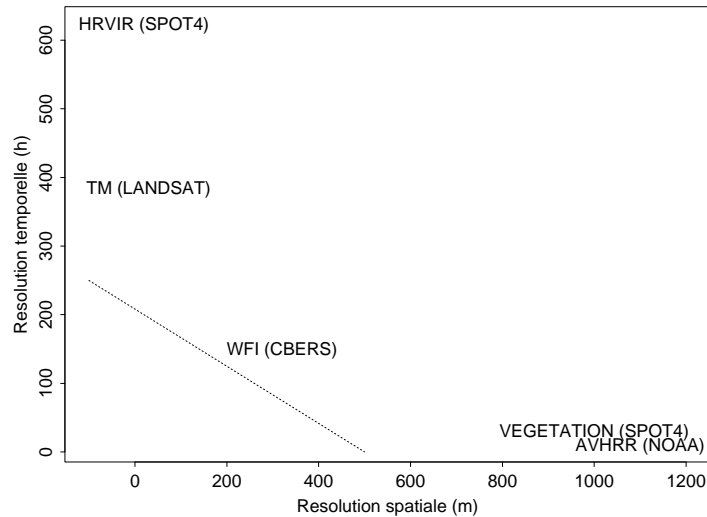


FIG. 1.2 – Illustration du compromis entre résolution spatiale et résolution temporelle à partir des données relatives aux satellites SPOT 4 (capteurs HRVIR et VEGETATION), NOAA (capteur AVHRR), CBERS (capteur WFI) et LANDSAT (capteur TM).

En représentant les satellites comme un point ayant pour coordonnées la résolution spatiale et la résolution temporelle, on constate qu'aucun point n'apparaît au-dessous de la ligne pointillée illustrant la contrainte. Par ailleurs, il est évident que le coin haut-droit du graphique, correspondant à des faibles résolutions spatiale et temporelle, n'est pas intéressant en pratique.

Compromis spatial-spectral

Le même type de contrainte existe entre la résolution spatiale et la résolution spectrale.

Une résolution spatiale élevée correspond au sol à des surfaces minimales observables de petite taille. Au niveau du capteur, cela se traduit par une diminution de la quantité d'énergie reçue pour une unité élémentaire au sol. Pour augmenter la quantité d'énergie atteignant le capteur sans réduire la résolution spatiale, il faut élargir l'intervalle de longueurs d'onde détecté par un canal du capteur. Cette opération réduit donc la résolution spectrale du capteur. Inversement, une résolution spatiale plus grossière permet une résolution spectrale plus fine.

L'illustration du compromis entre résolution spatiale et résolution spectrale pour un capteur satellitaire (Fig. 1.3) amène le même commentaire que la figure 1.2. Aucun point n'apparaît dans le coin proche de l'origine du graphique, ce qui montre la difficulté de

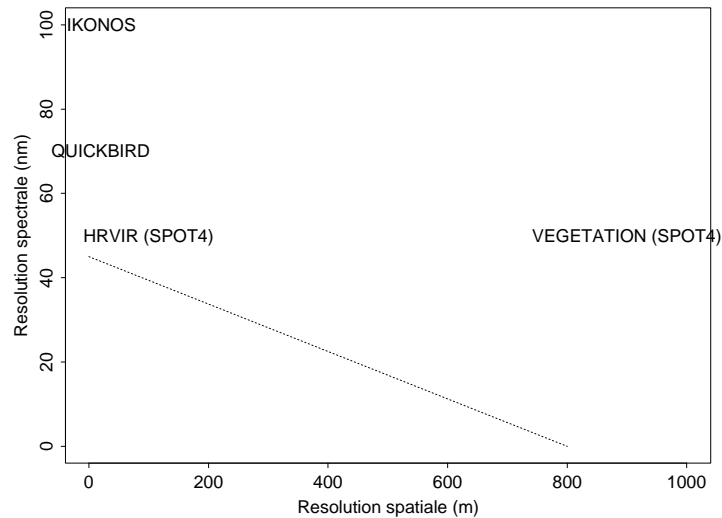


FIG. 1.3 – Illustration du compromis entre résolution spatiale et résolution spectrale à partir des données relatives aux satellites IKONOS, SPOT 4 (capteurs HRVIR et VEGETATION) et QUICKBIRD (multispectral).

disposer à la fois de hautes résolutions spatiale et spectrale.

Volume d'informations

La quantité d'informations recueillies apparaît également comme une contrainte à l'utilisation de données satellites. La manipulation des données implique non seulement le stockage sur la plate-forme satellitaire mais aussi la transmission au sol de ces données.

À titre d'exemple, considérons un capteur dont la fauchée est de 1000 km et la résolution spatiale de 100 m. Avec ce capteur, on souhaite acquérir une image centrée sur la France assimilable à un carré de 1000 km sur 1000 km. Une telle image est donc constituée de 10000×10000 pixels dans chaque canal d'acquisition. Si l'on considère un codage des pixels sur 1 octet (8 bit) et compte tenu du fait qu'une image en couleur est le résultat de la superposition de trois canaux, la taille d'une telle image est donc de $3 \times 10000 \times 10000$ octets, soit 300 Mo. Cela correspond en pratique à stocker seulement deux images sur un CD-ROM.

Conclusion sur les contraintes

Compte-tenu des diverses contraintes liées à la construction des capteurs et à la manipulation des données, un capteur optimal selon les trois résolutions précédemment évoquées n'existe pas.

Nous excluons de notre étude le recours à des capteurs aéroportés pour lesquels les contraintes exprimées précédemment n'ont pas forcément lieu d'être. Leur inconvénient majeur est de ne pas offrir la même exhaustivité spatiale que les capteurs satellitaires sur de grandes surfaces.

La section suivante présente quelques objectifs liés à l'agriculture. Dans chacun des cas de figures proposés, les caractéristiques techniques des capteurs permettant d'obtenir

des informations pertinentes s'imposent naturellement.

1.1.2 Problématiques liées à l'agriculture

Les enjeux du suivi de cultures sont différents selon la taille de la zone d'étude. Dans cette section, nous évoquons ces enjeux dans différents cas de figures sur une échelle allant de la parcelle agricole à la couverture mondiale. Nous présentons dans chaque cas quels peuvent être les apports de la télédétection spatiale.

1.1.2.1 Échelle parcellaire

Au niveau d'une parcelle agricole, les problèmes rencontrés sont liés à la mise en œuvre d'une agriculture de précision. On pourra consulter à ce sujet les dossiers de l'Institut Technique des Céréales et des Fourrages (ITCF, [101]). Le principe de l'agriculture de précision réside essentiellement dans la prise en compte des hétérogénéités intra-parcellaires pour ajuster au mieux les pratiques culturales aux besoins de la plante tout en respectant l'environnement. En repérant précisément les endroits qui nécessitent un apport en engrais, on le limite à ces zones. Ceci implique une diminution du coût engendré par l'achat d'engrais et le respect de l'environnement.

Les apports de la télédétection dans ce cas de figure sont limités d'une part, par la résolution spatiale des capteurs et d'autre part, par la durée entre l'acquisition des données et leur mise à disposition au niveau de l'exploitation agricole. Pour l'agriculture de précision, la télédétection aéroportée semble actuellement préférable, en complément de données terrain.

1.1.2.2 Bassin de collecte

À un niveau supérieur à la parcelle ou au regroupement local de parcelles, se situent les bassins de collecte. Sous ce terme, on peut englober des entités administratives comme des départements ou des régions, mais également d'autres zones plus ou moins vastes non délimitées administrativement (bassin parisien par exemple).

À cette échelle, l'enjeu principal réside dans l'estimation des productions agricoles. Pour cela, en considérant une production comme le produit d'une superficie par un rendement, on distingue deux types de tâches à accomplir :

- calculer la superficie des surfaces cultivées,
- estimer le rendement des cultures par un suivi temporel.

Superficie de surfaces cultivées

Généralement, les mesures satellitaires à haute résolution spatiale permettent d'obtenir des résultats satisfaisants pour l'estimation des superficies, notamment par des techniques de classification qui seront présentées ultérieurement (*Cf.* section 1.3.2).

Estimation du rendement

L'estimation du rendement est beaucoup plus complexe. En effet, ce dernier est grandement influencé par les conditions météorologiques sujettes à des variations inter-annuelles parfois importantes. De plus, d'éventuels accidents climatiques (sécheresse, inondation)

et/ou risques naturels (maladies, attaque parasitaire, invasion d'insectes...) peuvent également influencer sur le rendement. C'est donc par un suivi régulier des cultures que l'on a accès à une estimation fiable du rendement.

L'apport de la télédétection spatiale réside dans l'utilisation de capteurs optiques à haute résolution temporelle, qui peuvent fournir, sous réserve de conditions d'ennuagement favorables, des informations quotidiennes sur la santé des cultures et ainsi aider à mesurer l'impact des phénomènes météorologiques et autres accidents sur la santé des plantes.

Les données satellitaires en elles-mêmes ne suffisent pas à produire une estimation de rendement. Elles sont généralement utilisées en complément d'un modèle de fonctionnement de cultures (Delécolle *et al.* [25]).

Cette stratégie est une alternative aux méthodes de prévision de rendement classiques basées sur des méthodes statistiques de régression. Dans la pratique, elles consistent en une liaison statistique entre des observations réalisées par le passé durant la période de développement des cultures avant la récolte, et le rendement final observé. Cette relation est alors appliquée à l'année en cours. Les observations sont soit liées à l'état de la culture (nombre de plants, d'épis...) soit de nature météorologique (température, hygrométrie...). En règle générale, ces méthodes donnent des résultats décevants [41] essentiellement par manque de robustesse vis-à-vis de conditions différentes par rapport aux années de référence.

Les travaux menés dans le cadre de cette thèse se situent essentiellement à l'échelle de bassins de collecte. Nous reviendrons plus en détail sur le suivi régulier des cultures et sur l'introduction de données de télédétection dans les modèles de fonctionnement de cultures dans la section 1.2.

1.1.2.3 Échelles nationale, continentale et mondiale

À une échelle supérieure à celle des bassins de collecte, on retrouve les nations, les continents, et au plus haut niveau, la totalité du globe terrestre.

À ces niveaux, une prévision précise et précoce de l'état des productions agricoles est un atout majeur dans la gestion du commerce international des matières premières agro-alimentaires. En termes de marché, cela revient à prévoir l'offre, qui peut varier fortement selon les régions et les années, et déterminer ainsi les prix. Dans cette optique, un pays, ou une union de pays, n'est pas seulement intéressé par sa propre production mais également par celle des autres pays à des fins de politique commerciale internationale.

La télédétection spatiale offre une méthode sûre et efficace de recueil d'informations objectives, exhaustives et homogènes :

- L'objectivité des informations est assurée par des mesures physiques liées à la présence et au développement des cultures.
- L'exhaustivité des informations est apportée par une couverture totale du globe par les satellites. Il est impossible de prétendre à une telle exhaustivité par des enquêtes sur le terrain.
- L'homogénéité résulte du fait que ce sont les mêmes moyens qui sont mis en œuvre pour l'ensemble du globe. Ceci permet de lever l'incertitude vis-à-vis des statistiques officielles calculées individuellement par chaque pays.

À titre d'exemple, disposer d'une carte comme celle représentée sur la figure 1.4 est un atout important pour la stratégie commerciale et peut également permettre de prévenir les crises alimentaires afin de mieux adapter des actions à caractère plus humanitaire.

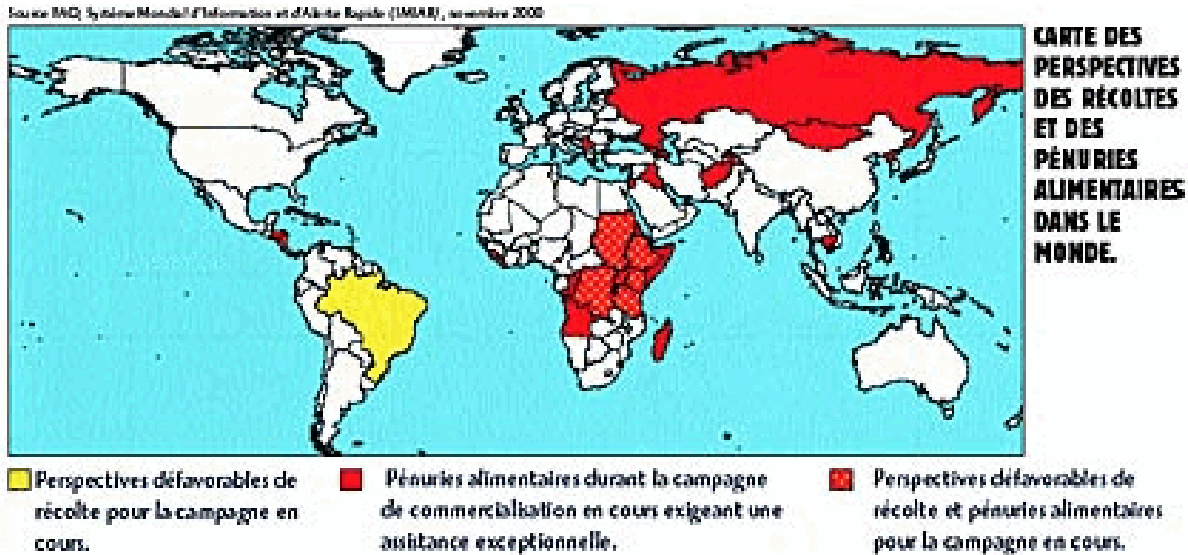


FIG. 1.4 – Carte des perspectives des récoltes et des pénuries alimentaires dans le monde, novembre 2000, FAO/SMIAR (Organisation des Nations Unies pour l'Agriculture et l'Alimentation / Système Mondial d'Information et d'Alerte Rapide sur l'alimentation et l'agriculture), [82].

Un autre exemple, au niveau européen, est le contrôle par télédétection spatiale des déclarations de surfaces cultivées effectuées par les exploitants agricoles subventionnés dans le cadre de la Politique Agricole Commune.

Dans ce même cadre, les informations satellitaires permettent de contrôler les surfaces cultivées par des pays en situation de surplus alimentaire. De plus, les amendes en cas de dépassement des quotas de production sont définies à partir d'estimation des productions, d'où l'intérêt d'en améliorer la précision.

1.2 Contexte de l'étude

Dans cette section, nous sortons du cadre général du suivi de cultures par télédétection spatiale pour préciser dans le cas de cette étude les étapes entre l'acquisition de données par le système embarqué sur SPOT 4 et l'estimation des paramètres d'un modèle d'évolution temporelle de l'indice de surface foliaire éventuellement introduits ensuite dans un modèle de fonctionnement de cultures.

Comme le décrit Baret [2], le lien entre la productivité d'une culture et sa réflectance dans certaines longueurs d'onde relève de deux relations de disciplines différentes :

- l'une d'ordre physique entre la structure du couvert végétal et sa réponse radiométrique,

- l'autre d'ordre biologique entre l'état de la culture et l'évolution de sa structure traduite par exemple par l'indice de surface foliaire.

Les étapes dans le traitement de données satellitaires reflètent la prise en compte de ces deux relations.

1.2.1 SPOT 4

1.2.1.1 Caractéristiques des capteurs HRVIR et VEGETATION

Le satellite SPOT 4 lancé en mars 1998 est une plate-forme qui dispose d'instruments dédiés à l'observation de la Terre. Il s'agit des capteurs HRVIR (Haute Résolution dans le Visible et l'InfraRouge) et VEGETATION. Ces deux capteurs diffèrent par leurs résolutions spatiale et temporelle, en revanche, ils effectuent des acquisitions dans des plages de longueurs d'onde communes, ce qui rend complémentaires les images issues des deux capteurs.

HRVIR est à haute résolution spatiale avec des pixels de 20 m sur 20 m et une fréquence d'acquisition de 26 jours. VEGETATION est un capteur à haute résolution temporelle, quotidienne, avec une résolution spatiale d'un kilomètre carré.

La figure 1.5 illustre de manière schématique la différence entre les deux capteurs caractérisée par la fauchée.

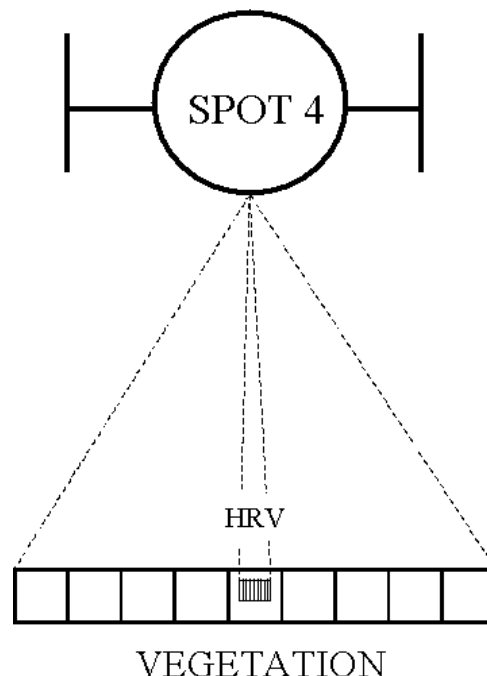


FIG. 1.5 – Représentation schématique de la différence de fauchée entre les capteurs HRVIR et VEGETATION.

1.2.1.2 Mesures de réflectances

Le principe des mesures de réflectances par un système optique passif est décrit sur la figure 1.6.

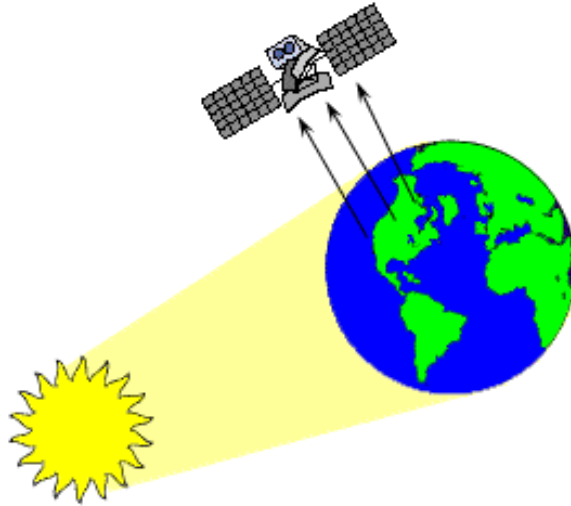


FIG. 1.6 – *Principe de fonctionnement d'un capteur passif [11].*

Contrairement aux systèmes actifs, principalement de type radar, les capteurs optiques passifs courtes longueurs d'onde ont besoin de la lumière du soleil pour effectuer les mesures. En pratique, le nombre d'images disponibles est donc directement dépendant des conditions atmosphériques, en particulier de l'enneigement.

Pour comprendre comment les mesures de réflectances diffèrent selon le type de sol rencontré, la figure 1.7 propose un zoom au niveau du sol de la figure 1.6.

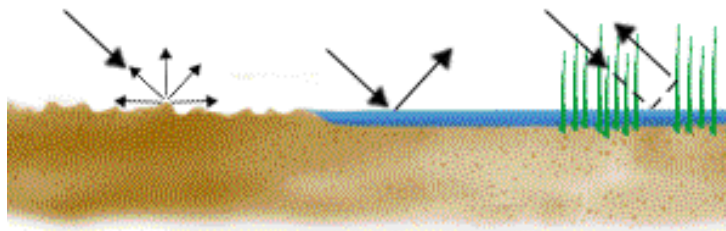


FIG. 1.7 – *Réflectances au niveau du sol sur trois surfaces différentes : sol nu, eau, végétation [11].*

L'étude de la végétation nécessite des mesures de réflectances dans les longueurs d'onde autour du rouge (R) et du proche infrarouge (PIR). La figure 1.8, extraite du guide de l'utilisateur de VEGETATION en ligne [81], présente les caractéristiques spectrales du capteur VEGETATION.

Les mesures de réflectances dans le proche infrarouge sont un témoin de l'activité photosynthétique de la plante. Les réflectances dans le rouge, en complément avec celles

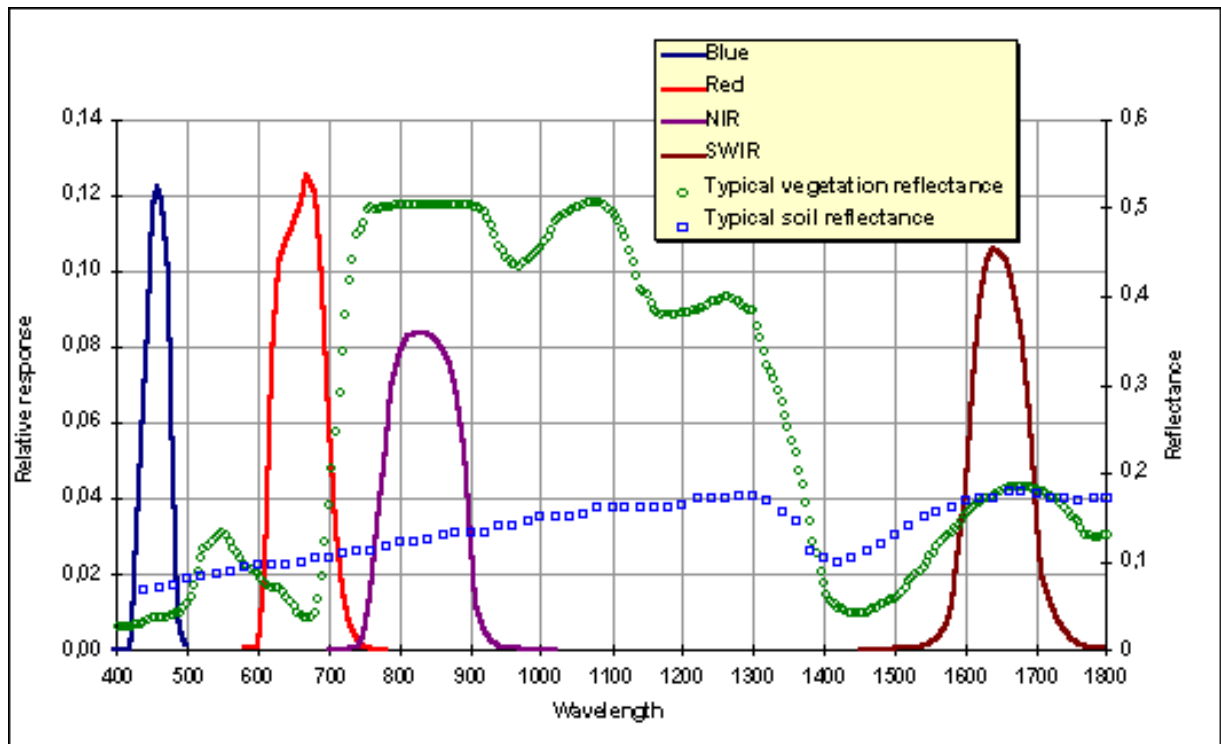


FIG. 1.8 – Caractéristiques spectrales du capteur VEGETATION. Extrait du *Vegetation User Guide en ligne* [81].

dans le proche infrarouge, permettent de mieux distinguer la végétation des autres types de sol.

1.2.2 Indices de végétation

1.2.2.1 Calcul et choix d'un indice de végétation

Même si elles permettent de distinguer le couvert végétal du sol nu, les réflectances dans le rouge et le proche infrarouge sont rarement utilisées directement. Elles sont en général combinées pour calculer des indices de végétation qui font le lien entre la réponse radiométrique d'une culture et sa structure végétale. De nombreux indices (Tab. 1.1) qui ont pour but de limiter l'influence de facteurs extérieurs comme par exemple la réflectance du sol, ont été proposés.

Parmi les indices de végétation présentés dans le tableau 1.1, un des plus couramment utilisés est l'indice de végétation de différence normalisée (Normalized Difference Vegetation Index, NDVI). Son utilisation est justifiée par Vignolles [88] par les raisons suivantes :

- le NDVI est faiblement sensible aux propriétés optiques du sol,
- il normalise l'effet des pentes,
- le profil temporel du NDVI permet de distinguer des types de couvert ayant une phénologie différente,

Dénomination	Formule	Référence
Rapport	$\frac{PIR}{R}$	[50, 89]
Indice de végétation normalisé	$NDVI = \frac{PIR-R}{PIR+R}$	[74, 86]
Indice de végétation transformé	$TVI = \sqrt{NDVI + 0.5}$	[24]
Indice de végétation perpendiculaire	$PVI = \sin(a) \times PIR \cos(a) \times R$	[71]

TAB. 1.1 – Quelques indices de végétation. Dans l'expression du PVI, a est une constante dépendant du type de sol.

- cet indice a été relié à divers caractères descriptifs de la physiologie du couvert dont l'indice de surface foliaire [3].

Le PVI fait également partie des indices les plus souvent utilisés. Il présente la particularité d'être calculé à partir d'une relation linéaire entre les réflectances dans le rouge et le proche infrarouge.

Dans la section suivante, nous nous intéressons à la dernière propriété évoquée du NDVI: la possibilité de le lier à l'indice de surface foliaire (LAI, Leaf Area Index) qui représente une mesure de la superficie du feuillage par unité de surface au sol. Autrement dit, c'est la surface foliaire totale d'une plante divisée par la surface de sol qu'elle occupe.

Une définition et un exemple de calcul sont donnés dans le tutoriel de télédétection du Centre Canadien de Télédétection [11]:

Le LAI est calculé en sélectionnant une surface couverte de végétation et en traçant un cube imaginaire de la surface du sol à la cime des arbres. Dans ce cube imaginaire, on mesure la surface d'un côté de toutes les feuilles vertes planes (ou la moitié de la surface totale des aiguilles de conifères). La superficie foliaire est divisée par la superficie au sol sélectionnée et le résultat représente l'indice LAI. Par exemple, si une surface au sol de 1 m^2 est sélectionnée et que la superficie foliaire est de 4 m^2 , l'indice LAI est de: $LAI = 4 \text{ m}^2 / 1 \text{ m}^2 = 4$.

1.2.2.2 Du NDVI au LAI

Dans la chaîne de traitement des données de télédétection, à la suite du calcul d'un indice de végétation, vient la conversion de cet indice en une caractéristique traduisant l'évolution de la structure végétale. Cette étape correspond à la relation entre l'état de la culture et l'évolution de sa structure. Nous utiliserons l'indice de surface foliaire. Baret et Guyot [3] ont mis en évidence une relation semi-empirique entre les indices de végétation et l'indice de surface foliaire:

$$IV = IV_{\infty} + (IV_s - IV_{\infty}) \times e^{(-K_{IV} \times LAI)} \quad (1.1)$$

où IV_s représente l'indice de végétation du sol nu, IV_{∞} la valeur asymptotique de l'indice de végétation quand le LAI tend vers l'infini (en pratique $LAI > 8$) et K_{IV} est un coefficient qui contrôle la pente de la relation.

L'inversion de la formule (1.1) permet alors de convertir un indice de végétation en LAI (Cf. [32] pour la conversion du NDVI et [33] pour le PVI).

1.2.3 Modèle d'évolution temporelle du LAI

Dans cette partie, nous décrivons le modèle d'évolution temporelle que nous utilisons par la suite. Nous en présentons également ses utilisations potentielles.

1.2.3.1 Modèle de Baret

Nous considérons le modèle paramétrique d'évolution temporelle du LAI proposé par Baret [2]. L'expression du modèle comme combinaison de deux termes en exponentielle permet de traduire les phases de croissance et de sénescence de la plante, comme le montre la figure 1.9.

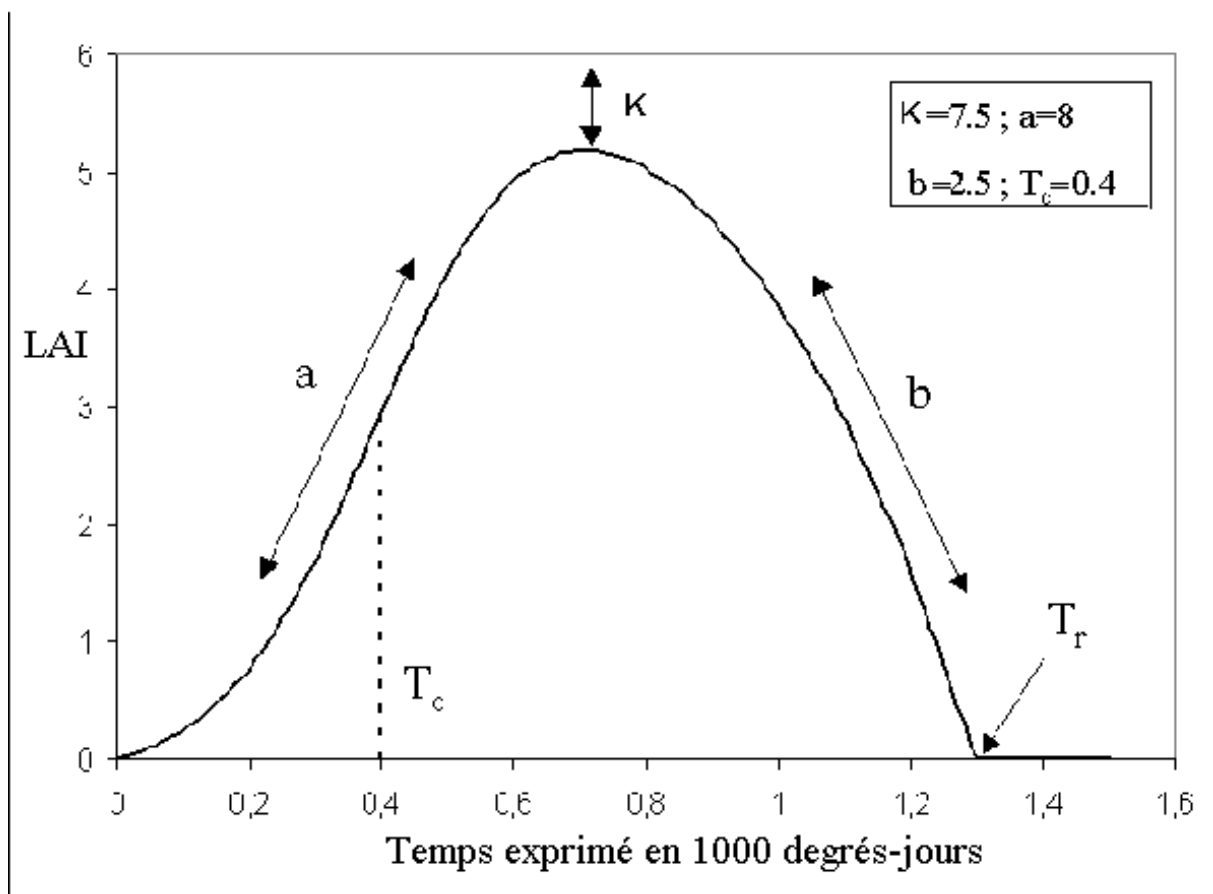


FIG. 1.9 – Allure du modèle d'évolution temporelle du LAI proposé par Baret [2] pour les valeurs de paramètres $K = 7.5$, $a = 8$, $b = 2.5$, $T_c = 0.4$ avec le temps exprimé en milliers de degrés-jours et la contrainte $T_r = \log(1 + \exp(a \times T_c))/b$.

Le modèle s'écrit

$$LAI(t) = f(t, \beta) = K \left(\frac{1}{1 + e^{-a(t-T_c)}} - e^{b(t-T_r)} \right) \quad (1.2)$$

où t représente le temps exprimé en somme de températures à partir d'une date fixe et $\beta = (K, a, b, T_c, T_r)$ est le vecteur des paramètres du modèle. Nous aurons recours à la contrainte $T_r = \log(1 + \exp(a \times T_c))/b$ pour faciliter les procédures d'estimation et notamment pour éviter d'estimer T_r inférieur à T_c , ce qui contredit l'interprétation des paramètres (Cf. ci-dessous) mais qui peut arriver lorsque l'on ne contraint pas le modèle.

La figure 1.10 est une représentation séparée des deux termes du modèle d'évolution.

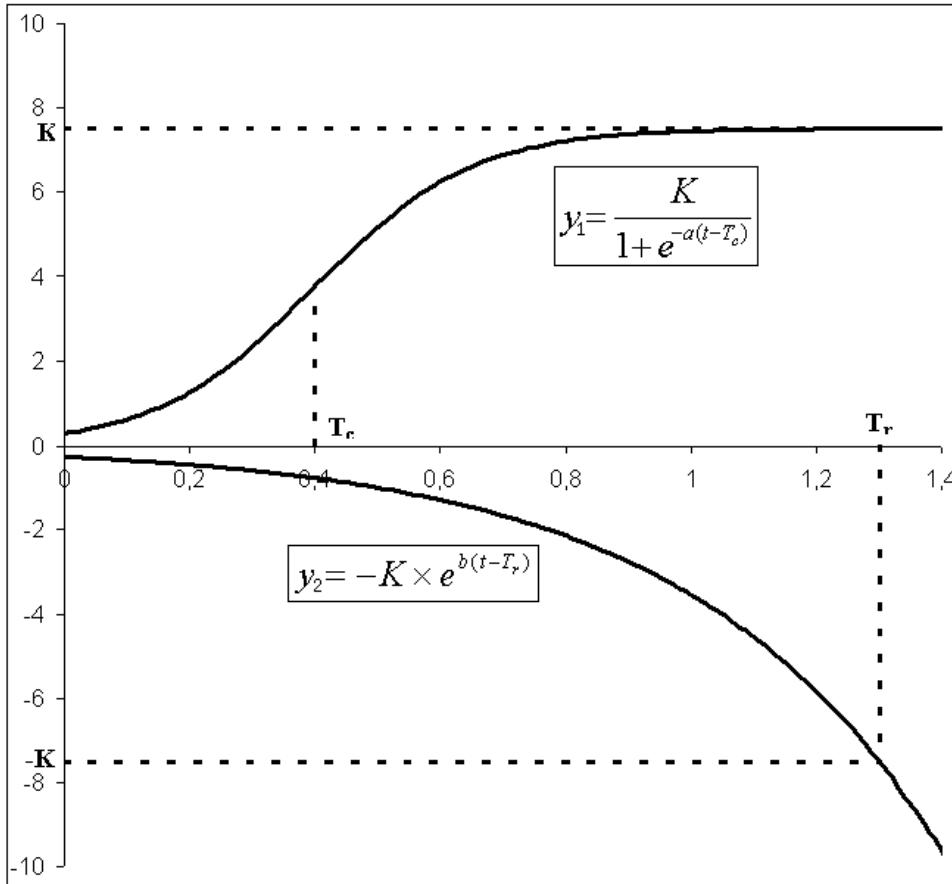


FIG. 1.10 – Distinction des deux sous-modèles du modèle de Baret liés respectivement à la croissance (y_1) et à la sénescence (y_2). Les valeurs des paramètres sont les mêmes que pour la figure 1.9 ($K = 7.5, a = 8, b = 2.5, T_c = 0.4$).

Les équations séparées du modèle global sont présentées dans le système

$$\begin{cases} y_1(t, \beta_1) = \frac{K}{1 + e^{-a(t-T_c)}} \\ y_2(t, \beta_2) = -K \times e^{b(t-T_r)} \end{cases} \quad (1.3)$$

où (y_1, β_1) et (y_2, β_2) représentent respectivement les données et les paramètres pour les phases de croissance et de sénescence.

La première équation du système (1.3) est un modèle logistique simple à trois paramètres (K, a, T_c). Il permet de modéliser la phase de croissance de la culture avec un point d'inflexion en $t = T_c$. La courbe représentative du modèle admet une asymptote horizontale en $y = K$.

La seconde équation de (1.3) est un modèle exponentiel négatif. Il vient s'ajouter au premier terme pour modéliser la phase de sénescence de la plante. Il contraint la courbe générale à s'annuler en $t = T_r$.

Avec les détails fournis par la décomposition du modèle, les paramètres du modèle d'évolution sont plus facilement interprétables en termes de croissance et de sénescence de la plante :

- Le paramètre K correspond au LAI maximal qui serait atteint théoriquement en l'absence de sénescence. En pratique, c'est un majorant du LAI maximal.
- Les paramètres T_c et T_r sont des indicateurs de moments-clés dans la période d'évolution de la plante: T_c est la durée entre la levée de la culture et l'accélération maximale de la croissance foliaire, cet indicateur est considéré comme un témoin de précocité; T_r marque la fin du cycle de vie de la plante ou la récolte.
- Les paramètres a et b sont des indicateurs des vitesses de croissance et de sénescence de la plante: plus le paramètre a est élevé plus la croissance est rapide, plus le paramètre b est élevé plus la sénescence est rapide.

1.2.3.2 Utilisation d'un modèle temporel du LAI

Nous évoquons ici deux utilisations potentielles des estimations des paramètres d'un modèle d'évolution temporelle de LAI: la contribution à la prévision de rendement *via* le forçage d'un modèle de simulation et l'interprétation des résultats sous forme cartographique.

Contribution à la prévision de rendement

Nous venons de voir que les données de télédétection peuvent conduire à l'expression de l'évolution de l'indice de surface foliaire d'une culture. Pour contribuer à la prévision de rendement, ce type d'information doit être utilisé en complément d'un modèle de fonctionnement (Delécolle *et al.* [25]), par exemple en forçant ce modèle à se caler sur la courbe d'évolution du LAI.

Les modèles de fonctionnement de cultures sont un moyen d'obtenir une prévision de rendement. L'objectif de ces modèles est de simuler les principaux processus intervenant dans le développement de la plante: croissance et bilans en eau, azote et carbone du système sol-culture. Une multitude de modèles existent, certains spécifiques à une culture, d'autres génériques. On pourra se référer à Brisson *et al.* [10] pour une revue de ces modèles.

Nous nous intéressons principalement au modèle STICS (Simulateur mulTIdisciplinaire pour les Cultures Standard, [10]) dont le but principal est de simuler l'effet du climat, du sol et des pratiques culturales sur la production en termes de quantité et de qualité. Avec un pas de temps journalier, STICS simule le comportement du système sol-culture sur une année. Un indice thermique, exprimé en degré-jour, contrôle le développement de la culture. Ce développement se traduit notamment par une évolution de

l'indice de surface foliaire. Les nombreuses variables d'entrée du modèle recouvrent les points suivants :

- le climat : températures (amplitude, maximum, minimum), précipitations, radiations solaires...
- les itinéraires techniques choisis par l'exploitant agricole : date de semis, irrigation, fertilisation...
- le sol : quantité de carbone, taux d'absorption d'eau, température du sol en fonction de la profondeur...
- la plante : nombre de grains, taux de croissance, profondeur des racines...

La prévision du rendement de la culture apparaît comme une des sorties du modèle de simulation.

Cartographie des paramètres

Les courbes ou les paramètres estimés peuvent également être interprétés directement en cartographiant les valeurs estimées (Fig. 1.11)

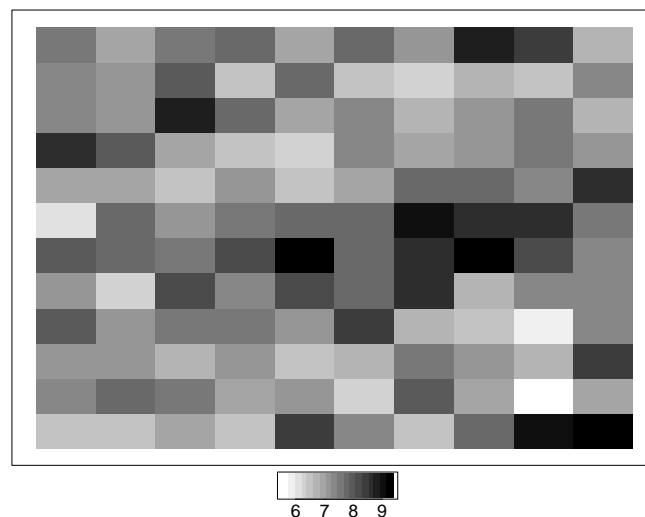


FIG. 1.11 – Cartographie des valeurs estimées du paramètre K du modèle de Baret sur une zone 12×10 km extraite des données traitées au chapitre 5.

Cette représentation permet de distinguer des zones où les valeurs estimées sont élevées (zone foncée à droite de l'image sous la diagonale bas-gauche/haut-droit). L'interprétation agronomique de ces fortes valeurs permet de dire par exemple que cette région a connu des conditions climatiques idéales pendant la phase de croissance permettant au blé d'atteindre une surface foliaire maximale.

1.3 Analyse de l'existant

Dans cette section, nous détaillons les deux types de données disponibles : haute résolution spatiale et haute résolution temporelle. Nous présentons ensuite les traitements

statistiques actuellement effectués sur chacun de ces types de données.

1.3.1 Les données

Pour assurer le suivi des cultures à partir de la plate-forme SPOT 4, on dispose des données issues des deux capteurs VEGETATION et HRVIR. La figure 1.12 illustre la différence entre des images issues des deux capteurs. Quelle que soit la grandeur représentée (réflectance, indice de végétation, LAI ...), les images HRVIR de résolution 20m sur 20m (à gauche sur la figure 1.12), permettent de distinguer nettement les différentes parcelles présentes sur la région étudiée. En revanche, c'est loin d'être le cas sur la même image dégradée à la résolution de VEGETATION (image de droite de la figure 1.12). Ce problème est rencontré lors d'acquisitions en Europe Occidentale où la taille des parcelles est inférieure à la résolution spatiale du capteur de fréquence quotidienne. C'est pourtant ce capteur qui dispose de la fréquence d'acquisition adéquate pour le suivi de cultures.

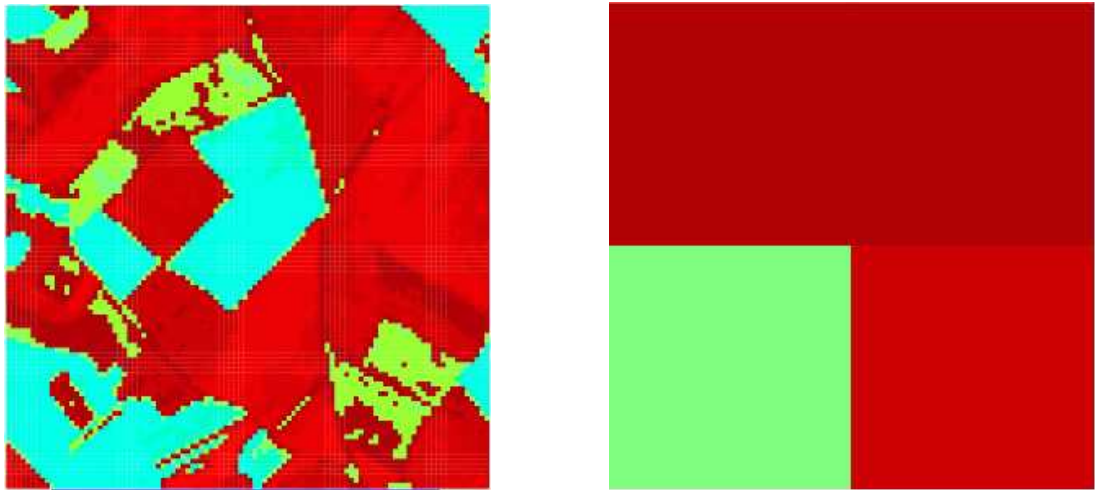


FIG. 1.12 – *Extrait (zone au sol: 2 × 2 km) d'une image SPOT/HRVIR de résolution 20m×20m acquise aux environs de la ville de Chartres le 18 mai 1992 et même extrait dégradé à la résolution de VEGETATION 1km×1km.*

Sur une période de six mois, on peut disposer approximativement de six images HRVIR et de plusieurs dizaines d'images VEGETATION (Fig 1.13).

Le nombre d'images disponibles pour chaque capteur est volontairement sous-estimé ici pour rappeler les difficultés d'acquisition en cas de conditions climatiques défavorables. Pour cette raison, au lieu d'utiliser directement les images quotidiennes, on a recours à des synthèses décadaires d'images. Ces synthèses périodiques consistent à produire une image à partir des dix images quotidiennes précédemment acquises en favorisant celles bénéficiant de conditions d'acquisition favorables. Cette opération revient à supprimer l'effet des nuages.

En pratique, les images HRVIR et VEGETATION sont utilisées à des fins différentes :

- définir le plan d'occupation des sols de la région d'étude grâce à la haute résolution spatiale,
- assurer le suivi des cultures grâce aux données fréquentes de VEGETATION.

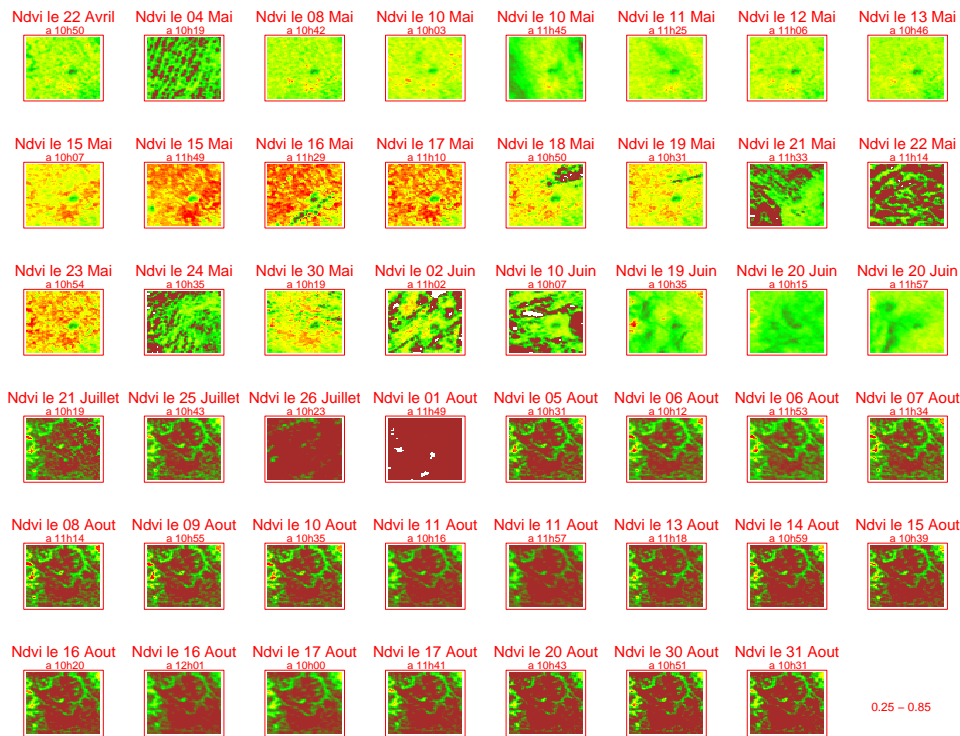


FIG. 1.13 – Valeurs de NDVI calculées à partir des mesures de réflectances dans le *R* et le *PIR* sur une zone d'environ 40×40 km autour de la ville de Chartres à différentes dates du 22 avril au 31 août 1998.

1.3.2 Traitement des données à haute résolution spatiale

Dans notre cas, l'intérêt majeur des données à haute résolution spatiale est leur aptitude à la détermination du plan d'occupation des sols.

Ceci est possible sous certaines conditions :

- la résolution spatiale du capteur permet de distinguer les différentes parcelles. C'est le cas avec HRVIR et sa résolution de 20 m sur 20 m, comme on le voit sur la figure 1.12 à gauche,
- la mesure effectuée par le capteur est différente pour chaque thème observé sur l'image.

Cette seconde condition n'est pas toujours respectée notamment si des cultures différentes sont observées à des stades phénologiques proches. Pour remédier à ce problème, des techniques de classification sont appliquées sur plusieurs images prises à diverses dates. En général, trois ou quatre images, prises à des périodes variées durant l'année, suffisent à déterminer avec succès un plan d'occupation des sols (Maselli *et al.* [61] et Viñas et Baulies [87]). La fréquence mensuelle des acquisitions à haute résolution spatiale n'est pas un obstacle pour atteindre cet objectif.

Diverses techniques de classification sont utilisées dans le cadre du traitement de données de télédétection. On pourra se référer à [41], chapitres 8 et 9, pour une revue de ces méthodes. Les figures 1.14 illustrent la différence entre les classifications de type supervisé

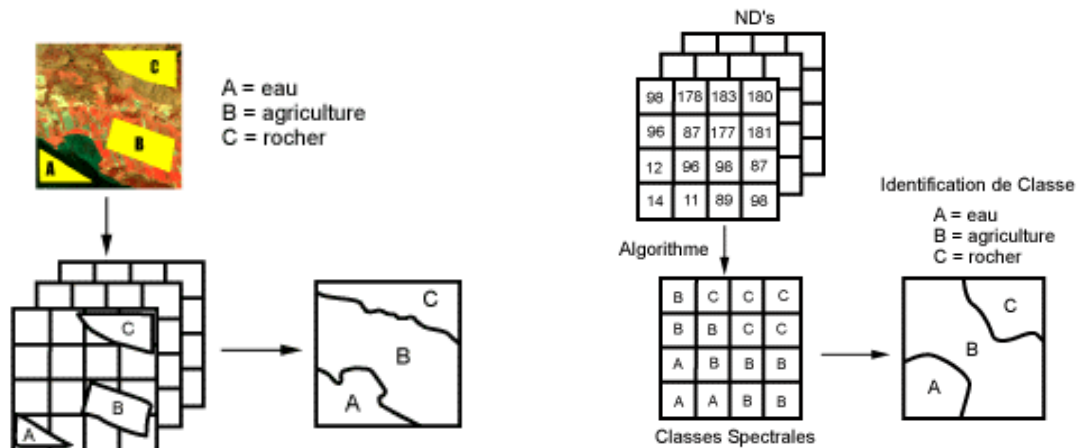


FIG. 1.14 – Illustration des deux types de méthodes de classification : à gauche, classification supervisée ; à droite, classification non supervisée [11].

et non supervisé. Dans le premier cas, l'utilisateur spécifie *a priori* le nombre de classes et l'algorithme recherche alors des objets proches d'objets de référence définis par un jeu d'apprentissage. Dans le second cas, l'utilisateur ne fournit pas d'information *a priori*, l'algorithme effectue une partition des données en fonction de leur structure. C'est ensuite à l'utilisateur de déterminer quels sont les thèmes correspondant à chaque élément de la partition.

1.3.3 Traitement des données à haute résolution temporelle

1.3.3.1 Hypothèse de base sur les pixels mixtes

Les données à haute résolution temporelle sont indispensables pour le suivi de cultures. Du fait de leur basse résolution spatiale (1 km^2), les parcelles, dont la taille peut être inférieure à 1 km^2 en Europe Occidentale, sont observées en mélange sous un même pixel. Ces pixels, appelés pixels mixtes ou mixels, nécessitent un pré-traitement afin d'extraire les informations relatives à chaque parcelle et donc à chaque culture.

Les techniques de désagrégation, traduction approximative de « unmixing », sont basées sur l'hypothèse que la réflectance mesurée sur un pixel mixte est la somme des réflectances des thèmes présents sous le pixel pondérée par le pourcentage d'occupation du sol de chaque thème [78]. Cette hypothèse se traduit par l'équation (1.4)

$$R_i = \sum_{k=1}^{N_c} x_i^k R_i^k \quad (1.4)$$

où les notations utilisées sont les suivantes :

- R_i représente la réflectance mesurée sur le pixel i , I désigne le nombre de pixels de l'image.

- N_c désigne le nombre de thèmes observés en mélange sur l'ensemble de l'image. Il est déterminé par le plan d'occupation des sols.
- Les x_i^k sont les pourcentages de surface occupée par la culture k sur le pixel i . Ces pourcentages représentent la traduction numérique du plan d'occupation des sols et sont supposés connus.
- Les R_i^k sont les valeurs à estimer, elles correspondent à la réflectance inconnue de la culture k sur le pixel i . On fait l'hypothèse d'homogénéité intra-pixel : les surfaces occupées par la même culture présentes sous un pixel ont des réflectances égales.

1.3.3.2 Résolution du problème de désagrégation

Nous présentons deux méthodes de résolution du problème de désagrégation. La première est simple dans sa mise en œuvre mais suppose que les valeurs de réflectances d'un thème sont identiques sur l'image. La seconde méthode de résolution est plus complexe mais permet d'exprimer une variabilité des valeurs de réflectances sur la zone d'étude.

Régression linéaire multiple

Cherchali *et al.* [13] simplifient le modèle exprimé en (1.4) en supposant que la réflectance d'une culture est la même sur tous les pixels à une date donnée.

Dans ces conditions, l'expression (1.4) devient :

$$\begin{cases} R_i &= \sum_{k=1}^{N_c} x_i^k \bar{R}_i^k + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{cases} \quad (1.5)$$

en supposant $R_i^k = \bar{R}_i^k$.

L'estimation des $(\bar{R}_i^k)_{k=1, \dots, N_c}$ se fait alors dans le cadre classique d'un modèle de régression linéaire multiple où $(R_i)_{i=1, \dots, I}$ est la variable à expliquer et x_i^k les variables explicatives. Cette procédure, attrayante par sa simplicité, présente le défaut de ne fournir qu'une valeur moyenne sur l'image de la réflectance d'une culture. En pratique, si, par exemple, les dates de semis sont différentes sur la région étudiée, ce résultat n'est pas satisfaisant.

Modèle à paramètres aléatoires

Pour remédier à cela, Faivre et Fischer [34] proposent un modèle linéaire à paramètres aléatoires pour prendre en compte la variabilité des R_i^k sur l'image.

Le modèle décrit en (1.6)

$$\begin{cases} R_i &= \sum_{k=1}^{N_c} x_i^k R_i^k + \varepsilon_i \\ R_i^k &\sim \mathcal{N}(\theta_k, \sigma_k^2) \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{cases} \quad (1.6)$$

complète le modèle décrit en (1.5) par une distribution de probabilité sur les R_i^k . Dans ce modèle les θ_k représentent les valeurs moyennes des réflectances pour chaque culture analogues aux \bar{R}_i^k estimées dans (1.5) ; les σ_k^2 traduisent la variabilité des R_i^k sur l'image.

Sous des hypothèses d'indépendance intra- et inter-pixels, on a :

$$R_i \sim \mathcal{N} \left(\sum_{k=1}^{N_c} x_i^k \theta^k, \sigma_\varepsilon^2 + \sum_{k=1}^{N_c} [x_i^k]^2 \sigma_k^2 \right). \quad (1.7)$$

L'estimation par maximum de vraisemblance se fait par un algorithme itératif alternant l'estimation des paramètres de moyenne (θ_k) et des paramètres de variance (σ_k^2 et σ_ε^2). Sous les hypothèses de normalité et d'indépendance, cet algorithme converge et les estimateurs obtenus sont équivalents aux estimateurs du maximum de vraisemblance (Anderson [1], Harville [43]).

Pour prédire les variations individuelles des R_i^k , les auteurs utilisent le meilleur prédicteur linéaire sans biais (Best Linear Unbiased Predictor, BLUP, Henderson [45]). Conditionnellement à la réflectance observée R_i sur le pixel i , le BLUP est estimé par :

$$\widehat{R}_i^k = \hat{\theta}^k + \frac{x_i^k \hat{\sigma}_k^2}{\widehat{\sigma}_i^2} [R_i - X_i \hat{\Theta}] \quad (1.8)$$

où $\hat{\Theta} = (\hat{\theta}^1, \dots, \hat{\theta}^p)^T$, $X_i = (x_i^1, \dots, x_i^p)$ et $\widehat{\sigma}_i^2$ l'estimation de $\tilde{\sigma}_i^2 = \sigma_\varepsilon^2 + \sum_{k=1}^{N_c} [x_i^k]^2 \sigma_k^2$.

On dispose ainsi d'une estimation de la réflectance de chaque culture pour chaque pixel et non plus d'une valeur moyenne sur l'image.

1.3.4 Ajustement d'un modèle d'évolution temporelle du LAI

L'étape d'ajustement d'un modèle temporel intervient lorsque l'on dispose pour chaque pixel et à chaque date des valeurs de LAI, ou d'une autre variable d'état, de la culture d'intérêt, donc après la mise en œuvre éventuelle de procédures de désagrégation ((Fig. 1.15). Les données numériques disponibles pour cette étape (Fig. 1.16) correspondent aux mesures du LAI pour chaque pixel à chaque date d'acquisition.

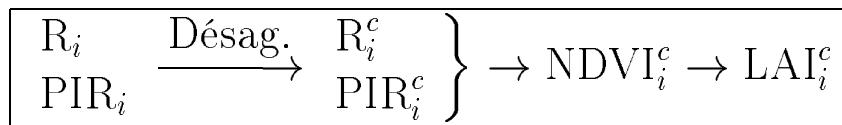


FIG. 1.15 – Chaîne de traitement des données depuis l'acquisition des réflectances dans le rouge (R_i) et proche infrarouge (PIR_i) à basse résolution spatiale pour chaque pixel i de 1 à I : extraction des réflectances propres à la culture d'intérêt (R_i^c et PIR_i^c) puis calcul du NDVI_i et du LAI_i de la culture.

Faivre et Delécolle [32] proposent d'ajuster le modèle de Baret (section 1.2.3.1) contraint par la relation $T_r = \log(1 + \exp(a \times T_c))/b$. en appliquant un modèle de régression non linéaire pour les données relatives à chaque pixel indépendamment. L'estimation des paramètres du modèle se fait par minimisation d'un critère des moindres carrés tenant compte du calcul du LAI par conversion du NDVI après désagrégation. Si l'on note g la fonction qui permet de convertir le LAI en NDVI (équation (1.1)), le critère de moindres carrés à minimiser s'écrit :

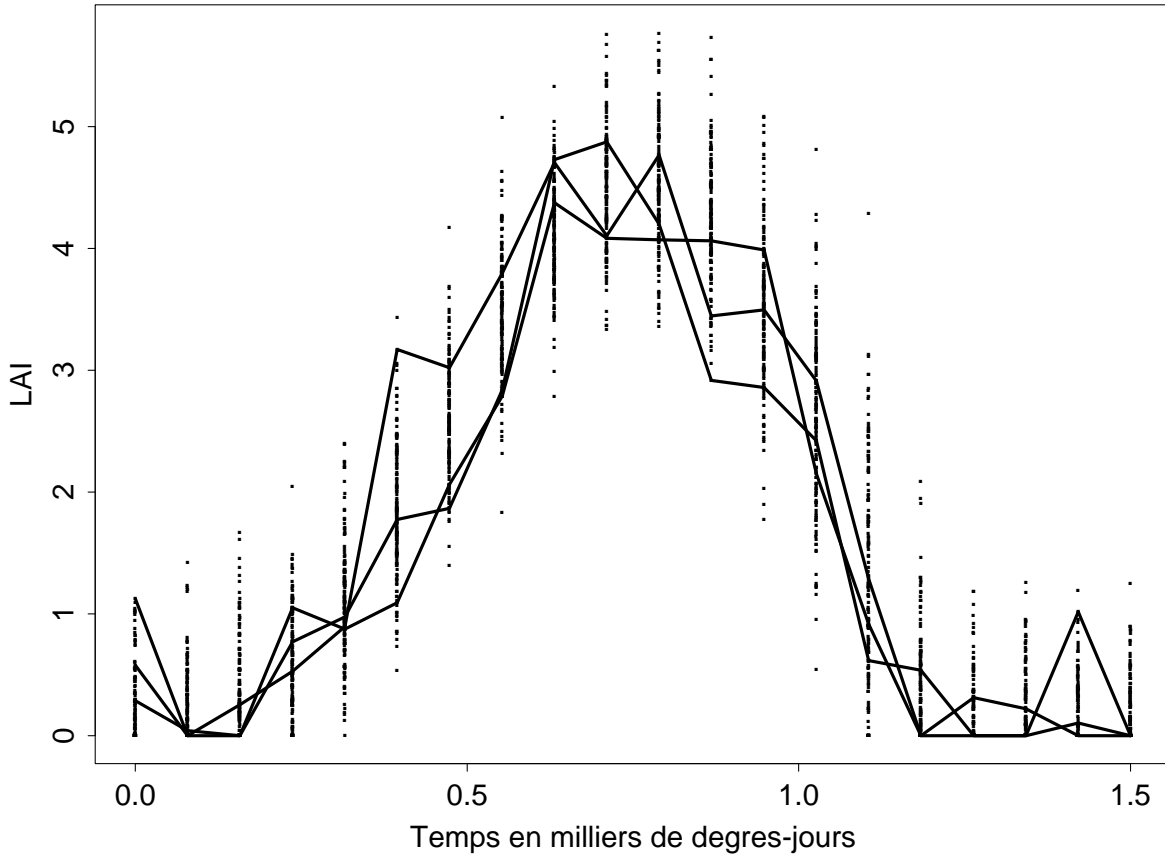


FIG. 1.16 – Exemple de données à traiter pour l'ajustement d'un modèle d'évolution temporelle du LAI; données simulées avec 1500 pixels et 50 dates accompagnées de la courbe pour 3 pixels. La simulation de données est détaillée au chapitre 4.

$$C(\beta) = C_{LAI}(\beta) + \tau \times C_{NDVI}(\beta) \quad (1.9)$$

avec

$$C_{NDVI}(\beta) = \sum_{j=1}^J [g(LAI_j) - \widehat{NDVI}_j]^2 \quad (1.10)$$

$$C_{LAI}(\beta) = \sum_{j=1}^J [(LAI_j) - g^{-1}(\widehat{NDVI}_j)]^2 \quad (1.11)$$

$$\tau = \frac{2}{[\widehat{NDVI}_\infty - \widehat{NDVI}_s]} \quad (1.12)$$

où \widehat{NDVI}_j désigne le NDVI de la culture considérée, calculé après désagrégation à la $j^{\text{ème}}$

date et J désigne le nombre d'observations de chaque pixel. $NDVI_{\infty}$ et $NDVI_s$ sont les constantes utilisées dans l'équation (1.1).

Les résultats obtenus permettent de proposer des cartographies des paramètres d'un modèle simple d'évolution temporelle du LAI. Ces représentations donnent à la fois une information spatiale et temporelle sur le développement des cultures.

L'inconvénient majeur de cette approche est de considérer séparément les données relatives à chaque pixel. Les résultats obtenus sur un pixel ne tiennent pas compte de ceux obtenus sur l'ensemble de l'image. On peut ainsi supposer un manque de robustesse vis-à-vis d'éventuels points aberrants et des difficultés à estimer les paramètres d'un modèle non linéaire si peu de dates sont disponibles.

1.4 Amorce des travaux de recherche

La présentation du contexte général de la télédétection spatiale pour l'agriculture permet de préciser le cadre de cette thèse. Dans un premier temps, nous évoquons les diverses pistes de travail potentiellement exploitables dans ce cadre. Puis, suite à l'examen de ces diverses possibilités, nous définissons la question de recherche qui sera développée dans cette thèse.

1.4.1 Pistes de recherche

Les premières pistes qui se présentent dans le contexte précédemment défini sont essentiellement liées à l'origine double des données de télédétection spatiale pour le suivi de cultures : haute résolution spatiale/faible répétitivité et haute résolution temporelle/faible résolution spatiale. Dans ces conditions, on peut chercher à quantifier l'apport de chaque source d'information afin d'optimiser leur utilisation.

Cette voie est une approche relativement générique qui n'est pas envisageable avant d'avoir étudié plus en détail le modèle de désagrégation qui se situe à la base de l'utilisation des images à haute résolution temporelle. On peut notamment s'intéresser à la robustesse du modèle de désagrégation vis-à-vis des hypothèses d'indépendance et de normalité des réflectances des cultures généralement requises pour simplifier les procédures de calcul. Or, pour des cultures voisines spatialement, donc aux conditions pédo-climatiques très proches, l'hypothèse de non-indépendance est à envisager. La prise en compte de structures de variance-covariance intra-pixel et inter-pixels nécessite l'adaptation de la procédure d'estimation déjà réalisée qui utilise très fortement les propriétés d'indépendance. La robustesse du modèle de désagrégation peut être également étudiée vis-à-vis de la connaissance du plan d'occupation des sols considéré comme connu parfaitement alors qu'il est en fait estimé. De même, les distributions des paramètres sont supposées gaussiennes ; d'autres lois de distribution peuvent être envisagées. Le modèle de désagrégation est utilisé à chaque date indépendamment les unes des autres. Une extension du modèle peut être envisagée pour instaurer la prise en compte de la dynamique temporelle.

L'ensemble des pistes évoquées ici est sous la contrainte forte de l'efficacité opérationnelle des méthodes utilisées. La taille des données qu'il faut considérer nécessite des méthodes algorithmiques efficaces. Dans le but de pouvoir fournir un outil opérationnel,

obtenir des temps de calcul non prohibitifs est une des conditions d'une utilisation des travaux. Nous fixons une limite supérieure approximative pour les temps de calcul autour d'une heure pour le problème standard consistant à traiter environ 1500 pixels sur quelques dizaines de dates.

1.4.2 Question de recherche

Dans les pistes évoquées précédemment, c'est l'extension du modèle de désagrégation vers la prise en compte d'une dynamique temporelle qui semble susceptible d'apporter une amélioration significative du traitement statistique des données de télédétection. Cette option paraît préférable à celle consistant à complexifier le modèle de désagrégation en essayant de lever ses hypothèses simplificatrices.

En effet, lever les hypothèses d'indépendance intra-pixel et inter-pixels présente le risque de compromettre l'aspect opérationnel des méthodes mises en œuvre. De plus le gain en précision n'est pas forcément garanti et les résultats du modèle de désagrégation sont déjà satisfaisants pourvu que l'erreur dans la détermination du plan d'occupation des sols ne soit pas trop importante (Faivre et Fischer [34]).

Pour améliorer la prise en compte de la dynamique temporelle, les travaux sont orientés vers l'étape de couplage des résultats de la désagrégation avec un modèle d'évolution temporelle des cultures. L'analyse de l'existant révèle que l'inconvénient majeur de la méthode existante réside dans le traitement individuel des pixels, des calculs de régression non linéaire sont appliqués sur les données relatives à chaque pixel séparément.

Dans le contexte du suivi de cultures par images satellites, la contribution des travaux de thèse se situent au niveau de l'estimation des paramètres d'un modèle d'évolution temporelle des cultures. La question de recherche peut alors se formuler de la façon suivante :

Comment améliorer le couplage de données de télédétection avec un modèle d'évolution temporelle d'un indice traduisant le développement de la culture ?

Dans cette formulation, le problème de la désagrégation n'est pas évoqué car il constitue un cas particulier parmi ceux que nous nous proposons de traiter. En effet, les pixels mixtes n'apparaissent que lorsque la taille des parcelles observées est inférieure à la résolution spatiale du capteur utilisé, ce qui n'est pas une généralité.

Dans le cadre de la question de recherche précédemment posée, nous nous attacherons à la question statistique sous-jacente qui peut être formulée ainsi :

Comment améliorer l'estimation des paramètres d'un modèle non linéaire pour l'ajustement à un ensemble de courbes ?

Par cette formulation, la portée des travaux n'est plus restreinte au problème de suivi de cultures par télédétection spatiale mais elle ouvre la porte vers d'autres domaines comme la pharmacocinétique ou l'épidémiologie, deux disciplines familières avec la modélisation de courbes non linéaires.

1.4.3 Proposition

L'étude de l'existant a permis de mettre en évidence les avantages et les inconvénients des méthodes utilisées dans les diverses étapes du traitement des données de télédétection.

Lors de la phase de désagrégation, l'introduction de paramètres aléatoires dans un modèle linéaire permet d'obtenir des valeurs de réflectances propres à chaque pixel et non plus une valeur moyenne sur l'image estimée dans le cadre d'une régression linéaire multiple.

La phase de couplage d'un modèle non linéaire à un ensemble de courbes est effectuée courbe par courbe, autrement dit, pixel par pixel. C'est ce point négatif de la procédure qui a été mis en exergue.

En tenant compte de ces deux remarques issues de l'analyse de l'existant, nous proposons d'effectuer la phase de couplage par un modèle non linéaire à paramètres aléatoires (Déjean *et al.* [30]).

Ce type de modèle permet de modéliser l'ensemble des courbes simultanément en créant un lien entre les pixels *via* une distribution de probabilité sur les paramètres du modèle. Même si un lien existe entre les pixels, il ne s'agit pas d'inclure une structure spatiale sur l'image. Sur une région où l'on étudie une culture en particulier, les différences qui peuvent apparaître dans le développement de la plante sont vraisemblablement dues aux pratiques culturales des exploitants agricoles (date de semis, utilisation d'engrais, pratiques d'irrigation ...) plutôt qu'à des structures spatiales sous-jacentes (type de sol, conditions climatiques ...). Cependant, si de telles structures sont effectivement à l'origine de la variabilité, le modèle à paramètres aléatoires permet de retrouver ces caractéristiques.

La présentation détaillée de ce type de modèles fait l'objet du chapitre suivant.

Chapitre 2

Le modèle non linéaire à paramètres aléatoires

L'utilisation de modèles non linéaires à paramètres aléatoires a été proposée et justifiée dans le chapitre précédent. L'objet de ce chapitre est de décrire ce qu'est un modèle non linéaire à paramètres aléatoires et d'en préciser les diverses spécifications rencontrées dans la littérature. Pour chacune de ces spécifications, nous présentons la formulation, les hypothèses sous-jacentes et les idées majeures des procédures d'estimation.

2.1 Généralités

L'idée des modèles à paramètres aléatoires est de modéliser les comportements d'un ensemble d'individus par une unique fonction dont les paramètres varient d'un individu à l'autre.

De cette façon, les modèles permettent de prendre en compte à la fois les variabilités intra- et inter-individus comme cela a été proposé pour la première fois par Sheiner *et al.* [80, 79]. Pour cela, le modèle comprend deux niveaux :

- le premier représente l'équation de régression et exprime la variabilité intra-individu à travers une structure de covariance,
- le second sert à modéliser la variabilité inter-individus.

Le terme hiérarchique est parfois utilisé pour faire référence à ces types de modèles [22].

Les différentes spécifications que nous décrirons par la suite diffèrent dans la formulation du second niveau. En revanche, quelle que soit l'approche envisagée, le premier niveau ressemble toujours à l'expression

$$\begin{cases} Y_i & = f(X_i, \beta_i) + \varepsilon_i \\ E(\varepsilon_i | \beta_i) & = \mathbf{0}_J \\ Cov(\varepsilon_i | \beta_i) & = V_i(\beta_i, \xi) \end{cases} \quad (2.1)$$

où $\mathbf{0}_J$ est un vecteur nul de taille J . La variabilité intra-individu est ici exprimée dans le cas le plus générique incluant l'hétéroscédasticité et des corrélations entre les diverses mesures

sur un individu à travers la matrice $V_i(\beta_i, \xi)$ où ξ représente le vecteur des paramètres de covariance.

Dans l'expression (2.1) du premier niveau du système hiérarchique, les notations utilisées sont les suivantes :

- $Y_i = (Y_{ij})_{j=1, \dots, J}$ représente le vecteur des observations effectuées sur l'individu i pour les J dates,
- f est la fonction de régression,
- X_i désigne la matrice des covariables,
- β_i est le vecteur des paramètres aléatoires.

Dans le cadre des modèles non linéaires à paramètres aléatoires, l'inférence peut porter sur deux types d'information :

1. à l'échelle de l'ensemble de la population étudiée,
2. à l'échelle de l'individu.

Dans le premier cas, nous parlerons d'informations ou de paramètres globaux ou de population. Dans le cadre du problème du suivi de cultures par images satellites, ces indications seront qualifiées de régionales ou à l'échelle de l'image. Dans le second cas, les informations individuelles seront qualifiées de locales ou à l'échelle du pixel.

Le choix entre ces deux types d'informations dépend de l'objectif poursuivi par l'utilisateur. Dans le cadre du suivi de cultures, les deux échelles présentent un intérêt. Pour la détection de zones à risque, ce sont les estimations individuelles à l'échelle du pixel qui fournissent les informations les plus pertinentes. En revanche, pour la prédiction de rendement d'une culture à une grande échelle, ce sont les informations globales qui ont un intérêt. La distinction entre ces deux objectifs pour le suivi de cultures n'entre pas dans le cadre de notre étude, aussi nous intéresserons-nous conjointement à l'inférence au niveau de la population et de l'individu.

2.2 Différentes spécifications pour l'expression de la variabilité inter-individus

L'expression de la variabilité inter-individus au second niveau peut prendre diverses formes, présentées dans les sections suivantes. L'ordre de présentation de ces approches suit le degré des restrictions nécessaires pour l'écriture du modèle : du non paramétrique qui ne suppose aucune hypothèse restrictive au bayésien qui ajoute une connaissance *a priori* à un modèle paramétrique.

2.2.1 Approche non paramétrique

2.2.1.1 Écriture du modèle

Mallet [58] propose une approche non paramétrique pour modéliser le caractère aléatoire des paramètres de la fonction non linéaire. Le niveau 1 du modèle étant écrit comme en (2.1), la spécification non paramétrique s'exprime simplement au niveau 2 par

$$\beta_i \sim H \quad (2.2)$$

où H représente une distribution de probabilités sans aucune restriction.

Le grand avantage de cette modélisation est de ne faire aucune hypothèse sur la forme de la distribution de probabilité des paramètres aléatoires.

2.2.1.2 Estimation

L'estimation de la densité des paramètres aléatoires dans le cadre non paramétrique, proposée par Mallet [58] et Schumitzky [75], se fait par maximum de vraisemblance. Il s'agit ici seulement de l'estimation de la densité et non pas des paramètres de variance (représentés dans l'écriture générique par ξ), supposés connus ou estimés lors d'une analyse préliminaire.

La vraisemblance marginale pour l'ensemble des données peut s'écrire

$$L(H) = \prod_{i=1}^I \int_{\mathcal{B}} p_i(\zeta) dH(\zeta) \quad (2.3)$$

avec les notations suivantes :

- \mathcal{B} représente le support de H supposé compact et inclus dans \mathbb{R}^k ,
- $p_i(\zeta) = p_{y_i|\beta}(y_i|x_{i1}, \dots, x_{iJ}, \zeta, \sigma_\varepsilon^2)$ désigne la densité conditionnelle de y_i sachant β_i ,
- ζ est la variable d'intégration.

Le maximum de vraisemblance H_{MV} est défini par :

$$H_{MV} \in \mathcal{H} / L(H_{MV}) \geq L(H), \quad \forall H \in \mathcal{H} \quad (2.4)$$

où \mathcal{H} est l'ensemble des densités de probabilités sur \mathcal{B} .

Soit \mathcal{H}^D l'ensemble des distributions de probabilités discrètes de \mathcal{H} avec un support d'au plus I points. À une distribution de \mathcal{H}^D , sont associés un ensemble de points $Z = \{\zeta_1, \dots, \zeta_I\}$ de \mathcal{B} et un ensemble de poids $\omega = \{\omega_1, \dots, \omega_I\}$ où $\omega_k \geq 0$ et $\sum_{k=1}^I \omega_k = 1$. Dans ces conditions, une distribution H de \mathcal{H}^D satisfait la relation (2.5) :

$$H = \sum_{k=1}^I \omega_k \delta(\zeta_k) \quad (2.5)$$

où $\delta(\zeta_k)$ est la mesure de Dirac qui charge le point ζ_k .

Les procédures d'estimation de la densité des paramètres aléatoires sont basées sur le résultat suivant :

➤ *L'estimateur du maximum de vraisemblance de la densité des paramètres aléatoires est une distribution discrète dont le support se caractérise par au plus I points de \mathcal{B} :*

$$H_{MV} \in \mathcal{H}^D. \quad (2.6)$$

Un schéma de la démonstration de ce résultat est proposé par Davidian et Giltinan [22]. On pourra se référer à Mallet [58] pour une version plus détaillée.

Par construction, le résultat de l'estimation est une distribution discrète. La distribution réelle étant en général supposée continue, ce résultat est lissé par exemple par une convolution avec un noyau gaussien.

La procédure d'estimation esquissée précédemment suppose que les paramètres de variance sont connus. En pratique, ces paramètres sont estimés lors d'une analyse préliminaire. Par exemple, la structure de covariance peut être estimée sur quelques individus pour lesquels on dispose d'un nombre d'observations important.

Par ailleurs, Mallet *at al.* [59] notent que cette procédure d'estimation ne fournit pas d'indicateurs sur la précision de l'estimation de la distribution. Dans ces conditions, on peut émettre des réserves sur l'interprétation des résultats, particulièrement sur des petits jeux de données.

2.2.2 Approche semi-paramétrique

L'approche semi-paramétrique est conceptuellement assez proche de la spécification non paramétrique décrite précédemment. En détaillant la procédure d'estimation, nous verrons en quoi la méthode présente par ailleurs un aspect paramétrique. Nous reviendrons sur la terminologie à la fin de cette section.

2.2.2.1 Écriture du modèle

Une autre approche de type non paramétrique a été proposée par Davidian et Gallant [20]. Le niveau 2 du modèle hiérarchique s'écrit

$$\begin{cases} \beta_i &= d(a_i, \beta, B_i) \\ B_i &\sim h, \quad h \in \mathcal{H}^L \end{cases} \quad (2.7)$$

où a_i représente les covariables susceptibles d'expliquer les variations inter-individus des paramètres, β désigne le paramètre moyen commun à l'ensemble des individus et les B_i désignent les effets aléatoires. La fonction d relie ces trois quantités au paramètre aléatoire β_i de façon éventuellement non linéaire.

L'écriture (2.7) permet de prendre en compte naturellement l'influence de covariables dans la variabilité des paramètres aléatoires. De plus, l'ensemble de définition de la densité h des paramètres aléatoires est restreint à une classe spécifique de densités notée \mathcal{H}^L . Cet ensemble défini et étudié par Gallant et Nychka [37] est essentiellement défini par un critère de différentiabilité des densités : $h \in \mathcal{H}^L$ doit être $p/2$ fois différentiable avec p la taille du vecteur β_i . La classe \mathcal{H}^L inclut la loi normale, des densités multi-modales, des densités dissymétriques ainsi que des distributions à queue plus ou moins fine que la loi normale. Cette classe exclut en revanche, des densités présentant des replis, des sauts ou des oscillations, traduisant des comportements peu en phase avec des phénomènes naturels.

2.2.2.2 Estimation

L'estimation conjointe des paramètres $(\beta, \sigma_\varepsilon^2)$ et de la densité des B_i est basée sur le maximum de vraisemblance.

La vraisemblance pour l'ensemble des données s'exprime comme le produit des densités marginales

$$\prod_{i=1}^I \int p_{y|\beta}(y_i|x_{i1}, \dots, x_{iJ}, a_i, \beta, \sigma_\varepsilon^2, B)h(B)dB \quad (2.8)$$

en supposant l'indépendance des individus.

La maximisation de l'expression (2.8) s'effectue simultanément sur les paramètres $(\beta, \sigma_\varepsilon^2)$ et sur la densité h . Les estimateurs sont obtenus en minimisant

$$L_{NPL}(\beta, \sigma_\varepsilon^2, h) = - \sum_{i=1}^I \log \int p_{y_\beta}(y_i|x_{i1}, \dots, x_{iJ}, a_i, \beta, \sigma_\varepsilon^2, B)h(B)dB \quad (2.9)$$

où l'indice NPL est mis pour Non Paramétrique Lisse.

L'estimation de h minimise $L_{NPL}(\beta, \sigma_\varepsilon^2, h)$ sur un ensemble spécifique de densités \mathcal{H}^L . Or, une des propriétés de l'ensemble \mathcal{H}^L est que toute densité peut s'écrire selon un développement en série. La propriété suivante montrée par Gallant et Nychka [37] précise ce résultat.

➤ *Toute densité appartenant à \mathcal{H}^L peut s'écrire comme le carré d'un polynôme de degré infini multiplié par la densité d'une loi normale :*

$$h(x) = \{P_\infty(L^{-1}x)\}^2 \times \mathcal{N}_p(0, LL')$$

où L est une matrice triangulaire supérieure.

L'expression d'une densité de \mathcal{H}^L par une série infinie est inutilisable en pratique. On procède donc par troncature pour traiter avec des polynômes de degré fini. Une densité de \mathcal{H}^L est alors approchée par $h_K(b)$ défini par

$$h_K(b) = \frac{\{P_K(L^{-1}b)\}^2 \times p(\mathcal{N}_p(0, LL'))}{\int P_K(z) \times p(\mathcal{N}_p(z, \text{Id}_p))dz} \quad (2.10)$$

où P_K est un polynôme de degré K et $p(\mathcal{N}(0, V))$ la densité de la loi normale $\mathcal{N}(0, V)$. L'expression (2.10) montre également que le cas $K = 0$ correspond à une distribution normale.

L'estimation de la densité revient ainsi à l'estimation des coefficients, en nombre fini, du polynôme. L'opposé de la log-vraisemblance approchée s'écrit

$$L_{NPL,K}(\beta, \sigma_\varepsilon^2, \eta) = - \sum_{i=1}^I \log \int p_{y_\beta}(y_i|x_{i1}, \dots, x_{iJ}, a_i, \beta, \sigma_\varepsilon^2, B)h(B|\eta)dB \quad (2.11)$$

avec η le vecteur des coefficients du polynôme.

Le problème à résoudre désormais est un problème d'optimisation standard en dimension finie. Pour simplifier le problème du calcul de l'intégrale dans l'expression de $L_{NPL,K}(\beta, \sigma_\varepsilon^2, \eta)$, Davidian et Gallant [21] proposent une approximation par quadrature gaussienne. On dispose ainsi des estimations des paramètres de population $(\hat{\beta}, \hat{\sigma}_\varepsilon^2)$ et de la densité (\hat{h}) . L'estimation des effets individuels est alors possible par maximisation en B de la densité *a posteriori* des B_i :

$$p_{y_\beta}(y_i | x_{i1}, \dots, x_{iJ}, a_i, \hat{\beta}, \hat{\sigma}_\varepsilon^2, B) \hat{h}(B | \eta) \quad (2.12)$$

d'où l'on déduit les paramètres individuels

$$\hat{\beta}_i = d(a_i, \hat{\beta}, \hat{B}_i). \quad (2.13)$$

Le problème du choix de K , le degré du polynôme inclus dans l'approximation d'une densité de \mathcal{H}^L , relève du cadre classique de la sélection de modèle dans lequel des critères comme AIC et BIC peuvent être utilisés. Cette procédure de détermination de la valeur K permet également de tester l'hypothèse selon laquelle la densité de probabilité des B_i est normale, cas correspondant à $K = 0$.

Terminologie

La terminologie généralement retenue pour cette approche est « semi-paramétrique » ou « semi-non paramétrique » pour signifier que cette modélisation repose sur des éléments à la fois issus du paramétrique et du non paramétrique.

L'aspect paramétrique réside dans l'estimation de la densité *via* un nombre fini de paramètres correspondant aux coefficients d'un polynôme de degré fini.

D'un autre côté, aucune hypothèse n'est supposée sur la forme explicite de la densité des paramètres aléatoires, ce qui donne un aspect non paramétrique à la méthode.

Toutefois, les auteurs préfèrent voir le sigle SNP interprété comme « Smooth NonParametric » (non paramétrique lisse) et non comme « SemiNonParametric », pour insister sur la classe de densités \mathcal{H}^L qui est à la base de leur méthode.

2.2.3 Approche paramétrique classique

Nous employons les termes « classique » et « bayésien » pour distinguer les deux approches de type paramétrique présentées dans cette section et la suivante.

2.2.3.1 Écriture du modèle

L'approche paramétrique proposée initialement par Beal et Sheiner [5], suppose connue la distribution de probabilité \mathcal{L} des paramètres aléatoires au vecteur de paramètres η près. Le système

$$\begin{cases} \beta_i &= d(a_i, \beta, B_i) \\ B_i &\sim \mathcal{L}(\eta) \end{cases} \quad (2.14)$$

caractérise le deuxième niveau du modèle hiérarchique pour une spécification paramétrique classique.

La loi \mathcal{L} est généralement choisie normale dans les applications. Certaines procédures d'estimation dépendent directement de cette hypothèse. Cependant d'autres lois ont également été utilisées. Wakefield [92] propose l'utilisation d'une loi de Student, qui étant une distribution à queue plus lourde que la normale est une alternative intéressante vis-à-vis de la robustesse à des points aberrants. Beal et Sheiner [6] proposent eux un mélange de lois normales pour prendre en compte une éventuelle multimodalité dans la distribution des effets aléatoires.

2.2.3.2 Estimation

Deux stratégies principales ont été développées pour l'estimation dans le cadre paramétrique classique. L'une est basée sur le traitement séparé des individus, l'autre, procède par maximum de vraisemblance sur l'ensemble des individus.

Traitement séparé des individus

Les procédures basées sur le traitement séparé des individus requièrent deux étapes :

- Estimer les paramètres individuels en considérant séparément chaque individu. Dans le cas des modèles non linéaires, on procède comme dans une régression non linéaire classique, par exemple par moindres carrés.
- Calculer à partir des valeurs individuelles, les paramètres de population représentés par η (moyenne et variance pour la loi normale) dans l'expression (2.14).

Diverses méthodes entrant dans ce cadre seront présentées dans le chapitre suivant consacré exclusivement à l'estimation dans le modèle paramétrique classique.

Maximum de vraisemblance

Dans le cadre générique où les lois de l'erreur et des effets aléatoires ne sont pas précisées, nous pouvons seulement exprimer la forme de la densité des observations.

Si on note $p(Y_i|a_i, \beta, B_i)$ la densité des observations pour l'individu i conditionnellement aux B_i et $p(B_i|\eta)$ la densité des effets aléatoires, alors la densité marginale des observations de l'individu i s'écrit :

$$V(Y_i|a_i, \beta, B_i) = \int p(Y_i|a_i, \beta, B_i)p(B_i|\eta)dB_i \quad (2.15)$$

où la notation \int symbolise une intégrale éventuellement multiple de la taille du vecteur B_i .

En supposant l'indépendance des Y_i , la densité marginale des observations de l'ensemble des individus s'écrit alors :

$$V(Y|a, \beta, B) = \prod_{i=1}^I V(Y_i|a_i, \beta, B_i) = \prod_{i=1}^I \left[\int p(Y_i|a_i, \beta, B_i)p(B_i|\eta)dB_i \right] \quad (2.16)$$

en effectuant le produit des densités individuelles et avec $Y = (Y_1, \dots, Y_I)$.

Pour une estimation par maximum de vraisemblance des paramètres globaux β et η , il faut donc maximiser la quantité (2.16). Cependant, l'intégrale ainsi définie n'admet en général pas d'expression analytique simple. Les stratégies basées sur le maximum de vraisemblance procèdent donc par approximation. Nous verrons dans le chapitre suivant en quoi consistent ces approximations et leur application au cas de densités gaussiennes pour la loi de l'erreur et des effets aléatoires.

2.2.4 Approche paramétrique bayésienne

2.2.4.1 Écriture du modèle

L'approche bayésienne, proposée par Racine-Poon [68] pour le modèle non linéaire, introduit un niveau supplémentaire dans le modèle hiérarchique paramétrique. Ce troisième niveau exprime une connaissance *a priori* sur les paramètres de la densité considérée pour les effets aléatoires. Les niveaux 2 et 3 du modèle séparés par « -- » s'écrivent

$$\left\{ \begin{array}{ll} \beta_i & = d(a_i, \beta, B_i) \\ B_i & \sim \mathcal{L}(\eta) \\ \text{--} & \\ (\beta, \eta) & \sim \mathcal{L}'_{\beta^*, \eta^*}(\beta, \eta) \end{array} \right. \quad (2.17)$$

où les deux premières lignes expriment la variabilité inter-individus de façon paramétrique par la loi \mathcal{L} de paramètre η . La troisième ligne de l'écriture (2.17) exprime l'information *a priori* apportée sur les paramètres par une loi \mathcal{L}' d'hyper-paramètres β^* et η^* .

Cette formulation est analogue à celle proposée par Lindley et Smith [53] et reprise par Louis [56] dans le cadre du modèle linéaire hiérarchique. Elle présente l'avantage de fournir une densité *a posteriori* des paramètres du modèle en plus de leur estimation. On pourra se référer à Robert [72] pour une monographie sur l'analyse statistique bayésienne.

2.2.4.2 Estimation

La non linéarité du problème à traiter implique des difficultés dans les calculs du modèle bayésien. Comme souvent dans ce cadre, on fait appel à des techniques de calcul comme les méthodes de Monte Carlo par chaînes de Markov (MCMC, Monte Carlo Markov Chain, *Cf.* par exemple Robert [73]) et principalement l'échantillonneur de Gibbs. L'application aux modèles hiérarchiques a été proposée par Wakefield [92] et Gelfand et Smith [38].

En incluant une hypothèse supplémentaire sur la distribution *a priori* des paramètres du modèle, l'approche bayésienne s'écrit par exemple

$$\left\{ \begin{array}{ll} \text{Niv. 1 :} & (Y_i | \beta_i, \sigma_\varepsilon^2) \sim \mathcal{N}(f(X_i, \beta_i), \sigma_\varepsilon^2 \text{Id}_J) \\ \text{Niv. 2 :} & \beta_i = \beta + B_i, \quad B_i \sim \mathcal{N}_p(0, \Gamma) \\ \text{--} & \\ \text{Niv. 3 :} & \beta \sim \mathcal{N}_p(\beta^*, H) \\ & \Gamma^{-1} \sim Wi\{(\rho D^*)^{-1}, \rho\} \quad \sigma_\varepsilon^{-2} \sim Ga(\nu_0/2, \nu_0 \tau_0/2) \end{array} \right. \quad (2.18)$$

avec au niveau 3 un *a priori* défini par une distribution normale pour le paramètre de moyenne β , une loi de Wishart pour l'inverse de la matrice de variance des paramètres aléatoires Γ , et une loi Gamma pour l'inverse de la variance résiduelle.

Dans ce cadre, il est possible d'exprimer la distribution de n'importe quel paramètre conditionnellement aux autres paramètres et aux données grâce au choix de lois conjuguées.

Avec les notations du système (2.18), les densités conditionnelles de β, Γ^{-1} et σ_ε^{-2} s'écrivent

$$\left\{ \begin{array}{l} (\beta|y, \sigma, \Gamma) \sim \mathcal{N}_p\{U(I\Gamma^{-1}\bar{\beta} + H^{-1}\beta^*), U\} \\ (\Gamma^{-1}|y, \beta, \sigma) \sim Wi\left(\left[\sum_{i=1}^I (\beta_i - \beta)(\beta_i - \beta)^T + \rho D^*\right]^{-1}, I + \rho\right) \\ (\sigma_\varepsilon^{-2}|y, \beta, \Gamma) \sim Ga\left(\frac{1}{2}(\nu_0 + IJ), \frac{1}{2}\left[\sum_{i=1}^I (y_i - f_i(\beta_i))^T (y_i - f_i(\beta_i)) + \nu_0\tau_0\right]\right) \end{array} \right. \quad (2.19)$$

avec $\bar{\beta} = I^{-1} \sum_{i=1}^I \beta_i$, $H_i = \sigma_\varepsilon^{-2} X_i^T X_i + \Gamma^{-1}$ et $U^{-1} = I\Gamma^{-1} + H^{-1}$.

L'application de l'échantillonneur de Gibbs (Geman et Geman [39]) permet de passer de ces lois conditionnelles aux lois marginales pour chaque paramètre β, Γ^{-1} et σ_ε^{-2} qui correspondent aux densités *a posteriori* conditionnellement aux données. Ce sont ces quantités qu'il faut maximiser pour estimer les paramètres.

Principe de l'échantillonneur de Gibbs

Considérons un ensemble de variables aléatoires $\{U_1, \dots, U_J\}$ pour lequel la densité jointe est notée $p_{U_1, \dots, U_J}(u_1, \dots, u_J)$. On dispose des densités conditionnelles

$p_{U_s|U_1, \dots, U_{s-1}, U_{s+1}, \dots, U_J}(u_s|u_1, \dots, u_{s-1}, u_{s+1}, \dots, u_J)$ pour $s = 1, \dots, J$.

Dans ces conditions, l'échantillonneur de Gibbs agit de manière itérative pour simuler une chaîne de Markov :

- Initialisation : $\{u_1^{(0)}, \dots, u_J^{(0)}\}$
- Étape l : simulation de $\{u_1^{(l)}, \dots, u_J^{(l)}\}$ selon le système

$$\left\{ \begin{array}{l} u_1^{(l)} \sim p(u_1|u_2^{(l-1)}, u_3^{(l-1)} \dots, u_J^{(l-1)}) \\ u_2^{(l)} \sim p(u_2|u_1^{(l)}, u_3^{(l-1)} \dots, u_J^{(l-1)}) \\ \vdots \\ u_J^{(l)} \sim p(u_J|u_1^{(l)}, u_2^{(l)} \dots, u_{J-1}^{(l)}) \end{array} \right. \quad (2.20)$$

La séquence ainsi générée ($\{U_1^{(T)}, \dots, U_J^{(T)}\}$ après T itérations) est une chaîne de Markov de distribution limite la densité jointe recherchée de $\{U_1, \dots, U_J\}$: $p_{U_1, \dots, U_J}(u_1, \dots, u_J)$.

Les inconvénients de ce procédé résident dans sa lourdeur due à la génération aléatoire de nombreuses valeurs et dans la détermination d'un critère d'arrêt : la distribution limite étant atteinte pour $T \rightarrow \infty$.

2.2.5 Discussion

Les différentes spécifications du modèle à paramètres aléatoires ont en commun la possibilité d'inclure des covariables dans l'expression des paramètres aléatoires. Cette prise en compte est immédiate dans les approches semi-paramétrique et paramétrique ; elle nécessite en revanche une adaptation de l'algorithme d'estimation pour l'approche non paramétrique.

Le problème qui peut se poser quand on exprime les paramètres aléatoires en fonction de covariables (a_i), d'un vecteur de paramètres fixés (β) et d'un vecteur d'effets aléatoires pour l'individu i (B_i) selon la relation

$$\beta_i = d(a_i, \beta, B_i) \quad (2.21)$$

est le choix des covariables susceptibles d'exprimer la variabilité des paramètres aléatoires.

Pour répondre à cela, Maitre *et al.* [57] et Mandema *et al.* [60] suggèrent une stratégie en trois étapes :

1. Ajuster le modèle sans covariable en considérant par exemple $\beta_i = \beta + B_i$,
2. calculer les $\hat{\beta}_i$ et sélectionner parmi des covariables potentielles par des méthodes graphiques :
 - soit en traçant les $\hat{\beta}_i$ en fonction des covariables potentielles, une à la fois [57],
 - soit en effectuant une sélection de variables selon un algorithme classique de type « stepwise » pour chaque composante du vecteur β_i pour éviter de considérer les covariables une par une [60].
3. ajuster le nouveau modèle incluant les covariables sélectionnées et revenir éventuellement à la deuxième étape.

La sélection des covariables étant effectuée, il reste à déterminer la densité de probabilité que l'on suppose sur les B_i . Les diverses spécifications du modèle à paramètres aléatoires présentées précédemment proviennent d'hypothèses différentes supposées sur les variations inter-individus.

L'approche non paramétrique présentée dans la section 2.2.1 est la moins restrictive des méthodes évoquées dans ce chapitre. Elle ne suppose aucune forme explicite pour la distribution des paramètres aléatoires et c'est là que réside son principal atout et en même temps son principal défaut. L'absence de restriction évite de proposer un modèle faux pour la distribution des paramètres aléatoires. Cependant, les résultats qu'elle fournit peuvent sembler un peu abstraits car pour le suivi de cultures, l'interprétation de paramètres est une pratique commune qu'il semble délicat de remettre en cause qui plus est pour l'introduction d'une nouvelle méthodologie. C'est principalement pour cette raison que nous ne retenons pas ce type d'approche.

L'approche qualifiée ici de semi-paramétrique présente *a priori* la meilleure formulation pour notre problème. En effet, estimer la densité des paramètres aléatoires parmi des distributions potentiellement multi-modales est intéressant si par exemple, sur la région étudiée, plusieurs variétés (en nombre inconnu) de la culture d'intérêt ont été semées. De même, une hétérogénéité spatiale des pratiques culturales peut impliquer une dissymétrie

de la distribution des paramètres aléatoires. Les informations qui permettent de déterminer précisément une loi de probabilité ne sont pas forcément facilement accessible ; dans ces conditions le recours à ce type d'approche est un avantage certain.

Cependant, ce travail étant une première étape dans l'utilisation de modèles à paramètres aléatoires pour le suivi de cultures par télédétection, nous nous restreindrons dans ce mémoire à un cas particulier de cette approche qui est le cadre paramétrique classique.

La formulation paramétrique bayésienne exprime une connaissance *a priori* sur les paramètres de l'expression de la variabilité inter-individus. Elle ajoute donc un niveau supplémentaire dans la spécification du modèle. Or, notre intention étant de diminuer les restrictions sur le modèle, nous préférons ne pas retenir ce type d'approche.

La section suivante est basée sur les méthodes d'estimation dans un modèle non linéaire à paramètres aléatoires entièrement paramétrique dans lequel nous nous plaçons pour la suite de l'étude.

Chapitre 3

Estimation dans le modèle paramétrique gaussien

Ce chapitre est entièrement consacré à l'estimation dans le cadre paramétrique classique des modèles non linéaires à paramètres aléatoires. Le choix du modèle paramétrique a été justifié dans la discussion du chapitre précédent. Nous nous concentrons ici sur le modèle où les densités de l'erreur et des effets aléatoires sont supposées gaussiennes. Ce modèle est présenté en détail dans une première partie. Nous présentons ensuite les diverses procédures d'estimation basées soit sur le traitement séparé des individus soit sur le maximum de vraisemblance. Enfin, nous nous intéressons aux propriétés des estimateurs, sujet pas toujours abordé dans la littérature où l'accent est généralement mis sur le comportement numérique des méthodes.

3.1 Généralités sur le modèle paramétrique gaussien

Cette section est destinée à présenter en détail le modèle avec lequel nous travaillerons par la suite. Nous présentons d'abord l'écriture du modèle accompagné de la justification des hypothèses sous-jacentes pour l'expression des différentes sources de variabilité prises en compte. Ensuite, nous explicitons les notations et le vocabulaire utilisés en les reliant au contexte du suivi de culture par images satellites. Pour terminer cette partie sur les généralités du modèle considéré, nous calculons la vraisemblance à partir de l'expression générique donnée dans le chapitre précédent.

3.1.1 Écriture du modèle

Le modèle que nous étudions dans ce chapitre s'écrit

$$\begin{cases} Y_i &= f(t, \beta_i) + \varepsilon_i \\ \beta_i &= \beta + B_i \\ B_i &\sim \mathcal{N}_p(0, \Gamma) \quad i.i.d. \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma_\varepsilon^2 \text{Id}_J) \quad i.i.d. \end{cases} \quad (3.1)$$

Pour mettre en évidence les distributions conditionnelles et marginales que l'on a à manipuler, on peut écrire de façon équivalente

$$\begin{cases} Y_i|B_i & \sim \mathcal{N}(f(t, \beta_i), \sigma_\varepsilon^2 \text{Id}_J) & i.i.d. \\ \beta_i & = \beta + B_i \\ B_i & \sim \mathcal{N}_p(0, \Gamma) & i.i.d. \end{cases} \quad (3.2)$$

Dans les écritures (3.1) et (3.2) du modèle, les notations utilisées sont les suivantes :

- L'indice i décrit les pixels. On note I le nombre de pixels.
- j est l'indice sur les dates. J est le nombre de dates supposé identique pour l'ensemble des pixels. En pratique, cette hypothèse se justifie par l'utilisation de synthèses décennales au lieu d'images quotidiennes qui peuvent être partiellement dégradées selon la couverture nuageuse et fournir ainsi un nombre différent d'acquisitions selon les pixels.
- Y_i représente le vecteur des J valeurs de LAI, ou d'un autre indice choisi pour étudier l'évolution temporelle de la culture, mesurée, ou calculée après désagrégation, sur le pixel i à la $j^{\text{ème}}$ date d'observation.
- f désigne le modèle non linéaire d'évolution temporelle, le modèle de Baret dans notre cas.
- $t = (t_1, t_2, \dots, t_J)$ est le vecteur des dates d'observation, généralement exprimée en degré-jour ou millier de degrés-jours. C'est la seule covariable que nous incluons dans le modèle.
- $\beta_i = (K_i, a_i, b_i, T_{ci}, T_{ri})$ est le vecteur des paramètres du modèle non linéaire pour le pixel i . Il est composée d'une partie commune à l'ensemble des pixels (β) et d'une partie propre au pixel (B_i) représentant l'écart du pixel par rapport à la moyenne de l'image.
- $f(t, \beta_i)$ désigne le vecteur défini par $(f(t_1, \beta_i), f(t_2, \beta_i), \dots, f(t_J, \beta_i))^T$.
- Γ est la matrice de variance-covariance des paramètres aléatoires. Elle sera généralement supposée diagonale et désignée par ses termes diagonaux $\sigma_\beta^2 = (\sigma_K^2, \sigma_a^2, \sigma_b^2, \sigma_{T_c}^2, \sigma_{T_r}^2)$. L'estimation d'une matrice non structurée est susceptible de compromettre l'aspect opérationnel des méthodes mises en œuvre : temps de calculs beaucoup plus importants et parfois non convergence de l'algorithme d'estimation comme constaté sur quelques petits exemples.
- Le terme ε_i désigne l'erreur résiduelle de variance $\sigma_\varepsilon^2 \text{Id}_J$.

Dans ce modèle, les paramètres à estimer sont :

- β la valeur moyenne des paramètres,
- $\Gamma = \text{diag}(\sigma_\beta)$ la matrice de variance des paramètres aléatoires,
- σ_ε^2 la variance résiduelle.

Ce sont les paramètres globaux selon la terminologie évoquée dans la section 2.1 du chapitre précédent. Nous nous intéressons également aux estimations individuelles associées aux paramètres β_i ou B_i .

3.1.2 Hypothèses du modèle

Les hypothèses sous-jacentes à l'écriture du modèle sont discutées selon l'expression des paramètres aléatoires, la variabilité inter-individus, la variabilité intra-individu.

- Expression des paramètres aléatoires : $\beta_i = \beta + B_i$
 - Aucune covariable ne figure dans l'expression des paramètres aléatoires. Nous supposons ainsi que les variations des paramètres entre les individus sont dues à des phénomènes aléatoires non imputables à une quelconque covariable. Ceci suppose une relative homogénéité des individus de la population étudiée. Pour le suivi de cultures, cela suppose notamment que la région étudiée est relativement homogène en termes de conditions climatiques, de sol, *etc.* Si la région d'étude est relativement restreinte, cette hypothèse est raisonnable. L'homogénéité doit également se manifester au niveau des pratiques culturales mises en œuvre sur cette région.
 - Nous supposons également que l'ensemble des paramètres du modèle est potentiellement aléatoire. Cette hypothèse évite de considérer une écriture de la forme $\beta_i = \beta + M_i B_i$ où M_i est une matrice composée de 0 et de 1 destinée à fixer certains paramètres. Par exemple, pour fixer les paramètres K , T_c et T_r , nous écririons :

$$\beta_i = \begin{pmatrix} \mathbf{K}_i \\ \mathbf{a}_i \\ \mathbf{b}_i \\ \mathbf{T}_{c_i} \\ \mathbf{T}_{r_i} \end{pmatrix} = \begin{pmatrix} \overbrace{K}^{\beta} \\ a \\ b \\ T_c \\ T_r \end{pmatrix} + \begin{pmatrix} \overbrace{0 \ 0 \ 0 \ 0 \ 0}^{M_i} \\ 0 \ 1 \ 0 \ 0 \ 0 \\ 0 \ 0 \ 1 \ 0 \ 0 \\ 0 \ 0 \ 0 \ 0 \ 0 \\ 0 \ 0 \ 0 \ 0 \ 0 \end{pmatrix} \begin{pmatrix} \overbrace{K_i}^{B_i} \\ a_i \\ b_i \\ T_{c_i} \\ T_{r_i} \end{pmatrix} = \begin{pmatrix} K \\ a + a_i \\ b + b_i \\ T_c \\ T_r \end{pmatrix}$$

Cependant, décider du caractère fixe d'un paramètre du modèle nécessiterait une connaissance très précise des processus de croissance et de sénescence de la culture considérée. Notre travail n'étant pas spécifique à un type de culture, nous ne souhaitons pas considérer ce cas de figure.

- Variabilité inter-individus : $B_i \sim \mathcal{N}_p(0, \Gamma)$ *i.i.d.*
 - L'hypothèse de normalité des effets aléatoires est la plus courante. Dans une première approche, nous ne considérons pas de lois plus complexes susceptibles de modéliser une éventuelle dissymétrie ou multi-modalité par exemple.
 - Les B_i sont supposés issus indépendamment d'une population distribuée selon une $\mathcal{N}_p(0, \Gamma)$.
- Variabilité intra-individu : $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \text{Id}_J)$ *i.i.d.*
 - L'hypothèse de normalité et de moyenne nulle sur les erreurs suppose que les corrections relatives au traitement des images ont été effectuées de manière satisfaisante et que seul réside un bruit blanc complètement aléatoire.
 - La structure de covariance intra-individu ne prend pas en compte une éventuelle hétéroscédasticité ni une corrélation des observations. La variabilité réside dans

le seul paramètre σ_ε^2 . Cette hypothèse se justifie comme la précédente en supposant que les diverses corrections apportées aux données ne laissent subsister qu'un bruit exempt de toute structure particulière.

3.1.3 Calcul de la vraisemblance

Pour compléter la présentation du modèle paramétrique que nous étudions, nous précisons l'expression de la vraisemblance des paramètres en poursuivant le calcul à partir de la relation (2.16) du chapitre précédent.

Dans notre cas où les distributions des paramètres aléatoires et des erreurs sont supposées gaussiennes, on obtient l'expression de la vraisemblance

$$V(\beta, \Gamma, \sigma_\varepsilon^2 | Y) = (2\pi)^{-\frac{IJ+Ip}{2}} \sigma_\varepsilon^{-IJ} |\Gamma|^{-\frac{I}{2}} \int \exp \left[\sum_{i=1}^I \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 - \frac{1}{2} B_i^T \Gamma^{-1} B_i \right) \right] dB \quad (3.3)$$

et de la log-vraisemblance

$$l(\beta, \Gamma, \sigma_\varepsilon^2 | Y) = -\frac{IJ+Ip}{2} \log(2\pi) - \frac{IJ}{2} \log \sigma_\varepsilon^2 - \frac{I}{2} \log |\Gamma| + \sum_{i=1}^I \left\{ \log \left[\int \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 - \frac{1}{2} B_i^T \Gamma^{-1} B_i \right) dB_i \right] \right\}. \quad (3.4)$$

Les expressions (3.3) et (3.4) illustrent la difficulté de calculer la vraisemblance du modèle à paramètres aléatoires. L'intégrale multiple (de la taille du vecteur B_i) à calculer n'a pas en général d'expression analytique simple. Les procédures d'estimation évoquées dans la suite suggèrent différentes façons de contourner ce problème.

3.2 Procédures d'estimation

Dans la section 2.2.3.2 du chapitre précédent, nous avons évoqué deux stratégies principales pour l'estimation dans le cadre du modèle paramétrique : le traitement séparé des individus et le maximum de vraisemblance. Avant de présenter en détail des méthodes issues des deux stratégies, nous proposons une classification de celles basées sur le maximum de vraisemblance.

3.2.1 Classification des méthodes

L'absence d'expression simple pour la vraisemblance implique le recours à des stratégies permettant de simplifier le problème. Pour les méthodes d'estimation utilisant comme critère le maximum de vraisemblance, nous proposons la classification du tableau 3.1.

La mise en œuvre de l'algorithme EM dans le cadre des modèles non linéaires à paramètres aléatoires produit selon un schéma itératif une suite d'estimateurs qui converge

Vrais. directe	Vrais. directe approchée	Vrais. indirecte
EM	Approximation laplacienne Monte Carlo Quadrature gaussienne	Pseudo-vraisemblance simulée Linéarisation

TAB. 3.1 – Classification des méthodes d'estimation selon le critère à optimiser.

vers la solution du maximum de vraisemblance. C'est la seule méthode qui vise à résoudre la maximisation de la vraisemblance du modèle initial (3.1). Cette stratégie est référencée en tant que « vraisemblance directe » dans le tableau 3.1.

La deuxième colonne du tableau 3.1 intitulée « vraisemblance directe approchée » contient les méthodes qui maximisent une approximation de la vraisemblance du modèle initial. Le problème du calcul de la vraisemblance résidant essentiellement dans le calcul d'intégrales, les techniques classiques d'intégration numérique (*Cf.* Dahlquist *et al.* [19] ou Davis et Rabinowitz [23]) ont été mises en œuvre pour approcher la vraisemblance. Les estimateurs ainsi calculés sont donc des estimateurs maximisant une approximation de la vraisemblance du modèle initial, d'où la terminologie employée : « vraisemblance directe approchée ».

Les méthodes recensées dans la troisième classe modifient le modèle initial en un modèle où la vraisemblance a une expression plus simple. Les estimateurs sont alors calculés en maximisant la vraisemblance du nouveau modèle. Nous faisons référence à ces méthodes en tant que « vraisemblance indirecte ».

3.2.2 Méthodes basées sur le traitement séparé des individus

Les méthodes basées sur les estimations individuelles ont été évoquées dans le chapitre précédent. Nous présentons deux procédures similaires dans l'estimations des paramètres individuels mais différentes pour le calcul des paramètres de population.

3.2.2.1 Estimation des paramètres individuels

L'estimation des paramètres individuels est effectuée à partir du premier niveau du modèle (3.1) ($Y_{ij} = f(t_j, \beta_i) + \varepsilon_{ij}$ et $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2) i.i.d.$)

Les paramètres individuels β_i sont estimés par minimisation d'un critère de moindres carrés :

$$\hat{\beta}_i = \arg \min_{\theta} \sum_{j=1}^J [Y_{ij} - f(t_j, \theta)]^2. \quad (3.5)$$

Sous des hypothèses classiques d'indépendance, d'homoscédasticité et de normalité de l'erreur, l'estimateur défini par (3.5) est équivalent à l'estimateur du maximum de vraisemblance (*Cf.* par exemple Huet *et al.* [48]).

Dans ces mêmes conditions, l'estimateur de σ_ε^2 est défini par :

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{J-p} \sum_{j=1}^J [Y_{ij} - f(t_j, \hat{\beta}_i)]^2 \quad (3.6)$$

où p est la taille du vecteur β_i .

Sous des conditions de régularité (Cf. par exemple Gallant [36]), on dispose de la normalité asymptotique de $\hat{\beta}_i$: $\hat{\beta}_i \approx \mathcal{N}(\beta_i, \Sigma_i)$. La matrice de variance asymptotique Σ_i , qui reflète la précision de l'estimateur $\hat{\beta}_i$ est estimée par $\hat{\Sigma}_i = \frac{1}{J} V_{\hat{\beta}_i}^{-1}$ (Cf. Huet *et al.* [47]) avec

$$V_{\hat{\beta}_i} = \frac{1}{\hat{\sigma}_\varepsilon^2} \sum_{j=1}^J \frac{\partial \hat{f}_j}{\partial \beta_i} \left(\frac{\partial \hat{f}_j}{\partial \beta_i} \right)^T \quad (3.7)$$

où $\hat{f}_j = f(t_j, \hat{\beta}_i)$ et $\partial \hat{f}_j / \partial \beta_i$ le vecteur d'éléments $\frac{\partial f}{\partial \beta_i}(t_j, \hat{\beta}_i)$ pour $l = 1, \dots, p$. Comme le notent Steimer *et al.* [83], l'incertitude sur l'estimation des paramètres est grande (*i.e.* valeurs élevées des termes de la matrice de variance asymptotique) quand on dispose de peu de données par individu et/ou d'un bruit résiduel élevé. À l'inverse, les estimateurs individuels sont fiables quand on dispose d'un nombre important de données peu bruitées pour chaque individu.

3.2.2.2 Estimation des paramètres globaux

Une fois estimés les paramètres individuels, on s'intéresse à l'estimation des paramètres globaux (β et Γ) du modèle (3.1). Nous présentons deux méthodes différant par la prise en compte ou pas de l'incertitude dans l'estimation des paramètres individuels.

Méthode standard

La méthode standard (Standard Two Stage, STS, Sheiner et Beal [4]) suppose que les $\hat{\beta}_i$ sont issus d'une loi normale $\mathcal{N}(\beta, \Gamma)$. Par conséquent, l'estimation par maximum de vraisemblance de β et Γ fournit les valeurs

$$\begin{cases} \hat{\beta}_{STS} &= \frac{1}{I} \sum_{i=1}^I \hat{\beta}_i \\ \hat{\Gamma}_{STS} &= \frac{1}{I-1} \sum_{i=1}^I (\hat{\beta}_i - \hat{\beta}_{STS})(\hat{\beta}_i - \hat{\beta}_{STS})^T \end{cases} \quad (3.8)$$

qui sont la moyenne et la variance empiriques des $\hat{\beta}_i$.

En procédant de la sorte, on néglige la variance asymptotique Σ_i des estimateurs individuels. Ceci a pour conséquence de surestimer systématiquement la matrice Γ . On peut s'en rendre compte en exprimant

$$\hat{\beta}_i \simeq \beta_i + \hat{\varepsilon}_i \quad (3.9)$$

où \hat{e}_i est de moyenne nulle et de matrice de variance Σ_i . La relation $\text{var}(\hat{\beta}_i) = \Gamma + \Sigma_i$ permet alors d'exprimer le biais de $\hat{\Gamma}_{STS}$:

$$E(\hat{\Gamma}_{STS}) = \Gamma + \frac{1}{I} \sum_{i=1}^I \Sigma_i. \quad (3.10)$$

Dans ces conditions, on peut débiaiser l'estimateur $\hat{\Gamma}_{STS}$ en lui retranchant la variance d'estimateur moyenne des $\hat{\beta}_i$. On obtient ainsi l'estimateur RTS (« Refined Two Stage ») proposé par Steimer *et al.* [83] défini par

$$\hat{\Gamma}_{RTS} = \hat{\Gamma}_{STS} - \frac{1}{I} \sum_{i=1}^I \tilde{\Sigma}_i. \quad (3.11)$$

Le grand avantage de cette méthode réside dans sa simplicité de mise en œuvre. Elle est cependant dépendante du nombre d'observations effectuées sur chaque individu pour pouvoir disposer d'estimations individuelles fiables.

Méthode globale

La méthode globale proposée par Steimer *et al.* [83] suppose que les $\hat{\beta}_i$ sont distribués conditionnellement aux β_i selon une loi $\mathcal{N}(\beta_i, \Sigma_i)$. Par analogie avec la relation (3.9), on écrit ici

$$\hat{\beta}_i \simeq \beta_i + \hat{e}_i \quad (3.12)$$

avec $E(\hat{e}_i) = 0$ et $\text{var}(\hat{e}_i) = \Sigma_i$.

On approche ainsi la distribution des $\hat{\beta}_i$ par $\mathcal{N}(\beta, \Sigma_i + \Gamma)$. On peut alors estimer β et Γ par maximum de vraisemblance appliqué aux pseudo-données $\hat{\beta}_i$, en minimisant

$$L_{GTS}(\beta, \Gamma) = \sum_{i=1}^I \log |\Sigma_i + \Gamma| + \sum_{i=1}^I (\hat{\beta}_i - \beta)' (\Sigma_i + \Gamma)^{-1} (\hat{\beta}_i - \beta). \quad (3.13)$$

en utilisant par exemple un algorithme d'optimisation classique de type Newton-Raphson (*Cf.* Demailly [26]).

3.2.3 Vraisemblance directe

Cette section est consacrée à la mise en œuvre de l'algorithme EM pour l'estimation par maximum de vraisemblance dans le modèle (3.1). Nous présentons d'abord le principe général de l'algorithme EM et certaines de ses variantes utilisées pour l'application aux modèles non linéaires à paramètres aléatoires (Walker, [93]).

Principe de l'algorithme EM

L'algorithme EM (Expectation Maximization, Dempster *et al.* [29]) est une méthode itérative pour le calcul du maximum d'une densité dite *a posteriori*, comme c'est le cas dans le modèle à paramètres aléatoires. Au lieu de maximiser directement cette quantité,

généralement difficilement calculable et manipulable, la méthode consiste en une série de maximisations de fonctions plus simples. Pour cela, l'idée de base est de compléter les données observées par des données manquantes (non observées).

Pour simplifier les expressions, nous utilisons des notations indépendantes du reste du texte pour présenter le principe de cet algorithme. Les notations utilisées sont les suivantes :

- y désigne les observations,
- θ est le paramètre sur lequel on doit maximiser la fonction objectif,
- Θ est l'espace de définition de θ ,
- z représente les données manquantes *i.e.* des données qui ne sont pas observées,
- Z est l'espace de définition des z ,
- $f(y|\theta)$ est la densité des observations conditionnellement à θ ; on note $y \sim f(y|\theta)$,
- la densité des données manquantes conditionnellement aux observations et au paramètre est notée $k(z|y, \theta)$; on note $z|y \sim k(z|y, \theta)$.

Avec ces notations, la vraisemblance des données complètes qui est la quantité à maximiser s'écrit

$$L(\theta|y, z) = f(y|\theta).k(z|y, \theta). \quad (3.14)$$

Partant d'une valeur initiale $\theta^{(0)}$, l'étape $(r + 1)$ de l'algorithme se déroule en deux temps :

- E : calcul de $Q(\theta, \theta^{(r)}) = E_{\theta^{(r)}} \log L(\theta|y, z) = \int_Z \log L(\theta|y, z)k(z|y, \theta^{(r)})dz$
M : recherche de $\theta^{(r+1)} = \arg \max_{\theta \in \Theta} Q(\theta, \theta^{(r)})$

L'algorithme ainsi défini possède la propriété fondamentale suivante :

➤ *La vraisemblance des observations augmente à chaque itération de l'algorithme :*

$$\forall r \in \mathbb{N}, L(\theta^{(r+1)}|y) \geq L(\theta^r|y) \quad (3.15)$$

Les hypothèses de régularité définies par Wu [97] :

- $\theta \in \Theta \subseteq \mathbb{R}^D$,
- l'application $L : \theta \mapsto L(\theta|y)$ est continue sur Θ , bornée et dérivable sur l'intérieur de Θ ,
- l'application $Q : (\theta, \theta') \mapsto Q(\theta, \theta')$ définie précédemment dans l'étape E de l'algorithme, est continue en θ et θ' ,
- $\Theta_a = \theta \in \Theta | L(\theta) \geq L(a)$ est compact $\forall a \in \Theta$ tel que $L(a) > 0$,

assurent la convergence de l'algorithme vers l'estimateur du maximum de vraisemblance.

Variantes de l'algorithme EM

Dans la pratique, l'algorithme EM présente certaines limites. Cet algorithme est notamment fortement dépendant vis-à-vis de la position initiale. D'autre part, il est pénalisé

par une vitesse de convergence lente, ce qui implique la nécessité de pouvoir réaliser à moindre coût les étapes E et M pour être opérationnel.

Pour pallier ces problèmes, des extensions de l'algorithme EM ont été proposées dont une grande partie est présentée par McLachlan et Krishnan [62]. Nous citons ici brièvement celles mises en œuvre dans le cadre de l'application au modèle mixte non linéaire proposée par Walker [93].

Expectation Conditional Maximization

Cet algorithme (ECM) proposé par Meng et Rubin [63] remplace l'étape M (si celle-ci est relativement compliquée et devient très coûteuse en temps de calcul) par une série de maximisations conditionnelles (Conditional Maximization, CM) plus rapides car agissant sur des espaces de dimensions inférieures.

On procède à des maximisations conditionnelles en découpant le vecteur des paramètres en considérant par exemple $\theta = (\alpha, \beta)$. Le système

$$\begin{cases} \alpha^{(r+1)} = \arg \max_{\alpha} \varphi[(\alpha, \beta^{(r)}), \theta^{(r)}] \\ \beta^{(r+1)} = \arg \max_{\beta} \varphi[(\alpha^{(r+1)}, \beta), \theta^{(r)}] \end{cases} \quad (3.16)$$

illustre le découpage du vecteur des paramètres et les maximisations qui en découlent.

Cet algorithme conserve les propriétés d'augmentation de la vraisemblance des observations à chaque itération (relation (3.15)) et de convergence sous des conditions similaires à celles de l'algorithme EM.

Monte Carlo Expectation Maximization

On fait appel à cette variante de EM (MCEM, Wei et Tanner [94]) quand le calcul de l'espérance est trop laborieux. Elle est alors approchée par une méthode de type Monte Carlo. Dans ces conditions, l'étape E consiste à simuler N_{MC} valeurs de données manquantes selon la densité $k(z|y, \theta)$ évaluée en $\theta = \theta^{(r)}$:

$$\begin{cases} z^{(1)} & \sim k(z|y, \theta^{(r)}) \\ & \vdots \\ z^{(N_{MC})} & \sim k(z|y, \theta^{(r)}). \end{cases} \quad (3.17)$$

La densité est alors approchée par

$$\frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \log L(\theta|y, z^{(k)}). \quad (3.18)$$

En pratique, Wei et Tanner [94] recommande de faire varier la quantité de N_{MC} en augmentant sa valeur au cours du déroulement de l'algorithme au lieu de la fixer.

L'algorithme EM ainsi modifié perd la propriété de croissance de la vraisemblance à chaque itération (relation (3.15)) mais dispose encore de la convergence sous des conditions de régularité déterminées par Chan et Ledolter [12]. La condition principale réside dans

la convergence uniforme en θ sur des espaces compacts de la quantité à maximiser dans MCEM (3.18) vers la quantité équivalente pour EM (Cf. étape M décrite précédemment).

Application au modèle mixte non linéaire

L'application d'un algorithme de type EM au modèle mixte non linéaire proposée par Walker [93] est basée sur la spécification des paramètres aléatoires (B_i) comme données manquantes. Cette considération conduit à la manipulation de la (log-)vraisemblance des données complètes, observations Y_i et paramètres aléatoires B_i , où n'apparaît plus de problème d'intégration.

En effet, avec les notations $\theta = (\beta, \Gamma, \sigma_\varepsilon^2)$ et $Z = (y, B)$, la quantité $Q(\theta^{(r)}, \theta)$ à maximiser à l'étape (r) s'écrit

$$Q(\theta^{(r)}, \theta) = E[\log V(\theta^{(r)}|Z)|y, \theta] \quad (3.19)$$

où $\log V(\theta|Z)$ désigne la log-vraisemblance des données complètes définie à une constante additive près par

$$\frac{N}{2} \log(\sigma_\varepsilon^2) - \frac{I}{2} \log |\Gamma| + -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^I \left(\sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 \right) - \frac{1}{2} B_i^T \Gamma^{-1} B_i. \quad (3.20)$$

La prise en compte des B_i dans les données complétées évite de procéder à une intégration pour calculer la vraisemblance marginale.

La mise en œuvre de l'algorithme consiste à calculer $\tilde{\theta}$ tel que

$$\frac{\partial}{\partial \theta'} Q(\theta', \theta)|_{\theta'=\tilde{\theta}} = 0. \quad (3.21)$$

En supposant la dérivation sous le signe intégral [49], l'équation à résoudre est la suivante :

$$E \left[\frac{\partial}{\partial \theta'} \log V(\theta|Z)|y, \theta \right]_{\theta'=\tilde{\theta}} = 0. \quad (3.22)$$

Si on se place dans un modèle à paramètres aléatoires où le niveau 2 de l'écriture hiérarchique s'écrit simplement

$$\beta_i \sim \mathcal{N}_p(\beta, \Gamma) \quad (3.23)$$

alors, l'étape M de l'algorithme peut être traité analytiquement en dérivant l'expression (3.20) par rapport aux paramètres β , Γ et σ_ε^2 [93].

Si on traite le cas plus général permettant l'inclusion éventuelle de covariables (a_i) à travers une relation non linéaire d dans les variations des paramètres aléatoires,

$$\begin{cases} \beta_i &= d(a_i, \beta, B_i) \\ B_i &\sim \mathcal{N}_p(0, \Gamma) \end{cases} \quad (3.24)$$

l'étape M s'effectue alors pour chaque paramètre $(\beta, \Gamma, \sigma_\varepsilon^2)$ conditionnellement aux deux autres selon l'algorithme ECM.

Dans les deux cas, l'étape E nécessite l'évaluation de quantités à base d'espérances mathématiques qui ne disposent pas d'expression analytique. Pour estimer ces quantités, Walker [93] suggère alors de procéder par un échantillonnage de type Monte Carlo (*Cf.* la méthode MCEM présentée précédemment).

Les propriétés de l'algorithme EM assurent la convergence vers le maximum de vraisemblance.

3.2.4 Vraisemblance directe approchée

Nous présentons dans cette partie les méthodes d'estimation qui consistent à maximiser une approximation de la vraisemblance. Nous évoquons successivement l'approximation laplacienne, la mise en œuvre de techniques de Monte Carlo et l'application de formules de quadrature gaussienne. Pour ces trois méthodes, nous présentons d'abord le principe de l'approximation sur un exemple simple puis nous détaillons son application à l'approximation de la vraisemblance (3.3) dans le modèle non linéaire à paramètres aléatoires.

3.2.4.1 Approximation laplacienne

L'approximation laplacienne a été proposée par Tierney et Kadane [84] et par Leonard *et al.* [52] pour l'approximation de densités marginales.

Son utilisation dans le cadre des modèles non linéaires mixtes (Pinheiro et Bates [66] et Wolfinger [95]) permet de simplifier le calcul de l'intégrale nécessaire à l'expression de la densité des observations.

Principe de l'approximation laplacienne

Soit à calculer l'intégrale

$$\int e^{-\psi(\theta)} d\theta, \quad (3.25)$$

avec $\theta \in \mathbb{R}^p$ et ψ une fonction de \mathbb{R}^p dans \mathbb{R} admettant un maximum unique en $\theta = \hat{\theta}$.

Le principe de l'approximation laplacienne exprimé par la relation (3.26) consiste à faire apparaître dans l'expression de l'intégrale la densité de la loi normale :

$$\int e^{-\psi(\theta)} d\theta \simeq e^{-\psi(\hat{\theta})} (2\pi)^{\frac{p}{2}} |\psi''(\hat{\theta})|^{-\frac{1}{2}}. \quad (3.26)$$

avec $\psi'' = \frac{\partial^2 \psi}{\partial \theta^T \partial \theta}$ la matrice hessienne de ψ .

L'approximation (3.26) repose sur le développement de Taylor de ψ au voisinage de $\hat{\theta}$ à l'ordre 2 ; le terme de premier ordre étant nul, par définition de $\hat{\theta}$. Ensuite, il s'agit d'une simple réécriture destinée à faire apparaître la densité d'une loi normale de moyenne $\hat{\theta}$ et de matrice de variance $(\psi'')^{-1}$.

Application au calcul de la densité

La vraisemblance des paramètres du modèle à paramètres aléatoires (3.3) peut s'écrire sous la forme

$$V(\beta, \theta | Y) = C(\sigma_\varepsilon^2, \Gamma) \prod_{i=1}^I \left\{ \int \exp[-F(\beta, \theta, Y_i, B_i)] dB_i \right\} \quad (3.27)$$

avec

$$F(\beta, \theta, Y_i, B_i) = \left(\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 + \frac{1}{2} B_i^T \Gamma^{-1} B_i \right) \quad (3.28)$$

et $C(\sigma_\varepsilon^2, \Gamma) = (2\pi)^{-\frac{I(J+p)}{2}} \sigma_\varepsilon^{2(-\frac{IJ}{2})} |\Gamma|^{-\frac{I}{2}}$, indépendant de B_i . Comme précédemment, nous notons $\theta = (\beta, \Gamma, \sigma_\varepsilon^2)$ le vecteur des paramètres à estimer.

Avec \hat{B}_i défini par $\arg \min_\eta F(\beta, \theta, Y_i, \eta)$, une approximation de la vraisemblance des paramètres est donnée par

$$V_{lap}(Y|\theta) = C(\sigma_\varepsilon^2, \Gamma) \prod_{i=1}^I \left\{ (2\pi)^{\frac{p}{2}} \times \exp[F(\beta, \theta, Y_i, \hat{B}_i)] \times |F''(\hat{B}_i)|^{-\frac{1}{2}} \right\} \quad (3.29)$$

où $F''(\hat{B}_i) = \frac{\partial^2 F(\beta, \theta, Y_i, B_i)}{\partial B_i \partial B_i^T} \Big|_{B_i=\hat{B}_i}$ est la hessienne de la fonction $F(\beta, \theta, Y_i, B_i)$ évaluée en $B_i = \hat{B}_i$

La hessienne de $F(\beta, \theta, Y_i, B_i)$ qui s'exprime

$$F''(\hat{B}_i) = \overbrace{-\frac{\partial^2 f_i(t_j, \beta + B_i)}{\partial B_i \partial B_i^T} \Big|_{B_i=\hat{B}_i} [Y_i - f_i(t_j, \beta + B_i)]}^{(*)} + \overbrace{\frac{\partial f_i(t_j, \beta + B_i)}{\partial B_i} \Big|_{B_i=\hat{B}_i} \times \frac{\partial f_i(t_j, \beta + B_i)}{\partial B_i^T} \Big|_{B_i=\hat{B}_i} + \Gamma^{-1}}^{(**)} \quad (3.30)$$

est alors approchée par $(**)$ car la contribution de $(*)$ est généralement négligeable par rapport à $(**)$.

Dans ces conditions, l'approximation laplacienne de la log-vraisemblance s'écrit

$$\begin{aligned} l_{lap}(\theta | Y) &= \log C(\sigma_\varepsilon^2, \Gamma) + \frac{Ip}{2} \log(2\pi) + \sum_{i=1}^I F(\beta, \theta, Y_i, \hat{B}_i) - \frac{1}{2} \sum_{i=1}^I \log |F''_{**}(\hat{B}_i)| \\ &= -\frac{IJ}{2} \log(2\pi \sigma_\varepsilon^2) - \frac{I}{2} \log |\Gamma| + \sum_{i=1}^I F(\beta, \theta, Y_i, \hat{B}_i) - \frac{1}{2} \sum_{i=1}^I \log |F''_{**}(\hat{B}_i)| \end{aligned} \quad (3.31)$$

où $F''_{**}(\hat{B}_i)$ désigne le terme $(**)$ de l'équation (3.30).

L'estimation des paramètres θ du modèle se fait alors par maximisation de l'expression (3.31) qui ne présente plus de calcul d'intégrales.

3.2.4.2 Monte Carlo par échantillonnage d'importance

L'échantillonnage d'importance (importance sampling), ou échantillonnage pondéré, est une méthode de réduction de variance dans le cadre des méthodes de Monte Carlo. Le principe de cet échantillonnage est détaillé dans une première partie, puis nous présentons son application au modèle mixte non linéaire proposée par Pinheiro et Bates [66].

Principe de l'échantillonnage d'importance

Le principe générique des méthodes de Monte Carlo consiste à écrire l'intégrale à calculer sous la forme d'une espérance mathématique que l'on peut approcher par la moyenne empirique grâce à la loi des grands nombres. Les calculs d'un algorithme de Monte Carlo peuvent être menés à partir d'un ensemble de points générés aléatoirement selon une loi de probabilité quelconque, généralement uniforme. Cependant, le choix de cette distribution influe sur l'efficacité de la méthode. Au lieu de générer les points selon une loi uniforme, l'échantillonnage d'importance concentre les points à des endroits bien spécifiques.

Si ψ est la fonction à intégrer et φ une densité de probabilité, l'intégrale de ψ peut s'écrire

$$S = \int \psi(x)dx = \int \frac{\psi(x)}{\varphi(x)}\varphi(x)dx = E \left[\frac{\psi(X)}{\varphi(X)} \right]. \quad (3.32)$$

L'intégrale S peut ainsi être approchée en simulant des $(x_k)_{k=1, \dots, N_{MC}}$ selon la densité de probabilité φ et en évaluant $\frac{\psi(\cdot)}{\varphi(\cdot)}$ en ces points.

Par la loi des grands nombres, on sait que l'on peut approcher S par une moyenne empirique (Cf. par exemple Davis et Rabinowitz [23] ou Lapeyre *et al.* [51]) :

$$\hat{S} = \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \frac{\psi(x_k)}{\varphi(x_k)} \xrightarrow[N_{MC} \rightarrow +\infty]{p.s.} S. \quad (3.33)$$

La variance de \hat{S} donnée par

$$\text{var}(\hat{S}) = \text{var} \left(\frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \frac{\psi(x_k)}{\varphi(x_k)} \right) = \frac{S^2}{N_{MC}} \int \left(\frac{\psi(x)}{S\varphi(x)} - 1 \right)^2 \varphi(x)dx \quad (3.34)$$

nous éclaire quant au choix de la distribution φ : il apparaît ainsi que plus les fonctions ψ et φ sont proches, plus le résultat est précis, le meilleur choix possible pour φ étant

$|\psi|$ normalisée pour en faire une densité. Autrement dit, cela consiste à sur-échantillonner dans les intervalles où la fonction à intégrer est grande.

Application au modèle mixte non linéaire

L'application de l'échantillonnage d'importance pour les modèles à paramètres aléatoires proposée par Pinheiro et Bates [66], est inspirée de Geweke [40].

La vraisemblance (3.27) est approchée grâce à l'approximation de l'intégrale

$$\int \exp[-F(\beta, \theta, Y_i, B_i)] dB_i$$

par une méthode de Monte Carlo par échantillonnage d'importance avec F définie par (3.28).

Le développement de Taylor de la fonction $F(\theta, Y_i, B_i)$ autour de \hat{B}_i qui maximise F

$$F(\theta, Y_i, B_i) \simeq F(\theta, Y_i, \hat{B}_i) + \frac{1}{2}(B_i - \hat{B}_i)^T F''(\hat{B}_i)(B_i - \hat{B}_i) \quad (3.35)$$

met en évidence que la fonction à intégrer peut être approchée à une constante près par la densité de la loi normale $\mathcal{N}(\hat{B}_i, [F''(\hat{B}_i)]^{-1})$. La mise en œuvre de l'échantillonnage d'importance nécessite de pouvoir simuler selon cette loi. En pratique, on génère aléatoirement des vecteurs z^* selon une $\mathcal{N}(0, \text{Id}_p)$ et on calcule des B_i^* par la relation $B_i^* = \hat{B}_i + [F''(\hat{B}_i)]^{-1/2} z^*$.

En procédant au changement de variable $B_i = \hat{B}_i + [F''(\hat{B}_i)]^{-1/2} z$, l'intégrale à calculer s'écrit

$$\int \exp \left[-F(\theta, Y_i, \hat{B}_i + [F''(\hat{B}_i)]^{-1/2} z) \right] |F''(\hat{B}_i)|^{-\frac{1}{2}} dz \quad (3.36)$$

Dans ces conditions, on applique la méthode de Monte Carlo avec

- $\psi(x) = \exp \left[-F(\theta, Y_i, \hat{B}_i + [F''(\hat{B}_i)]^{-1/2} x) \right]$ et
- $\varphi(x) = (2\pi)^{-\frac{p}{2}} \exp \left[-\frac{1}{2} x^T x \right]$ i.e. la densité de la loi $\mathcal{N}(0, \text{Id}_p)$.

On approche alors l'intégrale (3.36) par

$$[F''(\hat{B}_i)]^{-\frac{1}{2}} \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \frac{\exp \left[-F(\theta, Y_i, \hat{B}_i + [F''(\hat{B}_i)]^{-1/2} z_k^*) \right]}{(2\pi)^{-\frac{p}{2}} \exp \left[-\frac{1}{2} z_k^{*T} z_k^* \right]} \quad (3.37)$$

où les z_k^* sont générés selon une $\mathcal{N}(0, \text{Id}_p)$.

La log-vraisemblance approchée par Monte Carlo s'exprime alors

$$l_{MC}(\theta|Y) = -\frac{IJ}{2} \log(2\pi \sigma_\varepsilon^2) - \frac{1}{2} \sum_{i=1}^I \text{Log} |F''(\hat{B}_i)| + \sum_{i=1}^I \log \left\{ \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} \exp \left[F(\theta, Y_i, B_i^*) + \frac{1}{2} z_k^{*T} z_k^* \right] \right\} \quad (3.38)$$

La maximisation en θ de l'expression (3.38) conduit à l'estimation des paramètres $\theta = (\beta, \Gamma, \sigma_\varepsilon^2)$.

3.2.4.3 Quadrature gaussienne

Les méthodes de quadrature sont des techniques relativement anciennes utilisées pour approcher les intégrales n'admettant pas de résolution analytique (*Cf.* par exemple Rappaz et Picasso [70] pour une introduction sur ces méthodes). Nous rappelons leur principe avant de montrer leur application au cas de la vraisemblance dans le cadre des modèles non linéaires à paramètres aléatoires proposée par Pinheiro et Bates [66].

Principe des formules de quadrature

Si ψ est une fonction continue sur l'intervalle $[a, b]$ et si w est une fonction de poids vérifiant $\int_a^b w(x)dx = 1$, la formule de quadrature

$$\int_a^b w(x)\psi(x)dx = \sum_{k=1}^{N_{QG}} \omega_k \psi(x_k) \quad (3.39)$$

est définie par la donnée de N_{QG} points $a \leq x_1 < x_2 < \dots < x_{N_{QG}} \leq b$ appelés points ou abscisses d'intégration et de N_{QG} réels $\omega_1, \omega_2, \dots, \omega_{N_{QG}}$ appelés poids de la formule de quadrature. Les abscisses et les poids de la quadrature sont définis de telle sorte que la formule de quadrature soit exacte pour des fonctions f polynomiales de plus haut degré possible. En particulier, les formules de quadrature à N_{QG} points dites gaussiennes optimisent le choix des abscisses et sont exactes pour des polynômes de degré inférieur ou égal à $2N_{QG} - 1$.

L'approximation d'intégrales définies sur un intervalle infini est également possible par ce type de méthodes (*Cf.* Davis et Rabinowitz [23], chapitre 3). Les abscisses et les poids de la quadrature sont alors déterminés à partir des polynômes de Laguerre, pour des intervalles du type $(a; \infty)$ ou $(-\infty; a)$, ou des polynômes d'Hermite pour $(-\infty; \infty)$.

Application à l'approximation de la densité

La vraisemblance exprimée en (3.3) peut s'écrire

$$V(\theta|Y) = (2\pi\sigma_\varepsilon^2)^{-\frac{IJ}{2}} \prod_{i=1}^I \underbrace{\left\{ \int (2\pi)^{-\frac{p}{2}} |\Gamma|^{-\frac{1}{2}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 - \frac{1}{2} B_i^T \Gamma^{-1} B_i \right) dB_i \right\}}_{(\blacktriangle)} \quad (3.40)$$

En effectuant le changement de variable $B_i = \Gamma^{\frac{1}{2}} z$, l'intégrale (\blacktriangle) à calculer dans l'expression (3.40) se met sous la forme

$$\int w(z)\psi(z)dz \quad (3.41)$$

avec

- $w(z) = (2\pi)^{-\frac{p}{2}} \exp(-\frac{1}{2}z^T z)$ la fonction de poids (densité d'une $\mathcal{N}(0, \text{Id}_p)$) et

$$\bullet \psi(z) = \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[Y_{ij} - f(t_j, \beta + \Gamma^{\frac{1}{2}} z) \right]^2 \right).$$

Dans ces conditions, en appliquant la formule de quadrature, l'intégrale (3.41) est approchée dans le cas unidimensionnel ($p = 1$) par

$$\sum_{k=1}^{N_{QG}} \omega_k \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[Y_{ij} - f(t_j, \beta + \Gamma^{\frac{1}{2}} z_k) \right]^2 \right) \quad (3.42)$$

où z_k et ω_k représentent respectivement les abscisses et les poids de la formule de quadrature définis à partir des polynômes de Hermite $\mathcal{H}_n(z) = (-1)^n e^{z^2} \frac{d^n}{dz^n} (e^{-z^2})$ par :

- les z_k sont les zéros de $\mathcal{H}_n(z)$,
- $\omega_k = \sqrt{\pi} \frac{2^{n+1} n!}{\mathcal{H}_{n+1}(z_k)}$.

Pour une quadrature à N_{QG} points, on utilise le polynôme de Hermite de degré N_{QG} .

Dans le cas multi-dimensionnel ($p > 1$), la structure du modèle à paramètres aléatoires permet de procéder à des applications successives des quadratures en dimension 1 (Pinheiro et Bates [66]). On approche alors l'intégrale multiple (3.41) par

$$\sum_{k_1=1}^{N_{QG}} \cdots \sum_{k_p=1}^{N_{QG}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[Y_{ij} - f(t_j, \beta + \Gamma^{\frac{1}{2}} z_{k_1, \dots, k_p}) \right]^2 \right) \prod_{l=1}^p \omega_{k_l} \quad (3.43)$$

avec $z_{k_1, \dots, k_p} = (z_{k_1}, \dots, z_{k_p})^T$ et ω_{k_l} les poids de la quadrature appliqué à l'intégrale sur le $l^{\text{ème}}$ élément de B_i .

En remplaçant l'intégrale par son approximation (3.43) dans l'expression de la log-vraisemblance, on obtient :

$$l_{QG}(\theta|Y) = -\frac{IJ}{2} \log(2\pi\sigma_\varepsilon^2) + \sum_{i=1}^I \left\{ \sum_{k_1=1}^{N_{QG}} \cdots \sum_{k_p=1}^{N_{QG}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[Y_{ij} - f(t_j, \beta + \Gamma^{\frac{1}{2}} z_{k_1, \dots, k_p}) \right]^2 \right) \prod_{l=1}^p \omega_{j_l} \right\} \quad (3.44)$$

On obtient les estimateurs des paramètres du modèle en maximisant l'expression (3.44) en β , Γ et σ_ε^2 .

Quadrature gaussienne adaptative

Pinheiro et Bates [66] propose une amélioration de l'approximation par quadrature en s'inspirant de l'échantillonnage d'importance. En effet, la quadrature gaussienne peut être interprétée comme une version déterministe des méthodes de Monte Carlo dans la mesure où les abscisses sont pré-définies et non générées aléatoirement. Pour cela, ils suggèrent de centrer les abscisses autour de \hat{B}_i et non plus de 0 comme précédemment dans la relation (3.42). Par ailleurs, le facteur d'échelle Γ est remplacé par $F''(\hat{B}_i)$.

Avec ces changements, la log-vraisemblance est approchée par

$$\begin{aligned}
l_{AQG}(\theta|Y) &= -\frac{IJ}{2} \log(2\pi\sigma_\varepsilon^2) - \frac{I}{2} \log |\Gamma| - \frac{1}{2} \sum_{i=1}^I \log |F''(\hat{B}_i)| \\
&+ \sum_{i=1}^I \left\{ \sum_{k_1=1}^{N_{QG}} \cdots \sum_{k_p=1}^{N_{QG}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[Y_{ij} - f(t_j, \beta + \hat{B}_i + [F''(\hat{B}_i)]^{\frac{1}{2}} z_{k_1, \dots, k_p}) \right]^2 \right) \prod_{l=1}^p \omega_{j_l} \right\}
\end{aligned} \tag{3.45}$$

3.2.5 Vraisemblance indirecte

Les deux méthodes qui consistent à modifier le modèle initial en un modèle où la vraisemblance est plus simple à maximiser sont présentées ici : il s'agit de la méthode de pseudo-vraisemblance simulée proposée par Nuñez et Concordet [64, 15] et de la populaire méthode de linéarisation (Lindstrom et Bates [54]).

3.2.5.1 Pseudo-vraisemblance simulée

L'idée de base de la pseudo-vraisemblance simulée est de considérer les Y_i issus d'une loi normale de moyenne $\mu_i(\theta)$ et de variance $V_i(\theta)$ définis par

$$\begin{cases} \mu_i(\theta) &= E[f(t, \beta_i)] \\ V_i(\theta) &= \text{var}[f(t, \beta_i)] + \sigma_\varepsilon^2 \text{Id}_J \end{cases} \tag{3.46}$$

avec les notations du modèle défini en début de chapitre par le système (3.1) et où θ représente le vecteur des paramètres à estimer $(\beta, \Gamma, \sigma_\varepsilon^2)$.

De cette façon, la fonction objectif à minimiser peut s'écrire

$$\bar{C}_I(\theta) = \frac{1}{I} \sum_{i=1}^I [Y_i - \mu_i(\theta)]^T V_i^{-1} [Y_i - \mu_i(\theta)] + \log |V_i(\theta)|. \tag{3.47}$$

En général, $\mu_i(\theta)$ et $V_i(\theta)$ n'ont pas de forme explicite. Pour évaluer la fonction objectif (3.47), Nuñez et Concordet [64, 15] proposent de procéder à une approximation de type Monte Carlo. Pour cela, $\mu_i(\theta)$ et $V_i(\theta)$ sont approchés par simulation selon les relations

$$\begin{cases} \bar{\mu}_{i, N_{MC}}(\theta) &= \frac{1}{N_{MC}} \sum_{n=1}^{N_{MC}} f(t_j, \beta_i^n) \\ V_{i, N_{MC}}(\theta) &= \frac{1}{N_{MC} - 1} \sum_{i=n}^{N_{MC}} [f(t_j, \beta_i^n) - \bar{\mu}_{i, N_{MC}}(\theta)] [f(\beta_i^n) - \bar{\mu}_{i, N_{MC}}(\theta)]' \end{cases} \tag{3.48}$$

où les β_i^n sont définis par

$$\beta_i^n = \beta + \Gamma^{\frac{1}{2}} z_i^n \quad \text{avec } z_i^n \sim \mathcal{N}(0, \text{Id}_p). \tag{3.49}$$

La fonction objectif (3.47) est alors approchée par

$$\bar{C}_I(\theta) \simeq \bar{C}_I^{NMC}(\theta) = \frac{1}{I} \sum_{i=1}^I [Y_i - \bar{\mu}_{i,NMC}(\theta)]^T V_{i,NMC}^{-1} [Y_i - \bar{\mu}_{i,NMC}(\theta)] + \log |V_{i,NMC}(\theta)|. \quad (3.50)$$

L'estimation de θ est donc définie par

$$\hat{\theta}_{PVS} = \arg \max_{\theta} \bar{C}_I^{NMC}(\theta).$$

3.2.5.2 Linéarisation

Le principe des méthodes de linéarisation est l'approximation du modèle non linéaire par un modèle linéaire. Cette stratégie est motivée par la possibilité d'exploiter les techniques existantes pour les modèles linéaires (*Cf.* Searle *et al.* [77] pour une revue sur les modèles linéaires mixtes). On bénéficie également du fait que dans un modèle linéaire hiérarchique, sous des hypothèses d'indépendance entre les paramètres aléatoires B_i et l'erreur ε_i , la distribution marginale des Y_i s'écrit explicitement. De plus, si l'on suppose que les distributions des B_i et des ε_{ij} sont normales alors la distribution de Y_i est également normale (*Cf.* par exemple Graybill [42]).

Deux méthodes largement répandues sont basées sur ce principe. La linéarisation est effectuée dans les deux cas par un développement de Taylor au premier ordre. La différence entre les deux méthodes réside dans le choix du point autour duquel on procède au développement :

- Beal et Sheiner [5] propose de linéariser autour de l'espérance des B_i *i.e.* zéro dans le modèle que nous avons décrit. Cette méthode est revue en détail par Davidian et Giltinan [22].
- Lindstrom et Bates [54] effectue le même développement autour d'une prédiction \hat{B}_i des B_i mise à jour à chaque étape d'un algorithme itératif.

La question du choix entre ces deux méthodes n'est pas uniformément résolue. Davidian et Giltinan ([22], section 6.7) soulignent que l'on peut s'attendre à de meilleurs résultats par la linéarisation autour des \hat{B}_i , mais que ce gain ne vaut pas toujours la charge de calculs supplémentaire par rapport au développement autour de 0. Par ailleurs, ils précisent, d'après Beal et Sheiner [6], que l'écart entre les deux méthodes diminue quand le nombre d'observations par individu (J) diminue. En effet, les \hat{B}_i sont d'autant plus resserrés autour de leur moyenne (0) que J est petit. Dans ces conditions, il nous semble préférable d'utiliser la méthode de Lindstrom et Bates [54] car nous disposons *a priori* d'un nombre d'observations par pixel important par rapport aux problèmes de pharmacologie traités habituellement par des modèles non linéaires à paramètres aléatoires.

La procédure d'estimation proposée par Lindstrom et Bates [54] consiste à alterner deux étapes :

- calcul des \hat{B}_i et des $\hat{\beta}$. Cette étape est présentée par la suite dans le paragraphe « Prédiction des effets individuels ».

- estimation des paramètres de variance par maximisation de l'approximation de la densité marginale des observations. Cette seconde étape fournit une nouvelle estimation pour les β . Elle est détaillée dans le paragraphe « Estimation des paramètres du modèle ».

Dans ce contexte, les estimateurs des paramètres de variance sont des estimateurs de pseudo-vraisemblance puisqu'ils maximisent la vraisemblance évaluée en une estimation préliminaire de β .

Prédiction des effets individuels

Les prévisions des B_i et de β sont obtenus en maximisant sur les B_i et β la fonction exprimant la densité *a posteriori* des observations

$$l(\beta, B_i | Y, \hat{\Gamma}, \hat{\sigma}_\varepsilon^2) = \sum_{i=1}^I \left[-\frac{1}{2\hat{\sigma}_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \beta + B_i)]^2 - \frac{1}{2} B_i^T \hat{\Gamma}^{-1} B_i \right] \quad (3.51)$$

une fois estimée Γ et σ_ε^2 .

Estimation des paramètres du modèle

L'approximation réside dans la linéarisation des résidus par un développement de Taylor autour des \hat{B}_i et des $\hat{\beta}$:

$$Y_{ij} - f(t_j, \beta + B_i) \simeq Y_{ij} - \left[f(t_j, \hat{\beta} + \hat{B}_i) + \hat{Z}(B_i - \hat{B}_i) + \hat{X}(\beta - \hat{\beta}) \right] \quad (3.52)$$

où $\hat{Z} = \frac{\partial f(t_j, \beta + B_i)}{\partial B_i^T} \Big|_{\hat{B}_i, \hat{\beta}}$ et $\hat{X} = \frac{\partial f(t_j, \beta + B_i)}{\partial \beta^T} \Big|_{\hat{B}_i, \hat{\beta}}$.

La notation vectorielle équivalente s'écrit

$$Y_i - f_i(t, \beta + B_i) \simeq Y_i - \left[f_i(t, \hat{\beta} + \hat{B}_i) + \hat{Z}_i(B_i - \hat{B}_i) + \hat{X}_i(\beta - \hat{\beta}) \right]$$

avec $f_i(t, \beta + B_i) = (f(t_1, \beta + B_i), f(t_2, \beta + B_i), \dots, f(t_J, \beta + B_i))^T$

Dans ces conditions, la distribution conditionnelle des $Y_i | B_i$ est alors approchée par une loi normale de moyenne linéaire en B_i :

$$Y_i | B_i \sim \mathcal{N} \left(f_i(t, \hat{\beta}, \hat{B}_i) + \hat{Z}_i(B_i - \hat{B}_i) + \hat{X}_i(\beta - \hat{\beta}), \sigma_\varepsilon^2 \text{Id}_J \right)$$

où \sim désigne une distribution approchée.

Cette expression ainsi que la distribution des B_i ($B_i \sim \mathcal{N}_p(0, \Gamma)$), Cf. équation (3.1)) permet d'approcher la distribution marginale des Y_{ij} par la loi normale de moyenne $f_i(t, \hat{\beta}, \hat{B}_i) - \hat{Z}_i \hat{B}_i - \hat{X}_i \hat{\beta} + \hat{X}_i \beta$ et de variance $\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i$.

On peut alors écrire

$$W_i \sim \mathcal{N} \left(\hat{X}_i \beta, \sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i \right). \quad (3.53)$$

avec $W_i = Y_i - f_i(t, \hat{\beta}, \hat{B}_i) + \hat{Z}_i \hat{B}_i + \hat{X}_i \hat{\beta}$

La vraisemblance des paramètres correspondant à la distribution approchée des W_i s'écrit donc

$$V_{lin}(\theta|W) = (2\pi)^{-\frac{IJ}{2}} \prod_{i=1}^I |\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i|^{\frac{1}{2}} \times \prod_{i=1}^I \left\{ \exp \left[-\frac{1}{2} (W_i - \hat{X}_i \beta)^T (\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i) (W_i - \hat{X}_i \beta) \right] \right\} \quad (3.54)$$

Par passage au log, l'expression à maximiser pour estimer $\beta, \sigma_\varepsilon^2$ et Γ peut s'écrire

$$l_{in}(\theta|W) = -\frac{IJ}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^I \left\{ \log |\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i| + (W_i - \hat{X}_i \beta)^T (\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i^T \Gamma \hat{Z}_i) (W_i - \hat{X}_i \beta) \right\} \quad (3.55)$$

La maximisation de cette quantité est alors effectuée par exemple par un algorithme de type Newton-Raphson (*Cf.* [26]) pour obtenir l'estimation des paramètres de variance et la valeur mise à jour de $\hat{\beta}$.

Maximum de vraisemblance restreint

Le passage dans le cadre d'un modèle linéaire permet également de procéder à une estimation par maximum de vraisemblance restreint (Restricted Maximum Likelihood, REML), notamment pour tenir compte dans l'estimation des paramètres de variance de la perte de degrés de liberté lors de l'estimation préalable des paramètres de moyenne.

La procédure d'estimation proposée par Lindstrom et Bates [54] est analogue à la précédente. La seule différence réside dans l'expression de la (log-)vraisemblance à maximiser. Pour estimer les paramètres de variance par REML, la fonction à maximiser s'écrit :

$$l_{in_{REML}}(\theta|W) = l_{in}(\theta|W) - \frac{1}{2} \sum_{i=1}^I \log \left| \hat{X}_i (\sigma_\varepsilon^2 \text{Id}_J + \hat{Z}_i \Gamma \hat{Z}_i^T)^{-1} \hat{X}_i \right|. \quad (3.56)$$

Dans notre cas, nous avons constaté très peu de différences entre le maximum de vraisemblance et le maximum de vraisemblance restreint pour l'algorithme de linéarisation sur quelques simulations. Nous conserverons l'estimation par maximum de vraisemblance.

Par ailleurs, Wolfinger [95] démontre que l'algorithme itératif de Lindstrom et Bates pour le maximum de vraisemblance restreint issu de la linéarisation peut être dérivé de l'approximation laplacienne de la vraisemblance du modèle non linéaire à paramètres aléatoires en supposant comme dans notre cas la normalité des paramètres aléatoires et de l'erreur.

3.2.6 Implémentation des méthodes

Les méthodes basées sur les régressions individuelles ont l'avantage d'une mise en œuvre quasi-immédiate. Il suffit de disposer d'une routine d'estimation dans un modèle de régression non linéaire classique pour la première phase de la méthode. Le calcul des paramètres globaux est ensuite immédiat pour la méthode standard. Pour la méthode globale, l'implémentation d'un algorithme de minimisation fournit les estimations des paramètres globaux. Cette simplicité de mise en œuvre et le moindre coût opérationnel de ce type de méthodes nous conduit à retenir la méthode standard pour comparaison avec la méthode de linéarisation, sur données simulées (*Cf.* chapitre 4) et réelles (*Cf.* chapitre 5).

Parmi les méthodes basées sur le critère du maximum de vraisemblance, la méthode de linéarisation est certainement la plus populaire grâce à son implémentation entretenue et largement distribuée par les auteurs. Ainsi, on retrouve l'algorithme d'estimation par linéarisation proposé par Lindstrom et Bates [54] dans la librairie `nlme` [67] disponible pour Splus [76] et R (*Cf.* par exemple [100]). L'adaptation pour SAS sous forme d'une macro-commande est décrite par Littell *et al.* [55] et expérimentée par Wolfinger [96].

Les autres méthodes ne bénéficient pas d'un outil prêt à l'emploi et nécessitent une implication de l'utilisateur pour leur mise en œuvre. Ce point sera évoqué dans le chapitre 4 pour l'étude comparative sur données simulées des méthodes de linéarisation et par approximation numérique.

3.3 Propriétés des estimateurs

Cette section propose un retour sur la bibliographie autour des modèles non linéaires mixtes avec comme axe de lecture les propriétés des estimateurs obtenus selon les diverses méthodes. En effet, cet aspect des méthodes est généralement peu évoqué dans la littérature où l'accent est mis sur le comportement numérique. Nous complétons les quelques résultats parus dans la littérature par une réflexion sur la convergence des estimateurs obtenus par approximation numérique de la vraisemblance. Ces diverses sources d'information sont ensuite réunies pour proposer une synthèse sur le sujet.

3.3.1 Au sujet de l'asymptotique

Avant de rentrer dans le détail des propriétés des estimateurs obtenus par les diverses méthodes, nous précisons les diverses façons selon lesquelles peut être envisagé le comportement asymptotique.

En effet, dans le cadre de l'analyse de données longitudinales sur plusieurs individus, dans lequel le problème de suivi de cultures par télédétection spatiale s'insère logiquement, on dispose de deux axes pour l'asymptotique :

- le nombre d'individus *i.e.* dans notre cas, le nombre de pixels (I),
- le nombre de mesures par individus *i.e.* le nombre d'images disponibles pour la zone d'étude. Nous supposons que le nombre d'acquisitions (J) est identique pour chaque pixel. Ce point a été justifié au chapitre 1 par l'utilisation de synthèses décadales.

Dans ces conditions, plusieurs cas de figures peuvent être considérés :

1. L'asymptotique repose simultanément sur les deux axes : $I \rightarrow \infty$ et $J \rightarrow \infty$. On dispose d'images composées d'un nombre important de pixels acquises très souvent sur une longue période.
2. Seul le nombre de pixels supporte l'asymptotique : $I \rightarrow \infty$ et $J < \infty$. Les acquisitions étant effectuées sur une période pré-définie, leur nombre ne peut pas augmenter indéfiniment.
3. Le nombre d'acquisitions par individu tend vers l'infini, pour un nombre relativement faible d'individus : $J \rightarrow \infty$ et $I < \infty$. Les images considérées ici représentent une zone relativement petite quelle que soit la résolution ; elles sont acquises en grand nombre.

Le nombre de pixels d'une image pouvant largement dépasser le millier, nous pouvons supposer systématiquement que nous sommes dans les conditions de l'asymptotique sur I . Nous rejetons donc la troisième situation. En ce qui concerne le nombre de dates, la situation dépend du niveau auquel on place l'asymptotique. En disposant d'images à haute résolution temporelle et donc de plusieurs dizaines de dates durant le cycle de vie de la culture d'intérêt, nous pourrions supposer disposer également de l'asymptotique sur J . Nous envisagerons cependant les propriétés dans le deuxième cas où J est borné.

3.3.2 Revue bibliographique

Relativement peu d'articles sont destinés à l'étude des propriétés des estimateurs dans les modèles non linéaires à paramètres aléatoires. Les références [27, 93, 64, 90, 14, 69, 98] sont quasiment les seules sur le sujet. Nous proposons ici une revue de ces quelques articles en précisant les propriétés des estimateurs considérés parmi les suivantes :

- convergence : on précisera l'axe asymptotique nécessaire à la convergence et si possible la vitesse de convergence,
- normalité asymptotique,
- efficacité.

3.3.2.1 Vraisemblance directe

L'algorithme EM mis en œuvre dans le cadre des modèles non linéaires à paramètres aléatoires évoqué précédemment (*Cf.* section 3.2.3) fournit l'estimateur du maximum de vraisemblance. On dispose dans ce cas de la théorie du maximum de vraisemblance pour déterminer les propriétés de cet estimateur. On pourra se référer à [16, 18] pour y trouver les démonstrations de la convergence, de la normalité asymptotique et de l'efficacité de l'estimateur du maximum de vraisemblance. Ces propriétés sont asymptotiques sur le nombre d'individus ; le nombre d'observations par individu peut être uniformément borné.

3.3.2.2 Vraisemblance directe approchée

Parmi les méthodes de la classe « vraisemblance directe approchée », l'approximation laplacienne est la seule pour laquelle on a étudié les propriétés statistiques de l'estimateur.

Vonesh [90] étudie l'estimateur du maximum de l'approximation laplacienne de la vraisemblance et exprime sa vitesse de convergence. Nous présentons d'abord la démonstration proposée par Vonesh et la commentons par la suite.

Démonstration proposée par Vonesh

Les calculs sont basés sur les relations

$$V(Y_i) = V_{lap}(Y_i) + \mathcal{O}\left(\frac{1}{J}\right) \quad (3.57)$$

et

$$I^{-1}U(\hat{\theta}_{lap}) = I^{-1}U(\theta^*) + I^{-1}H(\theta^*)(\hat{\theta}_{lap} - \theta^*) + \mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta^*)^2 \quad (3.58)$$

exprimant respectivement la précision de l'approximation laplacienne et un développement de Taylor au premier ordre de la dérivée première de la vraisemblance avec les notations $U(\theta) = \frac{\partial l(\theta)}{\partial \theta}$ et $H(\theta) = \frac{\partial^2 l(\theta)}{\partial \theta^2}$.

Pour le développement de Taylor exprimé en (3.58), Vonesh suppose que $\hat{\theta}_{lap}$ est un point à l'intérieur d'un voisinage de la vraie valeur du paramètre θ^* .

De l'expression de l'approximation de la densité marginale pour l'individu i (3.57), on déduit l'expression de la log-vraisemblance du paramètre θ en supposant l'indépendance des individus :

$$l(\theta) = l_{lap}(\theta) + \mathcal{O}\left(\frac{1}{J}\right). \quad (3.59)$$

Ensuite, sous les conditions classiques de régularité, on a

$$I^{-1}H(\theta)(\hat{\theta}_{lap} - \theta) + \mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta)^2 = \mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta) \quad (3.60)$$

et par ailleurs pour les mêmes raisons

$$I^{-1}U(\theta^*) = \mathcal{O}_p(I^{-\frac{1}{2}}). \quad (3.61)$$

On dispose ainsi de deux expressions de $I^{-1}U(\hat{\theta}_{lap})$:

$$\begin{cases} I^{-1}U(\hat{\theta}_{lap}) &= \mathcal{O}_p(I^{-\frac{1}{2}}) + \mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta^*) \\ I^{-1}U(\hat{\theta}_{lap}) &= I^{-1}U_{lap}(\hat{\theta}_{lap}) + \mathcal{O}\left(\frac{1}{J}\right). \end{cases} \quad (3.62)$$

On déduit du système (3.62), l'expression de $\mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta^*)$:

$$\mathcal{O}_p(1)(\hat{\theta}_{lap} - \theta^*) = I^{-1}U_{lap}(\hat{\theta}_{lap}) + \mathcal{O}_p(I^{-\frac{1}{2}}) + \mathcal{O}\left(\frac{1}{J}\right) = \mathcal{O}_p(I^{-\frac{1}{2}}) + \mathcal{O}\left(\frac{1}{J}\right) \quad (3.63)$$

car $I^{-1}U_{lap}(\hat{\theta}_{lap}) = 0$ par définition de $\hat{\theta}_{lap}$ qui maximise l_{lap} .

La vitesse de convergence est alors exprimée en simplifiant par $\mathcal{O}_p(1)$:

$$\hat{\theta}_{lap} - \theta^* = \mathcal{O}_p \left(\max \left\{ \frac{1}{\sqrt{I}}, \frac{1}{J} \right\} \right). \quad (3.64)$$

Par la suite, Vonesh montre également l'équivalence entre $\hat{\theta}_{lap}$ et l'estimateur du maximum de vraisemblance ; d'où les propriétés de normalité asymptotique et d'efficacité de $\hat{\theta}_{lap}$.

Commentaires

Selon nous, des erreurs entachent cette démonstration. La première est qu'à aucun moment, la convergence de $\hat{\theta}_{lap}$ vers θ^* n'est démontrée explicitement. Il s'agit plutôt du calcul de la vitesse de convergence de l'estimateur une fois supposée la convergence. De plus, dans la relation (3.63), $(\hat{\theta}_{lap} - \theta^*)$ est exprimé en simplifiant par $\mathcal{O}_p(1)$. Or, La définition d'un $\mathcal{O}_p(1)$ comme un terme que l'on peut majorer par n'importe quelle constante, fait que cette simplification est erronée si le terme en question tend vers zéro.

3.3.2.3 Vraisemblance indirecte

Les estimateurs issus des deux méthodes qualifiées de « vraisemblance indirecte » ont des propriétés qui ont été démontrées dans la littérature. Nous évoquons successivement la pseudo-vraisemblance simulée et la linéarisation.

Pseudo-vraisemblance simulée

La procédure d'estimation proposée par Nuñez et Concordet [64, 15] fournit un estimateur ayant les propriétés de convergence presque sûre et de normalité asymptotique sur I . De plus, Nuñez donne la vitesse de convergence de cet estimateur en fonction du nombre d'individus I et de la taille N_{MC} de l'échantillonnage de Monte Carlo. La relation

$$\frac{\sqrt{I}}{N_{MC}} (\hat{\theta}_I^{N_{MC}} - \theta^*) \xrightarrow[N_{MC} \rightarrow \infty]{I \rightarrow \infty, p.s.} 0 \quad (3.65)$$

montre par ailleurs que la convergence est atteinte à J fixé. Autrement dit, il n'est pas nécessaire de disposer d'un grand nombre d'observations par individu.

La convergence est démontrée à partir du lemme exprimé dans la relation

$$\sup_{\theta \in \Theta} |\bar{C}_I^{N_{MC}}(\theta) - E_{\theta^*}[\bar{C}_I(\theta)]| \xrightarrow[N_{MC} \rightarrow \infty]{I \rightarrow \infty} 0 \quad (3.66)$$

qui assure que le critère à minimiser, dont la définition est rappelée dans l'équation (3.50), est un contraste *i.e.* une fonction des observations qui converge vers une fonction déterministe identifiant la vraie valeur du paramètre.

L'inégalité de Jensen

$$\varphi(E(X)) \leq E(\varphi(X)) \quad \text{si } \varphi \text{ est convexe,} \quad (3.67)$$

et l'inclusion

$$\left\{ \hat{\theta}_I^{NMC} \notin \mathcal{B} \right\} \subset \left\{ \inf_{\Theta \setminus \mathcal{B}} \bar{C}_I^{NMC}(\theta) \geq \bar{C}_I(\theta) \right\} \quad (3.68)$$

où \mathcal{B} désigne une boule ouverte de Θ centrée sur θ^* , permettent de conclure sur la convergence.

La normalité asymptotique résulte d'une application du théorème central limite. On pourra se référer à Nuñez [64] pour l'expression de la variance asymptotique.

Linéarisation

Malgré l'existence et la popularité de plusieurs méthodes de linéarisation, très peu de références existent sur l'étude des propriétés des estimateurs issus de telles techniques. Un article de Demidenko [27], rarement référencé dans la littérature est quasiment le seul avec Ramos et Pantula [69] à étudier les propriétés des estimateurs obtenus par la méthode de linéarisation selon Lindstrom et Bates [54]. Cette technique est pourtant une des plus couramment employée grâce à sa mise en œuvre facilement disponible. Les propriétés sont évoquées en distinguant les deux situations asymptotiques :

1. $I \rightarrow \infty$ et $J < \infty$,
2. $I \rightarrow \infty$ et $J \rightarrow \infty$.

Les propriétés des estimateurs issus de procédés de linéarisation sont parfois exprimés (*Cf.* Vonesh et Carter [91]) sous réserve que l'approximation linéaire est exacte ou autrement dit que le modèle est linéaire. Les modèles d'évolution de culture que nous envisageons étant non linéaires, nous ne ferons pas une telle hypothèse.

Première situation asymptotique

Ramos et Pantula [69] et Demidenko [27] utilisent la même technique pour montrer la non-convergence des estimateurs obtenus par linéarisation dans le cas où l'asymptotique porte seulement sur le nombre d'individus avec un nombre d'observations par individu constant et fini.

Les auteurs considèrent un modèle exponentiel simple à un paramètre

$$\begin{cases} y_{ij} &= e^{\beta_i} + \varepsilon_{ij} \\ \beta_i &= \beta + B_i \\ B_i &\sim \mathcal{N}_p(0, \gamma^2) \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{cases} \quad (3.69)$$

qui présente l'avantage de permettre le calcul analytique de certaines expressions.

Dans ces conditions, et en supposant connus les paramètres de variance γ^2 et σ_ε^2 , l'algorithme de Lindstrom et Bates se résume à la première étape d'estimation par moindres carrés pénalisés :

$$\hat{\beta}_{LB} = \arg \max_{\beta, B_i} \left\{ \frac{1}{I} \sum_{i=1}^N \left(\frac{1}{J} \sum_{j=1}^J (y_{ij} - e^{\beta+B_i})^2 + \frac{B_i^2}{\gamma^2} \right) \right\} \quad (3.70)$$

Les équations qui permettent d'estimer les effets aléatoires \hat{B}_i et $\hat{\beta}_{LB}$ s'écrivent :

$$\begin{cases} e^{2(\beta+B_i)} - e^{\beta+B_i} \bar{y}_i + \frac{1}{J\gamma^2} B_i = 0 & i = 1, \dots, I \\ \frac{1}{I} \sum_{i=1}^I (e^{2(\beta+B_i)} - e^{\beta+B_i} \bar{y}_i) = 0 \end{cases} \quad (3.71)$$

En sommant les I équations de la première ligne du système (3.71) et en intégrant dans la seconde ligne, il vient que $\hat{\beta}_{LB}$ satisfait la relation (3.72) :

$$\frac{1}{I} \sum_{i=1}^I \hat{B}(\beta, \bar{y}_i) = 0 \quad (3.72)$$

où $\hat{B}(\beta, \bar{y}_i)$ est solution de la première ligne de (3.71).

Pour calculer le biais de l'estimateur $\hat{\beta}_{LB}$, on s'appuie sur le résultat suivant (Cf. par exemple [46]) :

➤ Soient u_1, u_2, \dots, u_I des variables aléatoires i.i.d. dont la distribution dépend d'un paramètre β . On note β_0 la vraie valeur du paramètre. L'estimateur $\hat{\beta}_I$ est défini par l'équation

$$\frac{1}{I} \sum_{i=1}^I S(u_i, \beta) = 0. \quad (3.73)$$

Dans ces conditions,

$$\lim_{I \rightarrow \infty} \hat{\beta}_I = \beta_* \quad (3.74)$$

où β_* est solution de l'équation $\bar{S}(\beta, \beta_0) = E_{\beta_0} S(u_1, \beta) = 0$.

Grâce au résultat précédent, utilisé en considérant $\bar{S}(\beta, \beta_0) = E_{\beta_0} \hat{B}(\beta, \bar{y})$, on peut calculer la limite de $\hat{\beta}_{LB}$ quand $I \rightarrow \infty$ et ainsi exprimer son biais asymptotique.

Deuxième situation asymptotique

Dans la situation où l'on considère que I et J tendent tous les deux vers l'infini, Demidenko [27] démontre que l'estimateur $\hat{\beta}_{LB}$ est équivalent à l'estimateur du maximum de vraisemblance et possède donc les propriétés de convergence, de normalité asymptotique et d'efficacité.

La démonstration de ce résultat est basé sur le fait que l'estimateur $\hat{\beta}_{STS}$ évoqué en 3.2.2 est lui aussi équivalent à l'estimateur du maximum de vraisemblance. Le plan de la démonstration est le suivant :

1. $\hat{\beta}_{LB}$ et $\hat{\beta}_{STS}$ sont équivalents quand $J \rightarrow \infty$ si on suppose connus les paramètres de variance σ_ε^2 et Γ ,

2. $\hat{\beta}_{STS}$ est convergent presque sûrement et asymptotiquement normal,
3. les procédures d'estimation *STS* et *LB* produisent des estimateurs convergents de σ_ε^2 et Γ ,
4. $\hat{\beta}_{STS}$ est équivalent à l'estimateur du maximum de vraisemblance quand $I \rightarrow \infty$ et $J \rightarrow \infty$ en remplaçant les paramètres de variance par leurs estimateurs convergents.

L'idée de base pour montrer le premier point réside dans l'écriture d'un pseudo-estimateur $\hat{\beta}(\delta)$ défini par :

$$\hat{\beta}(\delta) = \arg \min_{\beta} \sum_{i=1}^I (||Y_i - f_i(\beta + B_i)||^2 + \delta B_i' \Gamma^{-1} B_i), \quad \delta > 0 \quad (3.75)$$

En faisant tendre δ vers zéro, en considérant $\delta = \frac{1}{J}$, on obtient l'équivalence entre $\hat{\beta}_{LB}$ et $\hat{\beta}_{STS}$. La suite réside essentiellement sur des résultats issus de la théorie sur la régression non linéaire et les modèles linéaires mixtes. On trouvera la démonstration détaillée dans Demidenko [28].

3.3.3 Contribution à l'étude des propriétés des estimateurs

Les méthodes d'estimation par approximation numérique de la vraisemblance sont généralement évoquées pour leur aptitude à fournir l'estimateur du maximum de vraisemblance exact (EMV) [14, 65].

Dans cette section, nous essayons de démontrer que ces estimateurs ont les mêmes propriétés que le maximum de vraisemblance en nous basant sur une démonstration des propriétés de l'EMV présentée par exemple dans Cox et Hinkley [16]. Nous ne pouvons cependant pas aboutir au résultat souhaité. L'absence de convergence uniforme des approximations numériques n'autorise pas une démonstration similaire à celle de l'EMV.

Les notations utilisées dans cette section sont les suivantes :

- $\theta = (\beta, \Gamma, \sigma_\varepsilon^2)$ représente l'élément générique,
- θ^* désigne la vraie valeur du paramètre,
- Θ est l'espace des paramètres, supposé compact.

Nous supposons que la vraie valeur du paramètre θ^* appartient à l'intérieur de l'espace Θ . Cette hypothèse évite de se retrouver sur la frontière de l'espace de définition du paramètre et remplace en quelque sorte l'hypothèse « Θ ouvert » parfois rencontrée lorsque que l'on s'intéresse aux composantes de la variance.

3.3.3.1 Convergence de l'estimateur du maximum de vraisemblance

Soit $\hat{\theta}_I^{MV} = \arg \max_{\theta \in \Theta} \frac{1}{I} \sum_{i=1}^I l(\theta; Y_i)$ l'estimateur du maximum de vraisemblance et soit \mathcal{B} une boule de rayon R centrée sur θ^* .

Si $\hat{\theta}_I^{MV}$ n'appartient pas à \mathcal{B} cela signifie que la vraisemblance atteint son maximum hors de \mathcal{B} ce qui s'écrit :

$$\left\{ \hat{\theta}_{MV} \notin \mathcal{B} \right\} \subset \left\{ \sup_{\Theta \setminus \mathcal{B}} \frac{1}{I} \sum_{i=1}^I l(\theta; Y_i) \geq \frac{1}{I} \sum_{i=1}^I l(\theta^*; Y_i) \right\}. \quad (3.76)$$

Or, la loi des grands nombres, la continuité de l et la compacité de Θ permettent d'écrire :

$$\sup_{\Theta \setminus \mathcal{B}} \frac{1}{I} \sum_{i=1}^I l(\theta; Y_i) - \frac{1}{I} \sum_{i=1}^I l(\theta^*; Y_i) \xrightarrow[I \rightarrow \infty]{p.s.} \sup_{\Theta \setminus \mathcal{B}} Q(\theta, \theta^*) \quad (3.77)$$

avec $Q(\theta, \theta^*) = E_{\theta^*} [l(\theta; Y_1) - l(\theta^*; Y_1)]$.

De plus, l'inégalité de Jensen (stricte avec $\theta \neq \theta^*$) implique que $Q(\theta, \theta^*)$ est strictement négatif :

$$Q(\theta, \theta^*) = E_{\theta^*} \left[\log \frac{V(\theta; Y_1)}{V(\theta^*; Y_1)} \right] < \log E_{\theta^*} \left[\frac{V(\theta; Y_1)}{V(\theta^*; Y_1)} \right] = 0 \quad (3.78)$$

avec $l(\theta; Y_1) = \log V(\theta; Y_1)$.

Les relations (3.77) et (3.78) impliquent alors que

$$P \left[\sup_{\Theta \setminus \mathcal{B}} \frac{1}{I} \sum_{i=1}^I l(\theta; Y_i) \geq \frac{1}{I} \sum_{i=1}^I l(\theta^*; Y_i) \right] \xrightarrow[I \rightarrow \infty]{p.s.} 0 \quad (3.79)$$

et par conséquent, d'après l'inclusion (3.76)

$$P \left[\hat{\theta}_I^{MV} \notin \mathcal{B} \right] \xrightarrow[I \rightarrow \infty]{p.s.} 0$$

ce qui nous permet de conclure à la convergence presque sûre de $\hat{\theta}_I^{MV}$ vers θ^* quand I tend vers l'infini.

3.3.3.2 Adaptation pour les méthodes d'estimation par approximation

Les méthodes d'intégration numérique s'appliquent sur la vraisemblance individuelle. Elles génèrent une erreur d'approximation ξ_i pour l'individu i . La vraisemblance approchée globale est donc liée à la vraisemblance par la relation (3.80) :

$$\frac{1}{I} \sum_{i=1}^I l_{app}(\theta; Y_i) = \frac{1}{I} \sum_{i=1}^I l(\theta; Y_i) + \frac{1}{I} \sum_{i=1}^I \xi_i. \quad (3.80)$$

Pour démontrer les propriétés des estimateurs du maximum d'une approximation de la vraisemblance de manière analogue à celles du maximum de vraisemblance, il faut montrer que l'erreur moyenne tend vers zéro :

$$\frac{1}{I} \sum_{i=1}^I \xi_i \xrightarrow[I \rightarrow \infty]{p.s.} 0. \quad (3.81)$$

Dans les expressions (3.80) et (3.81), la notation ξ_i est un abus d'écriture. Pour l'application des formules de quadrature gaussienne, l'erreur d'approximation dépend du nombre d'abscisses N_{QG} considéré. Pour l'approximation par Monte Carlo, la taille de l'échantillonnage N_{MC} influence également l'erreur. La notation ξ_{iN} où N est mis pour N_{QG} ou N_{MC} , correspond alors mieux à l'expression du terme d'erreur.

Dans ces conditions, les propriétés de convergence des deux méthodes d'approximation considérées permettent d'écrire le résultat suivant :

$$\xi_{iN} \xrightarrow[N \rightarrow \infty]{} 0. \quad (3.82)$$

Pour l'approximation par Monte Carlo, ce résultat est obtenu par la loi des grands nombres [51, 85] ; il s'agit d'une convergence presque sûre. Pour la quadrature, la convergence des formules gaussiennes sur un espace d'intégration infini est donnée par Davis et Rabinowitz [23].

L'expression (3.82) s'écrit

$$\forall \varepsilon > 0, \exists N_0(i) \text{ tel que } \forall N > N_0(i), |\xi_{iN}| < \varepsilon \quad (3.83)$$

avec la notation $N_0(i)$ qui exprime que c'est le terme d'erreur propre à l'approximation pour chaque individu qui tend vers zéro.

Si on disposait d'un résultat de convergence uniforme (3.84) pour les approximations,

$$\forall \varepsilon > 0, \exists N_0^* \text{ tel que } \forall i, \forall N > N_0^*, |\xi_{iN}| < \varepsilon \quad (3.84)$$

on aurait alors la relation requise (3.85)

$$\frac{1}{I} \sum_{i=1}^I \xi_{iN} \xrightarrow[N \rightarrow \infty]{I \rightarrow \infty} 0. \quad (3.85)$$

À notre connaissance, des résultats de convergence uniforme ne sont pas disponibles pour les deux méthodes d'approximation envisagées. L'obstacle majeur de ce type de résultats est la dimension infinie de l'ensemble de définition des i qui varient de 1 à I avec $I \rightarrow \infty$.

Par ailleurs, en pratique, le nombre d'abscisses des quadratures N_{QG} est généralement relativement faible et ne permet pas d'atteindre un résultat asymptotique.

3.3.4 Synthèse sur les propriétés des estimateurs

L'objectif de cette section est d'établir un bilan sur les propriétés des divers estimateurs évoqués dans le cadre de l'approche paramétrique non bayésienne. Les six méthodes évoquées précédemment sont référencées dans le tableau 3.2. Elles sont mises en relation avec les propriétés de convergence, de normalité et d'efficacité des estimateurs qu'elles fournissent. La situation asymptotique nécessaire pour bénéficier des propriétés est précisée dans chaque cas : soit le nombre d'individus tendant vers l'infini suffit (situation **1**), soit le nombre d'observations par individus doit également tendre vers l'infini (situation **2**).

Méthode d'estimation	Convergence	Normalité	Efficacité	Référence
EM	①	●	●	[93]
PVS	①	●	○	[64]
TS	②	●	●	[28]
Linéarisation	②	●	●	[28]
Laplace	②?	●	●	[90]
Monte Carlo	?	?	?	
Quadrature gaussienne	?	?	?	

TAB. 3.2 – *Récapitulatif des propriétés des estimateurs obtenus par les diverses méthodes dans le cadre paramétrique non bayésien : ① et ② indiquent la situation asymptotique requise pour la convergence, ● indique une propriété démontrée, ○ une absence de propriété démontrée, et ? une absence de résultat pour la propriété en question.*

3.4 Conclusion

Parmi les méthodes présentées dans ce chapitre, il semble que les méthodes d'estimation par linéarisation et traitement séparé des individus soient les plus aptes à répondre à notre problème à condition de pouvoir disposer d'un nombre important de dates (plusieurs dizaines). Nous envisagerons également l'estimation selon les trois méthodes d'estimation par approximation numérique de la vraisemblance pour lesquelles nous ne disposons pas de résultats fiables sur les propriétés. Par ailleurs, les deux premières méthodes listées dans le tableau 3.2 peuvent présenter un intérêt si nous disposons de peu de dates, mais celles-ci sont reconnues pour être très coûteuses en temps de calcul. Or, une de nos contraintes initiales est de garantir le caractère opérationnel des méthodes mises en œuvre.

Nous retenons donc pour étude par simulation, hormis les méthodes par approximation, des méthodes dont les propriétés des estimateurs sont asymptotiques sur le nombre d'individus et sur le nombre de dates par individus. Ce point illustre un paradoxe en rapport avec l'utilisation des modèles non linéaires à paramètres aléatoires. En effet, leur utilisation est parfois justifiée par le faible nombre de dates par individus qui empêche de procéder à des analyses individuelles. Or les méthodes d'estimation auxquelles on a recours n'ont pas de propriétés à distance finie sur le nombre de dates. Pour conclure sur l'estimation dans les modèles non linéaires à paramètres aléatoires, disons qu'en termes de popularité, les méthodes par linéarisation sont les plus répandues. La disponibilité d'outils comme le logiciel NONMEM [6] et la librairie nlme [67] en est la principale raison.

Chapitre 4

Étude par simulation de méthodes d'estimation des paramètres d'un modèle non linéaire à paramètres aléatoires

Le but de ce chapitre est de présenter trois séries de simulations menées pour valider dans notre cas, l'utilisation de la méthode d'estimation par linéarisation dans les modèles non linéaires à paramètres aléatoires. Les questions posées sont les suivantes :

- A. Parmi les méthodes d'estimation par linéarisation (Lindstrom et Bates [54] et approximation numérique (laplacienne, Monte Carlo, quadrature gaussienne, Pinheiro et Bates [66]), laquelle assure le meilleur compromis entre précision des résultats et temps de calculs ?
- B. L'estimation par linéarisation dans un modèle à paramètres aléatoires est-elle préférable à une approche par pixel de type STS [83] classiquement utilisée en télédétection ?
- C. L'estimation par linéarisation est-elle perturbée par une structuration spatiale du bruit ? Par ailleurs, permet-elle de retrouver la répartition spatiale des paramètres aléatoires sans être explicitement prise en compte dans le modèle ?

La première section de ce chapitre présente le principe de simulation de valeurs de LAI. Ensuite, les trois sections suivantes proposent les séries de tests permettant de répondre aux questions ci-dessus.

4.1 Simulation de valeurs de LAI

Pour simuler des valeurs de LAI pour I pixels à J dates, nous générons I vecteurs de paramètres (β_i pour le pixel i) et IJ termes de bruit selon des lois normales :

$$\begin{cases} (\beta_i)_{i=1,\dots,I} \sim \mathcal{N}_p(\beta, \Gamma) & i.i.d. \\ (\varepsilon_{ij})_{\substack{i=1,\dots,I \\ j=1,\dots,J}} \sim \mathcal{N}(0, \sigma_\varepsilon^2) & i.i.d. \end{cases} \quad (4.1)$$

Dans l'expression (4.1), β est un vecteur de taille p et $\Gamma = \text{diag}(\sigma_\beta^2)$ est la matrice de variance des paramètres, supposée diagonale c'est-à-dire avec des paramètres non corrélés.

Les valeurs de LAI sont simulées à partir du modèle de Baret (Cf. 1.2.3.1) selon la relation :

$$LAI_{ij} = f(t_j, \beta_i) + \varepsilon_{ij} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (4.2)$$

où $(t_j)_{j=1, \dots, J}$ représente le temps en millier de degrés-jours à la $j^{\text{ème}}$ date.

Le problème de la simulation selon la relation (4.2) est qu'elle peut engendrer des valeurs de LAI fortement négatives pour certaines valeurs de paramètres, ce qui est en désaccord avec des mesures d'indice de surface foliaire positives ou nulles.

Pour pallier ce problème et coller davantage aux données réelles que représentent des valeurs de LAI, nous effectuons une remise à zéro des valeurs négatives selon la relation

$$LAI_{ij} = [f(t_j, \beta_i) + \varepsilon_{ij}] \wedge 0 \quad i = 1, \dots, I \quad j = 1, \dots, J \quad (4.3)$$

où $a \wedge b$ désigne $\sup(a, b)$.

Or sur quelques manipulations préalables à la mise en œuvre du plan d'expérience, il apparaît que cette remise à zéro provoque une sous-estimation systématique de la variance résiduelle. La figure 4.1 permet de mieux comprendre pourquoi la remise à zéro après bruitage des valeurs simulées provoque cette sous-estimation.

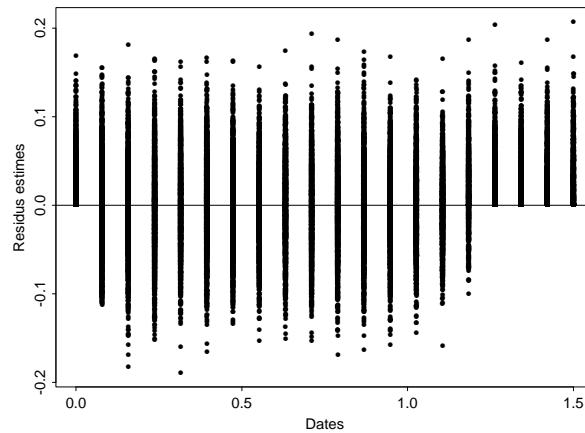


FIG. 4.1 – Représentation des résidus d'un ajustement par linéarisation en fonction de la covariable temps exprimée en milliers de degrés-jours.

La représentation des résidus estimés selon la covariable temps montre que dans les parties où les valeurs de LAI sont faibles (temps proche de 0 ou supérieur à 1,2 millier de degrés-jours), les résidus révèlent une structure particulière qui supprime une partie de leur variabilité, d'où la sous-estimation. Ce biais sur la variance résiduelle a également un effet sur l'estimation des paramètres.

Devant ces problèmes, nous avons modifié la phase de simulation en effectuant la remise à zéro de valeurs négatives non plus après bruitage des valeurs mais avant comme le montre la relation (4.4)

$$LAI_{ij} = [f(t_j, \beta_i) \wedge 0] + \varepsilon_{ij} \quad i = 1, \dots, I \quad j = 1, \dots, J. \quad (4.4)$$

La figure 4.2 illustre la différence entre les deux opérations de remise à zéro des valeurs négatives.

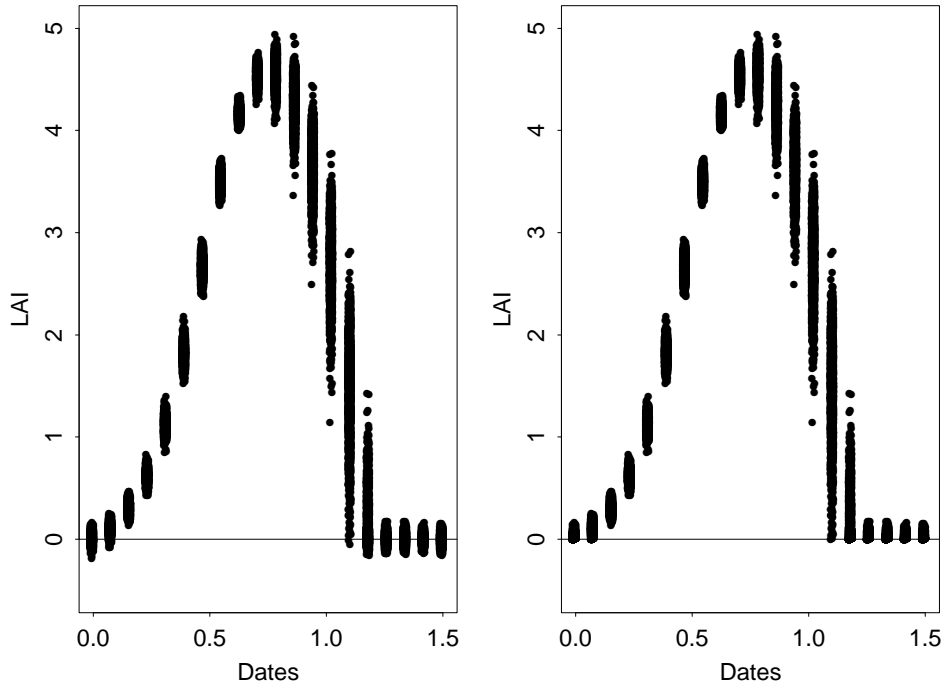


FIG. 4.2 – Illustration de la remise à zéro sur des données simulées. À gauche la remise à zéro est effectuée avant bruitage, à droite après bruitage.

Comme on peut le voir sur le graphique de droite de la figure 4.2, l'ajout du terme de bruit aux données simulées remises à zéro implique l'apparition de valeurs négatives de LAI dont la valeur absolue est directement liée à la variance résiduelle (Fig. 4.3).

Ces valeurs négatives de LAI, relativement faibles en valeur absolue, peuvent se justifier dans la pratique par l'obtention du LAI, non pas par mesure directe mais par conversion d'un indice de végétation. C'est le cas, lorsqu'en inversant la relation (1.1) ($IV = IV_\infty + (IV_s - IV_\infty) \times e^{(-K_{IV} \times LAI)}$), IV est inférieur à IV_s .

4.2 Comparaison de méthodes d'estimation par linéarisation et approximation numérique

La nécessité de la mise en œuvre d'un plan d'expérience pour la comparaison de méthodes d'estimation dans les modèles non linéaires à paramètres aléatoires se justifie

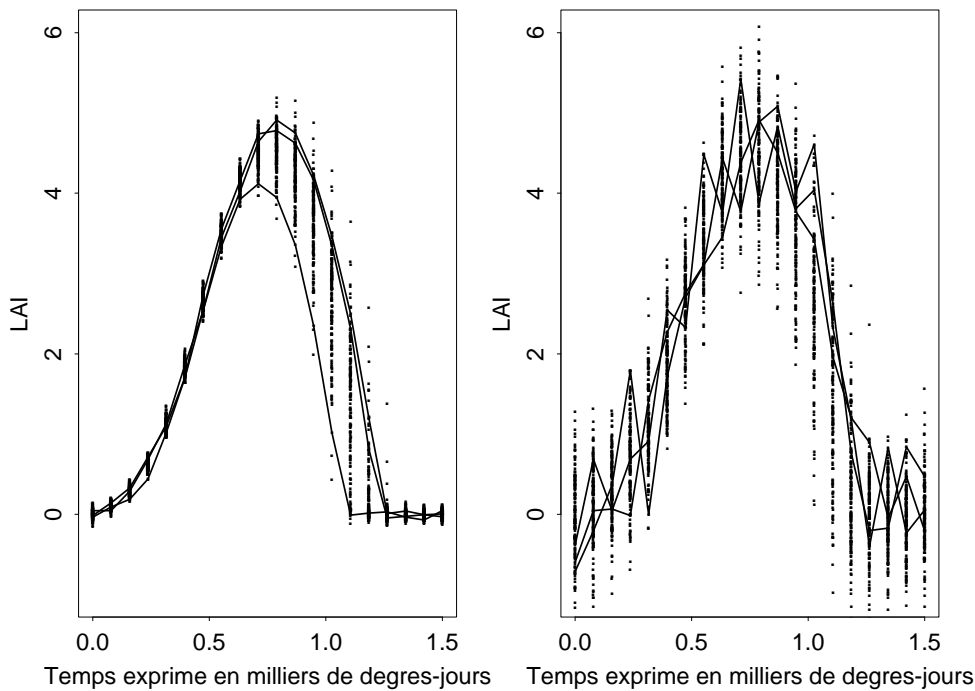


FIG. 4.3 – Valeurs simulées de LAI pour 1500 pixels et 20 dates dans une configuration faiblement bruitée (à gauche), fortement bruitée à droite et courbe d'évolution pour 3 pixels sur les 2 graphiques.

principalement par le fait que les tests comparatifs proposés dans la littérature sont effectués sur des jeux de données de petite taille par rapport au problème du suivi de cultures : quelques dizaines d'individus sur une dizaine de dates contre plus de mille pixels observés sur plusieurs dizaines de dates. Dans ces conditions, il est difficile de juger du caractère opérationnel des méthodes sur notre type de problèmes. Mise à part la méthode de linéarisation qui est implémentée dans la librairie `nlme` de Splus présentée en annexe (annexe B, section 5.2), les autres méthodes d'estimation par approximation de la vraisemblance ne bénéficient pas d'un outil utilisable directement (Cf. annexe B, section 5.3 pour leur mise en œuvre). Nous avons exclu de notre étude l'estimation par l'algorithme EM et par pseudo-vraisemblance simulée : deux méthodes reconnues comme coûteuses en temps de calcul.

4.2.1 Plan d'expérience

4.2.1.1 Motivations

Le plan d'expérience mis en œuvre est destiné à étudier le comportement des méthodes d'estimation dans divers cas de figure correspondant à :

- a. une zone d'étude homogène : les évolutions des cultures sont similaires sur toute l'image, les paramètres sont alors simulés constants sur l'ensemble des pixels,

- b. une zone d'étude hétérogène : les évolutions des cultures présentent une variabilité significative que l'on traduit en simulation par des paramètres aléatoires.

Chacun des deux cas de figures ci-dessus est croisé avec la prise en compte de deux valeurs de bruit résiduel pour traduire des corrections quasi-parfaites qui ne laissent subsister qu'un léger bruit entièrement aléatoire ou un bruit plus important dû éventuellement à la mise en œuvre de techniques de désagrégation (Cf. section 1.3.3.2).

4.2.1.2 Mise en œuvre

Suite à des premiers essais effectués avec le modèle de Baret à 4 paramètres ($T_r = \log(1 + \exp(a \times T_c))/b$), nous avons décidé de limiter le nombre de paramètres à estimer à 3, en considérant le paramètre K comme une constante, pour diminuer les temps de calcul et ainsi satisfaire la contrainte opérationnelle.

Config.	σ_a	σ_b	σ_{T_c}	σ_ε
1	0.1	0.1	0.01	0.05
2	0.1	0.1	0.01	0.5
3	0	0	0	0.05
4	0	0	0	0.5

TAB. 4.1 – Plan d'expérience à 2 facteurs (écart-type des paramètres aléatoires $\sigma_\beta = (\sigma_a, \sigma_b, \sigma_{T_c})$ et écart-type résiduel σ_ε) pour la comparaison des méthodes d'estimation par linéarisation et approximation : 100 répétitions par configuration.

Les données sont simulées (Tab. 4.1) selon des paramètres aléatoires (configurations 1 et 2) ou fixes (configurations 3 et 4), puis bruitées selon un bruit blanc gaussien indépendamment sur chaque pixel (équation (4.1)), faiblement (configurations 1 et 3) ou fortement (configurations 2 et 4). Les valeurs retenues pour les variances ($\sigma_\beta^2 = (0.1^2, 0.1^2, 0.1^2, 0.01^2)$) ont été préalablement déterminées selon une approche par pixel à partir d'un petit jeu de données réelles. Les deux valeurs retenues pour faire varier la variance du bruit sont $\sigma_\varepsilon^2 = 0.05^2$ et $\sigma_\varepsilon^2 = 0.5^2$. Elles correspondent environ à 1% et 10% du LAI maximal en termes d'écart-type.

Le nombre de pixels est fixé à 1500 pour l'ensemble des configurations. Cela correspond à une surface au sol de 1500 km² (soit environ de 30 à 40 km de côté) pour un capteur de résolution kilométrique. Le nombre de dates est fixé à 20. La valeur 20 correspond à une acquisition tous les 10 jours (ou à une synthèse décadaire) pendant environ 6 mois. Les dates sont simulées également réparties entre 0 et 1500 degrés-jours. La moyenne (β) des paramètres individuels (β_i) est fixée à $\beta = (8, 7, 3, 0.5)$ avec le temps exprimé en milliers de degrés-jours. Ces valeurs correspondent approximativement à celles observées sur un petit jeu de données réelles.

L'objectif de cette série de tests réside dans la comparaison des méthodes sur leur aptitude à estimer les paramètres globaux (moyenne et variance) dans un modèle non linéaire à paramètres aléatoires. L'aspect opérationnel de ces méthodes est également un critère important pour la comparaison. La méthode de linéarisation sert de référence au niveau des résultats. Nous cherchons à voir si l'une des méthodes d'estimation par

approximation de la vraisemblance améliore les résultats de la linéarisation avec des temps de calcul non prohibitifs.

Paramétrage des méthodes

Parmi les trois méthodes testées en complément de la linéarisation, deux d'entre elles nécessitent un paramétrage supplémentaire. La méthode de Monte Carlo requiert la taille de l'échantillonnage ; la quadrature gaussienne, le nombre d'abscisses dans chaque direction de l'espace des paramètres.

La détermination de ces paramètres est très largement influencée par la nécessité de garantir des temps de calcul « raisonnables » sur des problèmes de grande taille, non seulement pour un traitement unique de données mais surtout pour la répétition de traitements sur une série de tests.

Dans ces conditions, la taille de l'échantillonnage pour la méthode de Monte Carlo est fixée à 100 et le nombre d'abscisses pour la quadrature à 5, ce qui, dans un espace des paramètres à 3 dimensions crée une grille de $5^3 = 125$ points.

Initialisation des paramètres

Les valeurs initiales fournies à l'algorithme de minimisation pour les méthodes par approximation sont les résultats de la linéarisation : l'algorithme d'estimation par linéarisation étant seulement initialisé pour les effets fixes selon l'interprétation des paramètres du modèle de Baret :

- Calculs de pentes pour a et b ,
- valeur maximale de LAI pour K ,
- moyenne entre le premier temps d'observation et le temps où le LAI maximal est atteint pour T_c .

L'initialisation des algorithmes d'estimation par approximation numérique par les résultats de la linéarisation traduit notre intention de tester si une méthode par approximation permet d'améliorer les résultats de la linéarisation.

4.2.2 Résultats des méthodes d'approximation

Les résultats des trois méthodes d'estimation par approximation de la vraisemblance sont comparées d'abord sur le plan de l'efficacité opérationnelle à travers le nombre d'évaluations de fonctions nécessaires pour atteindre la convergence de l'algorithme de minimisation. Ensuite, nous présentons la comparaison des estimations obtenues par chaque méthode dans les quatre configurations. Les critères de biais et d'erreur quadratique moyenne correspondent respectivement aux propriétés de convergence et d'efficacité évoquée dans le tableau 3.2. La normalité est étudiée au travers des représentations des boîtes à pattes des estimations obtenues pour chaque configuration.

4.2.2.1 Aspect opérationnel

Un moyen de mesurer l'efficacité opérationnelle des méthodes, indépendamment de la machine sur laquelle on exécute le programme, est de compter le nombre d'évaluations de la fonction `mod_baret` (Cf. annexe B, section 5.3). Le tableau 4.2 présente ces indications

pour les trois méthodes dans les quatre configurations. Les valeurs présentées sont les moyennes sur les 100 simulations dans chaque configuration.

Config.	Laplace	MCIS	QG
1	80 000	8 500 000	41 500 000
2	100 000	8 300 000	52 500 000
3	92 000	8 700 000	55 200 000
4	96 000	8 200 000	30 300 000

TAB. 4.2 – Nombre moyen d'évaluations de la fonction `C mod_baret` jusqu'à la convergence de l'algorithme de minimisation pour les méthodes d'approximation laplacienne (Laplace), Monte Carlo par échantillonnage d'importance (MCIS) et quadrature gaussienne (QG).

En temps réel, sur une machine de type *Sun Ultrasparc 4 × 400 Mhz*, ces valeurs correspondent à une vingtaine de secondes pour l'approximation laplacienne, un peu plus de 10 minutes pour Monte Carlo et environ une heure pour la quadrature gaussienne.

La mise en œuvre de l'estimation par linéarisation étant différente, nous ne disposons pas d'informations équivalentes au nombre d'évaluations de fonctions. La comparaison du temps de calcul de `nlme` avec celui des méthodes d'approximation n'est pas non plus significatif car ses sorties contiennent plus d'information que la minimisation des approximations de la vraisemblance, notamment les estimations individuelles qui nécessitent un traitement supplémentaire sur les résultats des approximations. À titre d'information, signalons que l'estimation par `nlme` dure environ 25 minutes.

4.2.2.2 Estimation des paramètres

Pour les quatre méthodes et dans les quatre configurations, nous estimons le biais ($\text{biais}(\hat{\theta})$) et l'erreur quadratique moyenne ($\text{EQM}(\hat{\theta}) = [\text{biais}(\hat{\theta})]^2 + \text{var}(\hat{\theta})$) des estimateurs respectivement par

$$\frac{1}{100} \sum_{n=1}^{100} (\hat{\theta}^{(n)} - \theta)$$

et

$$\left[\frac{1}{100} \sum_{n=1}^{100} (\hat{\theta}^{(n)} - \theta) \right]^2 + \frac{1}{99} \sum_{n=1}^{100} (\hat{\theta}^{(n)} - \bar{\hat{\theta}})^2$$

où $\hat{\theta}^{(n)}$ représente l'estimation de θ à la $n^{\text{ème}}$ répétition. Une erreur quadratique moyenne faible caractérise un estimateur qui assure un bon compromis entre biais et variance. Dans les tableaux 4.3 à 4.6, nous utilisons la notation « $a.e^{-p}$ » pour $a \times 10^{-p}$. Nous complétons les tableaux relatifs au biais estimé par le test de Student de nullité de la moyenne des biais sur les 100 répétitions. Le symbole « * » indique un test significatif c'est-à-dire que le biais est significativement différent de 0 avec un niveau de risque de 5%.

En complément, pour illustrer la distribution des estimations, nous représentons la boîte à pattes des estimations des paramètres de moyenne et d'écart-type calculées pour les 100 répétitions dans chaque configuration. La vraie valeur est notée sur le graphique par une ligne pointillée. Les graphiques des estimations des effets fixes et des variances des paramètres sont disposés de gauche à droite selon les méthodes linéarisation (lin), Laplace (lapla), Monte Carlo (noté MCIS pour Monte Carlo Importance Sampling) et quadrature gaussienne (gaus), et de haut en bas respectivement pour les paramètres a , b et T_c .

Configuration 1

Les estimations obtenues par Monte Carlo et linéarisation sont globalement les moins biaisées sur les paramètres de moyenne et d'écart-type (Tab. 4.3). Cependant, la non significativité du biais par la méthode de Monte Carlo s'explique par une variance d'estimateur élevée qui apparaît dans l'erreur quadratique moyenne. Ce critère d'EQM est uniformément en faveur de la méthode de linéarisation.

Par ailleurs, les boîtes à pattes (Fig. 4.4, page 83), montrent que, mise à part la linéarisation, les méthodes présentent des résultats relativement éloignés d'une distribution normale.

Biais	lin	lapla	mcis	gaus	EQM	lin	lapla	mcis	gaus
a	$8.e^{-4*}$	$-4.e^{-4}$	$-8.e^{-3}$	$-1.e^{-2}$	a	$9.e^{-6}$	$2.e^{-4}$	$6.e^{-3}$	$1.e^{-2}$
b	$1.e^{-4}$	$-4.e^{-4}$	$-3.e^{-3}$	$1.e^{-2*}$	b	$7.e^{-6}$	$4.e^{-5}$	$7.e^{-4}$	$2.e^{-3}$
T_c	$-8.e^{-6}$	$7.e^{-6}$	$-5.e^{-4}$	$-7.e^{-4}$	T_c	$7.e^{-8}$	$6.e^{-7}$	$4.e^{-5}$	$5.e^{-5}$
σ_a	$3.e^{-4}$	$4.e^{-3*}$	$1.e^{-5}$	$5.e^{-2*}$	σ_a	$6.e^{-6}$	$8.e^{-5}$	$2.e^{-5}$	$3.e^{-3}$
σ_b	$5.e^{-4*}$	$4.e^{-3*}$	$2.e^{-4}$	$5.e^{-2*}$	σ_b	$4.e^{-6}$	$8.e^{-5}$	$1.e^{-5}$	$3.e^{-3}$
σ_{T_c}	$1.e^{-5}$	$1.e^{-3*}$	$-7.e^{-5}$	$5.e^{-3*}$	σ_{T_c}	$3.e^{-8}$	$6.e^{-6}$	$6.e^{-7}$	$3.e^{-5}$
σ_ε	$-8.e^{-6}$	$6.e^{-4*}$	$-2.e^{-4}$	$3.e^{-2*}$	σ_ε	$4.e^{-8}$	$2.e^{-6}$	$2.e^{-6}$	$8.e^{-4}$

TAB. 4.3 – Biais et erreur quadratique moyenne des estimations des paramètres de moyenne et d'écart-type dans la configuration 1.

Configuration 2

Dans la configuration 2 qui diffère de la précédente par une augmentation de l'écart-type résiduel dans la simulation des données, les estimations sont globalement plus biaisées (Tab. 4.4). Seule la méthode de Monte Carlo et l'approximation laplacienne (pour T_c) ne présentent pas de biais. Cependant, le critère d'EQM reste inférieur pour la linéarisation comme dans la configuration précédente.

Les boîtes à pattes (Fig. 4.5, page 84) illustrent par exemple la grande variabilité des estimations de Monte Carlo, ce qui est en accord avec la forte valeur d'EQM.

Configuration 3

La configuration 3 est la configuration pour laquelle les données sont le plus « propres » : les paramètres sont simulés constants et le bruit ajouté aux données est faible ($\sigma_\varepsilon = 0.05$).

Dans ces conditions privilégiées, les estimations des paramètres de moyenne par linéarisation, approximation laplacienne et quadrature gaussienne ne sont pas biaisées (Tab. 4.5).

Biais	lin	lapla	mcis	gaus	EQM	lin	lapla	mcis	gaus
a	$2.e^{-2*}$	$2.e^{-2*}$	$-0.5*$	$5.e^{-2*}$	a	$4.e^{-4}$	$5.e^{-3}$	3	$1.e^{-2}$
b	$1.e^{-2*}$	$1.e^{-2*}$	$-4.e^{-2}$	$4.e^{-2*}$	b	$2.e^{-4}$	$2.e^{-4}$	0.5	$6.e^{-3}$
T_c	$7.e^{-4*}$	$4.e^{-4}$	$-2.e^{-2}$	$2.e^{-2*}$	T_c	$1.e^{-6}$	$8.e^{-6}$	$1.e^{-2}$	$1.e^{-5}$
σ_a	$1.e^{-2*}$	$1.e^{-2*}$	$-7.e^{-3}$	$7.e^{-2*}$	σ_a	$9.e^{-4}$	$9.e^{-4}$	$4.e^{-3}$	$6.e^{-3}$
σ_b	$-4.e^{-3*}$	$-5.e^{-3*}$	$-2.e^{-2*}$	$7.e^{-2*}$	σ_b	$1.e^{-4}$	$1.e^{-4}$	$2.e^{-3}$	$6.e^{-3}$
σ_{T_c}	$2.e^{-3*}$	$-6.e^{-2*}$	$-2.e^{-2*}$	$8.e^{-3*}$	σ_{T_c}	$2.e^{-5}$	$3.e^{-2}$	$4.e^{-3}$	$7.e^{-5}$
σ_ε	$-3.e^{-4}$	$-8.e^{-4*}$	$-2.e^{-2*}$	$0.4*$	σ_ε	$5.e^{-6}$	$1.e^{-5}$	$6.e^{-3}$	0.1

TAB. 4.4 – *Biais et erreur quadratique moyenne des estimations des paramètres de moyenne et d'écart-type dans la configuration 2.*

Par contre, même dans ces conditions favorables, Monte Carlo ne peut pas garantir une très bonne précision. La taille de l'échantillonnage relativement faible (une centaine de points seulement) est la raison de ces résultats en deçà des autres méthodes. Par ailleurs sur les boîtes à pattes (Fig. 4.6, page 85), on peut constater que la distribution des estimations issues de la linéarisation est relativement proche d'une loi normale.

L'estimation des écarts-types des paramètres apparaît systématiquement biaisée pour les quatre méthodes (Tab. 4.5). Ceci s'explique par le fait que les méthodes d'estimations fournissent obligatoirement une valeur strictement positive et ne peuvent donc pas estimer précisément un écart-type nul. Les résultats sur l'EQM apparaissent là encore en faveur de la linéarisation. L'approximation laplacienne obtient également de bons résultats.

Malgré le biais systématique sur les estimations des écarts-types des paramètres, on peut constater sur les boîtes à pattes (Fig. 4.6, page 85) que les valeurs estimées sont globalement regroupées proche de zéro.

Biais	lin	lapla	mcis	gaus	EQM	lin	lapla	mcis	gaus
a	$2.e^{-4}$	$-4.e^{-4}$	-0.3	$-6.e^{-4}$	a	$1.e^{-6}$	$2.e^{-4}$	3	$1.e^{-4}$
b	$2.e^{-4}$	$-7.e^{-4}$	$-0.2*$	$-7.e^{-3}$	b	$8.e^{-7}$	$2.e^{-5}$	0.4	$4.e^{-3}$
T_c	$9.e^{-6}$	$1.e^{-4}$	$-4.e^{-2*}$	$-6.e^{-6}$	T_c	$5.e^{-9}$	$4.e^{-6}$	$2.e^{-2}$	$3.e^{-7}$
σ_a	$-6.e^{-4*}$	$-7.e^{-4*}$	$-2.e^{-3*}$	$-1.e^{-4*}$	σ_a	$9.e^{-7}$	$1.e^{-6}$	$3.e^{-5}$	$2.e^{-7}$
σ_b	$-2.e^{-4*}$	$-2.e^{-4*}$	$-9.e^{-4*}$	$-2.e^{-2}$	σ_b	$1.e^{-7}$	$1.e^{-7}$	$8.e^{-6}$	$1.e^{-2}$
σ_{T_c}	$-2.e^{-5*}$	$-4.e^{-5}$	$-6.e^{-4*}$	$-6.e^{-6*}$	σ_{T_c}	$3.e^{-9}$	$5.e^{-8}$	$7.e^{-6}$	$8.e^{-10}$
σ_ε	$3.e^{-5}$	$5.e^{-5*}$	$-3.e^{-2*}$	$3.e^{-2*}$	σ_ε	$4.e^{-8}$	$5.e^{-8}$	$1.e^{-2}$	$2.e^{-2}$

TAB. 4.5 – *Biais et erreur quadratique moyenne des estimations des paramètres de moyenne et d'écart-type dans la configuration 3.*

Configuration 4

Les résultats de cette configuration (Tab. 4.6) qui diffère de la précédente par une augmentation de l'écart-type résiduel amènent à peu près les mêmes commentaires que la

précédente :

- les estimations des paramètres de moyenne ne sont pas biaisées sauf pour Monte Carlo,
- les estimations des paramètres d'écart-types sont systématiquement biaisées car on estime une valeur nulle par des valeurs strictement positives,
- selon le critère d'EQM, c'est toujours la linéarisation qui présente les meilleurs résultats,
- les boîtes à pattes (Fig. 4.7, page 86) s'interprètent également de la même façon que dans la configuration précédente.

Biais	lin	lapla	mcis	gaus	EQM	lin	lapla	mcis	gaus
a	$2.e^{-3}$	-0.1	-0.4^*	$-1.e^{-3}$	a	$1.e^{-4}$	0.6	3	$1.e^{-2}$
b	$2.e^{-3}$	$-6.e^{-2}$	$-7.e^{-2}$	$-3.e^{-2}$	b	$8.e^{-5}$	0.1	0.5	$5.e^{-2}$
T_c	$8.e^{-5}$	$-2.e^{-3}$	$-4.e^{-2*}$	$4.e^{-4}$	T_c	$5.e^{-7}$	$3.e^{-4}$	$2.e^{-2}$	$1.e^{-5}$
σ_a	$-6.e^{-3*}$	$-7.e^{-3*}$	$-2.e^{-2*}$	$-2.e^{-3*}$	σ_a	$9.e^{-5}$	$1.e^{-4}$	$1.e^{-3}$	$2.e^{-5}$
σ_b	$-2.e^{-3*}$	$-5.e^{-3*}$	$-6.e^{-3*}$	$-1.e^{-2}$	σ_b	$1.e^{-5}$	$2.e^{-4}$	$3.e^{-4}$	$4.e^{-3}$
σ_{T_c}	$-3.e^{-4*}$	$-9.e^{-2*}$	$-2.e^{-3*}$	$-8.e^{-5*}$	σ_{T_c}	$3.e^{-7}$	$4.e^{-2}$	$6.e^{-5}$	$1.e^{-7}$
σ_ε	$4.e^{-4*}$	$4.e^{-4}$	$-2.e^{-2*}$	0.4^*	σ_ε	$4.e^{-6}$	$6.e^{-6}$	$1.e^{-2}$	0.2

TAB. 4.6 – *Biais et erreur quadratique moyenne des estimations des paramètres de moyenne et d'écart-type dans la configuration 4.*

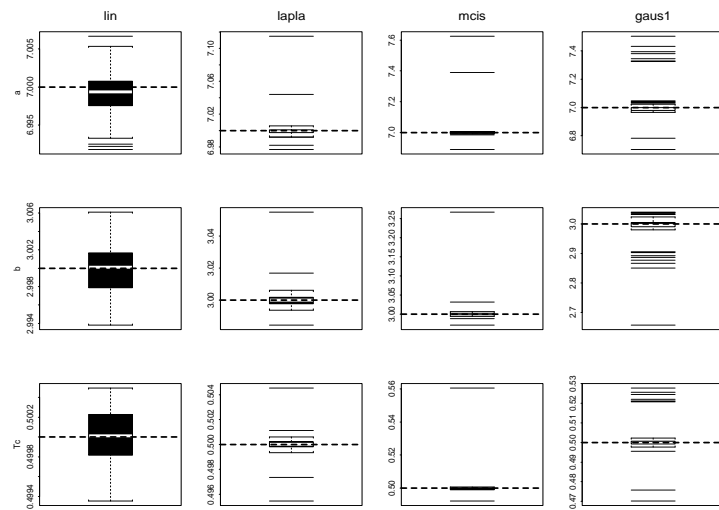
4.2.3 Conclusion des tests

En conclusion de cette série de tests, nous retiendrons que, dans notre cas, l'estimation par linéarisation fournit de très bons résultats sur un critère d'erreur quadratique moyenne et qu'aucune des autres méthodes testées ne permet d'en améliorer significativement la qualité uniformément sur l'ensemble des configurations.

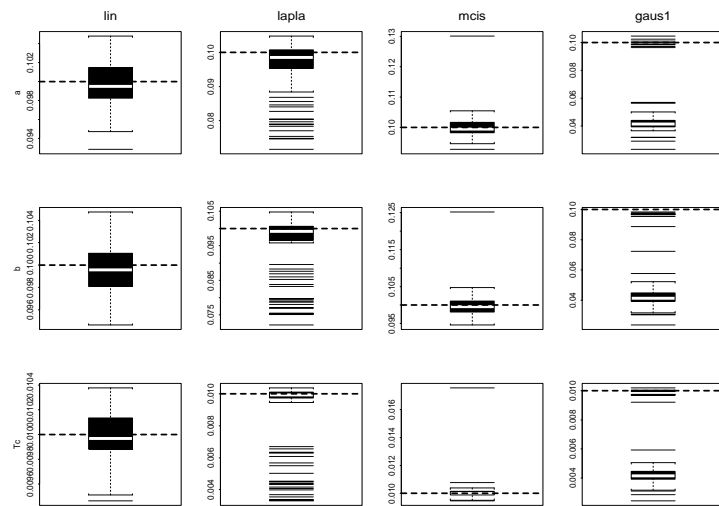
Cependant, nous n'en tirons pas la conclusion générale consistant à affirmer que la méthode de linéarisation est la meilleure. En effet, les méthodes de Monte Carlo et de quadrature gaussienne sont pénalisées par une taille d'échantillon relativement faible (100 points pour Monte Carlo, 5 abscisses dans chaque dimension pour la quadrature), mais ceci nous est imposé par la contrainte du caractère opérationnel (temps de calcul pas trop important) des algorithmes.

Or, dans le cadre du suivi de cultures, la taille du jeu de données à traiter étant très importante, il est difficile de favoriser ces deux méthodes en augmentant la taille de l'échantillonnage sans compromettre fortement l'aspect opérationnel.

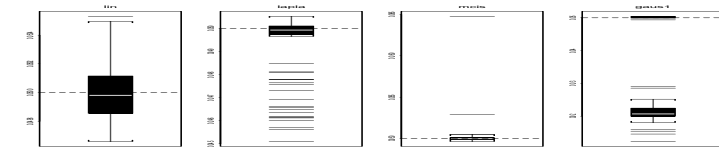
C'est pourquoi, au vu de ces résultats, nous validons l'utilisation de l'algorithme d'estimation par linéarisation dans le cadre de notre étude.



(a)

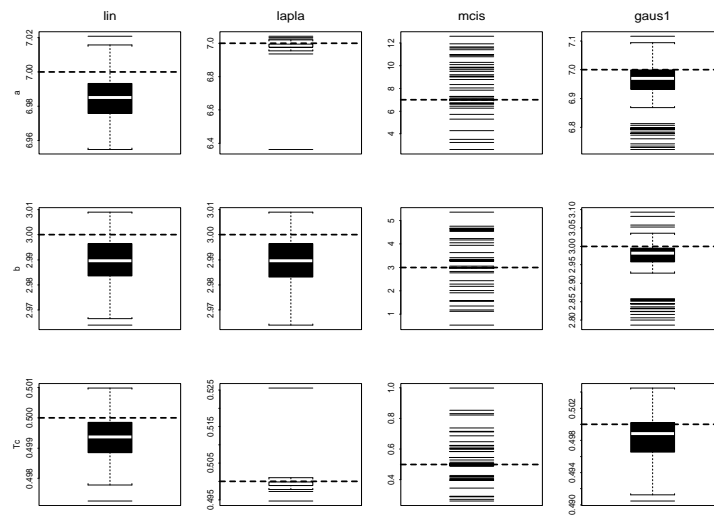


(b)

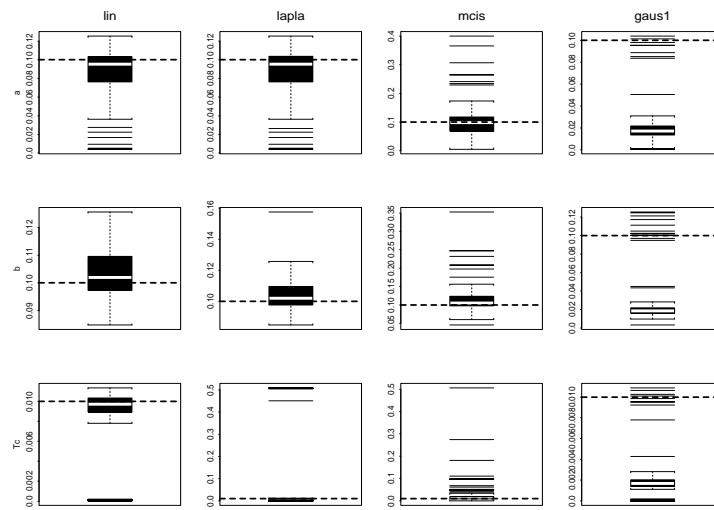


(c)

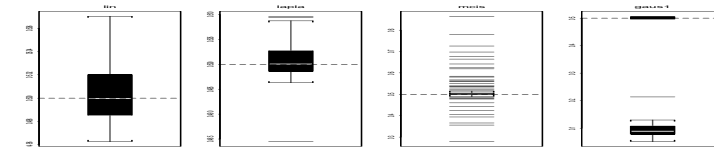
FIG. 4.4 – Boîtes à pattes des estimations des paramètres de moyennes β (a), et d'écart-type σ_β (b) et σ_ϵ (c) dans la configuration 1.



(a)

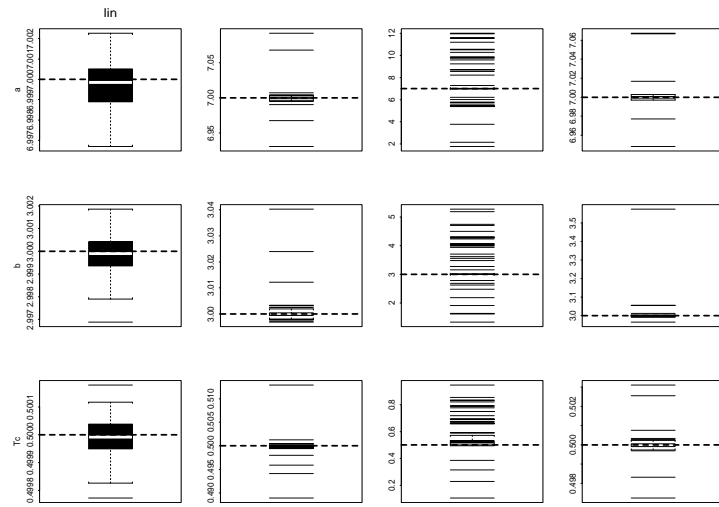


(b)

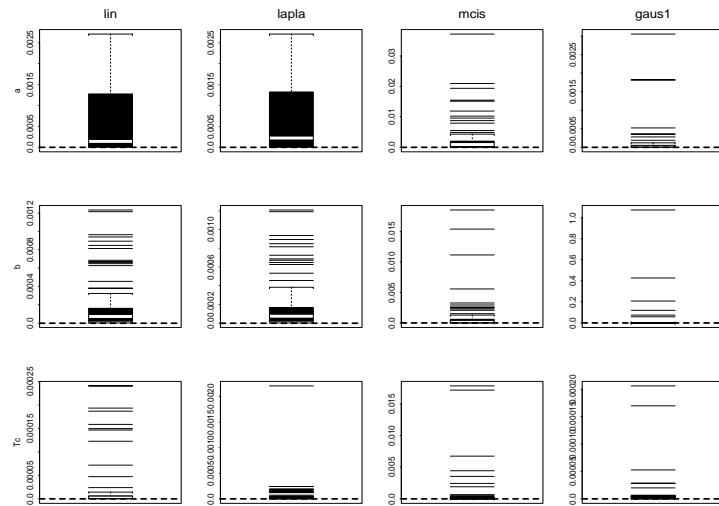


(c)

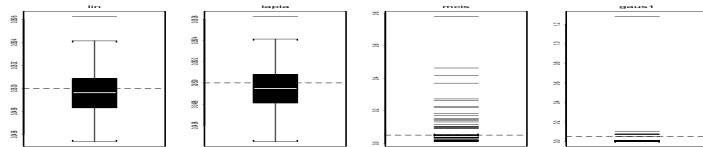
FIG. 4.5 – Boîtes à pattes des estimations des paramètres de moyennes β (a), et d'écart-type σ_β (b) et σ_ϵ (c) dans la configuration 2.



(a)

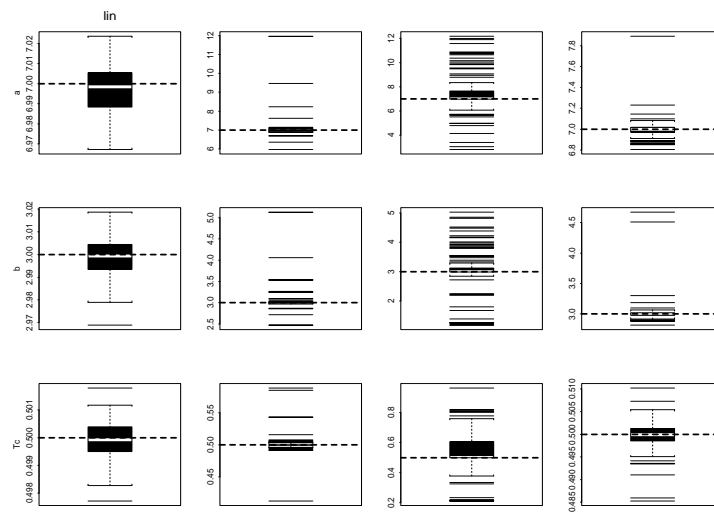


(b)

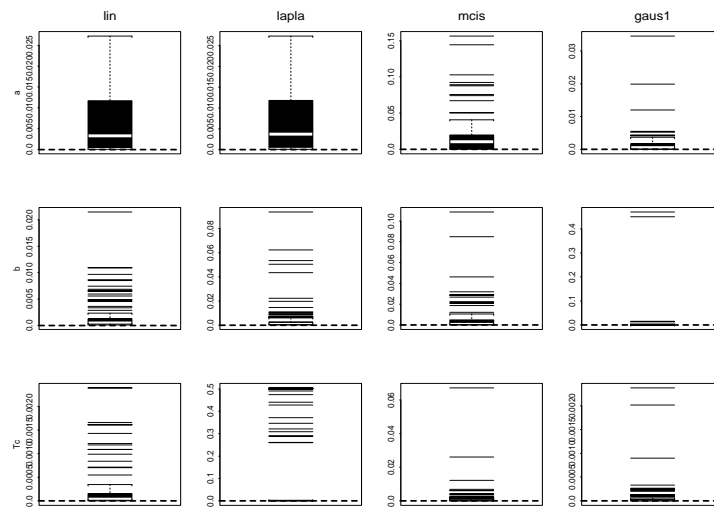


(c)

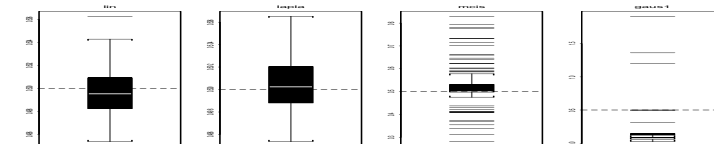
FIG. 4.6 – Boîtes à pattes des estimations des paramètres de moyennes β (a), et d'écart-type σ_β (b) et σ_ϵ (c) dans la configuration 3.



(a)



(b)



(c)

FIG. 4.7 – Boîtes à pattes des estimations des paramètres de moyennes β (a), et d'écart-type σ_β (b) et σ_ϵ (c) dans la configuration 4.

4.3 Comparaison d'estimations par pixel et par linéarisation dans un modèle non linéaire à paramètres aléatoires

Dans cette section, nous comparons sur données simulées deux approches pour l'estimation dans un modèle à paramètres aléatoires : une de type STS (*Cf.* section 3.2.2) basée sur les estimations individuelles par pixel et l'autre, la linéarisation (*Cf.* section 3.2.5.2), traitant tous les pixels simultanément. Le plan d'expérience mis en œuvre permet de tester les deux approches dans diverses configurations faisant varier le nombre de dates, le nombre de paramètres à estimer et l'écart-type résiduel. Cette série de simulation est reprise dans [31].

4.3.1 Plan d'expérience

4.3.1.1 Motivations

La mise en œuvre d'un plan d'expérience vise à apporter des éléments de comparaison sur le comportement des deux approches, par pixel et par linéarisation, sur les questions suivantes :

- a. Quelle est l'influence du nombre de paramètres à estimer sur la précision des estimateurs en fonction de la qualité des données ?
- b. Quelle est l'influence du nombre de dates sur la précision des estimateurs en fonction de la qualité des données ?
- c. Dans le cas de données fortement bruitées, comment évolue la précision des estimations en fonction du nombre de dates ?

4.3.1.2 Mise en œuvre

Pour répondre aux questions définies précédemment, nous construisons un plan d'expérience à 3 facteurs à travers 7 configurations (Tab. 4.7).

- a. Sous-plan 1 : les configurations 1, 2, 3 et 4, à nombre de dates égal, permettent d'étudier les effets croisés du nombre de paramètres à estimer et de la variance du bruit.
- b. Sous-plan 2 : avec les configurations 3, 4, 5 et 6, nous étudions les effets croisés entre le nombre de dates et la variance du bruit.
- c. Sous-plan 3 : enfin, les configurations 6, 4 et 7 sont destinées à montrer l'effet du nombre de dates sur la précision des estimateurs dans le cas de données fortement bruitées.

Les valeurs retenues pour les facteurs du plan d'expérience sont les suivantes :

- Pour prendre en compte les aléas climatiques, notamment l'enneigement qui rend parfois impossible l'exploitation de données, nous considérons plusieurs valeurs pour

Configuration	Nombre de dates	Paramètres à estimer	σ_ε^2
1	20	(K, a, b, T_c)	0.05^2
2	20	(K, a, b, T_c)	0.5^2
3	20	(a, b, T_c)	0.05^2
4	20	(a, b, T_c)	0.5^2
5	10	(a, b, T_c)	0.05^2
6	10	(a, b, T_c)	0.5^2
7	40	(a, b, T_c)	0.5^2

TAB. 4.7 – Plan d’expérience à 3 facteurs : nombre de dates (10, 20 et 40), paramètres à estimer ((K, a, b, T_c) et (a, b, T_c)) et variance du bruit ($\sigma_\varepsilon = 0.05^2$ et 0.5^2). 100 répétitions par configuration.

le nombre de dates : 10, 20 et 40. Autour de la valeur de référence de 20 dates (1 image tous les 10 jours pendant 6 mois), 40 et 10 dates représentent respectivement des acquisitions plus ou moins nombreuses car effectuées dans des conditions d’enregistrement plus ou moins bonnes. On dispose ainsi d’une image tous les 5 jours ou tous les 20 jours durant 6 mois. Les dates sont simulées également réparties entre 0 et 1500 degrés-jours pour 20 et 40 dates. Les 10 dates sont sélectionnées aléatoirement parmi les 20 précédemment définies.

- Dans la phase d’estimation, nous utilisons le modèle de Baret en considérant soit $p = 4$ paramètres (K, a, b, T_c) , soit $p = 3$ paramètres (a, b, T_c) à estimer. Dans ce dernier cas, le paramètre K est supposé caractéristique de la plante étudiée sur la région et donc fixé pour l’ensemble des pixels. La moyenne (β) et la variance (σ_β^2) des paramètres individuels (β_i) sont respectivement fixées à $\beta = (8, 7, 3, 0.5)$ et $\sigma_\beta^2 = (0.1^2, 0.1^2, 0.1^2, 0.01^2)$ comme dans la précédente série.
- Les deux valeurs retenues pour faire varier la variance du bruit sont les mêmes que précédemment ($\sigma_\varepsilon^2 = 0.05^2$ et $\sigma_\varepsilon^2 = 0.5^2$).

Par ailleurs, pour toutes les configurations, le nombre de pixels (I) est fixé à 1500 comme dans la série précédente.

Initialisation des algorithmes

Les deux procédures d’estimation sont implémentées dans un algorithme itératif qui nécessite une initialisation des paramètres. Seuls les effets fixes sont initialisés pour l’estimation par linéarisation. Pour les régressions par pixel, deux options ont été étudiées sur quelques simulations :

- initialisation commune à l’ensemble des pixels,
- initialisation spécifique à chaque pixel.

Les résultats obtenus dans les deux cas étaient très proches en termes de taux de convergence et de valeurs obtenues. Nous avons conservé l’initialisation commune à l’en-

semble des pixels qui est plus rapide.

Critères de comparaison

Les critères de comparaison sont calculés à partir des vraies valeurs que nous connaissons par la simulation.

Pour la comparaison des paramètres régionaux, nous utilisons un critère de biais exprimé en pourcentage relatif :

$$BPR(\hat{\theta}) = \left(\frac{1}{N} \sum_{n=1}^N \frac{\hat{\theta}^{(n)} - \theta}{\theta} \right) \times 100 \quad (4.5)$$

où θ est mis pour la vraie valeur de β , σ_β ou σ_ε , et $\hat{\theta}^{(n)}$ le paramètre estimé pour le n^{ème} jeu de données avec l'une des deux méthodes.

Pour les paramètres individuels, nous utilisons un critère d'erreur quadratique moyenne de prédiction que nous estimons par :

$$EQMP(\hat{\beta}_i) = \frac{1}{N} \sum_{n=1}^N \left(\frac{1}{I} \sum_{i=1}^I \left(\tilde{\beta}_i^{(n)} - \beta_i^{(n)} \right)^2 \right) \quad (4.6)$$

où $\beta_i^{(n)}$ désigne le vrai vecteur de paramètres sur le pixel i pour le n^{ème} jeu de données et $\tilde{\beta}_i^{(n)}$ est l'estimation correspondante issue de l'une des deux méthodes.

4.3.2 Résultats

Dans les tableaux présentés dans cette section, LIN et RTS (variante de STS qui débiaise l'estimation de la variance des paramètres, équation 3.11) font référence respectivement aux méthodes d'estimation par linéarisation et par les régressions par pixel. Avant de présenter et commenter les résultats relatifs aux critères de comparaison, nous donnons quelques informations liées à la convergence des procédures d'estimation.

4.3.2.1 Convergence

L'algorithme de Gauss-Newton utilisé pour la minimisation dans les régressions par pixel peut rencontrer des difficultés pour atteindre la convergence (Tab. 4.8).

Dans les configurations faiblement bruitées (1, 3, 5), les régressions ne convergent pas pour environ 10 pixels sur 1500, soit pour un peu moins de 1% des pixels. Ce chiffre atteint près de 10% dans les configurations fortement bruitées. Les pixels concernés par ce problème sont retirés du calcul des paramètres globaux. Pour la linéarisation, dès que les paramètres régionaux sont estimés, l'estimation des paramètres individuels est effectuée pour l'ensemble des pixels.

L'estimation par linéarisation est environ dix fois plus lente que par régressions par pixel. Cependant, ces durées ne sont pas rédhibitoires. Sur une machine de type *Sun Ultrasparc 4 × 400 Mhz*, l'estimation pour 1500 pixels, 20 dates et 4 paramètres à estimer avec une structure de variance diagonale ne prend pas plus de trente minutes en temps réel avec la fonction `nlme` de `Splus`.

Configuration	Nombre de pixels
1	12
2	110
3	12
4	113
5	11
6	109
7	122

TAB. 4.8 – Nombre de pixels sur 1500 au total sur lesquels l’algorithme des moindres carrés n’a pas convergé dans chaque configuration.

4.3.2.2 Paramètres de moyenne

Pour commenter le premier sous-plan d’expérience formé par les configurations 1 à 4, nous prenons comme référence la configuration 3 où les deux méthodes obtiennent des résultats relativement proches.

Config.	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
1	0.1	-0.1	-0.1	0.03	0.3	0.03	0.04	0.04
2	10	-8	-6	2	250	45	40	8
3		-0.01	-0.02	-0.004		0.01	0.01	-0.002
4		-0.4	-0.6	-0.2		1	2	0.005
5		-0.01	-0.03	-0.01		0.1	0.1	0.01
6		-0.5	-1	-0.6		0.5	10	3
7		-0.3	-0.4	-0.1		0.5	0.7	-0.004

TAB. 4.9 – Estimation du biais en pourcentage relatif (BPR , équation (4.5)) pour les estimations des paramètres régionaux par LIN et RTS.

Avec une variance d’erreur faible, estimer 4 paramètres au lieu de 3, provoque une erreur relative 10 fois plus grande pour LIN et environ 2 fois plus grande pour RTS qui obtient les meilleurs résultats. En revanche, une forte variance résiduelle est plus nuisible à RTS ; l’erreur relative de RTS dans la configuration 4 est 100 fois plus grande que dans la configuration 3, et environ 40 à 60 fois pour LIN. Si l’on combine à la fois 4 paramètres à estimer et une forte variance résiduelle (configuration 2), les résultats deviennent très mauvais pour RTS qui tombe sur ces cas pathologiques.

Pour le second sous-plan (configurations 3 à 6), la configuration 3 peut également servir de référence. Diminuer le nombre de dates (configuration 3 vers 5) a peu d’influence sur LIN si la variance résiduelle est faible ; pour RTS, l’erreur relative est jusqu’à 10 fois plus importante que pour LIN, selon le paramètre.

Dans le cas des configurations bruitées avec 3 paramètres (configurations 4, 6 et 7), les résultats sont proches pour 40 et 20 dates (configurations 7 et 4) sur a et b et meilleur

pour RTS sur T_c . Le passage de 40 à 10 dates provoque une erreur relative 2 fois plus grande pour LIN. Pour RTS, ce chiffre atteint 100 sur le paramètre T_c .

4.3.2.3 Paramètres de variance

L'estimation de l'écart type résiduel (σ_ε) pose peu de problèmes. L'erreur relative atteint il est vrai 15% par RTS dans la configuration 2, mais ne dépasse pas 2% dans les autres cas. On peut noter que quelle que soit la configuration, LIN présente un pourcentage d'erreur toujours inférieur à celui de RTS. Pour RTS, le facteur essentiel est le nombre de dates plutôt que l'intensité de la variance résiduelle ; les meilleurs résultats sont obtenus dans la configuration 7. Pour LIN, l'erreur relative n'excède pas 0.3% (configuration 5).

Config.	LIN					RTS				
	σ_K	σ_a	σ_b	σ_{T_c}	σ_ε	σ_K	σ_a	σ_b	σ_{T_c}	σ_ε
1	-0.3	-0.02	-0.1	-0.1	0.004	-4	-1	-0.04	-0.1	-1.5
2	40	-80	-5	8	0.2	*	*	*	*	15
3		-0.02	-0.1	-0.05	0.01		0.02	0.04	-0.01	-1
4		-12	5	-20	0.08		170	85	30	-1.3
5		-0.2	-0.6	-0.2	0.3		-7	-0.7	-2	-1.8
6		-30	6	-26	0.2		*	*	*	6
7		0.5	-1	1	-0.04		50	30	8	-0.8

TAB. 4.10 – Estimation du biais en pourcentage relatif (BPR, équation (4.5)) pour les estimations des paramètres de variance par LIN et RTS corrigé selon la relation (3.11), (page 49).

Pour l'estimation de la matrice Γ , la correction des résultats de RTS peut produire des résultats aberrants (marqués par * dans le tableau 4.10) si la variance correspondant à la précision des estimateurs ($\tilde{\Sigma}_i$) est supérieure à la variance globale des estimateurs par pixels ($\text{var}(\hat{\beta}_i)$). Les fortes valeurs des éléments de la matrice $\tilde{\Sigma}_i$ proviennent du faible nombre de dates (par rapport au nombre de paramètres à estimer) et de l'écart-type résiduel élevé (Cf. section 3.2.2.1).

Les résultats aberrants sont obtenus dans les configurations 2 et 6. Dans ces cas, LIN obtient ses pires résultats (80% pour σ_a dans la configuration 2). Dans les autres configurations fortement bruitées, l'erreur relative de RTS est de 10 à 100 fois plus grande que pour LIN. Dans les configurations peu bruitées, RTS obtient de meilleurs résultats avec 20 dates et 3 paramètres à estimer. Avec 4 paramètres (configuration 1), l'erreur relative de RTS est multipliée par 10 alors que celle de LIN reste à peu près la même. Dans la configuration 5, avec 10 dates, l'erreur relative de LIN est environ 10 fois plus petite que pour RTS.

4.3.2.4 Paramètres individuels

Quelle que soit la configuration, l'erreur quadratique moyenne de prédiction (EQMP, équation (4.6)) reste quasiment inchangée pour LIN (Tab. 4.11). Le seul changement si-

gnificatif apparaît dans la configuration 2, avec une EQMP 10 fois plus grande que dans les autres configurations pour le paramètre a .

Config.	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
1	0.02	0.03	0.04	0.0004	0.1	0.06	0.007	0.00003
2	0.8	0.4	0.07	0.0004	80000	400	50	0.02
3		0.04	0.04	0.0004		0.002	0.001	0.00001
4		0.02	0.03	0.0002		0.4	0.2	0.001
5		0.03	0.04	0.0003		0.005	0.004	0.00003
6		0.02	0.03	0.0002		14	1	0.01
7		0.02	0.03	0.0002		0.1	0.06	0.0004

TAB. 4.11 – Estimation de l’erreur quadratique moyenne de prédiction (EQMP, équation (4.6)) pour les paramètres individuels estimés par LIN et RTS.

De manière générale, les résultats de RTS sont meilleurs que LIN sur des données faiblement bruitées. En revanche, une forte variance résiduelle est très préjudiciable à RTS : par exemple, de la configuration 3 à la configuration 4 l’EQMP est multipliée par 100. La combinaison de quatre paramètres et d’une forte variance résiduelle produit des résultats médiocres. La très grande valeur d’EQMP pour le paramètre K est principalement due à 5% des pixels où les valeurs estimées atteignent 250 voire plus. Dans ces cas, la croissance apparaît relativement lente, ce qui implique un a faible. Pour compenser ceci, le paramètre K est alors fortement augmenté et le pont d’inflexion théorique de la phase de croissance T_c est déplacé au-delà de la date de récolte. Cependant, une courbe obtenue avec des valeurs à première vue extravagantes reste relativement proche d’une courbe produite par des valeurs plus conformes aux valeurs simulées (Fig. 4.8). Nous conservons toutefois notre intérêt pour l’estimation des paramètres dans le but de proposer une cartographie des valeurs estimées.

Sur les configurations 3 à 6, on peut noter que la diminution du nombre de dates en conservant une faible variance résiduelle n’augmente pas de beaucoup l’erreur de RTS. En revanche, avec une forte variance résiduelle, RTS devient imprécis quand seulement 10 dates sont disponibles.

Dans les configurations fortement bruitées, on peut constater que RTS et LIN obtiennent des résultats comparables sur 40 dates. Dès que ce nombre diminue, les résultats de RTS se dégradent ; l’EQMP de RTS est multipliée par 4 quand le nombre de dates diminue de 40 à 20 et par 100 de 20 à 10 dates alors que pour RTS l’EQMP varie très peu.

4.3.3 Conclusion

Les résultats présentés dans la section précédente illustrent quelques remarques générales en relation avec les facteurs du plan d’expérience. Les estimations locales ou régio-

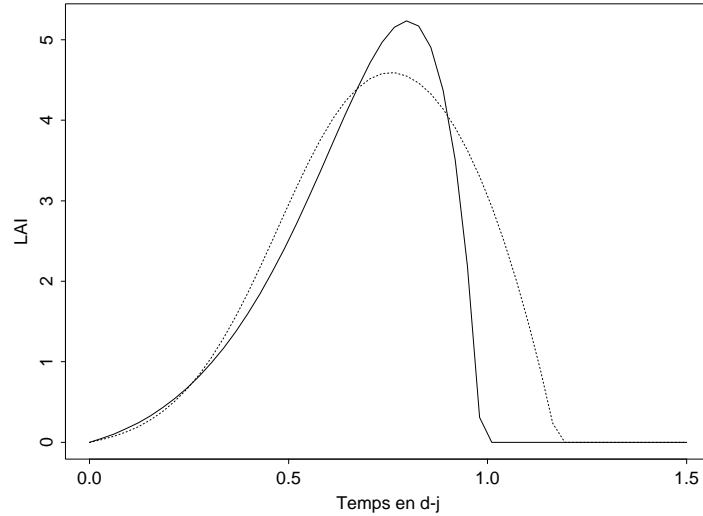


FIG. 4.8 – Courbes de LAI tracées à partir de deux vecteurs de paramètres : ($K = 7421, a = 3.17, b = 3.1, T_c = 1.8$) (trait plein) et $K = 8, a = 7, b = 3, T_c = 0.5$ (trait pointillé).

nales issues des deux méthodes sont plus précises quand :

1. le nombre de paramètres à estimer est faible,
2. le nombre d'observations par individu est élevé,
3. les données ne sont pas trop fortement bruitées.

Dans ces conditions, le traitement par pixel obtient globalement de meilleurs résultats. Toutefois, que ce soit sur les estimations régionales ou locales, la linéarisation est moins sensible que les régressions par pixel à la dégradation de ces conditions optimales.

Pour les estimations individuelles, LIN ne donne pas de résultats aberrants comme peut en produire RTS dont l'EQMP peut devenir très élevée dans les configurations fortement bruitées (80000 pour K dans la configuration 2). La raison de cette plus grande robustesse de LIN réside dans le calcul des prédictions individuelles. En effet, rappelons que les effets individuels sont calculés par minimisation sur les B_i de la fonction

$$\sum_{i=1}^I \left[-\frac{1}{2\hat{\sigma}_\varepsilon^2} \sum_{j=1}^J [Y_{ij} - f(t_j, \hat{\beta} + B_i)]^2 - \frac{1}{2} B_i^T \hat{\Gamma}^{-1} B_i \right]$$

une fois estimés les paramètres globaux $(\beta, \Gamma, \sigma_\varepsilon^2)$ (Cf. section 3.2.5.2). Cette fonction ressemble à la quantité minimisée dans l'algorithme de moindres carrés de RTS, avec un terme de pénalisation $(B_i^T \hat{\Gamma}^{-1} B_i)$. Ce terme agit comme une contrainte pour l'estimation des B_i et empêche la méthode de linéarisation de fournir des valeurs aberrantes. De plus, le fait d'estimer les paramètres individuels β_i en distinguant la moyenne globale β et la variation individuelle B_i implique nécessairement des valeurs resserrées autour de β .

En pratique, devant l'incertitude liée au bruit sur les données, on aura plutôt intérêt à utiliser la méthode d'estimation par linéarisation qui présente une plus grande robustesse vis-à-vis d'une augmentation de l'écart-type résiduel.

4.4 Effet d'une structure spatiale sur l'estimation par linéarisation

Dans les deux précédentes séries de tests sur données simulées, les valeurs pour chaque pixel sont simulées indépendamment. La motivation essentielle de cette troisième série est de mettre en œuvre l'estimation par linéarisation dans un modèle à paramètres aléatoires sur des données simulées avec une structure spatiale. Nous envisageons deux cas de figures : la structure spatiale porte soit sur le bruit ajouté aux données (section 4.4.2), soit sur les paramètres aléatoires (section 4.4.3). Nous présentons dans un premier temps, le principe de simulation d'une structure spatiale.

4.4.1 Simulation d'une structure spatiale

Nous présentons brièvement les principes de base de la génération de nombres aléatoires sur une grille selon une structure spatiale. Pour plus de détails, on peut consulter Cressie [17], chapitre 6. Par commodité, les notations utilisées dans cette section sont indépendantes du reste du document.

Le procédé de simulation d'une structure spatiale est basé sur la relation

$$Z = \mu + B(Z - \mu) + \epsilon, \quad (4.7)$$

qui s'écrit en notation développée

$$Z(s_i) = \mu_i + \sum_{j=1}^I b_{ij}(Z(s_j) - \mu_j) + \epsilon_i \quad (4.8)$$

où les notations utilisées sont les suivantes :

- Id_I est la matrice identité de taille $I \times I$, avec I le nombre de pixels de l'image.
- $B = (b_{ij})_{\substack{i=1,\dots,I \\ j=1,\dots,I}}$ est la matrice qui exprime la dépendance spatiale entre les pixels.

On la suppose symétrique et de diagonale nulle ($(b_{ii})_{i=1,\dots,I} = 0_I$).

- $Z = (Z(s_1), \dots, Z(s_I))'$, désigne le vecteur aléatoire dont on souhaite simuler des observations. Il est indexé par la position connue s_i de ces éléments : $s_i = (l, c)$ représente la position de l'élément i de Z sur la grille à l'intersection de la ligne l et de la colonne c ,

- $\mu \in \mathbb{R}_I$ est la moyenne de la variable aléatoire Z .
- ϵ est une variable aléatoire de \mathbb{R}_I distribuée selon une loi normale de moyenne nulle et de variance $\sigma^2 \text{Id}_I$. Les éléments de ϵ sont aussi indexés par leur position s_i .

Les relations (4.7) et (4.8) illustrent de façon explicite la dépendance entre les valeurs $Z(s_i)$ simulées sur chaque pixel.

La transformation de la relation (4.7) en

$$(\text{Id}_I - B)(Z - \mu) = \epsilon. \quad (4.9)$$

met en évidence le procédé de génération des données. En effet, en supposant la matrice $(\text{Id}_I - B)$ inversible, on a

$$Z = \mu + (\text{Id}_I - B)^{-1}\epsilon$$

ce qui est équivalent à

$$Z \sim \mathcal{N}(\mu, \sigma^2(\text{Id}_I - B)^{-1}(\text{Id}_I - B^T)^{-1}).$$

Comme nous l'avons précisé dans les notations, la dépendance spatiale est exprimée par la matrice B . Nous montrons sur un exemple simple comment elle est déterminée.

Exemple de détermination de la matrice B

Pour simplifier le problème, nous posons $B = \rho W$ où $\rho \in \mathbb{R}$ et W est une matrice de même taille que B et composée de 0 et de 1.

Considérons par exemple, une grille 4×4 sur laquelle nous souhaitons simuler des données selon une structure de voisinage avec les plus proches voisins à gauche, à droite, en haut et en bas (Table 4.12).

s_1	s_2	s_3	s_4
s_5	s_6	s_7	s_8
s_9	s_{10}	s_{11}	s_{12}
s_{13}	s_{14}	s_{15}	s_{16}

TAB. 4.12 – Représentation d'une grille 4×4 sur laquelle on a représenté la dépendance du pixel s_{11} avec ses plus proches voisins horizontaux et verticaux.

La matrice W correspondant à la dépendance spatiale exprimée sur la table 4.12 est alors définie par :

$$W = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \end{matrix} & \left(\begin{array}{cccccccccccccccc} \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot & 1 & \cdot \end{array} \right) \end{matrix} \quad (4.10)$$

où les 0 ont été remplacés par des « . » pour alléger l'écriture. L'affichage des indices de lignes et de colonnes permet un repérage plus facile des relations entre pixels.

Si l'élément (i, j) de la matrice W vaut 1, cela signifie que les pixels s_i et s_j sont voisins et qu'il existe donc entre eux une relation de dépendance.

Analyse de structures spatiales

Le variogramme est une fonction de structure qui traduit la corrélation entre les données spatiales. Il est défini par

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

où h représente un accroissement.

Le variogramme empirique est alors défini par

$$\hat{\gamma}(h) = \frac{1}{2|\mathcal{I}(h)|} \sum_{i \in \mathcal{I}(h)} [Z(s_i) - Z(s_j)]^2 \quad s_j \in \mathcal{V}_h(s_i)$$

où $\mathcal{I}(h)$ désigne l'ensemble des pixels admettant des pixels voisins à distance h (les pixels au bord de l'image n'en admettent pas forcément), $|\mathcal{I}(h)|$ le cardinal de cet ensemble et $\mathcal{V}_h(s_i)$ l'ensemble des pixels éloignés de s_i d'une distance h .

Dans notre cas, nous calculons les valeurs du variogramme séparément pour des voisinages en ligne et en colonne selon le système :

$$\begin{cases} \hat{\gamma}_l(h) = \frac{1}{2|\mathcal{I}_l(h)|} \sum_{i \in \mathcal{I}_l(h)} [Z(l_i, c_i) - Z(l_i + h, c_i)]^2 \\ \hat{\gamma}_c(h) = \frac{1}{2|\mathcal{I}_c(h)|} \sum_{i \in \mathcal{I}_c(h)} [Z(l_i, c_i) - Z(l_i, c_i + h)]^2 \end{cases}$$

où (l_i, c_i) représente les coordonnées du pixel i et $\mathcal{I}_l(h)$ (resp. $\mathcal{I}_c(h)$) désigne l'ensemble des pixels admettant des pixels voisins en ligne (resp. en colonne) à distance h .

Le variogramme résultant est donc une matrice que l'on peut représenter sous forme d'image (Fig. 4.9) ou par courbes représentant chaque ligne du variogramme (Fig. 4.10).

Les figures 4.9 et 4.10 illustrent que si la corrélation spatiale est très forte (image de gauche, $\rho = 0.99$), les valeurs du variogramme croissent régulièrement : plus des pixels sont proches (h petit), plus leurs valeurs sont liées et donc faiblement variables (faible valeurs du variogramme). En revanche, si la corrélation spatiale est faible, comme c'est le cas sur l'image de droite, la variance est à peu près constante quel que soit la distance de voisinage considérée et la corrélation est faible.

4.4.2 Structure spatiale sur le bruit

L'introduction d'une structure spatiale sur le bruit ajouté aux données traduit un comportement de dépendance de l'erreur vis-à-vis du voisinage. En pratique, cela peut

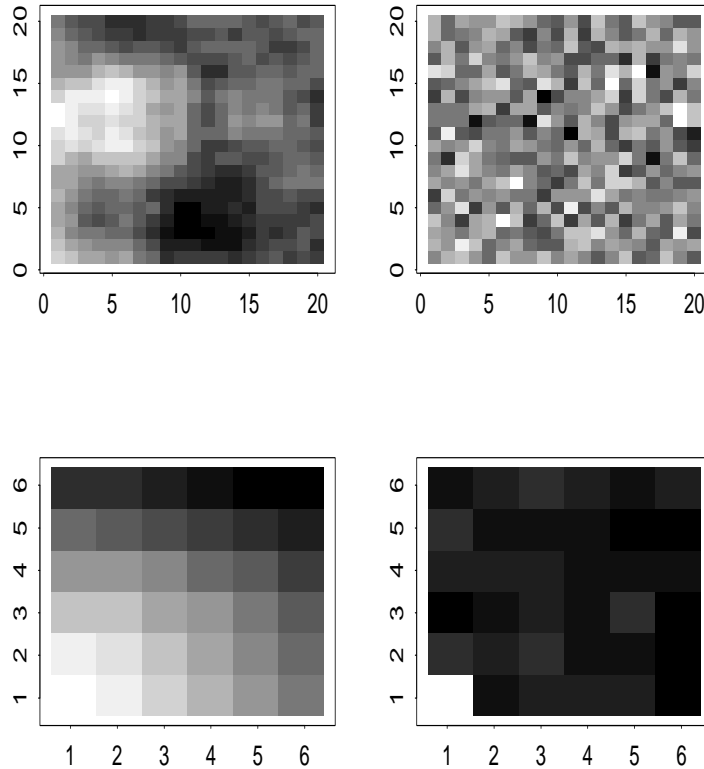


FIG. 4.9 – Représentation de deux images 20×20 simulées (en haut) et de leur variogramme empirique associé pour $h = 1, \dots, 5$ (en bas, valeur croissante du blanc vers le noir) : à gauche $\rho = 0.99$, à droite $\rho = 0.01$.

provenir de la phase de désagrégation. Selon les thèmes observés en mélange avec la culture d'intérêt et leur pourcentage d'occupation sous le pixel mixte, l'erreur commise est susceptible de présenter une certaine hétérogénéité spatiale. Par exemple, il est plus facile de retrouver la réflectance du blé sous un pixel mixte composé à 80% de blé que sous un pixel avec 20% de blé. L'erreur due à la désagrégation varie donc selon les zones de l'image en distinguant par exemple les zones à dominante rurale ou urbaine. La mise en œuvre sur données simulées permet de tester l'influence d'une telle structure spatiale dépendante du plan d'occupation des sols sur l'estimation des paramètres du modèle.

4.4.2.1 Plan d'expérience

Le principe de simulation des données est analogue à ceux précédemment mis en œuvre. Nous simulons d'abord I ($I = 1600$ pour disposer d'une image carrée 40×40) vecteurs de paramètres aléatoires selon une loi normale $((\beta_i)_{i=1, \dots, I} \sim \mathcal{N}_p(\beta, \Gamma)$, Cf. première équation du système (4.1)). Les valeurs pour β et Γ sont fixées à $(8, 7, 3, 0.5)$ et $\text{diag}(0.1^2, 0.1^2, 0.1^2, 0.01^2)$. Le nombre d'images J est fixé à 20.

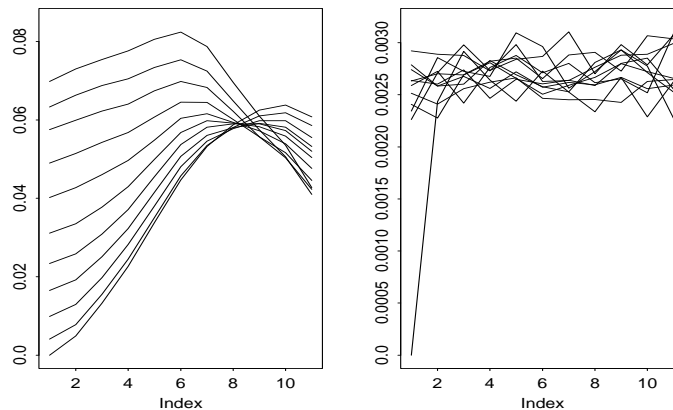


FIG. 4.10 – Représentation des courbes représentant les lignes des variogrammes associés aux images de la figure 4.9 pour $h = 1, \dots, 10$.

La seule différence réside dans la génération du terme d'erreur qui comporte ici une structure spatiale auto-régressive.

Génération de l'erreur

Contrairement aux cas précédents, nous ne générons pas directement $I \times J$ termes d'erreur, mais nous procédons en deux phases :

1. génération de I termes d'erreur selon une structure spatiale analogue à celle de la table 4.12: $\varepsilon_1(i) = (\text{Id}_I - \rho W)^{-1} \epsilon$ avec $\rho = 0.8$ et $\epsilon \sim \mathcal{N}(0, \sigma_1^2 \text{Id}_I)$,
2. génération de $I \times J$ termes d'erreur $\varepsilon_2(i)$ selon une loi normale centrée de variance σ_2^2 avec $\sigma_2^2 < \sigma_1^2$.

L'erreur totale sur un pixel i est la somme des deux termes $\varepsilon_1(i)$ et $\varepsilon_2(i)$ ainsi générés.

La première phase crée la structure spatiale de base pour une image. La seconde permet de conserver la structure initiale en la modifiant par un léger bruit résiduel dû aux corrections variable au cours du temps. Nous faisons l'hypothèse que la structure liée au plan d'occupation des sols est plus forte que celle liée par exemple aux corrections atmosphériques variables au cours du temps.

Cette procédure est une alternative aux deux cas extrêmes consistant soit à ajouter systématiquement le même bruit (phase 1 uniquement) à chaque image soit à régénérer une nouvelle structure à chaque date. Ces deux solutions sont peu en phase avec des conditions réelles : la première parce qu'elle suppose que l'erreur ne varie pas au cours du temps, la seconde parce qu'elle modifie la structure spatiale initiale et donc ne respecte plus la situation au sol.

Pour les simulations, nous conservons les valeurs 0.5 et 0.05 pour les écarts-types résiduels. Ces valeurs sont décomposées en σ_1 et σ_2 selon le tableau 4.13.

Nous simulons 20 jeux de données dans chacune des deux configurations.

Cfg.	σ	σ_1	σ_2
1	0.5	0.45	0.05
2	0.05	0.045	0.005

TAB. 4.13 – Valeurs des écarts-types σ_1 et σ_2 pour les deux phases de la simulation en fonction de l'écart-type global σ dans les configurations fortement et légèrement bruitées.

4.4.2.2 Résultats

Nous utilisons les mêmes critères que dans la section 4.3 : biais exprimé en pourcentage relatif pour les paramètres globaux (Cf. équation (4.5)) et erreur quadratique moyenne de prédiction pour les paramètres individuelles (Cf. équation (4.6)).

Faible variance du bruit

Critère	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
$BPR(\beta)$	-0.003	0.004	0.002	-5.10^{-6}	-0.03	-0.01	-0.005	-5.10^{-5}
$BPR(\sigma_\beta)$	-0.06	0.04	-0.01	-0.001	-0.4	-0.3	-0.08	-7.10^{-5}
$BPR(\sigma_\varepsilon)$	0.003				0.013			
$EQMP(\beta)$	0.05	0.02	0.04	0.0004	0.3	0.1	0.03	2.10^{-6}

TAB. 4.14 – Estimation du biais en pourcentage relatif (BPR , équation (4.5)) pour les estimations des paramètres régionaux et de l'erreur quadratique moyenne de prédiction ($EQMP$, équation (4.6)) pour les paramètres individuels par LIN et RTS dans la configuration faiblement bruitée.

L'erreur commise par LIN sur les données faiblement bruitées est globalement inférieure à celle commise par RTS sur les estimations des paramètres régionaux (Tab. 4.14). En revanche, les résultats sur les paramètres individuels sont partagés. Ils sont meilleurs pour LIN pour K et a mais moins bons pour b et T_c . Cependant, quelle que soit la méthode d'estimation, le pourcentage d'erreur et l'erreur quadratique restent relativement faible.

On peut dire que si la variance du bruit est faible, sa répartition selon une structure spatiale ne fausse pas l'estimation des paramètres régionaux ou locaux.

Forte variance du bruit

Pour les données fortement bruitées (Tab. 4.15), les erreurs commises par les deux méthodes d'estimation sont plus importantes que sur les données faiblement bruitées. On observe toujours que LIN est plus précis que RTS.

Si l'on s'intéresse exclusivement à l'estimation par linéarisation, on constate que le pourcentage d'erreur ne dépasse pas 1.2% en valeur absolue, valeur atteinte pour l'estimation de l'écart-type du paramètre K . Sur les paramètres individuels, l'EQMP est maximale pour K mais reste encore relativement faible.

Même avec une variance élevée, un bruit structuré spatialement ne fausse pas plus l'estimation des paramètres régionaux et locaux que si le bruit est réparti aléatoirement.

Critère	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
$BPR(\beta)$	-0.01	0.2	0.1	0.01	0.9	-2	-1	-0.001
$BPR(\sigma_\beta)$	-1.2	0.1	-0.04	-0.002	-2.5	-3.2	-1.5	-0.003
$BPR(\sigma_\varepsilon)$	0.01				0.2			
$EQMP(\beta)$	1.5	0.04	0.06	0.0003	7	15	3.5	7.10^{-5}

TAB. 4.15 – Estimation du biais en pourcentage relatif (BPR , équation (4.5)) pour les estimations des paramètres régionaux et de l'erreur quadratique moyenne de prédiction ($EQMP$, équation (4.6)) pour les paramètres individuels par LIN et RTS dans la configuration fortement bruitée.

Les valeurs de biais et d'erreur quadratique moyenne inférieures à celles obtenues dans la série précédente s'explique par une variabilité empirique plus faible ($0.45^2 + 0.05^2 < 0.5^2$).

4.4.3 Structure spatiale sur les paramètres aléatoires

L'introduction d'une structure spatiale sur les paramètres traduit des comportements proches pour le développement de la culture d'intérêt sur des pixels voisins. La simulation de données à partir d'une structure spatiale sur les paramètres permet de tester si l'estimation par linéarisation *via* un modèle à paramètres aléatoires retrouve cette structure.

4.4.3.1 Plan d'expérience

La génération de données avec une structure spatiale sur les paramètres aléatoires est effectuée selon le système (4.11)

$$\left\{ \begin{array}{l} (K_i)_{i=1,\dots,I} \sim \mathcal{N}(K, \sigma_K^2 V) \\ (a_i)_{i=1,\dots,I} \sim \mathcal{N}(a, \sigma_a^2 V) \\ (b_i)_{i=1,\dots,I} \sim \mathcal{N}(b, \sigma_b^2 V) \\ (T_{c_i})_{i=1,\dots,I} \sim \mathcal{N}(T_c, \sigma_{T_c}^2 V) \\ (\varepsilon_{ij})_{\substack{i=1,\dots,I \\ j=1,\dots,J}} \sim \mathcal{N}(0, \sigma_\varepsilon^2). \end{array} \right. \quad (4.11)$$

où $V = (\text{Id}_I - \rho W)^{-1}(\text{Id}_I - \rho W^T)^{-1}$ avec $\rho = 0.8$. Dans cette expression W représente la matrice indicatrice des dépendances spatiales représentées dans la table 4.12 pour les quatre paramètres. Comme précédemment, les valeurs moyennes pour les paramètres sont fixées à $(8, 7, 3, 05)$. Les variances $(\sigma_K^2, \sigma_a^2, \sigma_b^2, \sigma_{T_c}^2)$ conservent également les mêmes valeurs que dans la série précédente $(0.1^2, 0.1^2, 0.1^2, 0.01^2)$.

L'erreur est supposée aléatoire et issue d'une loi normale centrée. Nous générons simultanément des valeurs pour les J images sans considération spatiale. Nous reprenons les deux valeurs 0.5 et 0.05 pour σ_ε .

4.4.3.2 Résultats

Dans la présentation des résultats, nous distinguons les paramètres régionaux et locaux. Pour les premiers, nous donnons comme précédemment le biais en pourcentage relatif (BPR, Cf. équation (4.5)) qui donne une indication globale de la précision des estimations ; à titre de comparaison, nous donnons également les résultats de l'estimation par régressions par pixel (RTS). Pour les seconds, nous nous intéressons exclusivement à l'aptitude de l'estimation par linéarisation à retrouver les structures spatiales.

Faible variance du bruit

Les valeurs du tableau 4.16 révèlent une bonne précision globale des deux méthodes d'estimation. Tous les BPR sont largement inférieurs à 1%. Ces résultats sont cohérents avec ceux déjà obtenus dans des configurations peu bruitées.

Critère	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
$BPR(\beta)$	-0.006	0.007	0.002	-0.0001	-0.02	-0.003	-0.002	-0.0002
$BPR(\sigma_\beta)$	-0.05	-0.05	-0.05	-0.005	-0.3	-0.2	-0.07	-0.006
$BPR(\sigma_\varepsilon)$	1.10^{-5}				0.0008			

TAB. 4.16 – Estimation du biais en pourcentage relatif pour les estimations des paramètres régionaux par LIN et RTS dans la configuration faiblement bruitée.

La représentation des valeurs de paramètres simulées et estimées sous forme d'image (Fig. 4.11) n'est pas forcément très explicite même si on peut quand même constater que l'on retrouve grossièrement les mêmes teintes dans les mêmes zones pour les quatre paramètres.

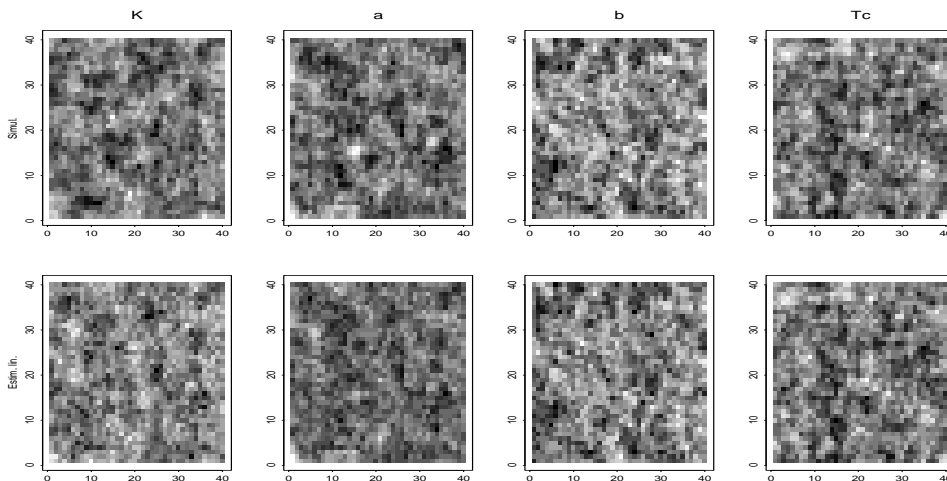
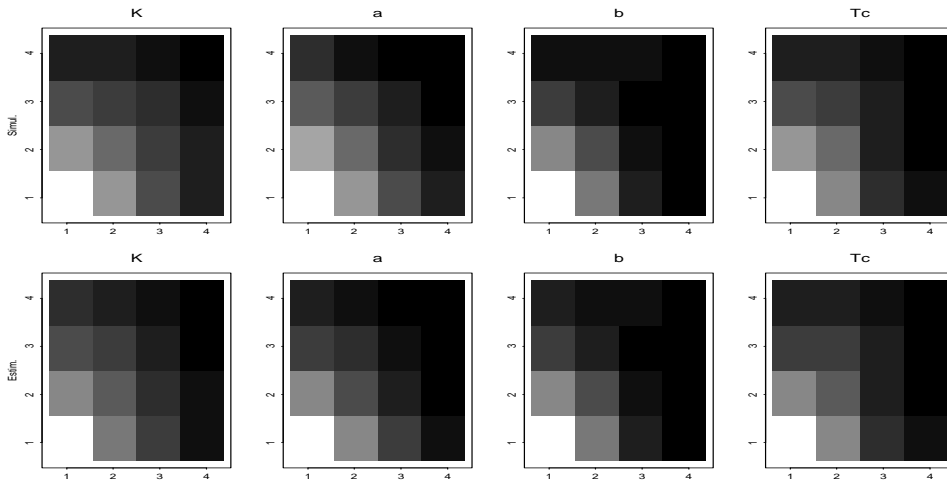
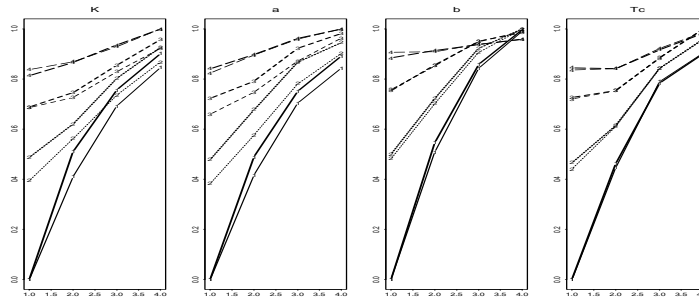


FIG. 4.11 – Représentations des variations autour des valeurs moyennes des paramètres individuels simulés (en haut) et estimés par linéarisation (en bas) pour K , a , b et T_c (de gauche à droite) dans la configuration faiblement bruitée.

Pour compléter l'analyse spatiale des résultats, nous construisons le variogramme empirique des données simulées et estimées pour les quatre paramètres (Fig. 4.12).



(a) Image des variogrammes : simulation en haut, estimation en bas.



(b) Courbes des variogrammes normalisés : simulation en trait fin, estimation en trait épais.

FIG. 4.12 – Variogrammes des valeurs simulées et estimées pour les quatre paramètres dans la configuration faiblement bruitée ($h = 1, \dots, 3$).

Quel que soit le paramètre, l'estimation par linéarisation semble fournir des valeurs corrélées selon la même structure spatiale que les simulations. Les résultats sur les paramètres b et T_c semblent particulièrement bons : les courbes relatives aux lignes du variogramme sont quasiment superposées.

Forte variance du bruit

Dans le cas d'une forte variance du bruit, les estimations globales par linéarisation ne sont plus aussi précises que précédemment mais restent tout à fait correctes. En revanche, les estimations issues des régressions par pixel présentent des erreurs relatives plus élevées notamment sur les écarts-types des paramètres.

Critère	LIN				RTS			
	K	a	b	T_c	K	a	b	T_c
$BPR(\beta)$	-0.8	0.6	0.2	-0.01	-13	-1.9	-0.7	-0.03
$BPR(\sigma_\beta)$	-0.1	0.03	-0.04	-0.006	-180	-8	-3	-0.1
$BPR(\sigma_\varepsilon)$	-0.001				-0.008			

TAB. 4.17 – Estimation du biais en pourcentage relatif pour les estimations des paramètres régionaux par LIN et RTS dans la configuration faiblement bruitée.

Les images de la figure 4.13 laissent supposer que l'estimation par linéarisation a plus de mal à retrouver les structures spatiales des paramètres avec des données plus fortement bruitées. En effet, on peut constater, essentiellement sur le paramètre K , que les images issues des estimations ne présentent quasiment aucun bloc de pixels homogènes, comme c'est le cas sur les données simulées.

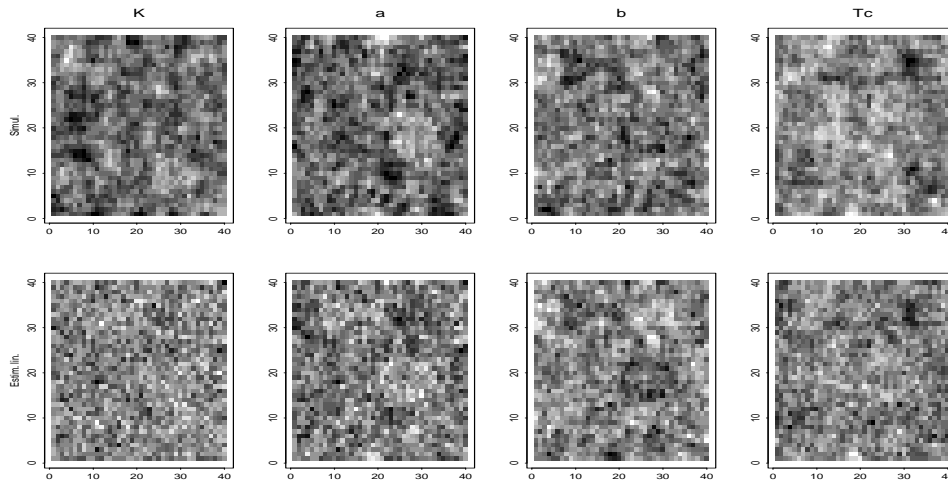


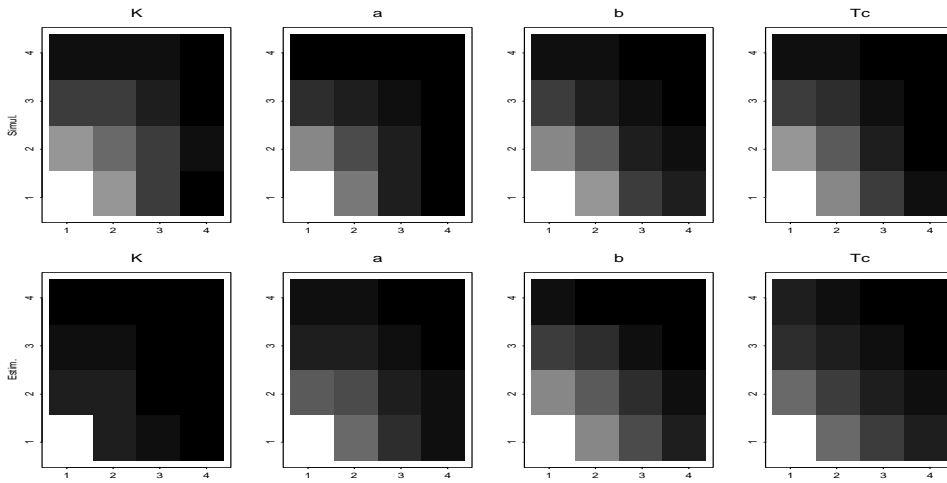
FIG. 4.13 – Représentations des variations autour des valeurs moyennes des paramètres individuels simulés (en haut) et estimés par linéarisation (en bas) pour K , a , b et T_c (de gauche à droite) dans la configuration fortement bruitée.

Ce point est confirmé par le variogramme des données estimées du paramètre K (Fig. 4.14). Les courbes relatives aux variogrammes sur données simulées et estimées sont relativement différentes. En revanche, les résultats sont corrects sur les autres paramètres.

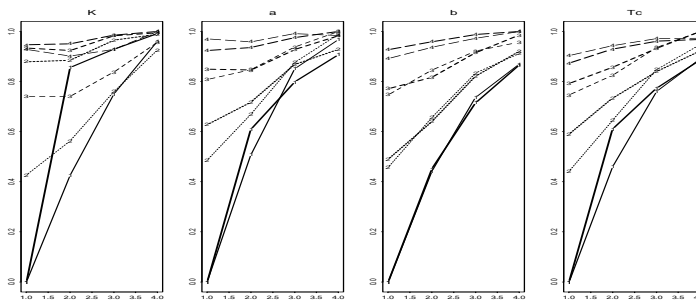
Synthèse

La synthèse graphique des résultats (Fig. 4.15) est obtenue en rangeant ligne par ligne les valeurs représentées sur les images des figures 4.11 et 4.13 dans un vecteur de taille 1600. On dispose ainsi de deux vecteurs, l'un pour les images simulées, l'autre pour les estimations, pour chaque paramètre.

La figure 4.15 illustre que les estimations individuelles par linéarisation varient dans le même sens que les valeurs simulées. Conformément à l'interprétation de la figure 4.11, on constate que, dans la configuration faiblement bruitée, les nuages de points ont une



(a) Image des variogrammes : simulation en haut, estimation en bas.



(b) Courbes des variogrammes normalisés : simulation en trait fin, estimation en trait épais.

FIG. 4.14 – Variogrammes des valeurs simulées et estimées pour les quatre paramètres dans la configuration fortement bruitée ($h = 1, \dots, 3$).

allure étirée selon une droite ce qui traduit l'aptitude de l'estimation par linéarisation à retrouver les valeurs individuelles simulées structurées spatialement. La seconde ligne de graphique illustre les difficultés rencontrées lorsque les données sont plus fortement bruitées. Les nuages de points sont moins étirés mais conservent quand même une forme généralement orientée le long d'une droite de pente positive.

4.4.4 Conclusion

Quelle que soit la variance de l'erreur, une structure spatiale sur celle-ci ne compromet pas l'estimation des paramètres. Si ceux-ci sont générées aléatoirement sans structure spatiale, l'estimation par linéarisation dans un modèle à paramètres aléatoires commet des erreurs relativement faibles.

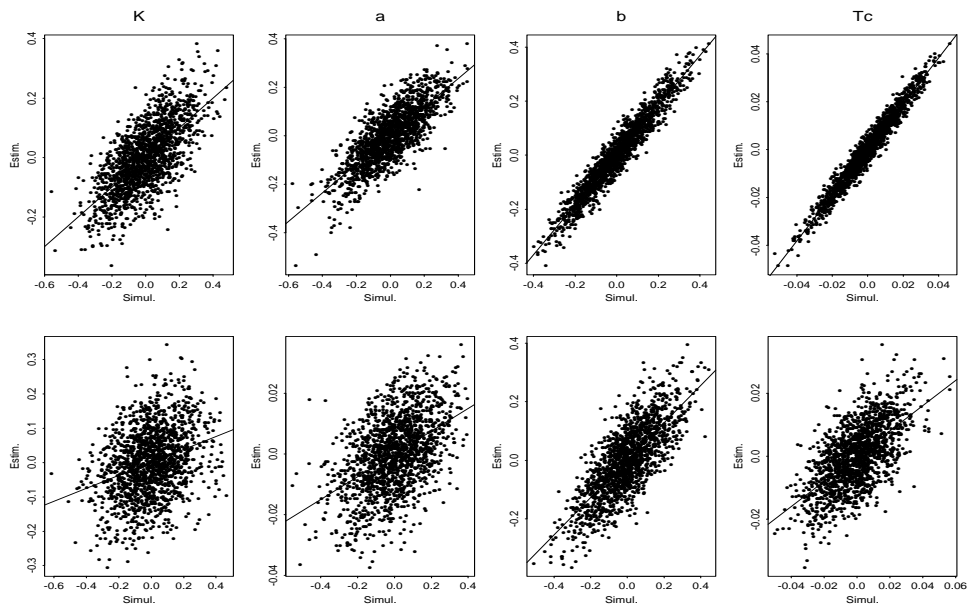


FIG. 4.15 – Représentation des variations individuelles autour des valeurs moyennes estimées par linéarisation en fonction des valeurs simulées, pour les quatre paramètres dans les configurations faiblement (en haut) et fortement bruitées (en bas). Les nuages de points sont complétés par la droite de régression ajustée par moindres carrés.

Si ce sont les paramètres aléatoires qui sont structurés spatialement, alors les estimations individuelles retrouvent la structure simulée si les données ne sont pas fortement bruitées.

Il ne semble donc pas indispensable d'inclure explicitement dans le modèle à paramètres aléatoires la prise en compte de telles structures.

Chapitre 5

Application à des données réelles

Ce chapitre propose la mise en œuvre des modèles non linéaires à paramètres aléatoires sur des données réelles. Les données à haute résolution spatiale que nous utilisons nous permettent de créer deux jeux de données correspondant à des images de type Végétation pour l'estimation des paramètres du modèle d'évolution temporelle de l'indice de surface foliaire. Les estimations issues de la méthode de linéarisation sont comparées à des régressions non linéaires appliquées sur chaque pixel séparément.

5.1 Les données

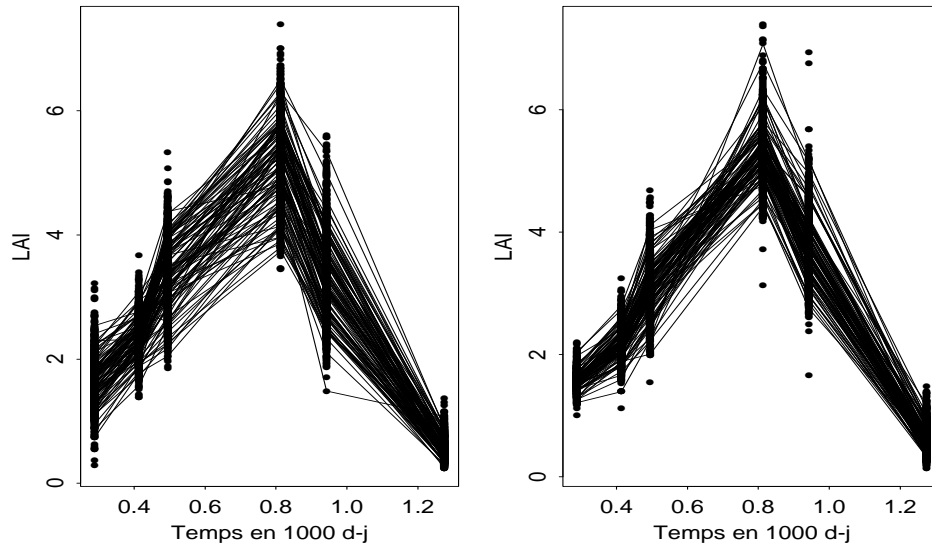
Les données proviennent de 6 scènes SPOT/HRV à haute résolution spatiale (pixel de $20\text{ m} \times 20\text{ m}$) prises entre le 19 mars et le 16 juillet 1996 (Tab. 5.1) dans les environs de la ville de Chartres.

Dates	19/03/96	04/04/96	19/04/96	05/06/96	16/06/96	16/07/96
Degré-jour	289	414	495	813	943	1275

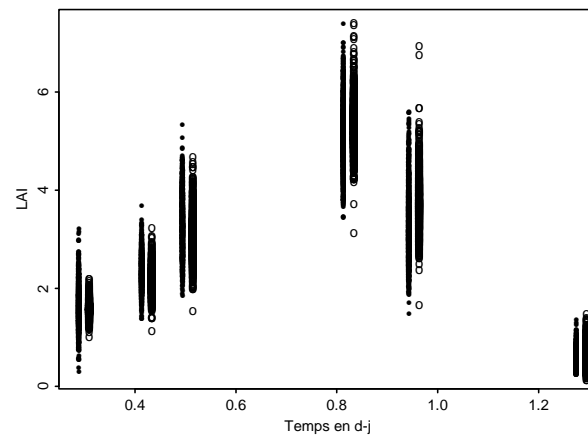
TAB. 5.1 – Équivalence entre les dates d'acquisition et les valeurs de somme de températures exprimées en degré-jour.

La région étudiée s'étend sur 1296 km^2 ($36\text{ km} \times 36\text{ km}$). Huit classes ont été répertoriées dans le plan d'occupation des sols: zone urbaine (4%), forêt (2%), cultures de printemps (3%), cultures d'été (3%), cultures permanentes (7%), pois (15%), colza (12%) et céréales d'hiver (53%). Nous nous focalisons sur l'étude de l'évolution des céréales d'hiver, thème qui représente la majeure partie du plan d'occupation des sols. Nous considérons directement les valeurs de LAI calculées à partir de ces réflectances *via* le PVI. Une sous-zone de la région observée de 120 km^2 est étudiée dans [31].

Grâce à la haute résolution spatiale des images que nous utilisons, nous mettons en œuvre deux stratégies pour obtenir deux jeux de données (Fig. 5.1) assimilables à des images de résolution kilométrique de type Végétation. Nous ne pouvons cependant pas remédier à la faible résolution temporelle des données initiales.



(a) Valeurs de LAI complétées par les courbes relatives à 100 pixels.



(b) Valeurs moyennées (●) et désagrégées (○).

FIG. 5.1 – Représentation des valeurs de LAI en fonction du temps exprimé en millier de degrés-jours : à gauche, les données moyennées, à droite les données désagrégées pour les 1205 pixels.

1. La première manipulation consiste à calculer les moyennes kilométriques des données à haute résolution spatiale du thème d'intérêt : $lai_i^c = \frac{1}{|L_i^c|} \sum_{l \in L_i^c} lai_{i,l}^c$ où lai_i^c est le LAI de la culture d'intérêt c sur le pixel kilométrique i , L_i^c est l'ensemble des pixels à haute résolution contenant la culture c sur le pixel kilométrique i , $|L_i^c|$ son cardinal et $lai_{i,l}^c$ le LAI de c sur le pixel l à haute résolution inclus dans le pixel de résolution kilométrique i . Nous faisons référence à ces données en tant que données moyennées.
2. La seconde méthode consiste à mettre en œuvre la procédure de désagrégation (Cf. section 1.3.3.2, page 24) sur les données relatives à l'ensemble des thèmes observés dégradées à la résolution VEGETATION :
 - Dégradation de l'image des valeurs de réflectances dans le rouge (R) et le proche infrarouge (PIR) à la résolution kilométrique : $R_i = \frac{1}{L} \sum_{l=1}^L R_{(i,l)}$ et $PIR_i = \frac{1}{L} \sum_{l=1}^L PIR_{(i,l)}$ pour $i = 1, \dots, I$, où R_i (resp. PIR_i) désigne la réflectance dans le rouge (resp. le proche infrarouge) sur le pixel kilométrique i et $R_{(i,l)}$ (resp. $PIR_{(i,l)}$) la réflectance dans le rouge (resp. le proche infrarouge) du pixel à haute résolution l contenu dans le pixel kilométrique i , avec L le nombre de pixels à haute résolution spatiale contenus dans un pixel kilométrique, soit $50 \times 50 = 2500$.
 - Désagrégation par paramètres aléatoires (Faivre et Fischer [34]) pour calculer les R_i^k (et les PIR_i^k) dans l'expression $R_i = \sum_{k=1}^{N_c} x_i^k R_i^k$ où les x_i^k sont connus par les images initiales non dégradées et N_c désigne le nombre de thèmes observés en mélange.

Nous récupérons ensuite les valeurs de LAI de la culture d'intérêt (lai_i^c) par conversion *via* le PVI.

Parmi les 1296 pixels kilométriques, nous en rejetons 91 pour lesquels le pourcentage d'occupation des sols du blé est inférieur à 10% car dans ces cas la désagrégation ne parvient pas à estimer les valeurs propres au blé.

On peut noter la bonne qualité de la désagrégation (Fig. 5.1) avec cependant une variabilité légèrement réduite des comportements (Fig. 5.1 (a)).

5.2 Traitement des données

Sur les deux jeux de données, nous ajustons le modèle de Baret à quatre paramètres (contraint par $T_r = \log(1 + \exp(a \times T_c)) / b$) en appliquant les deux procédures d'estimation déjà mises en œuvre sur données simulées (Cf. section 4.3) : la linéarisation (lin.) et les régressions par pixel (STS).

Ces deux méthodes sont initialisées selon la procédure suivante :

- Détermination de valeurs approximatives communes aux deux jeux de données d'après l'interprétation des paramètres du modèle d'évolution. Cette première phase donne les valeurs : $\beta_0^* = (K_0^*, a_0^*, b_0^*, T_{c_0}^*) = (7.4, 7.4, 2.4, 0.45)$.
- Amélioration de ces valeurs initiales (Fig. 5.2) pour chaque jeu de données par estimation des paramètres du modèle dans une régression non linéaire sur l'ensemble des pixels initialisée par β_0^* . On obtient alors les valeurs : $\beta_0^m = (7.91, 6.32, 2.10, 0.42)$

pour les données moyennées et $\beta_0^d = (8.63, 6.00, 2.19, 0.47)$ pour les données désagrégées.

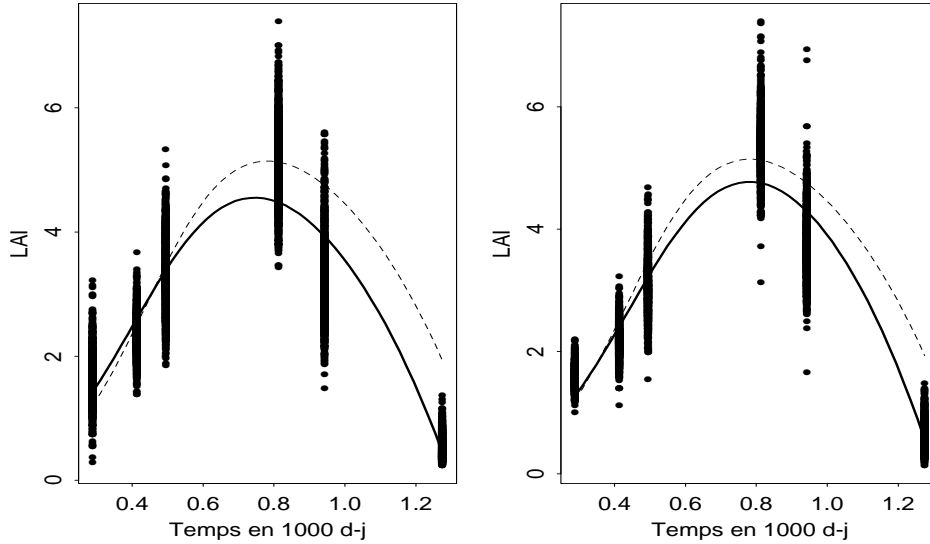


FIG. 5.2 – Représentation sur les données moyennées (à gauche) et désagrégées (à droite) de la courbe définie par les valeurs initiales issues de la régression non linéaire globale β_0^m et β_0^d (trait plein). La courbe en trait pointillé illustre l'initialisation de cette régression β_0^* .

5.3 Résultats

Les résultats sont présentés en deux parties. D'abord, des indicateurs statistiques chiffrés permettent de comparer les deux approches sur l'estimation des paramètres globaux et d'apprécier la qualité des estimateurs issus des régressions par pixel pour les deux jeux de données. Ensuite, les estimations individuelles sont comparées sur la base de graphiques.

5.3.1 Paramètres globaux

Les trois premières lignes des tableaux d'estimation des paramètres globaux ($\hat{\beta}$, $\hat{\sigma}_\beta$ et $\hat{\sigma}_\varepsilon$) sont les résultats de l'estimation par linéarisation ; ils représentent respectivement la moyenne et l'écart-type estimés pour chaque paramètre et l'écart-type résiduel. La matrice de variance des paramètres aléatoires est supposée diagonale pour l'estimation par linéarisation. En effet, un premier test laissant cette matrice non structurée n'a pas convergé à cause de l'inversion en cours d'algorithme d'une matrice singulière. L'initialisation des paramètres de variance n'a pas permis de lever cette difficulté.

Les quatre lignes suivantes sont les résultats des régressions par pixel. Ils comprennent :

- $\tilde{\beta}$: la moyenne des estimations par pixel des paramètres individuels,

- $\tilde{\Sigma}^{1/2}$: l'écart-type d'estimateur moyen des $\tilde{\beta}_i$,
- $s(\tilde{\beta}_i)$: l'écart-type empirique moyen des $\tilde{\beta}_i$,
- $\tilde{\sigma}_\varepsilon$: l'écart-type résiduel moyen.

Pour le calcul des quantités relatives à STS ($\tilde{\beta}$, $\tilde{\Sigma}^{1/2}$, $s(\tilde{\beta}_i)$ et $\tilde{\sigma}_\varepsilon$), seuls les pixels sur lesquels l'algorithme de minimisation de la régression a convergé sont conservés.

Moy.	K	a	b	T_c
$\hat{\beta}$	7.2	7.1	2.3	0.42
$\hat{\sigma}_\beta$	0.63	10^{-14}	10^{-17}	10^{-10}
$\hat{\sigma}_\varepsilon$	0.57			
$\tilde{\beta}$	9.2	6.4	2.2	0.43
$\tilde{\Sigma}^{1/2}$	15.0	6.6	1.8	0.12
$s(\tilde{\beta}_i)$	6.6	1.4	0.5	0.04
$\tilde{\sigma}_\varepsilon$	0.72			

Désag.	K	a	b	T_c
$\hat{\beta}$	8.1	6.4	2.3	0.46
$\hat{\sigma}_\beta$	0.4	10^{-4}	10^{-8}	10^{-9}
$\hat{\sigma}_\varepsilon$	0.49			
$\tilde{\beta}$	9.6	6.1	2.2	0.47
$\tilde{\Sigma}^{1/2}$	17.5	5.5	1.6	0.11
$s(\tilde{\beta}_i)$	13.3	0.6	0.2	0.06
$\tilde{\sigma}_\varepsilon$	0.66			

TAB. 5.2 – Estimation des paramètres globaux de moyenne et d'écart-type par nlme et nls sur les données moyennées (à gauche) et désagrégées (à droite) avec 4 paramètres.

5.3.1.1 Paramètres de moyenne β

Les problèmes de non convergence concernent 67 pixels sur les données moyennées et 63 sur les données désagrégées. Cette absence de convergence des régressions individuelles sur certains pixels révèle un handicap des régressions par pixels par rapport à l'estimation par linéarisation qui fournit systématiquement les prédictions individuelles pour tous les pixels. L'étude restreinte aux données relatives à ces pixels ne laisse apparaître aucune allure spécifique susceptible d'explicitier ces problèmes. Nous supposons que l'algorithme de minimisation utilisé dans les régressions ne parvient pas à déterminer le minimum une fois arrivé sur une zone relativement plate.

Comparaison des résultats : STS / linéarisation

Les estimations issues de la linéarisation sont nettement différentes de celles obtenues par STS (Tab. 5.2) : elles sont inférieures pour K et T_c et supérieures pour a . Les variations relatives des paramètres, quel soit le jeu de données, s'expliquent par leurs corrélations (Tab 5.3) : par exemple, la corrélation négative entre K et a illustre que quand K augmente, a diminue.

De plus, les courbes obtenues avec ces valeurs sont également nettement différentes (Fig. 5.3).

Ces valeurs s'expliquent par des valeurs élevées de K sur certains pixels (caractérisés par une croissance relativement lente : a faible, donc K élevé, Cf. section 4.3.2.4) qui font augmenter l'estimation moyenne globale de K (diminuer cette même quantité pour a et b et augmenter pour T_c).

	K	a	b		K	a	b
a	-0.97			a	-0.99		
b	-0.95	0.98		b	-0.96	0.99	
T_c	0.57	-0.54	-0.39	T_c	0.80	-0.78	-0.67

TAB. 5.3 – Corrélation entre les estimations des 4 paramètres selon nlme : à gauche sur les données moyennées, à droite sur les données désagrégées.

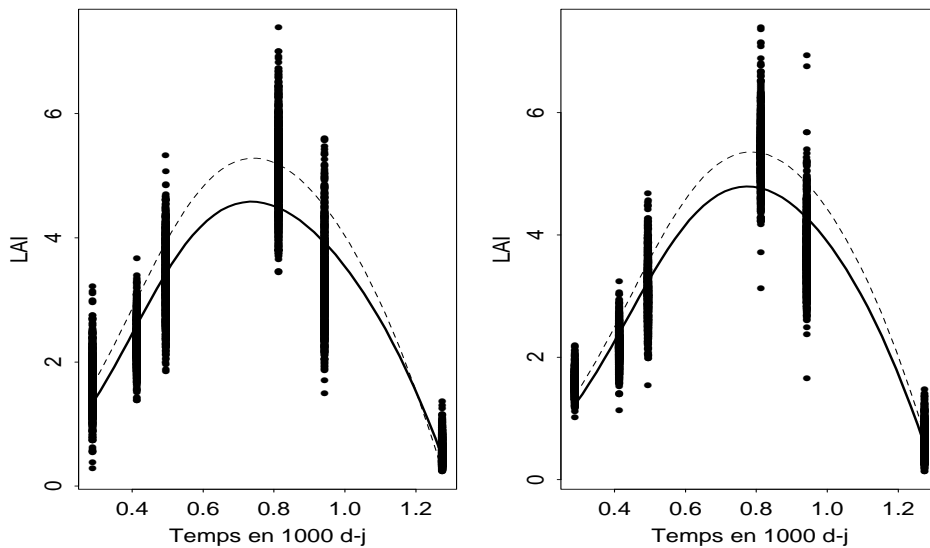


FIG. 5.3 – Courbes tracées avec les estimations des paramètres de moyenne par linéarisation (ligne pointillée) et STS (ligne continue) sur les données moyennées (à gauche) et désagrégées (à droite).

En effet, les valeurs médianes des estimations issues des régressions par pixel sont (8.0,6.4,2.2,0.4) sur les données moyennées. Ce qui indique que seuls quelques pixels sont responsables de l'écart important entre la linéarisation et STS.

Par ailleurs, les valeurs obtenues par linéarisation produisent des courbes quasiment superposables avec celles tracées à partir des valeurs initiales résultant de la régression sur l'ensemble des pixels malgré des valeurs différentes ($\beta_0^m = (7.91, 6.32, 2.10, 0.42)$ pour les données moyennées et $\beta_0^d = 8.63, 6.00, 2.19, 0.47$) pour les données désagrégées).

Nous en déduisons que le comportement de certains pixels peut venir fausser l'estimation globale des paramètres de moyennes par STS alors que l'estimation par linéarisation présente une plus grande robustesse.

Comparaison des résultats: moyenne / désagrégation

La valeur estimée du paramètre K sur les données moyennées est supérieure à celle estimée sur les données désagrégées quelle soit la méthode (Tab. 5.2). Les relations entre les autres paramètres sont cohérentes avec les tableaux des corrélations entre les estimations des paramètres (Tab. 5.3). Ces différences sont illustrées sur la figure 5.4.

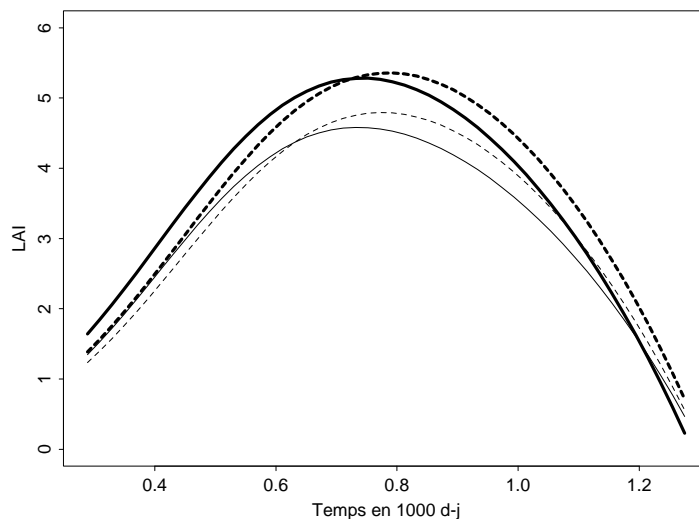


FIG. 5.4 – Courbes tracées à partir des estimations de β : ligne continue (resp. pointillée) pour les données moyennées (resp. désagrégées), trait fin (resp. épais) pour la linéarisation (resp. STS).

La représentation des deux jeux de données (Fig. 5.1) illustre que les données désagrégées atteignent un LAI maximum plus élevé que les données moyennées. Ce qui a tendance à faire augmenter la valeur estimée de K . De plus, les valeurs globalement plus élevées à la cinquième date implique une sénescence plus lente *i.e.* un b plus faible.

5.3.1.2 Écart-type des paramètres σ_β

Quelles que soient les données, les seules informations sur la variance des paramètres nous sont fournies par la méthode de linéarisation. En effet, pour les régressions par pixel, les grandes valeurs de $\tilde{\Sigma}^{1/2}$ dues essentiellement au faible nombre de dates par rapport au nombre de paramètres à estimer, ne permettent pas de débiaiser l'estimation de la variance.

Les valeurs de $\hat{\sigma}_\beta$ montrent que la linéarisation concentre la variabilité sur le seul paramètre K . Les autres estimations des écarts-types, inférieures à 10^{-4} , peuvent être considérées comme nulles. Les corrélations très élevées entre les estimations des paramètres est certainement la raison pour laquelle, l'essentiel de la variabilité est concentrée sur un seul paramètre : agir sur un paramètre permet de modifier l'allure générale de la courbe aussi bien qu'en modifiant tous les paramètres. Le report sur le paramètre K est peut-être dû au fait que ce paramètre entre dans le modèle de Baret par un produit alors que les autres paramètres figurent en exponentielle et sont donc contraints à des variations plus limitées.

Ce qui apparaît également dans la comparaison des deux tableaux 5.2 est la diminution de l'écart-type $\hat{\sigma}_\beta$ des paramètres lors de l'estimation sur les données désagrégées si nous considérons les valeurs inférieures à 10^{-4} comme nulles. L'écart type $\hat{\sigma}_K$ passe de 0.63 à 0.4 sur les données désagrégées. Ceci semble cohérent avec l'aspect plus resserré des

courbes issues de la désagrégation (Fig. 5.1 (a)).

5.3.1.3 Écart-type résiduel σ_ϵ

De manière générale, l'écart-type résiduel est relativement élevé, égal ou supérieur à celui simulé dans les configurations fortement bruitées (chapitre 4).

Des problèmes liés à la superposition temporelle des images et éventuellement à des corrections atmosphériques non homogènes peuvent expliquer cette valeur élevée sur les données moyennées. Ces effets semblent en partie réduits par la désagrégation.

Par ailleurs, dans les deux cas, le terme résiduel est aussi alimenté par une erreur probable de modélisation.

La valeur plus élevée de l'écart-type résiduel par STS est certainement due au fait qu'il s'agit d'une moyenne d'écart-types résiduels relativement élevés car calculés pour des régressions visant à estimer quatre paramètres avec seulement six dates.

5.3.2 Paramètres individuels

Dans cette section, nous nous intéressons à la comparaison des méthodes d'estimation par linéarisation et par régressions par pixel pour l'estimation des paramètres individuels.

Nous nous focalisons sur le paramètre K qui concentre la variabilité du vecteur des paramètres.

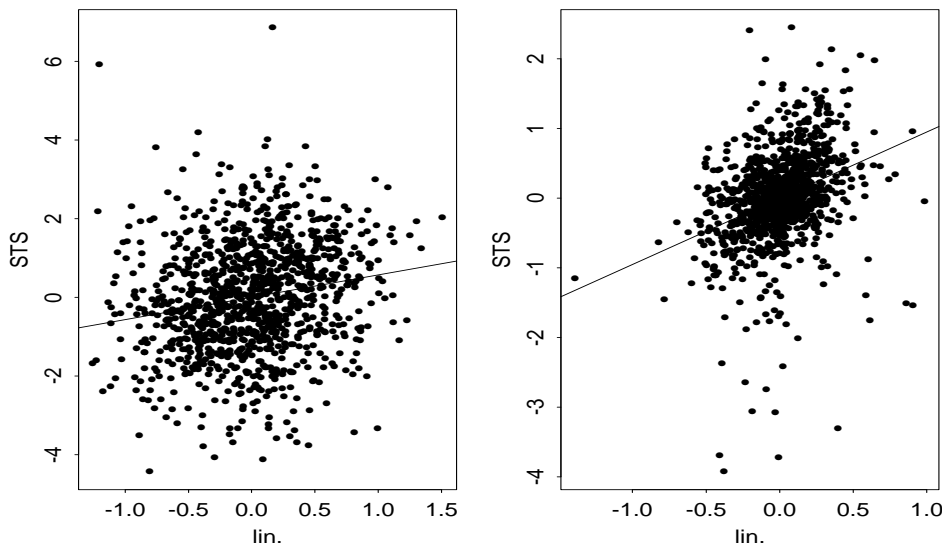


FIG. 5.5 – Représentation des variations individuelles autour de la valeur moyenne prédites par STS en fonction des mêmes quantités estimées par linéarisation (lin.) pour le paramètre K complétée par la droite de régression ajustée par moindres carrés aux nuages de points sur les données moyennées (à gauche) et désagrégées (à droite).

La représentation des estimations individuelles de STS en fonction des résultats de la linéarisation pour le paramètre K (Fig. 5.5) indique que les deux méthodes ne varient pas

vraiment dans le même sens, malgré les droites de régression de pente positive.

Au vu de ces graphiques, nous pouvons cependant dire que l'étendue des valeurs estimées par linéarisation est plus faible que celle des résultats des régressions par pixel.

Les difficultés d'estimation du paramètre K sont certainement liées à la répartition des dates d'acquisition. En effet, on peut constater un écart important entre la troisième et la quatrième date (Fig 5.1, page 108), période à laquelle l'indice de surface foliaire atteint ses valeurs maximales ; or, ce sont ces valeurs qui donnent les informations pertinentes pour estimer K qui s'interprète, rappelons-le, comme un majorant du LAI. Dans ces conditions, ces premiers résultats sur données réelles ne remettent pas en cause les résultats des simulations mais illustrent la nécessité de disposer d'acquisitions fréquentes et justifient le recours à des capteurs à haute résolution temporelle / basse résolution spatiale.

5.4 Modification du modèle

Pour pousser plus loin l'étude des données réelles, nous cherchons à neutraliser l'effet des corrélations entre les estimations des paramètres. Pour cela, nous conservons seulement les deux paramètres de pente (a et b) dans le modèle. Nous fixons le point d'inflexion de la phase de croissance, relativement stable sur l'ensemble des pixels pour les deux jeux de données, en posant $T_c = 0.45$. Bien que supportant la majorité de la variance des paramètres dans l'étude précédente, nous fixons le paramètre K pour les raisons évoquées dans les simulations : ce paramètre est susceptible d'être caractéristique de la variété culturale supposée unique sur la région étudiée ; nous fixons sa valeur à 8.5.

5.4.1 Résultats

5.4.1.1 Paramètres globaux

Moy.	a	b	Désag.	a	b
$\hat{\beta}$	5.9	2.1	$\hat{\beta}$	6.0	2.1
$\hat{\sigma}_\beta$	0.56	0.18	$\hat{\sigma}_\beta$	0.33	0.11
$\hat{\sigma}_\varepsilon$	0.58		$\hat{\sigma}_\varepsilon$	0.50	
$\tilde{\beta}$	6.0	2.1	$\tilde{\beta}$	6.6	2.4
$\tilde{\Sigma}^{1/2}$	0.61	0.23	$\tilde{\Sigma}^{1/2}$	0.95	0.48
$s(\tilde{\beta}_i)$	0.81	0.30	$s(\tilde{\beta}_i)$	0.81	0.40
$\tilde{\sigma}_\beta$	0.53	0.19	$\tilde{\sigma}_\beta$	///	///
$\tilde{\sigma}_\varepsilon$	0.61		$\tilde{\sigma}_\varepsilon$	0.54	

TAB. 5.4 – Estimation des paramètres globaux de moyenne et d'écart-types par linéarisation et STS sur les données moyennées (en haut), désagrégées (en bas).

Nous nous focalisons sur l'estimation des paramètres d'écart-type car les commentaires des résultats sur les paramètres de moyenne sont globalement similaires à ceux de l'étude à quatre paramètres.

Sur les données moyennées, on peut débiaiser les estimations des σ_β . On constate alors que les résultats sont très proches de ceux obtenus par linéarisation. La variance des paramètres est essentiellement reportée sur le paramètre a ce qui semble cohérent avec les données (Fig. 5.1 (a)) qui sont plus variables durant la croissance que pendant la sénescence notamment sur les données moyennées. Sur les données désagrégées, on constate une diminution globale de l'écart-type des paramètres comme sur les résultats de l'estimation de quatre paramètres.

Par rapport aux résultats précédents (Tab. 5.2), l'écart-type résiduel est globalement inférieur. La diminution du nombre de paramètres à estimer est la raison de cette diminution car σ_ϵ est proportionnel à $(N - p)^{-1}$ où N est le nombre d'observations qui ne varie pas entre nos deux configurations et p le nombre de paramètres à estimer qui passe de quatre à deux.

5.4.1.2 Paramètres individuels

La représentation des résultats de la linéarisation en fonction de ceux des régressions par pixel (Fig. 5.6) illustrent plus nettement que pour le paramètre K (Fig. 5.5) que les deux méthodes obtiennent des résultats qui varient dans le même sens. L'allure des nuages de points est clairement allongée selon une droite de pente positive.

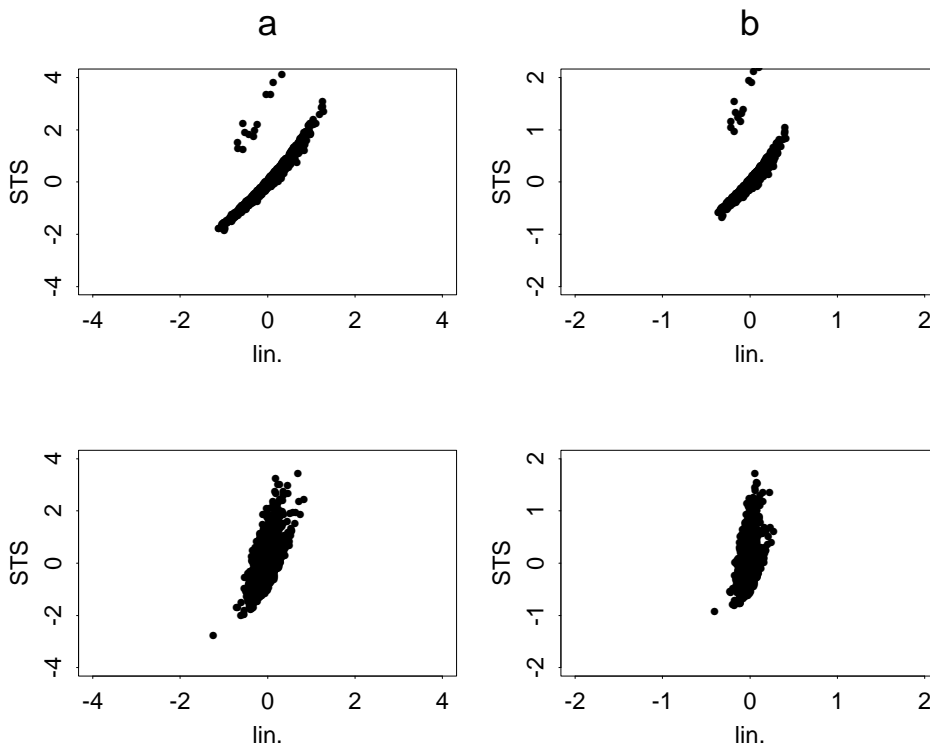


FIG. 5.6 – Représentation des variations individuelles autour de la valeur moyenne prédites par STS en fonction des mêmes quantités estimées par linéarisation (lin.) pour les paramètres a et b sur les données moyennées (à gauche) et désagrégées (à droite).

Sur les données moyennées, parmi les 1205 pixels, 13 ont un comportement particulier :

ce sont les points représentatifs de valeurs estimées entre -0.2 et 0.2 par linéarisation et supérieures à 1 par STS. La représentation des courbes relatives à ces pixels (Fig. 5.7) montre que les 13 points en question se caractérisent par une faible valeur de LAI à la sixième date. Ce qui implique une sénescence relativement rapide c'est-à-dire un b élevé. Dans ces cas, la linéarisation maintient les valeurs de b relativement proche des autres pixels alors que les régressions par pixel estiment des valeurs significativement différentes. La forte corrélation entre les paramètres a et b implique des comportements similaires sur les deux paramètres.

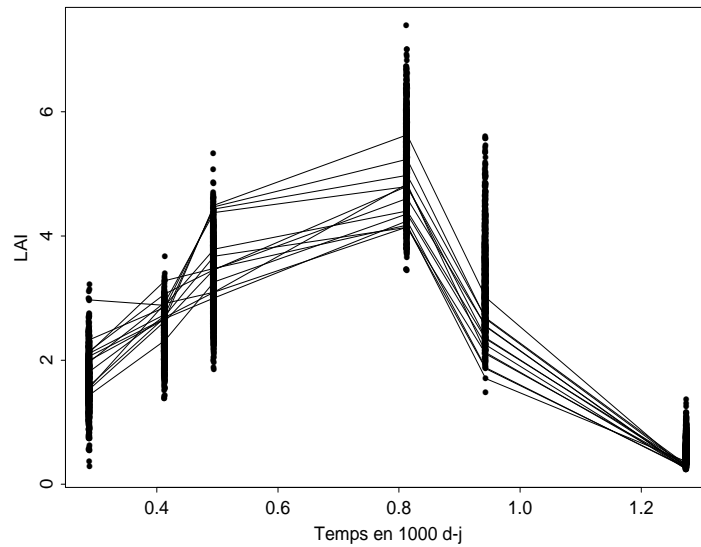


FIG. 5.7 – Représentation des courbes liant les valeurs de LAI des 13 pixels distingués sur les résultats comparés de la linéarisation et STS (Fig. 5.6) sur les données moyennées.

Sur les données désagrégées, les résultats des deux méthodes d'estimation ne présente pas une allure aussi allongée que sur les données moyennées. Cependant, il semble que l'on puisse distinguer deux sous-groupes dans les nuages de points : un premier très étiré comme sur les données moyennées, un second, correspondant aux 13 pixels distingués précédemment, caractérisé par des valeurs estimées par STS relativement élevées par rapport aux résultats de la linéarisation.

D'une manière générale, au vu des échelles des axes des graphiques de la figure 5.6, les estimations issues de la linéarisation apparaissent moins variables que les estimations par pixel. Ce comportement semble se rapprocher d'estimateurs à rétrécisseurs (Hébel *et al.* [44]).

5.5 Conclusion

Nous avons vu que quelle que soit la méthode d'estimation employée et quel que soit le jeu de données (moyenné ou désagrégé), l'écart-type résiduel est toujours de l'ordre de celui simulé dans les configurations bruitées au chapitre 4. Dans ces conditions, nous

avons conclu que l'estimation par linéarisation était préférable car plus robuste vis-à-vis d'un bruit important. Pour cette raison, dans la pratique, nous conseillons d'avoir recours à une procédure d'estimation comme la linéarisation traitant l'ensemble des pixels simultanément plutôt que pixel par pixel.

Pour conclure l'étude des données réelles nous proposons une utilisation potentielle de l'estimation des paramètres du modèle d'évolution du LAI : cartographier les estimations des paramètres permet de détecter rapidement des zones d'évolution différente (Fig. 5.8).

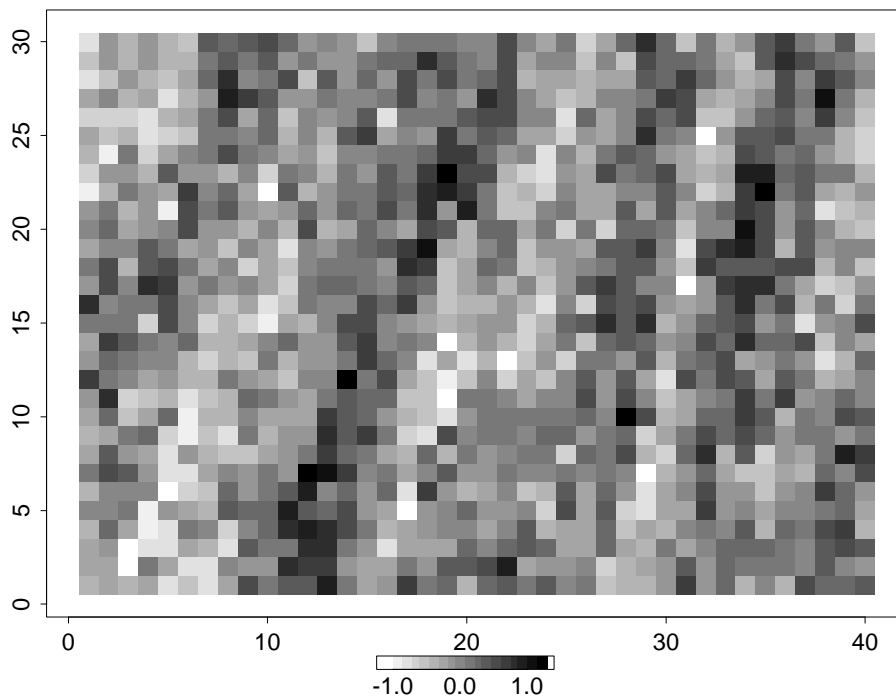


FIG. 5.8 – Simulation d'une cartographie des variations autour de la valeur moyenne des estimations par linéarisation du paramètre a avec 4 paramètres sur les données moyennées. 5 valeurs ont été retirés pour pouvoir afficher une image 40×30 .

En représentant les variations autour de la moyenne, l'image offre une vision des variations en positif ou en négatif par rapport au comportement moyen sur la zone étudiée. Les agronomes traduisent cette information sur les variations des paramètres en termes de différences dans le développement des cultures. Cette étape dans le traitement des données permet d'initier la mise en œuvre d'une enquête plus approfondie sur le terrain.

Sur la figure 5.8, on peut noter par exemple que le coin en haut à gauche composé d'environ 10×10 pixels, est relativement clair. Ceci signifie que les valeurs estimées du paramètre a sont plus faible que la moyenne dans cette zone. D'un point de vue agronomique, on peut ainsi dire que les céréales d'hiver ont été soumises à un stress, sécheresse par exemple, qui a limité leur vitesse de croissance. Ce constat permet alors d'étudier la mise en œuvre d'outils d'irrigation dans cette zone pour les saisons à venir.

De plus, le variogramme (Fig. 5.9) des estimations de K représentées sur la figure 5.8

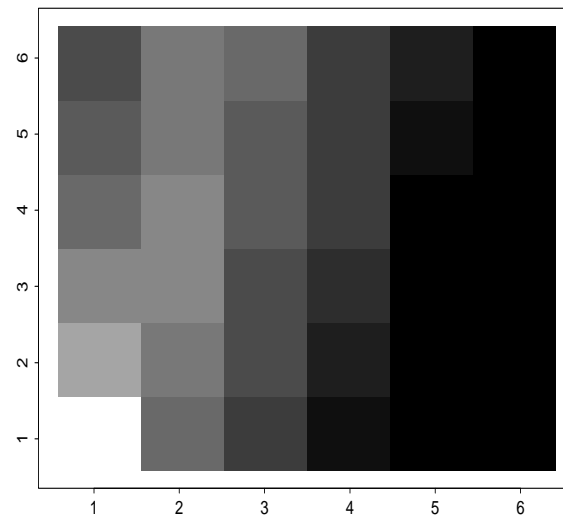


FIG. 5.9 – Variogramme des données de la figure 5.8.

confirme que les variations spatiales ne sont pas isotropes conformément aux structures linéaires obliques proches de la verticale que l'on voit sur la figure 5.8. Cet indice peut également initier une enquête terrain pour expliquer ce phénomène.

Conclusion et perspectives

1 Conclusion

En conclusion de ce mémoire, nous re-situons les travaux de recherches à l'interface des domaines de l'agronomie, de la télédétection spatiale et de la statistique. Les enjeux du problème qui a initié la thèse sont essentiellement d'ordre agronomique : ils sont liés à la prévision du rendement d'une culture. La télédétection apparaît comme un moyen susceptible de contribuer à atteindre cet objectif par un recueil sûr et efficace de données assurant un suivi journalier de la culture. La statistique intervient alors pour exploiter au mieux cette source d'information.

Devant les lacunes de la méthodologie actuelle qui traite séparément l'évolution de chaque pixel, nous proposons l'utilisation d'un modèle à paramètres aléatoires capable de traiter l'ensemble des pixels simultanément et d'exprimer explicitement les variabilités intra-pixel au cours du temps et inter-pixels.

L'originalité des travaux de recherches qui en découlent réside d'une part, d'un point de vue applicatif, dans l'apport d'une nouvelle méthodologie pour le traitement de données de télédétection, d'autre part, d'un point de vue méthodologique dans la contribution à l'étude des modèles non linéaires à paramètres aléatoires.

Application

Les plans d'expérience mis en œuvre pour l'étude par simulation informe les utilisateurs de données satellites sur le compromis entre quantité et qualité des images à acquérir ainsi que sur la méthode à employer pour estimer les paramètres d'un modèle d'évolution temporelle : l'estimation par linéarisation dans un modèle à paramètres aléatoires est plus robuste qu'un ensemble de régressions non linéaires individuelles face à des conditions défavorables (faible nombre de dates, bruit résiduel important).

Méthodologie

Au niveau méthodologique, nous contribuons à l'étude des modèles non linéaires à paramètres aléatoires en présentant une revue bibliographique détaillée des diverses spécifications rencontrées dans la littérature et des méthodes d'estimation dans le cadre paramétrique. Nous complétons cette revue bibliographique par une classification des méthodes d'estimation fondées sur le maximum de vraisemblance. Nous proposons également une synthèse des propriétés théoriques des estimateurs et nous mettons ainsi en évidence deux situations asymptotiques différentes nécessaires selon les méthodes pour obtenir la convergence : le nombre d'individus et le nombre d'observations par individu tendent vers l'infini ou bien le nombre d'observations par individu peut être uniformément borné. Ces

considérations révèlent un paradoxe quant à l'utilisation des modèles non linéaires à paramètres aléatoires : on justifie parfois leur utilisation par un faible nombre d'observations par individu alors que l'on estime avec une méthode (la linéarisation, Lindstrom et Bates [54]) dont l'estimateur est convergent asymptotiquement sur le nombre d'observations par individu.

L'apport des simulations sur le plan méthodologique réside essentiellement dans l'application des modèles non linéaires à paramètres aléatoires à des problèmes de grande taille. Les exemples traités dans la littérature sont généralement petits par rapport à notre problème de suivi de cultures par images satellites. Dans ce cadre, le choix d'une méthode d'estimation est largement contraint par son aspect opérationnel.

2 Perspectives

Dans les perspectives de ce travail, il est essentiel d'envisager une poursuite de l'étude autour des modèles non linéaires mixtes. De nombreux problèmes sont ouverts. Citons par exemple, la définition de tests sur la matrice de variance des paramètres pour déterminer le caractère fixe ou aléatoire d'un paramètre (Bonneu et Hyrien [8]) et la mise en œuvre de méthodes de sélection des covariables permettant de décrire les variations des paramètres aléatoires (Concordet [14]).

L'élaboration de méthodes d'estimation avec de bonnes propriétés théoriques à distances finies est également souhaitable pour rompre le paradoxe qui consiste à se tourner vers les modèles à paramètres aléatoires pour des raisons liées au manque d'observations par individu et à utiliser ensuite une procédure d'estimation dont les propriétés à distance finie ne sont pas garanties.

Les autres perspectives que l'on peut considérer se situent dans le domaine applicatif. On peut notamment envisager d'employer les modèles à paramètres aléatoires dans d'autres cas de figures que le suivi de cultures tout en restant en relation avec la télédétection. L'exemple *a priori* directement exploitable est l'étude de la croissance des forêts pour lesquelles il existe des modèles d'évolution. Le phénomène naturel étudié dans ce cas a une dynamique différente de l'évolution d'une culture, ce qui peut induire une problématique différente. La différence réside également dans la surface au sol occupée par un forêt par rapport à celle d'une parcelle agricole : se pose alors le problème du choix de la résolution des images à acquérir.

Que ce soit pour le suivi de cultures ou un autre problème lié à l'évolution temporelle d'un processus, un question intéressante consiste à déterminer au mieux les dates d'acquisition des images satellites. L'intérêt en termes de coût est évident pour celui qui dispose d'un budget limité pour acheter des images. Ce problème s'insère directement dans la planification expérimentale dans le cadre des modèles mixtes non linéaires.

Une autre piste pour la poursuite des travaux est suggérée par l'absence de prise en compte dans nos travaux d'une structure de voisinage dans les divers traitements des données. Même si l'estimation par linéarisation dans un modèle à paramètres aléatoires permet de retrouver des zones homogènes de paramètres sur une image, il semble que l'expression explicite d'une structure de voisinage inter-pixels pour traduire des comportements proches entre des pixels voisins puisse fournir une information supplémentaire

dans l'interprétation des résultats.

Pour étendre l'application des modèles à paramètres aléatoires au delà de l'usage post-désagrégation, on peut envisager de coupler les phases de désagrégation et d'estimation dans un modèle temporel d'évolution en un seul modèle. Formellement, l'écriture d'une tel modèle ne pose pas de problèmes et peut être inspirée de Fischer [35] où le mélange de deux thèmes est étudiée par des régressions non linéaires. En pratique, la non disponibilité de modèles d'évolution du LAI pour l'ensemble des thèmes observés en mélange nécessite des adaptations et le recours éventuel à des techniques issues du non paramétrique.

Il serait également intéressant de lier les problèmes rencontrés dans le cadre du suivi de cultures avec d'autres domaines. En épidémiologie par exemple, les questions que l'on peut se poser sont assez proches de notre situation. Est-il préférable de contraindre le patient à passer un examen à l'hôpital et obtenir ainsi une information précise mais rare car contraignante ou bien de confier certaines mesures au patient et obtenir une information plus fréquente mais forcément moins fiable? La mise en relation de ce type de problématiques est susceptible d'apporter un éclairage intéressant sur les divers domaines d'étude.

Bibliographie

- [1] Anderson, T. W. – Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, vol. 1, 1973, pp. 135–141.
- [2] Baret F. – *Contribution au suivi radiométrique de culture de céréales*. – Thèse d'université, Université de Paris-Sud Orsay, 1986.
- [3] Baret F. et Guyot G. – Potentials and limits of vegetation indices for lai and apar assessment. *Remote Sensing of Environment*, vol. 35, 1991, pp. 161–173.
- [4] Beal, S. L. et Sheiner, L. B. – The NONMEM system. *American Statistician*, vol. 34, 1980, pp. 118–119.
- [5] Beal, S. L. et Sheiner, L. B. – Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, vol. 8, 1982, pp. 195–222.
- [6] Beal, S. L. et Sheiner, L. B. – *NONMEM User's Guides*. – NONMEM Project Group, University of California, San Fransisco, 1992.
- [7] Bonn F. et Rochon G. – *Précis de télédétection: Volume 1, principes et méthodes*. – Presses de l'Université du Québec, AUPELF, 1991.
- [8] Bonneu M. et Hyrien O. – Test dans les modèles non lineaires à effets mixtes. *In: XXXI Journées de Statistique, 1999*. Société Française de Statistique, pp. 263–266.
- [9] Bouman, B. A. – The linking of crop growth models and multi-sensor remote sensing data. *In: Proceedings Fifth International Colloquium on Physical Measurements and Sig natures and Remote Sensing, 1991*. ESA SP-319, pp. 583–588.
- [10] Brisson N., Mary B., Ripoche D., Jeuffroy M-H., Ruget F., Nicoulaud B., Gate P., Devienne-Barret F., Antonioletti R., Durr C., Richard G., Beaudoin N., Recous S., Tayot X., Plenet D., Cellier P., Machet J-M., Meynard J-M. et Delécolle R. – Stics: a generic model for the simulation of crops and their water and nitrogen balances. 1. theory and parameterization applied to wheat and corn. *Agronomie*, vol. 18, 1998, pp. 311–346.
- [11] Centre canadien de télédétection. – Notions fondamentales de télédétection, 2001. <http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/tutorf.html>.
- [12] Chan, K. S. et Ledolter J. – Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, vol. 90, 1995, pp. 242–252.

- [13] Cherchali S., Amram O. et Flouzat G. – Retrieval of temporal profiles of reflectances from simulated and real NOAA-AVHRR data over heterogeneous landscapes. *International Journal of Remote Sensing*, vol. 21, n4, 2000, pp. 753–775.
- [14] Concordet D. – Modèle non linéaire à effets mixtes : quelques questions ouvertes. *In: XXXIII Journées de Statistiques, 2001*, pp. 43–48.
- [15] Concordet D. et Nuñez O. – A simulated pseudo-maximum likelihood estimator for nonlinear mixed models. *Computational Statistics and Data Analysis*, à paraître 2002.
- [16] Cox, D. R. et Hinkley, D. V. – *Theoretical statistics*. – Chapman & Hall, 1974.
- [17] Cressie, N. A. – *Statistics for Spatial Data*. – John Wiley & Sons, 1993, *Wiley series in probability and mathematical statistics*.
- [18] Dacunha-Castelle D. et Duflo M. – *Probabilités et statistiques*. – Masson, 1983 volume 2. Problèmes à temps mobiles.
- [19] Dahlquist G., Bjorck A. et Anderson N. – *Numerical methods*. – Prentice-Hall, 1974, *Prentice-Hall series in automatic computation*.
- [20] Davidian M. et Gallant, R. A. – Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 20, 1992, pp. 529–555.
- [21] Davidian M. et Gallant, R. A. – The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, vol. 80, 1993, pp. 475–488.
- [22] Davidian M. et Giltinian D.M. – *Nonlinear Models for Repeated Measurement Data*. – Chapman & Hall, 1995, *Monographs on Statistics and Applied Probability*.
- [23] Davis, P. J. et Rabinowitz P. – *Methods of Numerical Integration*. – Academic Press, 1984, *Computer Science and Applied Mathematics*.
- [24] Deering, D. W., Rouse, J. W., Haas, R. H. et Schell, J. A. – Measuring forage production of grazing units from landsat MSS data. *In: Proceedings 10th International Symposium Remote Sensing of Environment, 1975*, pp. 1169–1178.
- [25] Delécolle R., Maas S., Guérif M. et Baret F. – Remote sensing and crop production models. present trends. *ISPRS, Photogrammetry and Remote Sensing*, vol. 47, 1992, pp. 154–161.
- [26] Demailly J-P. – *Analyse numérique et équations différentielles*. – Presses Universitaires de Grenoble, 1991, *Collection Grenoble Sciences*.
- [27] Demidenko E. – Asymptotic properties of nonlinear mixed-effects models. *Lecture notes in statistics*, vol. 122, 1997, pp. 49–62.
- [28] Demidenko E. – Asymptotic properties of nonlinear mixed-effects models, 1998. <http://www.math.dartmouth.edu/~m50w99/resume.html>.

- [29] Dempster, A. P., Laird, N. M. et Rubin, D. B. – Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, vol. 39, 1977, pp. 1–38.
- [30] Déjean S., Faivre R. et Goulard M. – Estimation dans les modèles mixtes non linéaires : revue de méthodes en vue du traitement de données satellites. In: *XXXII Journées de Statistiques, 2000*, pp. 287–290.
- [31] Déjean S., Faivre R. et Goulard M. – Modèle non linéaire à paramètres fixes ou aléatoires de l'évolution des cultures observées par télédétection. *Journal de la Société Française de Statistique*, 2002.
- [32] Faivre R. et Delécolle R. – Synthetic map of crop leaf area index dynamics estimated with satellite data. In: *MODSIM'97: International congress on modelling and simulation, 1997*, éd. par McDONALD A.D. et ALEER M. Mc., pp. 228–233.
- [33] Faivre R., Delécolle R. et Guérif M. – Synthetic mapping of crop dynamics with pseudo-vegetation data. In: *ALPS'99 conference, Méribel, France, 1999*, éd. par CNES.
- [34] Faivre R. et Fischer A. – Predicting crop reflectances using data observing mixed pixels. *Journal of Agricultural, Biological and Environmental Statistics*, vol. 2, n1, 1997, pp. 87–107.
- [35] Fischer A. – A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. *Remote Sensing of Environment*, vol. 48, 1994, pp. 220–230.
- [36] Gallant, A. R. – *Nonlinear statistical models*. – John Wiley & Sons, 1987, *Wiley series in probability and mathematical statistics*.
- [37] Gallant, A. R. et Nychka, D. W. – Semiparametric maximum likelihood estimation. *Econometrica*, vol. 55, 1987, pp. 363–390.
- [38] Gelfand, A. E. et Smith, A. F. – Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, vol. 85, 1990, pp. 398–409.
- [39] Geman S. et Geman D. – Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 6, 1984, pp. 721–741.
- [40] Geweke J. – Bayesian inference in econometric models using monte carlo integration. *Econometrica*, vol. 57, 1989, pp. 1317–1339.
- [41] Girard M-C. et Girard C. – *Traitement des données de télédétection*. – Dunod, 1995.
- [42] Graybill, F. A. – *Theory and application of the linear model*. – Duxbury press, 1976.
- [43] Harville, D. A. – Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, vol. 72, 1977, pp. 320–340.

- [44] Hébel P., Faivre R., Goffinet B. et Wallach D. – Shrinkage estimators applied to prediction of french winter wheat yield. *Biometrics*, vol. 49, 1993, pp. 281–293.
- [45] Henderson, C. R. – Best linear unbiased estimation and prediction under a selection model. *Biometrics*, vol. 31, 1975, pp. 247–268.
- [46] Huber, P. J. – *Robust Statistics*. – New York: Wiley, 1981.
- [47] Huet S., Bouvier A., Gruet M-A. et Jolivet E. – *Statisticals Tools for Nonlinear Regression*. – Springer, 1996, *Springer Series in Statistics*.
- [48] Huet S., Jolivet E. et Messean A. – *La régression non-linéaire, méthodes et applications en biologie*. – INRA Editions, 1992.
- [49] Jamshidian M. et Jennrich, R. I. – Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, vol. 88, 1993, pp. 221–228.
- [50] Knippling, E. B. – Physical and physiological bases for the reflectance of visible and near infrared radiation from vegetation. *Remote Sensing of Environment*, vol. 1, 1970, pp. 155–159.
- [51] Lapeyre B., Pardoux E. et Sentis R. – *Méthodes de Monte-Carlo pour les équations de transport et de diffusion*. – Springer, 1998, *Mathématiques & Applications*.
- [52] Leonard T., Hsu, J. S. et Tsui, K. W. – Bayesian marginal inference. *Journal of the American Statistical Association*, vol. 84, 1989, pp. 963–974.
- [53] Lindley, D. V. et Smith, A. F. – Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society B*, vol. 34, 1972, pp. 1–42.
- [54] Lindstrom, M. J. et Bates, D. M. – Nonlinear mixed effects models for repeated measures data. *Biometrics*, vol. 46, 1990, pp. 87–107.
- [55] Littell R., Milliken G., Stroup W. et Wolfinger R. – *SAS System for Mixed Models*. – SAS Institute Inc., 1996.
- [56] Louis, T. A. – Using empirical Bayes methods in biopharmaceutical research. *Statistics in Medicine*, vol. 10, 1991, pp. 811–829.
- [57] Maitre, P. O., Buhner M., Thomson D. et Stanski, D. R. – A three-step approach combining bayesian regression and NONMEM population analysis: application to midazolam. *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 19, 1991, pp. 377–384.
- [58] Mallet A. – A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, vol. 73, 1986, pp. 645–656.
- [59] Mallet A., Mentré F., Steimer J-L. et Lokiec F. – Nonparametric maximum likelihood estimation for population pharmacokinetics, with application to cyclosporine. *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 16, 1988, pp. 311–327.

- [60] Mandema, J. W., Verotta D. et Sheiner, L. B. – Building population pharmacokinetic/pharmacodynamic models. *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 20, 1992, pp. 511–529.
- [61] Maselli F., Rodolfi A. et Conese C. – Fuzzy classification of spatially degraded Thematic Mapper data for the estimation of sub-pixel components. *International journal of remote sensing*, vol. 17, 1996, pp. 537–551.
- [62] McLachlan, G. J. et Krishnan T. – *The EM algorithm and extensions*. – John Wiley & Sons, 1997, *Wiley series in probability and statistics*.
- [63] Meng X-L. et Rubin, D. B. – Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, vol. 80, 1993, pp. 267–278.
- [64] Nuñez O. – *Estimation des modèles non linéaires mixtes*. – Thèse d’université, Université Paul Sabatier, 1998.
- [65] Pinheiro, J. C. – Model building using covariates in nonlinear mixed-effects models. *In: Modèles mixtes et biométrie, 2002*. Société Française de Biométrie.
- [66] Pinheiro, J. C. et Bates, D. M. – Approximations to the log-likelihood function in the nonlinear mixed-effects models. *Journal of Computational and Graphical Statistics*, vol. 4, 1995, pp. 12–35.
- [67] Pinheiro, J. C. et Bates, D. M. – *Mixed-Effects Models in S and S-PLUS*. – Springer, 2000, *Statistics and Computing*.
- [68] Racine-Poon A. – Bayesian approach to nonlinear random effects models. *Biometrics*, vol. 41, 1985, pp. 1015–1023.
- [69] Ramos, R. Q. et Pantula, S. G. – Estimation of nonlinear random coefficient models. *Statistics & Probability Letters*, vol. 24, 1995, pp. 49–56.
- [70] Rappaz J. et Picasso M. – *Introduction à l’analyse numérique*. – Presses polytechniques et universitaires romandes, 1998.
- [71] Richardson, A. J. et Wiegand, C. L. – Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, vol. 43, 1977, pp. 1541–1552.
- [72] Robert C. – *L’analyse statistique bayésienne*. – Economica, 1992.
- [73] Robert C. – *Methodes de Monte Carlo par chaines de Markov*. – Economica, 1996.
- [74] Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. et Harlan, J. C. – *Monitoring the vernal advancement and retrogradation of natural vegetation, Type III, Final Report*. – Rapport technique, NASA/GSFC, 1974.
- [75] Schumitzky A. – Nonparametric EM algorithms for estimating prior distributions. *Applied Mathematics and Computation*, vol. 45, 1991, pp. 141–157.
- [76] Sciences Statistical. – *Splus User’s Manual*. – Seattle: StatSci, a division of Mathsoft, Inc., 1993.

- [77] Searle S., Casella G. et McCulloch C. – *Variance components*. – Wiley, New York, 1992.
- [78] Settle, J. J. et Drake, N. A. – Linear mixing and the estimation of ground cover proportions. *International journal of remote sensing*, vol. 14, 1993, pp. 1159–1177.
- [79] Sheiner L., Rosenberg B. et Marathe, V. V. – Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 8, 1977, pp. 635–651.
- [80] Sheiner L., Rosenberg B. et Melmon K. – Modeling of individuals pharmacokinetics for computer-aided drug dosing. *Computers and Biomedical Research*, vol. 5, 1972, pp. 441–459.
- [81] SPOT IMAGE. – VEGETATION user guide, 1998. http://www.spotimage.fr/data/images/vege/VEGETAT/book_1/e_frame.htm.
- [82] SPOT IMAGE. – La prévision des récoltes dans le monde, un enjeu pour l’union européenne? *Spot Magazine*, vol. 33, 2001.
- [83] Steimer J-L., Mallet A., Golmard J-L. et Boisvieux J-F. – Alternative approaches to estimation of population pharmacokinetic parameters: Comparison with the non-linear mixed effect model. *Drug Metabolism Review*, vol. 15, 1984, pp. 265–292.
- [84] Tierney L. et Kadane, J. B. – Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, vol. 81, 1986, pp. 82–86.
- [85] Toulouse, P. S. – *Thèmes de probabilités et statistiques*. – Dunod, 1999, *Agrégation de mathématiques*.
- [86] Tucker C. – Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of environment*, vol. 8, 1979, pp. 127–150.
- [87] Viñas O. et Baulies X. – 1:250 000 land-use map of Catalonia (32,000 km²) using multi-temporal Landsat-TM data. *International journal of remote sensing*, vol. 16, 1995, pp. 129–146.
- [88] Vignolles C. – *Modélisation de la production agricole à moyenne échelle - Apport du couplage des mesures par satellite à haute résolution spatiale et à haute fréquence temporelle*. – Thèse d’université, Université Paul Sabatier, 1996.
- [89] Viollier M., Belsher T. et Loubersac L. – Signatures spectrales des objets du littoral. In: *Proceedings of the 3rd Int. Coll. on spectral signatures of objects in Remote Sensing, 1985*. ESA SP 247, pp. 253–257.
- [90] Vonesh, E. F. – A note on the use of laplace’s approximation for nonlinear mixed-effects models. *Biometrika*, vol. 83, 1996, pp. 447–452.
- [91] Vonesh, E. F. et Carter, R. L. – Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics*, vol. 48, 1992, pp. 1–17.

- [92] Wakefield, J. C. – The bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association*, vol. 91, n433, 1996, pp. 62–75.
- [93] Walker S. – An EM algorithm for nonlinear random effects models. *Biometrics*, vol. 52, 1996, pp. 934–944.
- [94] Wei, G. C. et Tanner, M. A. – A monte carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, vol. 85, 1990, p. 704.
- [95] Wolfinger, R. D. – Laplace’s approximation for nonlinear mixed models. *Biometrika*, vol. 80, 1993, pp. 791–795.
- [96] Wolfinger, R. D. – Comment : Experiences with the SAS macro NLINMIX. *Statistics in Medicine*, vol. 16, 1997, pp. 1258–1259.
- [97] Wu, C. F. – On the convergence properties of the EM algorithm. *Annals of Statistics*, vol. 11, 1983, pp. 95–103.
- [98] Zhiyu G., Bickel P. et Rice J. – *The histogram method for nonlinear mixed effects models*. – Rapport technique, University of California, Berkeley, 1998.
- [99] . – Glossaire des termes officiels de la télédétection aérospatiale. *Bulletin de la Société française de photogrammétrie et de télédétection*, vol. 112, 1988, p. 44.
- [100] . – The R project for statistical computing, 1997. <http://www.r-project.org>.
- [101] . – ITCF, agriculture de précision: Cap sur le 21e siècle, 2001. http://www.itcf.fr/dossiers/dos1_ch0.html.
- [102] . – NLME :Mixed-effects in S and S-PLUS, 2001. <http://nlme.stat.wisc.edu/>.

Annexe A : Glossaire de télédétection spatiale

L'annexe A est rattachée au chapitre 1. Elle propose un glossaire de termes techniques liées à la télédétection spatiale.

Bande spectrale : *Intervalle du spectre des longueurs d'onde du rayonnement électromagnétique.*

Capteur : *Instrument qui recueille de l'énergie provenant de la scène (ou cible) visée et fournit un signal électrique correspondant, à des fins de mesure. Un capteur passif courtes longueurs d'onde se sert du Soleil comme source d'énergie pour illuminer la cible, à l'inverse d'un capteur actif qui produit sa propre énergie.*

Fauchée : *Largeur de la bande de terrain visée par un système de détection.*

Indice de végétation : *Indicateur empirique permettant d'identifier et de suivre l'évolution temporelle des couverts végétaux. Un indice de végétation permet également d'évaluer certains paramètres du couvert comme l'indice de surface foliaire. Les indices de végétation sont calculés soit à partir de mesures de réflectance sur le terrain, soit de comptes numériques fournis par des données satellitaires.*

Infrarouge : *Partie du spectre électromagnétique s'étendant de $0,7 \mu\text{m}$ à 1 mm de longueur d'onde. En télédétection, on le subdivise généralement en infrarouge proche, moyen et thermique.*

LAI : *(Leaf Area Index), l'indice de surface foliaire est défini comme la moitié de la surface foliaire totale par unité de surface du sol.*

Pixel : *La plus petite surface homogène constitutive d'une image enregistrée, définie par les dimensions de la taille d'échantillonnage. On parle de pixel mixte ou mixel quand plusieurs thèmes sont observés en mélange sous un même pixel.*

Réflectance : *Ratio obtenu à partir de la mesure de la quantité de lumière réfléchie dans une longueur d'onde donnée par une surface donnée par rapport à celle obtenue par la même lumière éclairant un échantillon de référence. Réflectance des plantes : Rapport de l'intensité du rayonnement réfléchi au rayonnement incident des surfaces végétales.*

Résolution spatiale : *Caractéristique d'un capteur liée à son aptitude à distinguer au sol des éléments de petite taille. Plus la résolution spatiale est grande plus les éléments discernables sont petits.*

Résolution spectrale : *Caractéristique d'un capteur liée à son aptitude à utiliser des bandes spectrales plus ou moins étroites. Plus la résolution spectrale est grande plus la bande spectrale est étroite.*

Résolution temporelle : *Fréquence d'acquisition d'une image de la même zone au sol.*

Rouge : *(longueur d'onde dans le \sim) Portion de spectre électromagnétique comprise entre 0,6 et 0,7 μm .*

Satellite : *(\sim artificiel) Engin mis en orbite autour de la Terre, d'une autre planète ou d'un satellite naturel, généralement porteur de matériel scientifique, militaire ou de télécommunication.*

Télédétection : *La télédétection est l'ensemble des connaissances et des techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci. Télé signifie "à distance" et détection veut dire "découvrir" ou "déceler".*

Visible : *(longueur d'onde dans le \sim) Portion de spectre électromagnétique comprise entre 0,4 et 0,7 μm .*

Annexe B : Mise en œuvre des programmes

L'annexe B est rattachée essentiellement au chapitre 4 consacré aux tests de méthodes d'estimation sur simulation. Nous présentons d'abord la simulation de valeurs de LAI (Section 4.1). Dans la deuxième partie, nous proposons une présentation de la fonction `nlme` qui est une implémentation de la méthode d'estimation par linéarisation de Lindstrom et Bates [54] largement utilisée dans les divers tests mis en œuvre. Ensuite, nous présentons l'implémentation des méthodes d'estimation par approximation numérique de la vraisemblance du modèle à paramètres aléatoires (Section 4.2). Puis, nous abordons les programmes établis pour la comparaison des estimations par linéarisation et par régressions non linéaires par pixel (Section 4.3). Enfin, nous décrivons le programme utilisé pour la simulation de structures spatiales (Section 4.4).

1 Simulation de valeurs de LAI

Pour la simulation de valeurs de LAI décrite dans la section 4.1, nous avons écrit la fonction `Plus simul.lai` qui prend en paramètres les arguments suivants :

- `nbpix` : le nombre de pixels,
- `nbdat` : le nombre de dates
- `fix.param` : le vecteur des valeurs de simulations de β ,
- `et.param` : le vecteur des écarts-types σ_β simulés,
- `et.err` : l'écart-type résiduel σ_ϵ simulé.

Les sorties de la fonction `simul.lai` comprennent non seulement les dates et les valeurs de LAI simulées mais aussi les graines du générateur aléatoire pour pouvoir re-simuler les mêmes valeurs sans les stocker intégralement.

2 La librairie `nlme`

2.1 Arguments de la fonction `nlme`

La librairie `nlme`, développée par Pinheiro et Bates [67], est un ensemble de fonctions, méthodes et classes pour l'analyse de modèles non linéaires à effets mixtes. Toutes les informations relative à cette librairie sont disponibles sur le site de l'université du

Wisconsin-Madison [102], d'où il est également possible de télécharger gratuitement le fichier nécessaire à l'installation.

Depuis Splus, un appel classique à la fonction ressemble à :

```
S+> nlme(model, fixed, random, data, cluster, start)
```

L'argument `model` permet de spécifier le modèle non linéaire à ajuster contenant les paramètres et les covariables. Dans le cas du modèle d'évolution temporelle de LAI avec 4 paramètres, cet argument s'écrit :

```
model = lai ~ mod.baret(K, a, b, Tc, Date)
```

où `mod.baret` est défini par :

```
{
lai <- K * (1/(1 + exp(- a * (t - Tc)))
           - exp(b * (t - (log(1 + exp(a * Tc)))/b)))
lai[lai < 0] <- 0
}
```

où la remise à zéro des valeurs négatives est exprimée par la ligne de commande `lai[lai < 0] <- 0`.

L'argument `fixed` est un objet de type `list` déterminant la composante fixe des paramètres du modèle (β) :

```
fixed = list(K ~ ., a ~ ., b ~ ., Tc ~ .)
```

L'argument `random` permet de spécifier, de la même façon, parmi les paramètres, ceux qui ont une composante aléatoire (B_i). Si rien n'est spécifié par l'utilisateur, `random` prend par défaut la même valeur que `fixed` c'est à dire que tous les paramètres du modèle ont une composante aléatoire.

Les données à analyser doivent être mise au format d'un objet de type `data.frame` dont on précise le nom par l'argument `data`. L'argument `cluster` sert à déterminer l'unité expérimentale sur laquelle les paramètres aléatoires varient :

```
cluster = ~ pixel
```

si `pixel` est le nom donné à la variable qui numérote les individus de l'étude.

Enfin, l'argument `start` permet de fournir des valeurs initiales à l'algorithme itératif d'estimation. Seule l'initialisation de la composante fixe des paramètres (β) est requise. L'initialisation éventuelle des paramètres aléatoires (B_i) se fait sous forme d'une matrice de taille $I \times p$, où I désigne le nombre d'individus et p le nombre de paramètres aléatoires. En pratique, il suffit de passer en paramètre un objet de type `list` de la façon suivante :

```
start = list(fixed = c(8,7,3,0.5), random = mat.init)
```

pour initialiser le vecteur de paramètres passé dans l'argument `fixed`, soit (K, a, b, T_c) et `mat.init` est une matrice $I \times p$.

L'aide en ligne sur la fonction `nlme` propose encore bien d'autres arguments optionnels dont on peut trouver une description détaillée dans [67], qui présente également des méthodes numériques et graphiques pour l'interprétation des résultats fournis par `nlme`.

2.2 Sorties de la fonction nlme

Le résultat de la fonction est un objet de type `list` contenant les objets suivants :

```
[1] "call"           "var.ran"           "var.fix"           "alpha"
[5] "delta"         "sigma"            "coefficients"     "loglik"
[9] "niter"         "sizetheta"        "criterion"        "nobs"
[13] "nclus"         "est.method"       "re.paramtr"       "re.structure"
[17] "re.block"     "serial.structure" "var.function"     "plist"
[21] "cluster"       "fitted.values"    "residuals"        "map"
```

Nous renvoyons à l'aide en ligne ou à [67] pour une description détaillée de l'ensemble de ces objets. Signalons ceux que nous utilisons pour calculer les critères de comparaison :

- `"coefficients"` contient deux objets (`"random"` et `"fixed"`) représentant respectivement l'estimation des B_i et de β ,
- `"var.ran"` est l'estimation de la matrice Γ ,
- `"sigma"` désigne l'estimation de l'écart-type résiduel σ_ε .

3 Comparaison des estimations par linéarisation et approximation numérique

Comme le précisent Pinheiro et Bates dans [67], chapitre 7, la méthode de linéarisation est la seule implémentée dans la librairie `nlme`. Dans [66], la comparaison des méthodes a été menée à partir de programmes en C, interfacés sous Splus, spécifiques aux modèles non linéaires considérés dans cet article. Pour mener à bien notre propre série de tests, nous avons adapté les programmes fournis par José C. Pinheiro, en y incluant les sous-routines C nécessaires à la définition du modèle de Baret.

La mise en œuvre de ces méthodes est représentée de façon simplifiée sur la figure 1.

L'interface utilisateur est sous Splus. La seule action à effectuer consiste à appeler un algorithme de minimisation, par exemple `ms`, pour la fonction `approx.loglik` convenablement paramétrée.

Cette fonction implémentée sous Splus effectue une sélection de cas à partir d'une chaîne de caractères passée en paramètre pour choisir la méthode de calcul de la vraisemblance approchée : respectivement `"lapla"`, `"impsa"` et `"gaus"` pour l'approximation laplacienne, Monte Carlo par échantillonnage d'importance et la quadrature gaussienne.

Selon l'appel effectué par `approx.loglik`, le traitement se poursuit dans le sous-programme C correspondant : `Laplacian_loglik`, `IS_loglik` ou `GQ_loglik`. La fonction sélectionnée fait alors appel à la sous-routine `mod_baret` qui permet de calculer la vraisemblance approchée pour ce modèle.

Les résultats sont récupérés sous Splus dans un objet de type `list`.

En pratique, une fonction développée sous Splus permet de gérer l'ensemble du processus, de la simulation (par la fonction `simul.lai`) à l'estimation des paramètres d'abord par `nlme` puis par chacune des trois autres méthodes.

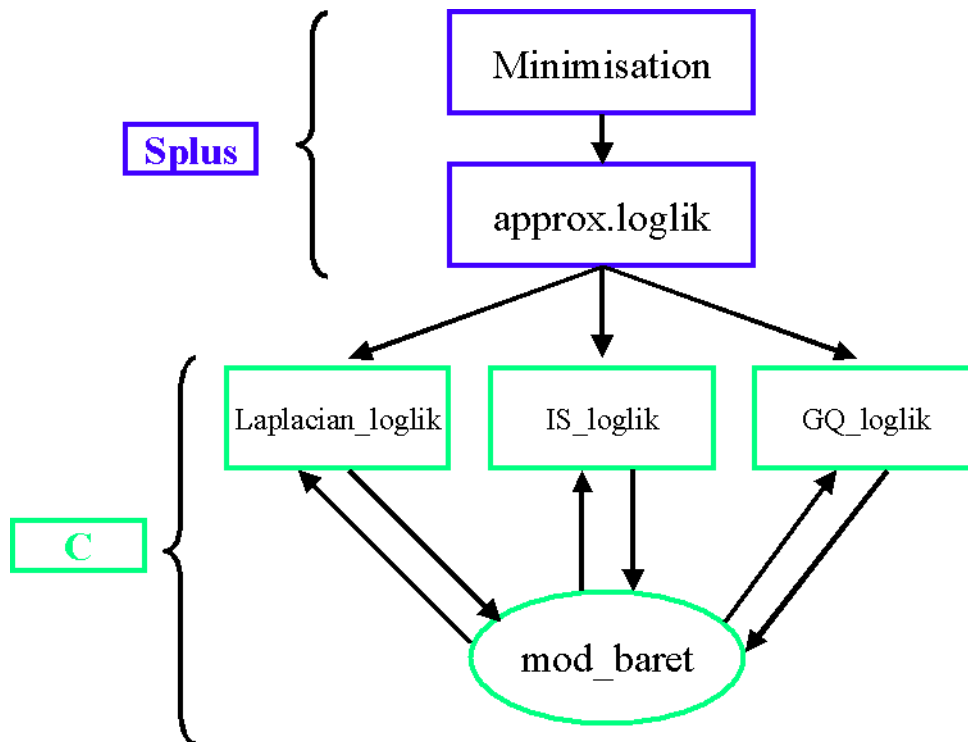


FIG. 1 – Architecture de la mise en œuvre des méthodes d'estimation par approximation de la vraisemblance.

Un appel à la fonction `test.lmg` (pour test des méthodes Laplace, Monte Carlo et quadrature Gaussienne) est de la forme :

```
test.lmg(nbpix=1500, nbdatt=20,
        simul.fixed, simul.stderr.param, simul.stderr.resid)
```

où `nbpix` et `nbdatt` correspondent respectivement au nombre de pixels à traiter et au nombre de dates d'observation des pixels ; ces arguments ayant les valeurs par défaut 1500 et 20.

Les arguments `simul.fixed` et `simul.stderr.param` sont des vecteurs de taille le nombre de paramètres du modèle (ici 3 dans la configuration du modèle de Baret que nous avons choisie). Ils désignent les valeurs des moyennes et écarts-types pour la simulation des paramètres. La variable `simul.stderr.resid` désigne l'écart-type du bruit que l'on ajoute aux données. Pour la configuration 1, l'appel à la fonction `test.lmg` est le suivant :

```
res.config1 <- test.lmg (simul.fixed = c(7,3,0.5),
                        simul.stderr.param = c(0.1,0.1,0.01),
                        simul.stderr.resid = 0.05)
```

Le résultat de cette fonction est stocké dans un objet de type `list` contenant les objets listés ci-dessous :

```
S+> names(res.config1)
[1] "config" "seed" "init" "horaires" "fcount" "resultats"
```

Ces objets correspondent d'une part aux informations relatives à la simulation des données et à l'initialisation de `nlme` (item 1) et d'autre part aux résultats de chaque méthode complétés par diverses informations (item 2) :

1. Informations relatives à la simulation des données et à l'initialisation de `nlme` :
 - “`config`” contient les vecteurs `simul.fixed` et `simul.stderr.param` et la valeur de `simul.stderr.resid`, passés en paramètres lors de la phase de simulation.
 - “`seed`” récupère la graine du générateur aléatoire pour la simulation des valeurs de LAI. Elle permet si besoin de régénérer exactement les mêmes valeurs sans les stocker intégralement.
 - “`init`” reçoit les valeurs initiales fournies à `nlme` pour les effets fixes.
2. Résultats de chaque méthode complétés par diverses informations :
 - “`resultats`” reçoit pour chaque méthode le vecteur des estimations des paramètres de population : moyenne, variances des paramètres et résiduelle.
 - “`fcount`” représente le nombre d'évaluations de fonctions nécessaires durant l'algorithme de minimisation pour les trois méthodes d'approximation.
 - “`horaires`” contient la durée en temps réel jusqu'à la convergence de l'algorithme de minimisation pour chaque méthode.

Paramétrage

Ces informations propres à une méthode sont passées en paramètres dans l'appel à la fonction correspondante *via* `approx.loglik`. Dans le cas de la méthode de Monte Carlo, la taille de l'échantillonnage est accompagnée d'un autre paramètre contenant la graine d'un générateur aléatoire. Pour la quadrature gaussienne, les abscisses et les poids associés sont calculés par une fonction FORTRAN extérieure au programme de base puis passés en paramètres dans `approx.loglik`.

4 Comparaison des estimations par pixel et par linéarisation

Pour mettre en œuvre cette série de tests, la fonction `comp.nlme.nls` a été développée sous Splus. Elle effectue la simulation des données (fonction `simul.lai`) et l'estimation selon les deux approches. Pour le modèle à paramètres aléatoires, l'estimation est exécutée par la méthode de linéarisation selon Lindstrom et Bates *via* la librairie `nlme`. Pour les régressions non linéaires par pixel, nous utilisons la fonction `nlsList` de Splus qui implémente une procédure d'estimation par moindres carrés pour chaque individus d'une population.

La fonction `comp.nlme.nls` prend en paramètre :

- le nombre de pixels,
- le nombre de dates,
- les valeurs du vecteur des moyennes β utilisées pour la simulation,
- les valeurs du vecteur des écarts-types σ_β utilisées pour la simulation,
- l'écart-type du bruit ajouté aux valeurs simulées.

Les résultats d'un appel à cette fonction sont stockés dans un objet de type `list` (objet géré par `Splus` pouvant contenir des variables de diverses natures : vecteur, matrice, chaîne de caractères ...) contenant les informations relatives à la simulation et les résultats obtenus par chaque méthode.

La fonction `comp.nlm.nls` gère également le calcul des valeurs initiales.

5 Simulation de structures spatiales

La mise en œuvre de la simulation de données structurées spatialement est effectuée par plusieurs routines FORTRAN interfacées sous `Splus` fournies par M. Goulard.

La fonction `simulg.S` assure l'interface avec l'utilisateur. Un appel type à cette fonction est le suivant :

```
simulg.S(n, m, lambda, dw, pasn, pasm, sig)
```

où les variables passées en paramètres sont :

- `n` et `m` : nombre de lignes et colonnes de la grille sur laquelle on souhaite simuler,
- `lambda` : valeur de la corrélation comprise strictement entre -1 et 1,
- `dw` : distance de corrélation. Cette valeur permet de définir le voisinage qui influence un pixel.
- `pasn` et `pasm` : pas de parcours de la grille en ligne et colonne,
- `sig` : écart-type de la variable aléatoire ϵ .

La suite du programme est essentiellement gérée par des routines FORTRAN pour en accélérer l'exécution notamment pour l'inversion de matrices. On récupère ensuite sous `Splus` les sorties du programme de simulation.

Comparison of estimation methods in the random parameter nonlinear models framework: application to the modelling of crop leaf area index temporal evolution from spatial remote sensing

Abstract

In the spatial remote sensing for agriculture framework, we deal with the estimation of a crop leaf area index evolution model from high temporal resolution images; the scope of the work is the improvement of yield prediction over a large area. Regular measurements can be considered as longitudinal data of a population of pixels. We propose a random parameter modelling to express between-pixels variability and to deal with the whole pixels simultaneously. We present parametric and non parametric expressions of the between-pixels variability. We focus on the parametric model for which we review estimation methods based either on a per-pixel approach or on the maximization of an approximated global likelihood. We establish a synthesis of estimators properties from these methods. A comparison on simulated data show that, in our case, the linearisation method is the best among approximation methods. Then, we make apparent the greater robustness of linearisation in relation with a per-pixel-based approach under bad conditions: low number of dates, high noise. We also exhibit the robustness of the linearisation towards non specification of spatial structure in the model. Finally, on a real dataset, we confirm that the linearisation method is more powerful than a per pixel based approach.

Key-words : nonlinear mixed effects model, longitudinal data, satellite image, SPOT/Végétation, crop monitoring.

Auteur : Sébastien DEJEAN

Titre : Comparaison de procédures d'estimation dans le cadre des modèles non linéaires à paramètres aléatoires : application à la modélisation de l'évolution temporelle de l'indice de surface foliaire de cultures observées par télédétection spatiale.

Directeurs de thèse : Jean-Marc Azaïs et Michel Meste.

Lieu et date de soutenance : Salle des thèses du bâtiment U4, Université Toulouse 3 - Paul Sabatier, 4 juillet 2002.

Résumé

Dans le cadre de la télédétection spatiale pour l'agriculture, nous nous intéressons à l'estimation d'un modèle d'évolution de l'indice de surface foliaire d'une culture à partir d'images à haute résolution temporelle ; l'enjeu étant d'améliorer la prévision de rendement sur une région agricole. Les acquisitions d'images sont assimilables à des données longitudinales sur une population de pixels. Nous proposons une modélisation par paramètres aléatoires pour exprimer la variabilité inter-pixels et traiter l'ensemble des pixels simultanément. Nous présentons les expressions paramétriques et non paramétriques de la variabilité inter-pixels. Nous nous attachons au modèle paramétrique pour lequel nous passons en revue les méthodes d'estimation basées soit sur des régressions non linéaires par pixel soit sur la maximisation d'une approximation de la vraisemblance globale. Nous établissons une synthèse des propriétés des estimateurs issus de ces diverses méthodes. Une comparaison sur données simulées valide dans notre cas l'emploi de la méthode par linéarisation parmi les méthodes d'approximation. Ensuite, nous montrons que l'estimation par linéarisation est plus robuste qu'une approche par pixel face à des conditions défavorables : faible nombre de dates, données bruitées. Puis, nous mettons en évidence la robustesse relative de la linéarisation par rapport à la non prise en compte d'une structure spatiale dans le modèle. Finalement, sur un jeu de données réelles, nous confirmons que l'estimation par linéarisation est plus performante qu'une approche par pixel.

Mots-clés : modèle mixte, régression non linéaire, données longitudinales, image satellite, SPOT/Végétation, suivi de cultures.

Discipline administrative : Mathématiques appliquées.

Laboratoire de rattachement : Laboratoire de Statistique et Probabilités - Université Paul Sabatier - 118, Route de Narbonne 31062 TOULOUSE cedex 4.