



HAL
open science

DEVELOPPEMENT D'UN ALGORITHME DE SUIVI DE LA VEGETATION A LARGE ECHELLE

Marie Weiss

► **To cite this version:**

Marie Weiss. DEVELOPPEMENT D'UN ALGORITHME DE SUIVI DE LA VEGETATION A LARGE ECHELLE. Physique de l'espace [physics.space-ph]. Université Nice Sophia Antipolis, 1998. Français. NNT: . tel-00707683

HAL Id: tel-00707683

<https://theses.hal.science/tel-00707683>

Submitted on 14 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

présentée en vue de l'obtention du titre de

**DOCTEUR DE L'UNIVERSITÉ DE
NICE-SOPHIA ANTIPOLIS**

Spécialité : Électronique

par

Marie WEISS

**DÉVELOPPEMENT D'UN
ALGORITHME DE SUIVI DE LA
VÉGÉTATION À LARGE ÉCHELLE**

*soutenue le 4 Décembre 1998 devant la
commission d'examen composée de :*

| | | |
|--------------------|---|-------------------|
| L. WALD | <i>Directeur de Recherches, EMP, Sophia Antipolis</i> | Président |
| M. LEROY | <i>Ingénieur CNES, CESBIO, Toulouse</i> | Rapporteur |
| B. PINTY | <i>Professeur, Université de Clermont-Ferrand</i> | Rapporteur |
| F. BARET | <i>Directeur de Recherches, INRA, Avignon</i> | Examineur |
| L. POLIDORI | <i>Ingénieur ALCATEL, Cannes La Bocca</i> | Examineur |
| M. STEVEN | <i>Professeur, Université de Nottingham</i> | Examineur |

Remerciements

À l'issue de cette thèse, je tiens à exprimer mes sincères remerciements à tous ceux qui ont contribué à sa réalisation.

Je remercie donc Frédéric Baret, qui a réalisé le suivi temporel de mon travail à fine échelle. Je le remercie pour tout ce qu'il m'a appris avec patience, pour son soutien et sa disponibilité.

Mes remerciements vont également à tous les autres membres du jury : Lucien Wald, qui a accepté d'être mon directeur de thèse universitaire ainsi que de présider le jury ; Marc Leroy, rapporteur, qui a montré un vif intérêt à ce travail tout au long de son déroulement, et dont les conseils ont été bénéfiques ; Bernard Pinty, qui, sans me connaître, a accepté d'endosser la lourde charge de rapporteur avec un esprit critique ; ses remarques ont été intéressantes et constructives ; Laurent Polidori, mon correspondant Aérospatiale, qui m'a permis de me "frotter" au monde de l'industrie, et qui m'a régulièrement encouragée pendant ces trois années ; Mike Steven, qui a eu la gentillesse d'accepter d'examiner un manuscrit dans une langue étrangère et de faire le voyage d'Angleterre pour assister à ma soutenance.

Je voudrais également remercier ceux qui ont assuré le financement de ce travail : l'aérospatiale de Cannes (fusionnée maintenant avec Alcatel Space Industry) et la région Provence-Alpes-Côte d'Azur. Je ne peux pas non plus oublier Bernard Seguin, et son successeur, Richard Delécolle, (les directeurs de la station de Bioclimatologie de l'INRA d'Avignon) qui m'ont accueillie dans leur laboratoire et m'ont permis d'y travailler dans la bonne humeur, et dans d'excellentes conditions (de la station Sun au ventilateur indispensable en été).

Je remercie aussi tout particulièrement mes deux "co-bureaux" : Thierry Fourty pour tous les conseils de "vieux routier" que j'ai pu prodiguer ensuite à son successeur, Frédéric Jacob. Je les remercie tous les deux pour les longues explications devant le tableau blanc ainsi que les moments de détente qui ont parsemé ces 3 années.

Je remercie ensuite mes fidèles et critiques lecteurs : Fred B., Albert, Jaco et les 15 personnages qui se bousculent dans sa tête, Fred H., Agnès, Nadine et Jean (merci Tonton!).

Un merci tout spécial à Nadine Bruguier (la "Faltaouette des Alpillettes") qui a toujours été à mon écoute dans les bons et les mauvais moments. Je la remercie pour les longues discussions et les fous-rires que nous avons partagés.

Enfin, je ne veux pas oublier non plus toutes les personnes qui ont participé au bon déroulement de cette thèse : Rosy "bip-bip" Olivier qui par son dynamisme et sa compétence, a permis de résoudre bien des problèmes administratifs ; Philippe Clastre ("Peuk-peuk" pour les intimes) et ses SOSI associés (Marie-Odile Ariès et Edith Larue), les "system reanimators" ; Christine Saumade (membre actif des escapades dans les Alpilles et ailleurs), Agnès Bégué (qui m'a enseigné avec patience le B.A-BA du transfert radiatif dans le couvert), Jacqueline et Jacques Saumade dont l'accueil et le sourire ont toujours été chaleureux ; mes compagnons de bonne fortune, Habiba, Carlos et Marisa à qui j'irais volontiers rendre visite dans leur pays respectif (!), ainsi que Patrice, Audrey et Karine ; le personnel permanent et non permanent de la station de bioclimatologie et d'ailleurs (Raphaëlle, Vanessa et Manu, Yannick, Jean-Pierre, Laurent, Jean-François, Dominique, Soraya, Renaud, Claudine, Sylvie, Marie, Fred Bi et Fred Ba, Véronique, Odile, Françoise, Ranga, Nathalie, Christophe, Mohammed, Agnès, Belkacem, Amadou, Julien, Serygne, Anne, Momo, Nicolas B., et tous les autres). Ils ont tous participé aux activités parfois intenses du club des pâtisseries. Pourvu que ça dure !

Il me reste à remercier ensuite ma famille et surtout celui qui, s'il n'était pas impliqué directement dans mon travail, a dû quand même en subir les conséquences (bonnes ou mauvaises) sur mon humeur quotidienne. À l'en croire, ce fût une très lourde tâche qui lui incombât!

Pour terminer, je dois attribuer une mention spéciale à l'inconnu de la parcelle de blé n° 120, qui je l'espère, ne m'en voudra pas trop de l'avoir exposé dans ce manuscrit.

Table des Matières

| | | |
|----------|---|-----------|
| 1 | Introduction. | 7 |
| 1.1 | Contexte général de l'étude | 7 |
| 1.2 | Les variables biophysiques des couverts végétaux | 11 |
| 1.2.1 | L'indice foliaire | 11 |
| 1.2.2 | La fraction de rayonnement photosynthétiquement actif absorbé par le couvert | 12 |
| 1.2.3 | La fraction de trou | 13 |
| 1.2.4 | La teneur en chlorophylle intégrée au niveau du couvert | 14 |
| 1.3 | Méthodes d'estimation des caractéristiques des couverts végétaux | 14 |
| 1.3.1 | Les indices de végétation | 15 |
| 1.3.2 | L'inversion de modèles de transfert radiatif | 19 |
| 1.3.3 | Les méthodes hybrides | 24 |
| 1.3.3.1 | Les tables de correspondance | 24 |
| 1.3.3.2 | Les réseaux de neurones artificiels | 25 |
| | État de l'art | 25 |
| | Principe | 25 |
| 1.4 | Objectif et plan de l'étude | 27 |
| 2 | Construction d'une base de données synthétiques. | 31 |
| 2.1 | Les modèles utilisés | 32 |
| 2.1.1 | Les modèles de réflectance du couvert | 32 |
| 2.1.1.1 | Les différents types de modèles | 33 |
| | Les modèles informatiques | 33 |
| | Les modèles géométriques | 33 |
| | Les modèles "milieu turbide" | 33 |
| | Les modèles hybrides | 34 |

| | | |
|----------|--|-----------|
| 2.1.1.2 | Le modèle SAIL | 34 |
| | Validité du modèle | 34 |
| | Rappels théoriques | 35 |
| | Variables d'entrée du modèle | 37 |
| 2.1.2 | Le modèle de propriétés optiques des feuilles : PROSPECT | 39 |
| 2.1.3 | Le modèle de propriétés optiques du sol : SOILSPECT | 39 |
| 2.1.4 | Le modèle d'orbitographie de <i>VEGETATION</i> | 40 |
| 2.2 | Rappels théoriques sur le calcul du $fAPAR$ et de la fraction de trou dans le formalisme de SAIL | 44 |
| 2.2.1 | Calcul du $fAPAR$ | 44 |
| 2.2.2 | Les fractions de trou $P_o(\theta_v), P_o(\theta_v, \theta_s)$ | 46 |
| 2.3 | Domaine de variation des variables d'entrée du modèle SAIL | 48 |
| 2.3.1 | Les variables de structure | 49 |
| | Le LAI | 49 |
| | L'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$ | 52 |
| | Le paramètre de hot spot | 53 |
| 2.3.2 | Les propriétés optiques des feuilles | 53 |
| 2.3.3 | Les propriétés optiques des sols | 53 |
| | Variables indépendantes de la longueur d'onde | 54 |
| | L'albédo de simple diffusion $\omega_s(\lambda)$ | 54 |
| 2.3.4 | La géométrie de mesure | 55 |
| 2.4 | Exemples de simulation | 57 |
| 3 | Estimation de variables biophysiques des couverts végétaux. | |
| | Calage des réseaux de neurones. | 65 |
| 3.1 | Choix d'un type de réseau de neurones | 66 |
| 3.1.1 | L'algorithme de rétropropagation du gradient de l'erreur | 67 |
| 3.1.2 | Amélioration des performances du réseau | 70 |
| 3.2 | Pré-traitement des variables d'entrée des réseaux neuronaux : | |
| | normalisation des données | 70 |
| 3.2.1 | Les modèles paramétriques de BRDF | 73 |
| 3.2.2 | Le modèle de Walthall | 75 |
| 3.2.3 | Le modèle MRPV | 75 |
| 3.2.4 | Comparaison des performances des deux modèles | 77 |
| | Restitution de l'information directionnelle | 78 |

| | | |
|-------|---|-----|
| | Stabilité des paramètres d'entrée des modèles en présence de nuages | 81 |
| | Conclusion | 83 |
| 3.3 | Méthodologie d'utilisation des réseaux de neurones : estimation du taux de couverture du sol | 85 |
| 3.3.1 | Calage du réseau de neurones | 85 |
| 3.3.2 | Performances de l'algorithme basé sur l'utilisation de réseau de neurones pour l'estimation de la fraction de trou au nadir | 87 |
| 3.3.3 | Performances de l'algorithme basé sur l'utilisation du <i>NDVI</i> pour l'estimation de la fraction de trou au nadir | 91 |
| | La relation " <i>NDVI/P_o(0)</i> " | 91 |
| | Calage de la relation | 93 |
| 3.3.4 | Conclusion | 97 |
| | Choix de la méthode d'estimation de variables biophysiques | 97 |
| | Influence de la présence de nuages sur l'estimation de la fraction de trou au nadir | 97 |
| 3.4 | Application à l'estimation des différentes variables biophysiques | 100 |
| 3.4.1 | Estimation de la fraction de trou dans des directions particulières $P_o(\theta_s), P_o(58^\circ)$ | 100 |
| 3.4.2 | Estimation de l'indice foliaire <i>LAI</i> | 101 |
| 3.4.3 | Estimation de l'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$ | 107 |
| 3.4.4 | Estimation des fractions de trou mono et bi-directionnelles $P_o(\theta_v), P_o(\theta_v, \theta_s)$ | 107 |
| | Estimation de la fraction de trou monodirectionnelle | 107 |
| | Estimation de la fraction de trou bidirectionnelle | 109 |
| 3.4.5 | Estimation de la fraction de rayonnement photosynthétiquement actif absorbé par le couvert <i>fAPAR</i> | 110 |
| 3.4.6 | Estimation de la chlorophylle intégrée au niveau du couvert <i>LAI.C_{ab}</i> | 111 |
| 3.4.7 | Conclusion sur l'estimation de variables biophysiques du couvert à l'aide de réseaux de neurones. | 113 |
| 3.5 | Sensibilité à la variabilité intrapixellaire | 115 |
| 3.5.1 | Le problème de l'hétérogénéité spatiale en télédétection | 115 |

| | | |
|----------|---|------------|
| 3.5.2 | Construction d'une base de données "pixels composites" | 116 |
| 3.5.3 | Estimation des variables biophysiques de "pixels composites" | 116 |
| | Estimation de la fraction de trou au nadir | 118 |
| | Estimation de l'indice foliaire | 118 |
| 3.5.4 | Conclusion sur la sensibilité à la variabilité intrapixellaire | 123 |
| 4 | Validation de l'algorithme de suivi des couverts végétaux. | 127 |
| 4.1 | Validation sur la campagne Grignon 1990 | 127 |
| 4.1.1 | Composition du jeu de données | 128 |
| | Agencement de la parcelle | 128 |
| | Mesure des variables biophysiques | 129 |
| | Mesures de réflectance | 130 |
| | Jeu de mesures disponibles | 130 |
| 4.1.2 | Construction d'un réseau de neurones adapté aux longueurs d'onde <i>CIMEL</i> | 132 |
| 4.1.3 | Validation de l'algorithme sur les données Grignon 1990 | 133 |
| | Le problème de l'échantillonnage directionnel | 133 |
| | Estimation de la fraction de trou au nadir | 133 |
| | Estimation de $P_o(\theta_s)$ et $P_o(58^\circ)$ | 136 |
| 4.1.4 | Conclusion | 138 |
| 4.2 | Validation sur la campagne Alpilles/ReSeDA 1997 | 139 |
| 4.2.1 | Composition du jeu de données | 140 |
| | Description des parcelles | 140 |
| | Mesure de variables biophysiques | 141 |
| | Mesure de réflectance bidirectionnelle | 143 |
| 4.2.2 | Construction d'un réseau de neurones adapté aux longueurs d'onde <i>PolDER</i> | 146 |
| 4.2.3 | Validation de l'algorithme sur les données Alpilles/ReSeDA 1997 | 151 |
| | Variables d'entrée du réseau de neurones : (ρ_o, ρ_h) dans les 4 bandes <i>PolDER</i> et θ_s | 152 |
| | Variables d'entrée du réseau de neurones : (ρ_o, ρ_h) dans 3 bandes <i>PolDER</i> (550 nm, 670 nm et 864 nm) et θ_s | 156 |
| 4.2.4 | Conclusion sur la campagne Alpilles/ReSeDA | 156 |

| | |
|---|------------|
| 5 Conclusion. | 161 |
| Bibliographie | 167 |
| Liste des symboles mathématiques | 181 |
| Glossaire | 187 |

Chapitre 1

Introduction.

1.1 Contexte général de l'étude

En décembre 1997, à Kyoto, 170 pays se sont réunis suite à un diagnostic établi ces dernières années par la communauté scientifique internationale démontrant les problèmes liés au changement climatique que nous subissons depuis le début du siècle. Contrairement au passé, ce changement climatique est très rapide (un siècle par rapport à des millénaires), et provoqué non seulement par des causes naturelles mais aussi pour la première fois par l'activité humaine (émission de gaz à effet de serre par l'industrie, les transports et l'agriculture).

Les conséquences du réchauffement brutal de la planète se font déjà sentir, semble-t-il, au travers de l'aggravation des phénomènes extrêmes : tempêtes, sécheresses, inondations, ... Ainsi, de grands programmes internationaux (World Climate Research Programme, International Geosphere Biosphere Programme) ont été mis en place afin de quantifier mais aussi de prévoir l'incidence de ces changements climatiques sur notre environnement. Des modèles de circulation globale (ou GCM) ont été développés pour simuler les changements de climat aussi bien sur des périodes courtes (prévisions météorologiques) qu'à plus long terme (changement global). Ils sont basés sur les équations fondamentales de la dynamique qui décrivent les mouvements atmosphériques à large échelle (Dickinson, 1984, Royer *et al.*, 1995).

Outre les océans, la biosphère continentale joue un rôle important dans le système climatique et donne les conditions aux limites inférieures pour les GCM (figure 1.1). En effet, elle détermine à la fois les cycles biogéochim-

iques (activité photosynthétique de la végétation pour le cycle du carbone, évapotranspiration et ruissellement des sols pour le cycle de l'eau) et les échanges énergétiques à l'interface sol-végétation-atmosphère (absorption de l'énergie solaire, émission dans l'infrarouge thermique). Les mécanismes physiques de ces processus sont maintenant assez bien connus et modélisés à l'échelle locale de la parcelle mais leur quantification et leur suivi à des échelles supérieures requises par les GCM (mailles de 30 km × 30 km au minimum) reste difficile à partir d'outils classiques.

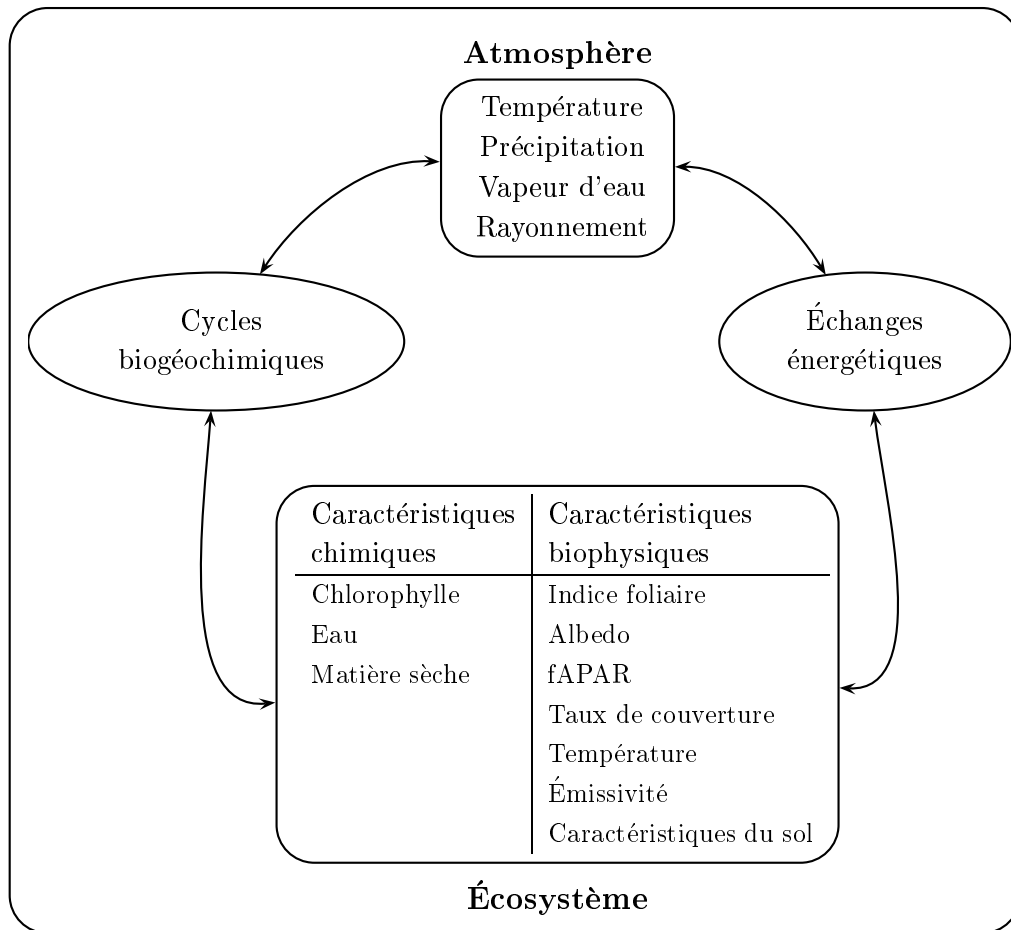


Figure 1.1 : Variables et processus physiques mis en jeu dans les interactions entre la biosphère continentale et l'atmosphère. Représentation schématique.

Depuis les années 70, l'avènement des premiers capteurs satellitaires destinés dans un premier temps à des fins météorologiques (*NOAA/AVHRR*) a permis de donner une réponse à ce type de besoin et d'élargir les champs de recherche sur l'observation de la terre. Le premier programme d'étude des ressources terrestres est décidé en 1969 avec le capteur américain *LANDSAT* lancé en 1972, suivi par le programme français *SPOT* en 1986. Les premiers résultats obtenus grâce à ces programmes ont permis de spécifier les caractéristiques spectrales, spatiales et temporelles des capteurs suivants pour étendre les champs d'investigation (observation des océans, de l'atmosphère et de la biosphère continentale) et acquérir des résultats de plus en plus fiables et précis. Depuis le début des années 1990, une nouvelle génération de capteurs embarqués sur des plateformes multi-instrumentées européenne (*ENVISAT*), française (*SPOT4*), américaine (*EOS*) et japonaise (*ADEOS*) voit le jour. L'observation du globe à partir de l'espace offre donc maintenant de nombreuses possibilités.

La télédétection regroupe l'ensemble des techniques utilisant de tels capteurs pour caractériser les surfaces naturelles (sols, cultures, forêts, mers, océans, ...) ou l'atmosphère, ainsi que leur évolution dans le temps. Cette caractérisation peut se faire aussi bien en termes d'occupation des sols et d'aménagement du territoire, qu'en termes de "mesure" ou de "quantification" d'une grandeur donnée, spécifique des surfaces considérées : par exemple, la concentration en chlorophylle pour les océans (phytoplancton), la concentration en aérosols pour l'atmosphère ou la surface foliaire pour les couverts végétaux. Ces quantités ne sont accessibles par télédétection que si elles interviennent dans les mécanismes qui gèrent le transfert radiatif. De plus, ces mécanismes font intervenir des composantes différentes selon le domaine spectral considéré : par exemple, la température de surface dans l'infrarouge thermique (Kerr *et al.*, 1992, Ottlé et Vidal-Madjar, 1992), l'humidité de surface dans le domaine des hyperfréquences (Champion *et al.*, 1996), la surface foliaire ou encore la quantité d'énergie absorbée pour la photosynthèse dans le domaine solaire (Asrar *et al.*, 1984).

L'étude que nous avons menée se restreint au domaine solaire et porte sur l'extraction de caractéristiques biophysiques des couverts végétaux, non seulement des cultures mais aussi des zones de végétation naturelle. Elle entre dans le cadre du programme préparatoire au capteur *VEGETATION* développé par le CNES et la communauté européenne. *VEGETATION* a été

lancé en mars 1998 sur la plateforme *SPOT4*. Ce capteur est destiné à trois domaines d'application particuliers :

1. *La cartographie des variables de surface.*

Ceci permet de déterminer les conditions aux limites inférieures des GCM. Les échelles spatiales (50 km) généralement utilisées dans ces modèles requièrent une description de l'occupation du sol et de sa variabilité avec un échantillonnage spatial de 8 à 10 km, ce qui est accessible grâce à la résolution de 1 km de *VEGETATION*.

2. *Le suivi des productions agricoles, pastorales ou forestières.*

L'utilisation des données de *VEGETATION* doit notamment, permettre d'évaluer les potentiels de production végétale : gestion de la production céréalière pour les pays exportateurs, suivi des ressources pastorales, évaluation des impacts de la déforestation, . . . Les images acquises par le capteur fournissent ainsi un outil d'aide à la décision pour des problèmes d'aménagement du territoire ou de politique agricole. La particularité de *VEGETATION* est de fournir des images simultanément au capteur *HRVIR* ayant une résolution spatiale de 20 m et les mêmes caractéristiques spectrales (excepté les bandes dans le bleu et le vert).

3. *Le suivi et la modélisation de la biosphère continentale.*

Le suivi temporel de la biosphère au moyen de la quantification de variables biophysiques ou biochimiques permet une meilleure compréhension des processus physiques du système climatique. Cela permet, à terme, d'améliorer la modélisation et la description des mécanismes du transfert radiatif, ou, plus généralement, des interactions sol-végétation-atmosphère.

VEGETATION est un capteur similaire à *NOAA/AVHRR* mais il présente de nombreux avantages :

- *au niveau spectral*, il possède deux bandes supplémentaires, une dans le bleu pour mieux caractériser l'atmosphère et une dans le moyen infrarouge pour mieux caractériser la végétation. Il n'a cependant pas de bande dans l'infrarouge thermique (3 pour *AVHRR*) pour l'étude du cycle de l'eau dans la végétation ;

- *au niveau technologique*, les distorsions géométriques dues à l'instrument sont beaucoup moins importantes, ce qui permet de mieux superposer des images acquises à différentes dates et d'exploiter une plus grande partie des images. De plus, contrairement à *VEGETATION*, il n'existe pas d'étalonnage fiable pour *NOAA/AVHRR* ;
- *l'analyse multi-échelle* est rendue possible grâce à la combinaison d'images *VEGETATION* (résolution spatiale de 1 km) et d'images *HRVIR* (résolution spatiale de 20 m) provenant de la même plateforme *SPOT4*.

VEGETATION a donc été conçu en grande partie pour réaliser un suivi des couverts végétaux par le biais de l'estimation des variables biophysiques qui les caractérisent. Ses propriétés spectrales et directionnelles ont été définies à partir de résultats provenant d'études réalisées sur les capteurs antérieurs. C'est pourquoi nous allons, dans un premier temps, décrire les variables biophysiques des couverts végétaux qui sont accessibles par télédétection. Nous présenterons ensuite un état de l'art des méthodes d'estimation de ces variables à partir de données satellitaires.

1.2 Les variables biophysiques des couverts végétaux

1.2.1 L'indice foliaire

L'indice foliaire (ou LAI^1) d'un couvert végétal est un paramètre adimensionnel qui est défini comme l'aire totale des faces supérieures des feuilles par unité de surface de sol. C'est une variable largement utilisée en écophysiologie, pour des modèles de fonctionnement de la végétation. Ceux-ci, appliqués à des cultures, donnent la possibilité d'effectuer un diagnostic de situation culturale en temps réel pour l'estimation de besoin en intrants par exemple, ou encore pour prédire le rendement en cours de culture (Guérif *et al.*, 1995). En outre, comme le LAI quantifie l'étendue de l'interface entre la plante et l'atmosphère, il est fortement impliqué dans les transferts de masses et d'énergie. Ainsi, c'est une variable couramment employée dans les modèles SVAT² (Sellers, 1985, Olioso *et al.*, 1997) ou dans les GCM (Dickinson, 1983, Sellers *et al.*, 1986, Royer

¹Leaf Area Index

²Soil-Vegetation-Atmosphere Transfer

et al., 1995). L'indice foliaire est, entre autres, l'une des principales variables qui gouvernent le transfert radiatif puisqu'il reflète la quantité de surface des feuilles qui interceptent la lumière.

1.2.2 La fraction de rayonnement photosynthétiquement actif absorbé par le couvert

Le rayonnement photosynthétiquement actif absorbé par le couvert est compris dans le domaine spectral 400 nm-700 nm. Ce domaine correspond en effet aux longueurs d'onde pour lesquelles il y a absorption par la chlorophylle (figure 1.2). Le $fAPAR^3$ est le rapport entre le rayonnement photosynthétiquement actif absorbé par le couvert et le rayonnement photosynthétiquement actif incident. À la différence du LAI , le $fAPAR$ varie tout au long de la journée puisqu'il dépend de l'éclairement solaire. C'est pourquoi, il faut considérer non pas un $fAPAR$ instantané mais le $fAPAR$ intégré tout au long de la journée.

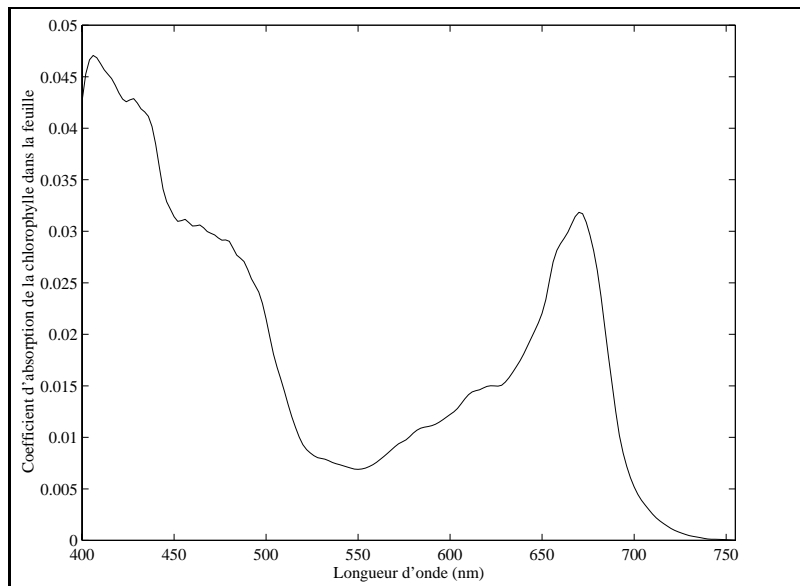


Figure 1.2 : Exemple-type de la variation spectrale du coefficient spécifique de l'absorption par la chlorophylle dans la feuille.

³fraction of Absorbed Photosynthetically Active Radiation

De la même façon que le LAI , le $fAPAR$ est utilisé dans les modèles de fonctionnement de culture pour estimer les rendements puisque la production de biomasse est le fruit de la photosynthèse (Monteith, 1972). Pour la même raison, il intervient aussi largement dans le cycle du carbone (Myneni et Williams, 1994).

1.2.3 La fraction de trou

La fraction de trou est une variable structurelle, qui dépend du LAI et de l'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$. On peut distinguer la fraction de trou monodirectionnelle qui correspond à la fraction de sol vu dans une direction θ_v donnée, de la fraction de trou bidirectionnelle qui est la portion de sol *éclairé* vu dans une direction donnée. En ce qui concerne la fraction de trou monodirectionnelle, trois directions sont particulièrement intéressantes :

- $P_o(\theta_v = 0)$ qui est la fraction de trou au nadir. C'est le complément de la fraction de couverture du sol C ($C = 1 - P_o(0)$) qui est une variable importante dans les modèles de fonctionnement de culture, ou encore les modèles SVAT. C'est aussi une variable explicitement utilisée dans les modèles de transfert radiatif pour les visées au nadir ;
- $P_o(\theta_v = \theta_s)$ qui est la fraction de trou dans la direction solaire. Elle correspond à $(1 - fAPAR_{\text{instantané}})$ si les feuilles sont supposées noires (réflectance et transmittance nulles). C'est également une variable explicitement employée dans la modélisation du transfert radiatif ;
- $P_o(\theta_v = 58^\circ)$ qui est indépendante de l'angle moyen d'inclinaison des feuilles (Bonhomme et Chartier, 1972) quelle que soit la distribution d'inclinaisons foliaires. $P_o(\theta_v = 58^\circ)$ fournit, par conséquent, un moyen potentiel d'estimer le LAI .

L'intérêt de la fraction de trou monodirectionnelle est qu'elle est un élément clé du transfert radiatif, et, par conséquent, fortement reliée au signal de réflectance. De plus, la connaissance de $P_o(\theta_v)$ permet d'accéder au LAI , comme au $fAPAR$ (Andrieu et Baret, 1993).

1.2.4 La teneur en chlorophylle intégrée au niveau du couvert

Les travaux de Yoder et Pettigrew-Crosby (1995) ont montré qu'il est difficile d'estimer indépendamment la biochimie foliaire et le *LAI*. En effet, la réflectance du couvert résulte d'une combinaison des réflectances du sol et de chacune des feuilles. C'est pourquoi on considère plutôt la teneur en chlorophylle intégrée au niveau du couvert, c'est-à-dire la teneur en chlorophylle des feuilles multipliée par l'indice foliaire. C'est une variable importante de la photosynthèse et donc du cycle du carbone (Sellers, 1985). Elle peut être en outre, une indicatrice du statut azoté des plantes (Everitt *et al.*, 1985).

1.3 Méthodes d'estimation des caractéristiques des couverts végétaux

Le problème posé est le suivant : on dispose d'images multispectrales provenant de capteurs situés entre 800 et 850 km au-dessus de la terre, visant une partie du globe terrestre. Dans le domaine solaire, ces images résultent de l'interaction rayonnement-matière le long du trajet entre le soleil et le capteur. Le problème revient donc à comprendre le trajet du rayonnement solaire du soleil vers la surface, puis de la surface vers le capteur, mais surtout, son interaction avec la surface.

Du fait de la complexité des milieux naturels et des processus physiques dont résulte la mesure de réflectance au niveau du capteur, il est nécessaire de les "modéliser", c'est-à-dire de traduire en langage mathématique une représentation simplifiée ou partielle de la réalité. Dans le cas des couverts végétaux dans le domaine solaire (300 nm - 3000 nm), la modélisation permet de relier des variables de structure ou des variables biophysiques caractéristiques du couvert à la réflectance obtenue au niveau du capteur. Les modèles peuvent fonctionner dans le sens direct (modélisation de la réflectance du couvert) mais il est primordial de pouvoir les faire fonctionner dans le sens inverse pour obtenir les variables biophysiques ou biochimiques à partir de la réflectance mesurée au niveau du capteur (figure 1.3). Les méthodes d'estimation des caractéristiques des couverts végétaux peuvent être divisées en deux catégories (Hall *et al.*, 1995):

- les méthodes "empiriques" qui sont basées sur l'ajustement de relations statistiques entre le signal fourni par le capteur et l'information que

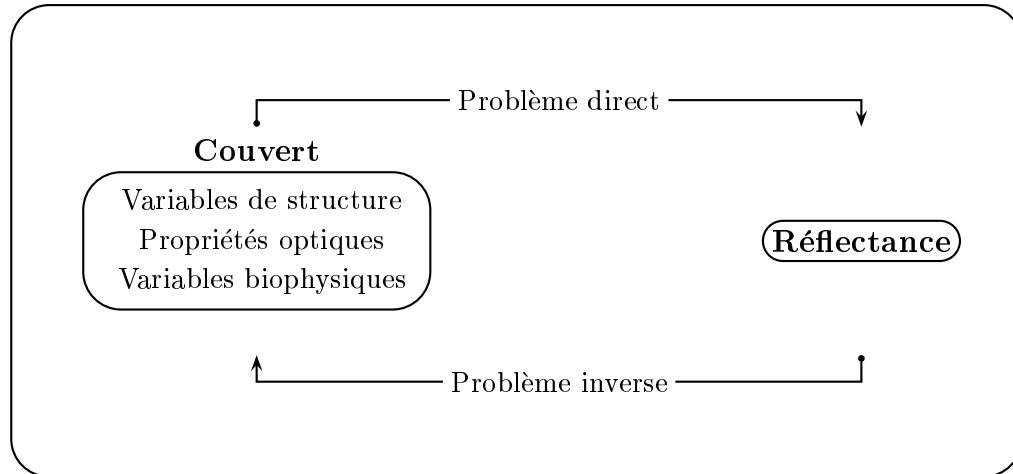


Figure 1.3 : Problèmes direct et inverse.

l'on désire acquérir sur la surface observée. Ces relations utilisent souvent les indices de végétation. Les bases de données sont "empiriques", c'est-à-dire issues d'expérimentations où sont mesurées simultanément les réflectances et les caractéristiques des couverts ;

- les méthodes "physiques" qui s'appuient sur les lois fondamentales de la physique (modèles de transfert radiatif basés sur les équations d'onde de Maxwell ou modèles SVAT basés sur le bilan d'énergie) ;

En combinant ces deux catégories, on peut en définir une troisième qui s'est développée ces dernières années et qui correspond aux méthodes "hybrides" qui consistent à établir des relations empiriques sur des jeux de données simulées par des modèles de transfert radiatif (utilisation de "look-up tables", réseaux de neurones).

1.3.1 Les indices de végétation

Le principe des indices de végétation consiste à réaliser une combinaison, linéaire ou non, de réflectances (ρ) dans différentes longueurs d'onde et à la relier à une variable caractéristique de la végétation (LAI , $fAPAR$, pourcentage de couverture du sol, ...). Ils s'appuient essentiellement sur les différences de propriétés optiques de la végétation dans le rouge (600-700 nm)

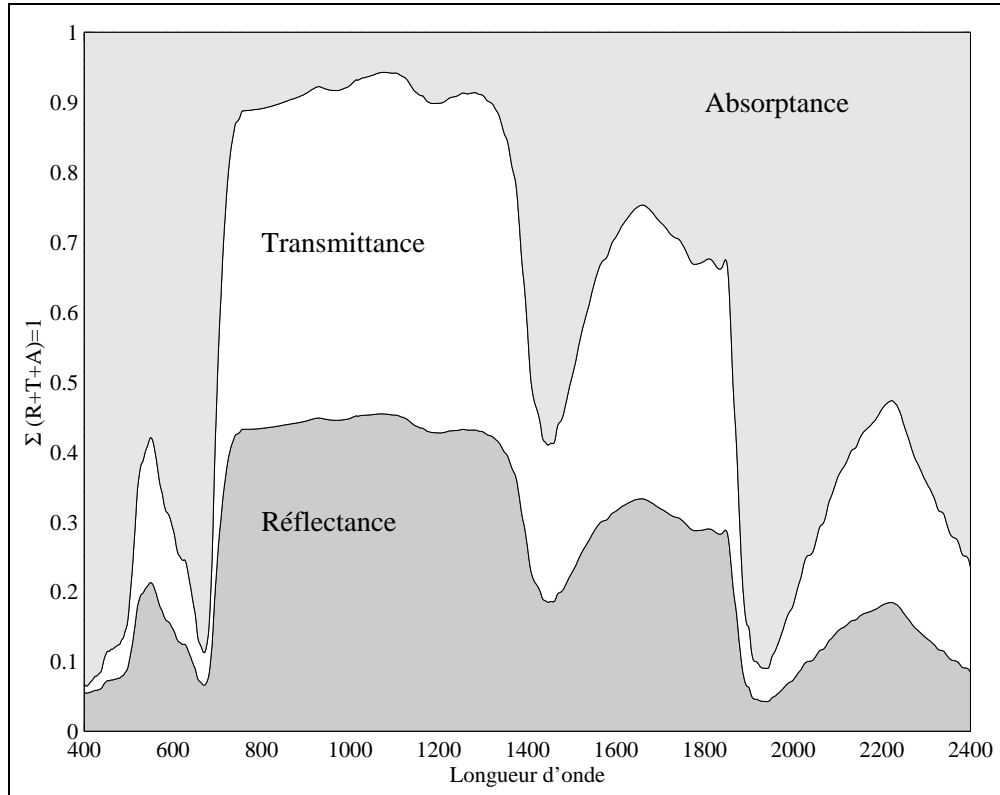


Figure 1.4 : Comportement radiatif de la végétation dans le domaine solaire (teneur en chlorophylle des feuilles : $35 \mu\text{g}/\text{cm}^2$, en eau = 0.01 cm^{-1} et en matière sèche = $5 \mu\text{g}/\text{cm}^2$).

et le proche infrarouge (700-1300 nm). En effet, dans le rouge (R), le pigment foliaire principal, la chlorophylle, présente un pic d'absorption (690 nm) alors que son absorption est nulle dans le proche infrarouge (PIR) à partir de 850 nm (figure 1.4). Les corrélations entre les indices de végétation et des variables comme le LAI ou le $fAPAR$ ont été étudiées à partir de nombreux travaux expérimentaux (Wiegand *et al.*, 1991, Asrar *et al.*, 1984, ...). Enfin, le lien qui existe entre les indices de végétation, le LAI et le $fAPAR$ a été démontré théoriquement par plusieurs auteurs (Sellers, 1987, Baret et Guyot, 1991, Myneni *et al.*, 1995).

Les premiers indices de végétation, largement utilisés ensuite pour

l'exploitation des données *NOAA/AVHRR* (Tucker, 1979) furent le *RVI*⁴ (Pearson et Miller, 1972, Eq 1.1) et le *NDVI*⁵ (Rouse *et al.*, 1974, Eq 1.2).

$$RVI = \frac{\rho_R}{\rho_{PIR}} \quad (1.1)$$

$$NDVI = \frac{\rho_{PIR} - \rho_R}{\rho_{PIR} + \rho_R} \quad (1.2)$$

Ces indices sont reliés aux variables biophysiques du couvert par des lois empiriques (régressions linéaires ou lois exponentielles) calées sur des jeux de données expérimentales :

- le *LAI* : Asrar *et al.*(1985), Bégué (1993), ...
- le *fAPAR* : Asrar *et al.*(1984), Sellers (1985), Baret et Olioso (1989), ...
- la biomasse : Tucker (1979), ...
- le pourcentage de couverture du sol : Carlson et Ripley (1997), Laurence et Ripple (1997), ...

Ces relations sont très sensibles aux différentes composantes, autres que la végétation, qui interviennent dans la mesure de réflectance. C'est la raison pour laquelle l'expression de ces indices a été modifiée pour s'affranchir au mieux des effets parasites (sol, atmosphère, structure des couverts, conditions d'illumination).

Ainsi, les propriétés optiques du sol, en particulier le fait qu'il soit plus ou moins clair, modifient les relations *LAI/NDVI*, *fAPAR/NDVI* (Huete *et al.*, 1988). Le *SAVI* (Soil Adjusted Vegetation Index, Huete *et al.*, 1988) et ses dérivés (*TSAVI*, Baret *et al.*, 1989 , *MSAVI*, Qi *et al.*, 1994) grâce à l'intervention d'un facteur d'étalonnage du sol permettent de minimiser les perturbations. D'autres types d'indices (*PVI*, Richardson et Wiegand, 1977, *WDVI*, Clevers, 1989) sont aussi utilisés. Ils sont basés sur des mesures de distance entre la réflectance du couvert et la droite des sols

⁴Ratio Vegetation Index

⁵Normalized Difference Vegetation Index

(Baret *et al.*, 1993) qui exprime la réflectance d'un sol nu dans le proche infrarouge en fonction de celle de ce même sol dans le rouge.

Les travaux se sont également focalisés sur les perturbations dues à l'atmosphère et aux nuages. L'algorithme du "Maximum Composite Value" (Tarpley *et al.*, 1984) sélectionne la valeur maximale du *NDVI* sur une séquence d'images correspondant, en général, à une fenêtre temporelle d'une dizaine de jours. Cela permet, en particulier, de s'affranchir des problèmes de nuages. Par ailleurs, Pinty et Verstraete (1992) ont développé un indice de végétation non linéaire, le *GEMI*, qui combine les réflectances dans le rouge et le proche infrarouge de manière quadratique de façon à minimiser les perturbations atmosphériques. D'autres indices de végétation comme l'*ARVI* (Kaufman et Tanré, 1992) ont aussi vu le jour. Ils utilisent la mesure de réflectance dans le bleu où les effets de l'atmosphère sont les plus importants (diffusion de Rayleigh et diffusion des aérosols). Enfin, le *SARVI* et le *MSARVI* (Huete et Liu, 1994) permettent à la fois de minimiser les effets du sol et de l'atmosphère. Il est cependant précisé que ces indices ne sont efficaces que lorsque les perturbations atmosphériques ont subi une correction partielle (en particulier pour la diffusion moléculaire et l'absorption de l'ozone).

L'ensemble de ces indices reste donc relativement sensible aux effets de sol et de l'atmosphère (Verstraete et Pinty, 1996). De plus, la variabilité des conditions d'illumination et la structure de la végétation nécessitent de caler les relations "variable biophysique/indice de végétation" pour chaque type de couvert (Jackson, 1986, Shibayama *et al.*, 1986, Baret et Olioso, 1989, Bégué, 1993). En outre, les indices de végétation saturent lorsque les couverts sont moyennement ou très denses. Cela implique qu'une même valeur d'indice peut correspondre à deux couverts très différents en terme de structure ou de densité de feuillage (Baret et Guyot, 1991, Myneni et Williams, 1994).

L'utilisation d'indices de végétation permet néanmoins d'avoir une assez bonne quantification des variables biophysiques, en particulier, lorsque l'on dispose d'une relation dont les paramètres sont calés pour le type de couvert que l'on observe. Cette méthode présente le grand avantage d'être rapide et facile à mettre en œuvre. Il est tout de même difficile d'avoir une estimation très précise des variables biophysiques du couvert du fait de la "simplicité" de la formulation des indices de végétation. De plus, ces indices sont classiquement calculés pour une seule configuration de visée (le nadir qui correspond à une visée verticale). Des travaux récents, comme ceux de Walter-

Shea *et al.* (1997) ont montré que certaines configurations angulaires sont plus adaptées que d'autres pour l'estimation du *LAI* et du *fAPAR* sur une parcelle de luzerne à partir du *NDVI*. De même, Roujean et Bréon (1995) ont montré que l'indice de végétation qu'ils ont développé, le *RDVI*⁶ permet de bien estimer le *fAPAR* mais pour une configuration géométrique bien précise (soleil à 45° et capteur à 60° dans le plan principal). Cependant, peu de travaux ont porté sur l'utilisation de la variation directionnelle de la réflectance dans les relations indicielles pour l'estimation des variables biophysiques des couverts végétaux. Pourtant, celle-ci apporte de l'information supplémentaire sur le type de couvert observé : en particulier, lorsque le couvert est observé dans la direction du soleil, on observe un pic de réflectance ou "hot spot" qui est directement relié à la structure du couvert (Kuusk, 1985).

1.3.2 L'inversion de modèles de transfert radiatif

Contrairement aux indices de végétation basés sur le comportement de la végétation dans certains domaines de longueurs d'onde et établis de façon empirique, la modélisation nécessite de comprendre les mécanismes d'interaction entre le rayonnement et la matière. Il faut donc décrire la façon dont le rayonnement lumineux est absorbé ou diffusé par les différents éléments qu'il traverse (atmosphère, couvert végétal, sol, océans, ...). Dans le domaine solaire, le système de télédétection comprend (Goel, 1988, figure 1.5) :

- une source de lumière directe, le soleil, caractérisée par son intensité et sa position et une source de lumière diffuse provenant de l'atmosphère ;
- un couvert végétal caractérisé par les propriétés optiques de ses composants (feuilles, tiges, troncs, ...) et des variables de structure ;
- un sol caractérisé par une rugosité liée à sa structure, des propriétés optiques intrinsèques liées au type de sol, une humidité et une structure ;
- un capteur caractérisé par ses propriétés spectrales, radiométriques, spatiales et directionnelles.

Chacune des composantes de ce système, ainsi que leurs interactions, sont mises en équation dans les modèles de transfert radiatif. Dans cette étude,

⁶Renormalized Difference Vegetation Index

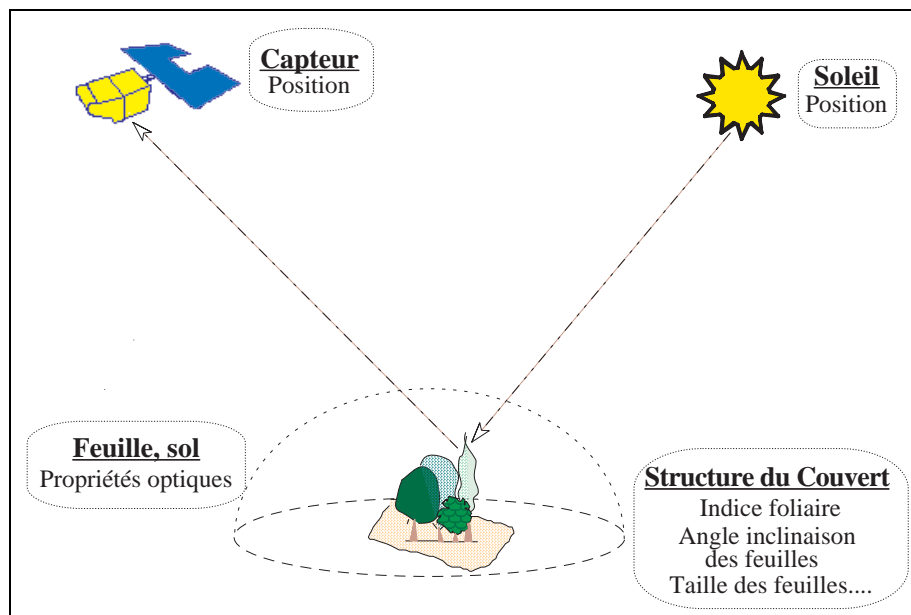


Figure 1.5 : Système de télédétection au niveau du couvert.

nous nous plaçons au niveau du couvert, c'est-à-dire que nous ne tenons pas compte des perturbations atmosphériques qui interviennent dans la trajectoire du rayonnement, *i.e.* on considère qu'elles ont été corrigées. Dans le domaine solaire, ces dernières sont dues à deux phénomènes (Tanré *et al.*, 1990) :

- la diffusion par les aérosols (diffusion de Mie) et par les molécules (diffusion de Rayleigh) qui modifie la trajectoire du rayonnement. L'importance de ces phénomènes décroît en fonction de la longueur d'onde λ (de l'ordre de $1/\lambda^4$ pour la loi de Rayleigh et de $1/\lambda$ pour celle de Mie) car ils sont fonction de la taille des particules. Contrairement à la diffusion des aérosols qui peut être très variable dans l'espace et dans le temps, la diffusion moléculaire est relativement constante et dépend essentiellement de la pression atmosphérique ;
- l'absorption gazeuse qui est essentiellement due à la présence d'ozone, d'eau, d'oxygène et de dioxyde de carbone. Ce phénomène peut être évité si les bandes spectrales des capteurs ne sont pas situées dans les bandes d'absorption de ces gaz. Dans le cas de *VEGETATION*, seuls

l'eau et l'ozone perturbent le signal, mais faiblement en comparaison de la diffusion des molécules et des aérosols ;

Le principal problème revient donc à corriger les effets des aérosols. Dans un premier temps, on cherche à détecter une information sur l'épaisseur optique. Celle-ci peut provenir d'autres capteurs ou de données climatologiques extérieures. Mais il existe également une méthode qui consiste à utiliser les propriétés des objets présents dans l'image (Koepke *et al.*, 1979, Kaufman et Sendra, 1988). On sélectionne des surfaces caractérisées par un très faible signal dans une longueur d'onde donnée : par exemple, une forêt dans le bleu et le rouge (correction atmosphérique des données *NOAA/AVHRR*) ou encore un lac dans le proche infrarouge (correction atmosphérique des données *LANDSAT/TM*). Pour ces pixels particuliers, dans les longueurs d'onde considérées, la contribution de l'atmosphère au signal obtenu au niveau du capteur est maximale. L'inversion d'un modèle de transfert radiatif dans l'atmosphère permet alors d'évaluer l'épaisseur optique en faisant l'approximation que la réflectance au niveau du couvert est nulle. Dans le cas de *VEGETATION*, on pourra, par exemple, utiliser l'algorithme SMAC⁷ développé par Rahman et Dedieu (1994), dans la bande bleue pour des pixels "type forêt". La deuxième étape de la correction atmosphérique consiste à inverser le modèle en utilisant l'épaisseur optique précédemment évaluée, pour déterminer la réflectance au niveau du couvert sur l'ensemble des pixels de l'image, dans tous les canaux du capteur.

Une fois la correction atmosphérique effectuée, on obtient des images de réflectance au niveau du couvert. Il est alors possible de modéliser le transfert radiatif au sein de la végétation et du sol (voir la revue de Myneni *et al.*, 1995 et l'état de l'art présenté au § 2.1.1). À partir des propriétés optiques et structurelles du sol et de la végétation, on définit une relation f , la plupart du temps non linéaire, qui permet d'accéder à la réflectance. La résolution du problème inverse consiste à déterminer la relation inverse f^{-1} qui relie la mesure aux caractéristiques du sol ou de la végétation (Verstraete *et al.* 1996, Eq 1.3) :

$$x_j = f_{ij}^{-1}(y_i) \tag{1.3}$$

⁷Simplified Method for the Atmospheric Correction

où $X = [x_j, j = 1, \dots, m]$ sont les variables du couvert végétal et $Y = [y_i, i = 1, \dots, n]$ sont les réflectances mesurées. L'indice i peut correspondre à plusieurs facteurs de variation comme la longueur d'onde ou les conditions d'observation (angles de visée ou position solaire). Pour que l'inversion mathématique de ces modèles soit possible, il est nécessaire de remplir certaines conditions :

- le nombre n de mesures indépendantes de réflectance doit être au moins égal au nombre m de paramètres du modèle. En pratique, du fait de la non-linéarité et des hypothèses faites dans la plupart des modèles, ainsi que de leur imprécision, il est fortement conseillé d'avoir $n > m$. Or, plus le modèle est précis, plus les équations qui le régissent sont complexes, et donc le nombre de paramètres important. De plus, le nombre de mesures est limité par la distribution angulaire définie par le capteur. Il faut ajouter que, pour une position angulaire donnée, le nombre de longueurs d'onde "physiquement indépendantes" ne dépasse pas une dizaine (Price, 1990, Hall *et al.*, 1995) ;
- Lorsqu'un modèle mathématique est non linéaire, il est fréquent qu'il ne soit pas analytiquement inversible. Il faut alors utiliser des méthodes d'optimisation pour estimer les paramètres qui, si elles permettent d'obtenir une solution probable, ne garantissent pas l'unicité de cette solution.

Les méthodes d'optimisation utilisées pour les inversions mathématiques reposent toutes sur la minimisation d'un critère ou d'une fonction d'erreur \mathcal{F} , qui traduit généralement une distance entre la variable \hat{Y} simulée par le modèle et la variable Y mesurée (Delécolle *et al.*, 1995). La méthode la plus courante est la minimisation du critère au sens des moindres carrés (Eq 1.4) :

$$\begin{aligned} \mathcal{F} &= \sum_i w_i [y_i - \hat{y}_i(x_{j,j=1,\dots,m})]^2 \\ &= \sum_i w_i [y_i - f_{ij}(x_{j,j=1,\dots,m})]^2 \end{aligned} \quad (1.4)$$

où w_i est le poids attribué à chaque mesure. L'utilisation d'une pondération permet par exemple, de favoriser les mesures les plus précises ou bien de

favoriser les cas de figures qui vont le plus influencer le résultat de l'estimation de la variable biophysique (Goel et Thomson, 1984).

La méthode consiste à calculer la distance \mathcal{F} entre la variable simulée à partir de valeurs initiales des paramètres du modèle, et les valeurs mesurées. Le processus est alors itératif : tant que la fonction d'erreur \mathcal{F} n'a pas atteint un minimum, les paramètres du modèle sont modifiés. Idéalement, si le modèle est mathématiquement inversible, \mathcal{F} est égale à 0. Cette méthode pose deux problèmes délicats :

- *les minima locaux* : lorsque les modèles ne sont pas linéaires, il est fréquent que le processus itératif soit stoppé à cause d'un minimum local qui donne une solution parmi l'ensemble des possibles. Ce phénomène dépend fortement des valeurs initiales des paramètres du modèle (Pinty et Verstraete, 1991). Il est donc important de donner des valeurs initiales proches de la solution et d'opérer l'inversion pour plusieurs jeux de paramètres initiaux. On peut, en outre, contraindre l'algorithme d'inversion en imposant, par exemple, une gamme de variation pour chacun des paramètres (Privette *et al.*, 1994) ;
- *le temps de calcul* : du fait de la nature itérative du processus, le temps de calcul pour l'inversion du modèle est important, et ce, d'autant plus que le nombre de paramètres à estimer est grand.

Pinty et Verstraete (1992) sont arrivés à la conclusion que l'on ne peut pas espérer retrouver plus de cinq ou six paramètres par inversion de modèle à l'aide d'une méthode d'optimisation. Lorsque l'on utilise des modèles plus complexes, l'une des solutions est de réaliser au préalable une analyse de sensibilité du modèle aux paramètres et de fixer les valeurs des paramètres les moins sensibles (Bicheron et Leroy, 1999). Un autre moyen d'améliorer les résultats est d'utiliser les mesures qui discrimineront le plus les variables que l'on cherche à estimer (par exemple, le proche infrarouge pour l'estimation de la surface foliaire, ou encore l'utilisation de mesures dans le plan principal pour l'estimation de paramètres de structure du couvert observé).

L'extraction de variables biophysiques par inversion de modèle de transfert radiatif permet d'obtenir de meilleures précisions qu'avec des indices de végétation. Cependant, elle présente aussi des inconvénients comme le

temps de calcul et l'initialisation des paramètres. Elle est, de plus, limitée par les hypothèses du modèle employé : il est souvent préférable d'utiliser certaines catégories de modèles pour certains types de couverts, comme par exemple, des modèles de type "géométrique" pour les forêts de conifères et des modèles de type "turbide" pour les prairies (Strahler *et al.*, 1995). C'est la raison pour laquelle des méthodes dites "hybrides" ont été développées dans le but de créer des algorithmes globaux, permettant d'appliquer un même algorithme, sans connaissance *a priori*, pour l'ensemble des couverts observés par un satellite.

1.3.3 Les méthodes hybrides

Les méthodes hybrides combinent des relations empiriques et des relations physiques entre réflectance mesurée et variable biophysique. Elles consistent à établir, souvent par le biais de la modélisation, des bases de données de référence combinant tous les types de couvert caractérisés par des variables de structure et des variables biophysiques, et les réflectances associées. Une loi empirique élaborée sur cette base de données permet alors, lorsque l'on présente une nouvelle mesure, d'attribuer les variables correspondantes.

1.3.3.1 Les tables de correspondance

C'est une méthode très récente dans le domaine de l'estimation des caractéristiques biophysiques des couverts végétaux. Elle a été développée par Gobron *et al.* (1997), Knyazikhin *et al.* (1998a et 1998b) pour des capteurs à venir (*MODIS*⁸, *MISR*⁹). L'approche consiste à réaliser une table de correspondance, c'est-à-dire une base de données la plus complète possible. Dans les cas cités, elle est obtenue à l'aide d'un modèle tridimensionnel de transfert radiatif qui offre une bonne précision de simulation, quelque soit le type de couvert considéré (prise en compte d'une hétérogénéité spatiale horizontale et verticale). Dans le cas des algorithmes développés pour les capteurs *MISR* et *MODIS*, les biomes sont divisés en 6 catégories et chaque catégorie correspond à une large variété de couverts. L'algorithme consiste alors, lorsqu'on présente un nouveau jeu de mesures, à lui attribuer les variables biophysiques les plus probables avec un intervalle de confiance. Cette méthode est en cours de validation sur des données aussi bien synthétiques qu'expérimentales.

⁸MODerate resolution Imaging Spectrometer

⁹Multi-angle Imaging Spectro-Radiometer

1.3.3.2 Les réseaux de neurones artificiels

État de l'art

Ils consistent à relier une sélection de variables d'entrée (réflectance) à une ou des variables de sorties du réseau (variables biophysiques). La relation est établie lors d'une phase d'apprentissage sur une base de données de référence. Peu de travaux de télédétection ont utilisé cette méthode dans notre domaine d'intérêt. Nabney et Bishop (1995) estiment la direction du vent à partir des données radar. Une autre méthode d'estimation de la direction du vent a été également développée par Thiria *et al.*(1993), puis validée par Mejia *et al.*(1998) sur des données provenant d'ERS1. Les résultats montrent que la méthode basée sur des réseaux de neurones est plus performante que les autres méthodes existantes. De même, Jin et Liu (1997) ont développé un réseau de neurones pour estimer des variables de biomasse à partir de données micro-onde passive (émissivité) et de télédétection active (coefficients de rétrodiffusion). Les résultats obtenus sont très satisfaisants. Kimes *et al.*(1997) ont aussi montré que les réseaux de neurones étaient performants pour l'estimation des paramètres caractéristiques des forêts (biomasse, âge et hauteur des arbres, ...) sur des données radar simulées. Dans le domaine solaire, Smith (1993) a développé un réseau neuronal pour estimer l'indice foliaire de végétations éparses à partir du NDVI et de la réflectance au nadir à 550 nm. Les résultats sont très satisfaisants et bien meilleurs en comparaison de ceux obtenus grâce à la simple technique de régression multiple. Ils ont, de plus, été validés sur des données expérimentales. Les travaux de Fourty et Baret (1997) montrent également une amélioration de la précision d'estimation de diverses variables biophysiques (teneur en eau, indice foliaire, contenu en matière sèche du couvert) en utilisant des réseaux de neurones plutôt qu'une technique de régression multiple. L'estimation des variables est réalisée à partir de réflectances au niveau du satellite, dans une sélection de bandes. Enfin Baret *et al.*(1995), ont développé une méthode d'estimation du pourcentage de sol nu performante, sur des cultures de betteraves à partir de mesures radiométriques.

Principe

Les réseaux de neurones sont des réseaux fortement connectés de processeurs élémentaires (neurones) fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie sur la base d'information qu'il reçoit (Touzet,

1992). Il est caractérisé par une fonction de transfert F , un biais b , une sortie unique x et un certain nombre d'entrées $[a_i, i = 1, \dots, n]$. À chacune de ces entrées est associé un poids w_i (figure 1.6).

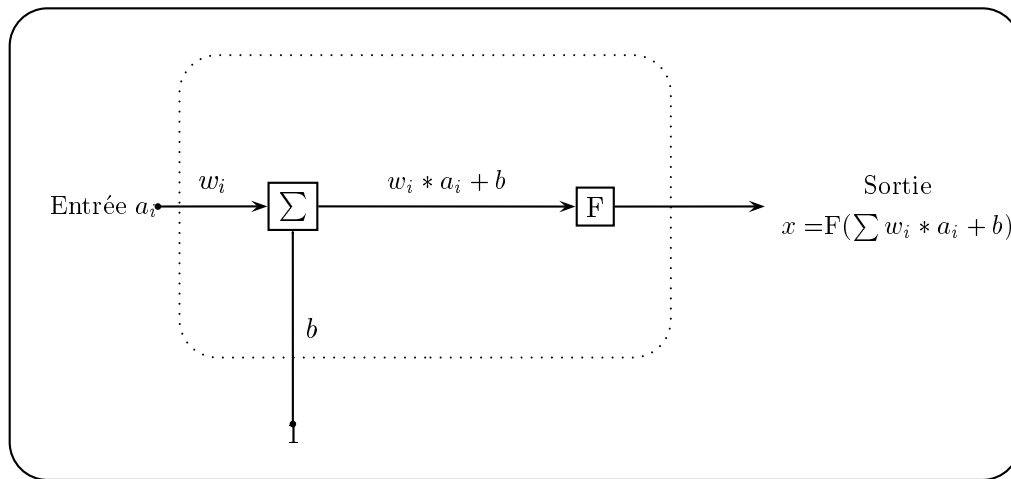


Figure 1.6 : Représentation schématique d'un neurone.

Le réseau de neurones (figure 1.7) est défini par

- *une architecture* : c'est-à-dire, pour un réseau multicouche, le nombre de couches qu'il comprend, le nombre de neurones pour chacune des couches (pour la première cela correspond au nombre d'entrées et pour la dernière, au nombre de sorties), et enfin la fonction de transfert F (qui est généralement, une fonction à seuil, linéaire ou bien de type sigmoïde)
- *une règle d'apprentissage* : c'est une boucle fermée qui consiste à présenter des exemples d'entrées au réseau, à calculer les sorties et à les comparer avec les sorties désirées. La comparaison permet, grâce à des techniques d'optimisation, de modifier les paramètres du réseau (poids et biais de chaque neurone) pour converger vers la solution.

L'apprentissage du réseau de neurones est une phase longue puisque, comme les méthodes d'optimisation, elle fait appel à des techniques itératives. Cependant, une fois terminée, le calcul de la sortie pour une nouvelle entrée est quasiment instantané. La plus grande limitation de ce type de méthode est le soin qu'il faut apporter à l'élaboration de la base d'apprentissage, constituée

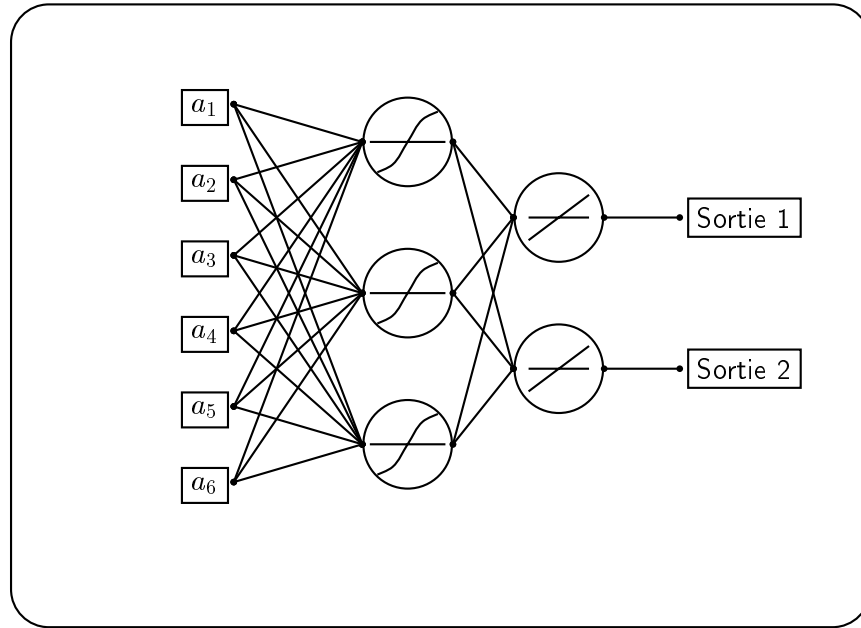


Figure 1.7 : Représentation schématique d'un réseau de neurones comprenant une couche d'entrée, une couche cachée à fonction de transfert sigmoïde et une couche de sortie à fonction de transfert linéaire.

d'exemples de couples (variables d'entrée, sortie désirée). En effet, comme pour les tables de correspondance, celle-ci doit être la plus représentative possible de l'ensemble des cas que l'on peut rencontrer. Le réseau de neurones est cependant capable, contrairement aux tables de correspondance, d'interpoler des solutions dans cette base d'apprentissage.

1.4 Objectif et plan de l'étude

L'objectif de cette thèse est de développer un algorithme global de suivi des couverts végétaux, c'est-à dire applicable aussi bien à des zones cultivées qu'à des zones de végétation naturelle. Ce suivi doit se faire au moyen de l'estimation de variables biophysiques du couvert à partir de données satellitaires. Cette étude se situe dans le cadre du programme préparatoire au capteur *VEGETATION/SPOT4* construit par la société Aérospatiale et opéré par le CNES depuis le printemps 1998. Elle doit cependant pouvoir s'étendre

à des images provenant de capteurs possédant des caractéristiques spectrales et directionnelles différentes.

Compte tenu des limitations bien établies pour les techniques basées sur les indices de végétation ainsi que l'inversion de modèles de transfert radiatif et des résultats prometteurs des méthodes hybrides, l'algorithme de suivi sera basé sur une approche semi-empirique combinant un modèle de transfert radiatif et des réseaux de neurones. Les performances de cet algorithme seront évaluées sur différents jeux de données, aussi bien simulées qu'expérimentales.

Ce travail s'organise autour de trois parties correspondant chacune à un chapitre :

Chapitre 2 : Construction d'une base de données synthétiques.

L'algorithme que nous souhaitons développer doit pouvoir s'appliquer à une large gamme de couverts végétaux et s'appuie sur l'utilisation de réseaux de neurones. La première phase consiste donc à établir une base d'apprentissage. Comme il n'existe pas de base de données expérimentale composée de mesures portant sur des couverts variés et réalisées dans les conditions requises pour notre étude (caractéristiques du capteur *VEGETATION*), il est nécessaire de créer une base de "données synthétiques". Ces données sont simulées à l'aide d'un modèle de transfert radiatif (SAIL) connu et validé par de nombreux auteurs. La variabilité des couverts simulés est obtenue grâce à une large gamme de variation des variables d'entrée du couvert. À la suite de cette étude, nous disposons d'une base de données composée de 1500 couverts végétaux observés à des latitudes et des dates variées. Chacun d'entre eux est caractérisé par des propriétés structurelles, optiques et biophysiques associées aux réflectances au niveau du couvert dans les bandes spectrales et les directions d'échantillonnage du capteur *VEGETATION*.

Chapitre 3 : Estimation de variables biophysiques des couverts végétaux. Calage des réseaux de neurones.

Cette partie correspond à la phase de développement de l'algorithme, basé sur l'utilisation de réseaux de neurones.

Pour chaque couvert, les données de réflectance sont d'abord norma-

lisées afin d'obtenir un nombre constant de variables d'entrée, et faciliter l'apprentissage des réseaux de neurones. Les réseaux de neurones sont ensuite calés sur une partie de la base de données synthétiques. Une première validation est alors effectuée sur l'autre partie afin de tester et comparer les performances de l'algorithme pour l'estimation de différentes variables biophysiques : le *LAI*, le *fAPAR*, la fraction de trou ainsi que la teneur en chlorophylle intégrée au niveau du couvert. Les performances des réseaux de neurones sont aussi comparées à celles d'une méthode classique basée sur le *NDVI*. De plus, l'influence de la variabilité intrapixellaire c'est-à-dire de la présence de couverts végétaux différents dans un même pixel (de taille $1,15 \text{ km} \times 1,15 \text{ km}$) sur l'estimation des variables est étudiée. Cette partie de l'étude permet de caler le réseau de neurones mais aussi de déterminer les variables biophysiques qui sont les mieux estimées.

Chapitre 4 : *Validation de l'algorithme de suivi des couverts végétaux.*

L'algorithme, calé sur des données synthétiques, doit être ensuite validé sur des données proches des conditions réelles. La validation est réalisée sur deux types de jeux de données. Dans un premier temps, elle est effectuée sur un jeu de données réelles établi sur une parcelle de betterave en 1990 avec des radiomètres CIMEL. Puis, une deuxième validation est réalisée, cette fois sur l'expérimentation Alpilles/ReSeDA qui se rapproche beaucoup plus des conditions réelles d'échantillonnage directionnel de VEGETATION. Pour chacune de ces expérimentations, il est nécessaire d'adapter les réseaux de neurones, en particulier aux bandes spectrales des capteurs utilisés durant les campagnes.

Enfin, cette étude s'achève par une conclusion rappelant les principaux résultats obtenus, les limites de la méthode et décrivant les perspectives que peut apporter ce travail.

Chapitre 2

Construction d'une base de données synthétiques.

L'utilisation de réseaux de neurones impose que nous ayons à notre disposition une base d'apprentissage composée d'exemples de couples d'entrées (réflectances mesurées et conditions d'observation) et de sorties (variables biophysiques des couverts végétaux). Celle-ci doit être la plus représentative de l'ensemble des cas que l'on peut rencontrer. Le capteur *VEGETATION*, de par ses caractéristiques orbitographiques, permet la couverture de l'ensemble des terres émergées avec une cadence quasi journalière. Il faut donc que le développement de notre algorithme soit basé sur un jeu de données le plus large possible comprenant une très large gamme de couverts végétaux. Ce jeu de données doit comprendre la réflectance du couvert mesurée par le capteur et les variables biophysiques du couvert correspondantes. Or, il n'existe pas à ce jour de base de données expérimentales suffisamment complète pour satisfaire ces besoins (variabilité de couverts, données de réflectance répondant aux caractéristiques de *VEGETATION*). Compte tenu de la taille d'un tel jeu de données, il est clair que l'acquisition d'une base expérimentale n'est pas envisageable. C'est la raison pour laquelle nous allons faire appel à des modèles de transfert radiatif. De nombreux auteurs utilisent ainsi des données simulées, et valident ensuite les résultats sur quelques jeux de données expérimentales (Choudhury, 1987, Bégué, 1993, Jacquemoud, 1993, Huete et Liu, 1994, Rahman, 1998, ...).

Ce chapitre est donc consacré à la construction de la base de données

qui va nous servir pour développer notre algorithme d'estimation de variables biophysiques des couverts végétaux.

Dans un premier temps, une revue des modèles de réflectance du couvert existants est présentée, suivie d'une description détaillée des modèles utilisés dans cette étude.

Nous définissons ensuite les variables biophysiques que nous voulons atteindre par le biais de notre algorithme et qui doivent donc être simulées.

Enfin, nous présentons la gamme de variation qui a été choisie pour chacune des variables des modèles utilisés, puis quelques résultats obtenus sur les simulations.

2.1 Les modèles utilisés

2.1.1 Les modèles de réflectance du couvert

La réflectance (ou facteur de réflectance) spectrale d'un couvert (notée ρ), pour un rayonnement incident (ou luminance) donné, est le rapport de la quantité de rayonnement diffusée par le couvert à la quantité de rayonnement réfléchi par un réflecteur diffusant parfaitement lambertien dans les mêmes conditions d'éclairement (L_i). Lorsqu'on n'utilise pas de surface de référence, mais qu'on mesure l'éclairement incident E , L_i est égal à E/π . La compréhension du signal de réflectance d'un couvert passe donc par la compréhension de l'interaction entre les radiations électromagnétiques et la matière pour une longueur d'onde donnée et consiste, par conséquent, à rechercher la solution des équations de Maxwell. Le couvert végétal est un milieu composé de nombreux éléments (feuilles, branches, tiges, fleurs, sol, fruits ...) dont l'agencement est complexe, et la résolution directe de ces équations n'est donc pas envisageable. Par contre, il est possible d'approcher la solution en utilisant des modèles de réflectance de couvert qui correspondent à une résolution approchée des équations de Maxwell.

L'objectif de ce paragraphe est de réaliser un bref état de l'art des différents types de modèles que l'on peut trouver dans la littérature. Ces modèles sont classés en 4 catégories (Goel, 1988).

2.1.1.1 Les différents types de modèles

Les modèles informatiques

Ce sont les modèles les plus précis et les plus réalistes mais aussi les plus gourmands en puissance et en temps de calcul. Une maquette tridimensionnelle du couvert végétal est d'abord construite en caractérisant chaque élément de la végétation par un ensemble de facettes élémentaires modélisées par des surfaces géométriques simples dont la position, l'orientation et les propriétés optiques sont bien déterminées. Différents types de modèles se basent sur cette structure pour déterminer avec précision le trajet de la lumière à l'intérieur du couvert : modèles de radiosité (Borel *et al.*, 1991) ou de lancer de rayons (Chelle, 1997) qui sont basés sur la méthode de Monte-Carlo (Ross et Marshak, 1991).

Les modèles géométriques

Ils assimilent le couvert végétal à un ensemble d'objets géométriques simples (cones, cylindres, sphères, ...). Chaque objet est caractérisé par sa position et ses dimensions (qui définissent précisément l'architecture de la canopée), ainsi que les propriétés optiques de sa surface (reflectance, transmittance et absorptance). La réflectance du couvert est alors déterminée en utilisant les lois de l'optique géométrique pour analyser les régions éclairées ou ombrées par le couvert. Ces modèles sont bien adaptés à la simulation de couverts peu denses, en particulier les forêts de conifères (Li et Strahler, 1992, Chen et Leblanc, 1997) ou les végétations éparées (Bégué, 1992).

Les modèles "milieu turbide"

Dans cette approche, les éléments de la végétation sont considérés comme des particules absorbantes et diffusantes distribuées aléatoirement, mais avec une orientation angulaire donnée. La description de l'architecture du couvert est donnée seulement par un indice foliaire et la distribution d'angles d'inclinaison des feuilles. Les modèles "milieu turbide" sont en général destinés à simuler la réflectance de couverts assez homogènes et denses, comme, par exemple, des prairies ou certaines cultures (Suits, 1972, Verhoef, 1984 et 1985, Kuusk, 1994).

Les modèles hybrides

Les modèles hybrides sont une combinaison des modèles précédents. Pour certains de ces modèles (Myneni *et al.*, 1992, Gastellu-Etchegorry *et al.*, 1996), le couvert végétal est discrétisé en cellules cubiques comprenant plus ou moins de végétation. Celle-ci est caractérisée, pour chaque cellule, par certaines propriétés optiques, une orientation angulaire et une densité de surface foliaire données. Ces modèles nécessitent beaucoup de temps de calcul puisqu'ils sont basés sur une description détaillée de la végétation. Ils sont particulièrement bien adaptés à la simulation de couverts épars ou de forêts. Le modèle d'España (1997), quant à lui, est une combinaison d'un modèle géométrique (réflectance simple) et d'un modèle "milieu turbide" (réflectance multiple), basée sur une maquette tridimensionnelle informatique d'un couvert de maïs. Enfin, il est possible de simuler des couverts hétérogènes grâce au couplage d'un modèle géométrique (arbres) et d'un modèle turbide (herbes) (Bégué *et al.*, 1996).

*Pour notre application, dans la mesure où nous cherchons à simuler une base de données importante (1500 couverts), il n'est pas envisageable d'utiliser des modèles basés sur une description très détaillée du couvert, essentiellement pour des raisons de temps de calcul et de construction de maquettes tridimensionnelles, mais aussi par la difficulté de fournir des descriptions très détaillées de couverts types (Baret *et al.*, 1997). Le modèle SAIL¹ que nous avons choisi a été développé par Verhoef (1984 et 1985). C'est un modèle "milieu turbide" qui présente un bon compromis entre son niveau de complexité, sa rapidité d'exécution et le réalisme des simulations obtenues.*

2.1.1.2 Le modèle SAIL

Validité du modèle

SAIL est un modèle utilisé dans de nombreuses applications et bon nombre de travaux portent sur le test de ses performances. Badwahr *et al.* (1985) ont mené une étude comparative sur des données de maïs et de soja entre le modèle de Suits (Suits, 1972) et SAIL montrant que celui-ci était plus précis. Par ailleurs, España (1997) a montré que, pour le maïs, SAIL était légèrement moins précis que son modèle hybride. Il présente néanmoins l'avantage d'être facile à mettre en oeuvre comparativement au modèle d'España qui nécessite

¹SAIL : Scattering by Arbitrary Inclined Leaves

la construction d'une maquette 3D. Enfin, SAIL a été validé sur des données expérimentales de soja (Goel et Thomson, 1984), de vergers (Goel et Deering, 1989) ou encore de maïs (Major *et al.*, 1992). De plus, Bicheron et Leroy (1999) ont montré que le modèle de Kuusk (1995), dérivé de SAIL, présente d'assez bonnes performances sur les végétations éparses et les forêts boréales.

Rappels théoriques

SAIL est dérivé des équations de Kubelka-Munk (1931) qui donnent une solution approchée de l'équation du transfert radiatif : elles décrivent le transfert de flux diffus dans un milieu absorbant ou diffusant, homogène et infiniment étendu, pour un rayonnement incident directionnel (Eq 2.1). Ce système a été complété par Suits (1972), en rajoutant une composante dans la direction d'observation (Eq 2.2) :

$$\left\{ \begin{array}{l} \frac{dE_-}{dz} = a_d E_- - \sigma_d E_+ - s_d E_s \\ \frac{dE_+}{dz} = \sigma_d E_- - a_d E_+ - s'_d E_s \\ \frac{dE_s}{dz} = k_e E_s \end{array} \right. \quad (2.1)$$

$$\frac{dE_v}{dz} = K_e E_v + u_d E_+ + \nu_d E_- + \omega_d E_s \quad (2.2)$$

où

- z représente la dimension verticale du couvert ;
- E_- (resp. E_+) est le flux diffus descendant (resp. montant), a_d (resp. σ_d) est le coefficient de diffusion (resp. rétrodiffusion) ;
- E_s est le flux directionnel incident, k_e le coefficient d'extinction du flux directionnel incident, s_d (resp. s'_d) est le coefficient de diffusion du flux directionnel incident contribuant à E_+ (resp. E_-) ;
- E_v est le flux directionnel dans la direction d'observation, K_e le coefficient d'extinction et ω_d, ν_d, u_d sont des coefficients de diffusion.

La résolution de ce système et le calcul des différents coefficients nécessitent deux hypothèses importantes :

- l'orientation azimutale des feuilles est aléatoire ;
- les feuilles sont lambertiennes, les propriétés optiques des 2 faces de la feuille étant identiques

Les coefficients s'expriment alors en fonction des caractéristiques du couvert (angle moyen d'inclinaison des feuilles, LAI , propriétés optiques des feuilles) et de la géométrie de mesure.

Verhoef (1985) montre que l'on peut passer de la solution générale au système suivant :

$$\begin{bmatrix} E^d(-1) \\ E^u(0) \end{bmatrix} = \begin{bmatrix} T_1 & R_1 \\ R_2 & T_2 \end{bmatrix} \begin{bmatrix} E^u(0) \\ E^d(-1) \end{bmatrix} \quad (2.3)$$

avec $E^d = \begin{bmatrix} E_s \\ E_- \end{bmatrix}$ et $E^u = \begin{bmatrix} E_+ \\ E_o \end{bmatrix}$ sachant que (-1) correspond au niveau du sol et (0) au sommet du couvert. Les matrices T_1 , T_2 (resp. R_1 , et R_2) sont les matrices de transmittance (resp. réflectance) du couvert et du sol. Le bouclage au niveau du sol se fait en imposant la condition suivante :

$$E^u(-1) = R_s E^d(-1), R_s \text{ étant la matrice de réflectance du sol}$$

Il s'en suit que :

$$E^u(0) = R_o^* E^d(0)$$

$$\text{avec } R_o^* = \begin{bmatrix} \rho_h(\theta_s) & \rho_h \\ \rho(\theta_v, \theta_s, \varphi) & \rho(\theta_v, \varphi) \end{bmatrix}$$

où, pour une longueur d'onde donnée, ρ_h est la réflectance bi-hémisphérique, $\rho(\theta_v, \varphi)$ la réflectance hémisphérique directionnelle, $\rho_h(\theta_s)$ est la réflectance directionnelle hémisphérique, et $\rho(\theta_v, \theta_s, \varphi)$ la réflectance bidirectionnelle. Cette équation permet d'obtenir la valeur de la réflectance bidirectionnelle, pour des angles d'incidence et d'observation donnés.

Variables d'entrée du modèle

Elles sont au nombre de 11 et peuvent être divisées en 3 catégories :

- Les variables de structure qui décrivent l'architecture du couvert : l'indice foliaire (LAI) et l'angle moyen d'inclinaison des feuilles ($\bar{\theta}_f$). Afin d'améliorer la précision du modèle, nous utilisons une version modifiée de SAIL prenant en compte le phénomène de hot spot (Andrieu *et al.*, 1997). Le hot spot (Kuusk, 1985) est observé dans un angle solide proche de la direction d'incidence (*i.e.* la direction solaire) : il n'y a pas d'ombrage mutuel des éléments de la végétation. Ceci se traduit par un pic de réflectance dans le plan solaire (ou plan principal). Ce pic dépend essentiellement de la structure du couvert (LAI et inclinaison des feuilles), et plus particulièrement de la taille et de la forme des feuilles (Qin et Goel, 1995). La variable s qui décrit le hot spot dans SAIL correspond au rapport entre la longueur des feuilles et la hauteur de la canopée.
- Les variables qui décrivent la géométrie de mesure, c'est-à-dire la position zénithale du soleil θ_s , la position zénithale de visée θ_v et la position azimutale relative entre le soleil et la visée $\phi = \phi_v - \phi_s$ (figure 2.1).
- Les variables qui décrivent les propriétés optiques des éléments du couvert : pour les feuilles, seules les réflectance ρ_f et transmittance τ_f hémisphériques sont requises (puisqu'elles sont supposées lambertiennes). Elles sont calculées grâce au modèle PROSPECT (Jacquemoud et Baret, 1990) que nous présentons un peu plus loin. Le sol est par contre décrit plus précisément par sa réflectance bidirectionnelle ($\rho_s(\theta_v, \theta_s, \phi)$), sa réflectance directionnelle hémisphérique ($\rho_s(\theta_s)$) et sa réflectance bi-hémisphérique (ρ_s) définies ci-après :

$$\begin{cases} \rho_s(\theta_s) &= \frac{1}{\pi} \int_0^{2\pi} d\phi \int_0^{\pi/2} \rho_s(\theta_v, \theta_s, \phi) \sin \theta_v \cos \theta_v d\theta_v \\ \rho_s &= 2 \int_0^{\pi/2} \rho_s(\theta_s) \cos \theta_s \sin \theta_s d\theta_s \end{cases}$$

Ces trois variables sont calculées grâce au modèle SOILSPECT développé par Jacquemoud *et al.*(1992)

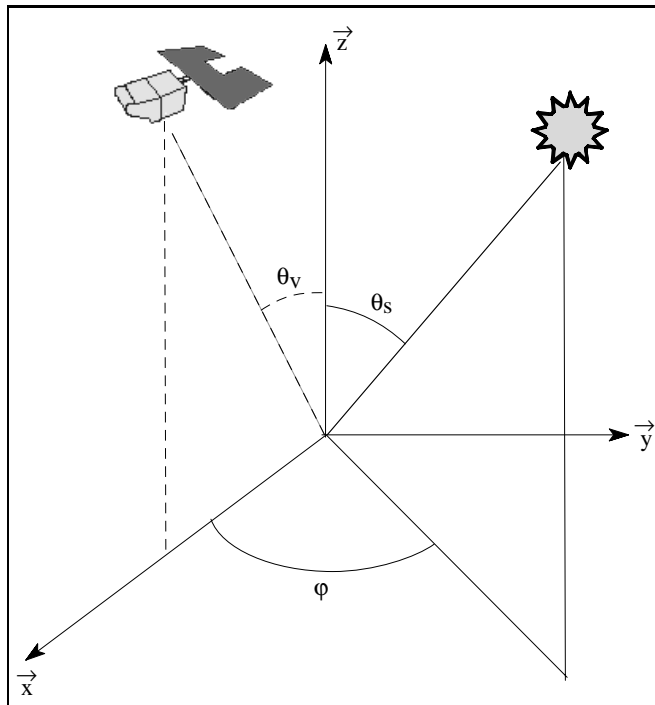


Figure 2.1 : Représentation en coordonnées sphériques de la géométrie de mesure. \vec{x} est la direction azimutale du soleil, \vec{y} est perpendiculaire à \vec{x} , et \vec{z} est la verticale au point observé.

2.1.2 Le modèle de propriétés optiques des feuilles : PROSPECT

PROSPECT (Jacquemoud et Baret, 1990), modifié par Baret et Fourty (1998), donne la variation spectrale des réflectance et transmittance hémisphériques des feuilles dans le domaine solaire (400 nm-2400 nm). La feuille est considérée comme un empilement de N couches diffusantes et absorbantes. La diffusion est caractérisée par un indice de réfraction (n) qui dépend de la longueur d'onde. Les composés absorbants (chlorophylle, eau et matière sèche) sont supposés répartis de façon homogène dans la feuille, ce qui permet d'évaluer un coefficient d'absorption global K_a égal à la somme des contributions de chaque composant : $K_a = \sum k_i C_i$ où k_i est le coefficient d'absorption spécifique du composant et C_i la teneur de ce même composant dans la feuille. Les variables d'entrée du modèle de propriétés optiques des feuilles sont donc :

- l'indice de structure N ;
- l'indice de réfraction n : il est supposé constant quelque soit le type de feuille et a été mesuré par Jacquemoud (1989) sur des feuilles de maïs albinos (figure 2.2) ;
- la teneur en chlorophylle a et b C_{ab} ($\mu g.cm^{-2}$), en eau C_e (cm^{-1}) et en matière sèche C_{ms} ($g.cm^{-2}$).

2.1.3 Le modèle de propriétés optiques du sol : SOILSPECT

SOILSPECT (Jacquemoud *et al.*, 1992) est un modèle de transfert radiatif dérivé du modèle de Hapke (1981) qui donne la variation spectrale de la réflectance bidirectionnelle du sol ($\rho_s(\theta_v, \theta_s, \phi)$) et, par conséquent la variation spectrale des réflectances directionnelle hémisphérique ($\rho_s(\theta_s)$) et bi-hémisphérique de ce même sol (ρ_s). L'expression de la réflectance bidirectionnelle pour une longueur d'onde λ donnée est :

$$\rho_s(\theta_v, \theta_s, \phi) = \frac{\omega_s(\lambda)}{4\pi} \frac{\mu_s}{\mu_s + \mu_v} ([1 + B(\xi)]P(\xi) + H(\mu_s)H(\mu_v) - 1) \quad (2.4)$$

où μ_s et μ_v sont les cosinus respectifs des angles zénithaux de visée et d'observation et $\xi = \arccos(\sin \theta_v \cos \phi \sin \theta_s + \cos \theta_v \cos \theta_s)$, l'angle de phase.

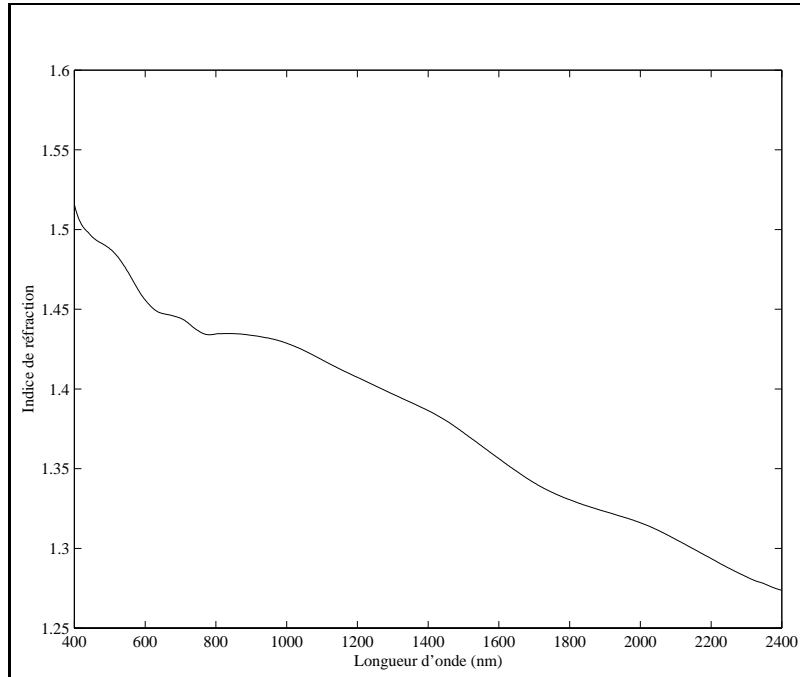


Figure 2.2 : Variation spectrale de l'indice de réfraction n des feuilles.

$\omega_s(\lambda)$ est l'albédo de simple diffusion ; il décrit le rapport entre l'énergie diffusée et l'énergie totale et dépend de la longueur d'onde. $B(\xi)$ est la fonction de rétrodiffusion qui décrit le phénomène de hot spot. La fonction de phase $P(\xi)$ représente la probabilité d'avoir une diffusion dans une direction donnée connaissant la direction d'incidence. Dans la modélisation de SOILSPECT, on suppose qu'elle est indépendante de la longueur d'onde. H est la fonction de Chandrasaekar.

$$B(\xi) = \frac{1}{1 + \frac{1}{h} \tan \frac{\xi}{2}}, P(\xi) = 1 + r_1 \cos \xi + r_2 \frac{3 \cos^2 \xi - 1}{2}, H(x) = \frac{1 + 2x}{1 + 2x \sqrt{1 - \omega_s(\lambda)}}$$

2.1.4 Le modèle d'orbitographie de *VEGETATION*

L'orbitographie d'un satellite, régie par les lois de Képler, permet, pour une latitude et une date données, de définir la configuration de visée du capteur. La latitude et la date déterminent aussi la position du soleil. *VEGETATION* est un satellite héliosynchrone (le plan de son orbite conserve un angle con-

stant avec la direction Terre-Soleil) et son orbite est quasi circulaire. Ceci assure une observation de pratiquement toute la surface du globe à une altitude constante et un éclairage solaire de la surface observée ne variant que selon les saisons et les angles de prises de vue. Le tableau 2.1 résume les caractéristiques orbitales de *VEGETATION* permettant de calculer à tout moment et en tout lieu la configuration de visée du capteur, sachant que le satellite possède un angle de balayage de $\pm 50.5^\circ$ ($\pm 60.7^\circ$ au sol) perpendiculairement à la trace, ce qui donne une trace au sol de 2250 km environ pour une résolution spatiale de 1.15 km. Le modèle d'orbitographie que nous avons utilisé est celui développé par Gilbert Saint, disponible sur le site web consacré à *VEGETATION* (<http://www-vegetation.cst.cnes.fr>).

| | |
|---|------------|
| Altitude moyenne | 832 km |
| Durée du cycle | 26 jours |
| Angle d'inclinaison | 98° |
| Heure solaire locale (noeud descendant) | 10:30 A.M. |

Tableau 2.1 : Caractéristiques orbitales du capteur *VEGETATION*.

La figure 2.3 montre le nombre de passages du satellite pour une latitude donnée durant tout le cycle d'observation. Pour des latitudes supérieures à 45° , un même point est observé au moins une fois et jusqu'à 5 fois par jour pour les latitudes extrêmes. Dans le cas de latitudes inférieures, la durée entre deux observations successives est inférieure à deux jours. L'orbitographie de *VEGETATION* est donc bien adaptée pour effectuer un suivi régulier de la végétation.

La figure 2.4 donne un exemple de configuration de visée en représentation polaire acquise sur un cycle de *VEGETATION* à une latitude de 26° , le 3 mai. L'angle azimutal relatif entre le satellite et le soleil (le plan principal, *i.e.* le plan solaire, étant situé sur le plan $0^\circ/180^\circ$) est repéré par l'angle φ . La valeur du rayon donne l'angle zénithal θ_v .

VEGETATION fournit des images dans quatre bandes spectrales : le bleu (430 nm-470 nm), le rouge (610 nm-680 nm), le proche infrarouge (780 nm-890 nm) et le moyen infrarouge (1580 nm-1750 nm). Dans cette étude, nous ne tenons pas compte de la largeur de chaque bande spectrale, pour des problèmes de temps de calcul, nous ne considérons le signal que dans la longueur d'onde

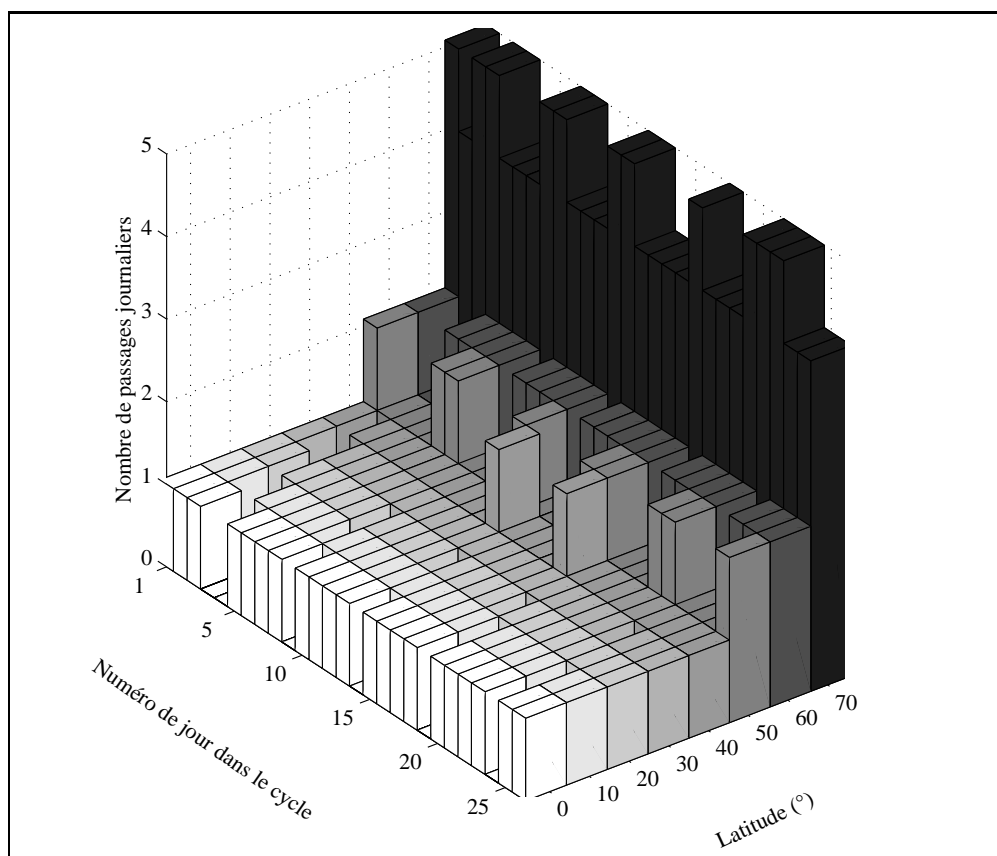


Figure 2.3 : Nombre de données acquises chaque jour durant un cycle de VEGETATION en fonction de la latitude, pour une longitude quelconque.

correspondant au centre de chaque bande (450 nm, 645 nm, 835 nm et 1665 nm).

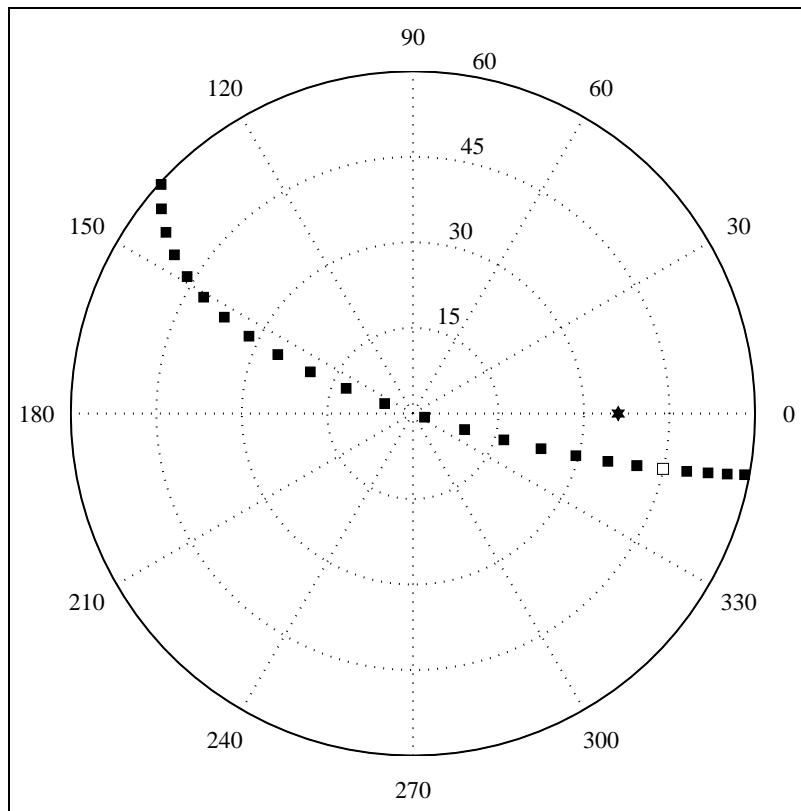


Figure 2.4 : Représentation de l'orbitographie du capteur VEGETATION, à 26° de latitude, le 3 mai. La position du soleil est repérée par une ★, celle du capteur par des ■. Le point repéré par une □ a pour coordonnées $(\theta_v = 45^\circ, \phi = 347.5^\circ)$.

Dans ce paragraphe, nous avons décrit l'ensemble des modèles que nous avons utilisés pour déterminer la réflectance des couverts végétaux de notre base d'apprentissage. Le suivi de la végétation à partir d'un signal satellitaire se fait au moyen de variables biophysiques de ces mêmes couverts. L'objectif du paragraphe suivant est de présenter ces variables et dans quelle mesure on peut penser les estimer à partir de données de réflectance. Cependant, il faut bien noter les limites de cette étude : nous ne prenons pas en compte

les imprécisions de chaque mesure qui peuvent être induites par la résolution radiométrique, par les corrections atmosphériques ou encore par la correction géométrique des images.

2.2 Rappels théoriques sur le calcul du $fAPAR$ et de la fraction de trou dans le formalisme de SAIL

L'objectif de cette étude est d'estimer des variables biophysiques du couvert végétal à partir de données de télédétection. Ces variables sont celles décrites au chapitre 1.2 :

- l'indice foliaire et la teneur en chlorophylle intégrée au niveau du couvert qui sont directement des variables d'entrée de SAIL,
- le $fAPAR$ et les fractions de trou mono et bi-directionnelles qui nécessitent des calculs supplémentaires pour les simulations.

L'objectif de ce paragraphe est de rappeler de quelle manière ces trois variables sont calculées dans le formalisme de SAIL.

2.2.1 Calcul du $fAPAR$

L'équation 2.5 exprime le rayonnement absorbé par le couvert pour une longueur d'onde donnée. Elle est établie à partir de la loi de conservation de l'énergie, c'est-à-dire que le rayonnement incident est égal à la somme des rayonnements transmis, réfléchi et absorbé par le sol et la végétation. Dans la cas du formalisme du modèle SAIL, on a :

$$a_\lambda \simeq F_d[1 - \rho_{dd} - \tau_{dd}(1 - r_{dd})] + F_s[1 - \rho_{sd} - \tau_{ss}(1 - r_{sd}) - \tau_{sd}(1 - r_{dd})] \quad (2.5)$$

où

- a_λ correspond à la fraction de rayonnement absorbé, dans une longueur d'onde donnée λ ;
- F_d (resp. F_s) est la fraction de rayonnement diffus d (resp. direct s) :
 $F_d + F_s = 1$;

- la lettre ρ (resp. r) se réfère à la réflectance du couvert (resp. du sol) ;
- la lettre τ se réfère à la transmittance du couvert ;
- les variables $\rho_{dd}, \tau_{dd}, r_{dd}, \rho_{sd}, \tau_{ss}, r_{sd}, \tau_{sd}$ sont les termes des matrices T_1, T_2, R_1 et R_2 de l'équation 2.3 et sont représentés sur la figure 2.5.

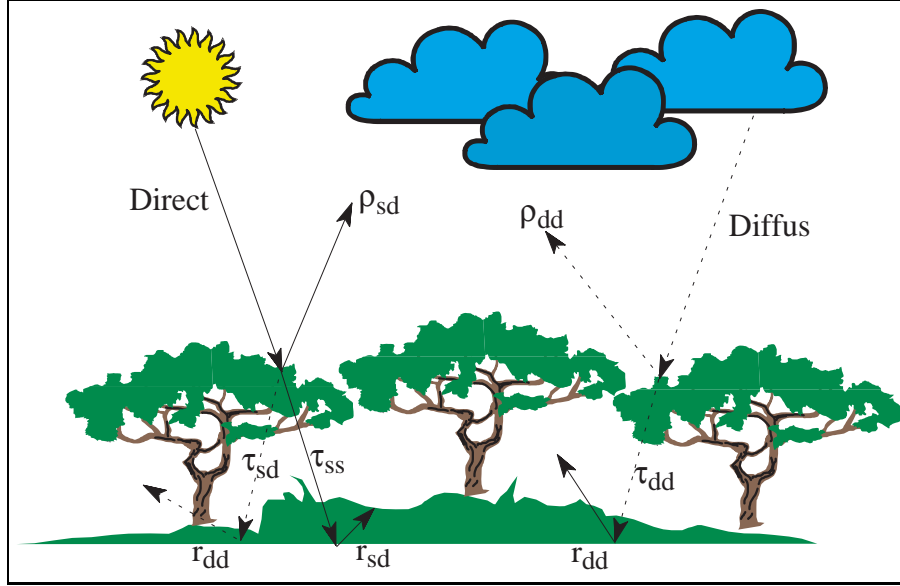


Figure 2.5 : Représentation schématique des coefficients du modèle SAIL caractérisant le mécanisme de l'absorption d'un couvert végétal. Les pointillés représentent des flux hémisphériques (atmosphère ou couvert), les traits pleins des flux directionnels.

Le $fAPAR(\theta_s)$ instantané est alors donné comme l'intégrale de a_λ sur 400 nm-700 nm (intégration numérique par intervalles de 10 nm chacun). Le $fAPAR$ est donc égal à l'intégrale de $fAPAR(\theta_s)$, pondéré par l'éclairement total incident E_{ts} du lever au coucher du soleil (eq 2.6).

$$\left\{ \begin{array}{l} fAPAR(\theta_s) = \int_{400nm}^{700nm} a_\lambda d\lambda \\ fAPAR = \frac{1}{\int_{t_1}^{t_2} E_{ts} \cos(\theta_s) d\theta_s} \int_{t_1}^{t_2} fAPAR(\theta_s) E_{ts} \cos(\theta_s) d\theta_s \end{array} \right. \quad (2.6)$$

où t_1 (resp. t_2) correspond au lever du soleil (resp. coucher du soleil).

2.2.2 Les fractions de trou $P_o(\theta_v), P_o(\theta_v, \theta_s)$

La fraction de trou est une variable structurelle, qui dépend du LAI et de l'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$. La fraction de trou monodirectionnelle $P_o(\theta_v)$ est la fraction de sol vue dans la direction de visée (Eq 2.9). Autrement dit, elle correspond à la transmittance du couvert, les feuilles étant supposées opaques.

Considérons une couche de feuilles du couvert de densité $LAD(z)^2$, la quantité d'énergie en haut de la couche d'épaisseur dz à la hauteur z , $E_{tv}(z + dz)$ ($E_{tv} = E_+$ (diffus) + E_v (direct)) est égale à la quantité d'énergie au bas de la couche $E_{tv}(z)$ moins la quantité d'énergie interceptée par la couche de feuilles. La section efficace S_{eff} de feuilles qui intercepte l'énergie montante du bas du couvert vers le haut est donnée par (figure 2.6) :

$$S_{eff} = LAD(z)dz.G(\theta_v) \quad (2.7)$$

$$G(\theta_v) = \int_0^{2\pi} \int_0^{\pi/2} \frac{f(\theta_f)}{2\pi} \sin \theta_f d\theta_f d\phi_f \vec{f} \cdot \vec{v} \quad (2.8)$$

où $G(\theta_v)$ représente la surface de feuille moyenne projetée dans la direction de visée θ_v . θ_f est l'angle d'inclinaison des feuilles et $f(\theta_f)$ est la fonction de distribution d'inclinaisons des feuilles. Par suite,

$$E_{tv}(z + dz) = E_{tv}(z) - \frac{E_{tv}(z)}{\cos \theta_v} . LAD(z)dz.G(\theta_v)$$

soit

$$\frac{E_{tv}(z + dz) - E_{tv}(z)}{E_{tv}(z)} = - \frac{LAD(z)dz.G(\theta_v)}{\cos \theta_v}$$

Sachant que $E_{tv}(H_c) = E_{tv}(0).P_o(\theta_v)$ où H_c est la hauteur du couvert et que $\int_0^{H_c} LAD(z)dz = LAI$, on en déduit, par intégration que

$$P_o(\theta_v) = exp[-LAI \cdot \frac{G(\theta_v)}{\cos(\theta_v)}] \quad (2.9)$$

²Leaf Area Density

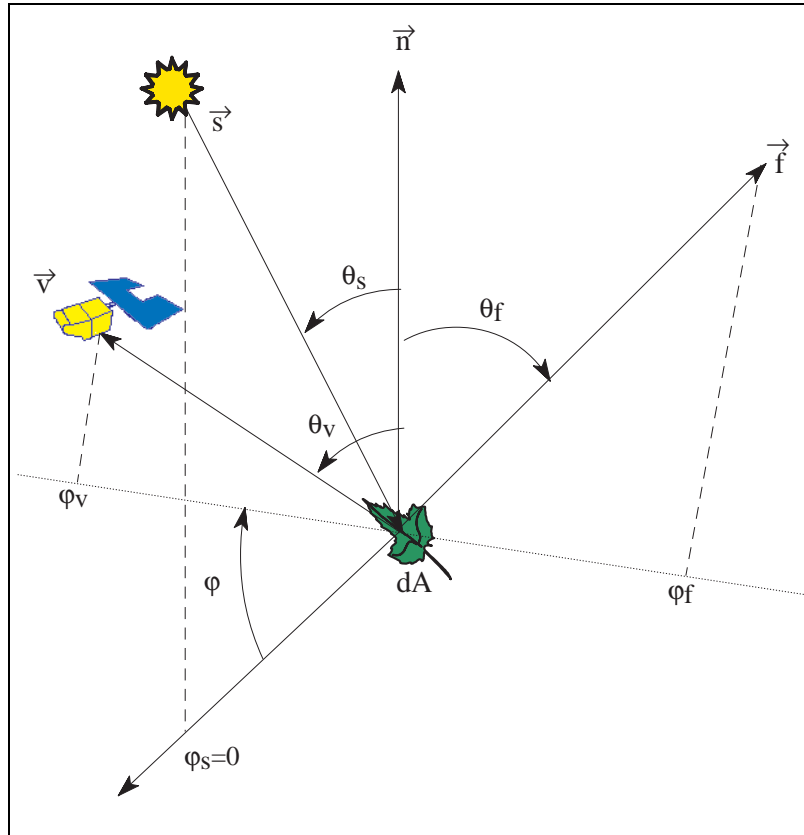


Figure 2.6 : Positions relatives des vecteurs de visée \vec{v} , d'incidence \vec{s} et de la normale à la feuille \vec{f} . dA est un élément de surface de feuille de la strate.

La fraction de trou bidirectionnelle $P_o(\theta_s, \theta_v)$ (Eq 2.10) correspond à la fraction de sol éclairé et vu dans la direction de visée. Nous avons utilisé l'expression de la fraction de trou bidirectionnelle donnée par Kuusk (1991), qui tient compte du phénomène de hot spot (Eq 2.10) grâce au facteur correctif κ . κ dépend du diamètre des feuilles d_f rapporté à la hauteur H_c du couvert.

$$P_o(\theta_s, \theta_v) = P_o(\theta_s) \cdot P_o(\theta_v) \cdot e^{\kappa} \quad (2.10)$$

avec

$$\left\{ \begin{array}{l} \kappa = LAI \sqrt{\frac{G(\theta_v)G(\theta_s)}{\mu_v \mu_s}} \frac{1-e^{-\alpha}}{\alpha} \\ \alpha = \frac{1}{s_l} (\tan \theta_s^2 + \tan \theta_v^2 - 2 \tan \theta_s \tan \theta_v \cos \varphi) \\ s_f = \frac{1}{H_c} \frac{\pi d_f G(\theta_v)}{4\mu_v \int_0^{2\pi} \int_0^{\pi/2} \frac{f(\theta_f) \sin \theta_f d\theta_f d\phi_f}{\sqrt{1+\tan^2 \theta_f \sin^2 \phi_f}}} \end{array} \right.$$

Dans la mesure où nous avons considéré une distribution des inclinaisons foliaires ellipsoïdale, l'expression de s_f est donnée par (Andrieu *et al.*, 1997) :

$$s_f = \frac{d_f}{H_c} \frac{\pi G(\theta_v)}{4\mu_v [1 + 0.357 \frac{\theta_f}{1.693 - \theta_f}]}$$

La fraction de trou monodirectionnelle est un élément clé de la simple diffusion au sein du couvert puisqu'elle correspond à la probabilité qu'un rayon lumineux atteigne directement le sol du haut du couvert. C'est donc une variable qui gouverne le transfert radiatif, très fortement corrélée au signal de réflectance. Bien que de nombreux auteurs s'accordent à dire que c'est effectivement une variable essentielle (Nilson, 1971, Li et Strahler, 1988, Albers *et al.*, 1990, Andrieu et Sinoquet, 1993, Baret *et al.*, 1993,...), il n'existe que peu de travaux permettant d'exploiter la fraction de trou. En effet, c'est une variable difficile à mesurer. L'estimation de la fraction de trou peut se faire à partir de photos hémisphériques (Bonhomme et Chartier, 1972; Baret *et al.*, 1993) ou bien grâce à des mesures radiométriques au-dessus de la canopée et au-dessous (Albers *et al.*, 1990). Baret *et al.* (1995) ont développé une méthode pour estimer la fraction de trou à partir de données de réflectance, basée sur l'utilisation de réseaux de neurones.

2.3 Domaine de variation des variables d'entrée du modèle SAIL

Pour obtenir une base d'apprentissage la plus représentative possible de ce qui peut exister dans la réalité, avec les limites qu'impose le modèle SAIL, il

faut déterminer le domaine de variation de chacune des variables d'entrée du modèle ainsi que la loi statistique d'échantillonnage qui lui est associée. Ceci a été réalisé en se basant sur des jeux de données expérimentales que nous avons compilés.

2.3.1 Les variables de structure

Le *LAI*

La gamme de variation du *LAI* sur les couverts végétaux est assez étendue, comprise entre 0 et 10. Grâce aux données acquises par des satellites tels que *NOAA/AVHRR*, des jeux de données de *LAI* et de *fAPAR* ont été établis à partir du *NDVI*. Nous avons utilisé la répartition de *LAI* issue du projet ISLSCP (International Satellite Land Surface Climatology Project) qui donne, en fonction du mois de l'année et des coordonnées géographiques d'un point, une valeur de *LAI* sur une grille de 1° par 1° (Sellers *et al.*, 1994 et 1996). Initialement, ces données étaient destinées à être utilisées comme entrées du modèle SiB2³ qui estime les transferts d'énergie et d'eau entre l'atmosphère et la biosphère. Elles sont donc suffisamment précises pour avoir une bonne idée de la répartition du *LAI* sur l'année (Los *et al.*, 1994). Cependant, il faut noter que les valeurs de *LAI* fournies ont été obtenues grâce à des données *NOAA/AVHRR* de résolution spatiale 1 km. Elles sont donc moyennées pour obtenir une valeur sur une maille de 1° par 1° . La figure 2.7 montre un exemple, pour le mois de mai 1987 de la répartition du *NDVI* et du *LAI*. On note bien la correspondance entre les fortes valeurs de *NDVI* (fort contraste entre le rouge et le proche infrarouge) et les fortes valeurs de *LAI* (saturation du signal dans le proche infrarouge). Les régions équatoriales montrent un fort taux de végétation alors que les zones tempérées présentent des *LAI* plus faibles.

La figure 2.8 présente la répartition du *LAI* par mois sur l'année 1987. Les résultats montrent peu de variations d'un mois sur l'autre avec une forte proportion de couverts à *LAI* faibles. Cependant, on note que la proportion de zones à fort *LAI* augmente pour les mois de mai à octobre. Cette évolution correspond à celle des cultures dans les zones tempérées de l'hémisphère nord. L'influence des zones cultivées de l'hémisphère sud est négligeable car elles correspondent à une surface beaucoup moins importante.

³Simple Biosphere model

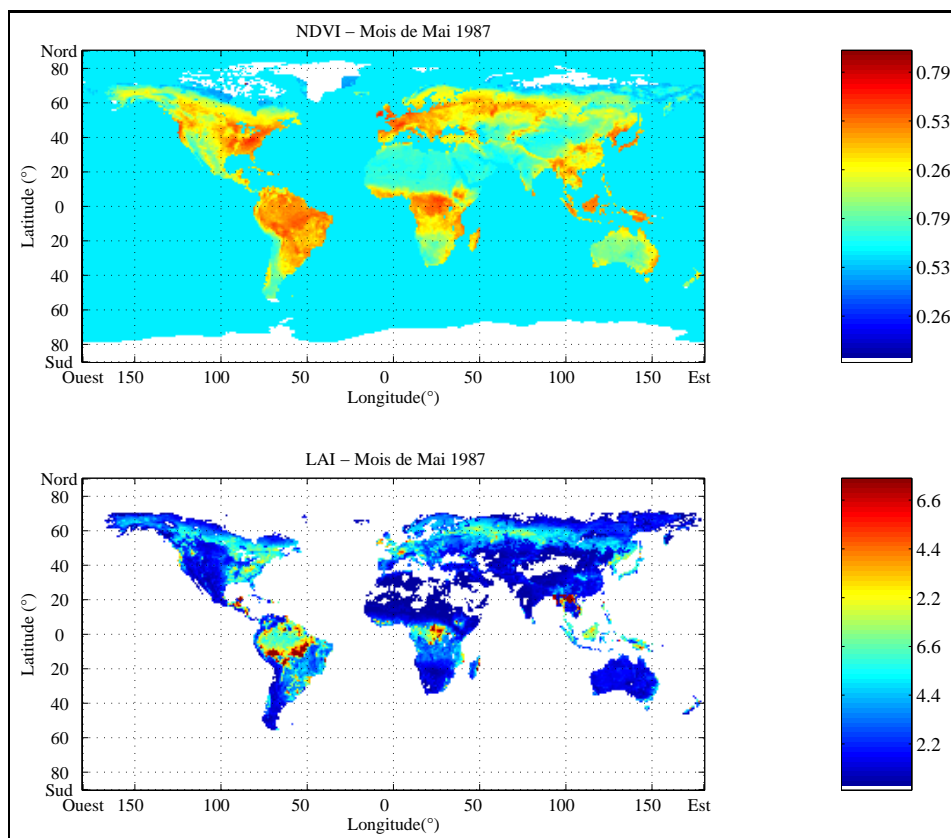


Figure 2.7 : Cartographie du *NDVI* et du *LAI* établie à partir des données ISLSCP pour le mois de mai 1987.

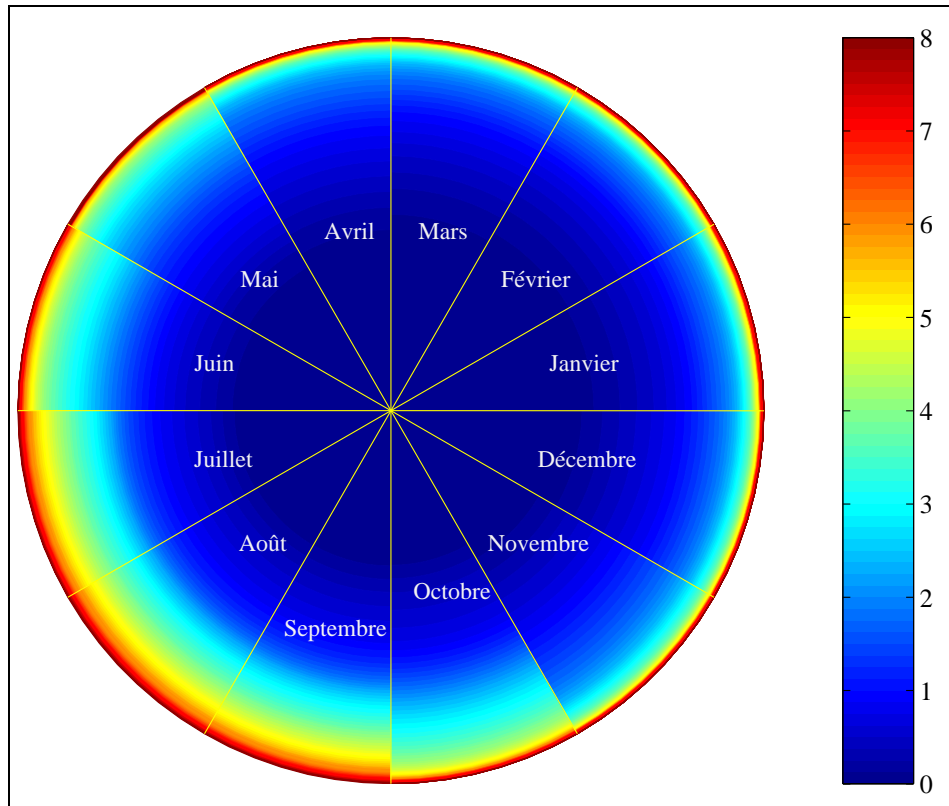


Figure 2.8 : Répartition du LAI sur l'année 1987. La surface d'une portion d'une couleur donnée est proportionnelle au nombre de points de la grille ISLSCP caractérisés par le LAI correspondant à cette couleur.

La valeur de LAI observée dépend donc de la latitude et de la date d'observation. C'est la raison pour laquelle nous avons tiré aléatoirement selon une loi uniforme une valeur de LAI dans la distribution ISLSCP pour une latitude et une date données. On remarquera (figure 2.9) que les faibles LAI sont majoritairement représentés.

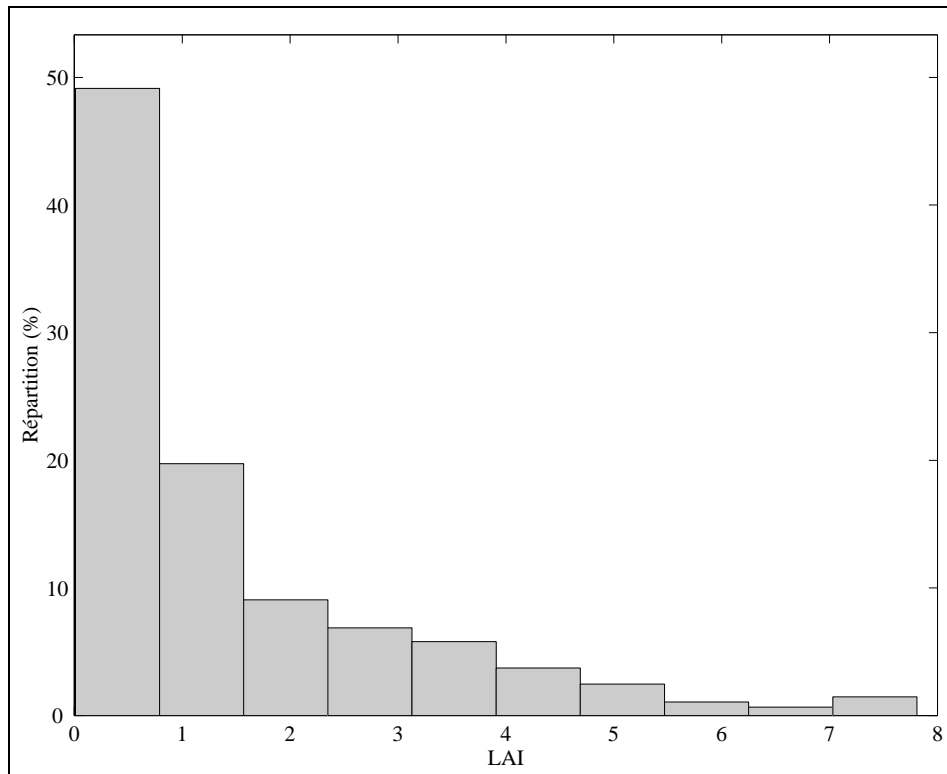


Figure 2.9 : Distribution du LAI sur la base d'apprentissage.

L'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$

Pour simuler l'ensemble des possibilités d'inclinaison des feuilles, nous avons choisi une fonction de distribution des angles foliaires ellipsoïdale (Campbell, 1986). $\bar{\theta}_f$ caractérise alors l'excentricité de l'ellipse dans laquelle sont distribués les angles zénithaux d'inclinaison des feuilles. Il peut varier entre 15° pour les couverts planophylles et 75° pour les couverts érectophylles.

Nous avons donc tiré aléatoirement selon une loi uniforme une valeur d'angle comprise entre ces deux extrêmes.

Le paramètre de hot spot

Étant donné qu'il est égal au rapport entre le diamètre d'une feuille et la hauteur du couvert, il est plausible de tirer toujours uniformément des valeurs entre 2 cas extrêmes qui sont d'une part un couvert composé de feuilles de petite taille ($s=0.01$) et d'autre part, un couvert composé de feuilles de grande taille ($s=1$) par rapport à la hauteur du couvert.

2.3.2 Les propriétés optiques des feuilles

Les deux variables requises par SAIL sont les réflectance (ρ_f) et transmittance (τ_f) bihémisphériques de la feuille. Pour les obtenir, nous avons besoin des 4 variables d'entrée du modèle PROSPECT. Pour chacun d'eux, nous avons utilisé une loi de tirage gaussienne dont la moyenne et l'écart-type sont fournis par le jeu de données LOPEX (Hosgood *et al.*,1995) composé d'un grand nombre d'espèces (tableau 2.2).

| | Min | Max | Moyenne | Ecart type |
|---------------------------|-------|-------|---------|------------|
| C_{ab} ($\mu g/cm^2$) | 10 | 80 | 50 | 16 |
| C_e (cm^{-1}) | 0.005 | 0.025 | 0.01 | 0.0024 |
| C_{ms} (g/cm^2) | 0.002 | 0.011 | 0.005 | 0.001 |
| N | 1.1 | 2.5 | 1.6 | 0.27 |

Tableau 2.2 : Variation des variables d'entrée de PROSPECT : C_{ab} est la teneur en chlorophylle a et b, C_e la teneur en eau, C_{ms} la teneur en matière sèche et N , le paramètre de structure.

2.3.3 Les propriétés optiques des sols

Les 3 variables requises par SAIL et simulées par SOILSPECT sont la réflectance bidirectionnelle du sol ($\rho_s(\theta_v, \theta_s, \varphi)$), la réflectance directionnelle hémisphérique correspondante ($\rho_s(\theta_s)$), intégration de la réflectance par rapport aux angles de visée, et la réflectance bihémisphérique du sol (ρ_s), intégration de la réflectance hémisphérique par rapport aux angles d'incidence.

SOILSPECT nécessite lui-même 4 variables d'entrée. Leur gamme de variation est définie à partir d'un jeu de données expérimentales (Jacquemoud *et al.*, 1991). Elles ont été réalisées sur 26 sols très différents, aussi bien en terme de type de sol, de rugosité (sols calcaires, sables, terreaux et galets) qu'en terme d'humidité.

Variables indépendantes de la longueur d'onde

Elles sont au nombre de trois : un paramètre de rugosité et deux paramètres de fonction de phase. Pour construire notre jeu de données, nous avons tiré des valeurs aléatoires selon une loi uniforme entre les extrêmes trouvés sur les 26 sols (tableau 2.3).

| | Min | Max |
|-------------------------|------|------|
| Rugosité h | 0 | 0.03 |
| Fonction de Phase r_1 | -1.8 | 1.8 |
| Fonction de Phase r_2 | 0.05 | 1 |

Tableau 2.3 : Variation des variables d'entrée indépendantes de la longueur d'onde de SOILSPECT.

L'albédo de simple diffusion $\omega_s(\lambda)$

Définir sa gamme de variation a posé plus de problèmes puisqu'il dépend à la fois du type de sol, de l'humidité et de la longueur d'onde. Pour que les spectres d'albédo de simple diffusion des sols de la base d'apprentissage ne se cantonnent pas à ceux des 26 sols de Jacquemoud *et al.*(1991), nous avons effectué une analyse en composantes principales (ACP). Ceci nous permet d'exprimer, par une simple combinaison linéaire, n'importe quel spectre d'albédo de simple diffusion ($\omega_s(\lambda)$) en fonction de trois spectres de base ($\omega_{s1}(\lambda), \omega_{s2}(\lambda), \omega_{s3}(\lambda)$) qui correspondent aux trois axes principaux de l'ACP :

$$\omega_s(\lambda) = x_s \omega_{s1}(\lambda) + y_s \omega_{s2}(\lambda) + z_s \omega_{s3}(\lambda) \quad (2.11)$$

La figure 2.10 présente le spectre de l'écart quadratique moyen, noté $RMSE^4$ entre la réflectance mesurée et celle reconstituée avec les 3 spectres de base de

⁴Root Mean Square Error

l'ACP (eq 2.11) en fonction de la longueur d'onde (eq 2.12). Le $RMSE$ est calculé sur l'ensemble des 26 sols considérés :

$$RMSE = \sqrt{\frac{1}{n} \sum_n (y - \hat{y})^2} \quad (2.12)$$

où y et \hat{y} sont les variables mesurées et estimées, n le nombre de données (26) et \bar{y} la moyenne de y sur les n échantillons.

Le $RMSE$ est globalement très faible, compris entre 0.005 et 0.035. On observe toutefois qu'il augmente dans les bandes d'absorption de l'eau centrées sur 1400 nm et 1900 nm (Irons *et al.*, 1989). Ceci est confirmé par la figure 2.11 où sont représentées les valeurs de $RMSE$ en fonction du type de sol considéré (argileux (1 à 9), sable (10 à 15), terreau (16 à 24), galets (25) et pouzzolane (26)). Les valeurs les plus fortes se retrouvent pour les sols humides. Ceci provient en partie du fait que lors de la mesure, le bruit instrumental est plus important pour les faibles valeurs de réflectance qui caractérisent des sols humides.

Compte tenu de ces résultats, nous pouvons considérer que l'utilisation des 3 axes principaux de l'analyse en composantes principales est suffisante pour obtenir une bonne précision sur le spectre d'albédo de simple diffusion d'un sol donné. Pour créer une variabilité de ω_s dans notre base d'apprentissage, nous avons donc effectué un tirage aléatoire selon une loi uniforme pour les 3 coordonnées (x_s, y_s, z_s) entre les valeurs minimales et maximales obtenues sur les 26 sols (tableau 2.4).

| | Min | Max |
|-------|--------|-------|
| x_s | 0.074 | 0.294 |
| y_s | -0.031 | 0.228 |
| z_s | -0.501 | 0.151 |

Tableau 2.4 : Variation des coordonnées des 26 sols sur les axes principaux.

2.3.4 La géométrie de mesure

Les angles zénithaux d'incidence et d'observation (θ_s, θ_v) ainsi que l'angle azimutal relatif (φ) sont définis par l'orbitographie du satellite en fonction de

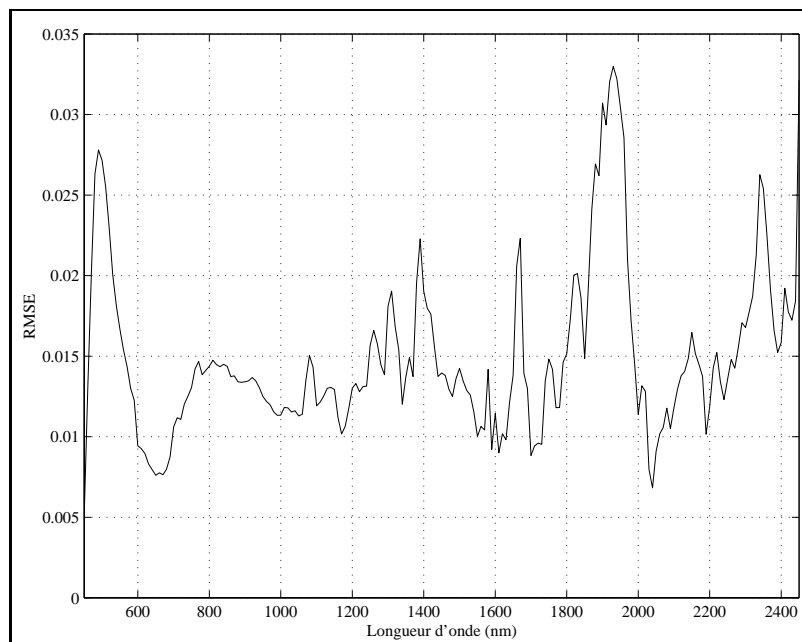


Figure 2.10 : Evolution du spectre du $RMSE$ entre la valeur de réflectance mesurée et celle estimée par ACP, pour les 26 sols du jeu de données.

la date et la latitude du point observé. Pour notre application, nous avons tiré uniformément la date d'acquisition des données ainsi que la latitude comprise entre l'équateur (0°) et le cercle polaire ($\approx 66^\circ$). Pour établir le cycle de 26 jours de *VEGETATION*, nous avons considéré que le capteur acquiert des images sur un intervalle de temps centré sur cette date. De cette manière, l'échantillonnage directionnel correspond au maximum des possibilités du capteur.

Nous avons également pris en compte la possibilité de présence de nuages lors de l'acquisition des données, qui rend les images inexploitable. Nous avons considéré des pourcentages de nébulosité en tirant aléatoirement 100%, 80%, 50% et 20% de données parmi toutes celles qu'aurait acquises le capteur par un ciel entièrement dégagé durant le cycle orbital.

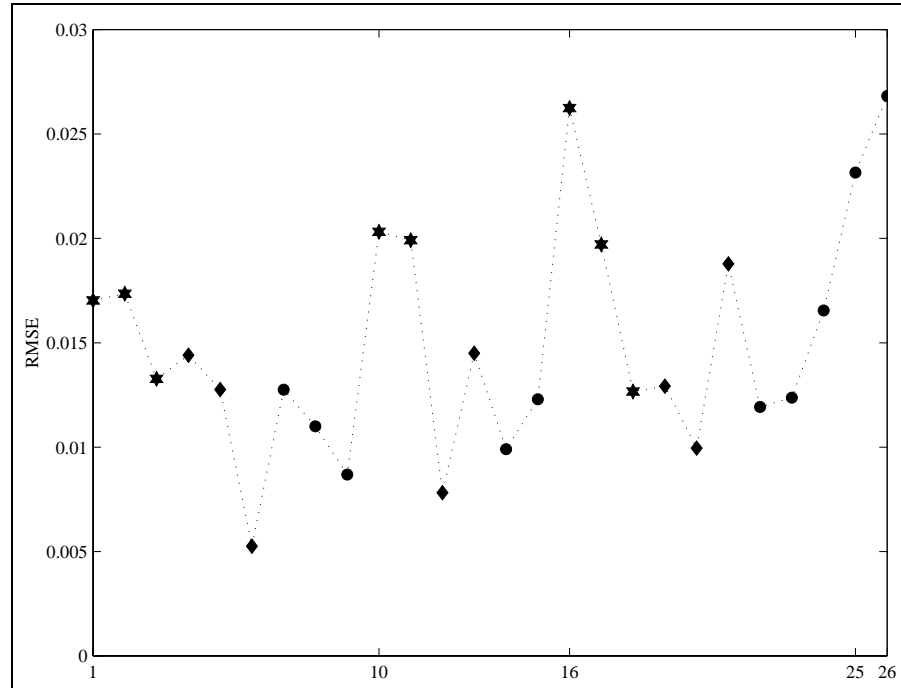


Figure 2.11 : Valeur du $RMSE$ entre le spectre mesuré et celui estimé par l'ACP pour chacun des 26 sols de base. Les ☆ correspondent aux sols très humides, les ◆ à une humidité moyenne et les ● aux sols secs.

2.4 Exemples de simulation

Le tableau 2.5 présente un récapitulatif de la gamme de variation de chacune des variables d'entrée du modèle SAIL. Les lois statistiques utilisées pour générer notre base d'apprentissage ont été choisies de manière à obtenir une meilleure précision pour les cas les plus probables (répartition ISLSCP du LAI , loi gaussienne pour les propriétés optiques des feuilles). Pour les variables d'entrée dont nous ne connaissons pas *a priori* la distribution, nous avons choisi une loi uniforme.

Compte tenu du fait que nous avons utilisé des jeux expérimentaux assez représentatifs de la réalité pour déterminer cette gamme de variation, nous pouvons considérer que notre base d'apprentissage est assez réaliste.

En définitive, cette base d'apprentissage est composée de 1500 couverts.

| Variable | Distribution | Min | Moyenne | Max | Écart-type |
|----------------------------|--------------|-------------|---------|-------|------------|
| LAI | ISLSCP | 0.01 | 1.45 | 7.8 | 1.6 |
| $\bar{\theta}_f(^{\circ})$ | Uniforme | 15 | 45 | 75 | 17 |
| Hot Spot s | Uniforme | 0.01 | 0.5 | 1 | 0.3 |
| $C_{ab}(\mu g c m^{-2})$ | Gaussienne | 10 | 48 | 80 | 16 |
| $C_e(c m^{-1})$ | Gaussienne | 0.005 | 0.01 | 0.017 | 0.002 |
| $C_{ms}(g c m^{-2})$ | Gaussienne | 0.002 | 0.005 | 0.009 | 0.001 |
| N | Gaussienne | 1.1 | 1.63 | 2.39 | 0.26 |
| x_s | Uniforme | 0.07 | 0.16 | 0.26 | 0.05 |
| y_s | Uniforme | -0.33 | -0.02 | 0.22 | 0.15 |
| z_s | Uniforme | -0.2 | -0.01 | 0.15 | 0.1 |
| h | Uniforme | 3.10^{-4} | 0.15 | 0.3 | 0.09 |
| r_1 | Uniforme | -1.8 | 0 | 1.8 | 1 |
| r_2 | Uniforme | -0.05 | 0.5 | 1 | 0.3 |
| Date | Uniforme | 1 | 183 | 365 | 102 |
| Latitude ($^{\circ}$) | Uniforme | -66 | -3 | 66 | 35 |

Tableau 2.5 : Tableau récapitulatif des domaines de variation des paramètres utilisés pour établir la base d'apprentissage.

Ces simulations ont été générées en tirant indépendamment (sauf pour la distribution du LAI qui dépend de la latitude) une valeur pour chacune des variables d'entrée de SAIL selon les lois décrites dans le tableau 2.5. Pour chacun des couverts simulés, nous avons :

- un jeu de réflectance $\rho(\theta_s, \theta_v, \varphi)$ pour la séquence d'angles de visée de *VEGETATION*, pour les quatre longueurs d'onde du capteur et pour une position du soleil donnée ;
- les fractions de trou mono- ($P_o(\theta_v)$) et bi-directionnelle ($P_o(\theta_s, \theta_v)$) simulées pour les angles de visée de *VEGETATION*, et la fraction de trou mono-directionnelle dans trois directions particulières : $P_o(\theta_v = 0^{\circ})$, $P_o(\theta_v = \theta_s)$, $P_o(\theta_v = 58^{\circ})$;
- le LAI , le $fAPAR$ intégré sur la journée et la teneur en chlorophylle intégrée au niveau du couvert : $LAI.C_{ab}$.

Cependant, avant d'utiliser cette base d'apprentissage, il est nécessaire d'effectuer quelques vérifications et de s'assurer de la cohérence des résultats. La figure 2.12 montre les réflectances obtenues pour les 4 longueurs d'onde de *VEGETATION*, ainsi que les fractions de trou mono- et bidirectionnelle obtenues dans le plan d'échantillonnage de *VEGETATION* ($\varphi = 36^\circ$), le 24 Octobre à 14° de latitude.

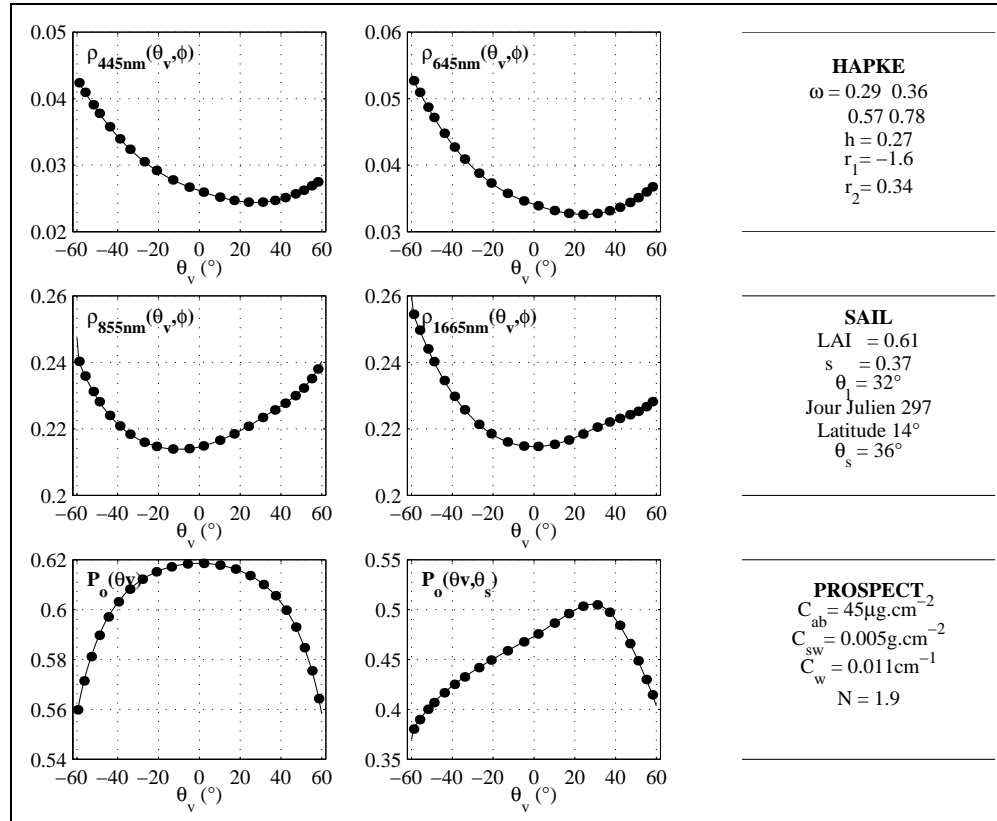


Figure 2.12 : Exemple de simulation lorsque *VEGETATION* échantillonne dans un plan éloigné du plan principal ($\varphi = 36^\circ$).

Les paramètres de propriétés optiques des feuilles du couvert considéré pour cette simulation se trouvent dans la moyenne (tableau 2.5) et le *LAI* est assez faible (0.6), ce qui explique les faibles réflectances obtenues dans le proche et le moyen infrarouge. Le plan d'échantillonnage ($\varphi = 36^\circ$) étant

éloigné du plan principal ($\varphi = 0^\circ$), on n'observe pas de pic marqué du hot spot. Le maximum de la fraction de trou monodirectionnelle, qui correspond à une visée au nadir, est représentatif du pourcentage de couverture du sol. Lorsque l'angle de visée augmente, cette fraction de trou diminue car la projection de végétation vue dans cet angle augmente. La fraction de trou bidirectionnelle correspond à la fraction de sol vu et éclairé, sa valeur est donc moins importante que celle de la fraction de trou mono-directionnelle. De plus, le maximum se trouve déplacé dans la direction solaire correspondant au hot spot.

On note le même comportement pour le couvert dont les propriétés optiques et les fractions de trou sont représentés figure 2.13. Cependant, bien que les propriétés optiques des feuilles soient similaires au couvert précédent, le LAI étant plus important, on observe une augmentation générale du niveau de réflectance dans le proche et le moyen infrarouge. De plus, on se trouve cette fois très près du plan principal ($\varphi = 2^\circ$), ce qui nous permet d'observer le phénomène de hot spot qui se traduit par une hausse du niveau de réflectance dans l'ensemble du plan avec un pic pour un angle zénithal de visée égal à l'angle zénithal solaire. Le pic observé est très étroit car le paramètre de hot spot est faible. Les valeurs de fractions de trou sont, par contre, plus faibles que pour le couvert de la figure 2.12 puisque, pour un angle moyen d'inclinaison des feuilles équivalent, le LAI est cinq fois plus grand. En outre, le pic de la fraction de trou bidirectionnelle est plus étroit car le paramètre de hot spot est bien plus petit.

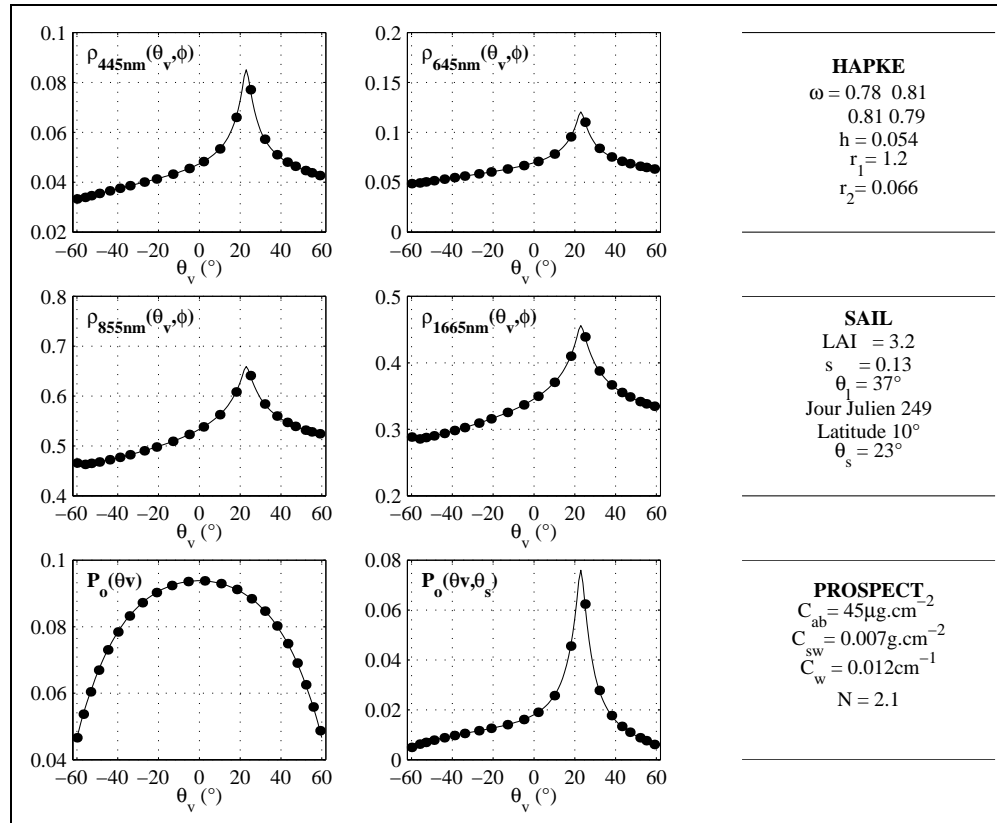


Figure 2.13 : Exemple de simulation lorsque *VEGETATION* échantillonne près du plan principal ($\varphi = 2^\circ$).



Construction d'une base de données synthétiques

-

Conclusions



Nous nous sommes attachés, dans ce chapitre, à construire une base de données synthétiques, c'est-à-dire simulées. Cette base de données contient, d'une part, les réflectances telles qu'elles seraient mesurées par le capteur VEGETATION pour un couvert donné, et d'autre part, les variables biophysiques correspondantes. Elle comprend 1500 simulations différentes. Cependant, compte tenu de la relative "simplicité" du modèle de réflectance de la végétation, les couverts ne sont pas décrits avec beaucoup de paramètres (3 paramètres structurels, 2 paramètres de propriétés optiques de feuilles et 3 paramètres de propriétés optiques du sol). Cela induit de fortes hypothèses sur le couvert simulé :

- *SAIL suppose que les plantes sont disposées aléatoirement de façon homogène dans le couvert, ce qui n'est pas toujours le cas, en particulier dans les cultures en rang, les couverts de type savanne ou certaines forêts ;*
- *le modèle ne tient pas compte de la présence de tiges ou de troncs, le couvert étant considéré comme une couche homogène de feuilles. De plus, nous n'avons considéré à chaque fois qu'une couche unique alors qu'il est possible de découper le couvert en couches homogènes horizontales pour simuler par exemple, une forêt dont le sous-bois est bien développé (une couche pour le sous-bois, une couche pour les feuilles d'arbres) ;*
- *les feuilles sont supposées lambertiennes (pas de variation directionnelle de la réflectance ou de la transmittance) et leurs propriétés optiques sur*

les 2 faces sont supposées identiques. De plus, leur orientation azimutale est supposée uniforme ce qui peut poser problème, par exemple, pour des plantes héliotropes comme le tournesol ou le soja.

Par ailleurs, compte tenu des lois statistiques d'échantillonnage que nous avons utilisées (loi ISLSCP pour le LAI, lois gaussiennes pour les propriétés optiques des feuilles), la précision d'estimation sera meilleure sur les cas les plus représentés dans la base (faibles LAI, valeurs moyennes du contenu en chlorophylle des feuilles). Le choix de cette représentativité est gouverné par notre connaissance expérimentale (base ISLSCP, jeu de données LOPEX) de la gamme de variation des variables.

La suite de l'étude consiste à utiliser cette base de données pour caler un algorithme d'extraction des caractéristiques biophysiques des couverts à l'aide de réseaux de neurones (chapitre 3). La validation de cet algorithme sur des jeux de données expérimentales (chapitre 4) permettra de montrer les limites imposées par le jeu de données synthétiques ainsi que celles de l'algorithme lui-même.

Chapitre 3

Estimation de variables biophysiques des couverts végétaux. Calage des réseaux de neurones.

L'objectif de ce chapitre est d'exploiter la base de données d'apprentissage élaborée au chapitre précédent pour développer un algorithme d'estimation des variables biophysiques des couverts végétaux. Comme il a été précisé en introduction (chapitre 2), il existe 3 types de méthodes d'estimation :

- les indices de végétation : c'est une méthode facile et rapide à mettre en oeuvre, mais qui ne peut pas être appliquée de façon globale car les paramètres des relations "indice de végétation/variable biophysique" dépendent de manière significative des conditions de mesures (sol, position solaire, ...) ;
- l'inversion de modèle de transfert radiatif : c'est une méthode permettant de prendre en compte explicitement l'effet des différents facteurs. Elle est cependant fastidieuse et longue à mettre en oeuvre. Elle pose en outre le problème du nombre de données disponibles par rapport au nombre de paramètres du modèle ;

- les méthodes hybrides qui représentent un bon compromis entre les deux méthodes précédentes, plus particulièrement en ce qui concerne le compromis “qualité d’estimation/temps de calcul”.

Dans le cas de l’estimation de variables biophysiques des couvert végétaux, les méthodes hybrides sont assez récentes mais les travaux déjà effectués sont prometteurs, comparativement aux indices de végétation. De plus, elles ont l’avantage une fois calées, d’être très rapides à mettre en œuvre par rapport aux méthodes d’inversion de modèles. Baret *et al.*(1995) ont développé un réseau de neurones pour estimer la fraction de trou au nadir $P_o(0)$ d’une culture de betterave en utilisant une base d’apprentissage composée de données aussi bien expérimentales que synthétiques. Dans cette étude, l’estimation de $P_o(0)$ se faisait à partir de réflectances mesurées au nadir par des radiomètres pour deux longueurs d’onde données (rouge et proche-infrarouge). Notre étude se trouve dans la continuité de ces travaux, dans la mesure où nous ajoutons une dimension directionnelle et deux longueurs d’onde supplémentaires du fait du capteur que nous considérons. De plus, notre objectif est d’évaluer dans quelle mesure les réseaux de neurones peuvent être utilisés pour estimer d’autres variables biophysiques. On cherche également à obtenir un algorithme global, capable d’estimer ces variables avec une précision acceptable sur n’importe quel type de couvert.

3.1 Choix d’un type de réseau de neurones

Comme nous l’avons déjà montré en introduction de ce mémoire, outre sa base d’apprentissage, un réseau de neurones est caractérisé par son architecture et sa règle d’apprentissage. Elles sont toutes deux imposées par le nombre mais aussi le type d’entrées et de sorties du neurone. Dans notre étude, nous utilisons un apprentissage supervisé, puisque dans la base d’apprentissage, nous connaissons la variable biophysique associée à chacun de nos jeux de réflectances. C’est la raison pour laquelle nous considérons des réseaux à rétro-propagation du gradient de l’erreur appelés aussi perceptrons multicouches (Rumelhart *et al.*, 1986).

3.1.1 L'algorithme de rétropropagation du gradient de l'erreur

On considère un réseau de neurones comprenant $N + 1$ couches et permettant d'estimer M variables $\hat{y}_{m,m=1,\dots,M}$. Il comporte donc M neurones pour la couche de sortie. L'algorithme de rétropropagation du gradient cherche à minimiser un critère d'erreur en ajustant les poids et les biais appropriés à chaque neurone du réseau. Pour simplifier la démonstration, on considère ici que les biais sont des poids qui correspondent à une entrée du réseau égale à 1. Par ailleurs, pour améliorer la compréhension du raisonnement, on ne considère que le cas d'un neurone (par exemple, le neurone 1 de la couche n_2), sachant que l'algorithme s'applique à l'ensemble des neurones du réseau (figure 3.1).

La procédure de minimisation est itérative et, à chaque itération $t_{t=0,\dots,T}$, on définit une fonction de coût C comme l'écart quadratique entre les sorties estimées par le réseau et les sorties mesurées $y_{m,m=1,\dots,M}$. D'après les notations que nous avons choisies, nous avons $\hat{y}_m = x_m^{n_{N+1}}$:

$$\mathcal{F} = \frac{1}{2} \sum_{m=1}^M (y_m - x_m^{n_{N+1}})^2 \quad (3.1)$$

Dans cet algorithme, il est nécessaire d'utiliser des fonctions de transfert continues et différentiables. Dans un premier temps, on initialise les poids à des valeurs aléatoires, puis pour chaque individu de la base d'apprentissage, on calcule le vecteur des sorties \hat{Y} obtenu par propagation, ainsi que la fonction de coût associée. Si celle-ci est supérieure à un seuil fixé par l'utilisateur, les poids sont modifiés par la procédure du gradient (μ est un gain fixé par l'utilisateur). Pour le neurone j de la couche $n_k + 1$, on a :

$$w_{ij}^{n_k}(t+1) = w_{ij}^{n_k}(t) - \mu \frac{\partial \mathcal{F}}{\partial w_{ij}^{n_k}(t)} \quad (3.2)$$

or

$$\frac{\partial \mathcal{F}}{\partial w_{ij}^{n_k}(t)} = \frac{\partial \mathcal{F}}{\partial a_j^{n_k}} \frac{\partial a_j^{n_k}}{\partial w_{ij}^{n_k}(t)} = -z_j^{n_k} \frac{\partial a_j^{n_k}}{\partial w_{ij}^{n_k}(t)} \quad (3.3)$$

$$= -z_j^{n_k} \frac{\partial \sum_p w_{pj}^{n_k}(t) x_p^{n_k}}{\partial w_{ij}^{n_k}(t)} = -z_j^{n_k} x_i^{n_k} \quad (3.4)$$

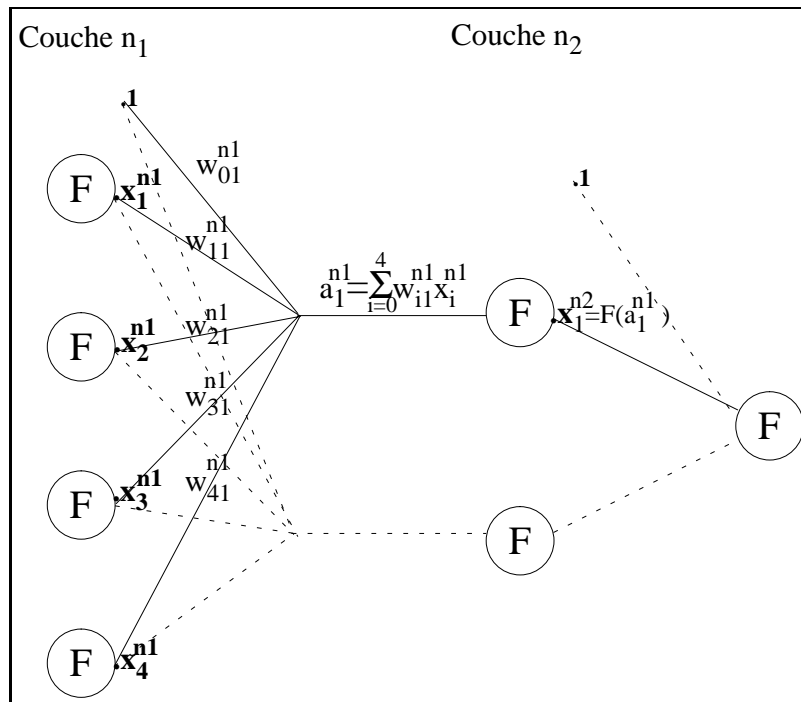


Figure 3.1 : Représentation schématisée du calcul de la sortie d'un réseau de neurone à 3 couches par propagation. $w_{ij}^{n_k}$ est le poids attribué à la sortie $x_i^{n_k}$ du neurone i de la couche n_k ($k = 1, \dots, N + 1$) vers le neurones j de la couche n_{k+1} . La somme $a_j^{n_k}$ des sorties $x_i^{n_k}$ de l'ensemble des neurones de la couche, pondérées par les poids $w_{ij}^{n_k}$ devient alors une entrée du neurone j de la couche n_{k+1} . F représente la fonction de transfert des neurones, de dérivée F' .

en posant $z_j^{n_k} = -\frac{\partial \mathcal{F}}{\partial a_j^{n_k}}$

- **Neurone j = neurone de la couche de sortie** ($n_k + 1 = N$) :

$$\begin{aligned} z_j^N &= -\frac{\partial \mathcal{F}}{\partial a_j^N} = -\frac{\partial \mathcal{F}}{\partial x_j^{N+1}} \frac{\partial x_j^{N+1}}{\partial a_j^N} \\ &= -\frac{1}{2} \frac{\partial (y_j - x_j^{N+1})^2}{\partial x_j^{N+1}} \frac{\partial F(a_j^N)}{\partial a_j^N} \\ &= (y_j - x_j^{N+1}) F'(a_j^N),_{j=1, \dots, M} \end{aligned}$$

- **Neurone j = Neurone d'entrée ou neurone de la couche caché :**

on suppose que les $z_j^{n_k+1}$ auxquels le neurone j envoie sa sortie $x_j^{n_k}$ sont connus (1 représente le nombre de neurones de la couche suivante) :

$$\begin{aligned} z_j^{n_k} &= -\frac{\partial \mathcal{F}}{\partial a_j^{n_k}} = \sum_l \left(-\frac{\partial \mathcal{F}}{\partial a_l^{n_k+1}} \frac{\partial a_l^{n_k+1}}{\partial a_j^{n_k}} \right) \\ &= \sum_l \left(z_l^{n_k+1} \frac{\partial a_l^{n_k+1}}{\partial x_j^{n_k+1}} \frac{\partial x_j^{n_k+1}}{\partial a_j^{n_k}} \right) = \sum_l \left(z_l^{n_k+1} w_{lj}^{n_k+1} F'(a_j^{n_k}) \right) \end{aligned}$$

Le passage à l'itération suivante se fait donc par :

Neurone de sortie :

$$w_{ij}^{N+1}(t+1) = w_{ij}^{N+1}(t) - \mu(y_j - \hat{y}_j) F'(a_j^N) x_i^N$$

Neurone d'entrée ou caché :

$$w_{ij}^{n_k}(t+1) = w_{ij}^{n_k}(t) - \mu x_i^{n_k} F'(a_j^{n_k}) \left[\sum_l z_l^{n_k+1} w_{lj}^{n_k+1}(t) \right]$$

Tant que l'erreur n'a pas atteint le seuil désiré, on modifie à nouveau les poids et les biais. Des améliorations ont été apportées à cet algorithme pour le faire converger plus rapidement. En plus du premier ordre (eq 3.2), on peut considérer le deuxième ordre qui donne une information supplémentaire sur la courbure de la surface d'erreur. C'est le cas de l'algorithme que nous utilisons,

basé sur l'optimisation de Levenberg-Marquardt (Finschi, 1996). Les fonctions de transfert associées aux neurones doivent donc être deux fois différentiables, ce qui est le cas lorsque l'on utilise des sigmoïdes (eq 3.5) ou des fonctions linéaires.

$$\text{log-sigmoïde : } f(x) = \frac{1}{1 + \exp(-x)} \quad (3.5)$$

$$\text{tangente-sigmoïde : } f(x) = \frac{2}{1 + \exp(-2x) - 1}$$

3.1.2 Amélioration des performances du réseau

Une façon simple d'améliorer les performances d'un réseau de neurones est de prétraiter les données en les normalisant par rapport à leur moyenne et leur écart-type. Ceci permet de recentrer les données autour d'une moyenne nulle, avec un écart-type de 1. La normalisation simultanée des données d'entrée et de sortie permet de penser que les poids et les biais du réseau seront de la même grandeur (Bishop, 1995).

De même, les poids et les biais sont tirés aléatoirement selon une loi uniforme dans un intervalle $[-\alpha, \alpha]$ (Rojas, 1996). La valeur de α ne doit pas être trop grande. En effet, si les poids et biais sont trop importants, cela rend les neurones correspondants inactifs car saturés. De plus, des valeurs trop importantes impliquent une augmentation de l'erreur qui est "rétropropagée" pour calculer à nouveau les poids et biais à l'itération suivante. Plusieurs solutions sont proposées pour la valeur de α (Wessels et Barnard, 1992, Drago et Ridella, 1992) : lorsque les entrées du réseau sont normalisées, α dépend, pour un neurone donné, du nombre d'entrées de ce neurone. Dans notre application, les neurones comportent au plus 9 entrées. Nous avons donc pris une valeur de 1 pour α , ce qui reste cohérent avec les valeurs proposées par les auteurs cités ci-dessus.

3.2 Pré-traitement des variables d'entrée des réseaux neuronaux : normalisation des données

Dans un premier temps, il est nécessaire d'adapter le réseau de neurones à notre application. En effet, le problème majeur consiste à tenir compte du fait

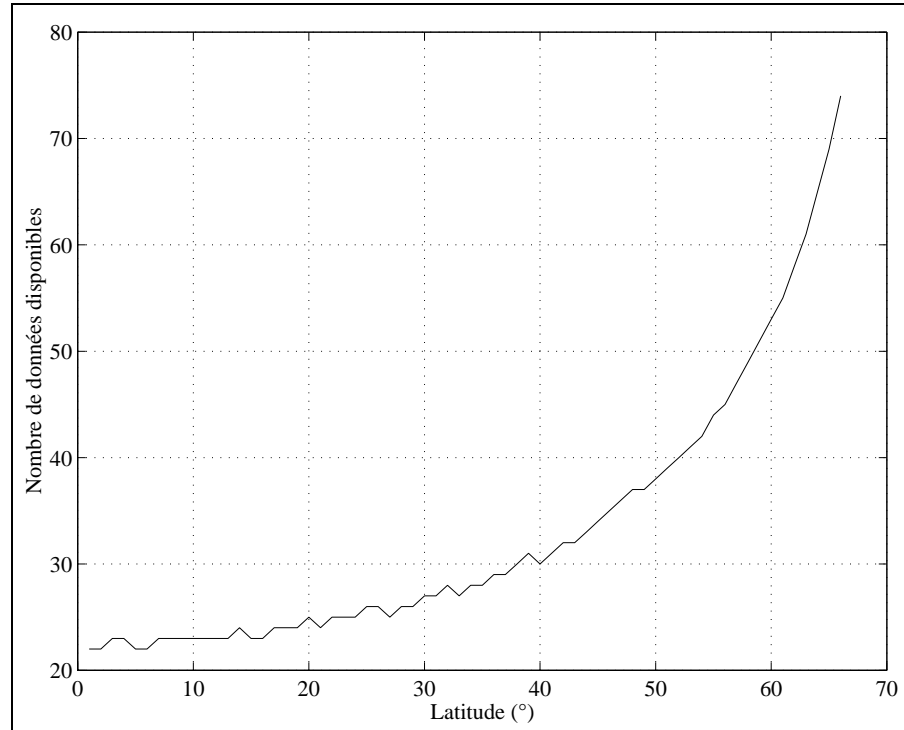


Figure 3.2 : Nombre de données de réflectance *VEGETATION* disponibles en fonction de la latitude, pour une longueur d'onde, sur un cycle de 26 jours.

que le nombre de données d'entrée du réseau, c'est-à-dire le jeu de réflectances dans les 4 longueurs d'onde de *VEGETATION* est variable selon la latitude, la date et la présence de nuages durant l'acquisition des données. Nous disposons donc d'un nombre d'entrées variable alors que le réseau de neurones en requiert un nombre constant.

Le cas le plus simple est de construire un réseau de neurones pour chaque configuration d'entrée. Cependant, étant donnée la variation du nombre de données disponibles en fonction de la latitude (figure 3.2), cette méthode serait beaucoup trop lourde à gérer. En admettant que ce nombre varie tous les degrés et qu'il n'y ait pas de nuage, il faudrait faire l'apprentissage d'une soixantaine de réseaux par variable biophysique à estimer.

Une autre solution proposée dans la littérature est de remplacer les données manquantes par une valeur moyenne de réflectance calculée sur le jeu

d'apprentissage (Steinmetz *et al.*, 1995). Ahmad et Tresp (1993) ont montré que cette méthode était loin d'être optimale, en particulier lorsque les données sont bruitées. Tresp *et al.*(1995) proposent de construire un algorithme permettant d'interpoler la donnée manquante à partir de celles qui sont déjà disponibles. Ceci peut se faire, par exemple, à l'aide d'un autre réseau de neurones ou bien en utilisant une fonction de distribution connue de la donnée d'entrée du réseau. Cette méthode ne semble pas envisageable dans notre cas, puisque les valeurs de réflectance en entrée du réseau correspondent à des configurations de visée et d'éclairement particulières, différentes pour chaque latitude et chaque jour considérés.

De plus, pour des latitudes élevées ($>55^\circ$), on peut avoir entre 40 et 70 données en entrée du réseau, ce qui va considérablement augmenter la complexité (nombre de couches et nombre de neurones par couche) du réseau, et donc le temps de calcul. Bishop (1995) note qu'en pratique, l'ajout de données en entrée du réseau peut en fait conduire à diminuer les performances de celui-ci. Il est alors conseillé de diminuer le nombre d'entrée en les prétraitant malgré la perte d'information que cela peut représenter.

La solution que nous avons adoptée consiste à garder un nombre constant d'entrées pour le réseau. Pour cela, on utilise une technique courante qui consiste à normaliser les données de réflectance par rapport aux angles de visée. En effet, Roujean *et al.* (1992) ont montré expérimentalement que les variations des données *NOAA/AVHRR* d'une date sur l'autre sont essentiellement dues à l'anisotropie de la surface observée lorsque l'on suppose que la correction atmosphérique a déjà été effectuée. Cette normalisation est très utile lorsque l'on veut comparer deux images de télédétection provenant d'un même capteur ou de deux capteurs dont les caractéristiques spectrales sont identiques mais correspondant à deux dates différentes. Actuellement, deux techniques sont utilisées :

- le MVC (Maximum Composite Value, Tarpley *et al.*, 1984) qui consiste à sélectionner sur une période pouvant aller d'une semaine à un mois, la donnée qui présente un *NDVI* maximum pour le pixel considéré ;
- l'autre technique consiste à calculer les paramètres (ou des paramètre dérivés) d'un modèle de BRDF¹ facilement inversible (Leroy et Roujean, 1994, Strahler *et al.*, 1995, Weiss *et al.*, 1999, ...).

¹Bidirectional Reflectance Distribution Function

Dans la mesure où nous disposons d'une variation angulaire de la réflectance, nous utiliserons la deuxième méthode. Le problème revient alors à choisir un modèle de BRDF et les paramètres directs ou dérivés de ce modèle. Les critères de choix sont :

- Un modèle linéaire ou semi-linéaire : dans un souci de précision et de rapidité d'exécution de l'inversion, la linéarité permet de réduire le processus d'inversion au calcul d'une matrice "pseudo-inverse" (car, en général, il y a plus de données que d'équations) ;
- Un nombre de paramètres limité : Lorsqu'on se situe à des latitudes faibles, par exemple à l'équateur, seulement 22 données sont disponibles. Si l'on considère 80% de nébulosité, ce qui peut arriver sous ce type de latitude, il ne reste que 4 données donc 4 équations pour l'inversion du modèle. Le modèle nécessite donc moins de 5 paramètres ;
- Le modèle doit bien s'ajuster sur les données et les reproduire avec une bonne précision pour conserver toute l'information directionnelle et spectrale contenue dans les réflectances de départ ;
- La valeur des paramètres du modèle doit rester aussi indépendante que possible de l'échantillonnage directionnel. En effet, elle ne doit pas varier que l'on inverse le modèle sur une partie (présence de nuages rendant les données inexploitable) ou bien sur l'ensemble des données acquises durant le cycle orbital.

3.2.1 Les modèles paramétriques de BRDF

Les modèles paramétriques de BRDF sont généralement des modèles empiriques ou semi-empiriques qui nécessitent peu de paramètres d'entrée. Leur avantage principal est la rapidité du processus d'inversion qui permet de traiter des images de télédétection pixel à pixel sur une période donnée, de manière opérationnelle. De nombreux modèles existent dans la littérature (Roujean *et al.*, 1992, Shibayama et Wiegand, 1985, Walthall *et al.*, 1985, Engelsen *et al.*, 1996, ...); ils peuvent souvent se mettre sous la forme d'une combinaison linéaire de noyaux \mathcal{N}_j , qui sont des fonctions de la géométrie de mesure (angles zénithaux solaire θ_s et de visée θ_v et l'angle azimutal relatif φ entre le capteur

et le soleil²) :

$$\rho(\theta_s, \theta_v, \varphi) = \sum_{i=1}^n a_i \mathcal{N}_i(\theta_s, \theta_v, \varphi) \quad (3.6)$$

Les coefficients a_i dépendent de la longueur d'onde. Lorsque l'on dispose de m mesures de réflectances correspondant à m configurations de mesure, on obtient le système :

$$\begin{bmatrix} \rho(\theta_{s1}, \theta_{v1}, \varphi_1) \\ \vdots \\ \rho(\theta_{sm}, \theta_{vm}, \varphi_m) \end{bmatrix} = \begin{bmatrix} \mathcal{N}_1(\theta_{s1}, \theta_{v1}, \varphi_1) & \cdots & \mathcal{N}_n(\theta_{s1}, \theta_{v1}, \varphi_1) \\ \vdots & \ddots & \vdots \\ \mathcal{N}_1(\theta_{sm}, \theta_{vm}, \varphi_m) & \cdots & \mathcal{N}_n(\theta_{sm}, \theta_{vm}, \varphi_m) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

soit

$$[\boldsymbol{\rho}] = [\mathcal{N}][\mathbf{a}]$$

L'inversion du modèle consiste à obtenir la matrice des paramètres $[\mathbf{a}]$ à partir de la géométrie de mesure et des réflectances correspondantes :

$$[\mathbf{a}] = [\mathcal{N}]^{-1}[\boldsymbol{\rho}] \quad (3.7)$$

La matrice $[\mathcal{N}]^{-1}$ est la matrice inverse des noyaux lorsque le nombre de mesures est égal au nombre de paramètres du modèle. Or, dans la plupart des cas, on dispose de plus de mesures que de paramètres, ce qui surdimensionne le problème et conduit à une minimisation au sens des moindres carrés par le biais d'une matrice pseudo-inverse.

Strahler *et al.*(1995) montrent que sur 19 modèles paramétriques de BRDF, le modèle de Walthall (1985) est globalement le plus performant sur un ensemble de couverts diversifiés (prairies, cultures, forêts, végétation éparse) pour délivrer les produits du capteur *MODIS*. Suite à une étude complémentaire que nous avons réalisée sur 4 modèles de BRDF (Baret *et al.*, 1997), nous avons choisi de tester ce modèle ainsi que le modèle MRPV développé par Engelsen *et al.*, en 1996.

²lorsque le modèle n'est pas réciproque, les noyaux ne dépendent que des angles de visée et les coefficients de la combinaison linéaire dépendent de la position solaire et de la longueur d'onde.

3.2.2 Le modèle de Walthall

Le modèle de Walthall est basé sur la similitude entre la forme globale d'une BRDF et le limaçon de Pascal caractérisé par deux paramètres. À cela, s'ajoute un troisième paramètre pour obtenir un ajustement plus précis sur la forme de la BRDF. D'abord élaboré sur des données simulées, il a ensuite été testé sur des données expérimentales de soja et de sols nus (Walthall *et al.*, 1985). L'expression de la réflectance bidirectionnelle est la suivante :

$$\rho(\theta_s, \theta_v, \varphi) = a_1(\theta_s) + a_2(\theta_s)\theta_v \cos\varphi + a_3(\theta_s)\theta_v^2$$

Le modèle n'étant pas réciproque, l'angle solaire n'est pas pris en compte dans l'expression des noyaux. Par contre, les paramètres dépendent de l'angle zénithal solaire (Walthall, 1997). Par ailleurs, a_1 représente la valeur de la réflectance au nadir $\rho_o(\theta_s)$ ($\theta_v = 0$). Cette équation s'intègre analytiquement pour calculer la réflectance directionnelle hémisphérique $\rho_h(\theta_s)$ (ou albédo spectral) qui est une variable importante caractéristique des couverts végétaux (Pinty et Verstraete, 1992):

$$\begin{aligned} \rho_h(\theta_s) &= \frac{1}{\pi} \int_0^{2\pi} d\varphi \int_0^{\frac{\pi}{2}} \rho(\theta_s, \theta_v, \varphi) \sin\theta_v \cos\theta_v d\theta_v \\ &= \rho_o(\theta_s) + a_3(\theta_s) \frac{\pi^2 - 4}{8} \end{aligned}$$

La réflectance bidirectionnelle peut donc finalement s'écrire comme :

$$\rho(\theta_s, \theta_v, \varphi) = \rho_o(\theta_s) + a_2(\theta_s)\theta_v \cos\varphi + 8 \cdot \frac{\rho_h(\theta_s) - \rho_o(\theta_s)}{\pi^2 - 4} \theta_v^2 \quad (3.8)$$

Elle est fonction de deux paramètres physiques ($\rho_o(\theta_s), \rho_h(\theta_s)$) et d'un paramètre empirique ($a_2(\theta_s)$).

3.2.3 Le modèle MRPV

C'est un modèle à trois paramètres dérivé du modèle de Rahman (Rahman *et al.*, 1993). Contrairement au modèle de Walthall, MRPV obéit au principe de réciprocity, c'est-à-dire que l'on peut intervertir les positions du capteur et du soleil sans que la valeur de réflectance soit modifiée. En outre, il prend explicitement en compte le phénomène de hot spot :

$$\rho(\theta_s, \theta_v, \varphi) = a_1 M(\theta_s, \theta_v, a_2) F_r(\xi, a_3) H_r(\bar{\rho}, \theta_s, \theta_v, \varphi) \quad (3.9)$$

où

$$- M(\theta_s, \theta_v, a_2) = \frac{\cos^{a_2-1} \theta_v \cos^{a_2-1} \theta_s}{(\cos \theta_s + \cos \theta_v)^{1-a_2}}$$

est dérivé de la fonction de Minnaert (1941) utilisée dans un premier temps pour décrire la réflectance de surface de la lune;

$$- F_r(\xi, a_3) = e^{-a_3 \cos \xi}$$

sert à moduler toutes les contributions en diffusion avant et arrière;

$\xi = \arccos(\cos \theta_v \cos \theta_s + \sin \theta_v \sin \theta_s \cos \varphi)$ est l'angle de phase;

$$- H_r(\bar{\rho}, \theta_s, \theta_v, \varphi) = 1 + \frac{1-\bar{\rho}}{1+(\tan^2 \theta_s + \tan^2 \theta_v - 2 \tan \theta_s \tan \theta_v \cos \varphi)^{1/2}}$$

est la fonction de hot spot; $\bar{\rho}$ est la réflectance moyenne mesurée.

L'équation 3.9 est linéarisable et on obtient :

$$\ln \frac{\rho(\theta_s, \theta_v, \varphi)}{H_r(\bar{\rho}, \theta_s, \theta_v, \varphi)} = \ln a_1 + (a_2 - 1) \ln [\cos \theta_s \cos \theta_v (\cos \theta_s + \cos \theta_v)] - a_3 \cos \xi \quad (3.10)$$

Le calcul analytique de la réflectance hémisphérique étant très complexe, il est nécessaire d'effectuer une intégration numérique de la BRDF. Pour cela, une fois le modèle inversé sur la séquence de données *VEGETATION*, on intègre la BRDF dans une quadrature de Gauss avec les paramètres a_i trouvés par inversion. Afin d'obtenir un bon compromis entre le temps et la précision du calcul numérique, nous avons choisi d'utiliser une quadrature de 576 couples d'angles (24 angles zénithaux \times 24 angles azimutaux (tableau 3.1) :

$$\begin{aligned} \rho_h(\theta_s) &= \frac{1}{\pi} \int_0^{2\pi} d\varphi \int_0^{\frac{\pi}{2}} \rho(\theta_s, \theta_v, \varphi) \sin \theta_v \cos \theta_v d\theta_v \\ &= \sum_{i=1}^{24} \left(W_{\theta_i} \sum_{j=1}^{24} W_{\varphi_j} \rho(\theta_s, \theta_i, \varphi_j) \right) \end{aligned}$$

| $\theta_v(^{\circ})$ | W_{θ} | $\varphi(^{\circ})$ | W_{φ} |
|----------------------|--------------|---------------------|---------------|
| 2.84 | 0.0032 | 5.62 | 0.0123 |
| 6.52 | 0.0073 | 12.90 | 0.0285 |
| 10.22 | 0.0115 | 20.23 | 0.0443 |
| 13.92 | 0.0156 | 27.57 | 0.0593 |
| 17.63 | 0.0196 | 34.91 | 0.0733 |
| 21.34 | 0.0236 | 42.25 | 0.0862 |
| 25.05 | 0.0274 | 49.60 | 0.0976 |
| 28.76 | 0.0312 | 56.94 | 0.1074 |
| 32.48 | 0.0348 | 64.29 | 0.1155 |
| 36.18 | 0.0382 | 71.63 | 0.1217 |
| 39.90 | 0.0415 | 78.98 | 0.1258 |
| 43.61 | 0.0447 | 86.32 | 0.1279 |
| 47.32 | 0.0476 | 93.67 | 0.1279 |
| 51.03 | 0.0504 | 101.01 | 0.1258 |
| 54.74 | 0.0529 | 108.36 | 0.1217 |
| 58.45 | 0.0552 | 115.71 | 0.1155 |
| 62.16 | 0.0573 | 123.05 | 0.1074 |
| 65.87 | 0.0591 | 130.39 | 0.0976 |
| 69.58 | 0.0607 | 137.74 | 0.0862 |
| 73.29 | 0.0620 | 145.08 | 0.0733 |
| 77.01 | 0.0631 | 152.42 | 0.0593 |
| 80.72 | 0.0639 | 159.76 | 0.0443 |
| 84.43 | 0.0645 | 167.09 | 0.0285 |
| 88.14 | 0.0647 | 174.36 | 0.0123 |

Tableau 3.1 : Angles et poids correspondant pour l'intégration de Gauss.

3.2.4 Comparaison des performances des deux modèles

Les 2 modèles ont été testés selon deux critères :

- Leur aptitude à bien reproduire la BRDF pour les angles de visée *VE-GETATION* ainsi que leur aptitude à extrapoler la BRDF dans d'autres angles de visée, pour un angle solaire donné ;
- la stabilité des valeurs des paramètres du modèle en fonction du nombre

de données que l'on utilise pour l'inversion. En effet, qu'il y ait présence de nuages ou non lors du cycle d'acquisition du capteur, il est souhaitable que les paramètres restent à peu près constants puisqu'ils correspondent à un même couvert observé dans les mêmes conditions (date, latitude).

Dans un premier temps, pour s'assurer de la cohérence des résultats obtenus pour les deux modèles, nous avons comparé $\rho_o(\theta_s)$ et $\rho_h(\theta_s)$ calculées par inversion des modèles avec les valeurs calculées lors des simulations. La figure 3.3 présente les histogrammes des résidus (écarts entre valeur mesurée et valeur observée) pour l'estimation des deux variables par les deux modèles. On tient compte ici de la séquence *VEGETATION* complète pour l'inversion du modèle.

Les deux modèles ont sensiblement le même comportement avec une meilleure estimation de la réflectance au nadir que de la réflectance hémisphérique. En effet, les caractéristiques orbitographiques de *VEGETATION* font que le capteur échantillonne près du nadir dans la plupart des cas (99,4% des couverts sont observés au moins une fois dans le cycle avec un angle zénithal de visée inférieur à 5°). La donnée de la réflectance au nadir est donc prise en compte lors de l'inversion des modèles. Par contre, pour avoir une bonne estimation de la réflectance hémisphérique, les angles de visée doivent être bien répartis sur l'ensemble de l'hémisphère, ce qui est rarement le cas puisque *VEGETATION* échantillonne avec un angle azimutal relatif relativement constant et un angle zénithal inférieur à 60° . C'est la raison pour laquelle l'histogramme des résidus est plus étalé pour la réflectance hémisphérique. Il est cependant bien centré sur 0, tout comme celui correspondant à la visée verticale. Les deux modèles présentent donc une bonne cohérence des résultats, et nous allons pouvoir, par la suite, évaluer leurs performances selon les deux critères mentionnés ci-dessus.

Restitution de l'information directionnelle

Afin de vérifier que la variation directionnelle de la réflectance est bien décrite par les paramètres du modèle, nous avons inversé les deux modèles sur la séquence *VEGETATION* pour un angle solaire et une nébulosité donnés, puis reconstitué la BRDF dans les 4 longueurs d'onde du capteur pour la même position solaire et pour :

- La séquence *VEGETATION* utilisée pour l'inversion ;

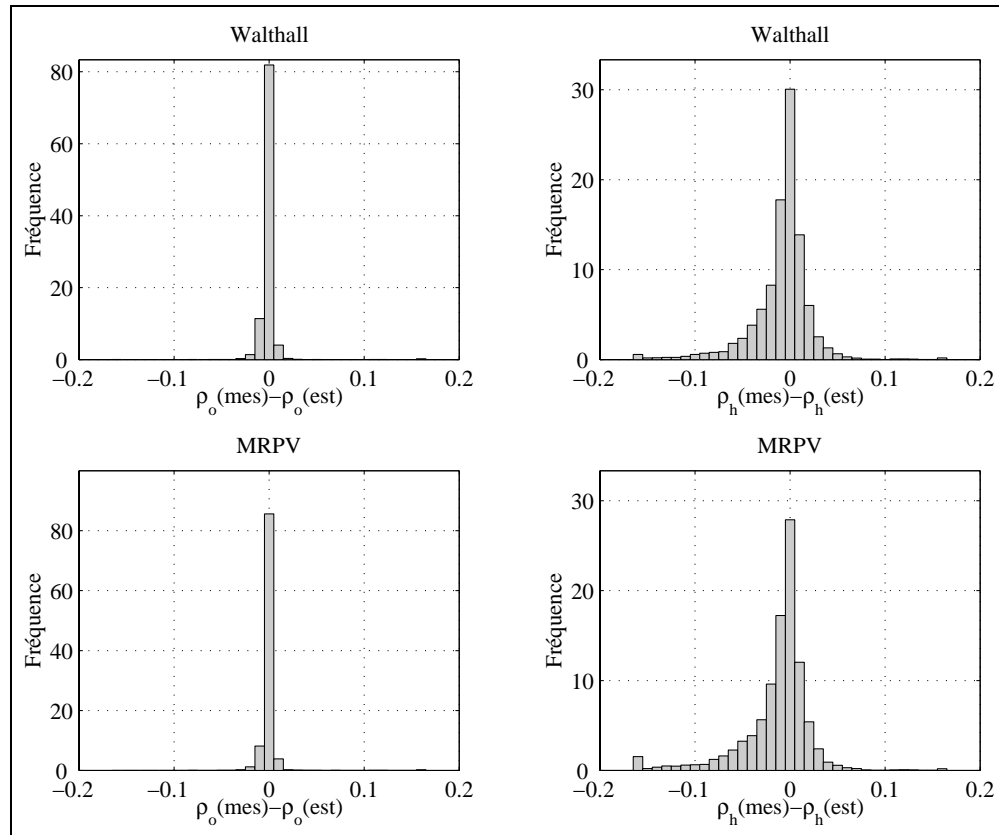


Figure 3.3 : Histogramme des résidus sur l'estimation de ρ_o et ρ_h par les modèles de Walthall et MRPV, lorsqu'on considère la séquence *VEGETATION* complète.

- La séquence *VEGETATION* complète correspondant aux 26 jours du cycle sans nébulosité ;
- Les 476 couples d'angles correspondant à la quadrature de Gauss, et qui permettent de bien couvrir tout l'hémisphère.

Dans chaque cas, nous avons évalué le *RMSE* entre la valeur "mesurée", c'est-à-dire provenant du jeu de 1500 simulations et la valeur estimée par les modèles à partir des paramètres obtenus par inversion. La figure 3.4 présente l'évolution du *RMSE* en fonction de la nébulosité pour les deux modèles. La

BRDF est bien restituée dans les angles de visée ayant servi pour l'inversion (en trait continu) et le $RMSE$ reste inférieur à 0.006. Il diminue lorsque la nébulosité augmente car le nombre de données utilisées pour l'inversion est réduit et les modèles s'ajustent donc mieux. Cependant, cette tendance est inversée lorsqu'on extrapole les valeurs de la BRDF à d'autres angles de visée que ceux utilisés pour l'inversion.

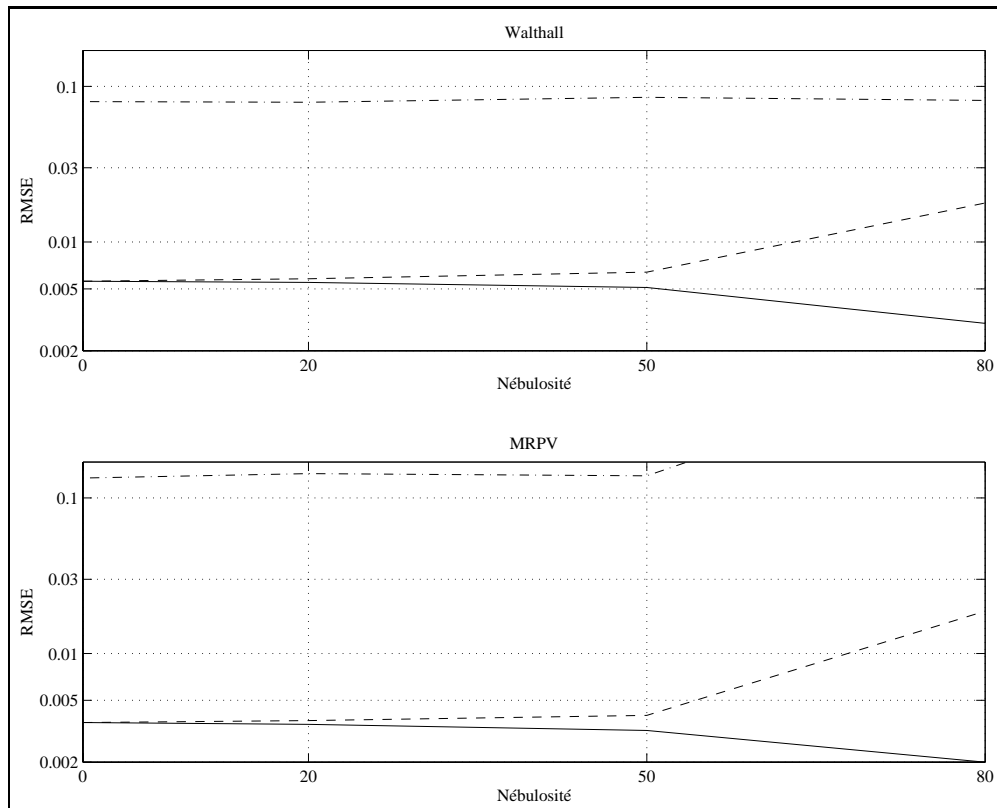


Figure 3.4 : Restitution de l'information directionnelle par les modèles de Walthall et MRPV (échelle semi-logarithmique) : — dans les angles *VEGETATION* ayant servi pour l'inversion; - - dans la séquence *VEGETATION* complète; -.- dans les 476 angles de Gauss.

Si l'on observe les capacités d'extrapolation dans la séquence *VEGETATION* (en pointillés), les deux modèles sont performants avec un $RMSE$ inférieur à 0.003, le modèle MRPV donnant des résultats légèrement supérieurs.

Par contre, lorsqu'il s'agit de restituer la BRDF dans tout l'hémisphère, le modèle de Walthall est plus performant ($RMSE < 0.1$) que le modèle MRPV ($RMSE$ de l'ordre de 0.2), en particulier lorsque la nébulosité est importante. On peut donc considérer que l'on ne perd quasiment pas d'information en utilisant les paramètres du modèle de Walthall ou MRPV en entrée du réseau, et on s'affranchit ainsi du problème de données manquantes. De plus, il faut noter que le modèle de Walthall extrapole mieux les valeurs de BRDF dans d'autres configurations de visée. On remarquera tout de même que, compte tenu de l'orbitographie de *VEGETATION*, très peu d'observations dans la direction du hot spot sont réalisées. Il est vraisemblable qu'avec un autre système de mesure (par exemple, *PolDER*), ou bien des caractéristiques orbitographiques différentes permettant des observations dans la direction du hot spot le modèle MRPV se révélerait meilleur.

Stabilité des paramètres d'entrée des modèles en présence de nuages

Comme nous l'avons déjà noté, il est important que la valeur des paramètres des modèles de BRDF varie peu selon qu'on inverse le modèle sur toutes les données *VEGETATION* ou sur seulement une partie. En effet, obtenir des paramètres différents correspondant à un même couvert, et donc une même variable biophysique, peut compliquer la tâche du réseau de neurones et l'induire en erreur. L'étude qui suit permet donc d'orienter non seulement le choix du modèle de BRDF mais aussi celui des paramètres retenus pour servir de variables d'entrée au réseau.

Pour analyser la stabilité des paramètres a_i des modèles de *BRDF*, nous avons calculé le $RMSE$ entre la valeur du paramètre pour une nébulosité donnée et celle obtenue pour une nébulosité nulle (figure 3.5a). En plus des 3 paramètres des modèles, nous avons aussi estimé les grandeurs physiques que sont la réflectance au nadir ρ_o (dans le cas de MRPV) et la réflectance hémisphérique ρ_h pour l'angle zénithal solaire utilisé pour l'inversion (figure 3.5a). Afin de pouvoir comparer les performances des 2 modèles et des paramètres les uns par rapport aux autres, nous avons normalisé les valeurs de $RMSE$ ($RMSE^*$) par rapport à l'amplitude du paramètre considéré sur les 1500 couverts (valeur calculée pour une nébulosité nulle).

Les valeurs de $RMSE$ relatif obtenues pour l'ensemble des paramètres des deux modèles sont satisfaisantes ($RMSE^* < 0.05$). Le niveau de $RMSE^*$ est globalement plus faible pour les paramètres du modèle de Walthall (<

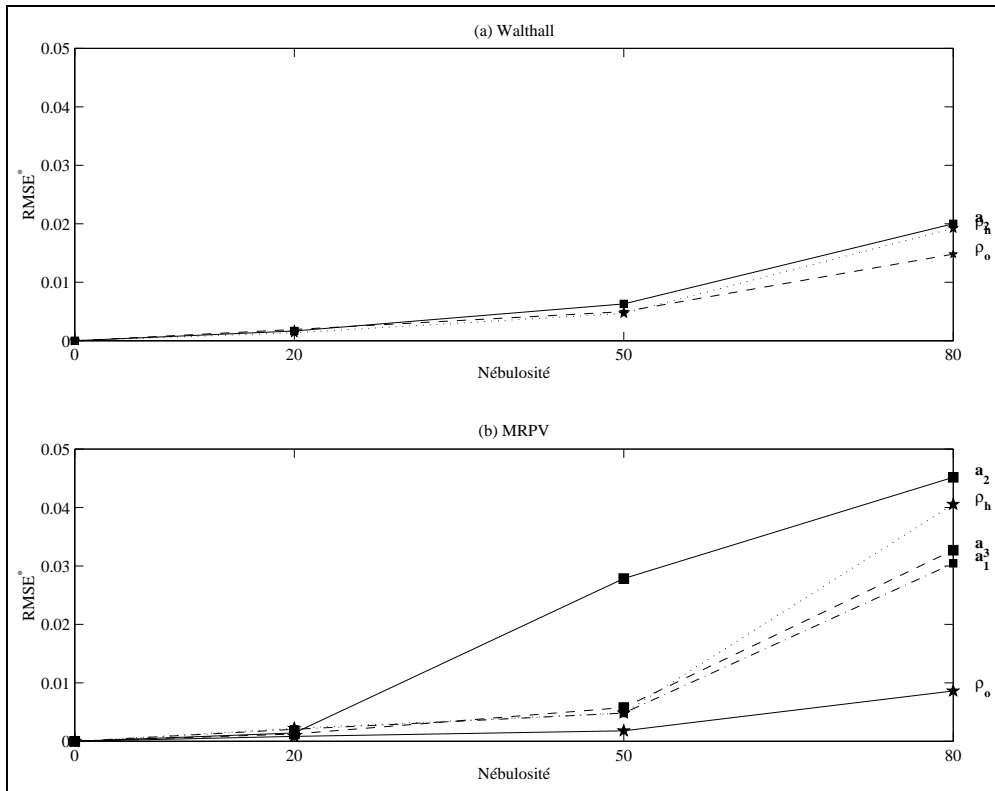


Figure 3.5 : Estimation des paramètres physiques (★) et empiriques (■) des modèles de Walthall et MRPV : stabilité de l'inversion en fonction de la nébulosité.

0.02) dont les paramètres physiques, plus particulièrement ρ_o , sont plus stables que le paramètre empirique a_2 . De plus, comme on pouvait s'y attendre, le modèle MRPV pose plus de problèmes pour l'estimation de la réflectance hémisphérique, puisqu'il extrapole moins bien dans les angles de Gauss que le modèle de Walthall. Il faut ajouter que la prise en compte du hot spot (possible seulement avec l'utilisation du modèle MRPV) par le modèle n'est pas nécessaire dans notre application pour deux raisons :

- *VEGETATION* échantillonne rarement dans le plan principal : 5% seulement des cas que nous avons simulés dans la base de données correspondent à une acquisition dans le plan principal $\pm 5^\circ$;

- Le phénomène de hot spot est très localisé et intervient donc avec un faible poids lors de l'intégration hémisphérique de la BRDF.

Au vu de ces résultats, nous avons choisi d'utiliser le modèle de Walthall, essentiellement pour sa capacité à bien extrapoler la BRDF dans d'autres angles que ceux utilisés pour l'inversion, en particulier quand peu de données sont disponibles. Dans l'ensemble, les 2 grandeurs physiques dérivées que sont ρ_o et ρ_h restent stables quel que soit le nombre d'angles de visée utilisés pour l'inversion. Nous choisissons donc ρ_o et ρ_h comme entrées du réseau de neurones.

Il peut également être intéressant de vérifier si, lors de la reconstruction de la BRDF, le modèle de Walthall ne donne pas des valeurs de *RMSE* qui varient suivant la position du soleil. En effet, la fraction de trou bidirectionnelle que nous souhaitons estimer dépend de θ_s . Bien que le modèle de Walthall ne prenne pas en compte explicitement la direction solaire, il existe une dépendance de ses paramètres d'entrée à la position du soleil car la réflectance dépend de la direction d'éclairement. La figure 3.6 montre la valeur du *RMSE* entre la BRDF reconstruite dans la séquence *VEGETATION* complète correspondant au cycle de 26 jours après inversion du modèle sur les données correspondant à une nébulosité variable (0%, 20%, 50% et 80%) pour différents angles zénithaux solaires. Le *RMSE* reste faible et augmente avec la nébulosité quelle que soit la fourchette d'angles solaires considérée. On note que plus l'angle zénithal solaire est réduit, plus le *RMSE* augmente. Ceci est simplement dû au nombre de données disponibles pour l'inversion qui passe de 12 en moyenne pour des angles solaires inférieurs à 20° correspondant à de faibles latitudes, à 25 pour des angles supérieurs à 60° aux plus fortes latitudes. Au vu de ce graphique, on ne constate donc pas de direction solaire privilégiée.

Conclusion

Pour obtenir un nombre constant de paramètres d'entrée requis par l'utilisation de réseaux de neurones, nous procédons dans un premier temps à une normalisation des données. Cette normalisation consiste à inverser le modèle de Walthall pour calculer les réflectances au nadir et hémisphérique dans chacune des longueurs d'onde caractéristiques du capteur. On peut noter en outre que cette normalisation est relativement indépendante de la quantité

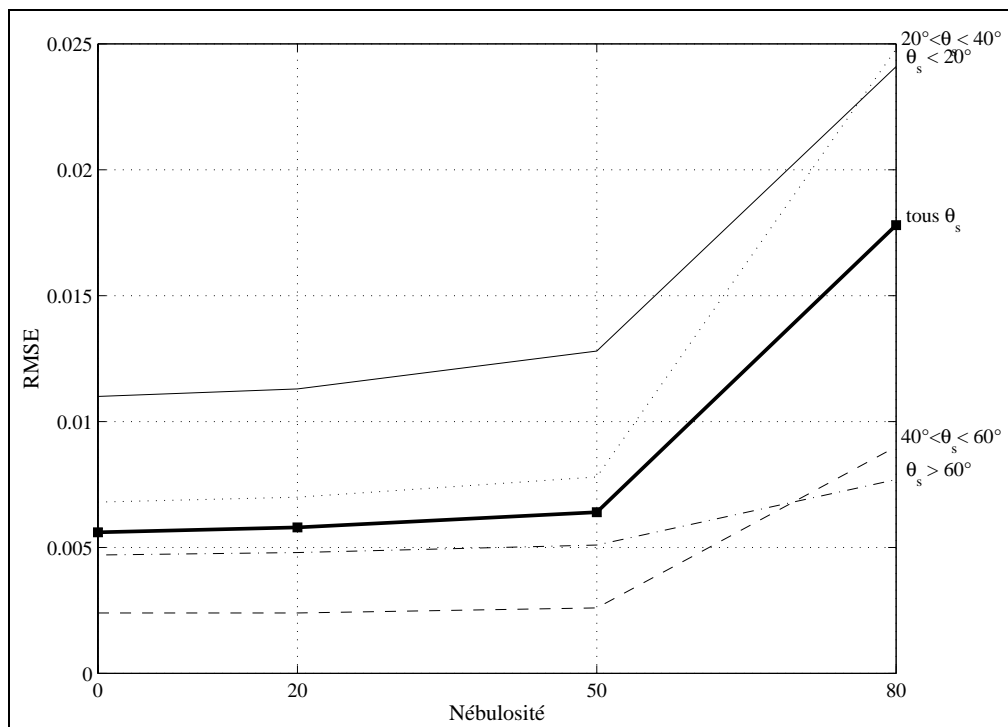


Figure 3.6 : Restitution de l'information directionnelle par le modèle de Walthall en fonction de la nébulosité, pour différentes gammes d'angles solaires.

de données dont on dispose pour l'inversion.

De plus, comme les paramètres du modèles de Walthall ainsi que ρ_o et ρ_h dépendent tous de la position solaire lors de l'acquisition des données VEGETATION, nous proposons d'ajouter l'angle zénithal solaire en entrée du réseau. Par ailleurs, l'angle azimutal solaire n'est pas utile car les deux grandeurs ρ_o et ρ_h utilisées en entrée du réseau ne dépendent pas de cette variable. En définitive, nous avons donc 9 variables d'entrée du réseau :

$$\left\{ \begin{array}{ll} (\rho_h, \rho_o)_{450nm} & (\rho_h, \rho_o)_{645nm} \\ (\rho_h, \rho_o)_{835nm} & (\rho_h, \rho_o)_{1665nm} \\ \theta_s & \end{array} \right.$$

3.3 Méthodologie d'utilisation des réseaux de neurones : estimation du taux de couverture du sol

Dans cette partie, nous décrivons la manière dont nous calons les réseaux de neurones pour estimer la fraction de trou au nadir $P_o(0)$. Cette variable est choisie pour illustrer l'approche utilisée, qui sera appliquée ensuite à l'estimation d'autres variables biophysiques du couvert. Les résultats obtenus pour la fraction de trou au nadir par les réseaux de neurones sont finalement comparés à la méthode classique, basée sur l'utilisation d'un indice de végétation.

3.3.1 Calage du réseau de neurones

Le calage d'un réseau de neurones consiste à déterminer les valeurs de poids et de biais (§3.1.2) qui permettent à l'estimation de $P_o(0)$ de s'ajuster au mieux à la valeur mesurée correspondante. Comme le recommande Bishop (1995), le jeu de données simulées que nous avons construit au préalable est divisé en trois sous-jeux indépendants composés de 500 couverts chacun :

- un jeu d'apprentissage : c'est le jeu qui contient les exemples qui permettent au réseau d'ajuster les poids et les biais de chaque neurone selon la loi de rétropropagation du gradient ;
- un jeu d'hyperspécialisation : il permet d'éviter que le réseau de neurones soit trop spécifique au jeu d'apprentissage (Sarle, 1995). Le phénomène d'hyperspécialisation provient du fait que l'on peut augmenter le nombre d'itérations pour obtenir de meilleurs résultats sur le jeu d'apprentissage mais cela peut se faire au détriment de la robustesse du réseau, qui risque alors d'avoir de piètres performances sur d'autres données. D'une certaine manière, l'hyperspécialisation d'un réseau amenuise ses facultés de généralisation. Le jeu d'hyperspécialisation est utilisé durant l'apprentissage. Toutes les cinquante itérations, on calcule le *RMSE* entre valeur mesurée et valeur estimée par le réseau de neurones pour les variables de sortie. Si, au bout de cent itérations, le *RMSE* diminue pour le jeu d'apprentissage mais augmente pour le jeu d'hyperspéciali-

sation, l'apprentissage du réseau est stoppé, considérant alors que l'apprentissage est optimal ;

- un jeu de validation : une fois que l'apprentissage du réseau est achevé, on présente les variables d'entrée du jeu de validation pour évaluer l'erreur commise sur l'estimation des variables biophysiques correspondantes.

Le problème du calage réside également dans le choix de l'architecture du réseau, c'est-à-dire, le nombre de couches, le nombre de neurones et les fonctions de transfert associés. Or, il n'existe pas de règle rigoureuse de choix de l'architecture. Elle dépend du nombre d'exemples de la base d'apprentissage et de la complexité de la relation "réflectance/variable biophysique" que l'on veut décrire (Rojas, 1996). La seule règle bénéficiant d'un consensus général est que le nombre d'exemples de la base d'apprentissage doit être assez important par rapport au nombre de neurones du réseau afin d'éviter l'hyperspécialisation. Malgré tout, le nombre de neurones doit être suffisant de manière à lui procurer une souplesse qui lui permette de s'ajuster à la complexité des relations entre variables d'entrée et variables de sorties du réseau.

Suite aux travaux de Baret *et al.*(1995), nous avons considéré un réseau de neurones à deux couches (pas de couche cachée) pour estimer la fraction de trou au nadir. Le nombre de neurones de la couche de sortie correspond au nombre de sorties (une seule, dans notre cas) et sa fonction de transfert est linéaire. Cela permet en effet d'obtenir une gamme de variation assez large sur les sorties de réseau, ce qui n'est pas le cas si l'on utilise des fonctions sigmoïdes. Pour la première couche, nous avons considéré des fonctions log-sigmoïdes et établi le nombre de neurones à 4. Si ce nombre de neurones est augmenté, le réseau se spécialise sur le jeu d'apprentissage ; s'il est diminué la précision d'estimation est amoindrie pour les trois jeux de données.

Lors de l'apprentissage, le réseau de neurones converge et se stabilise au bout d'une centaine d'itérations (figure 3.7). Une fois l'apprentissage stoppé, on observe que le réseau de neurones estime logiquement mieux la fraction de trou sur le jeu d'apprentissage que sur celui d'hyperspécialisation. Cependant, la valeur de l'écart quadratique moyen reste faible dans les deux cas (< 0.05). Au bout des 100 itérations, le *RMSE* reste bien stable pour les deux jeux même si l'on observe une très légère augmentation pour celui du jeu d'hyperspécialisation qui conduit à l'arrêt de l'apprentissage au bout de la

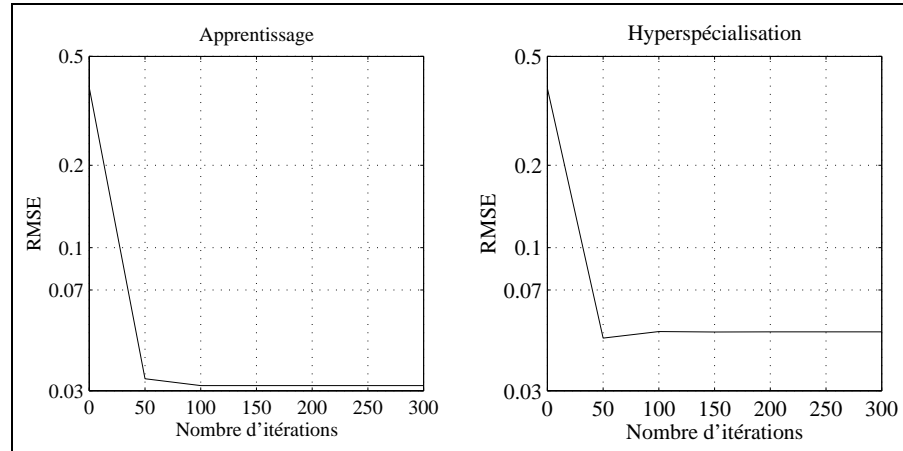


Figure 3.7 : Convergence du réseau de neurones pour l'estimation de la fraction de trou au nadir.

300ième itération.

Nous avons également testé l'effet de l'initialisation des poids et des biais sur les performances d'estimation du réseau de neurones. Nous avons construit 500 réseaux initialisés avec des poids et des biais différents mais équivalents par ailleurs (architecture, règle d'apprentissage). La figure 3.8 présente l'histogramme de répartition de l'écart quadratique moyen obtenu sur le jeu d'apprentissage pour les 500 réseaux. Selon l'initialisation, le réseau converge plus ou moins vite mais aboutit quasiment aux mêmes performances : 90% des réseaux estiment la fraction de trou au nadir avec une précision absolue comprise entre 0.03 et 0.04.

3.3.2 Performances de l'algorithme basé sur l'utilisation de réseau de neurones pour l'estimation de la fraction de trou au nadir

La figure 3.9 présente la procédure à suivre pour estimer la fraction de trou au nadir en utilisant l'algorithme que nous avons développé. Il se déroule en cinq étapes :

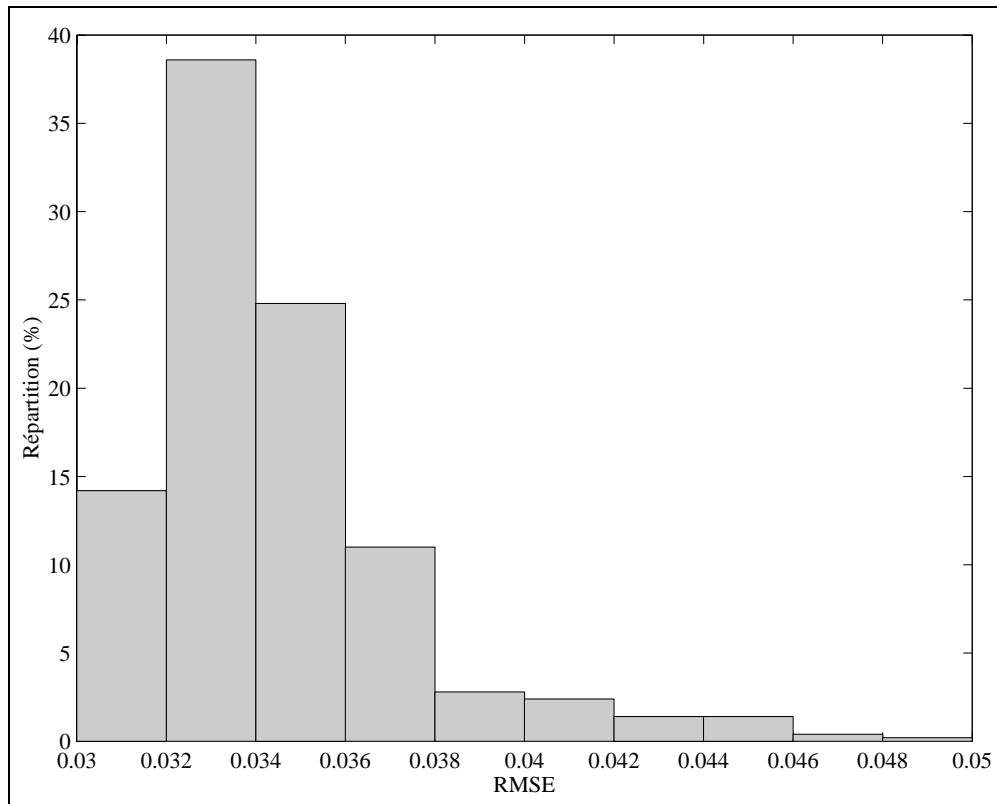


Figure 3.8 : Histogramme de répartition du $RMSE$ calculé entre les valeurs mesurées et les valeurs estimées par le réseau de neurones sur le fichier d'apprentissage.

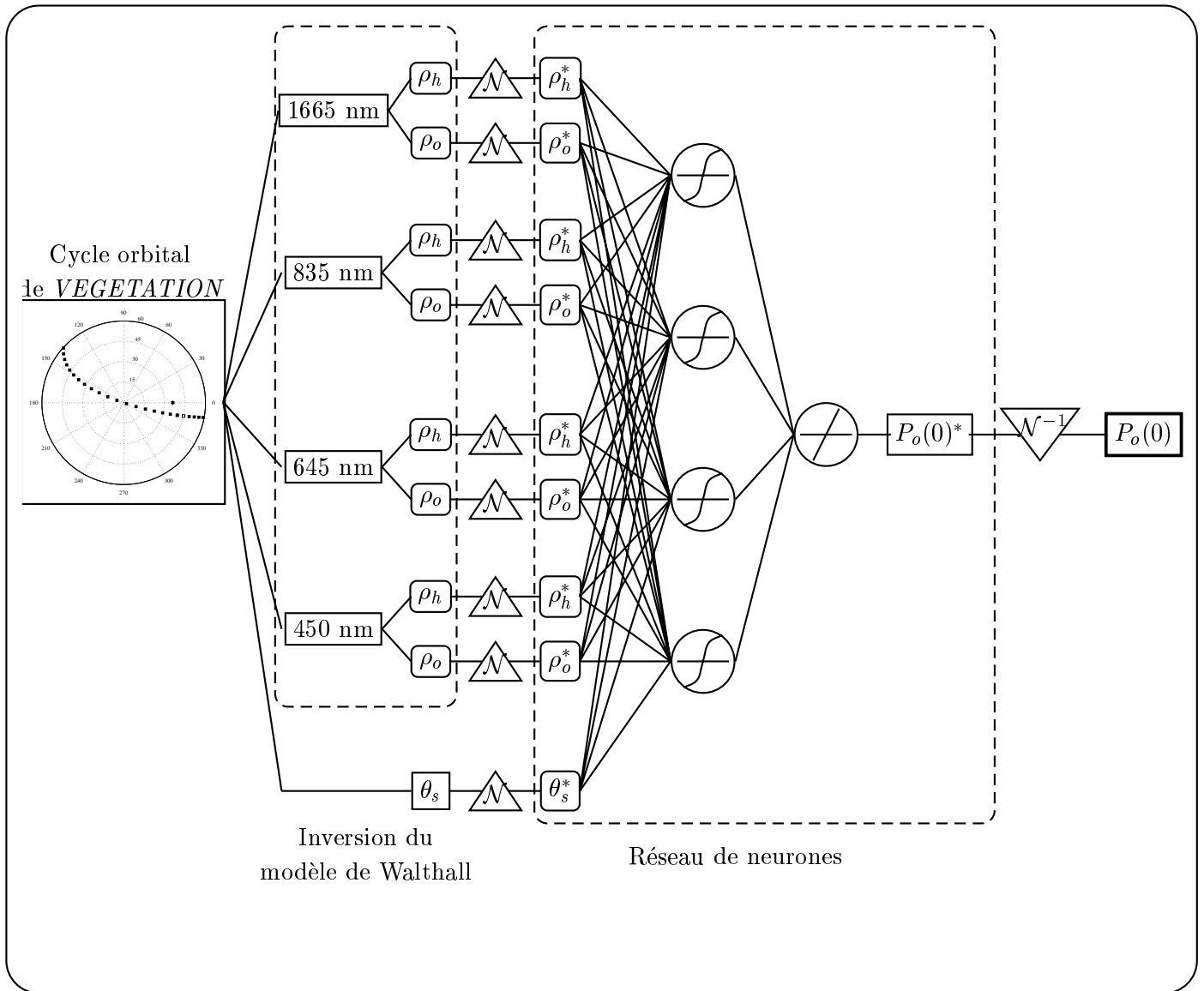


Figure 3.9 : Procédure d'estimation de la fraction de trou au nadir à l'aide d'un réseau de neurones. Le caractère \mathcal{N} (resp. \mathcal{N}^{-1}) correspond à la normalisation des variables d'entrée (resp. dénormalisation de la variable de sortie).

1. Inversion du modèle de Walthall sur les données de réflectance correspondant aux caractéristiques spectrales et directionnelles de *VEGETATION* ;
2. Calcul des réflectances au nadir et hémisphérique dans les quatre longueurs d'onde, pour l'angle zénithal solaire correspondant à l'acquisition *VEGETATION* ;
3. Normalisation de chacune des variables x d'entrée (ρ_o, ρ_h, θ_s) et de sortie (variable biophysique souhaitée) du réseau par rapport à leur moyenne \bar{x} et leur écart-type σ calculés sur le jeu d'apprentissage (eq 3.11) ;

$$x_{norm} = \frac{x - \bar{x}}{\sigma} \quad (3.11)$$

4. Passage dans le réseau ;
5. Dénormalisation de la variable biophysique estimée en sortie du réseau (à partir des moyenne et écart-type calculés sur le jeu d'apprentissage).

Les performances d'estimation sont exprimées en termes de *RMSE* et du paramètre statistique T (Eq 3.12) calculés entre la variable "mesurée" (c'est-à-dire provenant de la base de données construite au chapitre 2), et la variable estimée par le réseau de neurones.

$$T = 1 - \frac{\sum_n (y - \hat{y})^2}{\sum_n (y - \bar{y})^2} \quad (3.12)$$

où y et \hat{y} sont les variables mesurées et estimées, n le nombre de données et \bar{y} la moyenne de y sur les n échantillons.

Le paramètre T mesure l'égalité de deux variables (plus il est proche de 1, meilleure est l'estimation). Il est donc égal au coefficient de détermination R^2 lorsque la droite de régression est la première bissectrice ($y = x$). Ainsi, quand le coefficient de corrélation est proche de 1 et que, parallèlement T en est éloigné, cela signifie qu'il y a peu de bruit mais que le biais est important. Compte tenu de son expression mathématique, le T est adimensionnel, à l'inverse du *RMSE*. C'est la raison pour laquelle, pour comparer les performances d'estimation de deux variables x exprimées dans des unités différentes (par exemple, le *LAI* et *LAI.C_{ab}*), nous considérerons également un *RMSE*

relatif ($RMSE^*$, Eq 3.13), en le normalisant par rapport à l'amplitude maximale de variation observée. Dans le cas de la fraction de trou qui est comprise entre 0 et 1, le $RMSE$ est égal au $RMSE$ relatif.

$$RMSE^* = \frac{RMSE}{\max(x) - \min(x)} \quad (3.13)$$

La figure 3.10a présente les résultats que nous avons obtenus sur l'estimation de la fraction de trou au nadir pour le jeu de validation. La précision d'estimation est très bonne avec un $RMSE$ de 0.04 et un T de 0.98. La valeur de T est proche du $R^2 = 0.99$, ce qui montre qu'il n'y a pas de biais. Les figures 3.10b, c et d donnent les histogrammes de répartition des $RMSE^{**}$ en fonction de la fraction de trou, du LAI et de l'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$. Pour que les résultats soient comparables, chaque classe représentée dans l'histogramme comporte le même nombre d'individus (50), et le $RMSE^{**}$ est égal au $RMSE$ divisé par la valeur médiane de la fraction de trou sur la classe de 50 individus considérée. La figure 3.10b montre que les fractions de trou élevées, correspondant à des sols pratiquement nus, sont globalement mieux estimées. Ceci est confirmé par l'histogramme de la figure 3.10c qui indique que les fortes valeurs de LAI , correspondant à des "couverts couvrants", induisent de plus fortes erreurs relatives $RMSE^{**}$. Ceci est dû au fait que d'une part, la réflectance des couverts végétaux dans le proche infrarouge sature pour des LAI supérieurs à 3 ; et d'autre part, que les couverts de fort LAI sont peu représentés dans notre base d'apprentissage. L'angle moyen d'inclinaison des feuilles n'a, par contre, que peu d'influence sur l'estimation puisqu'il donne une répartition du $RMSE^{**}$ quasiment uniforme.

3.3.3 Performances de l'algorithme basé sur l'utilisation du $NDVI$ pour l'estimation de la fraction de trou au nadir

La relation " $NDVI/P_o(0)$ "

Comme nous l'avons présenté au chapitre 1, de nombreux indices de végétation ont été développés. Nous avons choisi de comparer notre algorithme à celui basé sur le $NDVI$ (Eq 1.2) pour différentes raisons :

- c'est l'indice le plus utilisé, particulièrement pour établir des cartes de paramètres biophysiques pour le suivi de la végétation, essentiellement à

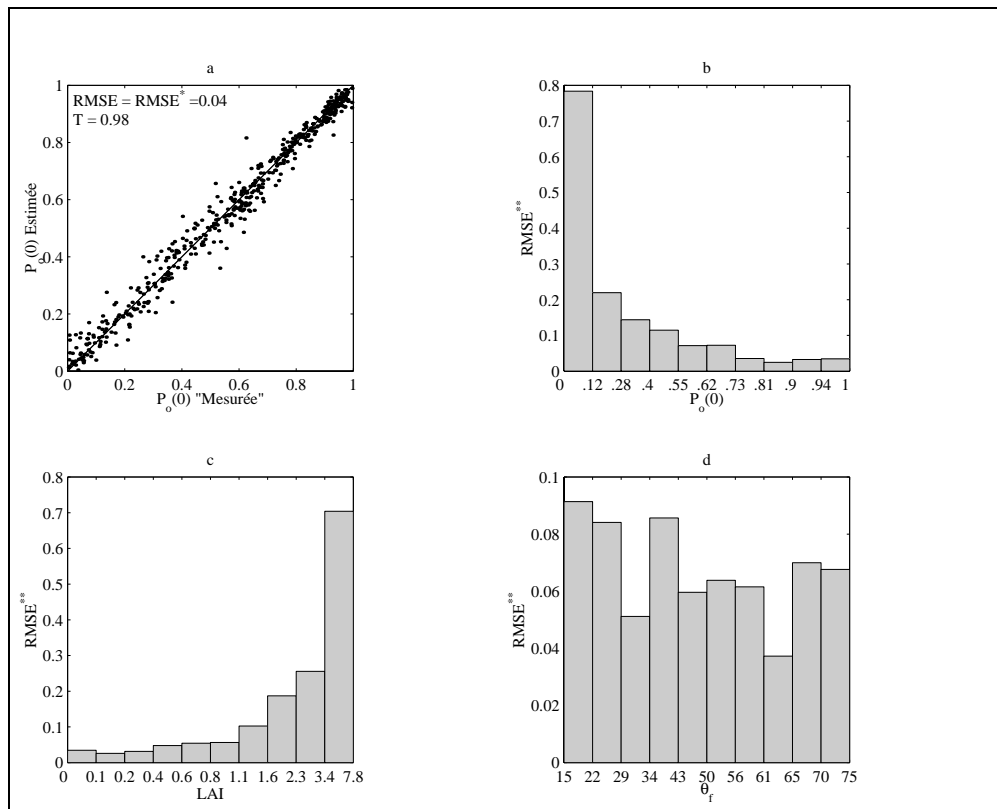


Figure 3.10 : Réseau de neurones : estimation de la fraction de trou au nadir sur le jeu de validation (a). Analyse des résidus en fonction de $P_o(0)$ (b), du LAI (c) et de l'angle moyen d'inclinaison des feuilles (d).

partir de données *NOAA/AVHRR* (algorithme du FASIR, Sellers *et al.*, 1994 pour l'estimation du $fAPAR$ et du LAI) ;

- compte tenu du fait que nous nous plaçons au niveau du couvert et que la correction des effets atmosphériques est supposée réalisée avec une bonne précision, il n'est pas utile d'utiliser des indices qui minimisent les effets atmosphériques ;
- de même, il est inutile de considérer des indices qui minimisent les effets de géométrie de visée car le calcul du $NDVI$ est fait pour un produit normalisé. Nous avons choisi la réflectance au nadir, car c'est celle qui

est la mieux estimée par inversion du modèle de Walthall (figure 3.3) ;

- de plus, les indices qui minimisent les effets du sol nécessitent la connaissance *a priori* de la droite des sols. Or, cette donnée est difficilement accessible en tout point du globe terrestre. C'est pourquoi on ne peut pas envisager d'utiliser facilement de tels indices.

Le $NDVI$, comme la fraction de trou au nadir, est reliée en première approximation au LAI par des lois de type Beer-Lambert (Asrar *et al.*, 1984, Nilson, 1971) :

$$P_o(0) = e^{-K_{P_o} \cdot LAI} \quad (3.14)$$

$$NDVI = NDVI_{\infty} + (NDVI_s - NDVI_{\infty})e^{-K_{NDVI} LAI} \quad (3.15)$$

où

$NDVI_{\infty}$ est la valeur asymptotique du $NDVI$ quand le LAI tend vers l'infini (en pratique, cette limite est atteinte pour un LAI de 8) ;

$NDVI_s$ est le $NDVI$ du sol nu ;

K_{NDVI} et K_{P_o} sont équivalents à des coefficients d'extinction.

Des équations précédentes, on peut déduire directement $P_o(0)$ en fonction du $NDVI$:

$$P_o(0) = \left[\frac{NDVI - NDVI_{\infty}}{NDVI_s - NDVI_{\infty}} \right]^K, K = \frac{K_{P_o}}{K_{NDVI}} \quad (3.16)$$

Calage de la relation

Pour comparer les performances de la méthode basée sur le $NDVI$ à celle basée sur le réseau de neurones, nous avons calé cette relation sur le jeu d'apprentissage. Les trois paramètres $[NDVI_s, NDVI_{\infty}, K]$ ont donc été évalués en utilisant la méthode du simplexe (Nelder et Mead, 1987). Cette méthode d'optimisation minimise une fonction d'erreur (norme de la différence) entre la fraction de trou au nadir "mesurée" et celle estimée par la relation 3.16.

Pour éviter les problèmes de minimum local, ainsi que minimiser l'effet de l'initialisation des paramètres, le calage de la relation 3.16 est fait en deux temps : $NDVI_\infty$ et $NDVI_s$ sont d'abord estimés par inversion de la relation 3.15, puis intégrés dans l'équation 3.16 pour obtenir la valeur de K . Les valeurs obtenues sont :

$$\begin{aligned} NDVI_\infty &= 0.8 \\ NDVI_s &= 0.2 \\ K &= 0.47 \end{aligned}$$

La figure 3.11a présente la relation “ $NDVI/LAI$ ” obtenue sur le jeu de validation. La loi de Beer-Lambert est donc une bonne approche de la réalité. Les valeurs mesurées (points) sont bien regroupés autour de la courbe décrite par l'équation 3.15. Toutefois, pour des LAI élevés ($LAI > 3$), le $NDVI$ sature et il n'y a plus moyen de faire la différence entre un couvert d'indice foliaire 4 et un autre d'indice 8. On observe le même phénomène lorsque l'on exprime la fraction de trou au nadir directement à partir du $NDVI$ où pour une même valeur de $NDVI$ (0.8), on peut observer des couverts ayant une probabilité de trous de 0 à 25%, ce qui correspond aussi à un LAI élevé (figure 3.11b). La fraction de trou est donc assez mal estimée par la relation 3.16. La figure 3.11c montre que la dispersion autour de la droite (1:1) entre la valeur de fraction de trous mesurée et celle estimée est importante. Le $RMSE$ est loin d'être satisfaisant (~ 0.17).

La figure 3.12 montre les répartitions des écarts quadratiques moyens (calculées comme pour les réseaux de neurones) en fonction des différentes variables qui régissent la relation $P_o(0)/NDVI$. Comme pour les réseaux de neurones, on observe une augmentation (resp. une diminution) du $RMSE^{**}$ en fonction du LAI (resp. de $P_o(0)$) due à la saturation du signal pour les forts LAI . Les valeurs de $RMSE^{**}$ obtenues avec le $NDVI$ sont nettement supérieures à celles obtenues avec les réseaux de neurones. L'histogramme de répartition de l'erreur en fonction du $NDVI$ est assez uniforme. L'erreur relative est cependant importante pour les faibles valeurs de $NDVI$.

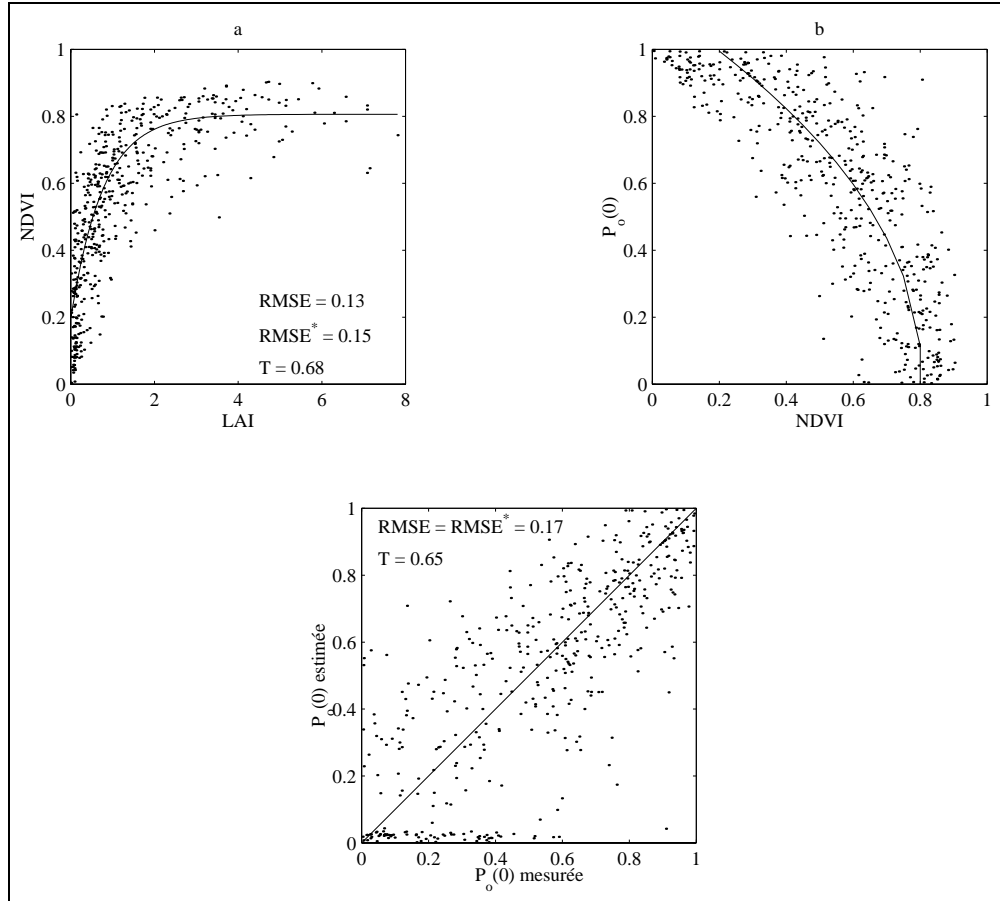


Figure 3.11 : *NDVI* : estimation de la fraction de trou au nadir sur le jeu de validation à l'aide du *NDVI*. (a) : ajustement de la relation *NDVI/LAI*, (b) : ajustement de la relation $P_o(0)/NDVI$, (c) : comparaison des valeurs de $P_o(0)$ estimées avec les valeurs mesurées.

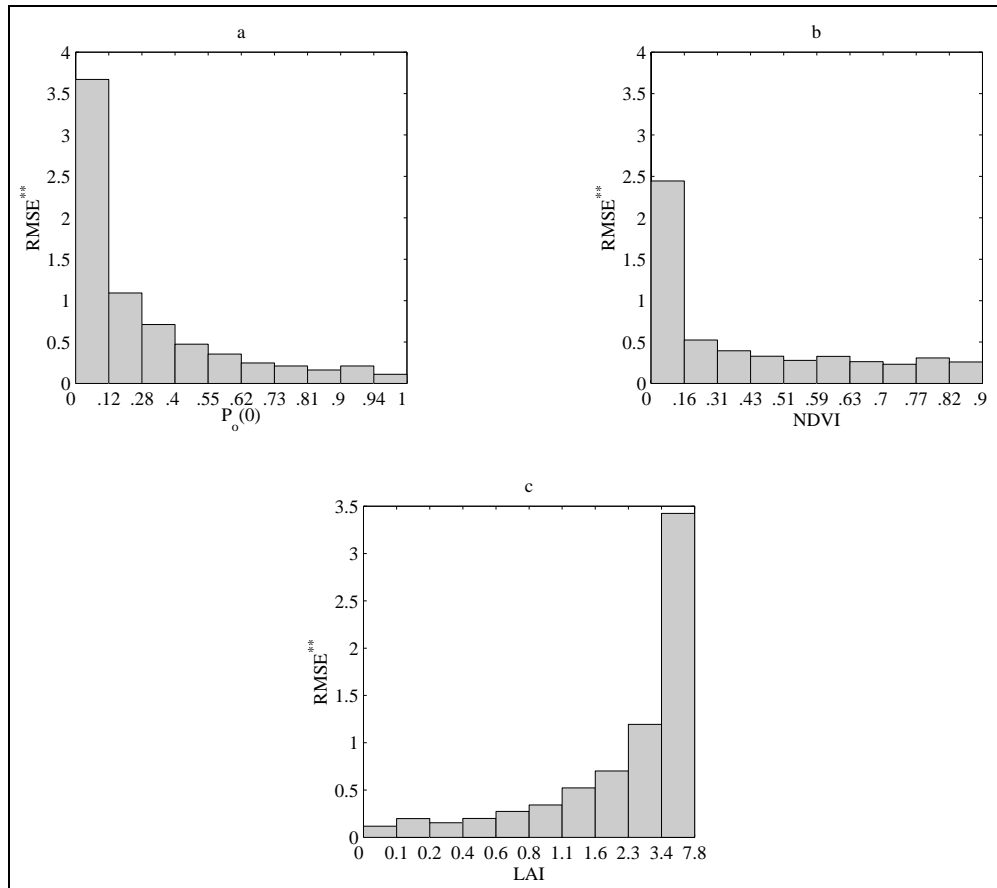


Figure 3.12 : Estimation de la fraction de trou au nadir sur le jeu de validation à l'aide du $NDVI$. Histogramme de répartition des $RMSE^{**}$ en fonction de la fraction de trou au nadir (a), du $NDVI$ (b) et du LAI (c).

3.3.4 Conclusion

Choix de la méthode d'estimation de variables biophysiques

Dans ce paragraphe, nous nous sommes attachés à estimer la fraction de trou au nadir en utilisant une méthode basée sur l'utilisation de réseaux de neurones et une autre basée sur l'utilisation du *NDVI*. Les paramètres employés pour chaque méthode ont été évalués sur le même jeu d'apprentissage. De même, les performances des deux méthodes ont été évaluées sur le même jeu de validation, indépendant du jeu d'apprentissage.

Les résultats obtenus montrent que le réseau de neurones estime $P_o(0)$ avec une bien meilleure précision que le *NDVI* (tableau 3.2). Cependant, on peut noter que pour les deux méthodes, il est plus difficile de déterminer la fraction de trou au nadir lorsque le *LAI* est élevé. Ceci est dû au fait que dans les deux cas, on utilise la réflectance dans le proche infrarouge qui, elle-même, présente une forte saturation pour les *LAI* importants.

| | <i>RMSE</i> | <i>T</i> |
|-----------------|-------------|----------|
| RN ^a | 0.04 | 0.98 |
| NDVI | 0.17 | 0.65 |

^aRN : Réseau de Neurones

Tableau 3.2 : Estimation de la fraction de trou au nadir. Comparaison de la méthode basée sur le *NDVI* et de celle basée sur l'utilisation de réseaux de neurones.

Des tests sur les autres variables biophysiques auxquelles nous nous intéressons (*LAI*, *fAPAR*, $P_o(\theta_s)$ et $P_o(58^\circ)$) montreraient certainement la supériorité des réseaux de neurones, car ceux-ci sont moins sensibles que le *NDVI* au phénomène de saturation que nous avons observé pour l'estimation de $P_o(0)$.

Influence de la présence de nuages sur l'estimation de la fraction de trou au nadir

Dans le paragraphe 3.2.4, nous avons choisi le modèle de Walthall pour calculer les variables d'entrée du réseau de neurones de manière à ce qu'elles dépendent le moins possible de la nébulosité. La figure 3.13 confirme que les

performances du réseau ne sont que très légèrement altérées en présence de nuages. En effet, le $RMSE$ (resp. le T), reste stable jusqu'à 50% de nuages et augmente (resp. diminue) faiblement pour 80% de nuages qui correspond à un nombre moyen de données disponibles pour l'inversion égal à 8 sur les 1500 cas simulés.

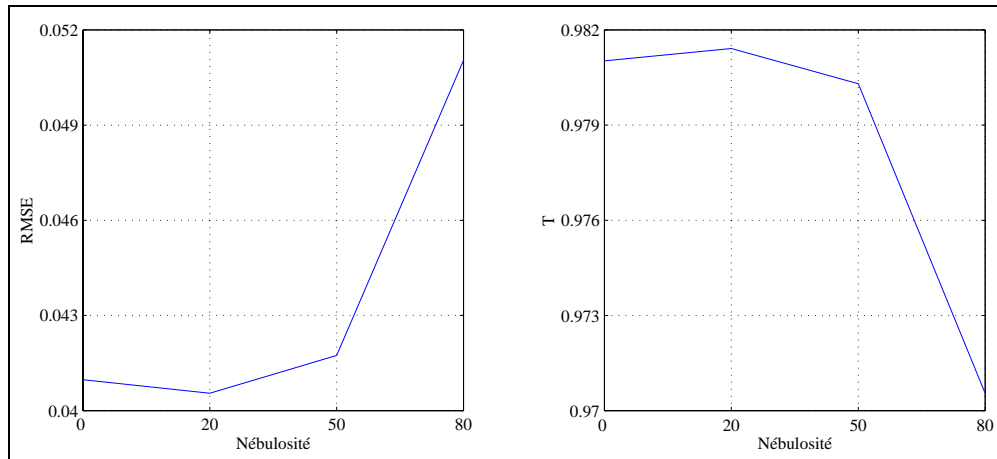


Figure 3.13 : Performances d'estimation de la fraction de trou au nadir en fonction de la couverture nuageuse (%).

Le programme ISCCP³ (Rossov *et al.*, 1988) a permis d'établir une répartition de la couche nuageuse sur l'ensemble du globe moyennée sur 8 ans. La figure 3.14 montre que la plupart des terres immergées ont une couverture nuageuse inférieure à 70%. La valeur de 80% que nous avons choisie pour tester notre algorithme correspond donc à des cas assez extrêmes. Il faut cependant noter que l'inversion du modèle de Walthall impose d'avoir au moins 3 journées non nuageuses sur les 26 jours du cycle orbital du capteur.

Dans la suite de cette étude, nous allons évaluer les performances du réseau de neurones pour estimer différentes variables biophysiques. Ceci nous permettra de choisir la variable biophysique la plus pertinente pour suivre la végétation à partir de données satellitales.

³International Satellite Cloud Climatology Project

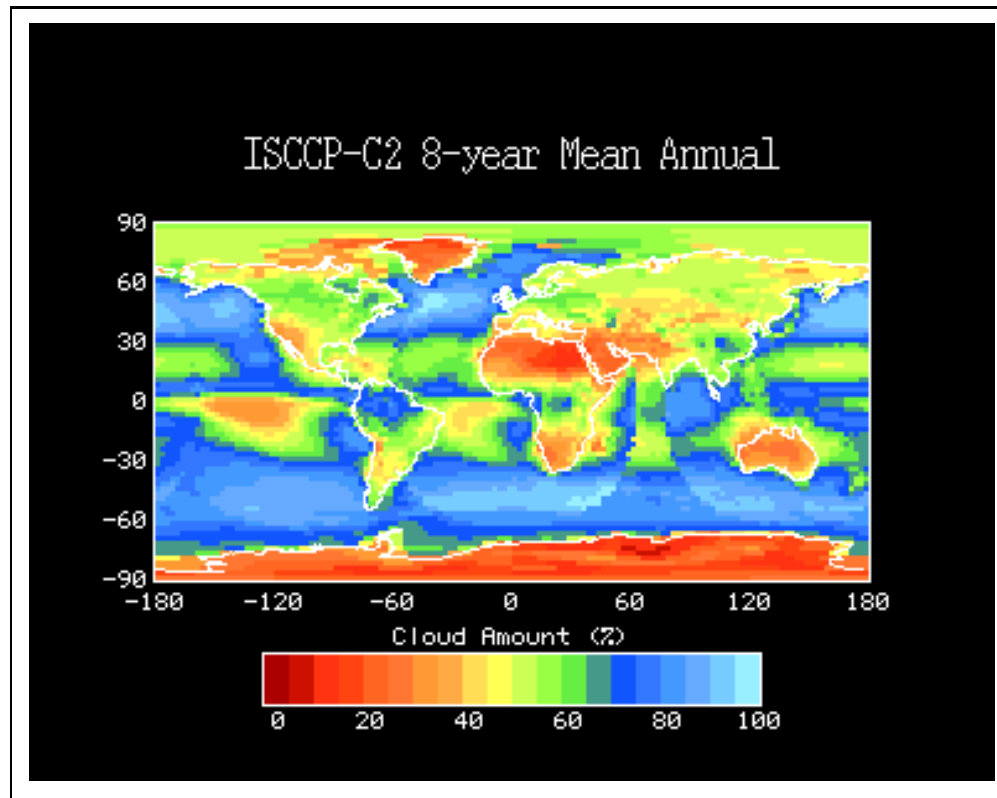


Figure 3.14 : Répartition nuageuse moyenne sur l'ensemble du globe terrestre établie sur 8 années (Images provenant de <http://isccp.gis.nasa.gov>).

3.4 Application à l'estimation des différentes variables biophysiques

Dans cette partie, nous appliquons la méthode basée sur l'utilisation de réseaux neuronaux pour estimer les variables biophysiques qui nous intéressent. Elle se déroule selon les étapes suivantes :

1. *Détermination de l'architecture du réseau* : pour chaque variable, nous débutons avec un réseau à deux couches contenant respectivement 4 et un neurones. Si le réseau se spécialise, on diminue le nombre de neurones ; dans le cas contraire et si les performances d'estimation ne sont pas satisfaisantes, nous augmentons le nombre de neurones jusqu'à obtenir un bon compromis. L'utilisation de fonctions tangente-sigmoïdes à la place des log-sigmoïdes pour la première couche est aussi envisagée ;
2. *Apprentissage du réseau* : nous évaluons l'influence de l'initialisation des poids et des biais pour chaque réseau sur les résultats (convergence et précision d'estimation) ;
3. *Validation du réseau* : la précision d'estimation du réseau est évaluée sur le jeu de validation ; les résultats obtenus sont ensuite discutés.

De plus, pour pouvoir comprendre et comparer les résultats que nous avons obtenus pour l'estimation des différentes variables, nous avons établi leur matrice de corrélation calculée sur les 500 cas du jeu de validation (tableau 3.3).

3.4.1 Estimation de la fraction de trou dans des directions particulières $P_o(\theta_s)$, $P_o(58^\circ)$

En plus de la fraction de trou au nadir, nous avons envisagé deux autres directions particulières : la fraction de trou dans la direction solaire (qui donne le complément du $fAPAR$ instantané) et la fraction de trou pour un angle de visée de 58° (qui donne une fraction de trou indépendante de la distribution des inclinaisons foliaires).

Les réseaux de neurones que nous avons construits pour ces deux variables possèdent la même architecture que celui de la fraction de trou au nadir avec des fonctions de transfert log-sigmoïdes. De même, l'initialisation des poids

| | LAI | $P_o(0)$ | $P_o(58^\circ)$ | $P_o(\theta_s)$ | $fAPAR$ | $LAI.C_{ab}$ | e^{-LAI} | $e^{-LAI.C_{ab}}$ |
|-------------------|-------|----------|-----------------|-----------------|---------|--------------|------------|-------------------|
| LAI | 1 | | | | | | | |
| $P_o(0)$ | -0.86 | 1 | | | | | | |
| $P_o(58^\circ)$ | -0.84 | 0.94 | 1 | | | | | |
| $P_o(\theta_s)$ | -0.82 | 0.93 | 0.94 | 1 | | | | |
| $fAPAR$ | 0.83 | -0.93 | -0.97 | -0.99 | 1 | | | |
| $LAI.C_{ab}$ | 0.89 | -0.79 | -0.77 | -0.76 | 0.80 | 1 | | |
| e^{-LAI} | -0.83 | 0.94 | 0.999 | 0.94 | -0.97 | -0.76 | 1 | |
| $e^{-LAI.C_{ab}}$ | -0.18 | 0.27 | 0.32 | 0.28 | -0.3 | -0.17 | 0.33 | 1 |

Tableau 3.3 : Coefficient de corrélation entre les variables biophysiques, pour les 500 cas du jeu de validation.

et des biais de chaque réseau n'a d'effet que sur le temps que met le réseau à converger.

La précision d'estimation pour $P_o(58^\circ)$ et $P_o(\theta_s)$ est aussi bonne que pour $P_o(0)$ (figures 3.15 et 3.16), puisque les trois variables sont fortement corrélées avec un coefficient de corrélation, R , supérieur à 0.93 (tableau 3.3). De même, les histogrammes de répartition des résidus montrent que les fractions de trou élevées sont mieux estimées. On retrouve naturellement la même tendance pour la répartition des résidus en fonction du LAI . L'angle moyen d'inclinaison des feuilles, par contre, n'a pratiquement pas d'influence sur la précision d'estimation. Dans le cas de la fraction de trou dans la direction solaire, nous observons que l'erreur est plus forte lorsque l'angle zénithal solaire est important. Il faut également noter que, comparativement à $P_o(0)$, les valeurs de $RMSE^{**}$ sont beaucoup plus fortes pour la première classe de 50 individus car la dynamique est beaucoup plus petite (0.12 pour $P_o(0)$ et 0.04 pour $P_o(\theta_s)$ et $P_o(58^\circ)$).

En définitive, la précision de l'estimation de la fraction de trou monodirectionnelle dans des directions particulières est tout à fait correcte, particulièrement en comparaison des méthodes basées sur le $NDVI$ (tableau 3.4).

3.4.2 Estimation de l'indice foliaire LAI

L'architecture du réseau à laquelle nous avons abouti est la même que pour la fraction de trou au nadir, c'est-à-dire une couche de 4 neurones à fonction

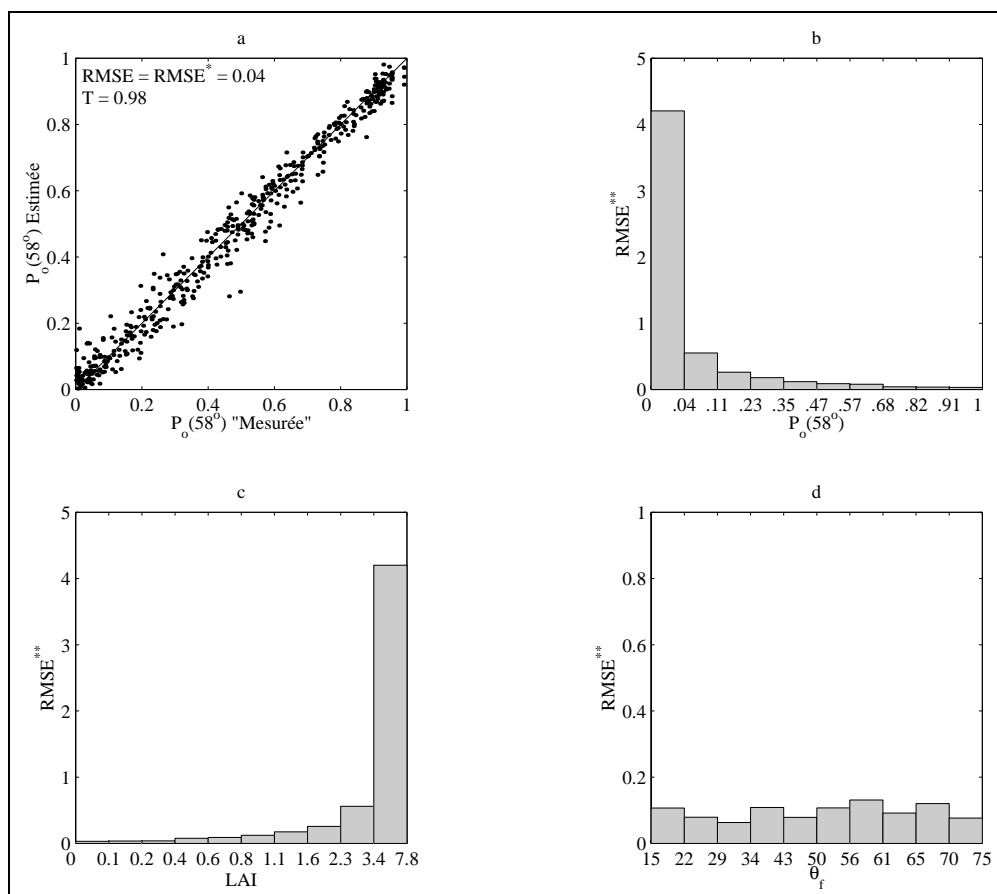


Figure 3.15 : Estimation de $P_o(58^\circ)$ (a et b). Analyse des résidus en fonction du LAI (c) et de l'angle moyen d'inclinaison des feuilles (d).

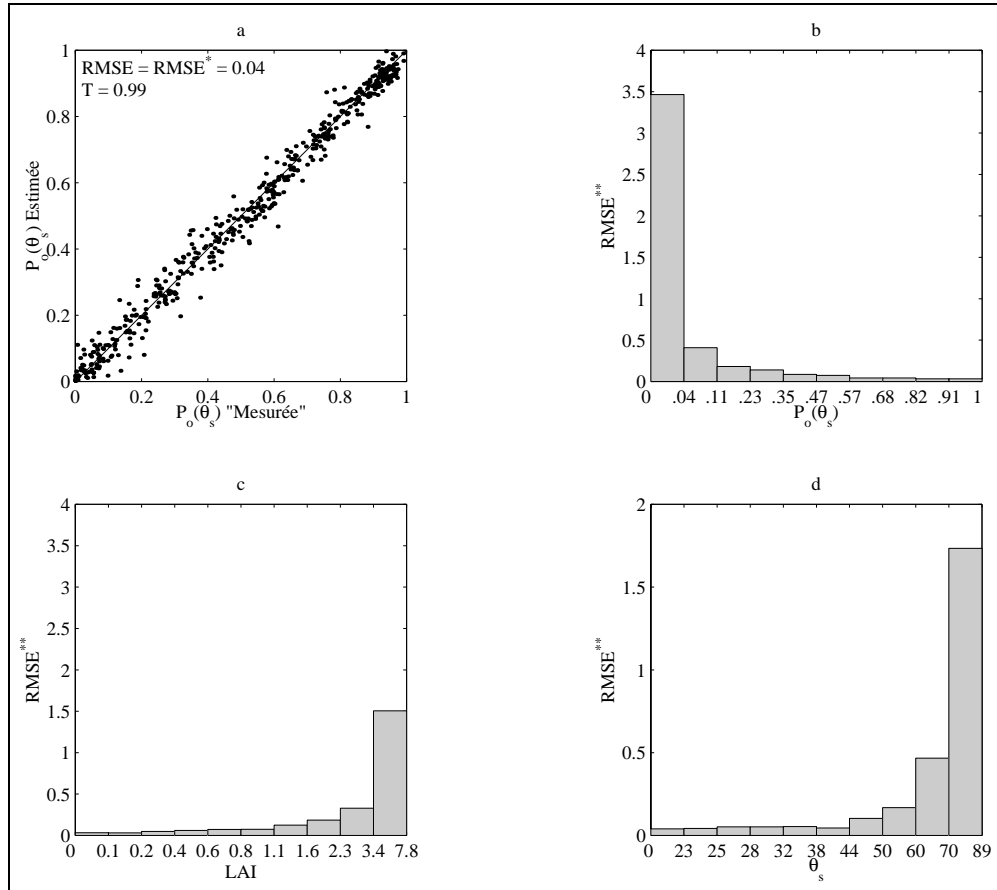


Figure 3.16 : Estimation de $P_o(\theta_s)$ (a et b). Analyse des résidus en fonction du LAI (c) et de l'angle zénithal solaire (d).

| | $P_o(0)$ | | $P_o(58^\circ)$ | | $P_o(\theta_s)$ | |
|------|----------|------|-----------------|------|-----------------|------|
| | $RMSE$ | T | $RMSE$ | T | $RMSE$ | T |
| RN | 0.04 | 0.98 | 0.04 | 0.98 | 0.04 | 0.99 |
| NDVI | 0.17 | 0.65 | 0.2 | 0.54 | 0.18 | 0.66 |

Tableau 3.4 : Estimation des fractions de trou monodirectionnelles dans des directions de visée particulières. Comparaison de la méthode basée sur le $NDVI$ et de celle basée sur l'utilisation de réseaux de neurones.

de transfert log-sigmoïde et une couche de 1 neurone à fonction de transfert linéaire. L'initialisation des poids et des biais du réseau n'a d'influence que sur le temps de convergence lors de l'apprentissage.

Les valeurs de $RMSE^4$ associées à l'estimation du LAI (figure 3.17) sont acceptables pour les indices foliaires faibles mais augmentent pour des LAI supérieurs à 3 (figure 3.17a et b). Ce phénomène de saturation, déjà noté pour la relation $NDVI/LAI$, est illustré par la figure 3.18 où nous avons tracé le spectre de réflectance au nadir pour des couverts de LAI croissants, possédant les mêmes caractéristiques biophysiques et structurelles par ailleurs (simulations SAIL). On observe bien que, sur le plateau du proche infrarouge, il est difficile de différencier le comportement des couverts de LAI élevé. Le fait que l'on ait choisi une distribution ISLSCP de l'indice foliaire pour la base d'apprentissage favorise également l'estimation des LAI faibles, largement plus représentés dans la base d'apprentissage. Malgré ce problème de saturation, on constate d'une part, que l'erreur relative d'estimation (différence entre LAI mesuré et LAI estimé, divisée par la valeur du LAI), reste entre -1 et 1 pour 85% des couverts (figure 3.17c) ; et, d'autre part que les résultats demeurent meilleurs qu'avec la méthode basée sur le $NDVI$, pour laquelle l'effet de saturation est plus important (tableau 3.5).

L'étape suivante consiste à étudier l'estimation de la fraction de trou monodirectionnelle. Dans la mesure, où les variables d'entrée du réseau de neurones sont calculées à partir de la séquence d'angles de visée $VEGETATION$, nous estimerons, dans un premier temps, la fraction de trou monodirectionnelle pour ces angles de visée. L'estimation directe de $P_o(\theta_v)$ n'est pas

⁴On ne considère pas ici le $RMSE^{**}$ du fait de la dynamique du LAI qui comprend des valeurs comprises entre 0 et 1 et des valeurs supérieures à 1.

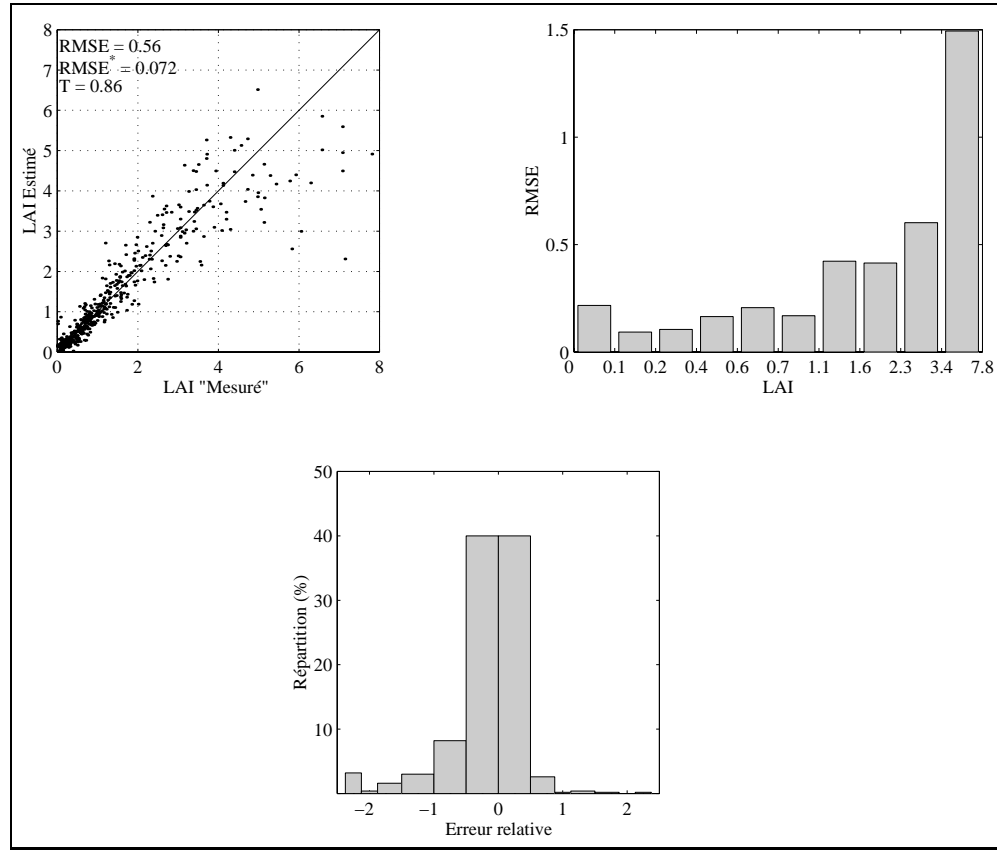


Figure 3.17 : Estimation du *LAI* (a). Analyse des résidus en fonction du *LAI* (b). Répartition des couverts en fonction de l'erreur relative (c).

| | $RMSE$ | T | $RMSE^{*a}$ |
|------|--------|------|-------------|
| RN | 0.56 | 0.86 | 0.07 |
| NDVI | 1.01 | 0.53 | 0.13 |

^avoir équation 3.13

Tableau 3.5 : Estimation du *LAI*. Comparaison de la méthode basée sur le *NDVI* et de celle basée sur l'utilisation de réseaux de neurones.

possible car les réseaux de neurones requièrent un nombre de sorties constant

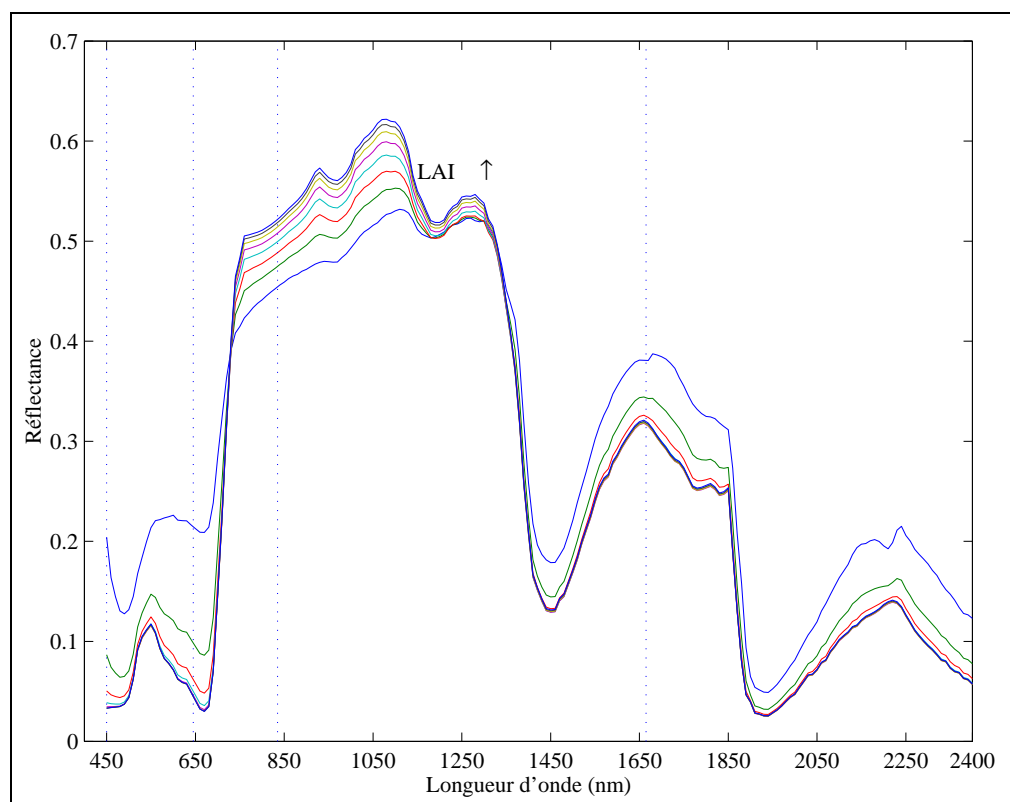


Figure 3.18 : Variation de la réflectance au nadir en fonction de la longueur d'onde pour des couverts de LAI croissants.

et le nombre d'angles de visée de *VEGETATION* varie selon la date et la latitude. Le seul moyen d'estimer la fraction de trou monodirectionnelle est donc de construire au préalable deux réseaux permettant d'estimer l'indice foliaire et l'angle moyen d'inclinaison des feuilles séparément, puis de calculer $P_o(\theta_v)$ d'après l'équation 2.9. Les paragraphes suivants sont donc consacrés à l'estimation de $\bar{\theta}_f$, puis de $P_o(\theta_v)$.

3.4.3 Estimation de l'angle moyen d'inclinaison des feuilles $\bar{\theta}_f$

Le réseau que nous avons construit possède la même architecture que ceux construits pour l'estimation des autres variables. La figure 3.19 montre que l'angle moyen d'inclinaison des feuilles, $\bar{\theta}_f$, est relativement bien estimé avec un *RMSE* relatif de 0.11 et un *T* de 0.85. Contrairement au *LAI*, on observe une dispersion qui dépend peu de la valeur de $\bar{\theta}_f$. Toutefois, les faibles valeurs d'angle moyen d'inclinaison des feuilles sont moins bien estimées. L'histogramme de répartition du *RMSE*** en fonction du *LAI* montre qu'il est plus difficile d'estimer $\bar{\theta}_f$ pour les faibles valeurs de *LAI*. En effet, il est plus délicat d'évaluer l'angle moyen d'inclinaison des feuilles lorsque le pourcentage de couverture du sol est faible et qu'il y a peu de feuillage.

3.4.4 Estimation des fractions de trou mono et bi-directionnelles

$$P_o(\theta_v), P_o(\theta_v, \theta_s)$$

Estimation de la fraction de trou monodirectionnelle

Dans ce paragraphe, nous considérons seulement la fonction de distribution de la fraction de trou monodirectionnelle dans les directions de visée correspondant à la séquence d'acquisition de *VEGETATION* utilisée pour évaluer les variables d'entrée du réseau de neurones. Comme nous l'avons présenté plus haut, nous avons, dans un premier temps, estimé le *LAI* et l'angle moyen d'inclinaison des feuilles à l'aide des deux réseaux de neurones précédents, puis calculé la fraction de trou monodirectionnelle. Le tableau 3.6 résume les résultats que nous avons obtenus pour les trois variables biophysiques : la précision d'estimation de $P_o(\theta_v)$ dans les angles de visée *VEGETATION* n'est pas très satisfaisante car la dispersion autour de la droite (1:1) n'est pas négligeable (figure 3.20).

Compte tenu que la précision d'estimation de la fraction de trou au nadir est satisfaisante, nous avons déterminé la variation directionnelle de la fraction

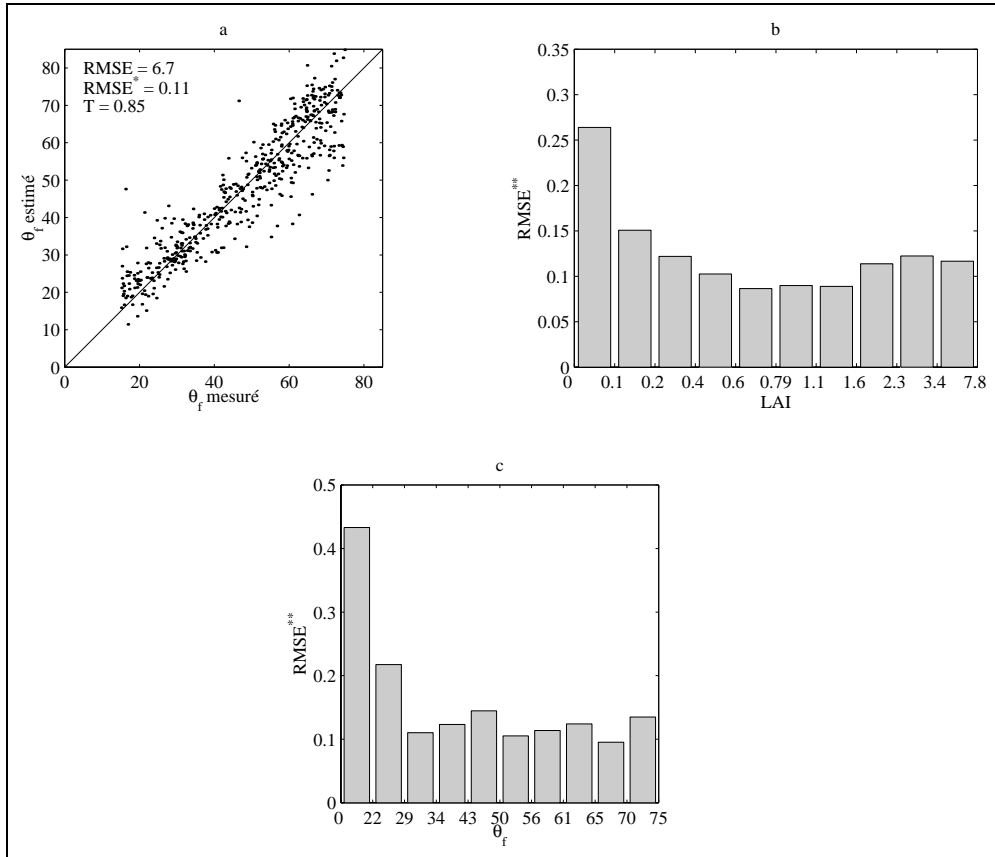


Figure 3.19 : Estimation de l'angle moyen d'inclinaison des feuilles (a). Analyse des résidus en fonction du LAI (b) et de $\bar{\theta}_f$ (c).

| LAI | | $\bar{\theta}_f$ | | $P_o(\theta_v)$ | |
|-------|----------|------------------|----------|-----------------|----------|
| T | $RMSE^*$ | T | $RMSE^*$ | T | $RMSE^*$ |
| 0.86 | 0.07 | 0.85 | 0.11 | 0.92 | 0.09 |

Tableau 3.6 : Estimation de la distribution de la fraction de trou monodirectionnelle dans les angles de visée $VEGETATION$.

de trou autour du nadir, plutôt que la fraction de trou monodirectionnelle elle-même. On observe également une très forte dispersion des points autour de

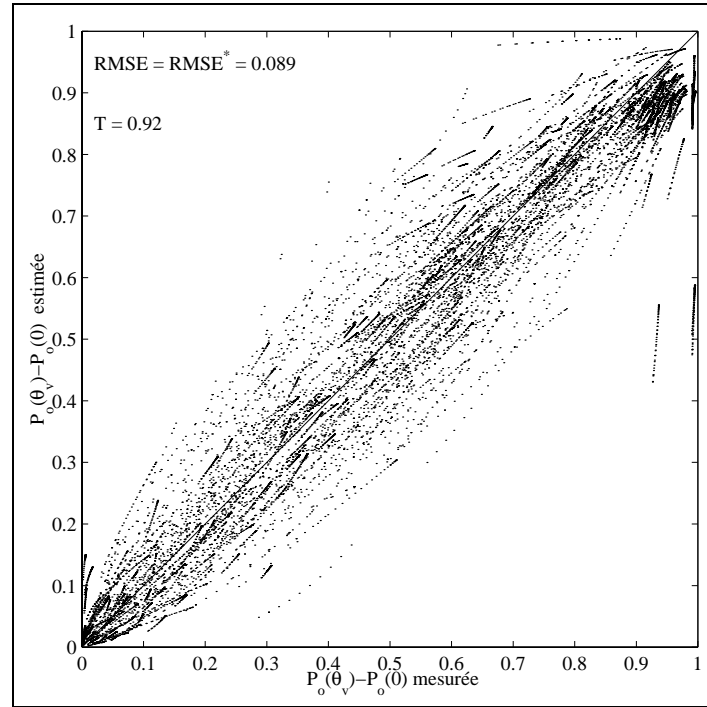


Figure 3.20 : Estimation de la fraction de trou monodirectionnelle.

la droite (1:1) (figure 3.21). Ces résultats, assez décevants, nous amènent à renoncer à extrapoler l'estimation de la distribution de la fraction de trou monodirectionnelle dans une quadrature de Gauss.

Estimation de la fraction de trou bidirectionnelle

Compte tenu que la fraction de trou monodirectionnelle est mal estimée, il n'est pas envisageable de construire un réseau de neurones pour l'estimation de la fraction de trou bidirectionnelle. De plus, $P_o(\theta_v, \theta_s)$ dépend non seulement du LAI et de l'angle moyen d'inclinaison des feuilles, mais aussi du paramètre de hot spot (eq 2.10). Or, sur les 1500 cas que nous avons simulés, nous avons constaté que seules 5% des $BRDF$ ont été échantillonnées dans un plan proche ($\pm 5^\circ$) du plan principal (figure 3.22). On peut également noter que les canopées situées au-delà d'une latitude de 30° (nord ou sud) ne seront jamais observés dans le hot spot par *VEGETATION*. Par ailleurs, les variables

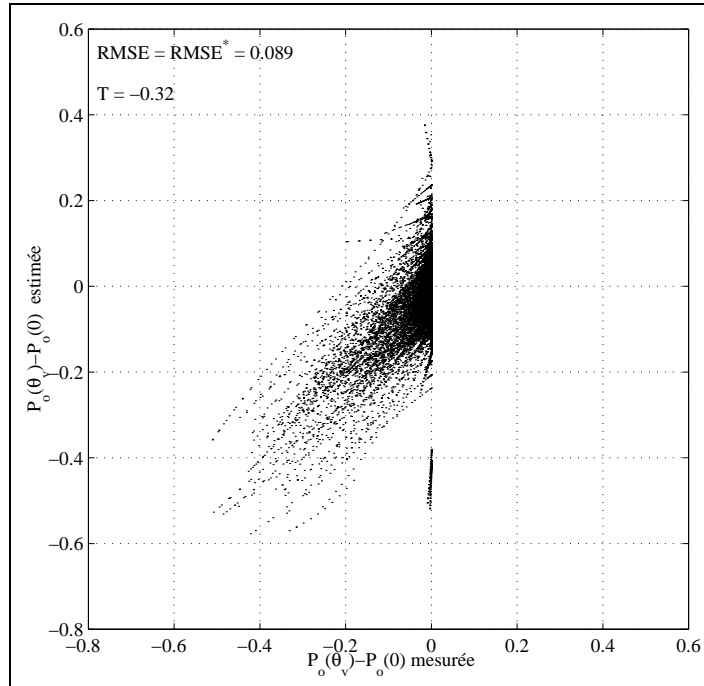


Figure 3.21 : Estimation de la variation de la fraction de trou monodirectionnelle autour du nadir.

d'entrée des réseaux de neurones ne sont pas représentatives du hot spot et sont obtenues par inversion du modèle de Walthall qui ne tient pas compte explicitement de ce phénomène.

3.4.5 Estimation de la fraction de rayonnement photosynthétiquement actif absorbé par le couvert $fAPAR$

Le $fAPAR$ étant très fortement corrélé (coefficient de corrélation de -0.99) avec la fraction de trou dans la direction solaire ($=fAPAR$ instantané), nous pouvons nous attendre à obtenir des résultats similaires (figure 3.23). Le réseau de neurones possède donc la même architecture (4 neurones à fonction de transfert log-sigmoïde et un neurone à fonction linéaire). La précision d'estimation est tout à fait correcte. Le résultat dépend peu de la position du soleil lors de l'acquisition de la séquence *VEGETATION*. Le réseau de neurones est, quoi qu'il en soit, plus performant que la méthode basée sur le $NDVI$ qui donne

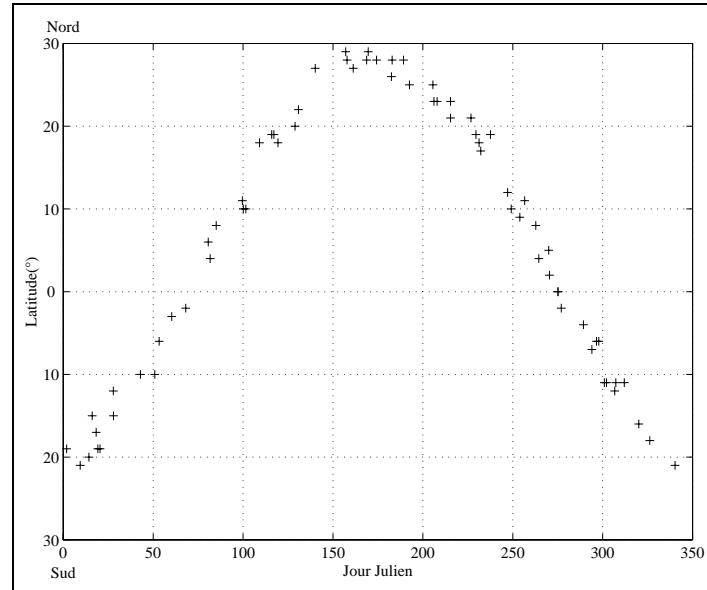


Figure 3.22 : Jour et latitude des couverts dont la $BRDF$ est échantillonnée dans le plan principal $\pm 5^\circ$.

un T de 0.55 au lieu de 0.98.

3.4.6 Estimation de la chlorophylle intégrée au niveau du couvert $LAI.C_{ab}$

Pour l'estimation de la chlorophylle intégrée au niveau du couvert, nous avons dû modifier la structure de notre réseau de neurones. En effet, cette variable n'intervient pas de façon directe dans le transfert radiatif. La relation "réflectance/ $LAI.C_{ab}$ " est donc plus complexe et nous avons dû rajouter une couche de neurones supplémentaire. Le réseau comporte ainsi 3 couches, 2 à fonction de transfert log-sigmoïde (4 et 3 neurones) et la dernière à fonction de transfert linéaire (1 neurone). La convergence est plus longue et requiert environ 2000 itérations. La figure 3.24 montre une forte dispersion pour les valeurs de $LAI.C_{ab}$ importantes. Les histogrammes de répartition des résidus en fonction de l'indice foliaire et du contenu en chlorophylle des feuilles⁵ mon-

⁵On ne considère pas ici le $RMSE^{**}$ du fait de la dynamique du $LAI.C_{ab}$ qui comprend des valeurs comprises entre 0 et 1 et des valeurs supérieures à 1.

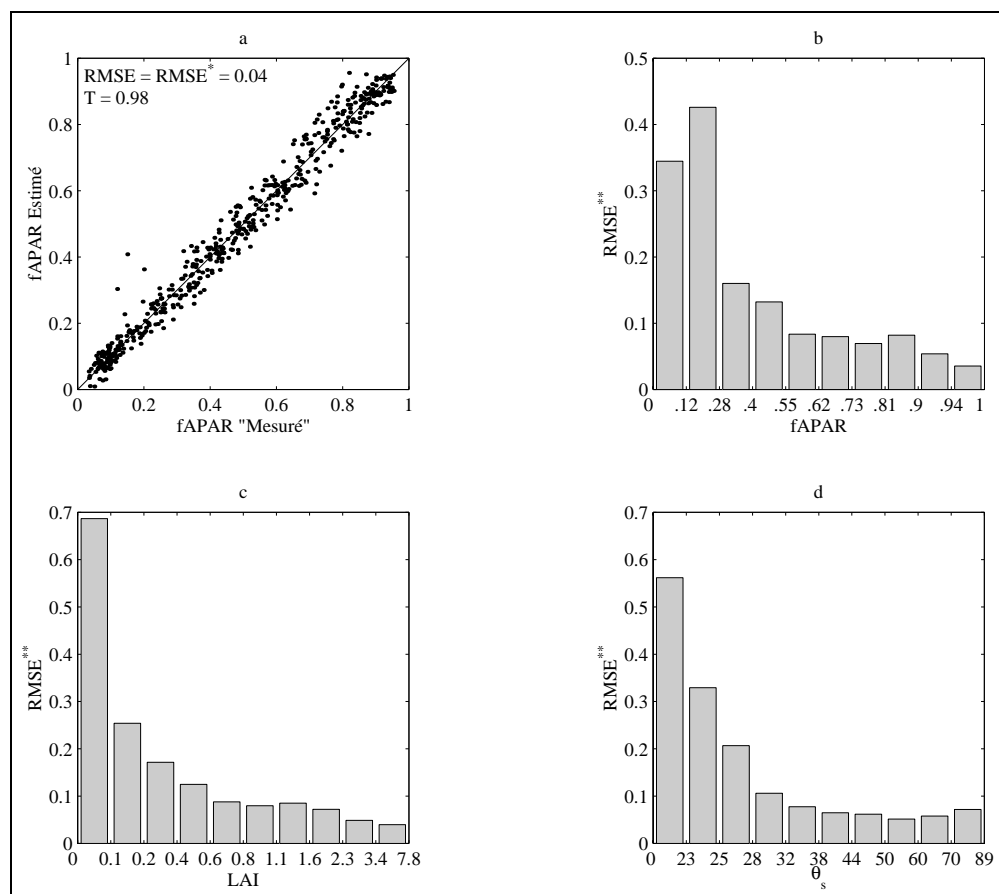


Figure 3.23 : Estimation du $fAPAR$ (a). Analyse des résidus en fonction du $fAPAR$ (b), du LAI (c) et de l'angle zénithal solaire (d).

trent que, comme pour le LAI , la dispersion est due à la saturation du signal. Peu de travaux ont porté sur l'estimation de cette variable. Yoder et Pettigrew (1995) ont estimé le contenu en chlorophylle intégré d'une forêt d'érables à partir de réflectances dans une sélection de longueurs d'onde. Ils obtiennent un coefficient de détermination R^2 entre 0.39 et 0.72 (qui dépend des longueurs d'onde sélectionnées). En comparaison de cette étude, nous obtenons un résultat acceptable sur des données simulées, avec un R^2 de 0.92.

3.4.7 Conclusion sur l'estimation de variables biophysiques du couvert à l'aide de réseaux de neurones.

Après avoir présenté notre méthode d'estimation de variables biophysiques du couvert à partir de données *VEGETATION*, nous avons évalué ses performances sur différentes variables biophysiques. Le tableau 3.7 résume les résultats que nous avons obtenus en terme de paramètres statistiques pour l'estimation de chacune des variables. Ces paramètres sont calculés pour deux nébulosités (0% et 80% de nuages).

L'influence du manque de données dû à la présence de nuages lors de l'acquisition de *VEGETATION* se fait plus ressentir pour le LAI , $\bar{\theta}_f$, $P_o(\theta_v)$ et $LAI.C_{ab}$. Bien que les valeurs de ρ_o et ρ_h soient stables par rapport au nombre de données disponibles pour l'inversion du modèle de Walthall, une légère imprécision commise sur celles-ci sera transférée à la variable estimée pour les réseaux de neurones les moins stables. C'est ce qui explique les moins bonnes performances pour LAI , $\bar{\theta}_f$, $P_o(\theta_v)$ et $LAI.C_{ab}$.

Le tableau 3.7 montre en outre, qu'il est préférable de construire un réseau de neurones pour estimer directement la fraction de trou dans une direction particulière, plutôt que de calculer celle-ci à partir des estimations du LAI et de l'angle moyen d'inclinaison des feuilles. Lorsque la fraction de trou monodirectionnelle est estimée à partir du LAI et de $\bar{\theta}_f$, une erreur commise sur le LAI peut être cumulée avec une erreur sur le $\bar{\theta}_f$, ce qui ne se produit pas si l'estimation est faite directement,

Il faut remarquer, de plus, que les précisions d'estimation obtenues avec les réseaux de neurones sont meilleures qu'avec la méthode basée sur le $NDVI$, lorsqu'il est possible de l'appliquer ($fAPAR$, $P_o(\theta_v = \theta_{vo})$, LAI).

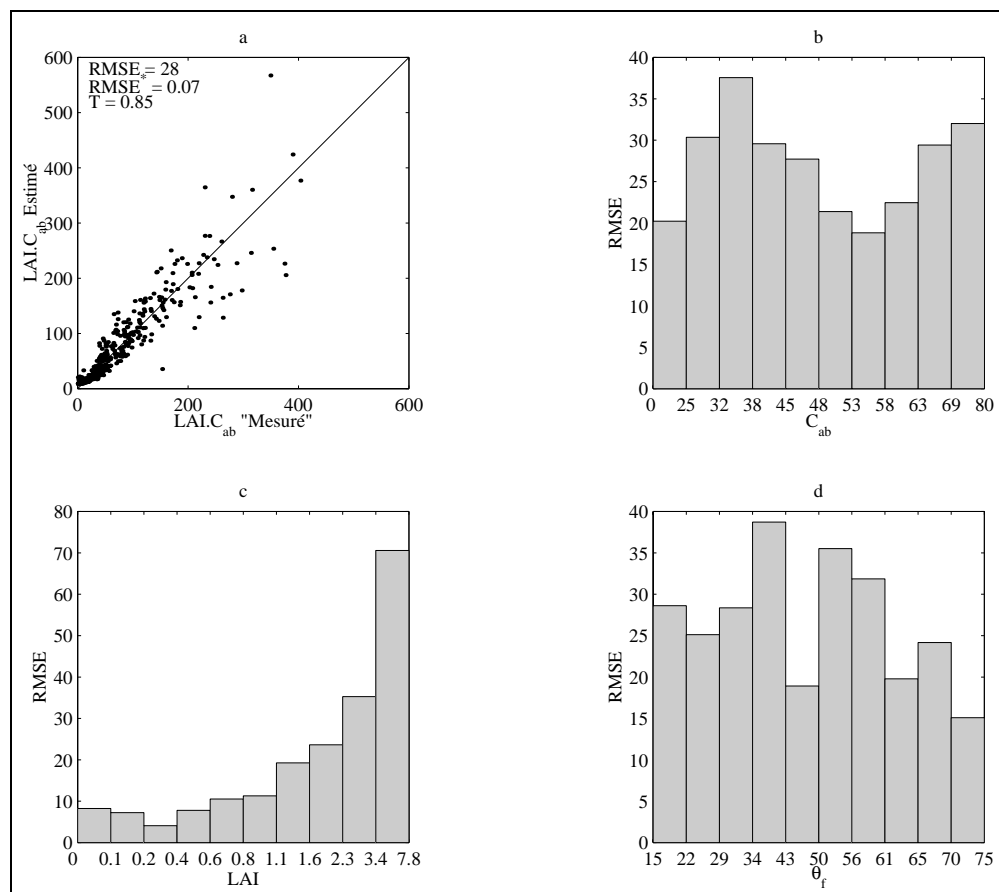


Figure 3.24 : Estimation de la chlorophylle intégrée au niveau du couvert (a). Analyse des résidus en fonction du contenu en chlorophylle des feuilles (b), du LAI (c) et de l'angle moyen d'inclinaison des feuilles (d).

| Variable | 0% nuages | | | 80% nuages | | |
|------------------|-------------------------|----------|-------------------------|------------|----------|------|
| | $RMSE$ | $RMSE^*$ | T | $RMSE$ | $RMSE^*$ | T |
| LAI | 0.55 | 0.07 | 0.86 | 0.64 | 0.08 | 0.81 |
| $\bar{\theta}_f$ | 6.7 | 0.11 | 0.85 | 8.4 | 0.14 | 0.76 |
| $P_o(\theta_v)$ | 0.09 | 0.09 | 0.92 | 0.11 | 0.11 | 0.87 |
| $P_o(0^\circ)$ | 0.04(0.08) ¹ | 0.04 | 0.98(0.91) ¹ | 0.044 | 0.051 | 0.97 |
| $P_o(58^\circ)$ | 0.04(0.08) ¹ | 0.04 | 0.98(0.92) ¹ | 0.054 | 0.054 | 0.97 |
| $P_o(\theta_s)$ | 0.04(0.11) ¹ | 0.04 | 0.99(0.88) ¹ | 0.044 | 0.044 | 0.98 |
| $fAPAR$ | 0.04 | 0.04 | 0.98 | 0.046 | 0.046 | 0.97 |
| $LAI.C_{ab}$ | 28 | 0.07 | 0.85 | 39 | 0.098 | 0.70 |

Tableau 3.7 : Performances d'estimation des variables biophysiques considérés dans la méthode basée sur les réseaux de neurones. ()¹ se rapporte aux paramètres statistiques de l'estimation de la fraction de trou dans des directions particulières à partir de l'estimation du LAI et de $\bar{\theta}_f$.

3.5 Sensibilité à la variabilité intrapixellaire

3.5.1 Le problème de l'hétérogénéité spatiale en télédétection

Les réflectances simulées dans notre base de données synthétiques correspondent, du fait de la modélisation de SAIL, à une couche de végétation homogène. Or, dans le cas d'images de télédétection, on est amené à raisonner en terme de pixel correspondant à une résolution spatiale donnée. L'homogénéité à considérer dépendra de la résolution spatiale du capteur :

- Haute résolution spatiale (*SPOT* -20m- ou *LANDSAT/TM* -120m-) : le problème d'hétérogénéité se situe au niveau de la parcelle : on observe la même culture mais celle-ci présente des aspects non uniformes, dus au labour, au semis, aux traitements azotés, etc ... ;
- Basse résolution spatiale (*VEGETATION*, *NOAA/AVHRR*, *POLDER*, ...) : même si l'hétérogénéité est toujours présente au niveau de la parcelle, elle existe aussi à une échelle nettement supérieure : la taille du pixel est assez grande pour que celui-ci contienne plusieurs parcelles mitoyennes qui, de plus, ne se rapportent pas forcément au même type de végétation. On parle alors de pixel composite.

Du fait de la résolution spatiale de *VEGETATION*, il est important d'étudier l'aspect composite des pixels qui va avoir le plus d'influence sur l'estimation des variables. C'est pourquoi nous avons construit une deuxième base de données permettant d'évaluer l'effet de l'homogénéité du pixel sur les performances de l'algorithme d'estimation des caractéristiques du couvert.

3.5.2 Construction d'une base de données "pixels composites"

Afin de maximiser l'hétérogénéité sur chaque pixel, nous avons considéré qu'il était composé de deux parties égales ayant chacune un niveau de *LAI* le plus éloigné possible. Pour ce faire, nous avons séparé la base de données initialement construite en deux parties : les couverts de *LAI* inférieur à 1.45 et ceux de *LAI* supérieur à 1.45. La valeur seuil de 1.45 correspond à la valeur moyenne de *LAI* sur l'ensemble des 1500 couverts simulés (tableau 2.5).

La nouvelle base de pixels composites a alors été simulée en 3 étapes (figure 3.25) :

1. Sélection des deux parties du pixel par tirage aléatoire dans chacun des deux sous-jeux de variables d'entrée du modèle de réflectance.
2. Simulation de la réflectance et des variables biophysiques associées de chacun des couverts pour une même date.
3. Calcul de la réflectance ($\rho = 1/2(\rho_1 + \rho_2)$) et des variables biophysiques correspondantes, $P_o(0)$, $fAPAR$, $LAI \dots$, ($varbio = 1/2(varbio_1 + varbio_2)$) du pixel hétérogène.

3.5.3 Estimation des variables biophysiques de "pixels composites"

Suite à l'étude que nous avons effectuée sur l'estimation des variables biophysiques dans les paragraphes précédents, nous ne considérerons que deux variables pour lesquelles les performances des réseaux sont satisfaisantes : la fraction de trou au nadir ainsi que l'indice foliaire. Les comportements des autres variables biophysiques envisagées seront déduits d'analogies notées entre variables : $P_o(0) \approx P_o(58^\circ) \approx P_o(\theta_s) \approx fAPAR$ et $LAI \approx LAI.C_{ab}$.

Les réseaux de neurones ont été testés sur 3 jeux de données :

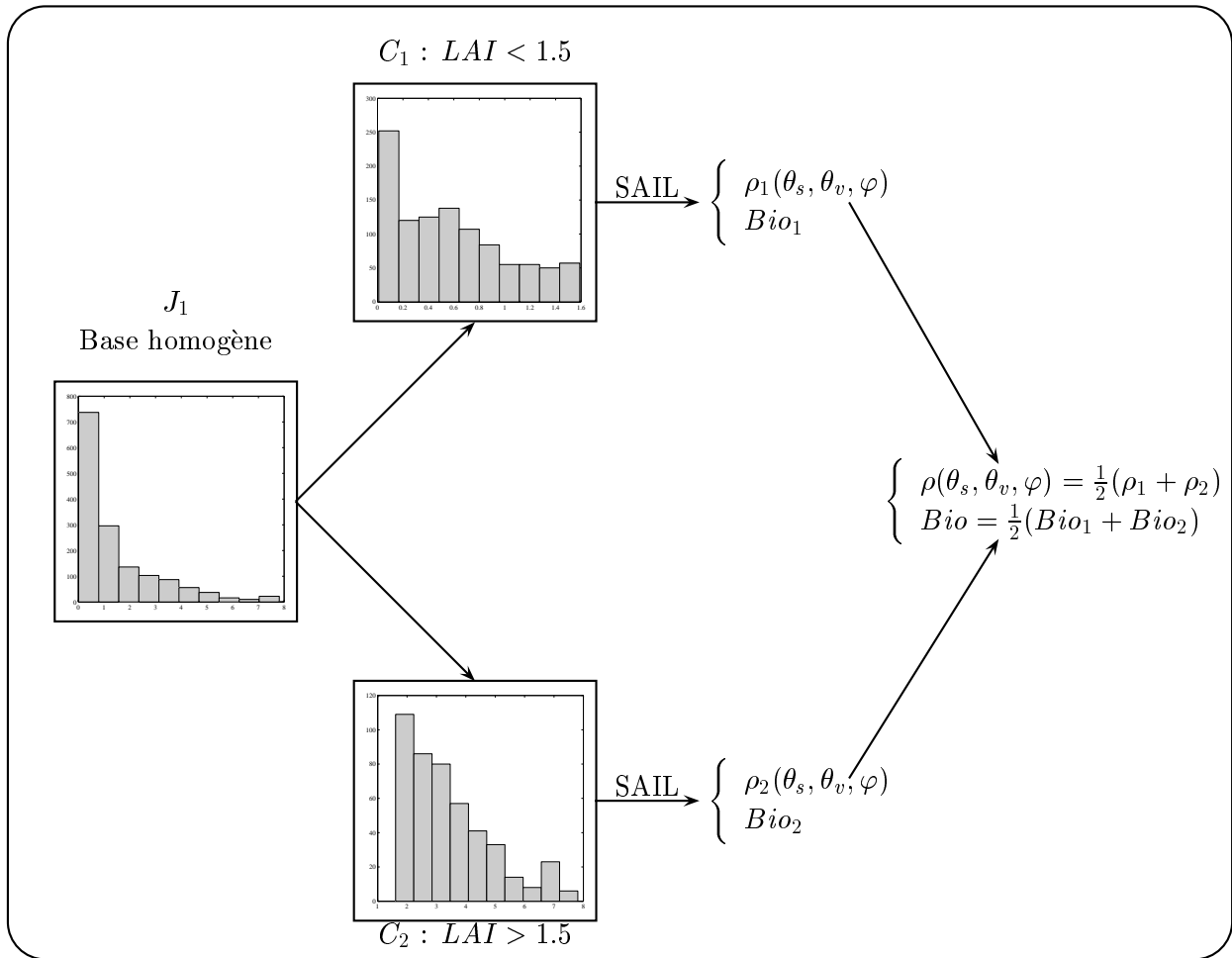


Figure 3.25 : Construction de la base de pixels composites.

J_1 : jeu de validation de la base de données initiale établie dans le chapitre 2 (500 cas);

J_2 : ensemble des couverts C_1 et C_2 de la base J_1 ayant servi à produire des pixels mixtes (1000 cas);

J_3 : composé de pixels mixtes (500 cas);

Estimation de la fraction de trou au nadir

Nous avons tracé la fraction de trou au nadir estimée par le réseau de neurones appris sur pixel homogène (paragraphe 3.3.2) en fonction de $P_o(0)$ “mesurée” de chacun des jeux de données que nous avons construits (figure 3.26). Les résultats sont similaires pour les trois jeux de données : les valeurs de $RMSE$ obtenues restent acceptables (≤ 0.08). Les différences que l’on observe entre les 3 jeux proviennent essentiellement des différences de distribution de la fraction de trou au nadir dans chacun d’eux. C’est la raison pour laquelle le T diminue de façon importante pour le jeu J_3 . En effet, d’une part, le réseau de neurones est plus performant pour les fractions de trou élevées qui sont bien représentées dans les jeux J_1 et J_2 alors qu’elles sont inexistantes dans le jeu J_3 . D’autre part la valeur de T , de par son expression, dépend de la dynamique de la fraction de trou sur l’ensemble des données. On peut donc conclure que l’estimation de la fraction de trou au nadir est satisfaisante, même sur les pixels mixtes.

Pour confirmer ces conclusions, nous avons simulé un jeu de données comprenant 5 couverts de LAI respectifs de 8, 4, 2, 1 et 0.5 et un sol nu. Les autres variables d’entrée de SAIL correspondent à un couvert moyen (voir le tableau 2.5 dans le chapitre 2). La réflectance et les variables biophysiques associées sont ensuite calculées pour un pixel hétérogène constitué pour moitié d’un des 5 couverts homogènes précédents, et pour l’autre moitié du sol nu. Le tableau 3.8 montre que les valeurs estimées de $P_o(0)$ sont très précises pour l’ensemble des simulations (à 0.04 près au maximum). Ce résultat confirme donc que l’hétérogénéité du pixel n’a que très peu d’incidence sur l’estimation de la fraction de trou au nadir.

Estimation de l’indice foliaire

La figure 3.27 donne le LAI estimé en fonction du LAI mesuré pour les trois jeux de données. Compte tenu de la dynamique importante du LAI (comparativement à la fraction de trou au nadir) sur l’ensemble des jeux de données, il est difficile de comparer, en termes de paramètres statistiques, les performances du réseau de neurones. Nous avons donc calculé les LAI minimal (0.8) et maximal (4.4) sur le jeu J_3 (plus basse dynamique), et sélectionné les données de J_1 et J_2 compris dans cette dynamique. Les résultats (tableau 3.9) montrent que l’estimation se détériore pour J_2 et J_3 . Toutefois, sur J_2 , on

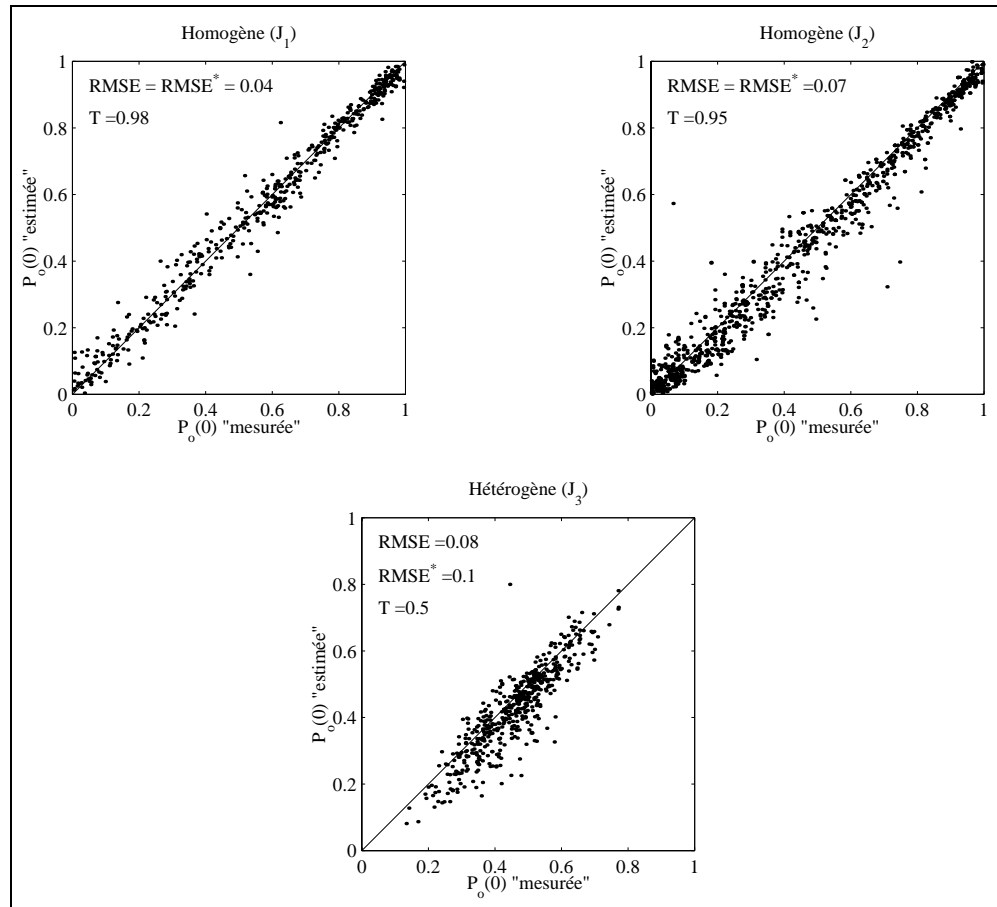


Figure 3.26 : Pixels mixtes : comparaison de l'estimation de la fraction de trou au nadir sur les trois bases de données.

| | Couvert | | Sol nu | | Pixel Mixte | |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| LAI | $P_o(0)_{mes}$ | $P_o(0)_{est}$ | $P_o(0)_{mes}$ | $P_o(0)_{est}$ | $P_o(0)_{mes}$ | $P_o(0)_{est}$ |
| 8 | 0.005 | 0.014 | 1 | 1.03 | 0.5 | 0.51 |
| 4 | 0.07 | 0.064 | 1 | 1.03 | 0.53 | 0.54 |
| 2 | 0.26 | 0.24 | 1 | 1.03 | 0.63 | 0.62 |
| 1 | 0.51 | 0.51 | 1 | 1.03 | 0.76 | 0.77 |
| 0.5 | 0.72 | 0.68 | 1 | 1.03 | 0.86 | 0.89 |

Tableau 3.8 : Effet de l'hétérogénéité sur l'estimation de la fraction de trou au nadir.

| | $RMSE$ | T | R^2 |
|-------|--------|-------|-------|
| J_1 | 0.29 | 0.74 | 0.79 |
| J_2 | 1.29 | -0.50 | 0.60 |
| J_3 | 0.98 | -0.70 | 0.15 |

Tableau 3.9 : Comparaison des performances d'estimation du réseau de neurones pour l'estimation de l'indice foliaire sur les trois jeux de données. Dans chaque cas, les paramètres statistiques sont calculés pour les pixels dont le LAI est compris dans la dynamique de J_3 .

observe un biais (tendance à surestimer) caractérisé par un T faible mais un R^2 relativement fort. Au contraire, la dispersion est beaucoup plus importante sur J_3 pour lequel le T comme le R^2 sont faibles. La répartition du LAI est encore différente entre les trois jeux de données (figure 3.28). C'est pourquoi, on observe une différence de $RMSE$ assez grande entre J_1 et J_2 . En effet, J_1 contient plus de couverts de LAI faibles, pour lesquels le réseau de neurones est plus performant.

De même que pour la fraction de trou au nadir, pour confirmer que l'hétérogénéité des pixels pose problème pour l'estimation du LAI , nous avons estimé l'indice foliaire sur le jeu de données comprenant 5 couverts et un sol nu (tableau 3.10). On retrouve bien l'effet de saturation pour le couvert de LAI égal à 8 (estimé à 4.5) sur le pixel homogène. Le LAI estimé à 0.15 pour le sol nu provient du fait que la base d'apprentissage J_1 ne comporte pas de sol totalement nu (LAI minimum de 0.01). Les résultats obtenus sur

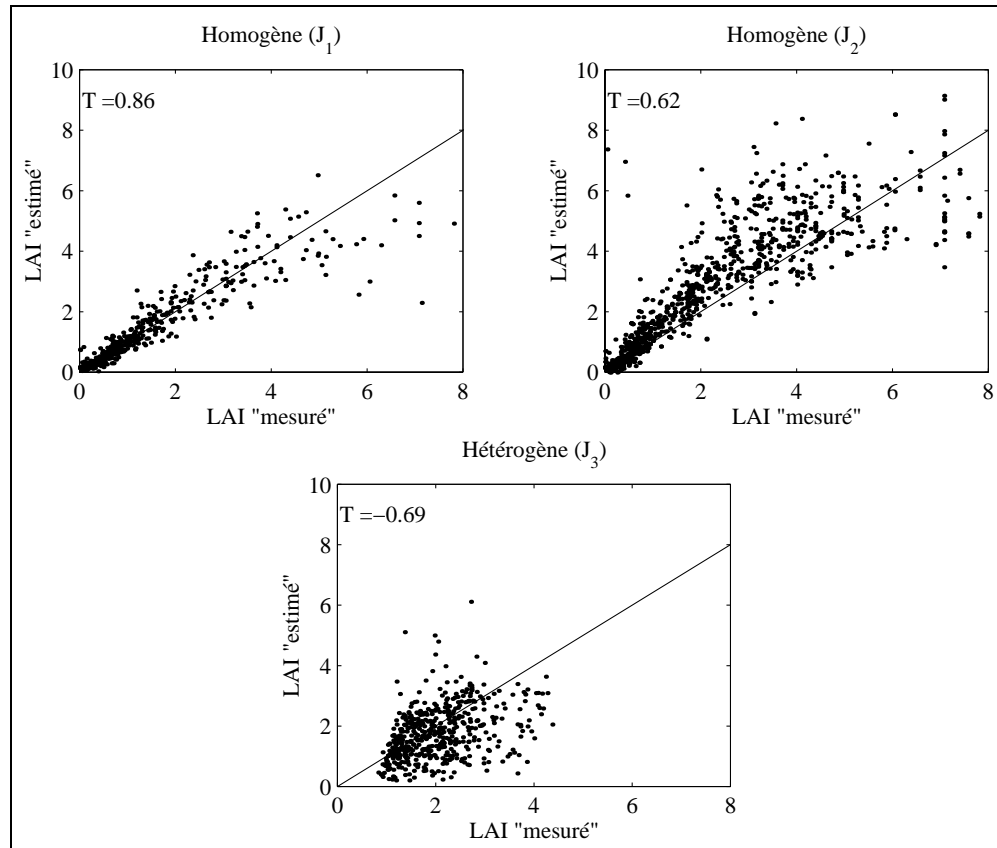


Figure 3.27 : Pixels mixtes : comparaison de l'estimation de l'indice foliaire sur les trois bases de données.

les pixels mixtes ne sont pas satisfaisants : l'estimation du LAI sur un pixel homogène de LAI 4 est assez bonne (3.57) alors qu'elle est très mauvaise pour un pixel mixte de même LAI (LAI estimé de 0.57). Il faut cependant noter que lorsque le niveau d'hétérogénéité du pixel est faible (" LAI de 0.5/sol nu", " LAI de 2/ LAI de 3"), la précision d'estimation reste assez bonne. Les mauvais résultats obtenus sur les pixels mixtes corroborent donc nos premières conclusions, à savoir que l'estimation du LAI sur des pixels composites n'est pas satisfaisante.

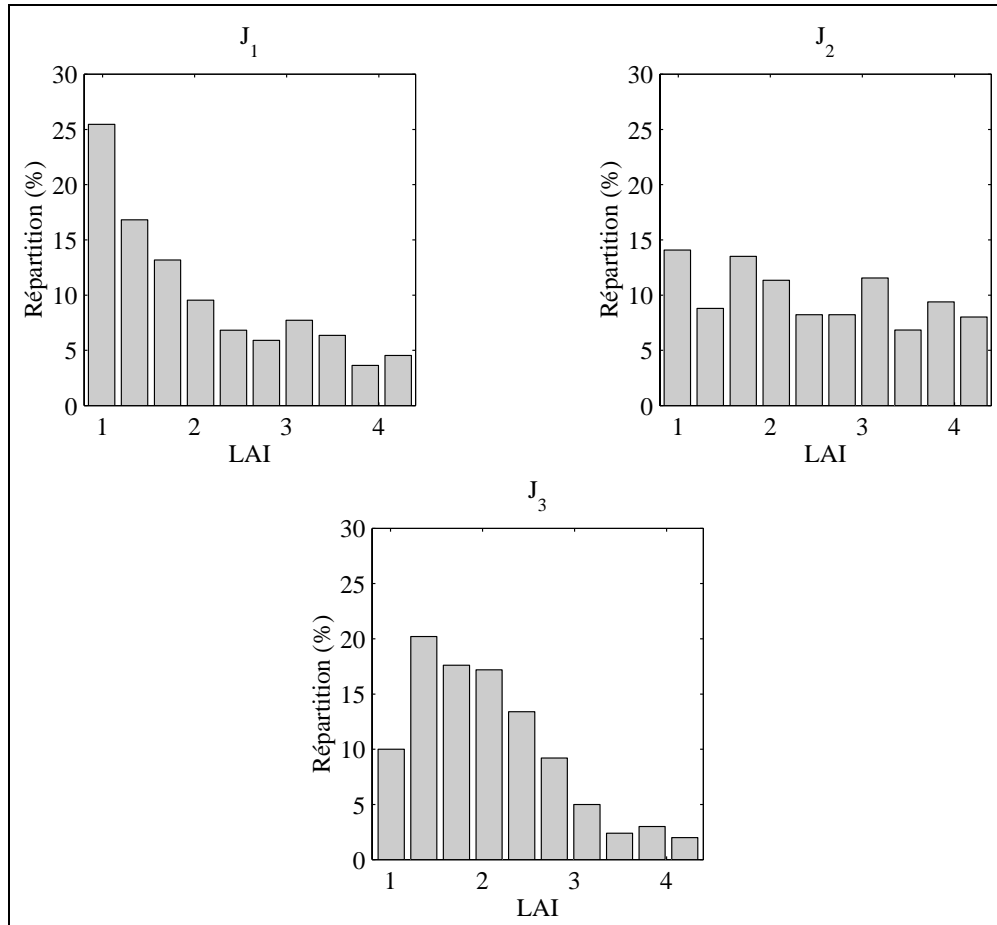


Figure 3.28 : Répartition du LAI sur les trois bases de données, pour les pixels entrant dans la dynamique de J_3 .

| Pixel Homogène 1 | | Pixel Homogène 2 | | Pixel Mixte | |
|------------------|-------------|------------------|-------------|-------------|-------------|
| LAI_{mes} | LAI_{est} | LAI_{mes} | LAI_{est} | LAI_{mes} | LAI_{est} |
| 8 | 4.47 | 0 | 0.15 | 4 | 0.67 |
| 4 | 3.57 | 0 | 0.15 | 2 | 0.57 |
| 2 | 2.31 | 0 | 0.15 | 1 | 0.41 |
| 2 | 2.31 | 3 | 3.09 | 2.5 | 2.69 |
| 1 | 1.03 | 0 | 0.15 | 0.5 | 0.28 |
| 0.5 | 0.36 | 0 | 0.15 | 0.25 | 0.21 |

Tableau 3.10 : Effet de l'hétérogénéité sur l'estimation de l'indice foliaire.

3.5.4 Conclusion sur la sensibilité à la variabilité intrapixelaire

Dans cette partie de l'étude, nous avons mis en évidence l'effet de la variabilité intrapixelaire sur l'estimation de la fraction de trou au nadir et du LAI . Nous avons montré que, pour $P_o(0)$, l'hétérogénéité du pixel influence peu les performances d'estimation du réseau de neurones, ce qui n'est pas le cas du LAI .



Estimation de variables biophysiques des couverts végétaux Calage de réseau de neurones

-

Conclusions



Ce chapitre constitue la phase de développement de l'algorithme d'estimation des variables biophysiques des couverts végétaux à partir de données de réflectance acquises lors d'un cycle orbital du capteur VEGETATION.

Dans une première partie, nous avons caractérisé le réseau de neurones : règle d'apprentissage, architecture et variables d'entrée. Ce dernier point est en effet critique car le nombre de données de réflectance varie selon la latitude du point observé et la nébulosité lors de la date d'acquisition. Aussi, nous avons étudié la possibilité d'utiliser des modèles linéaires de BRDF pour normaliser ces données. Au final, les variables d'entrée du réseau sont les réflectances au nadir et hémisphérique pour les longueurs d'onde du capteur, ainsi que la position zénithale du soleil.

Dans une deuxième partie, nous avons montré l'intérêt d'employer des réseaux de neurones plutôt qu'une méthode plus classique basée sur l'utilisation d'un indice de végétation. La suite de l'étude a alors consisté à déterminer les variables biophysiques qui sont les mieux estimées par les réseaux de neurones. Ainsi, les performances du réseau sont très satisfaisantes pour l'estimation de la fraction de trou dans des directions de visée particulières et la fraction de rayonnement photosynthétiquement actif absorbé par le couvert. L'indice foliaire, comme la chlorophylle intégrée au niveau du couvert, posent plus de problèmes, essentiellement à cause d'un phénomène de saturation pour les fortes valeurs d'indice foliaire. Toutefois, les résultats obtenus pour le LAI

sont meilleurs qu'en utilisant l'indice de végétation. Enfin, la distribution de la fraction de trou monodirectionnelle (resp. bidirectionnelle) en fonction des directions de visée (resp. et de l'angle solaire) est mal estimée par la méthode des réseaux de neurones, essentiellement à cause du manque d'information directionnelle dans les variables d'entrée du réseau.

Enfin, dans la troisième partie de ce chapitre, dans le but de nous rapprocher des conditions réelles (résolution spatiale de 1.15 km pour VEGETATION), nous avons évalué les performances du réseau sur des pixels mixtes. L'étude montre que l'estimation de la fraction de trou au nadir n'est pas affectée, contrairement à celle du LAI.

À la lumière des résultats précédents, les variables retenues pour le suivi de la végétation sont la fraction de trou dans des directions de visée particulières ($P_o(0)$, $P_o(58^\circ)$, $P_o(\theta_s)$) et le fAPAR. En effet, ce sont les variables les mieux estimées et les moins sensibles à la variabilité intrapixelaire induite par les caractéristiques du capteur.

À ce stade de l'étude, nous avons travaillé uniquement avec des données simulées, ce qui nous a amené à faire parfois de fortes hypothèses :

- les couverts qui composent les pixels (mixtes ou non) sont supposés homogènes dans la modélisation de SAIL, ce qui est loin d'être le cas sur les cultures en rang, les savanes ou encore les forêts ;
- les réflectances utilisées pour déterminer les variables d'entrée du réseau de neurones ont été évaluées au niveau du couvert. Aussi, nous n'avons tenu compte ni du bruit radiométrique induit par le capteur, ni du bruit induit par les corrections atmosphériques ;
- l'acquisition des réflectances se fait durant un cycle de 26 jours pendant lequel on suppose que l'état de la végétation reste inchangé, hypothèse forte dans le cas de beaucoup de couverts (cultures saisonnières, forêts d'arbres à feuilles caduques, ...).

Il est donc nécessaire de valider cet algorithme sur des données expérimentales afin d'évaluer l'impact de certaines de ces hypothèses sur les performances des réseaux de neurones.

Chapitre 4

Validation de l'algorithme de suivi des couverts végétaux.

L'acquisition de données dans des conditions réelles est soumise à de nombreuses perturbations dues à l'atmosphère, aux caractéristiques intrinsèques des capteurs, au problème de coregistration des images d'une date à l'autre, etc ... L'algorithme que nous avons développé a été conçu à partir de données simulées par un modèle qui impose des hypothèses simplificatrices. C'est pourquoi il est nécessaire de le valider avec des données réelles pour évaluer ses qualités comme ses faiblesses, et envisager ainsi des améliorations pour le rendre plus performant.

Dans la mesure où *VEGETATION* n'a été lancé qu'au mois de mars 1998, il n'était pas envisageable, pour des raisons de délai essentiellement, de mener une expérimentation à partir de cette date afin d'acquérir un jeu de données pour la validation de l'algorithme. C'est pourquoi les données expérimentales utilisées dans cette étude proviennent d'autres capteurs. Pour cette phase de validation, nous avons considéré deux expérimentations : la campagne Grignon 1990 (Malthus, 1990) et la campagne Alpilles/ReSeDA¹ (Prévoit *et al.*, 1998).

4.1 Validation sur la campagne Grignon 1990

Cette expérimentation a été réalisée dans des conditions assez différentes du contexte dans lequel nous nous sommes placés. Toutefois, ce jeu de données

¹Remote Sensing Data Assimilation

a été utilisé par Baret *et al.*(1995) pour élaborer et tester des réseaux de neurones en comparaison de la méthode basée sur le *NDVI* pour estimer la fraction de trou au nadir. La comparaison avec les résultats qu'ils ont obtenus est donc intéressante.

La campagne de mesures s'est déroulée au printemps 1990 sur une parcelle de Betterave irriguée à la station de Bioclimatologie de l'INRA de Grignon (latitude 48°51' nord, longitude 1°58' est). Cette expérimentation avait quatre objectifs principaux :

1. Rechercher de nouvelles méthodes pour estimer les variables biophysiques de la végétation (pourcentage de couverture du sol, indice foliaire et biomasse) en présence de perturbations (couleur du sol ou manque de chlorophylle pour la végétation).
2. Acquérir un jeu de données pour tester des modèles de réflectance de couvert et faire des analyses de sensibilité aux variables biophysiques, comme aux paramètres environnementaux.
3. Affiner les techniques de télédétection pour la mesure des propriétés optiques des couverts végétaux.
4. Montrer les avantages que peut présenter la haute résolution spectrale pour l'analyse de la végétation en télédétection.

4.1.1 Composition du jeu de données

Agencement de la parcelle

Afin d'obtenir une variabilité satisfaisante du pourcentage de couverture, la culture a été semée à deux densités :

- une densité normale : un espace inter-rangs de 50 cm avec une distance inter-plants de 25 cm ;
- une densité double : un espace inter-rangs de 25 cm avec une distance inter-plants de 12,5 cm. En outre, pour augmenter encore la variabilité au sein de la parcelle, un éclaircissage supplémentaire a été effectué en retirant certains plants (tableau 4.1).

La parcelle a été divisée en trois : une partie "saine" qui n'a pas reçu de traitement, la deuxième partie a reçu un traitement herbicide léger, et la

troisième un traitement herbicide conséquent. Ces différents traitements ont permis de réduire la teneur en chlorophylle des feuilles du couvert végétal, et donc, de modifier la réflectance du couvert. En définitive, la parcelle est divisée en 24 placettes, c'est-à-dire en trois parties comprenant chacune 8 placettes de densités de semis différentes.

| Densité initiale | Espace inter-rangs (cm) | Espace inter-plants (cm) |
|-----------------------------|------------------------------------|-------------------------------------|
| Normale | 25 | 50 |
| Double | 12.5 | 25 |
| Double | 12.5 | 50 |
| Double | 37.5 | 50 |
| Double | 25 | 25 |
| Double | 25 | 50 |
| Double | 50 | 50 |
| Double | 12.5 | 75 |

Tableau 4.1 : Espace inter-rangs et inter-plants correspondant aux différentes configurations de semis.

Mesure des variables biophysiques

La fraction de trou au nadir

La fraction de trou au nadir a été estimée à partir de photographies digitalisées prises à la verticale de chaque placette. La mesure de $P_o(0)$ a été faite du même endroit que les mesures de réflectance correspondantes. La précision de la mesure est à 0.02.

L'indice foliaire

La surface foliaire de cinq plants sélectionnés au hasard a été d'abord mesurée pour chaque placette, en utilisant aussi une technique d'analyse d'images. L'indice foliaire a ensuite été calculé comme la surface moyenne des cinq plants multipliée par le nombre de plants dans chaque placette, et rapportée au m^2 . La précision de la mesure est de l'ordre de 10%.

L'angle moyen d'inclinaison des feuilles

L'inclinaison des feuilles a été mesurée *in situ* avec un inclinomètre électronique sur 16 placettes. L'angle d'inclinaison des feuilles a été mesuré sur chacune des feuilles de trois plants sélectionnés au hasard. La précision d'estimation est de l'ordre de 10° .

Le tableau 4.2 résume l'ensemble des mesures faites sur les 24 placettes de la parcelle. On a donc une gamme de variation assez large de la fraction de trou au nadir, comme du *LAI*. On distingue bien, par ailleurs, la différence de chlorophylle entre les 3 sous parcelles (de 8 placettes chacune) due au traitement herbicide : traitement herbicide "lourd" (placettes 1 à 8) et "léger" (placettes 9 à 16) qui ont le même angle moyen d'inclinaison des feuilles et les placettes saines (17 à 24) pour lesquelles l'angle augmente (feuilles plus érigées).

Mesures de réflectance

Les réflectances ont été mesurées grâce à cinq radiomètres *CIMEL* qui permettent d'obtenir la luminance réfléchie dans les trois bandes *SPOT* : vert (620-680 nm), rouge (500-590 nm), et proche infrarouge (790-890 nm). La conversion "luminance/réflectance" a été effectuée grâce à des mesures d'éclairement faites avec un sixième radiomètre placé au-dessus d'une surface de référence.

Les radiomètres, situés à 4.75 m au-dessus du sol, étaient positionnés de manière à mesurer la réflectance au nadir et à 45° au nord, sud, est et ouest. Les mesures ont été effectuées durant une même journée (19 juillet 1990), environ toutes les heures (12 positions solaires au total).

Jeu de mesures disponibles

Pour chacune des 24 placettes de la parcelle de betterave, nous disposons de 5 mesures de réflectance bidirectionnelle pour 12 positions solaires, ainsi que du *LAI*, de $\bar{\theta}_f$ et de $P_o(0)$:

$$\left\{ \begin{array}{l} (\rho(0, 0), \rho(45, 90), \rho(45, 180), \rho(45, 270), \rho(45, 0))_{(\theta_s, \varphi_s)} \\ P_o(0), LAI, \bar{\theta}_f \end{array} \right.$$

L'utilisation de ces mesures va permettre une première validation de nos résultats sur des données expérimentales. En effet, les conditions de mesure

| Placette | LAI | $\bar{\theta}_f$ | $P_o(0)$ | C_{ab} |
|----------|-------|------------------|----------|----------|
| 1 | 0.40 | 34.6 | 0.73 | 13.1 |
| 2 | 0.59 | 34.6 | 0.58 | 13.1 |
| 3 | 2.37 | 34.6 | 0.12 | 13.1 |
| 4 | 1.59 | 34.6 | 0.25 | 17.9 |
| 5 | 1.19 | 34.6 | 0.32 | 13.1 |
| 6 | 0.30 | 34.6 | 0.76 | 13.1 |
| 7 | 2.37 | 34.6 | 0.17 | 13.1 |
| 8 | 1.59 | 34.6 | 0.20 | 17.9 |
| 9 | 0.38 | 34.6 | 0.68 | 23.0 |
| 10 | 0.29 | 34.6 | 0.71 | 23.0 |
| 11 | 2.28 | 34.6 | 0.06 | 23.0 |
| 12 | 1.44 | 34.6 | 0.15 | 23.8 |
| 13 | 0.76 | 34.6 | 0.44 | 23.0 |
| 14 | 1.14 | 34.6 | 0.29 | 23.0 |
| 15 | 2.28 | 34.6 | 0.13 | 23.0 |
| 16 | 1.44 | 34.6 | 0.19 | 23.8 |
| 17 | 0.48 | 51.8 | 0.71 | 32.9 |
| 18 | 0.36 | 51.8 | 0.85 | 32.9 |
| 19 | 2.85 | 51.8 | 0.14 | 32.9 |
| 20 | 1.76 | 51.8 | 0.24 | 34.3 |
| 21 | 1.42 | 51.8 | 0.35 | 32.9 |
| 22 | 1.42 | 51.8 | 0.67 | 32.9 |
| 23 | 2.85 | 51.8 | 0.12 | 32.9 |
| 24 | 1.76 | 51.8 | 0.19 | 34.3 |

Tableau 4.2 : Variables biophysiques mesurées sur la campagne Grignon 90.

sont relativement différentes du cadre de notre étude. Nous avons ici à notre disposition les 3 bandes *SPOT* (rouge, vert, proche infrarouge) qui diffèrent des bandes spectrales de *VEGETATION*. Nous disposons d'une bande différente (vert au lieu de bleu) dans le visible mais pas de bande dans le moyen infrarouge. La bande du proche infrarouge est pratiquement la même pour les deux capteurs (780-890 nm pour *VEGETATION*). Cette bande est celle

qui présente le plus d'intérêt car la réflectance des feuilles est maximale (absorption des feuilles très faible dans ce domaine). Elle contient donc beaucoup d'information sur la végétation. De plus, lors de la campagne Grignon, la distribution des angles de visée est très différente d'une séquence d'acquisition de *VEGETATION* : on ne dispose que de 5 mesures à angle zénithal constant avec un angle azimutal variable. Nous avons donc dû, dans un premier temps, adapter notre réseau de neurones aux longueurs d'onde des radiomètres *CIMEL*.

4.1.2 Construction d'un réseau de neurones adapté aux longueurs d'onde *CIMEL*

Pour réaliser l'apprentissage du réseau de neurones, nous avons utilisé la base de données du chapitre 2. Nous avons effectué des simulations supplémentaires pour obtenir les réflectances dans les longueurs d'onde *CIMEL*. Les réflectances bidirectionnelles ont été calculées pour des séquences d'acquisition de *VEGETATION*, puis normalisées selon la procédure de prétraitement décrite au paragraphe 3.2. De même, la base de données a été divisée en trois parties égales, correspondant chacune à un jeu d'apprentissage, d'hyperspécialisation et de validation. La procédure d'estimation des variables biophysiques est celle décrite figure 3.9.

Compte tenu des conclusions du chapitre précédent et des mesures au sol dont nous disposons (fraction de trou au nadir, *LAI* et angle moyen d'inclinaison des feuilles), nous avons construit 3 réseaux à deux couches (4 neurones log-sigmoïde, 1 neurone linéaire) permettant d'estimer $P_o(0)$, $P_o(58^\circ)$ et $P_o(\theta_s)$. Les paramètres statistiques obtenus pour les trois variables sur le jeu de validation sont donnés dans le tableau 4.3. Le fait qu'on utilise des longueurs d'onde différentes de celles de *VEGETATION* n'a que peu d'influence sur les résultats. En effet, pour les deux capteurs, on dispose d'une longueur d'onde à la fois dans le rouge et le proche infrarouge qui sont les deux principaux domaines de longueurs d'onde permettant de différencier le sol de la végétation. Par conséquent, la précision d'estimation obtenue pour les trois fractions de trou est très satisfaisante. Comme pour le réseau de neurones construit pour *VEGETATION*, nous avons vérifié que l'initialisation des poids et des biais n'avait pas d'influence sur les performances d'estimation des variables (§3.3).

| | <i>RMSE</i> | <i>T</i> |
|-----------------|-------------|----------|
| $P_o(0)$ | 0.049 | 0.97 |
| $P_o(\theta_s)$ | 0.044 | 0.98 |
| $P_o(58^\circ)$ | 0.046 | 0.98 |

Tableau 4.3 : Paramètres statistiques pour l'estimation de la fraction de trou à partir de réflectances bidirectionnelles *VEGETATION* dans les longueurs d'onde CIMEL.

4.1.3 Validation de l'algorithme sur les données Grignon 1990

Le problème de l'échantillonnage directionnel

Nous avons montré dans le paragraphe précédent que l'utilisation des longueurs d'onde des radiomètres *CIMEL* ne change pratiquement pas les performances d'estimation des réseaux de neurones calés à partir des longueurs d'onde de *VEGETATION*. Les conditions expérimentales de la campagne Grignon 1990 présentent pourtant une différence importante au niveau de l'échantillonnage directionnel de la réflectance. On ne dispose en effet que de cinq mesures de réflectance effectuées dans les mêmes angles de visée pour chaque placette, pour douze positions solaires différentes. Ceci est pratiquement à l'opposé des conditions d'acquisition de *VEGETATION* pour lesquelles on a généralement un angle zénithal solaire quasiment constant et plusieurs angles de visée. Pour les 1500 séquences du jeu de données simulées, nous avons calculé ρ_o et ρ_h dans les trois longueurs d'onde à partir de la séquence d'acquisition de *VEGETATION* et pour les cinq angles de visée de l'expérimentation. Nous obtenons un écart quadratique moyen de 0.01 pour la réflectance au nadir et de 0.02 pour la réflectance hémisphérique. Le nombre de données disponibles n'est donc pas un facteur limitant pour l'estimation de variables biophysiques sur la campagne Grignon 1990. Ceci confirme le résultat obtenu précédemment sur la robustesse du modèle de Walthall.

Estimation de la fraction de trou au nadir

La figure 4.1 montre les résultats que nous avons obtenus sur l'ensemble de la parcelle pour l'estimation de la fraction de trou au nadir. On note que $P_o(0)$ est surestimée avec un biais parfois important pour les couverts ayant

une faible teneur en chlorophylle. Cette surestimation de la fraction de trou au nadir pour les parcelles ayant subi un traitement herbicide peut être due au fait que la diminution de chlorophylle provoque une hausse de réflectance dans le rouge. Le réseau de neurones interprète les réflectances de ces placettes comme étant celles d'un couvert sain, ayant les mêmes propriétés spectrales et par conséquent, comportant une proportion de sol nu plus importante. Le *RMSE* obtenu sur les placettes saines est satisfaisant (0.097 contre 0.04 pour le jeu de validation simulé) ; le biais est faible puisque les valeurs de R^2 et T sont similaires. On peut remarquer en outre, que, même pour les parcelles saines, le contenu en chlorophylle des feuilles est faible par rapport à la distribution que nous avons utilisée pour l'apprentissage du réseau : en effet, seuls 4,4% des cas du jeu d'apprentissage contiennent entre $10 \mu gcm^{-2}$ et $20 \mu gcm^{-2}$ de chlorophylle (traitement lourd), 9,6% entre $20 \mu gcm^{-2}$ et $30 \mu gcm^{-2}$ (traitement léger) et 14% entre $30 \mu gcm^{-2}$ et $40 \mu gcm^{-2}$ (parcelle saine). De plus, la différence entre les résultats obtenus sur les données simulées et les données expérimentales est due aux hypothèses du modèle SAIL : en particulier, la parcelle de betterave ne correspond pas tout à fait à un ensemble de particules absorbantes et diffusantes disposées aléatoirement dans une couche de végétation.

Ces résultats peuvent être comparés à ceux obtenus par Baret *et al.*(1995) qui estiment aussi la fraction de trou au nadir à l'aide de réseaux de neurones possédant la même architecture que la notre (figure 4.2). Cependant, les variables d'entrée du réseau ne sont que les réflectances au nadir dans le rouge et le proche infrarouge. L'apprentissage de ce réseau a été réalisé sur des données simulées avec SAIL, sur une gamme de couverts beaucoup plus réduite (distribution gaussienne de LAI de moyenne 0.86 et d'écart-type 0.65 ; distribution gaussienne de C_{ab} de moyenne $32 \mu gcm^{-2}$ et d'écart-type 7; distribution gaussienne de θ_l de moyenne 30° et d'écart-type 3°). C'est pourquoi, le réseau de neurones de Baret *et al.*, 1995 est, par construction, mieux adapté aux cultures de betterave. Il donne donc de meilleurs résultats. Par contre, on observe toujours une tendance du réseau à surestimer $P_o(0)$ pour les parcelles ayant subi un traitement herbicide.

Comme pour le jeu de données simulées, nous avons testé la stabilité de l'estimation de la fraction de trou au nadir par rapport à l'initialisation des poids et biais des réseaux (figure 4.3). Sur le jeu de données simulées, le *RMSE* reste inférieur à 0.1 quel que soit le réseau, 80% des cas donnent

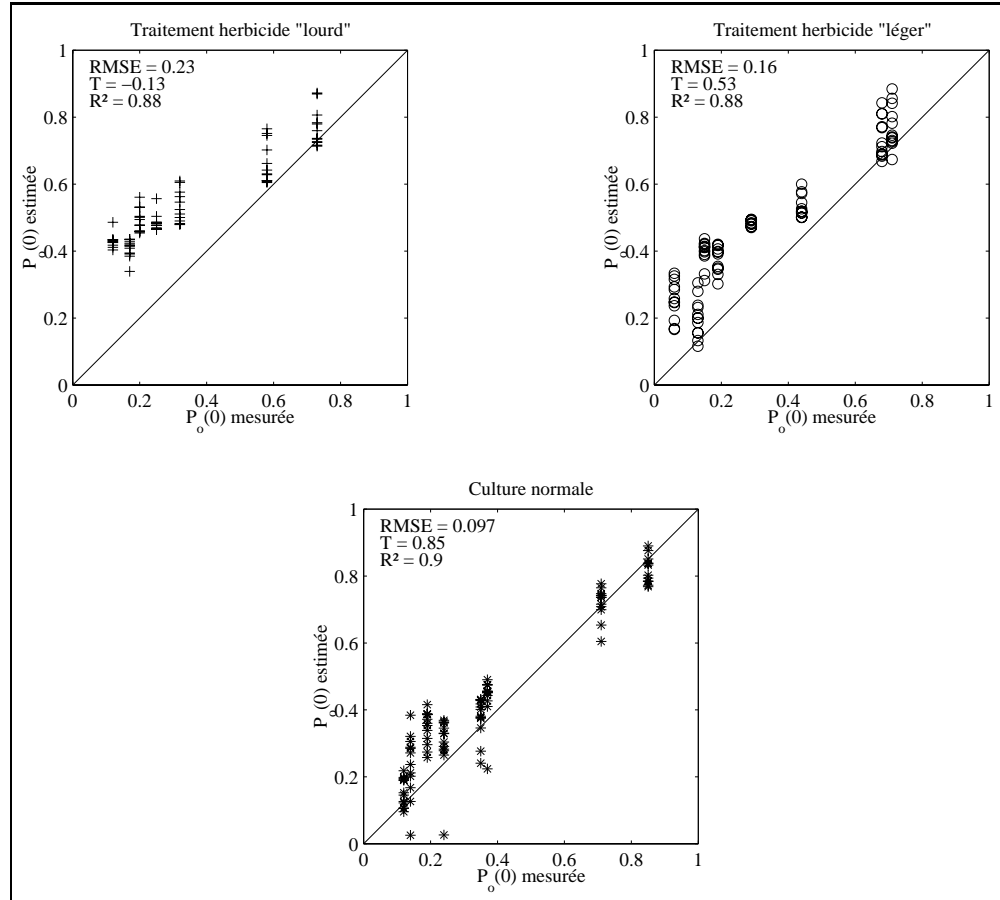


Figure 4.1 : Estimation de $P_o(0)$ sur la parcelle Grignon 1990 à partir de notre réseau de neurones. Les groupes de points correspondent à la variation due à la position du soleil (12 positions).

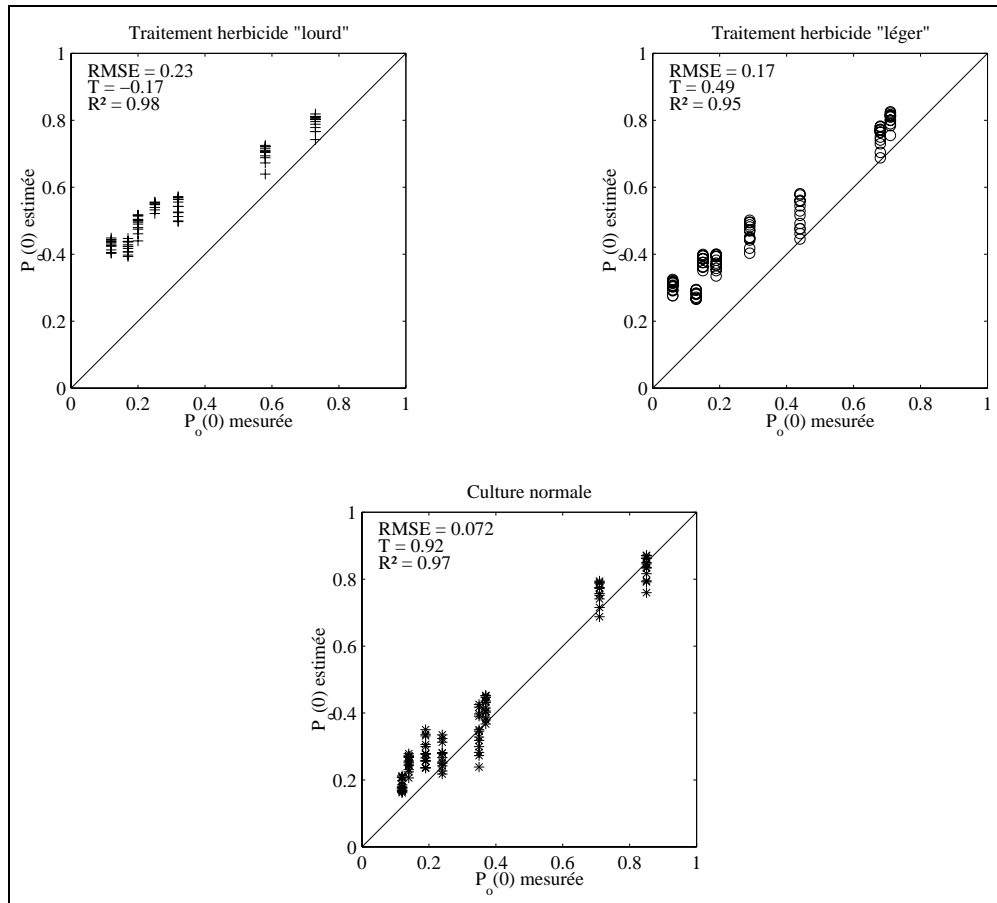


Figure 4.2 : Estimation de $P_o(0)$ sur la parcelle Grignon 1990 avec le réseau de neurones de Baret *et al.*, 1995.

un *RMSE* inférieur à 0.06. L'histogramme est plus étalé sur les données expérimentales mais 76% des réseaux donnent un écart quadratique moyen inférieur à 0.13, ce qui reste satisfaisant.

Estimation de $P_o(\theta_s)$ et $P_o(58^\circ)$

Le tableau 4.4 donne les paramètres statistiques de l'estimation de la fraction de trou dans les trois directions (nadir, angle zénithal solaire et angle zénithal de 58°) sur les mesures de la campagne Grignon. L'estimation de

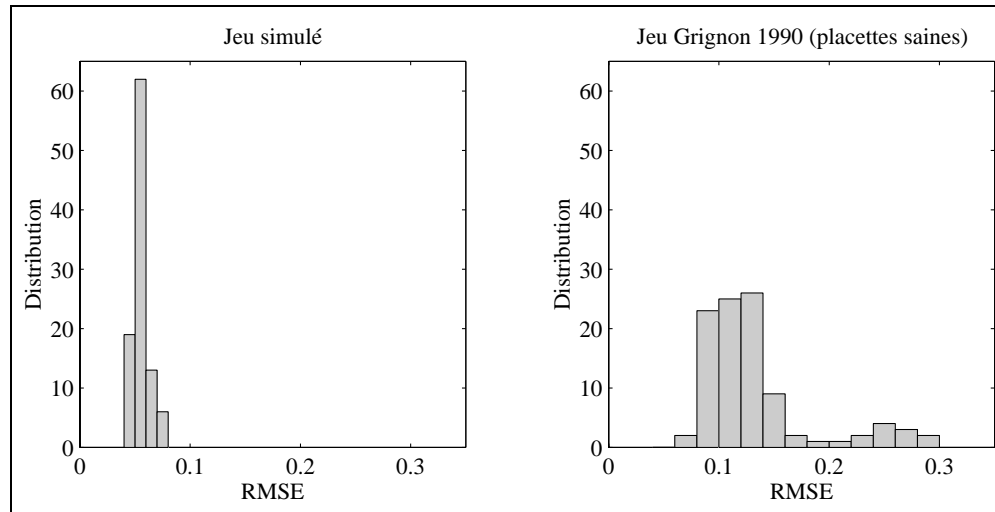


Figure 4.3 : Distribution de l'écart quadratique moyen obtenu sur l'estimation de la fraction de trou au nadir pour une centaine de réseaux différents en poids et biais.

$P_o(58^\circ)$ est moins précise que celle de la fraction de trou au nadir, même si l'on observe la même tendance à la surestimation pour les parcelles ayant subi un traitement herbicide.

| | Traitement lourd | | | Traitement léger | | | Parcelles saines | | |
|-----------------|------------------|-------|-------|------------------|------|-------|------------------|------|-------|
| | $RMSE$ | T | R^2 | $RMSE$ | T | R^2 | $RMSE$ | T | R^2 |
| $P_o(0)$ | 0.23 | -0.13 | 0.88 | 0.16 | 0.53 | 0.88 | 0.10 | 0.85 | 0.9 |
| $P_o(\theta_s)$ | 0.2 | 0.1 | 0.89 | 0.09 | 0.86 | 0.89 | 0.08 | 0.88 | 0.89 |
| $P_o(58^\circ)$ | 0.24 | -0.38 | 0.69 | 0.15 | 0.57 | 0.69 | 0.12 | 0.71 | 0.79 |

Tableau 4.4 : Paramètres statistiques pour l'estimation de la fraction de trou sur la campagne Grignon 1990.

Par contre, concernant l'estimation de la fraction de trou au zénith solaire, on observe aussi la même tendance, mais seulement pour les parcelles ayant subi le traitement herbicide lourd. Pour l'ensemble des parcelles, le nuage de points est mieux centré sur la droite (1:1), particulièrement pour les parcelles saines ou ayant subi un traitement herbicide léger (figure 4.4). L'estimation

de $P_o(\theta_s)$ est donc légèrement plus précise, ce qui confirme les résultats que nous avons sur les jeux de données simulées.

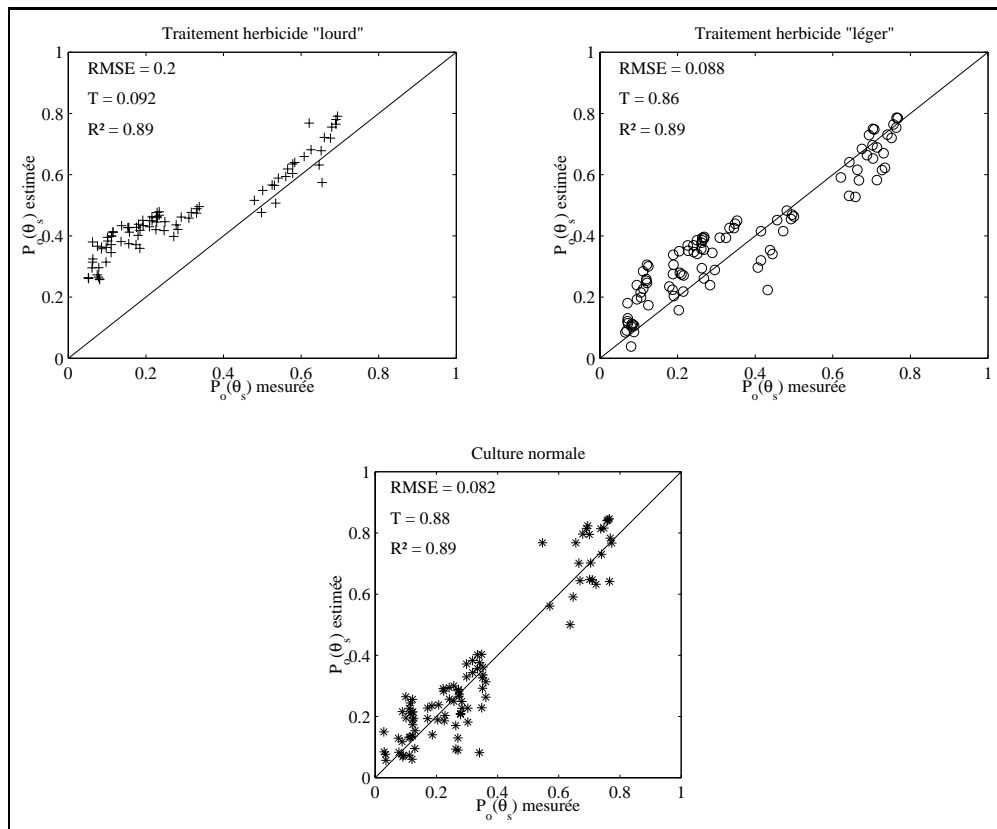


Figure 4.4 : Estimation de $P_o(\theta_s)$ sur la parcelle Grignon 1990.

4.1.4 Conclusion

L'estimation de la fraction de trou dans des directions de visée particulières sur les données de la campagne de mesures Grignon 1990 est satisfaisante. La différence d'acquisition de mesures de réflectance entre les radiomètres CIMEL de l'expérimentation et le capteur *VEGETATION* ne semble pas avoir d'impact important sur les résultats. Cependant, la composition de notre jeu d'apprentissage, comme on pouvait s'y attendre, influe de manière relativement significative sur les résultats. À cela, on peut trouver plusieurs explica-

tions. D'une part, le réseau de neurones peut assimiler le couvert "malade" à un couvert sain possédant les mêmes propriétés spectrales et comportant donc un plus fort pourcentage de sol nu. D'autre part, cette surestimation peut être due aux hypothèses inhérentes au modèle SAIL (problème de structure ou non prise en compte des propriétés directionnelles des feuilles). Enfin, on peut penser que l'utilisation d'autres longueurs d'onde, permettant de discerner les pigments bruns des feuilles, pourrait améliorer les performances du réseau de neurones.

4.2 Validation sur la campagne Alpilles/ReSeDA 1997

Les conditions d'acquisition des mesures de réflectance bidirectionnelle sur l'expérimentation Grignon 1990 étaient loin des conditions réelles du capteur *VEGETATION*. En effet, d'une part, seuls cinq angles de visée étaient pris en compte pour douze positions solaires et, d'autre part, il n'y avait ni effet atmosphérique, ni perturbations dues aux caractéristiques du capteur puisque les mesures étaient ponctuelles, effectuées à l'aide de radiomètres situés à 5 m au-dessus du sol. En outre, on ne considérait qu'une seule culture, même si une certaine variabilité des paramètres biophysiques avait été artificiellement introduite.

La campagne Alpilles/ReSeDA (Prévot *et al.*, 1998) au cours de laquelle les mesures bidirectionnelles ont été effectuées par le capteur *POLDER*² aéroporté (Deschamps *et al.*, 1994) correspond mieux aux conditions de mesures *VEGETATION*. Cette campagne de mesures a plusieurs objectifs :

1. L'estimation de caractéristiques de la végétation et du sol à partir de données de télédétection (méthodes d'inversion) ;
2. L'assimilation de données de télédétection dans les modèles de culture ;
3. L'assimilation de données de télédétection dans les modèles *SVAT*.

L'expérimentation a été réalisée sur une surface de 5 km × 5 km dans le sud-est de la France, située à 43°47' de latitude nord et 4°45' de longitude est (figure 4.5). Les mesures ont été réalisées durant tout le cycle cultural sur des cultures d'hiver (blé), d'été (maïs, tournesol) ou encore annuelles (luzerne).

²Polarization and Directionality of the Earth's Surface

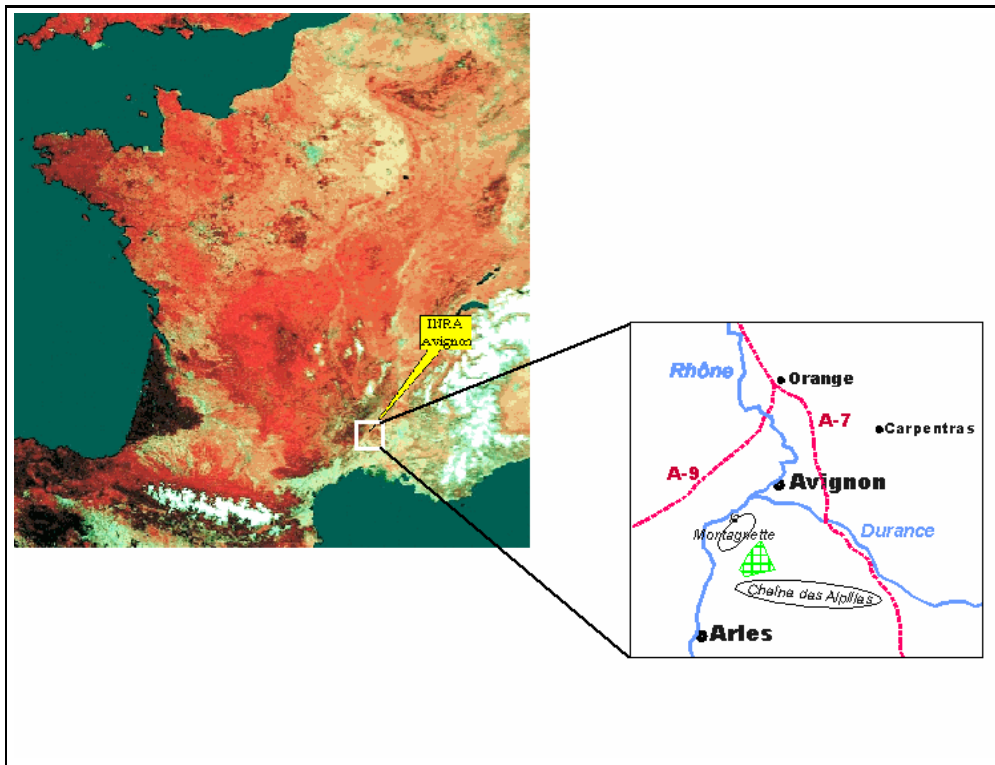


Figure 4.5 : Localisation du site de mesures Alpilles/ReSeDA (image NOAA/AVHRR du 9 mai 1987).

4.2.1 Composition du jeu de données

Compte tenu de la dimension importante de la base de données de la campagne Alpilles/ReSeDA qui dépasse largement le cadre de notre étude, nous ne décrivons dans ce paragraphe que les données qui serviront à valider notre algorithme.

Description des parcelles

La campagne de mesures a été effectuée sur un carré de 5 km × 5 km. Cependant, pour des raisons évidentes d'emploi du temps, seules certaines parcelles ont été suivies. Les mesures dont nous avons besoin ont été acquises environ une fois par semaine, durant tout le cycle cultural, sur 15 parcelles :

7 de blé, 4 de tournesol, 3 de maïs et une de luzerne (figure 4.6).

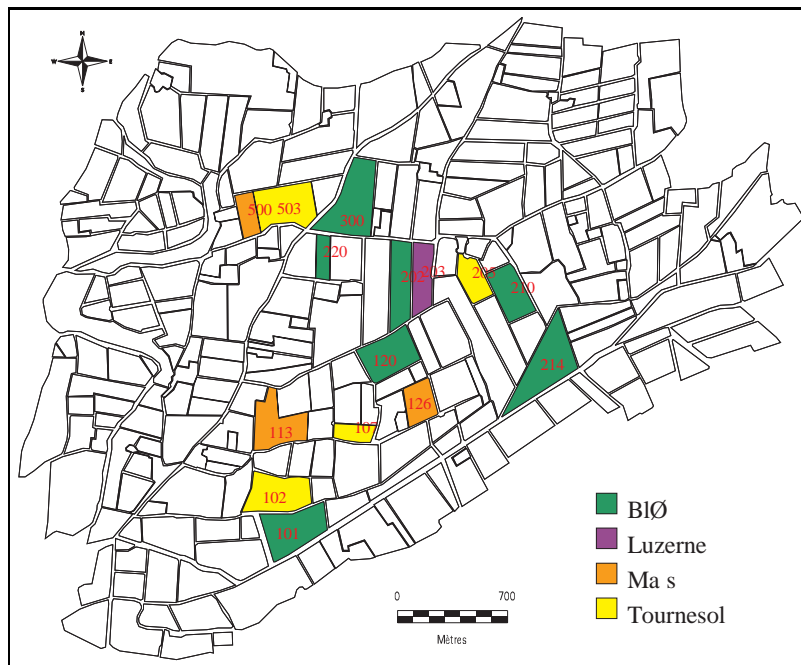


Figure 4.6 : Site de mesures Alpilles/ReSeDA. Localisation des parcelles.

Mesure de variables biophysiques

Pour réaliser le suivi des parcelles, les variables biophysiques ont été mesurées à l'aide de l'instrument *LAI2000*. Cet appareil mesure la transmittance du couvert dans cinq directions zénithales grâce à une lentille de type "fisheye". Les mesures sont faites pour des longueurs d'onde inférieures à 490 nm, ce qui permet de considérer que les feuilles sont noires. La transmittance est calculée comme le rapport entre une mesure du rayonnement incident au-dessus de la canopée (mesure 'A') et une mesure de rayonnement au niveau de la dernière strate de feuilles vertes (mesure 'B').

Pour obtenir une bonne représentativité du couvert, dix placettes sont échantillonnées au hasard dans la parcelle. Sur chacune d'entre elles, sont effectuées dix mesures 'B', pour une mesure 'A', selon le protocole décrit dans la figure 4.7, qui tient compte du fait que les cultures sont en rang. L'utilisateur

se tient toujours dos au soleil. Un cache noir couvrant la moitié de la lentille permet de l'isoler du capteur.

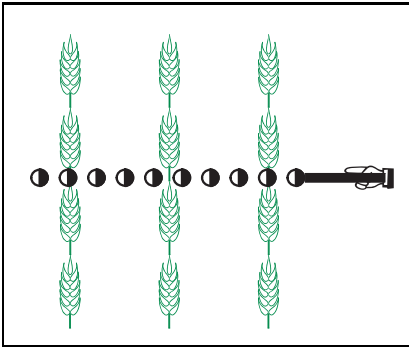


Figure 4.7 : Protocole de mesure des variables biophysiques sur Alpillles/ReSeDA.

Pour obtenir les variables biophysiques (fraction de trou, LAI , angle moyen d'inclinaison des feuilles), trois hypothèses sont prises en compte :

1. Lors de la mesure 'B', le rayonnement incident ne varie pas de manière significative par rapport à la mesure 'A' ;
2. Les feuilles sont assez petites et réparties de manière aléatoire et la distribution d'inclinaisons ($\xi(\theta_f, \theta_v)$) des feuilles est ellipsoïdale (Campbell, 1986). Cette hypothèse est sans doute forte compte tenu du caractère agrégé ou régulier de la plupart des couverts ;
3. L'orientation azimutale des feuilles est aléatoire. Cette hypothèse peut être critiquée, en particulier dans le cas du tournesol.

Le LAI , et l'angle moyen d'inclinaison des feuilles sont déduits de la fraction de trou monodirectionnelle assimilée à la transmittance mesurée, à partir de l'équation suivante :

$$P_o(\theta_v) = \exp(-LAI \cdot \xi(\bar{\theta}_f, \theta_v)) \quad (4.1)$$

L'indice foliaire et $\bar{\theta}_f$ sont obtenus à partir de l'équation 4.1 par une méthode d'optimisation (algorithme du simplex) en prenant comme valeurs initiales $LAI = 2$ et $\bar{\theta}_f = 45^\circ$. La figure 4.8 présente un exemple de l'évolution temporelle du LAI mesuré sur la parcelle de luzerne n° 203. On observe bien une évolution en dents de scie, correspondant aux périodes de croissance entrecoupées par les fauches (jours 163, 197 et 230).

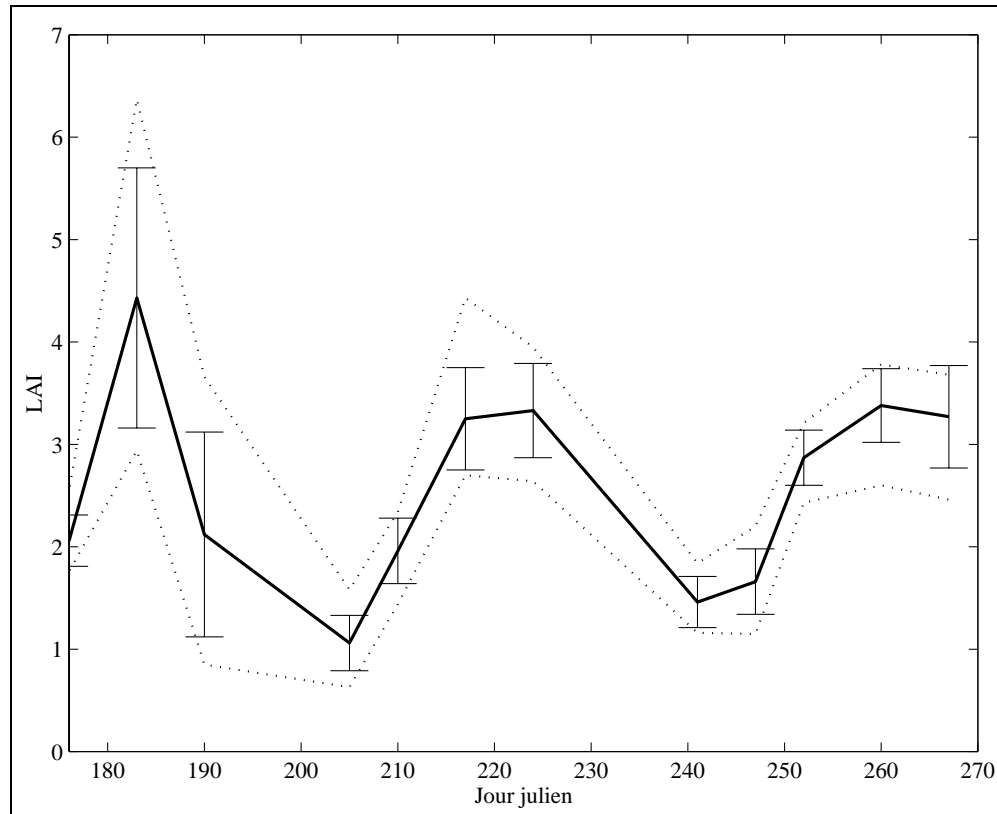


Figure 4.8 : Evolution temporelle du LAI de la parcelle de luzerne n°203 : moyenne et écart-type (traits pleins), valeurs minimale et maximale (pointillés).

Mesure de réflectance bidirectionnelle

PolDER est un radiomètre conçu pour mesurer la directionnalité et la polarisation du rayonnement solaire réfléchi par le système Terre-atmosphère. C'est une caméra CCD, munie d'une optique télécentrique à large champ, et d'une roue porte-filtres permettant d'obtenir des mesures polarisées ou non, dans différentes bandes spectrales. En ce qui nous concerne, nous n'utiliserons que les bandes non polarisées qui sont centrées sur : 443 nm (20 nm de large), 550 nm (20 nm de large), 670 nm (20 nm de large) et 864 nm (40 nm de large).

Le capteur était monté à bord d'un avion PIPER Aztec volant à 3000

m d'altitude, fournissant une taille de pixel de l'ordre de 20 m. Les images successives pendant le vol permettent d'observer le même pixel selon différents angles de visée. Le plan de vol est constitué de 5 axes : 4 axes parallèles au plan principal, et un axe dans le plan perpendiculaire au plan solaire. Les vols avaient lieu approximativement au midi solaire.

Les images dont nous disposons ont subi deux traitements successifs effectués par le CESBIO³ à Toulouse :

- des corrections géométriques : elles sont établies grâce à un modèle géométrique, à partir des paramètres d'attitude de l'avion (tangage, roulis, cap) mesurés avec précision grâce à la présence d'une centrale inertielle à bord ;
- des corrections atmosphériques : elles sont obtenues grâce à l'algorithme *SMAC*⁴ (Rahman and Dedieu, 1994), basé sur l'inversion d'un modèle physique du transfert radiatif dans l'atmosphère. Lorsque les mesures n'étaient pas disponibles, les paramètres de l'inversion (épaisseur optique à 550 nm, pression, contenu en vapeur d'eau et en Ozone) ont été fixés à des valeurs standard.

Toutes les images sont géoréférencées en projection Lambert II étendu sur un carré de 5 km par 5 km repéré par son coin inférieur droit (1867 km nord, 792 km est). Nous avons ensuite extrait les réflectances et les angles de visée associés à chacune des parcelles que nous étudions pour chaque journée de vol. Un masque a donc été établi à partir d'une carte de l'institut géographique national géoréférencée dans le même système, permettant d'identifier les pixels de l'image de chaque parcelle considérée. Pour un pixel de l'image, le calcul des angles de visée (θ_v, φ_v) est donné par (figure 4.9) :

$$\begin{cases} \theta_v = \arctan(d/a) \\ \varphi_v = \arctan(-dx/dy) \end{cases} \quad (4.2)$$

où

- a est l'altitude du capteur,

³Centre d'Etudes spatiales de la BIOSphère

⁴Simplified Method for the Atmospheric Correction

- dx (resp. dy) représente la différence, en mètres, entre l'abscisse (resp. l'ordonnée) du pixel et celle de l'avion, données en coordonnées Lambert II étendu,
- $d = \sqrt{dx^2 + dy^2}$

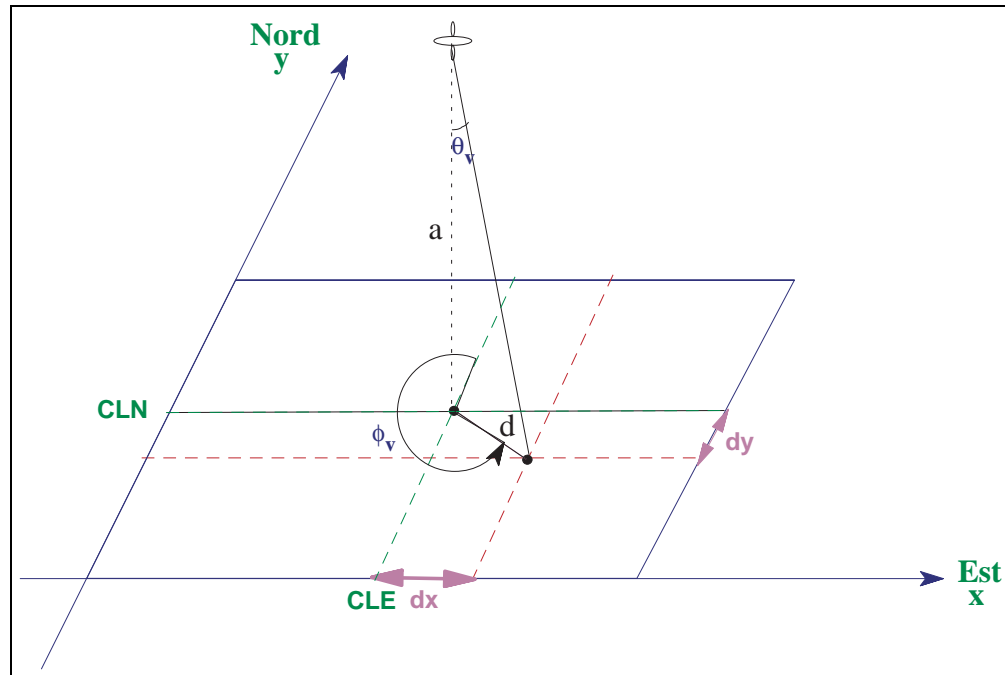


Figure 4.9 : Calcul des angles de visée pour un pixel donné d'une image POLDER géoréférencée. CLN (resp. CLE) correspond à la coordonnée Lambert nord (resp. est) de l'avion.

Afin d'éviter les perturbations dues à la fonction de transfert de modulation (*FTM*) du capteur ainsi que les résidus des corrections géométriques, nous avons éliminé systématiquement les pixels de contour des parcelles (figure 4.11). En effet, la *FTM* est due au fait que le signal reçu ne provient pas uniquement du pixel visé, mais subit aussi l'influence des pixels voisins. Aussi, l'effet de la *FTM* est faible lorsque l'on considère une surface radiométriquement homogène, et forte lorsque l'on considère la transition entre deux surfaces ayant un comportement radiométrique différent.

Nous avons observé peu de variabilité des angles de visée et de la *BRDF* associés à chaque pixel d'une même parcelle. Nous considérerons donc la *BRDF* moyennée sur l'ensemble des pixels (figure 4.12). Les figures qui suivent résument les différentes étapes du traitement des images *PolDER*. Pour la bande spectrale centrée sur 550 nm, on observe bien une baisse du niveau de réflectance entre l'image corrigée seulement géométriquement et l'image corrigée géométriquement et atmosphériquement. Les fortes valeurs de réflectance des parcelles situées dans la partie haute de l'image proviennent du fait qu'elles sont observées dans la direction du hot spot. La petite zone rouge située dans la partie basse à droite de l'image correspond à des bâtiments d'une zone industrielle.

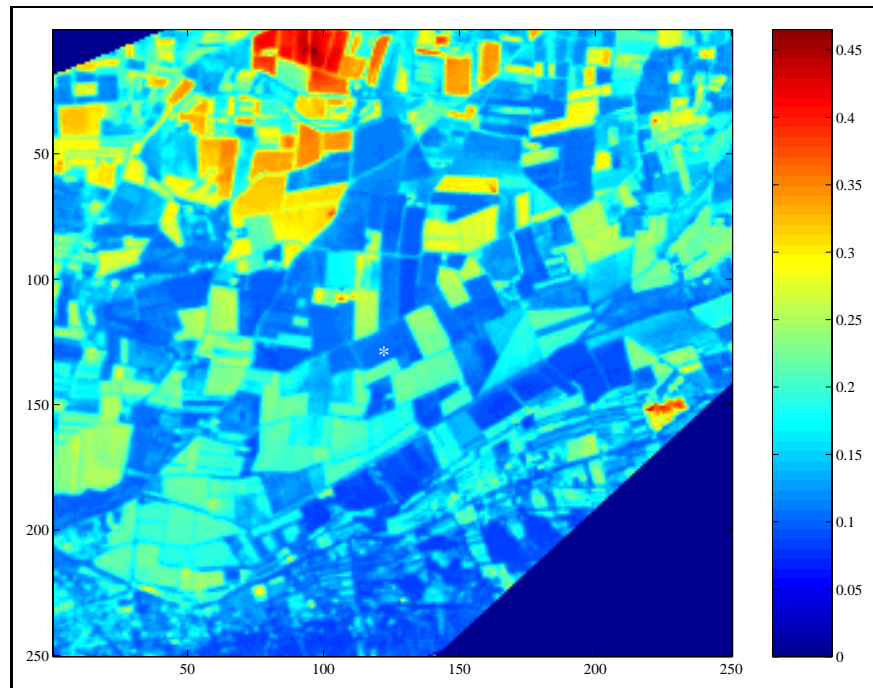
Si l'on considère le comportement radiométrique de la parcelle de blé n° 120, on observe bien le contraste entre le visible (niveau de réflectance bas) et le proche infrarouge (niveau de réflectance moyen) qui montre que le couvert est bien développé (*LAI* de l'ordre de 2, figure 4.13). La figure 4.11 fait ressortir les problèmes liés au contour des parcelles et la nécessité d'effectuer une érosion. Enfin, on observe la faible variabilité des angles de visée, comme de la réflectance, sur l'ensemble des pixels de la parcelle (figure 4.12). On note bien deux pics de réflectance dus au hot spot lorsque l'observation se fait dans le plan principal ($\cos \varphi_v = -1$).

Compte tenu du caractère récent de la campagne, nous ne disposons, à ce jour, que de la date du 10 avril 1997 pour laquelle nous avons à la fois des données terrain et des mesures de réflectance *PolDER* aéroportées.

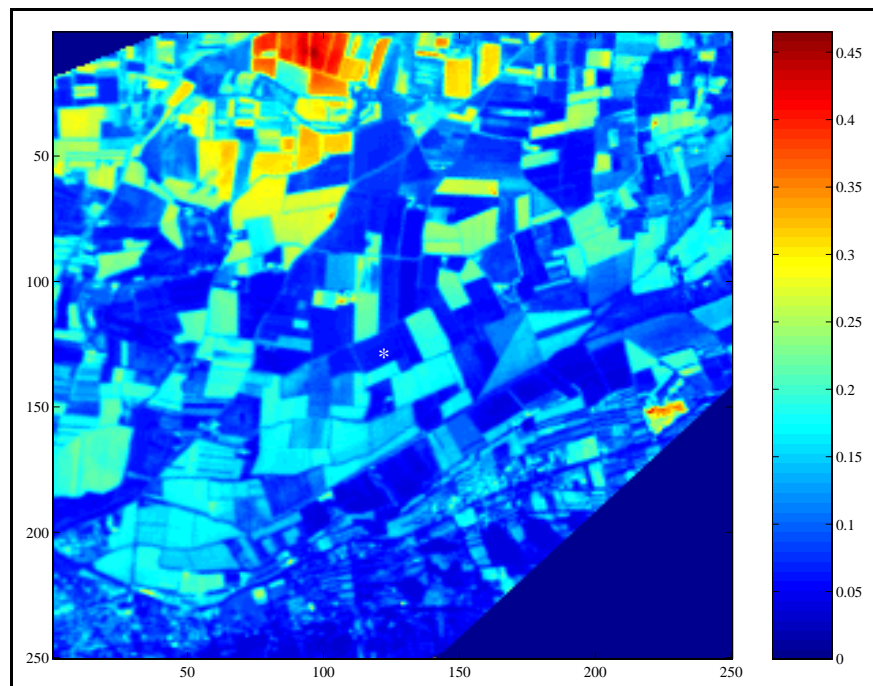
4.2.2 Construction d'un réseau de neurones adapté aux longueurs d'onde *PolDER*

De même que pour la campagne Grignon 1990, les longueurs d'onde disponibles sont différentes de celles de *VEGETATION*. Nous avons donc utilisé la base de données des variables d'entrée de *SAIL* du chapitre 2, pour obtenir le jeu de *BRDF* dans les longueurs d'onde *PolDER* (bleu, vert, rouge, proche infrarouge) correspondant aux 1500 couverts.

L'acquisition des données au niveau directionnel (figure 4.14) est plus proche des conditions de *VEGETATION* que lors de la campagne Grignon. Même si le passage de *PolDER* a lieu vers 12 heures, contrairement à *VEGETATION* qui passe à 10h30, la gamme de variation des latitudes des couverts

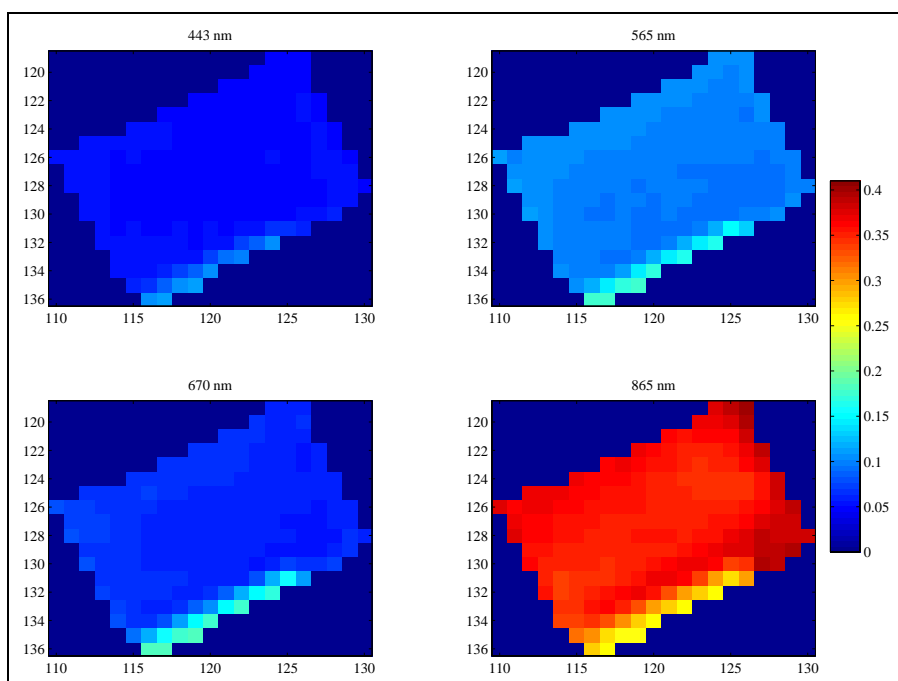


a. Image *PolDER* de la région des Alpilles géoréférencée et corrigée géométriquement à 550 nm.

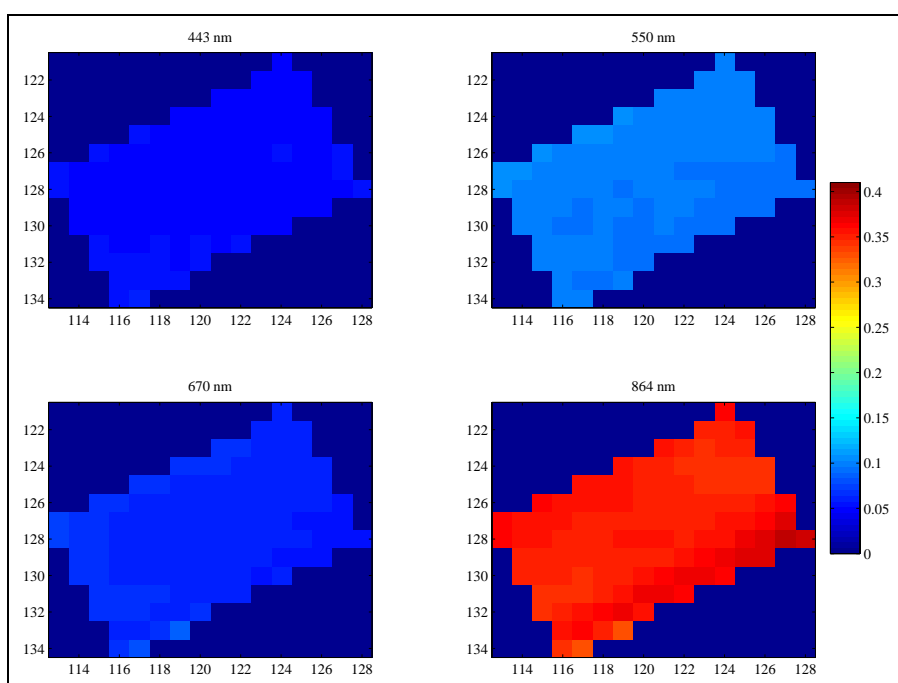


b. Image *PolDER* de la région des Alpilles géoréférencée et corrigée géométriquement et atmosphériquement à 550 nm.

Figure 4.10 : Effets des traitements géométrique et atmosphérique des images *PolDER* (10 avril 1997). La parcelle n° 120 est repérée par une étoile.



a. Extraction des réflectances bidirectionnelles de la parcelle de blé n° 120.



b. Erosion des contours de la parcelle de blé n° 120.

Figure 4.11 : Images en réflectance de la parcelle de Blé n° 120 pour des angles de visée : $\theta_v = 6^\circ$, $\varphi_v = 16^\circ$ (10 avril 1997).

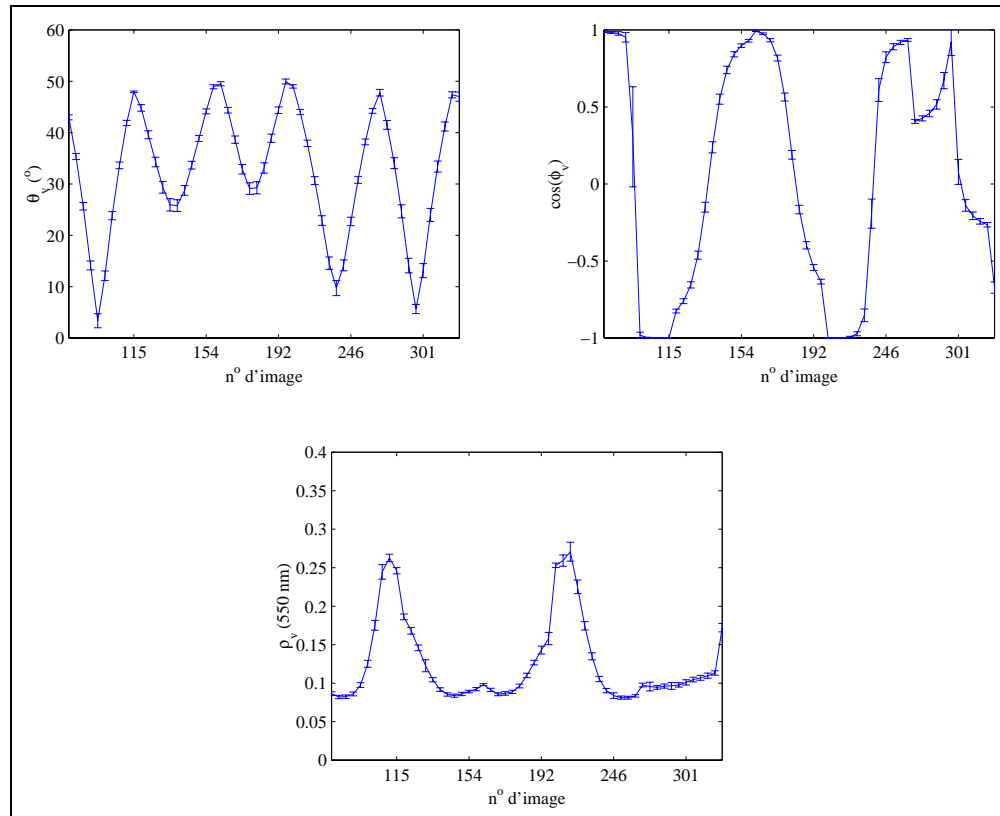


Figure 4.12 : Extraction de la *BRDF* de la parcelle de blé n° 120, moyenne et écart-type sur l'ensemble des pixels de la parcelle (550 nm), le 10 avril 1997 ($\theta_s = 37^\circ$).

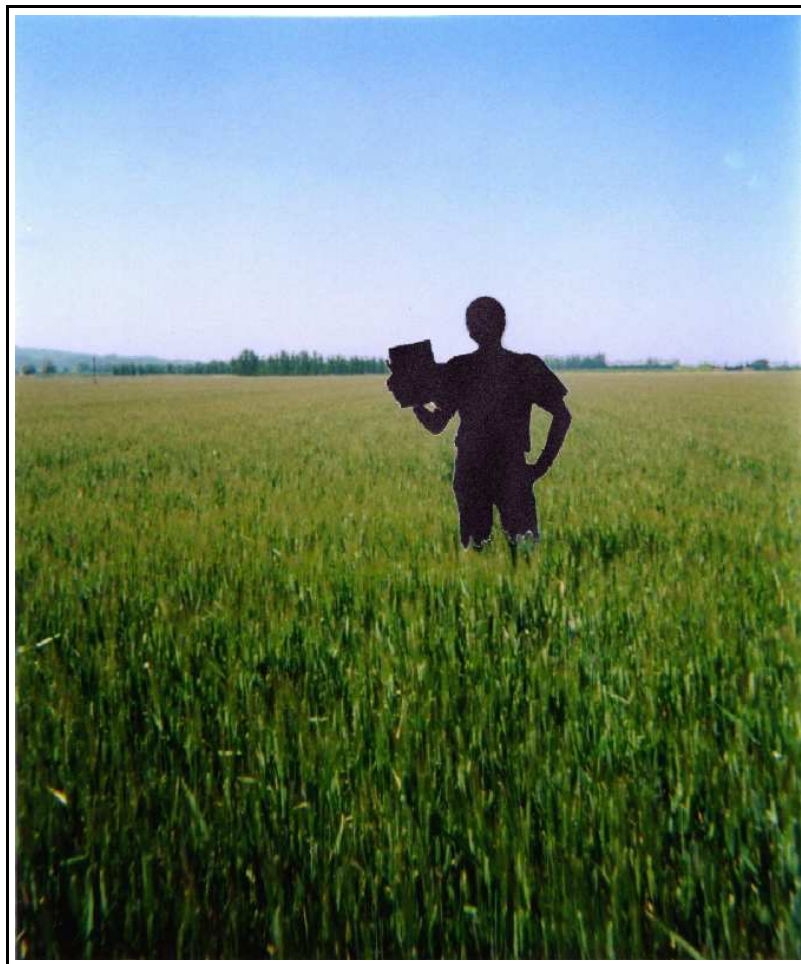


Figure 4.13 : Parcelle de blé n°120, le 23 avril 1997.

observés dans la base d'apprentissage est assez large pour couvrir l'ensemble des positions solaires de l'expérimentation. *PolDER* aéroporté permet d'acquérir plus de données directionnelles que *VEGETATION* et, de plus, échantillonne dans le hot spot. Nous avons montré toutefois que, d'une part, l'échantillonnage directionnel de *VEGETATION* est suffisant pour avoir une bonne estimation de ρ_o et ρ_h ; et, d'autre part l'acquisition de données dans le hot spot n'a que peu d'effet sur le calcul de ces deux grandeurs (§ 3.2.4). Nous avons en effet calculé le *RMSE* entre les valeurs de ρ_o et ρ_h obtenues par inversion du modèle de Walthall en tenant compte du hot spot, et celles obtenues en éliminant les données de réflectance proches du plan principal ($\pm 10^\circ$). Nous obtenons peu de différence avec un *RMSE* de 0.002 pour ρ_o et de 0.01 pour ρ_h . Nous avons donc utilisé l'ensemble des données acquises par *PolDER* aéroporté pour réaliser l'inversion du modèle de Walthall. Il faut cependant noter qu'il est plus prudent, lorsque l'on dispose de beaucoup de données dans le hot spot, de ne pas tenir compte de celles-ci afin d'obtenir une meilleure précision sur les variables d'entrée du réseau de neurones.

Nous avons donc réalisé l'apprentissage d'un autre réseau de neurones alimenté par les réflectances hémisphériques et au nadir dans les 4 bandes non polarisées de *PolDER*, complété par l'angle zénithal solaire lors de l'acquisition des données. Nous obtenons des performances similaires à celles du réseau de neurones construit pour les longueurs d'onde *VEGETATION*: $RMSE = 0.06$, $T = 0.96$ et $R^2 = 0.98$. Comme précédemment, nous avons construit une centaine de réseaux qui ne diffèrent que par les poids et les biais des neurones qui les composent : le *RMSE* entre fraction de trou mesurée et estimée sur le jeu de validation simulé reste stable, autour de 0.05, quel que soit le réseau de neurones employé.

4.2.3 Validation de l'algorithme sur les données Alpilles/ReSeDA 1997

Contrairement à la campagne Grignon où les mesures de $P_o(0)$, LAI et $\bar{\theta}_f$ étaient indépendantes, les valeurs d'indice foliaire et d'angle moyen d'inclinaison des feuilles de la campagne Alpilles/ReSeDA, nécessaires pour le calcul de $P_o(58^\circ)$ et $P_o(\theta_s)$, sont déduites de la mesure de $P_o(0)$. C'est pourquoi on ne considérera ici que l'estimation de la fraction de trou dans la direction du nadir.

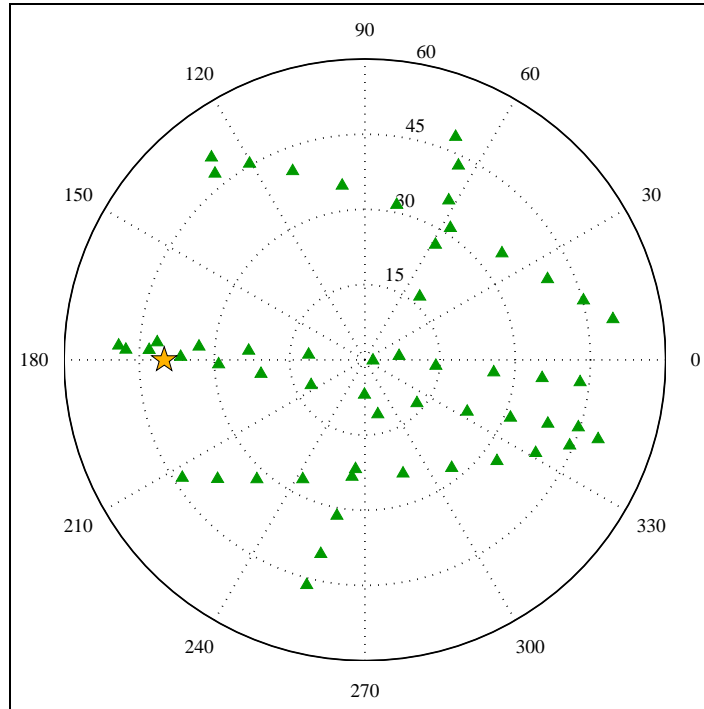


Figure 4.14 : Représentation polaire des angles de visée *POLDER* (triangles) pour le pixel central de la parcelle de blé n° 120, le 10 avril 1997. Le soleil est repéré par une étoile.

Le 10 avril 1997, les parcelles de tournesol et de maïs se trouvaient encore à l'état de sol nu, les semis ayant été effectués entre la mi-avril et la mi-mai. Pour ces parcelles, nous avons imposé une valeur de fraction de trou égale à 1 car les mesures au *LAI2000* n'étaient pas disponibles. Les mesures au sol et aéroportées sont disponibles sur un total de 12 parcelles correspondant à du blé ($P_o(0)$ moyenne), de la luzerne ($P_o(0)$ faible) du maïs et du tournesol.

Variables d'entrée du réseau de neurones : (ρ_o, ρ_h) dans les 4 bandes *POLDER* et θ_s .

La figure 4.15a montre les résultats que nous avons obtenus avec le réseau de neurones du paragraphe précédent. La précision d'estimation est satisfaisante (*RMSE* de 0.09), on observe un léger biais avec un T de 0.94 et un

coefficient de détermination de 0.99. Les paramètres statistiques obtenus sur ce jeu sont légèrement pessimistes dans la mesure où nous avons considéré une fraction de trou mesurée de 1 sur les parcelles de sol nu alors qu'en réalité, elle est légèrement inférieure.

Ces résultats sont néanmoins plutôt optimistes car l'analyse de sensibilité en fonction de l'initialisation des réseaux de neurones indique une instabilité considérable des performances d'estimation (figure 4.15b) : moins de 10% des réseaux estiment la fraction de trou avec un *RMSE* de l'ordre de 0.1. Comme la précision d'estimation sur les données Grignon 1990 est satisfaisante, l'hypothèse la plus probable que l'on puisse émettre est que l'instabilité est due au bruit associé aux mesures de réflectance qui a un poids beaucoup plus important dans la campagne Alpilles/ReSeDA.

Afin de confirmer cette hypothèse, nous avons bruité les données du jeu de validation simulé. Nous avons considéré deux sources de bruit :

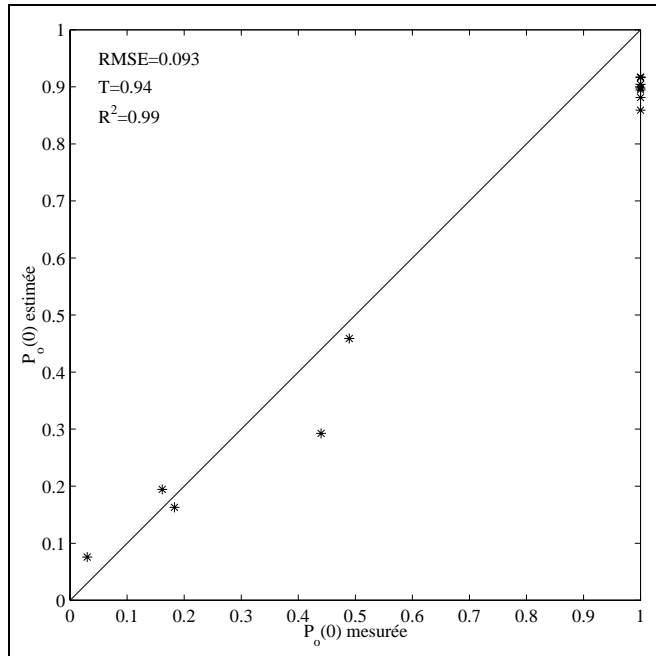
- le bruit résiduel (biais) ϵ_a des corrections atmosphériques qui perturbe essentiellement le signal dans le bleu : pour que le niveau de bruit soit réaliste, nous nous sommes appuyés sur les travaux de Baret *et al.*(1997) basés sur des simulations de réflectance en haut de l'atmosphère à l'aide de modèles de transfert radiatif dans les couverts végétaux et l'atmosphère (tableau 4.5) ;
- le bruit dû à la mesure (étalonnage, correction géométrique, capteur, chaîne de traitement). Nous avons fait l'hypothèse assez réaliste d'un bruit gaussien correspondant à une valeur relative ϵ_r de 5% de la réflectance moyenne $\bar{\rho}$ calculée sur le jeu de validation dans chaque bande.

Finalement, la réflectance bruitée ρ_b est donnée par :

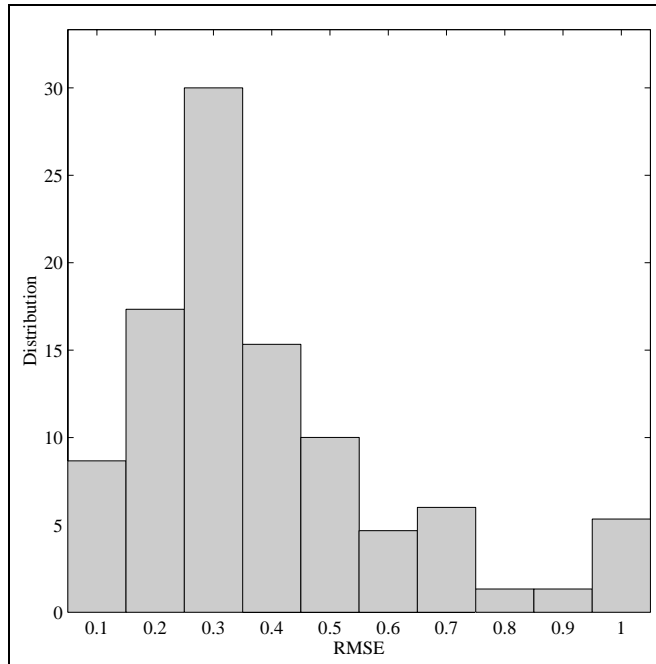
$$\rho_b = \rho + \bar{\rho}\epsilon_r\mathcal{N}(0, 1) + \epsilon_a \quad (4.3)$$

où $\mathcal{N}(0, 1)$ est la loi normale centrée réduite.

Nous avons alors testé les 100 réseaux de neurones sur le jeu de validation simulé et le jeu simulé et bruité artificiellement (figure 4.16). L'histogramme de répartition de l'écart quadratique moyen entre $P_o(0)$ mesurée et $P_o(0)$ estimée



a. Estimation de la fraction de trou au nadir.



b. Distribution de l'écart quadratique moyen obtenu sur l'estimation de la fraction de trou au nadir pour une centaine de réseaux différents en poids et biais.

Figure 4.15 : Validation sur les données expérimentales du 10 avril 1997 de la campagne Alpilles/ReSeDA. Les variables d'entrée des réseaux sont données à 443 nm, 670 nm, 864 nm.

| ϵ_a | | | |
|--------------|--------|--------|--------|
| 443 nm | 550 nm | 670 nm | 864 nm |
| 1.35% | 0.9% | 0.75% | 1.05% |

Tableau 4.5 : Niveau de bruit résiduel dû aux corrections atmosphériques (d'après Baret *et al.*(1997)).

est beaucoup plus étalé dans le cas des données bruitées. Par ailleurs, la précision d'estimation est dégradée puisqu'on passe d'une valeur de *RMSE* médiane de 0.04 à 0.09 pour les données bruitées.

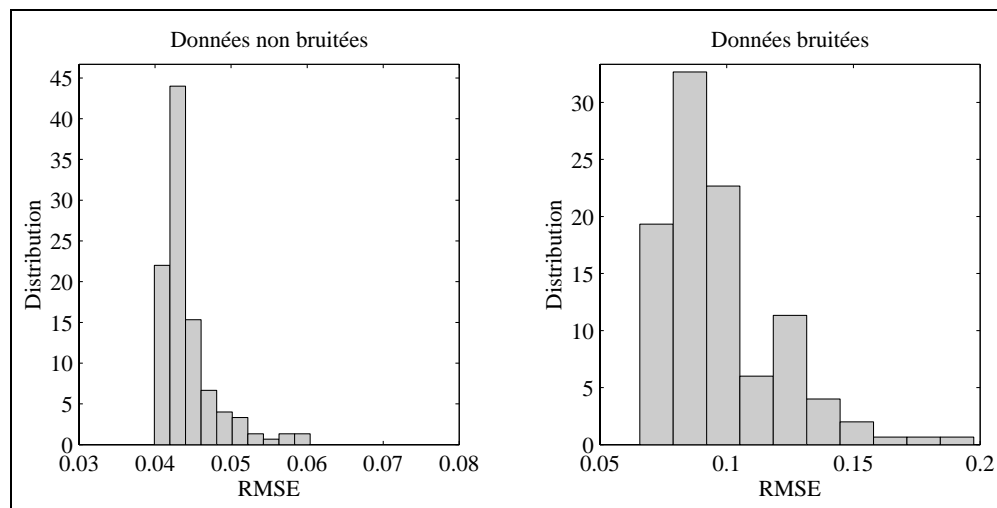


Figure 4.16 : Distribution de l'écart quadratique moyen obtenu sur l'estimation de la fraction de trou au nadir, sur le jeu de simulation bruité, pour une centaine de réseaux différents en poids et biais. Les longueurs d'onde utilisées en entrée sont : 443 nm, 550 nm, 670 nm et 864 nm.

Cela confirme donc l'hypothèse que l'instabilité des réseaux de neurones sur les données expérimentales est due à un niveau de bruit non négligeable sur les données Alpilles/ReSeDA. Par ailleurs, ce bruit semble important, puisque nous obtenons des *RMSE* beaucoup plus élevés sur les données expérimentales que sur les données simulées et bruitées artificiellement. Nous savons en effet que, si les corrections géométriques sur les données Alpilles sont précises, la

correction atmosphérique est plus grossière par manque de mesures “terrain” des paramètres atmosphériques.

Afin de réduire les problèmes dûs à l’atmosphère, nous proposons de construire un réseau de neurones utilisant les données seulement dans le vert, le rouge et le proche infrarouge. En effet, les résultats obtenus sur le jeu de simulation associé à la campagne Grignon 1990 (longueurs d’onde *SPOT*) montrent que la précision d’estimation n’est pas affectée lorsqu’on ne prend pas en compte les données dans le bleu.

Variables d’entrée du réseau de neurones : (ρ_o, ρ_h) dans 3 bandes *PolDER* (550 nm, 670 nm et 864 nm) et θ_s .

Nous avons directement simulé une centaine de réseaux de neurones initialisés avec des poids et des biais différents. La figure 4.17 montre que les résultats sur le jeux de validation simulé et bruité artificiellement sont satisfaisants ($RMSE < 0.06$) quelque soit le réseau de neurones employé. Concernant les données Alpilles/ReSeDA, 80% des réseaux estiment la fraction de trou au nadir avec un écart quadratique moyen inférieur à 0.07. L’étalement de l’histogramme des $RMSE$ a donc été considérablement réduit en excluant les variables d’entrée des réseaux neuronaux à 443 nm.

La figure 4.18 présente un exemple d’estimation de $P_o(0)$ sur les 12 parcelles, pour le 10 avril. Ces résultats corroborent donc encore l’hypothèse que l’utilisation de la bande bleue, pour les données Alpilles/ReSeDA, introduit une grande incertitude sur l’estimation de la fraction de trou.

4.2.4 Conclusion sur la campagne Alpilles/ReSeDA

Les performances du réseau de neurones dépendent fortement du niveau de bruit observé sur les données : lorsqu’il est important le réseau de neurones est incapable de prédire avec précision la fraction de trou. C’est ce que l’on observe en utilisant les réflectances de la bande bleue de *PolDER*, fortement contaminées par l’atmosphère. Par contre, si le niveau de bruit est “raisonnable”, la précision d’estimation est satisfaisante. C’est le cas lorsque l’on utilise les trois autres canaux de *PolDER* qui sont aussi contaminés, mais à un degré moindre, par les résidus de la correction atmosphérique et de la correction

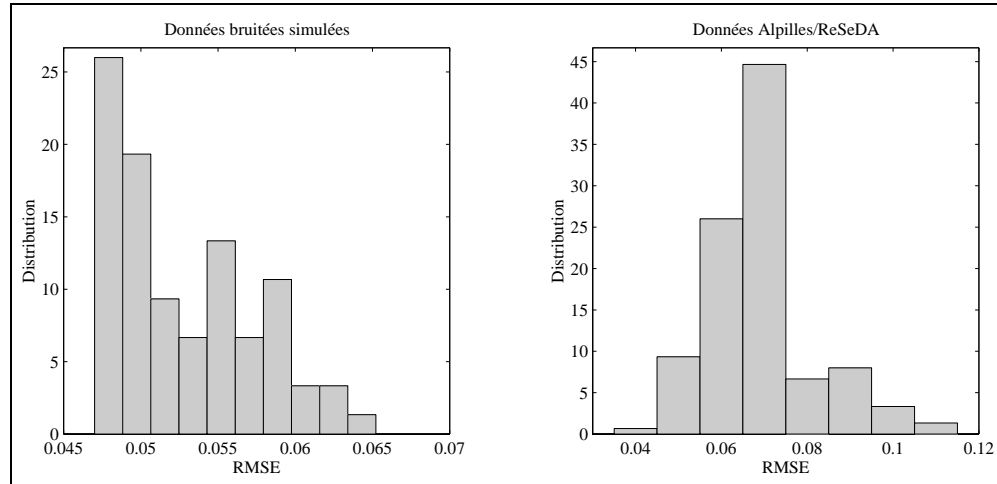


Figure 4.17 : Distribution de l'écart quadratique moyen obtenu sur l'estimation de la fraction de trou au nadir, sur le jeu de simulation bruité, pour une centaine de réseaux différents en poids et biais. Les longueurs d'onde utilisées en entrée sont : 550 nm, 670 nm et 864 nm.

géométrique, ainsi que par le bruit inhérent au capteur (bruit radiométrique, traitement de la chaîne d'acquisition).

Nous avons montré que la précision d'estimation sur le jeu de validation simulé est sensiblement la même que l'on utilise 3 ou 4 longueurs d'onde alors qu'elle est considérablement améliorée sur les données expérimentales. Ceci indique donc que l'utilisation des données de réflectance dans le bleu comme variables d'entrée du réseau de neurones représente plus une source d'erreur qu'un apport d'information supplémentaire pour l'estimation de la fraction de trou.

Des validations complémentaires devront être réalisées sur la campagne Alpilles/ReSeDA afin d'exploiter l'ensemble des données, qui, à ce jour, ne sont pas encore toutes traitées. Ainsi, les autres dates d'acquisition de données aéroportées permettront de valider l'algorithme sur d'autres cultures.

L'algorithme de suivi de la végétation étant initialement développé pour *VEGETATION*, nous avons construit un réseau de neurones n'utilisant pas les données de réflectance dans le canal bleu. Les résultats obtenus sur le jeu de validation simulé ($RMSE = 0.046$ et $T = 0.98$) permettent de confirmer que

la précision d'estimation est inchangée avec ou sans le bleu. Reste à valider ce réseau sur des données expérimentales provenant de l'acquisition de données *VEGETATION* embarqué sur la plateforme *SPOT4*.

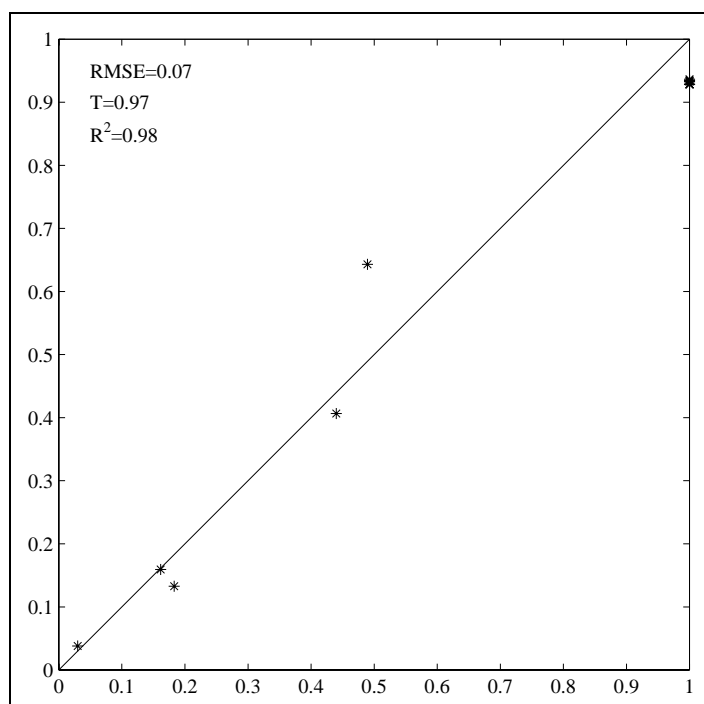


Figure 4.18 : Exemple d'estimation de la fraction de trou au nadir. Validation sur les données expérimentales du 10 avril 1997 de la campagne Alpilles/ReSeDA. Les variables d'entrée du réseau sont données à 550 nm, 670 nm et 864 nm.

◇

Validation de l'algorithme de suivi de la végétation sur des données expérimentales

Conclusions

◇

Le tableau 4.6 résume l'ensemble des validations effectuées sur notre algorithme de suivi de la végétation. Les résultats sont satisfaisants quel que soit le jeu de données employé, avec un RMSE inférieur à 0.1. Des validations complémentaires, dans les conditions réelles (VEGETATION embarqué sur SPOT4), sont nécessaires pour deux raisons principales :

- les problèmes liés à la fenêtre temporelle lors de l'acquisition de données ;*
- l'observation de la végétation naturelle, et tout particulièrement des forêts ;*

Au terme de cette étude, nous proposons donc un algorithme de suivi de la végétation basé sur l'estimation de la fraction de trou au nadir à partir des réflectances hémisphériques et au nadir à 645 nm, 835 nm et 1665 nm mesurées par VEGETATION, associées à l'angle solaire mesuré lors de l'acquisition des données.

| Conditions d'observation | Modèle SAIL | Modèles SAIL+bruit | Grignon 1990 Radiomètre CIMEL | Alpilles/ReSeDA POLDER aéroporté | VEGETATION SPOT4 |
|---------------------------------|-----------------------|---------------------------|--------------------------------------|---|--------------------------------|
| Rés. spatiale | 1.15 km | 1.15 km | 1 m | 20m | 1.15 km |
| Éch. directionnel | 20 à 70 données | 20 à 70 données | 5 données | ≈ 50 données | 20 à 70 données |
| Éch. temporel | instantané | instantané | instantané | instantané | 26 jours |
| B. spectrales | B, R, PIR, MIR | R, PIR, MIR | V, R, PIR | V, R, PIR | R, PIR, MIR |
| Bruit instrumental | non | oui | non | oui | oui |
| Atmosphère | non | oui | non | oui | oui |
| Couvert | Pix. homogène & mixte | Pix. homogène | betterave | blé, luzerne, sols nus | cultures, végétation naturelle |
| $RMSE(P_o(0))$ | 0.05 | 0.06 | 0.1 | 0.07 | ? |

Tableau 4.6 : Validation de l'algorithme de suivi de la végétation.

Chapitre 5

Conclusion.

Cette étude a tenté de répondre à un des objectifs de la mission *VEGETATION* portant sur le suivi à large échelle de la végétation. Il s'agit en particulier de suivre les variations spatio-temporelles de certaines variables biophysiques de surface utilisées ensuite comme limites inférieures aux modèles de flux de masse et d'énergie à l'échelle régionale ou globale.

Le suivi des couverts végétaux est réalisé à partir de variables biophysiques estimées grâce aux données de télédétection. Compte tenu des limitations des techniques actuelles d'extraction de ces variables biophysiques (indices de végétation et inversion de modèles physiques), nous avons développé un outil basé sur l'utilisation de réseaux de neurones, bénéficiant des avantages des deux techniques précédentes. L'algorithme ainsi développé doit pouvoir être appliqué de façon globale, sans connaissance *a priori* des surfaces observées.

Comme nous ne disposons pas de banque de données expérimentales assez large pour satisfaire au besoin de cette étude, nous avons élaboré une base de données à partir de simulations de modèles de transfert radiatif dans les couverts végétaux. La diversité des couverts est assurée par une large gamme de variation des variables d'entrée des différents modèles, établie à partir d'une connaissance expérimentale de leur distribution.

Cette base de données synthétique nous a permis de développer des réseaux de neurones permettant d'estimer différentes variables biophysiques des couverts végétaux : la fraction de trou dans des directions de visée particulières, l'indice foliaire, la fraction de rayonnement photosynthétiquement actif ab-

sorbé par le couvert, l'angle moyen d'inclinaison des feuilles, et la chlorophylle intégrée au niveau du couvert. L'estimation se fait à partir des réflectances bidirectionnelles mesurées par *VEGETATION* durant un cycle orbital de 26 jours, associées à la valeur de l'angle zénithal solaire lors de l'acquisition de ces données. Un prétraitement des données de réflectances bidirectionnelles est proposé de manière à avoir, en entrée des réseaux neuronaux, deux grandeurs physiques caractéristiques de la fonction de distribution de la réflectance bidirectionnelle : la réflectance au nadir et la réflectance hémisphérique. Compte tenu de la résolution spatiale de *VEGETATION*, l'influence de l'hétérogénéité intrapixel a été analysée pour différentes variables biophysiques. Il apparaît que les variables de type fraction de trou (y compris le *fAPAR*) sont beaucoup moins sensibles que les variables de type *LAI* à l'hétérogénéité du pixel. Nous avons, par ailleurs, montré l'apport des réseaux de neurones en comparaison à une technique basée sur l'utilisation d'un indice de végétation.

Nous avons finalement validé l'algorithme sur des données expérimentales. Les données de la campagne Grignon 1990 ont permis de montrer la robustesse de notre méthode lorsque le nombre de données bidirectionnelles acquises durant le cycle orbital est faible (basses latitudes, présence de nuages). Par ailleurs, elles ont aussi permis de confirmer que les performances d'estimation des variables biophysiques sont affectées non seulement par la représentativité du couvert et des conditions d'observations qui lui sont associées dans la base d'apprentissage, mais aussi par les hypothèses inhérentes aux modèles de transfert radiatif utilisés pour générer la base d'apprentissage des réseaux de neurones. Les données de la campagne Alpilles/ReSeDA 1997, plus proches des conditions réelles d'acquisition de *VEGETATION*, ont permis, en outre, de mettre en évidence la sensibilité de l'algorithme au bruit qui n'était pas pris en compte lors de la construction de la base d'apprentissage simulée. Toutefois, moyennant certaines précautions, les résultats obtenus sur les deux campagnes de mesures sont satisfaisants et aboutissent à un écart quadratique moyen inférieur à 0.1. L'intérêt d'utiliser des réseaux de neurones apparaît donc clairement, en particulier vis à vis des approches classiques basées sur l'utilisation d'indices de végétation.

Au terme de ce travail, subsistent néanmoins certaines limitations dues aux choix et aux hypothèses que nous avons faits au cours de notre démarche :

- *Les limites imposées par les modèles utilisés*

L'utilisation de modèles de transfert radiatif pour générer notre base d'apprentissage induit nécessairement des hypothèses simplificatrices puisque, par définition, un modèle correspond à une schématisation de la réalité. Ainsi, les couverts végétaux présentant des structures particulières (cultures en rang, forêts boréales, ...), sont mal représentés par SAIL qui suppose une disposition aléatoire des éléments de la végétation. De plus, le comportement non lambertien des feuilles n'est pas pris en compte par le modèle ;

- *Les limites imposées par notre démarche*

Dans la mesure où nous avons choisi d'utiliser des réseaux neuronaux, la représentativité de la base d'apprentissage est un facteur limitant. En effet, si l'on soumet au réseau de neurones un jeu de réflectances éloigné des caractéristiques des couverts de la base d'apprentissage, l'algorithme ne sera pas capable d'extrapoler la valeur de la variable biophysique correspondante. En outre, nous n'avons pas tenu compte, dans la base d'apprentissage, des différentes sources de bruit induites par les étapes de l'acquisition du signal (atmosphère, capteur, chaîne de traitement). Enfin, nous avons considéré qu'il n'y avait pas d'évolution significative de la végétation dans une fenêtre temporelle de 26 jours, ce qui, dans le cas des surfaces cultivées ou des forêts à feuilles caduques, n'est pas très réaliste. Toutefois, dans la mesure où nous avons montré que les performances du réseau restent stables même pour un nombre restreint de données, on peut envisager de réduire la fenêtre temporelle d'observation à une dizaine de jours, sans affecter la précision de l'estimation des variables biophysiques.

Un travail conséquent est donc encore indispensable pour restreindre, lorsque c'est possible, l'effet de ces limitations et rendre l'algorithme que nous avons développé opérationnel :

- *Sur le plan expérimental*

Il est nécessaire de valider l'algorithme, d'une part, sur des données expérimentales provenant de *VEGETATION* embarqué sur *SPOT 4* ; et,

d'autre part, sur d'autres types de surfaces que des cultures, plus particulièrement des forêts. L'utilisation d'une variable biophysique pour suivre les couverts végétaux apparaît ici comme essentielle. Elle garantit de réelles possibilités de validation. En particulier, la mesure de la fraction de trou au nadir, même à l'échelle du pixel *VEGETATION* est réaliste, le protocole de mesure *in situ* étant des plus simples.

En outre, lors des différentes validations effectuées dans cette étude, l'acquisition des réflectances bidirectionnelles est réalisée sur une même journée. Il faut donc aussi évaluer, l'effet de la taille de la fenêtre temporelle sur des données expérimentales.

Enfin, la problème du pixel mixte a été abordé sur des données simulées mais il serait aussi souhaitable de valider nos résultats sur des données expérimentales.

- *Sur le plan théorique :*

Il faut en particulier s'attacher à évaluer la robustesse de l'algorithme face à des données bruitées, car, si nous avons détecté un problème de stabilité sur la campagne Alpilles/ReSeDA, cela n'a pas été le cas pour Grignon. Pourtant les deux jeux de données sont naturellement bruités et/ou biaisés, bien sur, à des niveaux très différents (prise en compte d'effets géométriques et atmosphériques pour la campagne Alpilles). La connaissance conjointe des sources et de la structure du bruit est essentielle pour optimiser le système. Il faut en effet dissocier le "bruit" atmosphérique résiduel, le bruit radiométrique induit par les problèmes d'étalonnage absolu, interbande et multitemporel, ainsi que les bruits induits par la précision géométrique.

Nous avons comparé la méthode basée sur les réseaux de neurones avec une technique utilisant les relations indicielles entre les réflectances et des variables biophysiques des couverts végétaux. Il serait intéressant d'effectuer aussi une comparaison avec des méthodes d'inversion de modèle ou encore des techniques basées sur l'utilisation de tables de correspondance. Cette comparaison doit porter à la fois sur les temps de calcul, la précision et la robustesse de l'estimation.

Enfin, pour améliorer les performances de l'algorithme, on pourrait envisager d'étudier les potentialités de l'apport de nouvelles bandes spectrales

autres que celles de *VEGETATION* pour l'estimation des variables biophysiques. En effet, dans la mesure où les variables d'entrée des réseaux de neurones sont normalisées, l'utilisation conjointe de données multi-capteurs n'est pas exclue (*VEGETATION*, *NOAA/AVHRR*, *MERIS*, *PoLDER*, *PRISM*, ...). Opérationnel, notre algorithme de suivi de la végétation pourrait être couplé avec des modèles de fonctionnement de culture pour améliorer la gestion de la production agricole mondiale par la prévision de rendement ou le diagnostic de situation culturale en temps réel.

Bibliographie

- [1] AHMAD S. and TRESP V., 1993. *Some solutions to the missing feature problem in vision*. Morgan Kaufman, San Mateo, CA, Hanson S.J. , Cowan J.D. , Giles C.L. edition.
- [2] ALBERS B., STRAHLER A., LI X., LIANG S., and CLARKE K., 1990. Radiometric measurements of gap probability in conifer tree canopies. *Remote Sensing of Environment*, 34:179–192.
- [3] ANDRIEU B. and BARET F., 1993. *In Crop structure and light microclimate - Characterization and Applications-*, Varlet-Grancher C. and Bonhomme R. and Sinoquet H., Paris, France, chapter Indirect methods of estimating crop structure from optical measurements, pp. 285–322. INRA edition.
- [4] ANDRIEU B., BARET F., JACQUEMOUD S., MALTHUS T., and STEVEN M., 1997. Evaluation of an improved version of SAIL model to simulate bidirectional reflectance of sugar beet canopies. *Remote Sensing of Environment*, 60:247–257.
- [5] ANDRIEU B. and SINOQUET H., 1993. Evaluation of structure description requirements for predicting gap fraction of vegetation canopies. *Agricultural and Forest Meteorology*, 65:207–227.
- [6] ASRAR G., FUCHS M., KANEMASU E., and HATFIELD J., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal*, 76:300–306.
- [7] ASRAR G., KANEMASU E.T., and YOSHIDA M., 1985. Estimates of leaf area index from spectral reflectance of wheat under different cultural practices and solar angle. *Remote Sensing of Environment*, 17:1–11.

- [8] BADWAHR G., BUNNIK N., and VERHOEF W., 1985. Comparative study of Suits and SAIL canopy reflectance models. *Remote Sensing of Environment*, 17:179–195.
- [9] BARET F., ANDRIEU B., FOLMER J., HANOCQ J., and SARROUY C., 1993. In *Crop structure and light microclimate - Characterization and Applications* -, Varlet-Grancher C. and Bonhomme R. and Sinoquet H., chapter Gap fraction measurement from hemispherical infrared photography and its use to evaluate PAR interception efficiency, pp. 359–371. INRA edition.
- [10] BARET F., CLEVERS J., and STEVEN M., 1995. The robustness of canopy gap fraction estimates from red and near-infrared reflectances : a comparison of approaches. *Remote Sensing of Environment*, 54:141–151.
- [11] BARET F. and FOURTY T., 1998. Estimation of leaf water content and specific leaf weight from reflectance and transmittance measurements. *Agronomie*, 17(9-10):455–464.
- [12] BARET F. and GUYOT G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 35:161–173.
- [13] BARET F., GUYOT G., and MAJOR D., 1989. TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *12th, Canadian Symposium on Remote Sensing and IGARSS'90*. Vancouver, p. 4.
- [14] BARET F., JACQUEMOUD S., and HANOCQ J., 1993. The soil line concept in remote sensing. *Remote Sensing Reviews*, 7:65–82.
- [15] BARET F. and OLIOSSO A., 1989. Estimation à partir de mesures de réflectance spectrale du rayonnement photosynthétiquement actif absorbé par une culture de blé. *Agronomie*, 9:885–895.
- [16] BARET F., WEISS M., LEROY M., HAUTECOEUR O., SANTER R., and BÉGUÉ A., 1997. Impact of surface anisotropies on the observation of optical imaging sensors. *Technical Report 11341/95/NL/CN*, ESA.

- [17] BÉGUÉ A., 1992. Modeling hemispherical and directional radiative fluxes in regular-clumped canopies. *Remote Sensing of Environment*, 40:219–230.
- [18] BÉGUÉ A., 1993. Leaf area index, intercepted photosynthetically active radiation, and spectral vegetation indices : a sensitivity analysis for regular-clumped canopies. *Remote Sensing of Environment*, 46:45–59.
- [19] BÉGUÉ A., PRINCE S., HANAN N., and ROUJEAN J., 1996. Shortwave radiation budget of sahelian vegetation. 2. Radiative transfer model. *Agricultural and Forest Meteorology*, 79:97–112.
- [20] BICHERON P. and LEROY M., 1999. A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model. *Remote Sensing of Environment*, 67:251–266.
- [21] BISHOP C., 1995. *Neural network for pattern recognition*. Oxford University Press. 504p.
- [22] BONHOMME R. and CHARTIER P., 1972. The interpretation and automatic measurement of hemispherical photographs to obtain sunlit foliage area and gap frequency. *Israël Journal of Agricultural Research*, 22(2):53–61.
- [23] BOREL C., GERSTL S., and POWERS B., 1991. The radiosity method in optical remote sensing of structured 3D surfaces. *Remote Sensing of Environment*, 36:13–44.
- [24] CAMPBELL G., 1986. Extinction coefficients for radiation in plant canopies calculated using an ellipsoïdal inclination angle distribution. *Agricultural and forest meteorology*, 36:317–321.
- [25] CARLSON T. and RIPLEY D., 1997. On the relation between NDVI, fractional vegetation cover and leaf area index. *Remote Sensing of Environment*, 62:241–252.
- [26] CHAMPION I., WIGNERON J., and BRÉDA N., 1996. Variables de fonctionnement issues des données de télédétection micro-onde. In *Actes de l'école-chercheurs INRA en bioclimatologie*, ed. INRA. volume 2, pp. 145–167.

- [27] CHELLE M., 1997. *Développement d'un modèle de radiosité mixte pour simuler la distribution du rayonnement dans les couverts végétaux*. Ph.D. thesis, Université de Rennes, France.
- [28] CHEN J. and LEBLANC S., 1997. A four-scale bidirectional reflectance model based on canopy architecture. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1316–1337.
- [29] CHOUDHURY B., 1987. Relationships between vegetation indices, radiation absorption and net photosynthesis evaluated by a sensitivity analysis. *Remote Sensing of Environment*, 22:209–233.
- [30] CLEVERS J., 1989. The application of a weighted infrared-red vegetation index for estimating leaf area index by correcting for soil moisture. *Remote Sensing of Environment*, 29:25–37.
- [31] DELÉCOLLE R., LOUBET B., and TCHAMITCHIAN M., 1995. Calibration, sensibilité, validation des modèles. In *Actes de l'école chercheurs INRA en bioclimatologie*, ed. INRA. volume 1, pp. 285–303.
- [32] DESCHAMPS P., BRÉON F., LEROY M., PODAIRE A., BRICAUD A., BURIEZ J., and SÈZE G., 1994. The POLDER mission : instrument characteristics and scientific objectives. *IEEE Transactions on Geoscience and Remote Sensing*, 32(3):598–611.
- [33] DICKINSON R., 1983. Land surface processes and climate-surface albedos and energy balance. *Advanced Geophysics*, 25:305–353.
- [34] DICKINSON R., 1984. *Modeling evapotranspiration for three dimensional global climate models.*, J. Hansen and T. Takahashi, American Geophysical Union, Washington DC, USA, pp. 58–72.
- [35] DRAGO G. and RIDELLA S., 1992. Statistically controlled activation weight initialization (SCAWI). *IEEE Transactions on Neural Network*, 3(4):627–631.
- [36] ENGELSEN O., PINTY B., VERSTRAETE M., and MARTONCHIK J., 1996. Parametric bidirectional reflectance factor models : evaluation, improvements and applications. *Technical Report EUR 16426*, European Commission, Ispra, Italy.

- [37] ESPAÑA M., 1997. *Simulation de la variation temporelle, directionnelle et spectrale de la réflectance des cultures de maïs à partir d'un modèle dynamique de la structure 3D du couvert*. Ph.D. thesis, Université de Marne-La-Vallée, France.
- [38] EVERITT J., RICHARDSON A., and GAUSMAN H., 1985. Leaf reflectance-chlorophyll relations in buffelgrass. *Photogrammetric Engineering and Remote Sensing*, 51:463–466.
- [39] FINSCHI L., 1996. *An implementation of the Levenberg-Marquardt Algorithm*. Eidgenössische Technische Hochschule Zürich, Zürich. Disponible sur le web à l'adresse : <http://www.ifor.math.ethz.ch/staff/finschi/Papers/LevMar.html>.
- [40] FOURTY T. and BARET F., 1997. Vegetation water and dry matter contents estimated from top of atmosphere reflectance data : a simulation study. *Remote Sensing of Environment*, 61:34–45.
- [41] GOBRON N., PINTY B., VERSTRAETE M., and GOVAERTS Y., 1997. Presentation and application of an advanced model for the scattering of light by vegetation in the solar domain. In *Physical Measurements and Signatures in Remote Sensing*, eds. G. Guyot and T. Phulpin. Balkema, Rotterdam, Courchevel, France, volume 1, pp. 267–273.
- [42] GOEL N., 1988. Models of vegetation canopy reflectance and their use in estimation of biophysical parameters from reflectance data. *Remote Sensing Reviews*, 4(Issue 1):1–213. ISBN : 3-7186-4824-5.
- [43] GOEL N. and DEERING D., 1989. Evaluation of a canopy reflectance model for LAI estimation through its inversion. *IEEE Transactions on Geoscience and Remote Sensing*, GE23:674–684.
- [44] GOEL N. and THOMPSON R., 1984. Inversion of a vegetation canopy reflectance model for estimating agronomic variables. III : Estimating using only canopy reflectance data as illustrated by the Suits model. *Remote Sensing of Environment*, 15:223–236.
- [45] GOEL N. and THOMSON R., 1984. Inversion of a vegetation canopy reflectance model for estimating agronomic variables. V. Estimation of

- leaf area index and average leaf angle using measured canopy reflectance. *Remote Sensing of Environment*, 14:77–111.
- [46] GUÉRIF M., COURAULT D., and BRISSON N., 1995. Assimilation de données de télédétection dans les modèles de fonctionnement des cultures. In *Actes de l'école-chercheurs INRA en bioclimatologie*, ed. INRA. volume 1, pp. 169–191.
- [47] HALL F., TOWNSHEND J., and ENGMAN E., 1995. Status of remote sensing algorithms for estimation of land surface state parameters. *Remote Sensing of Environment*, 51:138–156.
- [48] HAPKE B., 1981. Bidirectional reflectance spectroscopy. 1. Theory. *Journal of Geophysical Research*, 86(B4):3039–3054.
- [49] HOSGOOD B., JACQUEMOUD S., ANDREOLI G., VERDEBOUT J., PEDRINI G., and SCHMUCK G., 1995. Leaf Optical Properties Experiment 93 (LOPEX93). *Technical Report Report EUR-16095-EN*, European Center, Joint Research Center, Institute for Remote Sensing Applications, Ispra (Italy).
- [50] HUETE A., 1988. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25:285–309.
- [51] HUETE A.R. and LIU H., 1994. An error and sensitivity analysis of the atmospheric and soil correcting variants of the NDVI for the MODIS-EOS. *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):897–905.
- [52] IRON J., WEISMILLER R., and PETERSEN G., 1989. *Theory and applications of optical remote sensing*, Wiley interscience, chapter Soil reflectance, pp. 66–106.
- [53] JACKSON R., 1986. Spectral response of architecturally different wheat canopies. *Remote Sensing of Environment*, 20:43–56.
- [54] JACQUEMOUD S., 1989. Modélisation des propriétés optiques des feuilles. Mémoire de diplôme d'études approfondies, Méthodes physiques en télédétection.

- [55] JACQUEMOUD S., 1993. Inversion of the PROSPECT+SAIL canopy reflectance model from AVIRIS equivalent spectra : Theoretical study. *Remote Sensing of Environment*, 44:282–292.
- [56] JACQUEMOUD S. and BARET F., 1990. PROSPECT : A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34:75–91.
- [57] JACQUEMOUD S., BARET F., and HANOCQ J., 1991. Validation d'un modèle de réflectance spectrale et directionnelle du sol. In *Proceedings of the 5th Colloquium - Physical Measurements and Signatures in Remote Sensing*. ESA 319, Courchevel, France, pp. 427–430.
- [58] JACQUEMOUD S., BARET F., and HANOCQ J., 1992. Modeling spectral and directional soil reflectance. *Remote Sensing of Environment*, 41:123–132.
- [59] JIN Y. and LIU C., 1997. Biomass retrieval from high dimensional active/passive remote sensing data by using artificial neural networks. *International Journal of Remote Sensing*, 18(4):971–979.
- [60] KAUFMAN Y. and SENDRA C., 1988. Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery. *International Journal of Remote Sensing*, 9(8):1357–1381.
- [61] KAUFMAN Y. and TANRÉ D., 1992. Atmospherically resistant vegetation index (ARVI for EOS-MODIS). *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):261–270.
- [62] KERR Y.H., LAGOUARDE J., and IMBERNON J., 1992. Accurate land surface temperature retrieval from AVHRR data with use of an improved split window algorithm. *Remote Sensing of Environment*, 41:197–209.
- [63] KIMES D., RANSON K., and SUN G., 1997. Inversion of a forest backscatter model using neural networks. *International Journal of Remote Sensing*, 18(10):2181–2199.
- [64] KNYAZIKHIN Y., MARTONCHIK J., DINER D., MYNENI R., VERTRAETE M., PINTY B., and GOBRON N., 1998. Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. *Journal of Geophysical Research*, 103(D24):32239–32256.

- [65] KNYAZIKHIN Y., MARTONCHIK J., MYNENI R., DINER D., and RUNNING S., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, 103(D24):32257–32275.
- [66] KOEPKE P. and QUENZEL H., 1979. Turbidity of the atmosphere determined from satellite calculation of optimum viewing geometry. *Journal of Geophysical Research*, 84:7847–7855.
- [67] KUBELKA P. and MUNK F., 1931. Ein betrag zur optik der farbanstriche. *Ann. Tech. Phys.*, 11:593–61.
- [68] KUUSK A., 1985. The hot spot effect of a uniform vegetative cover. *Sovietic Journal of Remote Sensing*, 3(4):645–658.
- [69] KUUSK A., 1991. In *Photon-Vegetation interactions*, Myneni R.B. and Ross J., chapter The hot spot effect in plant canopy reflectance, pp. 139–159. Springer-Verlag edition.
- [70] KUUSK A., 1994. A multispectral canopy reflectance model. *Remote Sensing of Environment*, 50:75–82.
- [71] KUUSK A., 1995. A fast invertible canopy reflectance model. *Remote Sensing of Environment*, 51:342–350.
- [72] LAWRENCE R. and RIPPLE W., 98. Comparisons among vegetation indices and bandwise regression in a highly disturbed heterogeneous landscape : Mount St. Helen, Washington. *Remote Sensing of Environment*, 64:91–102.
- [73] LEROY M. and ROUJEAN J., 1994. Sun and view angle corrections on reflectances derived from NOAA/AVHRR data. *IEEE Transactions on Geoscience and Remote Sensing*, 32(3):684–697.
- [74] LI X. and STRAHLER A., 1986. Geometrical-optical bidirectional reflectance modeling of conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing*, GE(24):906–919.

- [75] LI X. and STRAHLER A., 1988. Modeling the gap probability of a discontinuous vegetation canopy. *IEEE Transactions on geoscience and remote sensing*, 26(2):161–170.
- [76] LOS S., JUSTICE C., and TUCKER C., 1994. A global 1° by 1° NDVI data set for climate studies derived from the GIMMS continental NDVI data. *International Journal of Remote Sensing*, 15(17):3493–3518.
- [77] MAJOR D., SCHAALJE G., WIEGAND C., and BLAD B., 1992. Accuracy and sensitivity analysis of SAIL model-predicted reflectance of maize. *Remote sensing of environment*, 41:61–70.
- [78] MALTHUS T., 1990. Anglo-french collaborative reflectance experiment. report and data. *Technical report*, University of Nottingham. 38p.
- [79] MEJIA C., THIRIA S., TRAN N., and CRÉPON M., 1998. Determination of the geophysical model function of the ERS-1 scatterometer by the use of neural networks. *Journal of Geophysical Research*, 103(C6):12853–12868.
- [80] MINNAERT M., 1941. The reciprocity principle in lunar photometry. *Astrophys. Journal*, 93:403–410.
- [81] MONTEITH J., 1972. Solar radiation and productivity of tropical ecosystems. *Journal of Applied Ecology*, 9:747–766.
- [82] MYNENI R., ASRAR G., and HALL F., 1992. A three-dimensional radiative transfer method for optical remote sensing and plant ecology. *Remote Sensing of Environment*, 41:105–121.
- [83] MYNENI R., HALL F., SELLERS P., and MARSHAK A., 1995. The interpretation of spectral vegetation indexes. *IEEE Transactions on geoscience and remote sensing*, 33(2):481–486.
- [84] MYNENI R., MAGGION S., IAQUINTA J., PRIVETTE J., GOBRON N., PINTY B., KIMES D., VERSTRAETE M.M., and WILLIAMS D., 1995. Optical remote sensing of vegetation : modelling, caveats, and algorithms. *Remote Sensing of Environment*, 51:169–188.
- [85] MYNENI R. and WILLIAMS D., 1994. On the relationship between FA-PAR and NDVI. *Remote Sensing of the Environment*, 49:200–211.

- [86] NABNEY I. and BISHOP C., 1995. Modelling wind direction from satellite scatterometer data. *Technical Report NCRG/95/011*, Aston university, Neural Computing Research Group, Birmingham B4 7ET, United Kingdom. Disponible sur <http://www.ncrg.aston.ac.uk/>.
- [87] NELDER J. and MEAD R., 1987. A simplex method for function minimization. *Computer Journal*, 7:308–313.
- [88] NILSON T., 1971. A theoretical analysis of the frequency of gaps in plant stands. *Agricultural Meteorology*, 8:25–38.
- [89] OLIOSO A., COURAULT D., CHAUKI H., WIGNERON J.P. OTTLÉ C., and VIDAL-MADJAR D., 1997. Estimation de l' évapotranspiration et de la photosynthèse à l'aide de modèles de TSVA et de la température de surface. In *Physical Measurements and Signatures in Remote Sensing*, eds. G. Guyot and T. Phulpin. Balkema, Rotterdam, Courchevel, France, volume 40, pp. 841–851.
- [90] OTTLÉ C. and VIDAL-MADJAR D., 1992. Estimation of land surface temperature with NOAA9 data. *Remote Sensing of the Environment*, 40:27–41.
- [91] PEARSON R. and MILLER L., 1972. Remote mapping of standing crop biomass for estimation of the productivity of the short-grass Prairie, Pawnee National Grasslands, Colorado. In *8th International Symposium on Remote Sensing of Environment*. ERIM, Ann Arbor MI, pp. 1357–1381.
- [92] PINTY B. and VERSTRAETE M., 1991. Extracting information on surface properties from bidirectional reflectance measurements. *Journal of Geophysical Research*, 96(D2):2865–2874.
- [93] PINTY B. and VERSTRAETE M., 1992. GEMI : a non-linear index to monitor global vegetation form satellites. *Vegetatio*, 101:15–20.
- [94] PINTY B. and VERSTRAETE M., 1992. On the design and validation of surface bidirectional reflectance and albedo models. *Remote Sensing of Environment*, 41:155–167.

- [95] PRÉVOT L., BARET F., CHANZY A., OLIOSO A., and *et al.*, 1998. Assimilation of multi-sensor and multi-temporal remote sensing data to monitor vegetation and soil : the Alpilles-ReSeDA project. In *IGARSS'98, International Geoscience and Remote Sensing Symposium*, ed. L. Tsang. Seattle.
- [96] PRICE J., 1990. On the information content of soil reflectance spectra. *Remote Sensing of Environment*, 33:113–119.
- [97] PRIVETTE J., MYNENI R., TUCKER C., and EMERY W., 1994. Invertibility of a 1-D discrete ordinates canopy reflectance model. *Remote Sensing of Environment*, 48:89–105.
- [98] QI J., CHEHBOUNI A., HUETE A., KERR Y., and SOROOSHIAN S., 1994. A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48:119–126.
- [99] QIN W. and GOEL N., 1995. An evaluation of hotspot models for vegetation canopies. *Remote Sensing Reviews*, 13:121–159.
- [100] RAHMAN H., 1998. Inferring spectral surface albedo from bidirectional reflectance measurements. *Geocarto International*, 13(1):41–46.
- [101] RAHMAN H. and DEDIEU G., 1994. SMAC : a Simplified Method for the Atmospheric Correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing*, 16(1):123–143.
- [102] RAHMAN H., PINTY B., and VERSTRAETE M., 1993. Coupled Surface-Atmosphere Reflectance (CSAR) model. 2. Semiempirical surface model usable with NOAA Advanced Very High Resolution Radiometer Data. *Journal of Geophysical Research*, 98(D11):20791–20801.
- [103] RICHARDSON A. and WIEGAND C., 1977. Distinguishing vegetation from soil background information. *Photogrammetry Engineering on Remote Sensing*, 43:1541–1552.
- [104] ROJAS R., 1996. *Neural Networks. A systematic introduction*. Springer-Verlag, Berlin Heidelberg New York. 502pp.
- [105] ROSS J. and MARSHAK A., 1991. *Photon-Vegetation interactions*, Springer-Verlag, chapter Monte-Carlo Methods, pp. 441–464.

- [106] ROSSOW W., GARDNER L., LU P., and WALKER A., 1988. International Satellite Cloud Climatology Project (ISCCP), documentation on cloud data. *Technical Report WMO/TD n° 226*, World Meteorological Organisation, Genève, Suisse. 78 p.
- [107] ROUJEAN J. and BRÉON F., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of the Environment*, 51:375–384.
- [108] ROUJEAN J., LEROY M., and DESCHAMPS P., 1992. A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97(D18):20455–20468.
- [109] ROUJEAN J., LEROY M., PODAIRE A., and DESCHAMPS P., 1992. Evidence of surface bidirectional effects from a NOAA/AVHRR multitemporal data set. *International Journal of Remote Sensing*, 13(4):685–698.
- [110] ROUSE J., HAAS R., SCHELL J., DEERING D., and HARLAN J., 1974. Monitoring the vernal advancement of retrogradation of natural vegetation. *Type III 371pp*, NASA/GSFC, Greenbelt MD.
- [111] ROYER J., OLIOSO A., and COURAULT D., 1995. Les modèles de circulation générale de l'atmosphère. In *Actes de l' école chercheurs INRA en bioclimatologie*, ed. INRA. volume 2, pp. 193–226.
- [112] RUMMELHART D., HINTON G., and WILLIAMS R., 1986. In *Parallel Data Processing*, Rummelhart D. and McClelland J. (Eds), Cambridge, chapter Learning internal representations by error propagation, pp. 318–362. The MIT press edition.
- [113] SARLE W., 1995. Stopped training and other remedies for overfitting. In *of the 27th Symposium on the Interface of computing science and statistics*. pp. 352–360. Disponible sur le web : <ftp://ftp.sas.com/pub/neural/inter95.ps.Z>.
- [114] SELLERS P., 1985. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, Vol 6, n 8:1335–1372.
- [115] SELLERS P., 1987. Canopy reflectance, photosynthesis, and transpiration. II. The role of biophysics in the linearity of their interdependence. *Remote Sensing of the Environment*, 21:143–183.

- [116] SELLERS P., LOS S., TUCKER C., JUSTICE C., DAZLICH D., COLLATZ G., and RANDALL D., 1994. A global 1 by 1 degree NDVI data set for climate studies. Part 2 : the generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing*, 15(17):3519–3545.
- [117] SELLERS P., LOS S., TUCKER C., JUSTICE C., DAZLICH D., COLLATZ G., and RANDALL D., 1996. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part 2 : the generation of global fields of terrestrial biophysical parameters from satellite data. *Journal of Climate*, 9(4):706–737.
- [118] SELLERS P., MINTZ Y., SUD Y., and DALCHER A., 1986. A simple biosphere model (SiB) for use within general circulation models. *Journal of the Atmospheric Sciences*, 43(6):505–531.
- [119] SHIBAYAMA M. and WIEGAND C., 1985. View azimuth and zenith and solar angle effects on wheat canopy reflectance. *Remote Sensing of Environment*, 18:91–103.
- [120] SHIBAYAMA M., WIEGAND C., and RICHARDSON A., 1986. Diurnal patterns of bidirectional vegetation indices for wheat canopies. *International Journal of Remote Sensing*, 7:233–246.
- [121] SMITH J., 1993. LAI inversion using a backpropagation neural network trained with multiple scattering model. *IEEE Transactions on Geoscience and Remote Sensing*, 31(5):1102–1106.
- [122] STEINMETZ V., CAUQUIL B., and BOCQUET F., 1995. Des neurones dans le champagne. *Ingenieries - EAT*, 3:23–28.
- [123] STRAHLER A., BARNSLEY M., D'ENTREMONT R., HU B., LEWIS P., LI X., MULLER J.P., BACKER SCHAAF C., WANNER W., and ZHANG B., 1995. MODIS BRDF/Albedo product : algorithm theoretical basis document. *Technical report*, NASA, Available at <http://ltpwww.gsfc.nasa.gov/MODIS>.
- [124] SUITS G., 1972. The calculation of the directional reflectance of a vegetative canopy. *Remote Sensing of Environment*, 2:117–125.

- [125] TANRÉ D., DEROO C., DUHAUT P., HERMAN M., MORCRETTE J., PERBOS J., and DESCHAMPS P., 1990. Description of a computer code to simulate the satellite signal in the solar spectrum : the 5Scode. *International Journal of Remote Sensing*, 11(4):659–668.
- [126] TARPLEY J., SCHNEIDER R., and MONEY R., 1984. Global vegetation indices from NOAA-7 meteorological satellite. *Journal of Climate Applied Meteorology*, 23:491–494.
- [127] THIRIA S., MEJIA C., and BADRAN F., 1993. A neural network approach for modeling nonlinear transfer functions: application for wind retrieval from spaceborne scatterometer data. *Journal of Geophysical Research*, 98(C12):22827–22841.
- [128] TOUZET C., 1992. *Les réseaux de neurones artificiels. Introduction au connexionisme*. Collection de l'ERIEE.
- [129] TRESP V., NEUMEIER R., and AHMAD S., 1995. *Advances in neural information processing systems 7*, MIT Press, Cambridge MA., chapter Efficient methods for dealing with missing data in supervised learning.
- [130] TUCKER C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment*, 8:127–150.
- [131] VERHOEF W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling : the SAIL model. *Remote Sensing of Environment*, 16:125–141.
- [132] VERHOEF W., 1985. Earth observation modeling based on layer scattering matrices. *Remote Sensing of Environment*, 17:165–178.
- [133] VERSTRAETE M. and PINTY B., 1991. Designing optimal spectral indexes for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, 24(5):1254–1264.
- [134] VERSTRAETE M., PINTY B., and MYNENI R., 1996. Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58:201–214.

- [135] WALTER-SHEA E., PRIVETTE J., CORNELL D., MESARCH M., and HAYS C., 1997. Relations between directional spectral vegetation indices and leaf area and absorbed radiation in alfalfa. *Remote Sensing of Environment*, 61:162–177.
- [136] WALTHALL C., 1997. A study of reflectance anisotropy and canopy structure using a simple empirical model. *Remote Sensing of Environment*, 61:118–128.
- [137] WALTHALL C., NORMAN J., WELLES J., CAMPBELL G., and BLAD B., 1985. Simple equation to approximate the bidirectional reflectance from vegetative canopies and bare soil surfaces. *Applied Optics*, 24(3):383–387.
- [138] WEISS M., BARET F., LEROY M., BÈGUÈ A., HAUTECOEUR O., and SANTER R., 1999. Hemispherical reflectance and albedo estimates from the accumulation of across track sun synchronous datarach. Submitted to the journal of geophysical research.
- [139] WESSELS L. and BARNARD E., 1992. Avoiding false local minima by proper initialization of connections. *IEEE Transactions on Neural Networks*, 3(6):899–905.
- [140] WIEGAND C., RICHARDSON J., ESCOBAR D., and GERBERMANN A., 1991. Vegetation indices in crop assessments. *Remote Sensing of Environment*, 35:105–119.
- [141] YODER B. and PETTIGREW-CROSBY R., 1995. Predicting nitrogen and chlorophyll content and concentration from reflectance spectra (400-2500nm) at leaf and canopy scales. *Remote Sensing of Environment*, 53:199–211.

Liste des symboles mathématiques

| | |
|-------------------|--|
| a_d | Coefficient de diffusion |
| a_λ | Fraction de rayonnement absorbé dans une longueur d'onde |
| $B(\xi)$ | Fonction de phase de rétrodiffusion du sol |
| C_{ab} | Teneur en chlorophylle a et b des feuilles |
| C_e | Teneur en eau d' une feuille |
| C_{ms} | Teneur en matière sèche d'une feuille |
| \mathcal{F} | Fonction de coût |
| d_f | Diamètre d'une feuille |
| E_{ts} | Flux total incident |
| E_v | Flux directionnel dans la direction de visée |
| E_- | Flux diffus descendant |
| E_+ | Flux diffus montant |
| E_s | Flux directionnel incident |
| E_{tv} | Flux total montant |
| $fAPAR$ | Fraction de rayonnement photosynthétiquement actif absorbé |
| $fAPAR(\theta_s)$ | fAPAR instantané |
| F | Fonction de transfert d'un neurone |
| F_d | Fraction de rayonnement diffus incident |
| F_s | Fraction de rayonnement direct incident |
| $f(\theta_f)$ | Fonction de distribution des inclinaisons foliaires |
| $G(\theta)$ | Surface de feuille moyenne projetée dans la direction θ |
| H | Fonction de Chandrasaekar |

| | |
|---------------------------|---|
| H_c | Hauteur du couvert |
| h | Paramètre de rugosité du sol |
| K | Coefficient d'extinction du $NDVI$ pour P_o |
| K_a | Coefficient d'absorption de la feuille |
| K_e | Coefficient d'extinction de E_v |
| k_e | Coefficient d'extinction de E_s |
| k_i | Coefficient d'absorption spécifique des composants de la feuille |
| K_{P_o} | Coefficient d'extinction du LAI pour P_o |
| K_{NDVI} | Coefficient d'extinction du LAI pour le $NDVI$ |
| LAD | Densité foliaire |
| LAI | Indice foliaire |
| $LAI.C_{ab}$ | Teneur en chlorophylle a et b intégrée au niveau du couvert |
| L_l | Luminance réfléchie par un réflecteur diffusant parfaitement lambertien |
| n | Indice de réfraction |
| N | Paramètre de structure des feuilles |
| $NDVI$ | Normalized Difference Vegetation Index |
| $NDVI_\infty$ | Valeur asymptotique du $NDVI$ quand le LAI tend vers l'infini |
| $NDVI_s$ | $NDVI$ du sol nu |
| \mathcal{N} | Noyau d'un modèle de BRDF |
| $P(\xi)$ | Fonction de phase du sol |
| $P_o(0)$ | Fraction de trou au nadir |
| $P_o(\theta_s)$ | Fraction de trou dans la direction solaire |
| $P_o(\theta_v)$ | Fraction de trou monodirectionnelle |
| $P_o(\theta_v, \theta_s)$ | Fraction de trou bidirectionnelle |
| r_1, r_2 | Paramètre de fonction de phase du sol |
| R^2 | Coefficient de détermination |
| $RMSE$ | Erreur au sens des moindres carrés |
| $RMSE^*$ | Normalisation du RMSE par rapport à l'amplitude de la variable considérée |
| $RMSE^{**}$ | Normalisation du RMSE par rapport à la valeur médiane de la variable considérée |
| s | Paramètre de hot spot |
| s_d | Coefficient de diffusion de E_s contribuant à E_- |
| s'_d | Coefficient de diffusion de E_s contribuant à E_+ |
| u_d | Coefficient de diffusion de E_+ contribuant à E_v |

| | |
|---------------------------------------|---|
| S_{eff} | Section efficace d'une feuille |
| T | Paramètre statistique représentatif de la dispersion des points autour de la droite (1:1) |
| \bar{x} | Moyenne de x |
| \hat{x} | Valeur de x estimée |
| x_s, y_s, z_s | Coordonnées du spectre d'albédo de simple diffusion provenant d'une analyse en composantes principales sur un jeu de données expérimentales |
| z | Dimension verticale du couvert |
| κ | Facteur correctif de hot spot |
| λ | Longueur d'onde |
| μ | Gain utilisé dans la procédure de rétropropagation du gradient |
| μ_s | Cosinus de l'angle zénithal solaire |
| μ_v | Cosinus de l'angle zénithal de visée |
| ν_d | Coefficient de diffusion de E_- contribuant à E_v |
| ω_d | Coefficient de diffusion de E_s contribuant à E_v |
| $\omega_s(\lambda)$ | Albédo de simple diffusion du sol |
| φ_f | Angle azimutal des feuilles |
| ρ | Réflectance spectrale du couvert |
| ρ_f | Réflectance spectrale bihémisphérique des feuilles |
| ρ_h | Réflectance spectrale bihémisphérique du couvert |
| ρ_s | Réflectance spectrale bihémisphérique du sol |
| $\rho_h(\theta_s)$ | Réflectance spectrale directionnelle hémisphérique du couvert |
| $\rho_s(\theta_s)$ | Réflectance spectrale directionnelle hémisphérique du sol |
| $\rho(\theta_v, \varphi)$ | Réflectance spectrale hémisphérique directionnelle du couvert |
| $\rho(\theta_v, \theta_s, \varphi)$ | Réflectance spectrale bidirectionnelle du couvert |
| $\rho_s(\theta_v, \theta_s, \varphi)$ | Réflectance spectrale bidirectionnelle du sol |
| σ | Écart type |
| σ_d | Coefficient de rétrodiffusion |
| τ_f | Transmittance spectrale bihémisphérique des feuilles |
| $\bar{\theta}_f$ | Angle d'inclinaison des feuilles |
| θ_f | Angle moyen d'inclinaison des feuilles |
| θ_s | Angle zénithal solaire |
| θ_v | Angle zénithal de visée |

- φ Angle azimutal relatif entre le soleil et la visée
 ξ Angle de phase

Résumé :

L'objectif de ce travail est de développer un algorithme utilisant des réseaux neuronaux pour estimer des variables biophysiques des couverts végétaux à partir des données du capteur VEGETATION : la fraction de trou dans différentes directions de visée, l'indice foliaire, la fraction de rayonnement photosynthétiquement actif absorbé par le couvert, l'angle moyen d'inclinaison des feuilles, et la teneur en chlorophylle intégrée au niveau du couvert. Une base de données a été élaborée à partir de simulations de modèles de transfert radiatif dans la végétation. Ceci nous a permis de développer des réseaux de neurones pour estimer les différentes variables biophysiques considérées. L'estimation se fait à partir des réflectances bidirectionnelles mesurées par VEGETATION durant un cycle orbital, associées à l'angle zénithal solaire lors de l'acquisition de ces données. L'analyse de sensibilité à l'hétérogénéité intrapixellaire montre que les variables de type fraction de trou et fAPAR sont beaucoup moins sensibles que les variables de type LAI. La validation de l'algorithme sur des données expérimentales montre la robustesse de la méthode lorsque le nombre de données bidirectionnelles acquises durant le cycle orbital est faible. Par ailleurs, elle confirme que les performances d'estimation des variables biophysiques sont affectées non seulement par la représentativité du couvert et des conditions d'observation qui lui sont associées dans la base d'apprentissage, mais aussi par les hypothèses inhérentes aux modèles de transfert radiatif utilisés pour générer cette base. Nous mettons également en évidence la sensibilité de l'algorithme à des bruits de natures diverses (mesure, correction atmosphérique, ...). Les résultats obtenus sur les données expérimentales sont satisfaisants et aboutissent à un écart quadratique moyen inférieur à 0,1. L'intérêt des réseaux de neurones apparaît donc clairement, en particulier vis à vis des approches basées sur les indices de végétation.

Abstract :

The objective of this study is to develop an algorithm for vegetation monitoring using neural networks. It consists in estimating canopy biophysical variables from the VEGETATION sensor reflectance data : gap fraction for different view angles, leaf area index, fraction of absorbed photosynthetically active radiation, mean leaf inclination angle and integrated chlorophyll content. A synthetic top of canopy BRDF catalog of homogeneous canopies was built thanks to well known radiative transfer models. This data base was used to train neural networks to estimate the biophysical variables from the solar zenith angle and the associated bidirectional reflectances collected by the sensor during its orbit cycle. We demonstrated that variables such as gap fractions or fAPAR were less sensitive to heterogeneous pixels than variables such as LAI. The validation of the algorithm on experimental data sets showed the robustness of neural network when few reflectance data are acquired (cloud occurrence or low latitudes). Neural nets thus provide gap fraction estimates within an accuracy better than 0.1. The interest of using neural network is demonstrated, especially when compared to vegetation indices based methods. Moreover, it appears that estimates of the biophysical variables depend not only on the representativity of the canopy and associated viewing conditions in the learning data base, but also on the basic assumption made within the radiative transfer model used to generate it. We additionally investigated the sensitivity of the algorithm to noisy data.