

Assimilation de données de télédétection pour le suivi des surfaces continentales : Mise en œuvre sur un site expérimental

THÈSE

présentée et soutenue publiquement le 13 avril 2007

pour l'obtention du

Doctorat de l'Université Paul Sabatier – Toulouse III

(Discipline: Télédétection spatiale et modélisation de la biosphère)

par

Joaquín Muñoz Sabater

Composition du jury

Rapporteurs : Martine Guérif, DR INRA
Mehrez Zribi, CR CNRS

Examineurs : Gianpaolo Balsamo, CEPMMT
Yann Kerr, Ingénieur CNES
Jean-Philippe Gastellu-Etchegorry , Professeur Université Paul Sabatier

Directeur de Thèse : Jean-Christophe Calvet, Météo-France/CNRM, GAME

Assimilation de données de télédétection pour le suivi des surfaces continentales : Mise en œuvre sur un site expérimental

THÈSE

présentée et soutenue publiquement le 13 avril 2007

pour l'obtention du

Doctorat de l'Université Paul Sabatier – Toulouse III

(Discipline: Télédétection spatiale et modélisation de la biosphère)

par

Joaquín Muñoz Sabater

Composition du jury

Rapporteurs : Martine Guérif, DR INRA
Mehrez Zribi, CR CNRS

Examineurs : Gianpaolo Balsamo, CEPMMT
Yann Kerr, Ingénieur CNES
Jean-Philippe Gastellu-Etchegorry , Professeur Université Paul Sabatier

Directeur de Thèse : Jean-Christophe Calvet, Météo-France/CNRM, GAME

Remerciements

Igual que una naranja valenciana está fomada de varias porciones en forma de luna, la consecución de una tesis es la unión de muchos y diversos componentes . Hace tres años y algunos meses llegaba a Toulouse un poco perdido... Ahora, tres años y algunos meses después, salgo de Toulouse con un doctorado y un valioso conocimiento adquirido. En esta sección quiero expresar mi reconocimiento a todos aquellos que han hecho posible esta tesis y facilitado mi vida en Toulouse.

Primeramente comenzaré por la parte técnica, parte necesaria para poder realizar un trabajo científico de esta magnitud. Mi agradecimiento a aquellos que me han empujado a explorar dentro de esta tema desconocido por mi hace poco más de tres años : la asimilación de observaciones de teledetección. En este sentido agradecer a mi director de tesis, Jean-Christophe Calvet, por haberme confiado este tema, así como las valiosas explicaciones y discusiones fructíferas con Lionel Jarlan, Christophe Rudiger y Jean François Mahfouf. Tampoco puedo olvidar las discusiones puntuales y consejos aportados en otros temas tratados dentro de esta tesis con Joel Noilhan, Jean François Bouyssel, Olivier Samain, Kauzar Saleh, Jean Louis Roujean y Sebastian Massart.

Un tercio de nuestras vidas de lunes a viernes (que cuándo no más), discurren al lado de las mismas personas, los compañeros de trabajo. Mi gran merci a las decenas de cafés, pausas y discusiones de innumerables temas con mi compañera de oficina durante tres años, Claire Sarrat, y "vecino de al lado", Pierre Lacarrere, algo así como mis hermana y papá de la météo, que me han hecho sin duda más llevadero el día a día dentro del CNRM. Agradecer de profundo corazón la inestimable ayuda de éste último no sólo en cuestiones con respecto a la météo, sino también en multitud de temas de la vida diaria "hors météo", en especial su admirable apoyo para hacerme un huequecito dentro del mundo del tenis francés, y siempre con un particular pero "très rigolo" sentido del humor. Cómo no, un gran merci también a Aurore Brut, por las agradables consultas post-café y a Pere Quintana por su interminable buen humor. Por último

dentro de esta categoría, mis gracias a Ronan Paugan por los momentos tan agradables y risas durante la hora sagrada de la comida.

Junto al côté científico del doctorado existe una parte "oculta", que aunque menos evidente es fundamental, sin la cual los trámites administrativos y los nunca ausentes problemas con nuestro querido y fiel compañero, el ordenador, se harían mucho más difíciles. En este sentido, un profundo gracias a la siempre maravillosa y simpática acogida de Laurent Jacquin y Daniel Olcese, al buen humor de Regine Darros y por supuesto, a la continua ayuda y sonrisa de Sylvie Donier, Serge Blin, Claude Cheroux y Michel Tyteca, entre otros.

Hace tres años y medio ponía mis dos pies y las cuatro ruedas de mi Ibiza rojo en Toulouse, pronunciando con dificultad lo único que podía decir en francés : " Parlez-vous anglais ? ". Arreglárselas en estas condiciones, hacerse un hueco en la vida diaria, ir a comprar el pan, abrirse una cuenta en un banco e inscribirse en torneos de tenis, no resulta fácil. Un gran y enorme abrazo a todos aquellos que han contribuido a hacer de mi vida en Toulouse una vida fácil y agradable (Pierre Lacarrere, Bertran Decharme, Thierry Elias, Anne Pirani, Irina Sandu, etc.), y sobretodo, de haberme dejado la sensación que volviendo a Toulouse me encontraré tan a gusto como en mi propia casa.

Finalmente, last but not least, gracias al apoyo incondicional de mi familia por cada una de mis decisiones, a pesar de que ello pueda implicar muchos kilómetros y piedras de por medio.

*pour Sara et Mariló
à Joaquín et Lola
para aquellos que creen en mí
a los que siempre estarán ahí : mis fieles amigos...*

Résumé

Le travail réalisé dans le cadre de cette thèse porte sur le thème de l'assimilation des données de télédétection pour le suivi des surfaces continentales. Les variables analysées sont l'état hydrique du sol et la biomasse de la végétation. Une estimation correcte de ces variables est primordiale pour les modèles de prévision numérique du temps et pour les modèles d'écosystèmes, car elles influencent fortement les échanges de chaleur, d'eau et de CO_2 entre la surface et la couche la plus basse de l'atmosphère. L'étude menée dans cette thèse aborde l'analyse de ces deux variables sur un type particulier de surface : la jachère du site expérimental SMOSREX, au sud-ouest de Toulouse. La période d'étude s'étend de 2001 à 2004. Cette période inclut les fortes sécheresses de 2003 et 2004. Plusieurs méthodes d'assimilation sont appliquées aux observations de l'humidité de la surface du sol et de l'indice foliaire (ou Leaf Area Index, LAI) pour corriger les erreurs des estimations de l'humidité de la zone racinaire et de la biomasse de la végétation pouvant être commises par le modèle de surface ISBA-A- g_s . Le modèle ISBA-A- g_s est une version du modèle opérationnel ISBA de Météo-France. Alors que ISBA est forcé par un LAI interpolé à partir des mesures récoltées sur le site, ISBA-A- g_s permet de simuler la biomasse et le LAI de la végétation à partir de la quantité de CO_2 assimilé par photosynthèse.

Ce travail de thèse a été réalisé en deux étapes. Dans une première étape, on a mis en œuvre et comparé plusieurs méthodes d'assimilation. La variable assimilée est l'humidité de la surface du sol et la variable analysée le contenu en eau du sol dans la zone racinaire. Les méthodes d'assimilation considérées sont basées sur des approches séquentielles (filtres de Kalman) et variationnelles. L'objectif est d'identifier une méthode d'assimilation qui soit compétitive en terme de qualité des analyses et, en même temps, en coût de calcul. En effet, la perspective est à terme d'intégrer cette méthode dans une plateforme opérationnelle à une échelle régionale. La deuxième partie a porté sur l'extension à la biomasse de la méthode retenue précédemment, à l'issue de l'intercomparaison. Dans ce cas, observations de l'humidité de la surface et du LAI sont assimilées simultanément. L'objectif est de quantifier les améliorations apportées par l'assimilation à l'humidité de la zone racinaire et à la biomasse de la végétation, et d'évaluer les effets d'une incertitude sur les précipitations.

Mots clés : assimilation, télédétection, humidité du sol, biomasse de la végétation.

Abstract

The research undertaken in the context of this thesis deals with the assimilation of remote sensing data for the monitoring of ground surface variables. The analysed variables are root-zone soil moisture content and above-ground vegetation biomass. An accurate estimation of these variables is essential for the initialisation of Numerical Weather Prediction and ecosystem models, as they strongly influence heat, water and CO₂ exchanges between the land surface and the atmosphere. The studies conducted in this thesis focus on the analysis of these two variables over a particular type of surface : the fallow of the SMOSREX experimental site, located south-west of Toulouse, France. The period under study extents from 2001 to 2004, which includes the severe droughts of 2003 and 2004. Several assimilation methods are applied for the assimilation of surface soil moisture and leaf area index (LAI) observations. These variables are assimilated in order to correct the possible shortcomings in the root-zone soil moisture and the above-ground vegetation biomass estimates obtained from the land surface model ISBA-A-g_s. The ISBA-A-g_s model is a variant of ISBA, the operational land surface model of Météo-France. While ISBA is forced with an interpolated LAI from ground measurements collected over the experimental site, ISBA-A-g_s allows to simulate the biomass and the LAI of the vegetation from the CO₂ assimilated by photosynthesis.

The studies are conducted in two phases. First, several assimilation techniques are implemented and compared. The assimilated variable is the surface soil moisture and the analysed variable is the root-zone water content. The employed assimilation techniques are based on either sequential (Kalman Filters) or variational approaches. The objective is to identify an assimilation method which is competitive in terms of quality of the analyses, and at the same time, in terms of computational cost. The envisaged work is to integrate this method into an operational platform at the regional scale. Second, the assimilation method, identified in the aforementioned comparison exercise, is extended to allow the analysis of the vegetation biomass. In this case, soil surface moisture and LAI observations are jointly assimilated. The objective is to quantify the improvements in the root-zone soil moisture and the above-ground vegetation biomass variables achieved through the assimilation procedure, and to evaluate the impact of having a large uncertainty in the precipitation data.

Keywords : assimilation, remote sensing, soil moisture, vegetation biomass.

Table des matières

Chapitre 1

Introduction

1.1	Contexte général	1
1.2	Le rôle des systèmes d'assimilation	3
1.3	Objectifs	3
1.4	Contexte particulier de l'étude	5
1.5	Plan de manuscrit	5

Chapitre 2

Vers l'assimilation de données dans les modèles de surface

2.1	Analyse de données : concepts	8
2.2	Méthodes d'analyse de données	8
2.2.1	Interpolation classiques, fonctions d'interpolation	8
2.2.1.1	Voisin le plus proche (Proximal/Nearest Neighbours)	9
2.2.1.2	Interpolation linéaire	9
2.2.1.3	Interpolation par polynômes	9
2.2.1.4	Interpolation par splines	11
2.2.1.5	Autres fonctions d'interpolation	11
2.2.1.6	Problèmes de l'analyse par fonctions d'interpolation	11
2.2.2	Interpolation avec ébauche	11
2.2.3	Méthodes statistiques	12
2.3	Estimation statistique linéaire	13
2.3.1	Définition d'un problème d'estimation	13
2.3.2	Cas trivial	13
2.3.3	Estimateurs linéaires à partir de deux observations	14
2.3.3.1	BLUE's : Best Linear Unbiased Estimators	14
2.3.3.2	Moindres carrées pondérées	17
2.3.3.3	Maximum de vraisemblance	17

2.3.3.4	Convergence des trois approches	18
2.3.4	Généralisation	19
2.3.5	Introduction de l'ébauche dans les estimateurs linéaires	21
2.4	Modèles dynamiques	24
2.5	Assimilation de données	26
2.5.1	Filtres de Kalman	26
2.5.2	Méthodes variationnelles	28
2.6	Modèles non-linéaires	30

Chapitre 3

SMOSREX et ISBA-A-g_s

3.1	Le site expérimental. Caractéristiques principales.	34
3.2	Évolution temporelle des variables de forçage atmosphérique	36
3.3	Humidité du sol et biomasse de la végétation	38
3.4	Mesures des instruments de télédétection	40
3.4.1	Températures de brillance en bande L	41
3.4.2	Réflectances	42
3.4.2.1	Réflectances journalières sur SMOSREX	43
3.4.2.2	Indices de végétation	44
3.4.3	Température infrarouge de la surface	45
3.5	Heating Rates	49
3.6	Le modèle de surface	54
3.6.1	ISBA-A-g _s	54
3.6.2	Simulations d'ISBA-A-g _s	56

Chapitre 4

Méthodes d'assimilation des observations de w_g dans ISBA-A-g_s

4.1	Introduction	59
4.2	"From near-surface to root-zone soil moisture using different assimilation techniques."	61
4.2.1	Résumé	61
4.2.2	Article	62
4.2.2.1	Introduction	62
4.2.2.2	Méthodologie	64
4.2.2.3	Résultats et discussion	74
4.2.2.4	Résumé et conclusions	81
4.2.2.5	Bibliographie	82

4.3	Une méthode dérivée de l'EnKF	85
-----	---	----

Chapitre 5

Assimilation des observations de w_g et LAI dans ISBA-A-gs en mode interactif

5.1	Introduction	90
5.2	Assimilation des w_g avec LAI interactif	90
5.2.1	Stratégie défensive	90
5.2.2	Stratégie offensive	93
5.2.3	Fenêtre séquentielle vs fenêtre glissant pour le 1D-VAR simplifié	94
5.2.4	Conclusion	95
5.3	Assimilation d'observations de LAI	96
5.4	"Joint assimilation of surface soil moisture and LAI observations using a simplified 1D-VAR : The SMOSREX case study."	98
5.4.1	Résumé	98
5.4.2	Article	99
5.4.2.1	Introduction	99
5.4.2.2	Méthodologie	100
5.4.2.3	Résultats	106
5.4.2.4	Résumé et discussion	114
5.4.2.5	Conclusion	115
5.4.2.6	Bibliographie	116
5.5	Assimilation d'observations des instruments de télédétection	119
5.5.1	Températures de brillance	119
5.5.1.1	Méthode directe	119
5.5.1.2	Inversion des T_B	122
5.5.2	Réfectances	125
5.5.3	Erreur des pseudo-observations	126
5.5.4	assimilation des produits de télédétection	128
5.5.5	Conclusion	129

Chapitre 6

Conclusions

6.1	Conclusions	131
6.2	Perspectives	134

Bibliographie	135
----------------------	------------

Annexe A Forçage Atmosphérique 2001-2004	141
---	------------

Annexe B Variables manquants du forçage atmosphérique	145
Annexe C Conversion des unités de l'humidité du sol	147
Annexe D Correction des éclairagements et luminances sur SMOSREX	149
Glossaire	151
Table des figures	153
Liste des tableaux	159

Chapitre 1

Introduction

Sommaire

1.1	Contexte général	1
1.2	Le rôle des systèmes d'assimilation	3
1.3	Objectifs	3
1.4	Contexte particulier de l'étude	5
1.5	Plan de manuscrit	5

1.1 Contexte général

Les modèles de prévision numérique du temps (ou Numerical Weather Prediction -NWP-models) utilisent notre connaissance des processus physiques se produisant dans l'atmosphère (formation des nuages, effet des aérosols, etc.) et à la surface (échanges de vapeur d'eau et de chaleur sur les océans et les continents), afin d'améliorer la représentation de la dynamique de l'atmosphère. A Météo-France, les systèmes de prévision numérique ARPEGE (modèle global) et ALADIN (modèle à zone limitée) fournissent des prévisions météorologiques de façon opérationnelle. Pour fonctionner, les NWP ont besoin de connaître un état initial de l'atmosphère ainsi que les conditions à sa limite inférieure. Sur les continents, la limite inférieure de l'atmosphère est le plus souvent constituée par l'interface avec le sol et la végétation. Cette interface est très importante pour les modèles atmosphériques, puisque l'interaction entre les différents éléments qui composent la surface (océans, sol, végétation, zones urbaines) et les premiers mètres de l'atmosphère va conditionner les flux verticaux de chaleur, d'eau, de CO_2 et de quantité de mouvement. Ces flux sont fournis par l'intermédiaire d'un schéma de surface (ou Land Surface Model -LSM-), qui a pour rôle de modéliser les processus biophysiques qui se produisent à la

surface.

Dans cette étude, le modèle ISBA-A-g_S est utilisé. Il s'agit d'une version du LSM opérationnel à Météo-France, le modèle ISBA, permettant de simuler la dynamique de la végétation et les flux de CO₂. Les flux fournis par le LSM sont conditionnés par la nature du sol, par son contenu en eau, et par la végétation (notamment par la quantité de biomasse présente). Ce travail de thèse a été consacré à l'étude et à l'analyse de deux variables clés de la surface : l'humidité de la zone racinaire (dorénavant notée w_2) et la biomasse de la végétation. Considérant l'importance d'avoir une bonne connaissance de la distribution spatiale et temporelle de ces deux variables, leur suivi est un sujet d'étude important. Ne disposant pas d'un réseau de mesure de ces deux variables (elles ne sont mesurées que sur un nombre limité de sites, très rarement de manière pérenne), une solution possible est de les estimer de façon indirecte à partir de mesures plus nombreuses et accessibles. Ainsi, l'humidité du sol peut être estimée par assimilation dans un LSM couplé à un modèle de la couche limite atmosphérique, à partir des observations synoptiques de la température et de l'humidité de l'air mesurés à 2 m de hauteur (Mahfouf, 1991). Cependant, ce lien est très indirect et de forts biais peuvent perturber l'estimation de w_2 . Une alternative possible est l'utilisation de techniques de télédétection, basées sur le bilan d'énergie à la surface, qui est lié plus directement avec le contenu en eau du sol. Les premiers essais ont été effectués par Wetzol *et al.*, (1984), qui a montré la sensibilité de la température de la surface à la variabilité de w_2 , mais seulement pour des sols relativement secs. L'utilisation de capteurs micro-ondes (actifs ou passifs) en bande L a pour avantage de donner accès à l'humidité du sol même en présence de nuages et d'une couverture végétale relativement dense, puisqu'à ces fréquences le signal est relativement peu atténué par l'atmosphère et par la végétation. Ainsi, Schmugge (1983) a montré qu'il est possible d'estimer l'humidité des 5 premiers cm de la surface du sol avec des capteurs micro-ondes actifs ou passifs. Plus récemment, Entekhabi *et al.*, (1995) a montré qu'il est possible d'estimer l'humidité de la surface au-dessous d'une fréquence de 10 GHz. A partir de 1995, de nombreuses études en micro-ondes, notamment en bande C et bande L, ont montré qu'il est possible d'estimer l'humidité du sol à partir de techniques de télédétection. Ces estimations de l'humidité du sol en surface peuvent être reliées au contenu en eau de couches plus profondes, peu ou pas observables par télédétection, en utilisant un LSM (Walker et Houser, 2001, Reichle *et al.*, 2001, Ni-Meister *et al.*, 2006, Muñoz Sabater *et al.*, 2007).

Les estimations de la biomasse de la végétation à partir de mesures de télédétection sont obtenues en les liant à des indices de végétation, tels que le Normalized Difference Vegetation Index

(NDVI) ou le Soil Wetness Variation Index (SWVI) Cayrol *et al.*, (2000). Cependant, une variable plus directement utilisable en météorologie pour les estimations du flux d'évapotranspiration est l'indice foliaire ou Leaf Area Index (LAI). Des estimations de LAI à 1 km de résolution spatiale sont obtenues à partir du MODerate-Resolution Imaging Spectroradiometer (MODIS) (Tian *et al.*, 2002a, Tian *et al.*, 2002b) ou bien à 250 m de résolution par MEdium Resolution Imaging Spectrometer (MERIS) (Baret *et al.*, 2004). Néanmoins, plusieurs LSM ont été couplés avec les modules de fonctionnement de la végétation et les estimations de LAI peuvent être liées à la biomasse de la végétation (Calvet, 2000a). Le suivi de la biomasse a un intérêt en agrométéorologie et pour la surveillance des écosystèmes naturels. En conséquence, l'utilisation conjointe d'un système de télédétection et d'un schéma de surface, permet de relier les quantités mesurées par les instruments de télédétection avec les variables pronostiques d'un schéma de surface de nouvelle génération comme ISBA-A-gs, en particulier, avec deux variables de la surface : w_2 et la biomasse de la végétation.

1.2 Le rôle des systèmes d'assimilation

Même si les techniques de télédétection peuvent fournir une estimation spatiale de plusieurs variables de la surface, elles ne peuvent pas le faire pour toutes et encore moins d'une manière continue. C'est pourquoi nous avons recours aux schémas de surface, capables de simuler l'évolution spatio-temporelle de ces variables sur un domaine déterminé. Le problème est que, habituellement, ces simulations peuvent s'écarter de la réalité à cause des incertitudes portant sur la connaissance des paramètres du sol et de la végétation, le forçage atmosphérique, et aussi à cause des imperfections des modèles. Les prévisions du modèle et les estimations déduites de mesures satellitaires sont deux types d'informations qui peuvent être associées pour obtenir un niveau de précision impossible à obtenir autrement Talagrand (1997). C'est le rôle des schémas d'assimilation qui, par le biais des observations de télédétection forcent les variables pronostiques des LSM. Par exemple, l'observation de l'humidité de la surface dérivée des diffusiomètres des satellites ERS-1 et ERS-2 (European Remote Sensing satellites) (Wagner *et al.*, 2003) a été utilisée par François *et al.*, (2003) pour réinitialiser l'humidité de la zone racinaire.

1.3 Objectifs

Dans cette thèse plusieurs méthodes d'assimilation sont mises en œuvre et appliquées à des variables susceptibles d'être dérivées des observations de télédétection (humidité de la couche superficielle du sol, dorénavant appelée w_g , et LAI) pour réinitialiser w_2 et la biomasse de la

végétation dans le LSM, ISBA-A- g_s (fig. 1.1).

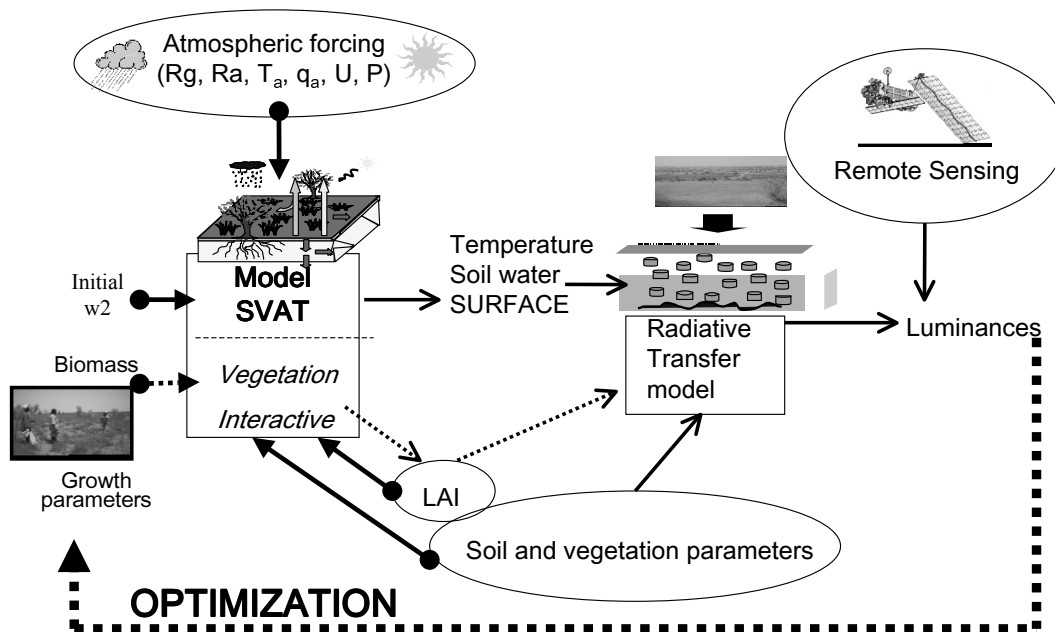


FIG. 1.1: Principe de l'assimilation de données de télédétection pour la réinitialisation des schémas de surface.

Les objectifs principaux sont de :

1. Sélectionner une méthode d'assimilation, permettant d'améliorer les estimations de w_2 et de la biomasse de la végétation du modèle de surface ISBA-A- g_s , basée sur l'information contenue dans les observations de w_g et LAI. Cette méthode doit pouvoir satisfaire aux contraintes opérationnelles : ne pas être excessivement coûteuse en temps de calcul, pouvoir s'appliquer à des LSM différents ou à plusieurs versions d'un LSM ;
2. Déterminer dans quelle mesure l'assimilation d'une seule variable (w_g ou LAI) suffit pour avoir une bonne estimation des deux variables à analyser (w_2 et biomasse) ;
3. Déterminer dans quelle mesure l'assimilation conjointe d'observations de w_g et de LAI améliore le suivi des deux variables à analyser (w_2 et biomasse).

A ces trois objectifs principaux on ajoutera des objectifs secondaires. Par exemple, évaluer si la mesure régulière des réflectances sur jachère nous permet d'obtenir un suivi temporel de l'état de la végétation, étudier l'effet des différents niveaux d'erreurs d'observation et du modèle sur les analyses de l'humidité de la zone racinaire, examiner les effets de l'introduction d'un forçage atmosphérique bruité sur la performance des schémas de surface et d'assimilation, etc.

1.4 Contexte particulier de l'étude

Pour évaluer les améliorations des simulations du LSM par l'assimilation des observations de w_g et LAI, il faut se confronter à la réalité. Pour cela, il est indispensable d'utiliser les données de campagnes de mesures à l'échelle locale. L'échelle locale constitue le domaine d'étude de cette thèse. Les observations utilisées pour l'assimilation et la validation sont issues de l'expérience SMOSREX (Surface Monitoring Of the Soil Reservoir EXperiment)(De Rosnay *et al.*, 2006). SMOSREX est l'une des expériences conduites dans le cadre de la préparation de la future mission satellitaire SMOS (Soil Moisture and Ocean Salinity) (Kerr *et al.*, 2001). SMOS permettra d'accéder à w_g et à sa dynamique à l'échelle globale. Les produits de niveau 2 (w_g) de SMOS pourront être assimilés pour analyser w_2 à une résolution spatiale de l'ordre de 30 km. SMOSREX est aussi une partie intégrante du programme de recherche PIRRENE : Programme Interdisciplinaire de Recherche sur la Radiométrie en ENvironnement Extérieur.

D'autre part, l'importance du développement de l'assimilation des produits de télédétection dans les schémas de surface a été reconnue au niveau Européen. Dans ce contexte, cette thèse a été conduite en lien avec le projet GEOLAND, qui représente la composante " végétation " de la contribution de la Commission Européenne à GMES (Global Monitoring of Environment and Security), dans le cadre du 6^{ième} PCRD. GMES est un projet ambitieux qui a pour but de réunir les compétences européennes pour développer des systèmes globaux de sécurité et de surveillance de l'environnement. L'assimilation des données de télédétection dans les LSM est l'une des priorités de l'une des composantes de GEOLAND : l'observatoire des Flux Naturels de Carbone (fig. 1.2), coordonné par le CNRM. Le suivi des surfaces continentales permettra d'évaluer la variabilité inter-annuelle du cycle de carbone terrestre aux échelles régionales et globales.

1.5 Plan de manuscrit

Le manuscrit est divisé en six chapitres. Suivant cette introduction, le chapitre 2 a pour but de fournir les éléments théoriques de base concernant l'ensemble des méthodes d'assimilation. On y trouvera une brève description des principales méthodes d'analyse de données ainsi que la base de la théorie d'estimation linéaire, à l'origine des méthodes d'assimilation. Le chapitre 3 décrit les types de données utilisées dans ce document. Une brève description des caractéristiques principales du schéma de surface ISBA-A-gs est aussi présentée dans ce chapitre. Les résultats sont présentés dans les chapitres 4 et 5. Le chapitre 4 est consacré au choix d'une méthode d'assimilation. Dans ce chapitre figure un article sous-pressé dans " Journal of Hydrometeorology " qui

The geoland project

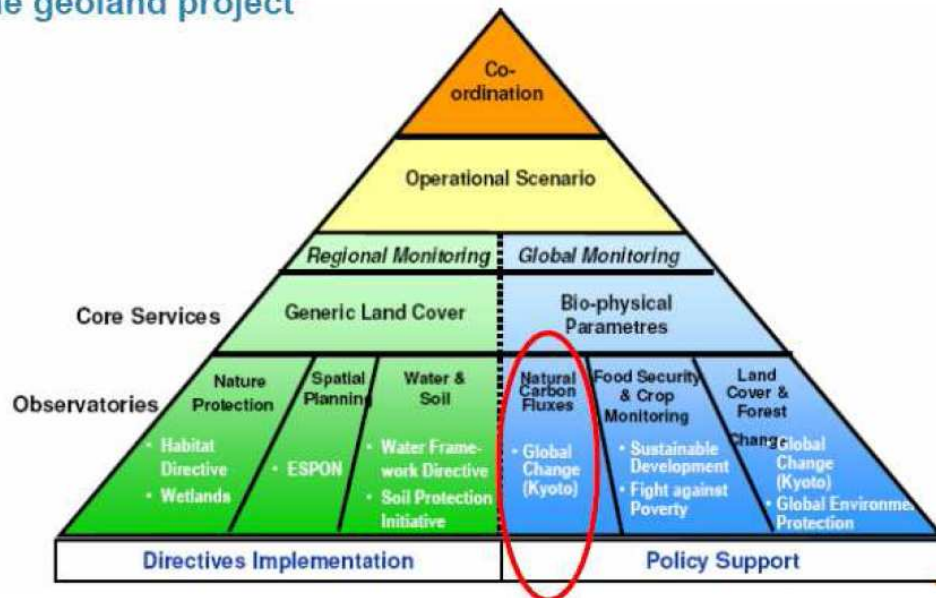


FIG. 1.2: Structure du projet GEOLAND (projet intégré GMES du 6^{ième} PCRD portant sur la végétation).
L'observatoire des flux naturels de carbone est indiqué.

présente une bonne partie des résultats obtenus sur l'application de différentes méthodes d'assimilation d'observations de w_g pour l'analyse de w_2 . De plus, une discussion sur l'implémentation d'une autre méthode d'assimilation est également présentée. Le chapitre 5 aborde l'assimilation d'observations de w_g et de LAI pour l'analyse conjointe de w_2 et de la biomasse de la végétation. Tout d'abord, plusieurs tests sont présentés pour estimer dans quelle mesure l'assimilation d'une seule variable (w_g ou LAI) suffit pour avoir une bonne estimation des deux variables à analyser (w_2 et biomasse). Un deuxième article (soumis) inclus dans ce chapitre aborde la question de l'assimilation conjointe d'observations de w_g et de LAI. La dernière partie du chapitre 5 présente une première évaluation de l'assimilation des produits directement dérivés des instruments de télédétection. Enfin, les conclusions générales et les perspectives de ce travail sont présentées dans le chapitre 6.

Chapitre 2

Vers l'assimilation de données dans les modèles de surface

Sommaire

2.1	Analyse de données : concepts	8
2.2	Méthodes d'analyse de données	8
2.2.1	Interpolation classiques, fonctions d'interpolation	8
2.2.2	Interpolation avec ébauche	11
2.2.3	Méthodes statistiques	12
2.3	Estimation statistique linéaire	13
2.3.1	Définition d'un problème d'estimation	13
2.3.2	Cas trivial	13
2.3.3	Estimateurs linéaires à partir de deux observations	14
2.3.4	Généralisation	19
2.3.5	Introduction de l'ébauche dans les estimateurs linéaires	21
2.4	Modèles dynamiques	24
2.5	Assimilation de données	26
2.5.1	Filtres de Kalman	26
2.5.2	Méthodes variationnelles	28
2.6	Modèles non-linéaires	30

Comme on l'a mentionné au chapitre précédent, les modèles de surface sont capables de fournir une estimation de l'état de la surface par le biais d'un nombre fini de variables pronostiques qui caractérisent le système. Les erreurs ou approximations introduites dans les équations qui paramétrisent le système, dans les variables atmosphériques, dans les paramètres du sol ou de

végétation, etc., font que cette estimation est imparfaite. Néanmoins, les méthodes d'analyses de données permettent d'améliorer l'estimation brute du modèle. Ce chapitre théorique est divisé en deux parties. Dans la première, une classification simplifiée des méthodes d'analyses de données (en mettant l'accent sur les méthodes utilisables pour la surface), est présentée. Dans la deuxième, on se focalise sur les méthodes d'analyse statistiques à travers de la théorie d'estimation linéaire. Cette partie a pour but de décrire, d'une façon simplifiée et pédagogique, les bases de la théorie d'estimation qui aboutissent aux méthodes optimales d'assimilation de données.

2.1 Analyse de données : concepts

Il existe de nombreuses définitions pour l'analyse des données. En général, on pourrait la définir comme un processus mathématique par lequel on essaie de décrire, de la façon la plus réaliste possible, l'état d'un système en combinant toute l'information dont on dispose (modèle, observations, incertitudes, etc.). L'analyse est le résultat de ce processus mathématique. La complexité de la méthode d'analyse utilisée dépend des caractéristiques physiques du système étudié, de la quantité et de la distribution spatio-temporelle des observations, de l'objectif de l'analyse, etc. Le cas le plus simple est celui où l'on fait l'hypothèse que les observations décrivant le système sont parfaites. Dans ce cas, on peut simplement produire une analyse en substituant les observations aux simulations du modèle. C'est le principe de la méthode d'*insertion directe* (Walker, 1999). Cependant, dans la réalité soit (i) les observations ne coïncident pas avec cet état du modèle dans l'espace ou dans le temps, ou bien (ii) à cause du processus d'observation, des erreurs de représentativité ou les erreurs instrumentales, les observations sont imparfaites. On peut classer les méthodes d'analyses de données en fonction de la quantité d'information qu'elles utilisent (observations, incertitudes sur les observations, du modèle, etc.).

2.2 Méthodes d'analyse de données

2.2.1 Interpolation classiques, fonctions d'interpolation

Quand l'état du modèle est complètement caractérisé par les observations (supposées parfaites), le problème d'analyse se réduit à un problème d'interpolation. Des méthodes d'interpolation qui n'utilisent que les observations peuvent être utilisées pour estimer la distribution d'une variable inconnue.

2.2.1.1 Voisin le plus proche (Proximal/Nearest Neighbours)

Il s'agit essentiellement de substituer, pour chaque variable du modèle, chaque point de l'estimation par la valeur de l'observation la plus proche. Dans ce type d'interpolation rentre aussi les polygones de Thiessen (Thiessen, 1911) et les méthodes Pycnophylactic (Tobler, 1979).

2.2.1.2 Interpolation linéaire

L'interpolation linéaire est une technique simple pour laquelle la valeur d'une variable f à un point particulier X est calculée à partir des observations qui l'encadrent dans l'espace et/ou dans le temps. Elle consiste à relier les deux observations, x_0 et x_1 , avec une ligne droite et à en déduire la valeur au point inconnu X par des triangles équivalents de Newton :

$$\frac{f(X) - f(x_0)}{X - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \quad (2.1)$$

$$f(X) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(X - x_0) \quad (2.2)$$

Cependant, le principal problème de l'interpolation linéaire est qu'elle n'est pas très précise. Pour l'interpolation linéaire, l'erreur d'estimation est proportionnelle au carré de la distance entre les observations. D'autres méthodes d'interpolation, comme l'interpolation par polynômes ou "spline interpolation" (décrits ci-dessous), permettent d'obtenir des fonctions d'interpolation plus lisses, où l'erreur d'estimation est inférieure à celle de l'interpolation linéaire.

2.2.1.3 Interpolation par polynômes

L'interpolation par polynômes est la généralisation de l'interpolation linéaire. Dans l'exemple précédent d'interpolation linéaire, on fait une erreur en approximant une courbe avec une ligne droite. Cette estimation peut être améliorée si on introduit certaines courbures à la ligne qui joint les deux observations. Le plus simple est un polynôme d'ordre deux : polynôme quadratique. Pour cela on a besoin de trois observations x_0, x_1 et x_2 . Dans ce cas, en utilisant les triangles de Newton, la valeur d'une variable quelconque au point X est donnée par :

$$f(X) = b_0 + b_1(X - x_0) + b_2(X - x_0)(X - x_1) \quad (2.3)$$

où

$$b_0 = f(x_0) \quad (2.4)$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \quad (2.5)$$

$$b_2 = \frac{\left(\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0} \right)}{x_2 - x_0} \quad (2.6)$$

Cette analyse peut être généralisée en ajustant un polynôme d'ordre n aux $n+1$ observations x_0 à x_n :

$$f_n(X) = b_0 + b_1(X - x_0) + b_2(X - x_0)(X - x_1) + \dots + b_n(X - x_0)(X - x_1) \dots (X - x_{n-1}) \quad (2.7)$$

où

$$b_0 = f(x_0) \quad (2.8)$$

$$b_1 = f(x_1, x_0) \quad (2.9)$$

$$b_2 = f(x_2, x_1, x_0) \quad (2.10)$$

$$\dots \quad (2.11)$$

$$b_n = f(x_n, x_{n-1}, \dots, x_1, x_0) \quad (2.12)$$

Chaque fonction f est, en général, calculée à partir des différences finies :

$$f(x_n, x_{n-1}, \dots, x_1, x_0) = \frac{f(x_n, x_{n-1}, \dots, x_2, x_1) - f(x_{n-1}, x_{n-2}, \dots, x_1, x_0)}{x_n - x_0} \quad (2.13)$$

Les polynômes d'interpolation de Lagrange, plus faciles à programmer, estiment la valeur de la fonction f au point X selon :

$$f_n(X) = \sum_{i=0}^n L_i(X) f(x_i) \quad (2.14)$$

où

$$L_i(X) = \prod_{j=0, j \neq i}^n \frac{X - x_j}{x_i - x_j}$$

Ce type d'interpolation surmonte les problèmes de l'interpolation linéaire, et permet d'obtenir une estimation plus lissée et ayant une erreur d'estimation inférieure à l'interpolation linéaire. Cependant, elles sont généralement coûteuses en temps de calcul et l'interpolation peut ne pas être exacte, notamment aux extrémités ("end points problem" ou "phénomène de Runge", voir par exemple Isaacson et Bishop Keller, 1994).

2.2.1.4 Interpolation par splines

L'interpolation par splines utilise des polynômes de bas-degré par intervalles et choisit les morceaux des polynômes qui s'adaptent le mieux aux observations, c.à.d., les polynômes qui sont les plus proches des observations. Par exemple, l'interpolation cubique par splines est cubique par segments. La fonction interpolante est plus facile à évaluer et sa dérivée seconde est nulle aux extrémités, ce qui évite le problème cité précédemment.

2.2.1.5 Autres fonctions d'interpolation

On peut construire toute une famille de fonctions interpolantes en choisissant différents types de fonctions : fonctions trigonométriques, fonctions rationnelles, etc. Si l'analyse comprend plusieurs variables, il existe des méthodes d'interpolation multivariées : interpolation bilinéaire, trilinéaire, etc. (Isaacson et Bishop Keller, 1994).

2.2.1.6 Problèmes de l'analyse par fonctions d'interpolation

L'analyse par fonctions interpolantes pose plusieurs problèmes : dans les systèmes non-linéaires, l'interpolation linéaire est une façon très brutale d'approximer la valeur d'une variable entre deux observations, surtout si les observations sont distantes dans le temps ou dans l'espace. Dans l'interpolation par polynômes, les extrêmes du champ peuvent donner des valeurs sans aucun sens physique. Un autre aspect important est la représentativité des valeurs interpolées : si la densité des observations est inférieure à celle de la grille du modèle, cela conduit à des analyses linéaires par morceaux. Si un point de la grille du modèle est très éloigné de l'observation la plus proche, l'interpolation sera très incertaine. L'interpolation peut aussi poser des problèmes si, au contraire, la densité d'observations est supérieure à celle de la grille. Les points d'analyse seront le résultat de l'interpolation des points au voisinage des points de grille, tandis que les observations qui sont éloignées des point de grille n'auront aucune influence, de sorte que l'analyse ne prendra pas en compte la distribution des observations.

2.2.2 Interpolation avec ébauche

On rentre ici dans une nouvelle catégorie d'interpolation lorsque on ajoute un terme supplémentaire issu d'une estimation *à priori* ou ébauche. Cela permet de résoudre en partie quelques-uns des problèmes précédents de l'interpolation. La base de ce type d'interpolation est l'*analyse de Cressman*, qui consiste à déterminer un état du modèle qui est égal aux observations au voisinage de l'endroit où elles sont disponibles, et qui est relaxé vers un état arbitraire x_b ailleurs.

Si on dénote x_b l'ébauche d'une variable modélisée x , et y_i un vecteur de p observations de la variable x , l'état du modèle x_a^j en chaque point de grille j est défini par :

$$x_a^j = x_b^j + \frac{\sum_{i=1}^n w_{i,j} [y_i - x_b^j]}{\sum_{i=1}^n w_{i,j}} \quad (2.15)$$

avec

$$w_{i,j} = \max(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}) \quad (2.16)$$

où $d_{i,j}$ est la distance entre les point i et j , et R est appelé "rayon d'influence", de façon que si la distance entre deux points est plus grande que R les observations n'ont pas de poids par rapport à l'ébauche. Les observations plus proches du point de grille ont un poids plus important. A mesure que la distance augmente, le poids des observations diminue. Notons que si le point de grille est au même endroit que l'observation, alors $d_{i,j} = 0$ et $w_{i,j} = 1$. La méthode de *corrections successives* ou de *nudging* (Bratseth, 1986) est une amélioration de la méthode de Cressman, en permettant que la fonction poids w soit inférieure à 1 lorsque $i=j$, et en conséquence relâchant le poids des observations par rapport à l'ébauche. De plus, elles prennent en compte la dimension temporelle de l'assimilation.

Les principaux avantages de ces techniques d'analyse de données sont la simplicité et la rapidité du calcul numérique, et en général elles sont plus précises que d'autres méthodes d'interpolation plus simple. Par contre, elles peuvent provoquer des problèmes de stabilité si la densité des points de grille est plus grande que les observations disponibles. Elles sont aussi assez sensibles aux erreurs d'observation, et ne prennent pas en compte les corrélations spatiales entre les observations. Le principal problème est que l'on ne sait pas bien comment spécifier les poids des corrections de manière optimale et donc ces fonctions sont caractérisées par leur empirisme (Bouttier et Courtier, 1999).

2.2.3 Méthodes statistiques

Les méthodes décrites dans les sections précédentes sont assez faciles à mettre en œuvre. Par contre, elles présentent certains problèmes identifiés précédemment que l'on peut (difficilement) résoudre avec des approches empiriques. Par contre, il est possible de développer des systèmes optimaux qui prennent en compte la qualité de l'information dont on dispose sur le système étudié, l'ébauche et les observations. Ce type de méthodes doit être capable d'accorder davantage de confiance au modèle lorsque les observations sont de mauvaise qualité et inversement. Cette approche probabiliste consiste, fondamentalement, à minimiser la distance qui existe

entre les observations et l'estimation du modèle sous contrainte des informations disponibles. Par exemple, on peut citer la méthode de *Krigeage* (Journel, 1977), dont le principe est le même que pour la méthode de Cressman, mais la fonction de poids est définie en prenant en compte les caractéristiques statistiques des variables à analyser. La section suivante se consacre à l'étude plus détaillée des méthodes d'interpolation statistique. En particulier, on va présenter la théorie statistique d'estimation linéaire qui aboutit aux équations générales d'estimation optimale.

2.3 Estimation statistique linéaire

Le problème de l'estimation des valeurs inconnues d'un processus stochastique relatif à un ensemble d'observations se posait déjà au début du XIX^{ième} siècle, quand le célèbre astronome Karl Friedrich Gauss essayait d'établir l'orbite de la planète Ceres (aujourd'hui un astéroïde). Gauss trouvait cette orbite en ajustant un ensemble d'observations astronomiques $\mathbf{y}^0$ à une fonction mathématique f , et en déterminant les paramètres de f qui minimisaient la somme des carrés des distances entre les observations $\mathbf{y}^0$ et la fonction f . Cette approche par moindres carrés constitue la base et le point de départ pour la recherche des méthodes d'assimilation optimale.

Cette partie du chapitre a pour but d'expliquer, d'une façon pédagogique et simplifiée, des fondements de la théorie d'estimation statistique linéaire qui constitue la base des méthodes d'assimilation de données. Les développements mathématiques seront appliqués à un exemple pratique, pour les rendre plus compréhensibles aux lecteurs qui ne sont pas familiarisés avec cette théorie.

2.3.1 Définition d'un problème d'estimation

Le problème que l'on aborde ici est l'estimation d'un ensemble d'inconnues $\mathbf{x}$ d'un processus aléatoire ou stochastique à partir d'un ensemble d'observations $\mathbf{y}^0$ d'un autre processus, aléatoire lui aussi. Imaginons, par exemple, le parcours d'une route passant à côté d'un village appelé Villarejo, constituant un point dangereux de la route et soumis à une limitation de vitesse fixée à 100 km/h. Deux agents de police contrôlent avec un radar la vitesse v^t des véhicules passant par ce point. L'inconnue x est la vitesse de la voiture v^t . Est-ce que les policiers devront arrêter et verbaliser le véhicule uniquement à partir de la mesure donnée par le radar ?

2.3.2 Cas trivial

A un instant donné les agents de police reçoivent une lecture radar v^0 d'une voiture roulant à une vitesse de 105 km/h. Forts de cette information, les policiers arrêtent la voiture pour un

excès de vitesse, en lui montrant qu'il a roulé à plus de 5 km/h au dessus de la vitesse autorisée. Le conducteur de la voiture affirme qu'il roulait à 97 km/h, selon son tableau de bord. Les agents de police n'ayant pas d'autre source d'information, pour eux, la meilleure estimation $\hat{v}^a$ de la vitesse réelle v^t de la voiture à l'instant t est donnée par v^0 . En conséquence, le conducteur est condamné à une amende pour excès de vitesse. Or, à la mesure v^0 du radar routier est associée une incertitude. Sur les spécifications du radar, suite à un grand nombre de mesures, l'écart type de ces mesures est fixé à 10 km/h. Cet écart type détermine l'erreur absolue d'observation du radar. En conséquence, la vitesse réelle n'est définie qu'à une incertitude près dans l'intervalle compris entre 95 km/h et 115 km/h et, l'automobiliste n'aurait peut-être pas dû être verbalisé.

2.3.3 Estimateurs linéaires à partir de deux observations

Imaginons que la Police ait reçu plusieurs contestations d'amendes établies au même endroit proche de Villarejo. Ils décident donc d'installer un autre radar, indépendant du premier, au même endroit. Ce nouvel appareil, plus précis, a une incertitude de mesure de 5 km/h. Le problème qui consiste à trouver le meilleur estimateur $\hat{v}^a$ de la vitesse de la voiture devient maintenant un peu plus complexe. Les policiers seront tentés maintenant d'utiliser la moyenne des deux mesures radar. Par exemple, si la lecture du deuxième radar donne 98 Km/h, la moyenne des deux observations est de 101.5 Km/h, et le conducteur sera encore verbalisé. Par contre, cette estimation sera incomplète tant qu'ils n'utiliseront pas l'information sur l'erreur des deux radars. Une bonne estimation doit tenir compte des incertitudes sur la mesure. Maintenant on va développer trois approches différentes de l'estimation "optimale" de la vitesse réelle de la voiture v^t avec toute l'information dont nous disposons. On verra que les trois approches coïncident en proposant des estimateurs identiques si on fait des hypothèses sur les moments statistiques et sur la forme de la fonction de distribution de probabilité des erreurs d'observation et d'estimation.

2.3.3.1 BLUE's : Best Linear Unbiased Estimators

Etant donné que les mesures radar ont pour but d'estimer la vitesse réelle v^t de la voiture et sachant que ces observations son imparfaites, on peut alors les exprimer comme suit :

$$v_1^0 = v^t + \epsilon_1 \quad (2.17)$$

$$v_2^0 = v^t + \epsilon_2 \quad (2.18)$$

où ϵ_1 et ϵ_2 représentent les erreurs associées aux moyennes des mesures de v_1^0 et v_2^0 (imperfections des capteurs radar, imperfections sur leur construction, etc.). On ne connaît pas la forme exacte

de la fonction de distribution de probabilité des erreurs ϵ_1 et ϵ_2 . Par contre, on peut définir une classe d'estimateurs optimaux en s'appuyant sur plusieurs hypothèses :

1. supposons que les deux radars sont calibrés de façon que l'on puisse s'assurer que l'espérance mathématique des erreurs ϵ_1 et ϵ_2 est nulle, c'est à dire, que les observations ne sont pas biaisées :

$$E\{\epsilon_1\} = 0 \quad E\{\epsilon_2\} = 0 \quad (2.19)$$

2. puisque chaque radar est indépendant de l'autre, on fait l'hypothèse que les erreurs des mesures ne sont pas corrélées :

$$E\{\epsilon_1 \cdot \epsilon_2\} = 0 \quad (2.20)$$

3. on suppose aussi que les moments d'ordre deux sont égaux aux carrés des écarts types, c'est à dire, la variance des observations :

$$E\{\epsilon_1^2\} = \sigma_1^2 \quad E\{\epsilon_2^2\} = \sigma_2^2 \quad (2.21)$$

L'objectif est de chercher une estimation optimale de la vitesse $\hat{v}^a$ comme une combinaison linéaire des observations v_1^0 et v_2^0 (à laquelle se réfère le " L " de BLUE, de l'anglais " Linear ") :

$$\hat{v}^a = C_1 v_1^0 + C_2 v_2^0 \quad (2.22)$$

où C_1 et C_2 sont des constantes inconnues, qui pondèrent chacune des deux observations. À l'estimation de $\hat{v}^a$ on associe une erreur d'estimation $\hat{\epsilon}^a$, dont l'espérance mathématique s'écrit :

$$\begin{aligned} E\{\hat{\epsilon}^a\} &= E\{\hat{v}^a - v^t\} \\ &= E\{C_1 v_1^0 + C_2 v_2^0 - v^t\} \\ &= C_1 \cdot E\{v^t + \epsilon_1\} + C_2 \cdot E\{v^t + \epsilon_2\} - E\{v^t\} \\ &= E\{v^t\} \cdot (C_1 + C_2 - 1) + C_1 \underbrace{E\{\epsilon_1\}}_0 + C_2 \underbrace{E\{\epsilon_2\}}_0 \\ &= E\{v^t\} \cdot (C_1 + C_2 - 1) \end{aligned} \quad (2.23)$$

On veut des estimateurs non biaisés et en conséquence l'espérance mathématique de l'incertitude de cet estimateur doit être égale à zéro (le " U " de BLUE, de l'anglais " Unbiased "). Ceci ne s'accomplira que si :

$$C_1 + C_2 = 1 \quad (2.24)$$

Cette dernière relation définit une famille infinie d'estimateurs linéaires. On choisira parmi eux celui qui donnera l'erreur d'estimation $\hat{\epsilon}_a$ la plus faible possible. Cette solution sera obtenue en minimisant la variance de $\hat{v}^a$ (le " B " de BLUE, de l'anglais " Best ") :

$$\min(\sigma_{\hat{v}^a}^2) = \min(E\{(\hat{v}^a - v^t)^2\}) \quad (2.25)$$

Considérant que :

$$\sigma_{\hat{v}^a}^2 = E\{(\hat{v}^a - v^t)^2\} \quad (2.26)$$

et étant donné l'éq.(2.22) :

$$\sigma_{\hat{v}^a}^2 = E\{(C_1 v_1^0 + C_2 v_2^0 - v^t)^2\} \quad (2.27)$$

étant donné les eqs. (2.17) et (2.18) :

$$\sigma_{\hat{v}^a}^2 = E\{(C_1 \cdot (v_t + \epsilon_1)^2 + C_2 \cdot (v_t + \epsilon_2)^2 - v^t)^2\} \quad (2.28)$$

étant donné l'éq. (2.19) :

$$\sigma_{\hat{v}^a}^2 = C_1^2 \cdot E\{\epsilon_1^2\} + C_2^2 \cdot E\{\epsilon_2^2\} + 2 \cdot C_1 \cdot C_2 \cdot E\{\epsilon_1 \cdot \epsilon_2\} \quad (2.29)$$

enfin, étant donné les eqs. (2.20), (2.21) et (2.24) :

$$\sigma_{\hat{v}^a}^2 = C_1^2 \cdot \sigma_1^2 + (1 - C_1)^2 \cdot \sigma_2^2 \quad (2.30)$$

Le minimum de la variance de l'estimateur $\hat{v}^a$ s'obtient lorsque on dérive (2.30) par rapport à C_1 (ou par C_2) :

$$\frac{\partial \sigma_{\hat{v}^a}^2}{\partial C_1} = 0 \quad (2.31)$$

Et avec l'expression (2.24) on obtient les valeurs des deux inconnues C_1 et C_2 :

$$C_1^2 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (2.32)$$

$$C_2^2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (2.33)$$

En conséquence l'estimateur linéaire, non biaisé et optimal au sens d'une variance minimum, est obtenu en substituant (2.32) et (2.33) dans (2.22) :

$$\hat{v}^a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \cdot v_1^0 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \cdot v_2^0 \quad (2.34)$$

En remplaçant σ_1 par 10 km/h, σ_2 par 5 km/h, v_1^0 par 105 km/h et v_2^0 par 98 km/h, l'estimation de la vitesse de la voiture est de 99.4 km/h. En conséquence, le conducteur ne doit pas être verbalisé. L'expression (2.34) peut être interprétée en terme de précision. Si on définit la précision comme l'inverse de la variance des observations, (2.34) peut s'écrire :

$$\frac{\hat{v}^a}{\sigma_a^2} = \frac{v_1^0}{\sigma_1^2} + \frac{v_2^0}{\sigma_2^2}; \quad \text{avec} \quad \frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad (2.35)$$

Cet exemple simple met en lumière trois résultats importants : (1) la précision totale est la somme des précisions individuelles, (2) les observations plus précises contribuent avec un poids plus important à l'estimateur optimal, (3) même si on a des observations avec des poids faibles, elles contribuent tout de même à améliorer la précision de l'estimation.

2.3.3.2 Moindres carrés pondérés

Les résultats précédents ont été obtenus à partir d'hypothèses sur la nature des erreurs d'observation, c.à.d., une espérance mathématique nulle et une matrice diagonale. Si on veut éviter de s'appuyer sur des hypothèses probabilistes pour estimer la vitesse de la voiture, on peut aussi utiliser un principe déterministe. Dans ce cas, on minimise une fonction quadratique $\mathcal{J}$ qui mesure la distance aux observations et qui prend en compte la précision de la mesure. Cette fonction, appelée *fonction coût*, prend la forme suivante :

$$\mathcal{J}(v) = \frac{(v - v_1^0)^2}{\sigma_1^2} + \frac{(v - v_2^0)^2}{\sigma_2^2} \quad (2.36)$$

L'estimateur $\hat{v}^a$ qui minimise $\mathcal{J}$ est obtenu par la relation suivante :

$$\frac{\delta \mathcal{J}(\hat{v}^a)}{\delta v} = 0 \quad (2.37)$$

On obtient facilement l'estimateur $\hat{v}^a$:

$$\hat{v}^a = \frac{\sigma_2^2 v_1^0 + \sigma_1^2 v_2^0}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} v_1^0 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} v_2^0 \quad (2.38)$$

qui produit exactement le même estimateur que (2.34).

2.3.3.3 Maximum de vraisemblance

Jusqu'ici on a vu deux approches différentes pour estimer la vitesse v^t de la voiture. L'approche déterministe est basée sur la minimisation d'une fonction quadratique $\mathcal{J}$ qui mesure l'écart entre les observations et l'estimation du modèle. L'approche probabiliste fait des hypothèses sur les moments d'ordre 1 et 2 des erreurs d'observation. Dans notre exemple, les radars routiers estiment la vitesse d'un véhicule en utilisant l'effet Doppler, pour lequel la vitesse du véhicule est proportionnelle à la différence de fréquence entre l'onde émise contre la voiture et l'onde reçue après la réflexion. Cette relation s'écrit (Ostdiek et Bord, 2005) :

$$v^0 = \frac{c}{2\nu} \cdot \Delta\nu^0 \quad (2.39)$$

où c est la vitesse de la lumière, ν la fréquence initiale de l'onde émise par le radar et $\Delta\nu^0$ la différence de fréquence entre l'onde émise et l'onde réfléchie. Imaginons maintenant que les

deux radars soient livrés avec un manuel décrivant les distributions de probabilités complètes $\Psi(v, \Delta\nu^0)$. C'est à dire, que pour une valeur fixe de la vitesse v , Ψ est une fonction de $\Delta\nu^0$. Le problème que l'on se pose ici est de déterminer la valeur correcte de la vitesse v^t à partir de l'observable $\Delta\nu^0$. L'estimation consiste ici à maximiser la vraisemblance de l'inconnu v , cet à dire, maximiser la fonction de distribution de probabilité Ψ qui est associé à l'observable $\Delta\nu^0$.

$$\max_v \Psi(v, \Delta\nu^0) \equiv \max_v \Psi(v, v^0) \quad (2.40)$$

On va considérer le cas particulier où la distribution $\Psi(v, v^0)$ est gaussienne et que la précision de l'observation $\Delta\nu^0$ est égale à $1/\sigma^2$. Dans ce cas :

$$\Psi(v, v^0) = \sigma^{-1} (2\pi)^{-1/2} \cdot \exp\left(-\frac{(v - v^0)^2}{2\sigma^2}\right) \quad (2.41)$$

Dans notre cas des deux observations indépendantes qui visent à estimer la même inconnue v , avec la précision des observations v_1^0 et v_2^0 respectivement égales à $(1/\sigma_1^2)$ et $(1/\sigma_2^2)$, la probabilité conjointe $\Psi(v, v_{12}^0)$ sera le produit des probabilités individuelles :

$$\Psi(v, v_{12}^0) = \Psi(v, v_1^0) \cdot \Psi(v, v_2^0) = \quad (2.42)$$

$$= \underbrace{(2\pi \cdot \sigma_1 \sigma_2)^{-1}}_K \exp\left[-\frac{1}{2} \left(\frac{(v - v_1^0)^2}{\sigma_1^2} + \frac{(v - v_2^0)^2}{\sigma_2^2} \right)\right] \quad (2.43)$$

L'estimateur de la vitesse $\hat{v}^a$ s'obtient en appliquant le critère du maximum de vraisemblance (2.40), ou son équivalent en prenant le logarithme népérien :

$$\max_v \Psi(v, v_{12}^0) \equiv \max_v \ln(\Psi(v, v_{12}^0))$$

Puisque le terme K de (2.43) est constant, le critère du maximum de vraisemblance se réduit à calculer le maximum de $-\frac{1}{2} \left(\frac{(v - v_1^0)^2}{\sigma_1^2} + \frac{(v - v_2^0)^2}{\sigma_2^2} \right)$, soit le minimum de $\frac{(v_1 - v_1^0)^2}{\sigma_1^2} + \frac{(v - v_2^0)^2}{\sigma_2^2}$. On retrouve la même fonction à minimiser que le cas (2.36) sous la contrainte d'utiliser des fonctions gaussiennes.

2.3.3.4 Convergence des trois approches

Dans la section précédente on a traité trois approches différentes pour obtenir une estimation optimale de la vitesse v^t de la voiture. On est arrivé à l'importante conclusion que les trois approches convergent vers les estimateurs linéaires, moyennant quelques hypothèses : dans une approche de type moindres carrés si le poids associé à une observation est égal à l'inverse de la variance (c.à.d., la précision), alors l'estimateur obtenu est équivalent à celui déduit de l'approche BLUE avec observations non biaisées et non corrélées. Egalement, si $\Psi(v, \Delta\nu^0)$ est une

fonction de distribution normale, les estimateurs du maximum vraisemblance sont équivalents aux estimateurs obtenus avec les deux autres approches.

Maintenant, on va généraliser la théorie pour un nombre p d'observations et un vecteur d'estimation composé de n variables.

2.3.4 Généralisation

On va maintenant définir le vecteur $\mathbf{x}$ comme celui qui contient toutes les variables qui décrivent l'état d'un système. Ce vecteur aura pour dimension $n \times 1$ si le nombre des variables d'état est égal à n . Par exemple, on pourra imaginer que pour une application hydrologique, ce vecteur peut être composé par les températures et humidités des différentes couches du sol d'un modèle de surface et en chaque point de grille d'un domaine limité. Concernant les observations, on va maintenant définir le vecteur $\mathbf{y}$ comme celui composé par toutes les observations disponibles et de dimension $p \times 1$. Le vecteur d'état $\mathbf{x}$ et le vecteur d'observations $\mathbf{y}$ sont liés par un opérateur d'observation $\mathcal{H}$, qui généralement est non-linéaire, et lui associe à chaque état $\mathbf{x}$ décrit par ses n composantes, un état observable du système décrit par ses p observations :

$$\mathbf{y} = \mathcal{H}(\mathbf{x}) \quad (2.44)$$

Étant donné que les observations sont une image non parfaite de l'état réel du système $\mathbf{x}^t$, à chaque observation est associée un erreur d'observation ϵ , de dimension $p \times 1$:

$$\underset{[p \times 1]}{\mathbf{y}} = \underset{[p \times n]}{\mathcal{H}}(\underset{[n \times 1]}{\mathbf{x}^t}) + \underset{[p \times 1]}{\epsilon} \quad (2.45)$$

Pour que le système soit déterminé la quantité d'information doit être supérieure ou égale au nombre d'inconnues représentant le vecteur d'état $\mathbf{x}$. Pour simplifier les calculs, on va faire l'hypothèse que la relation (2.45) est linéaire et que la fonction de distribution de probabilité de ϵ est gaussienne :

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x}^t + \epsilon \quad (2.46)$$

$$\Psi_{\epsilon} \sim \mathcal{N}(0, \underset{[p \times p]}{\mathbf{R}}) \quad (2.47)$$

c.à.d., que l'espérance mathématique des erreurs d'observation est égale à zéro et la covariance des erreurs d'observation est représentée par la matrice $\mathbf{R}$ de dimension $p \times p$, dont les termes de la diagonale représentent les variances de chaque observation.

Le problème que l'on veut résoudre maintenant est l'obtention de l'estimateur optimal le plus proche possible de l'état réel du système en utilisant un ensemble d'observations disponibles.

On va d'abord suivre l'approche BLUE décrite en 2.3.3.1. Dans ce cas, cette relation s'écrit :

$$\underset{[n \times 1]}{\hat{\mathbf{x}}^a} = \underset{[n \times p]}{\mathbf{K}} \cdot \underset{[p \times 1]}{\mathbf{y}} \quad (2.48)$$

où $\mathbf{K}$ est la matrice des inconnues de dimension $n \times p$ à déterminer. L'écart entre l'état réel $\mathbf{x}^t$ et l'estimé $\hat{\mathbf{x}}^a$ est défini par ϵ^a . Son espérance mathématique est :

$$E\{\hat{\epsilon}^a\} = E\{\hat{\mathbf{x}}^a - \mathbf{x}^t\} \quad (2.49)$$

$$\begin{aligned} &= E\{\mathbf{K} \cdot \mathbf{y} - \mathbf{x}^t\} \\ &= E\{\mathbf{K} \cdot (\mathbf{H}\mathbf{x}^t + \epsilon) - \mathbf{x}^t\} \\ &= E\{\mathbf{x}^t \cdot (\mathbf{K} \cdot \mathbf{H} - \mathbf{I}) + \mathbf{K} \cdot \epsilon\} \\ &= (\mathbf{K} \cdot \mathbf{H} - \mathbf{I}) \cdot \underbrace{E\{\mathbf{x}^t\}}_{\mathbf{x}^t} + \mathbf{K} \cdot \underbrace{E\{\epsilon\}}_0 \end{aligned} \quad (2.50)$$

Pour que l'estimateur $\hat{\mathbf{x}}^a$ soit non biaisé, l'espérance mathématique de son erreur doit être nulle. On en déduit facilement à partir de la relation (2.50) la condition suivante :

$$\mathbf{K} \cdot \mathbf{H} = \mathbf{I} \quad (2.51)$$

Cette famille d'estimateurs reste encore infinie tant que la matrice $\mathbf{K}$ est inconnue. L'optimalité est obtenue en minimisant la matrice de covariance d'erreur, en accord avec le critère (2.25). À partir des hypothèses faites jusqu'ici, la variance de l'estimateur $\hat{\mathbf{x}}^a$ peut être réécrite comme suit :

$$\sigma_{\hat{\mathbf{x}}^a}^2 = E\{(\hat{\mathbf{x}}^a - \mathbf{x}^t) \cdot (\hat{\mathbf{x}}^a - \mathbf{x}^t)^T\} \quad (2.52)$$

$$\begin{aligned} &= E\{(\mathbf{K}(\mathbf{H}\mathbf{x}^t + \epsilon) - \mathbf{x}^t) \cdot (\mathbf{K}(\mathbf{H}\mathbf{x}^t + \epsilon) - \mathbf{x}^t)^T\} \\ &= E\{[\underbrace{(\mathbf{K}\mathbf{H} - \mathbf{I})\mathbf{x}^t + \mathbf{K}\epsilon}_{(2.51)}] \cdot [\underbrace{(\mathbf{K}\mathbf{H} - \mathbf{I})\mathbf{x}^t + \mathbf{K}\epsilon}_{(2.51)}]^T\} \\ &= \mathbf{K} \cdot \underbrace{E\{\epsilon \cdot \epsilon^T\}}_{(2.47)} \cdot \mathbf{K}^T \\ &= \mathbf{K} \cdot \mathbf{R} \cdot \mathbf{K}^T \end{aligned} \quad (2.53)$$

Le minimum de la variance de la norme de l'estimation d'erreur par rapport à $\mathbf{K}$ est obtenu en annulant sa dérivée. Il est obtenu pour $\mathbf{K}$ (Lewis, 1986) :

$$\mathbf{K} = (\mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H})^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \quad (2.54)$$

Teunissen (1999) donne une interprétation géométrique de cette équation et démontre que le minimum entre les observations et l'état réel du système est la projection oblique du vecteur d'observation sur la direction du vecteur $\mathbf{H}\mathbf{x}$.

Si on récapitule, jusqu'à ici on a obtenu pour le cas général de n variables d'état et p observations un estimateur optimal avec les hypothèses suivantes :

1. la relation entre le vecteur d'état et le vecteur d'observation à travers l'opérateur $\mathcal{H}$, est linéaire ;
2. les erreurs d'observation sont distribuées selon une loi normale.

Avec ces deux contraintes, en substituant (2.54) dans (2.48) on obtient l'estimateur optimal du système :

$$\hat{\mathbf{x}}^a = (\mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H})^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{y} \quad (2.55)$$

L'estimateur (2.55) est obtenu en suivant une approche BLUE. C'est à dire, aux deux hypothèses précédentes il faut ajouter les hypothèses sur les deux premiers moments du vecteur des observables $\mathbf{y}$. De la même façon, on pourrait suivre des approches différentes, voire une approche de moindres carrés (section 2.3.3.2) ou de maximum de vraisemblance (section 2.3.3.3). Pour simplifier le développement théorique de ce chapitre, cette partie n'est pas développée ici, et le lecteur est invité à consulter les démonstrations de Bouttier et Courtier (1999) et Cohn (1997). Dans tous les cas, sous plusieurs hypothèses, les trois types d'estimateurs sont équivalents (voir le tableau-résumé 2.1).

TAB. 2.1: Présentation synthétique des 3 méthodes d'estimation linéaire.

méthode	Principe d'estimation	hypothèses	estimateur $\hat{\mathbf{x}}^a$
BLUE's	$\mathbf{x} = \mathbf{K}\mathbf{y}, E\{\mathbf{x}\} = \mathbf{x}, \sigma_{\mathbf{x}}^2 \min$	$E\{\epsilon\} = 0, E\{\epsilon\epsilon^T\} = \sigma_{\mathbf{x}}^2$	$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}$
Moindres carrés	$\min_{\mathbf{x}} (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x})$	$\mathbf{y} = \mathbf{H}\mathbf{x} + \epsilon$	$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}$
Maximum vraisemblance	$\max_{\mathbf{x}} \Psi(\mathbf{x}, \mathbf{y})$	$\Psi_{\epsilon} \sim N(0, \mathbf{R})$	$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}$

2.3.5 Introduction de l'ébauche dans les estimateurs linéaires

Imaginons maintenant un schéma de surface bi-couche appliqué à toute la France. On considère que le nombre de variables d'état est réduit à quatre par point de grille : humidités et températures de deux couches à chaque point de grille, avec une résolution de quelques kilomètres. On peut imaginer que le vecteur d'état sera assez grand et que le nombre d'observations peut être inférieur au nombre de variables d'état. De plus, ces observations sont irrégulières dans le temps et dans l'espace. Pour résoudre le problème d'estimation, il faut fournir d'autres sources d'information qui permettent de déterminer le problème d'estimation. Typiquement, cela peut être possible grâce à l'utilisation d'une information *a priori*, par exemple une simulation précédente du modèle de surface, qui constitue une première solution, et non la seule, de l'état

de la surface et de la même dimension. En ajoutant les observations à l'ébauche du modèle, l'information totale est forcément de dimension supérieure à la dimension de l'état du système, et en conséquence, le problème d'estimation devient déterminé. Etant donné que l'ébauche est aussi une estimation imparfaite de l'état réel de notre système, on peut l'exprimer comme suit :

$$\underset{[n \times 1]}{\mathbf{x}^b} = \underset{[n \times n]}{\mathbf{I}} \cdot \underset{[n \times 1]}{\mathbf{x}^t} + \underset{[n \times 1]}{\boldsymbol{\epsilon}_b} \quad (2.56)$$

On remarque que dans ce cas l'équivalent de l'opérateur d'observation $\mathbf{H}$ est la matrice identité de même dimension que le nombre de variables d'état. De plus, on pourra se servir d'autres sources d'information disponibles, comme des mesures des variables de la surface fournies par des capteurs satellitaires. Pour l'exemple que l'on a exposé dans ce chapitre, on pourrait imaginer que la Police souhaite mettre en place un système automatique de contrôle de vitesse sur tous les points noirs du réseau routier français. Imaginons que l'on ait 1000 points très fréquentés à contrôler entre autoroutes et routes nationales. Le budget disponible ne permettant pas la mise en œuvre de 1000 radars de la dernière technologie, nous allons ajouter aux 50 radars dont on dispose et aux autres mesures manuelles de toutes sortes, le calcul de vitesses déduites du système GPS (Global Positioning System). L'idée est d'obtenir la plus grande quantité d'information possible avec une estimation de leur incertitude.

Pour prendre en compte une ébauche dans les équations que l'on a développées ultérieurement, on va ajouter l'exposant ⁽⁰⁾ à tous les vecteurs et matrices qui se réfèrent aux observations, $\mathbf{z}$ sera le vecteur qui rassemble tout les types d'information (ébauche et observations), $\mathcal{H}$ et $\boldsymbol{\epsilon}$ représentent respectivement les opérateurs d'observation et la matrice d'erreur de covariance d'ébauche et d'observation :

$$\underset{[n+p \times 1]}{\mathbf{z}} = \begin{bmatrix} \underset{[n \times 1]}{\mathbf{x}^b} \\ \underset{[p \times 1]}{\mathbf{y}^0} \end{bmatrix}; \quad \underset{[n+p \times n]}{\mathcal{H}} = \begin{bmatrix} \underset{[n \times n]}{\mathbf{I}} \\ \underset{[p \times n]}{\mathcal{H}'} \end{bmatrix}; \quad \underset{[n+p \times 1]}{\boldsymbol{\epsilon}} = \begin{bmatrix} \underset{[n \times 1]}{\boldsymbol{\epsilon}^b} \\ \underset{[p \times 1]}{\boldsymbol{\epsilon}^0} \end{bmatrix} \quad (2.57)$$

Comme dans le paragraphe précédent, on fait l'hypothèse de linéarité de l'opérateur d'observation $\mathcal{H}$. Ceci conduit à la relation du même type que (2.46) :

$$\begin{bmatrix} \underset{[n \times 1]}{\mathbf{x}^b} \\ \underset{[p \times 1]}{\mathbf{y}^0} \end{bmatrix} = \begin{bmatrix} \underset{[n \times n]}{\mathbf{I}} \\ \underset{[p \times n]}{\mathbf{H}} \end{bmatrix} \cdot \underset{[n \times 1]}{\mathbf{x}^t} + \begin{bmatrix} \underset{[n \times 1]}{\boldsymbol{\epsilon}^b} \\ \underset{[p \times 1]}{\boldsymbol{\epsilon}^0} \end{bmatrix} \quad (2.58)$$

Finalement, par rapport à la distribution de probabilité des erreurs d'ébauche et pour sim-

plifier les calculs, on fait la même hypothèse que pour (2.47) :

$$\Psi_{\epsilon^b} \sim \mathcal{N}(0, \mathbf{P}_{[n \times n]}) \quad (2.59)$$

où $\mathbf{P}$ représente la matrice de covariance de l'ébauche. Pour en déduire des estimateurs optimaux, on applique une approche BLUE dont les hypothèses de probabilité sur la matrice d'erreur sont :

$$\bullet \quad E\{\epsilon\} = E\left\{\begin{bmatrix} \epsilon^b \\ \epsilon^0 \end{bmatrix}\right\} = 0 \quad (2.60)$$

$$\bullet \quad E\{\epsilon\epsilon^T\} = E\left\{\begin{bmatrix} E\{\epsilon^b(\epsilon^b)^T\} & E\{\epsilon^b(\epsilon^0)^T\} \\ E\{\epsilon^0(\epsilon^b)^T\} & E\{\epsilon^0(\epsilon^0)^T\} \end{bmatrix}\right\} = \begin{bmatrix} \mathbf{P} & 0 \\ 0 & \mathbf{R} \end{bmatrix} = \tilde{\mathbf{R}} \quad (2.61)$$

On veut un estimateur linéaire avec le vecteur d'observation généralisé $\mathbf{z}$:

$$\hat{\mathbf{x}}^a = \mathbf{K} \cdot \mathbf{z} = \mathbf{K} \cdot \begin{bmatrix} \mathbf{x}^b \\ \mathbf{y}^0 \end{bmatrix} \quad (2.62)$$

Les mêmes arguments que ceux de la section précédente sont encore valables ici pour déterminer l'expression de la matrice inconnue $\mathbf{K}$, spécifiée par la relation (2.54). En substituant cette expression dans (2.62) on obtient :

$$\begin{aligned} \hat{\mathbf{x}}^a &= (\mathbf{H}^T \tilde{\mathbf{R}}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \tilde{\mathbf{R}}^{-1} \cdot \begin{bmatrix} \mathbf{x}^b \\ \mathbf{y}^0 \end{bmatrix} \\ &= \left(\begin{bmatrix} \mathbf{I} \\ \mathbf{H} \end{bmatrix}^T \begin{bmatrix} \mathbf{P}^{-1} & 0 \\ 0 & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{H} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{H} \end{bmatrix}^T \begin{bmatrix} \mathbf{P}^{-1} & 0 \\ 0 & \mathbf{R}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}^b \\ \mathbf{y}^0 \end{bmatrix} \\ &= (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \cdot (\mathbf{P}^{-1} \mathbf{x}^b + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^0) \end{aligned} \quad (2.63)$$

On peut faire une interprétation importante de l'estimateur $\hat{\mathbf{x}}^a$ si dans le deuxième terme de l'expression précédente on ajoute et on retranche $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{x}^b$. De cette façon (2.63) devient :

$$\hat{\mathbf{x}}^a = (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \cdot [\mathbf{x}^b \cdot (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) + \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^0 - \mathbf{H} \mathbf{x}^b)] \quad (2.64)$$

$$= \mathbf{x}^b + (\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^0 - \mathbf{H} \mathbf{x}^b) \quad (2.65)$$

De cette façon on peut interpréter que la meilleure estimation de $\hat{\mathbf{x}}^a$ est d'ajouter à l'ébauche $\mathbf{x}^b$ un terme qui est proportionnel à la différence entre le vecteur d'observation $\mathbf{y}^0$ et l'image de l'ébauche dans l'espace des observations. Ce terme est appelé vecteur d'innovation. Le terme $(\mathbf{P}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \cdot \mathbf{H}^T \mathbf{R}^{-1}$ s'appelle la matrice de gain et on remarque qu'il est composé

uniquement des erreurs d'observation et d'ébauche, c'est à dire, l'innovation est multipliée par un facteur qui a un poids statistique optimal en fonction des incertitudes des différentes sources d'information. Dans l'assimilation de données, il est très fréquent de ré-écrire l'équation (2.64) de la façon suivante (Bouttier et Courtier, 1999) :

$$\hat{\mathbf{x}}^a = \mathbf{x}^b + \mathbf{H}\mathbf{P}^T(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}^0 - \mathbf{H}\mathbf{x}^b) \quad (2.66)$$

Pour démontrer cette relation il faut faire la transformation suivante :

$$\begin{aligned} \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}) &= (\mathbf{P}^1 + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \mathbf{P} \mathbf{H}^T \\ &= \mathbf{H}^T + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{P} \mathbf{H}^T \end{aligned}$$

Le même estimateur est obtenu lors d'une approche de type moindres carrés ou de maximum de vraisemblance. Pour le premier, il suffit de redéfinir la fonction coût $\mathcal{J}$ pour prendre en compte l'ébauche selon :

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2} \cdot [(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^0 - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^0 - \mathbf{H}\mathbf{x})] \quad (2.67)$$

où on a utilisé la matrice $\mathbf{B}$ (de l'anglais "background") pour définir la matrice de covariance d'ébauche pour des fonctions variationnelles. Ensuite, le gradient de cette fonction par rapport à $\mathbf{x}$ est calculé. Le lecteur peut consulter Desroziers (2001) pour une description plus détaillée. Pour le maximum de vraisemblance, on fait également l'hypothèse que l'erreur d'ébauche suit une distribution gaussienne. Dans ce cas, la fonction conjointe de probabilité s'écrit :

$$\begin{aligned} \Psi(\boldsymbol{\epsilon}^b, \boldsymbol{\epsilon}^0) &= \Psi(\boldsymbol{\epsilon}^b) \cdot \Psi(\boldsymbol{\epsilon}^0) \\ &= (2\pi)^{-(p+n)/2} (\mathbf{P}\mathbf{R})^{-1/2} \\ &\quad \cdot \exp \left[-\frac{1}{2} \left((\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{-1}(\mathbf{x} - \mathbf{x}^b) + (\mathbf{y}^0 - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y}^0 - \mathbf{H}\mathbf{x}) \right) \right] \end{aligned} \quad (2.68)$$

Le maximum de cette fonction de distribution de probabilité est obtenue en minimisant l'expression entre crochets. Le résultat est similaire aux estimateurs BLUE et variationnel en minimisant la fonction $\mathcal{J}$ de (2.67).

2.4 Modèles dynamiques

Jusqu'ici on a développé d'une façon simplifiée une théorie pour l'obtention des estimateurs statistiques linéaires à partir d'un ensemble d'observations, et aussi d'autres informations dont on dispose comme, par exemple, une ébauche issue d'un modèle dynamique. Toutes ces informations étaient disponibles au même temps i d'observation. Mais comment déterminer l'estimateur

optimal d'un état si les observations sont disponibles sur une échelle de temps non-fixe, avec un système qui évolue dans le temps ? Dans notre exemple initial, imaginons que 2 km après Villarejo il existe un autre radar routier d'une ancienne génération, peu précis. Il est souhaitable de compléter la mesure de ce radar par une estimation *a priori* en utilisant l'autre radar, plus moderne. Cette partie de la route présente une descente en ligne droite, ce qui nous permet de supposer que les véhicules suivent un mouvement rectiligne uniformément accéléré (MRUA). On peut ainsi calculer *a priori* la vitesse v^t du véhicule selon l'équation des MRUA : $v(2 \text{ km}) = v(0 \text{ km}) + a \cdot [t(2 \text{ km}) - t(0 \text{ km})]$, où a est l'accélération de la voiture. Qu'elle est la meilleure estimation de la vitesse de la voiture au deuxième contrôle ?

Le problème que l'on se pose ici est l'estimation de l'état réel $\mathbf{x}^t$ d'un système dynamique à chaque fois qu'il y a une observation ou un jeu d'observations disponibles (estimation séquentielle) ou à certains instants (estimation temporelle). Les développements mathématiques sont très semblables à ceux que l'on a fait auparavant. Ainsi, l'état du système au temps de mesure $i + k$ est obtenu à partir de :

$$\mathbf{x}_{i+k} = \mathcal{M}(\mathbf{x}_i) + \omega_i; \quad \Psi_{\omega_i} \sim \mathcal{N}(0, \mathbf{Q}_i)_{[n \times n]} \quad (2.70)$$

où $\mathcal{M}$ représente le modèle mathématique qui permettra de passer de l'état $\mathbf{x}_i$ à l'état $\mathbf{x}_{i+k}$ et w_i est le terme qui prend en compte toutes les imperfections du modèle de propagation dynamique, que l'on suppose de moyenne nulle et de covariance $\mathbf{Q}$. Le modèle $\mathcal{M}$ est généralement non-linéaire, mais dans un premier temps on va le considérer comme étant linéaire. Concernant le vecteur qui rassemble toutes les informations sur les observations et les informations *a priori*, il est lié à l'état réel du système au temps d'observation i selon la relation :

$$\begin{bmatrix} \mathbf{x}_i^b \\ \mathbf{y}_i^0 \end{bmatrix}_{[n+p \times 1]} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_i \end{bmatrix}_{[n+p \times n]} \cdot \mathbf{x}_i^t_{[n \times 1]} + \begin{bmatrix} \boldsymbol{\epsilon}_i^b \\ \boldsymbol{\epsilon}_i^0 \end{bmatrix}_{[n+p \times 1]} \quad (2.71)$$

Comme on a fait jusqu'ici, on considère que les distributions d'erreurs d'ébauche et d'observations sont gaussiennes, avec une espérance mathématique nulle et les matrices de covariance à l'instant i respectivement égales à $\mathbf{P}_i$ et $\mathbf{R}_i$:

$$\Psi_{\boldsymbol{\epsilon}_i^b} \sim \mathcal{N}(0, \mathbf{P}_i)_{[n \times n]}; \quad \Psi_{\boldsymbol{\epsilon}_i^0} \sim \mathcal{N}(0, \mathbf{R}_i)_{[p \times p]} \quad (2.72)$$

En appliquant les mêmes hypothèses que dans la section 2.3.5 on arrive rapidement aux estimateurs optimaux (Talagrand, 1997) :

$$\hat{\mathbf{x}}_i^a = \mathbf{x}_i^b + (\mathbf{P}_i^{-1} + \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i)^{-1} \cdot \mathbf{H}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i^0 - \mathbf{H}_i \mathbf{x}_i^b) \quad (2.73)$$

$$= \mathbf{x}_i^b + (\mathbf{H}_i \mathbf{P}_i \mathbf{H}_i^T + \mathbf{R}_i)^{-1} (\mathbf{y}_i^0 - \mathbf{H}_i \mathbf{x}_i^b) \quad (2.74)$$

De même, une approche variationnelle (de moindres carrés), on obtiendra le même estimateur précédent au temps i en minimisant la fonction coût suivante :

$$\mathcal{J}(\mathbf{x}_i) = \frac{1}{2} \cdot [(\mathbf{x}_i - \mathbf{x}_i^b)^T \mathbf{B}_i^{-1} (\mathbf{x}_i - \mathbf{x}_i^b) + (\mathbf{y}_i^0 - \mathbf{H}_i \mathbf{x}_i)^T \mathbf{R}_i^{-1} (\mathbf{y}_i^0 - \mathbf{H}_i \mathbf{x}_i)] \quad (2.75)$$

et l'estimateur optimal sera à l'évidence celui donné par (2.73).

2.5 Assimilation de données

L'assimilation de données est aussi un problème d'estimation. Elle rassemble toute l'information fournie par les observations et le modèle simulant la dynamique du système pour décrire, de façon la plus cohérente et précise possible, l'état d'un système. L'assimilation de données est déjà utilisée dans de nombreuses applications, notamment dans les domaines de la météorologie et de l'océanographie. Son utilisation en hydrologie est plus récente. La majorité des méthodes d'assimilation est fondée, en réalité, sur l'application de l'estimation statistique linéaire, décrite précédemment. Ces méthodes ont été développées sous deux formes différentes, pour l'essentiel :

- si l'état du système est analysé séquentiellement chaque fois que l'on trouve une observation ou un ensemble d'observations, on parlera d'**assimilation séquentielle**. Dans cette catégorie entrent toutes les méthodes dérivées du Filtre de Kalman standard ;
- si l'état initial d'un système est ajusté globalement à toutes les observations qui sont disponibles pendant la période d'assimilation, on parlera, généralement, d'**assimilation variationnelle**.

Une caractéristique importante d'une bonne méthode d'assimilation est que non seulement elle fournit une estimation "optimale" de l'état du système, mais qu'elle caractérise aussi l'incertitude avec laquelle cet état a été estimé. Maintenant on va présenter les algorithmes généraux du Filtre de Kalman et des méthodes variationnelles, utilisés plus loin.

2.5.1 Filtres de Kalman

Le problème que l'on veut résoudre ici est l'estimation de l'état d'un système et de son incertitude associée, chaque fois que l'on dispose d'un ensemble d'observations et en utilisant une première estimation du modèle dynamique. Pour être cohérent avec la notation utilisée dans ce document, on va écrire l'état prévu par notre système $\mathbf{x}^f$, où l'exposant (f) fait référence au mot anglais "forecast". La première étape consiste à initialiser le filtre. Ensuite deux étapes suivent : les variables d'état et leur covariance sont propagées en temps, ensuite l'analyse est effectuée.

1. Initialisation du filtre de Kalman :

Imaginons qu'à un instant donné la seule information que l'on a de l'état d'un système est déduite seulement de la prévision d'un modèle dynamique. Par exemple, à partir d'un schéma de surface on pourrait imaginer les champs de température, d'humidité de la surface ainsi que les flux de chaleur et d'humidité échangés avec l'atmosphère. Les erreurs sur la paramétrisation de la physique des processus de surface, sur les variables de forçage atmosphérique, *etc.*, font que cette estimation au temps i est imparfaite. Les écarts par rapport à l'état vrai $\mathbf{x}_i^t$ sont représentés par $\boldsymbol{\epsilon}_i^f$:

$$\mathbf{x}_i^f = \mathbf{x}_i^t + \boldsymbol{\epsilon}_i^f \quad (2.76)$$

Les deux premiers moments sont supposés connus :

$$E\{\boldsymbol{\epsilon}_i^f\} = 0 \quad (2.77)$$

$$E\{\boldsymbol{\epsilon}_i^f (\boldsymbol{\epsilon}_i^f)^T\} = \mathbf{P}_0^f \quad (2.78)$$

Avec cette seule information, la meilleure estimation de l'état du système au temps i est l'ébauche (analogie avec la section 2.3.2) et la covariance de l'estimation est égale à la covariance de l'ébauche :

$$\mathbf{x}_i^a = \mathbf{x}_0^f \quad (2.79)$$

$$\mathbf{P}_i^a = \mathbf{P}_0^f \quad (2.80)$$

2. Étape d'évolution jusqu'à la prochaine observation au temps $i + 1$:

Dans cette étape, l'état du système évolue temporellement grâce au modèle physique $\mathcal{M}$, qui en première approximation est supposé linéaire :

$$\mathbf{x}_{i+1}^f = \mathbf{M}\mathbf{x}_i^a + \omega_i \quad (2.81)$$

avec $\omega_i \sim \mathcal{N}(0, \mathbf{Q}_i)$. De même, la matrice de covariance du système est propagée de l'instant i à l'instant $i + 1$ selon la loi de propagation des variances :

$$\mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q}_i \quad (2.82)$$

où $\mathbf{Q}_i$ est la covariance des erreurs dues aux imprécisions du modèle mathématique, selon (2.70).

3. Étape d'analyse :

C'est dans cette étape que l'information des observations est introduite pour être combinée

avec l'estimation du modèle dynamique. Les mêmes arguments que ceux utilisés dans la section 2.4 peuvent être employés ici pour obtenir l'estimateur $\mathbf{x}_{i+1}^a$ (noter que le chapeau a été enlevé pour être en cohérence avec la notation des deux articles présentés plus loin). Une autre approche cherche l'état analysé comme une combinaison linéaire de l'état prévu par le modèle et la différence pondérée entre le vecteur d'observations au temps $i+1$, $\mathbf{z}_{i+1}$, et la prédiction $\mathbf{H}_{i+1} \cdot \mathbf{x}_{i+1}^f$, que l'on appelle *vecteur d'innovation* :

$$\mathbf{x}_{i+1}^a = \mathbf{x}_{i+1}^f + \mathbf{K}_{i+1} \underbrace{(\mathbf{z}_{i+1} - \mathbf{H}_{i+1} \cdot \mathbf{x}_{i+1}^f)}_{\text{innovation}} \quad (2.83)$$

où $\mathbf{K}_{i+1}$ est la matrice du gain inconnu. Brown et Hwang (1997) justifient cette équation de façon probabiliste. De toutes les expressions possibles de $\mathbf{K}_{i+1}$ on cherche celle qui va minimiser l'erreur d'analyse :

$$\min \mathbf{P}_{i+1}^a = \min E\{(\mathbf{x}_{i+1}^a - \mathbf{x}_{i+1}^f)(\mathbf{x}_{i+1}^a - \mathbf{x}_{i+1}^f)^T\} \quad (2.84)$$

Pour cela, on remplace dans la formule (2.84) $\mathbf{x}_{i+1}^a$ par son expression trouvée en (2.83) et on annule la dérivée par rapport à $\mathbf{K}_{i+1}$. Ainsi, on obtient la matrice de gain :

$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^f \mathbf{H}_{i+1} (\mathbf{H}_{i+1} \mathbf{P}_{i+1}^f \mathbf{H}_{i+1}^T + \mathbf{R}_{i+1})^{-1} \quad (2.85)$$

Finalement, pour que le Filtre soit complet, une estimation de la matrice de covariance d'erreur d'analyse $\mathbf{P}_{i+1}^a$ doit être fournie :

$$\mathbf{P}_{i+1}^a = E[(\mathbf{x}_{i+1}^a - \mathbf{x}_{i+1}^f)(\mathbf{x}_{i+1}^a - \mathbf{x}_{i+1}^f)^T] = (\mathbf{I} - \mathbf{K}_{i+1} \mathbf{H}_{i+1}) \mathbf{P}_{i+1}^f \quad (2.86)$$

2.5.2 Méthodes variationnelles

Les filtres de Kalman ne sont pas toujours aisément utilisables, surtout lorsque la dimension du vecteur d'état est très grande. Par exemple, la dimension de ce vecteur pour les modèles météorologiques est de l'ordre de $n = 10^6 - 10^7$. L'inversion de la matrice $\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R}$ de (2.85), de dimension p , est très coûteuse ainsi que l'étape de propagation de la matrice de covariance d'état (2.82). Un autre désavantage des méthodes qui corrigent d'une façon séquentielle une estimation du modèle dynamique, est que les observations à un instant donné n'ont aucune influence sur les analyses précédentes. Les méthodes variationnelles sont une alternative aux méthodes séquentielles, elles évitent les désavantages cités précédemment. Le principe est de corriger l'état initial à un instant donné i , en minimisant l'écart entre la trajectoire du modèle et les observations disponibles sur toute la période d'assimilation, entre les instants i et $i+k$. Sans entrer dans les détails, on présente ci-dessous, comme pour les filtres de Kalman, les étapes générales des méthodes variationnelles :

1. Initialisation de l'état du système :

Bien qu'il existent plusieurs façon d'initialiser l'état initial du système avec une méthode variationnelle, une manière habituelle est de le prendre égal à la prevision du modèle au temps i :

$$\mathbf{x}_i^0 = \mathbf{x}_i^f \quad (2.87)$$

Cependant, la matrice de covariance d'erreur est estimée en suivant plusieurs approches empiriques et est souvent supposée constante pendant toute la période d'étude. Cette matrice est notée $\mathbf{B}$, de l'anglais "background".

2. Propagation des variables pronostiques à l'instant $i + k$ à partir de l'ébauche du modèle au temps i :

$$\mathbf{x}_{i+k}^f = \mathbf{M}_{i,i+k} \cdot \mathbf{x}_i^f \quad (2.88)$$

On remarque que pour simplifier les calculs on a utilisé l'opérateur de propagation linéaire $\mathbf{M}$. Supposons qu'entre l'instant i et l'instant $i + k$ il existe un nombre k d'observations tel que l'opérateur de propagation $\mathbf{M}_{i,i+k}$ soit le produit matriciel des opérateurs linéaires (dans notre cas) entre deux observations :

$$\mathbf{M}_{i,i+k} = \mathbf{M}_{i+k,i+k-1} \circ \mathbf{M}_{i+k-1,i+k-2} \circ \dots \circ \mathbf{M}_{i-j,i-j-1} \circ \dots \circ \mathbf{M}_{i-2,i-1} \circ \mathbf{M}_{i-1,i} \quad (2.89)$$

Chacun de ces opérateurs représente la propagation de l'état du système entre deux observations :

$$\mathbf{x}_{i+j-1}^f = \mathbf{M}_{i+j,i+j-1} \cdot \mathbf{x}_{i+j}^f \quad (2.90)$$

3. Correction de l'état initial $\mathbf{x}_i^f$:

On va maintenant redéfinir les vecteurs $\mathbf{z}$, $\mathbf{H}$ et ϵ de façon que toutes les observations, aux différents pas de temps, soient prises en compte :

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}_i^b \\ \mathbf{y}_{i+1}^0 \\ \vdots \\ \mathbf{y}_j^0 \\ \vdots \\ \mathbf{y}_{i+k}^0 \end{bmatrix}; \quad \mathbf{H} = \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_{i+1}^0 \\ \vdots \\ \mathbf{H}_j^0 \\ \vdots \\ \mathbf{H}_{i+k}^0 \end{bmatrix}; \quad \epsilon = \begin{bmatrix} \epsilon_i^b \\ \epsilon_{i+1}^0 \\ \vdots \\ \epsilon_j^0 \\ \vdots \\ \epsilon_{i+k}^0 \end{bmatrix} \quad (2.91)$$

La relation entre le vecteur d'observation global $\mathbf{z}$ et l'état réel $\mathbf{x}^t$, sous forme matricielle, s'écrit :

$$\mathbf{z} = \mathbf{H}\mathbf{x}^t + \epsilon \quad (2.92)$$

Une dernière hypothèse est que l'on considère, dans un premier temps, que les erreurs de mesure sont indépendantes, ce qui n'est pas toujours vrai. Ces erreurs sont rassemblées dans la matrice $\mathbf{R}$. La matrice d'erreur d'ébauche ou "background" $\mathbf{B}$ reste, habituellement, fixe à chaque pas de temps :

$$\mathbf{B} = \mathbf{B}_i; \quad \mathbf{R} = \begin{bmatrix} \mathbf{R}_{i+1} & \dots & 0 \\ & \ddots & \\ \vdots & \mathbf{R}_j & \vdots \\ & & \ddots \\ 0 & \dots & \mathbf{R}_{i+k} \end{bmatrix} \quad (2.93)$$

Finalement, l'équation de correction de l'état initial a la même forme que pour les filtres de Kalman :

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}(\mathbf{z} - \mathbf{H} \cdot \mathbf{x}_i^f) \quad (2.94)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) \quad (2.95)$$

mais, maintenant, les matrices $\mathbf{H}$ et $\mathbf{R}$ sont définies chaque fois que l'on trouve une observation entre les instants i et $i + k$. Un résumé de ces équations est présenté à la section 4.2.2.2 (méthodologies d'assimilation).

L'équation précédente de correction de l'état initial (ou "background") est obtenue en minimisant la fonction coût de (2.75). Étant donné que les variables de contrôle définissent l'état initial du système, la minimisation est effectuée par rapport à ces variables. Numériquement, cette minimisation est réalisée à travers un processus itératif, où chaque étape nécessite la connaissance explicite des valeurs locales de dérivées partielles de $\mathcal{J}$ par rapport à chaque variable de contrôle. Cette méthode classique de résolution numérique du problème de minimisation de la fonction coût $\mathcal{J}$ est appelée la méthode des "adjoint equations". Cependant, on verra que les méthodes dérivées de principes variationnels utilisées dans cette thèse évitent le calcul explicite du modèle adjoint en faisant plusieurs hypothèses simplificatrices, en particulier, sur l'opérateur d'observation $\mathcal{H}$.

2.6 Modèles non-linéaires

La théorie décrite jusqu'ici nous a fourni une façon de calculer les estimateurs optimaux d'un système en utilisant toute l'information disponible (ébauche, modèle, observations, incertitudes). Elle est basée sur l'hypothèse que le système dynamique et l'opérateur d'observation suivent un

comportement linéaire. Pour eux, les deux premiers moments décrivent complètement le processus. Cependant, ceci est une contrainte forte qui se rapporte à des systèmes idéaux. Dans le monde réel, les processus physiques qui caractérisent l'atmosphère, les océans ou la biosphère continentale sont très non-linéaires. L'estimation non-linéaire est un problème bien plus compliqué. Elle nécessite le calcul de la propagation de la fonction de densité de probabilité (pdf) complète du système, laquelle ne peut généralement pas être décrite avec un nombre fini de paramètres (Julier et Uhlmann, 2004). En pratique, le calcul des estimateurs pour des systèmes non-linéaires repose sur des approximations sur le modèle de propagation dynamique du système ou de la matrice de covariance associée. Le Filtre de Kalman étendu (EKF) où le 1D-VAR simplifié (utilisés dans cette thèse) sont des méthodes de ce type. Pour les systèmes non-linéaires, les équations générales de propagation du système entre les instant de temps i et $i + k$, et du processus d'observation à l'instant $i + k$, sont les suivants :

$$\mathbf{x}_{i+k}^f = \mathcal{M}(\mathbf{x}_i) + \omega \quad (2.96)$$

$$\mathbf{y}_{i+k} = \mathcal{H}(\mathbf{x}_{i+k}) + \epsilon \quad (2.97)$$

où $\mathcal{M}$ représente le modèle dynamique non-linéaire, $\mathcal{H}$ l'opérateur d'observation non-linéaire, et ω et ϵ représentant les erreurs commises par la paramétrisation du modèle dynamique et le processus d'observation, respectivement. Pour les filtres de Kalman, la propagation explicite de la matrice de covariance d'erreur associée à la prévision $\mathbf{x}_{i+k}$ a besoin de certaines hypothèses de linéarité sur l'opérateur $\mathcal{M}$, par exemple, elle peut-être la matrice jacobienne $\tilde{\mathbf{M}}$ de $\mathcal{M}$. Dans ce cas, cette équation s'écrit comme :

$$\tilde{\mathbf{P}}_{i+k}^f = \tilde{\mathbf{M}}\tilde{\mathbf{P}}_i^a\tilde{\mathbf{M}}^T + \tilde{\mathbf{Q}}_i \quad (2.98)$$

Ainsi, $\tilde{\mathbf{P}}_{i+k}^f$ s'exprime comme une fonction du jacobien $\tilde{\mathbf{M}}$ et des approximations des matrices de covariance d'erreur d'analyse $\tilde{\mathbf{P}}_i^a$ et modèle $\tilde{\mathbf{Q}}_i$, obtenues dans une étape précédente. La recherche des estimateurs optimaux dans le cas non-linéaire suit le même raisonnement que pour le cas linéaire. Ainsi, l'estimateur optimal à l'instant $i + k$ est le suivant :

$$\mathbf{x}_{i+k}^a = \mathbf{x}_{i+k}^f + \tilde{\mathbf{K}}_{i+k} \cdot [\mathbf{y}_{i+k}^0 - \mathcal{H}(\mathbf{x}_{i+k}^f)] \quad (2.99)$$

$$\tilde{\mathbf{K}}_{i+k} = \tilde{\mathbf{P}}_{i+k}^f \tilde{\mathbf{H}}_{i+k}^T [\tilde{\mathbf{H}}_{i+k} \tilde{\mathbf{P}}_{i+k}^f \tilde{\mathbf{H}}_{i+k}^T + \mathbf{R}_{i+k}]^{-1} \quad (2.100)$$

où la matrice $\tilde{\mathbf{H}}$ est une approximation de l'opérateur d'observation non-linéaire $\mathcal{H}$, par exemple par une approximation numérique par différences finies comme à la méthode 1D-VAR traitée dans cette thèse. Au chapitre 4 et 5, plusieurs méthodes d'assimilation basées sur la théorie décrite dans ce chapitre seront testées avec des observations de l'humidité de la surface et de

LAI. La section "methodological discussion" à la page 69, discute les principales différences existant entre les quatre méthodes d'assimilation de données utilisées dans cette thèse.

Chapitre 3

SMOSREX et ISBA-A-g_s

Sommaire

3.1	Le site expérimental. Caractéristiques principales.	34
3.2	Évolution temporelle des variables de forçage atmosphérique	36
3.3	Humidité du sol et biomasse de la végétation	38
3.4	Mesures des instruments de télédétection	40
3.4.1	Températures de brillance en bande L	41
3.4.2	Réfectances	42
3.4.3	Température infrarouge de la surface	45
3.5	Heating Rates	49
3.6	Le modèle de surface	54
3.6.1	ISBA-A-g _s	54
3.6.2	Simulations d'ISBA-A-g _s	56

SMOSREX est l'acronyme de "Surface MONitoring of the Soil Reservoir EXperiment". Il s'agit du nom de la campagne de mesure de longue durée dont les données ont été utilisées pour atteindre les objectifs scientifiques décrits dans l'introduction. Ce chapitre a plusieurs objectifs :

1. fournir une brève description des caractéristiques du site instrumental de SMOSREX et de ses objectifs scientifiques principaux (De Rosnay *et al.*, 2006) ;
2. présenter le jeu de données acquis sur SMOSREX, tant les variables qui composent le forçage atmosphérique que les variables modélisées par le LSM et celles obtenues à partir d'instruments de télédétection ;
3. présenter les caractéristiques principales du LSM employé dans ce travail : ISBA-A-g_s.

3.1 Le site expérimental. Caractéristiques principales.

Le site expérimental de SMOSREX est situé à quelques dizaines de kilomètres au sud-ouest de Toulouse, sur le site du Fauga-Mauzac de l'ONERA (Office National d'Études et de Recherches Aérospatiales). Le dispositif instrumental de SMOSREX a été mis en place par le CNRM, le CESBIO et l'ONERA. Les mesures automatiques sur jachère ont été réalisées et analysées par le CNRM (station météorologique et profils d'humidité et de température dans le sol), le CESBIO et l'INRA ont réalisé les mesures manuelles de biomasse et de LAI.

Sur ce site, des observations de télédétection sont réalisées en continu sur deux types de surfaces : sol nu et jachère. Les zones observées par les instruments de télédétection couvrent les deux types de surface sur, approximativement, 40 m par 10 m. Des arbustes contournant ces deux parcelles sont situés suffisamment loin pour considérer qu'il n'ont pas d'effet sur les mesures de télédétection ni sur l'état hydrique du sol. Les instruments de télédétection sont placés sur un portique de mesure à une hauteur de 15 m. À cette distance de la surface, il est inutile de faire des corrections atmosphériques. La jachère de SMOSREX Cette dernière constitue le type d'occupation du sol considéré dans cette étude d'assimilation. Ce type de surface comporte plusieurs espèces de plantes et montre une forte hétérogénéité spatiale qui produit parfois des observations de biomasse très dispersées. Selon la dernière version du Dictionnaire de l'Académie française, la jachère est définie de la façon suivante :

"l'état d'une terre labourable qu'on cesse temporairement d'ensemencer pour la laisser reposer ou, par ext., qu'on laisse sans culture, à l'abandon".

Dans le cas de SMOSREX, il s'agit d'une prairie naturelle qui est coupée une fois par an (en hiver). Le couvert végétal est constitué de plusieurs espèces de plantes herbacées (type C3). L'étude d'observations de télédétection sur prairie est importante du fait que de vastes étendues sont occupées par ce type d'occupation du sol, à l'échelle du globe, sous tous les climats. Par exemple, on pourrait citer les Pampas de l'Amérique du sud ou, en France, de nombreuses régions montagneuses. Le dispositif expérimental est présenté par la fig. 3.1. La partie sud du site occupé par le dispositif expérimental de SMOSREX est consacrée aux mesures sur jachère. La partie nord est consacrée au sol nu, qui n'a pas été traité dans cette étude.

SMOSREX est conçu pour atteindre plusieurs objectifs scientifiques, dont les principaux sont :

1. La modélisation de l'émission micro-onde en bande-L de la prairie et du sol nu ;

2. L'évaluation d'algorithmes d'inversion visant à estimer l'humidité du sol à partir de mesures de températures de brillance en bande L (disponibles en bipolarisation et pour plusieurs angles d'incidence) ;
3. Le développement de l'assimilation multi-spectrale des données de télédétection dans les modèles de surface. Cette thèse est consacrée à ce dernier objectif.

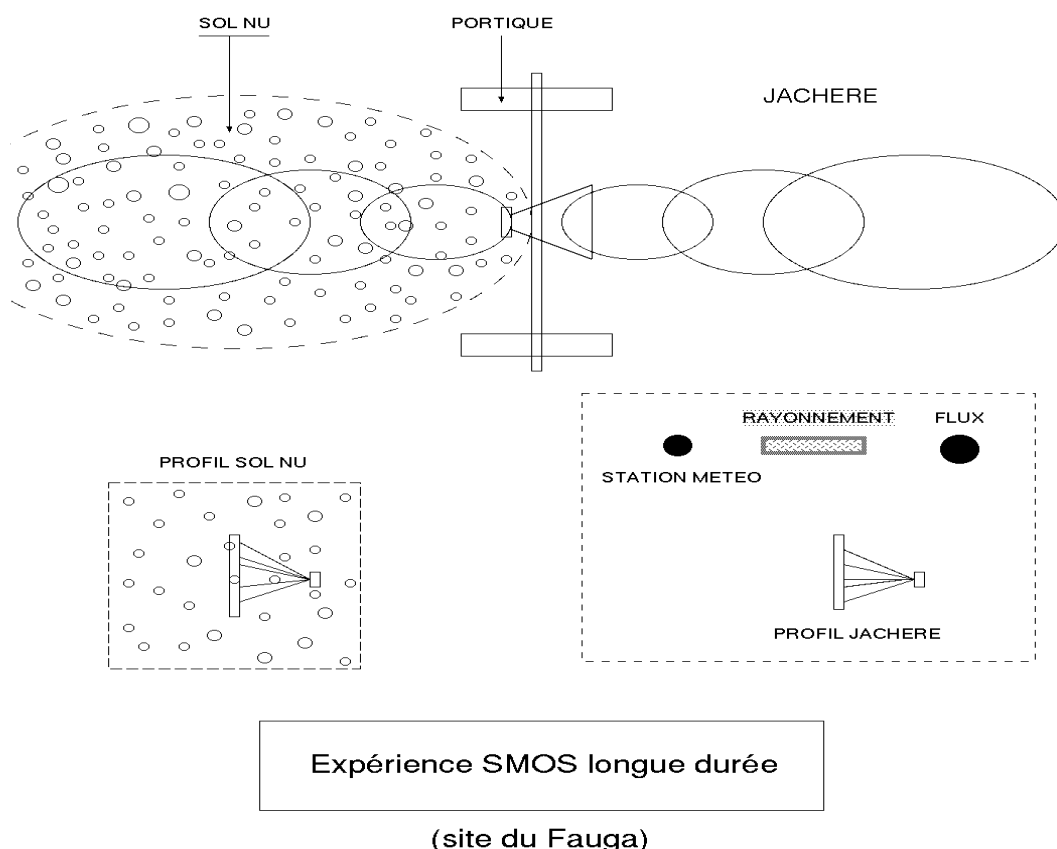


FIG. 3.1: Instrumentation de l'expérience SMOSREX. Les instruments de télédétection (le radiomètre en bande L, LEWIS, le luminancemètre CIMEL et les pyromètres Heitronics) sont installés sur le portique de mesure.

L'un des aspects les plus importants et novateurs de SMOSREX est la longue durée de cette expérience, qui permet de réaliser des études de modélisation et d'assimilation couvrant non seulement les effets saisonniers, mais aussi la variabilité interannuelle. Cet aspect avait été traité en partie par MUREX (Monitoring the Usable soil Reservoir Experimentally), mais sur une période plus courte, et avec un dispositif expérimental moins complet que celui de SMOSREX (Calvet *et al.*, 1999, Calvet et Noilhan, 2000b). SMOSREX fournit des mesures des variables atmosphériques et de surface (sol et végétation) depuis l'année 2001 et de télédétection depuis l'année 2003. Il est envisagé de continuer l'expérience jusqu'en 2008.

L'instrumentation de SMOSREX permet de créer une base continue de données météorologiques et radiométriques. Par rapport à nos objectifs, ces données vont nous permettre : (1) d'imposer au LSM toutes les variables du forçage atmosphérique (y compris le rayonnement infrarouge incident, très rarement mesuré dans les stations météorologiques), (2) de tester et valider plusieurs schémas d'assimilation grâce aux données de télédétection et, (3) de contrôler les variables qui caractérisent la dynamique de la végétation, comme le LAI ou le NDVI (Reed *et al.*, 1994).

3.2 Évolution temporelle des variables de forçage atmosphérique

Dans le cadre de ce travail, un forçage atmosphérique continu dans le temps a été obtenu pour SMOSREX (Annexe A). Cette section présente les résultats de l'analyse de ces données. Les figs. 3.2 et 3.3 montrent l'évolution temporelle des variables du forçage.

On observe, en général, que la moyenne mensuelle du rayonnement solaire incident obtenue pendant les mois de printemps et d'été est plus élevée pour 2003 et 2004, ainsi que le maximum journalier de température de l'air. L'évènement marquant de cette période est la canicule de 2003. La fig. 3.3 montre que la précipitation totale cumulée pour l'année 2003 sur le site de SMOSREX est de moins de 600 mm, clairement inférieure de 50 à 100 mm aux autres années traitées. En fait, 2004 a été aussi caractérisée par une longue période sèche, s'étendant de mai à septembre, mais on voit que la distribution mensuelle des événements de précipitation est très irrégulièrement distribuée ; contrairement à l'été, le printemps a été très humide. En résumé, la période 2001-2004 considérée ici offre une grande variabilité interannuelle : 2001 est une année " normale " proche de la climatologie, 2002 est une année humide, 2003 et 2004 comportent des périodes très sèches. Ceci est un atout important pour tester l'assimilation de données dans des conditions contrastées.

3.2. Évolution temporelle des variables de forçage atmosphérique

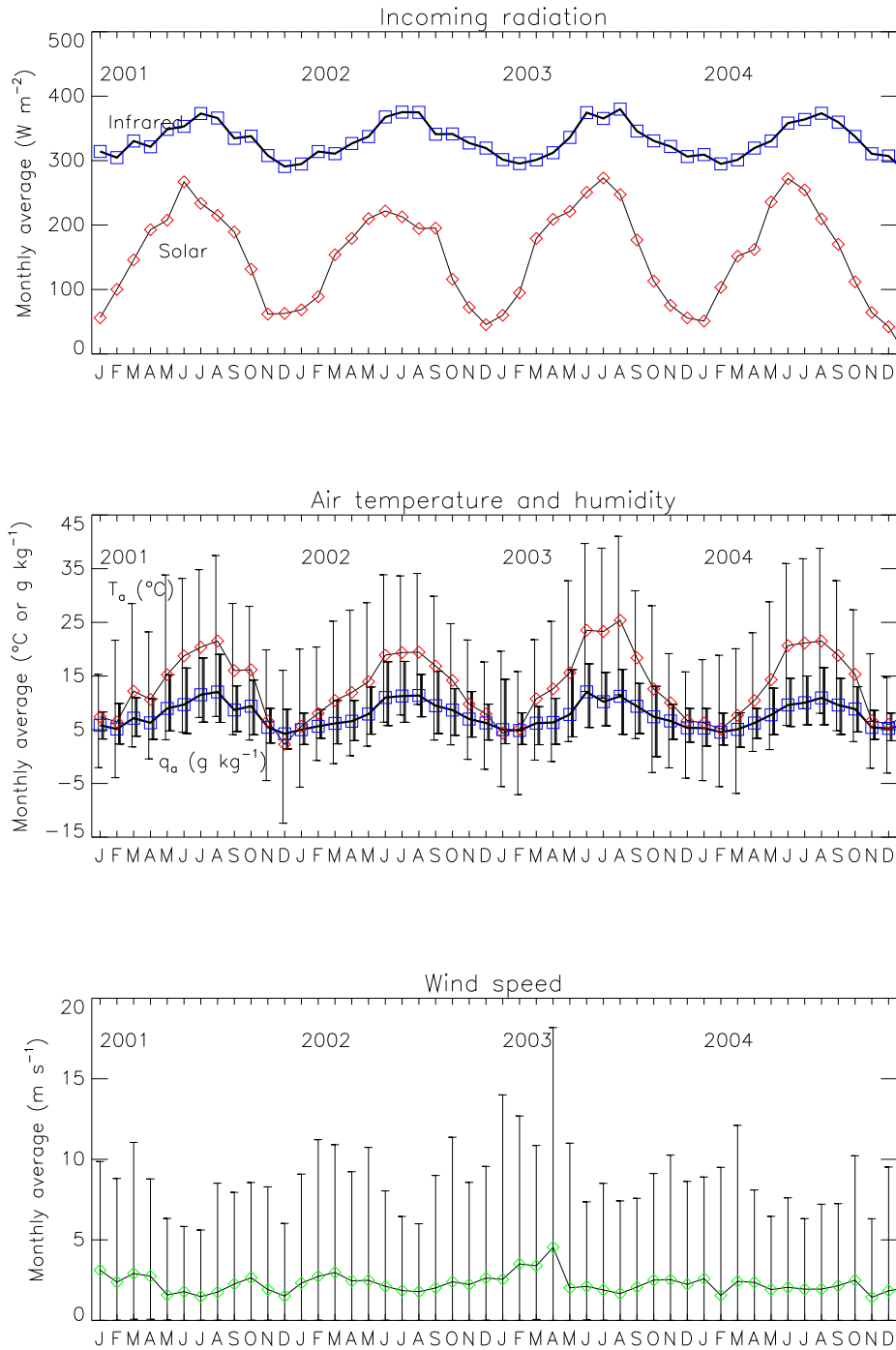


FIG. 3.2: Evolution temporelle des variables du forçage atmosphérique de SMOSREX, pendant la période 2001-2004. Les barres verticales indiquent les valeurs maximum et minimum enregistrées pour chaque mois.

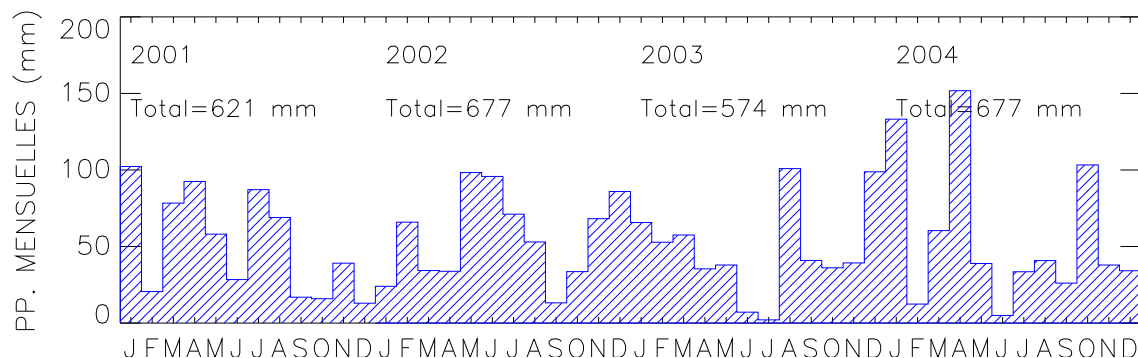


FIG. 3.3: Précipitations cumulées mensuelles et annuelles sur le site de SMOSREX de 2001 à 2004.

3.3 Humidité du sol et biomasse de la végétation

Deux variables observables par télédétection et susceptibles d'être assimilées ont été mesurées : le LAI et l'humidité superficielle du sol (w_g) intégrée sur les 6 premiers cm du sol. Les mesures réalisées sur SMOSREX nous permettent également de valider les variables analysées par les différentes méthodes d'assimilation testées : le contenu en eau intégré du sol (w_2), jusqu'à une profondeur de 95 cm, et la biomasse de la végétation.

La fig. 3.4 montre l'humidité de la couche superficielle du sol (moyenne des mesures effectuées par quatre sondes Theta Probe disposées verticalement en surface), la moyenne de toutes les mesures effectuées à différentes profondeurs (c.à.d., le contenu en eau intégré du sol, de 0 à -95 cm), et le profil de l'humidité volumique du sol. Les sections " Automatic measurements " et " Experimental site and data set " des articles aux pages 62 et 99, décrivent plus en détail comment les mesures de l'humidité du sol ont été effectuées (voir aussi De Rosnay *et al.*, 2006). On remarque l'assèchement presque total du sol durant la période de sécheresse de l'année 2003, et également en 2004, caractérisée par de faibles précipitations pendant la période Mai-Septembre. Ainsi, des humidités volumiques de seulement 20% sont mesurées à une profondeur de 90 cm pendant l'été 2003 et l'automne 2004 et les valeurs minimales de l'humidité sont très inférieures aux minima de 2001 et 2002. Ces résultats sont en accord avec les caractéristiques des variables atmosphériques présentées dans les figs. 3.2 et 3.3.

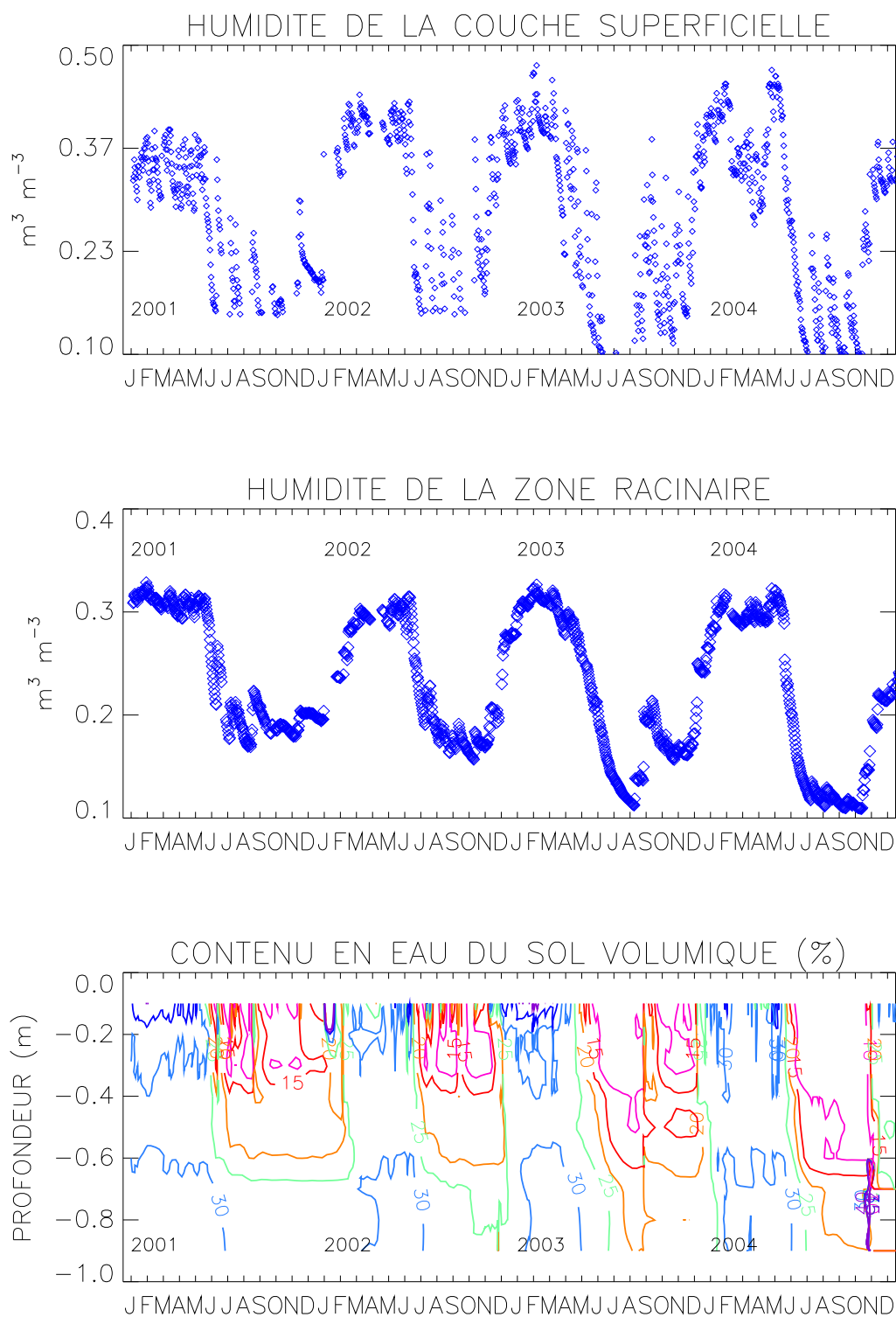


FIG. 3.4: Evolution temporelle de (du haut vers le bas) l'humidité de la couche superficielle du sol (0 à -6 cm), du contenu intégré en eau du sol, du profil d'humidité de la zone racinaire.

La fig. 3.5 représente l'évolution temporelle des observations de deux variables de la jachère de SMOSREX : le LAI et la hauteur du couvert végétal. Le trait continu représente le résultat de l'application d'une fonction d'interpolation aux observations. Pour les deux variables on observe un comportement particulier provoqué par la canicule de 2003 : un double cycle de LAI est observé en 2003.

La fig. 3.6 présente la biomasse aérienne sèche mesurée sur la jachère de SMOSREX. Deux caractéristiques apparaissent : (1) le cycle végétatif évolue peu l'hiver et, en conséquence, peu de mesures ont été réalisées en hiver et, (2) on observe une forte dispersion des mesures, surtout pour 2001 et 2002 (cela est dû à l'hétérogénéité spatiale de la jachère et à une modification du protocole de mesure en 2003). Dans la suite de l'étude, ce sont les valeurs interpolées qui sont utilisées. La Figure 3.6 présente la biomasse aérienne sèche mesurée sur la jachère de SMOSREX.

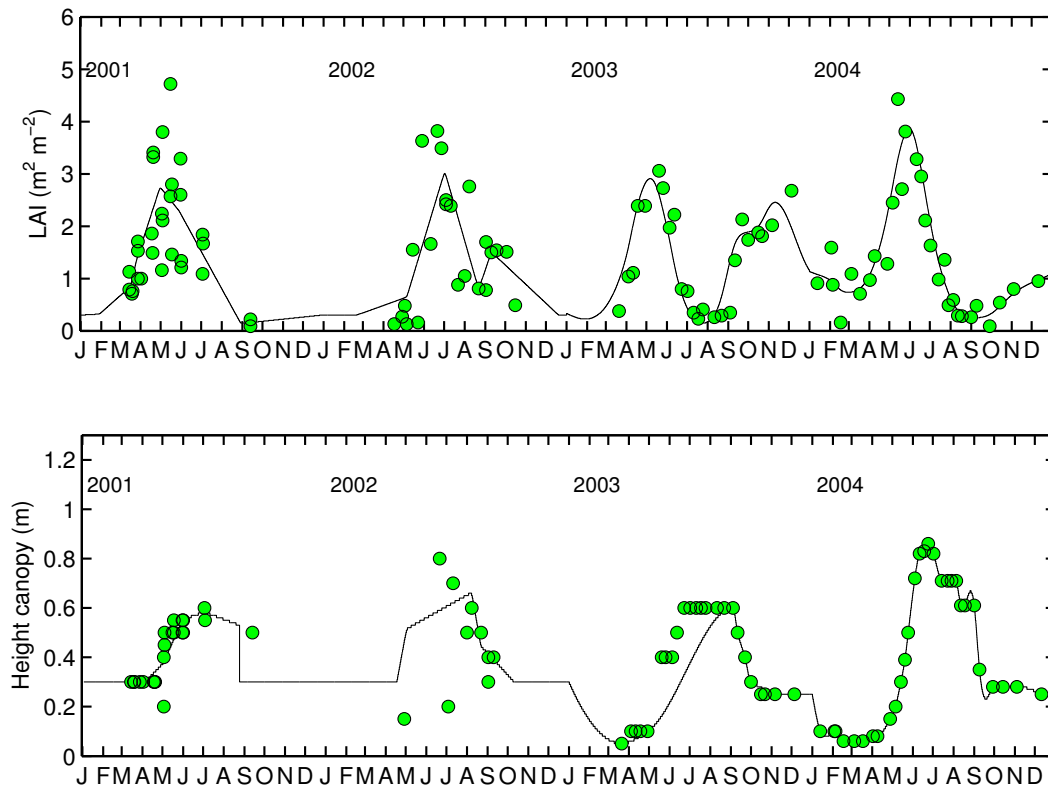


FIG. 3.5: Dynamique de l'évolution du LAI et de la hauteur de la jachère sur le site de SMOSREX, pendant la période 2001-2004.

3.4 Mesures des instruments de télédétection

Un élément essentiel de l'expérience SMOSREX est l'obtention de mesures de télédétection en continu pendant plusieurs années. En effet, l'étude de la stabilité des modèles d'émission

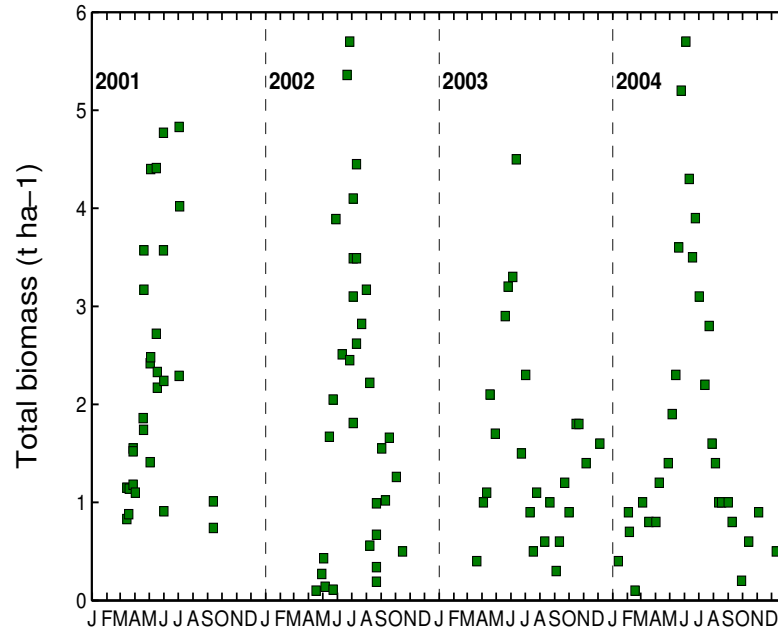


FIG. 3.6: Biomasse aérienne mesurée sur la jachère de SMOSREX.

micro-ondes de la surface et leur réponse aux phénomènes de courte durée (précipitations ou évaporation intenses, rosée, gel du sol, etc.) requiert plusieurs cycles annuels de mesure (De Rosnay *et al.*, 2006). De même, pour étudier la réponse des schémas d'assimilation aux processus de basse fréquences (croissance de la végétation, assèchement du sol pendant l'été, etc.), il est indispensable de disposer de longues séries. Les observations de télédétection de SMOSREX sont obtenues depuis l'année 2003. Ces données sont présentées ci-dessous.

3.4.1 Températures de brillance en bande L

L'instrument le plus innovant de SMOSREX est le "L-band radiometer for Estimating Water In Soils (LEWIS, Lemaitre *et al.*, 2004). LEWIS est un radiomètre en bande L bi-polarisation (polarisations V et H). Sa grande stabilité (0.05 K) et résolution (0.1-0.2 K) font de lui un radiomètre particulièrement précis. De plus, l'usage a montré qu'il est extrêmement fiable. LEWIS a été placé au sommet d'un portique de mesure de 15 m de hauteur (fig. 3.7) et un système automatique développé par le CNRM lui permet de fournir des mesures en continu de l'émission micro-onde en bande L, sur jachère et sur sol nu, sous différents angles d'incidence (entre 20 et 60°). Ce dispositif permet d'obtenir une série de températures de brillance (T_B) sur plusieurs cycles annuels, qui permettra à terme de réaliser l'un des principaux objectifs de SMOSREX, l'assimilation des T_B dans ISBA-A- g_s sur plusieurs cycles annuels contrastés. Cette série de mesures a déjà montré son intérêt pour le développement de la modélisation de l'émission en bande

L des surfaces naturelles, car la jachère présente des caractéristiques particulières, comme la présence d'une litière formée de débris végétaux (Saleh *et al.*, 2006a), que l'on trouve rarement sur les couvert agricoles, les plus souvent étudiés jusqu'alors. La série temporelle d'observations de T_B pour l'année 2003 est présentée au chapitre 5.5.1.

Radiomètre LEWIS:
L-band radiometer for
Estimating Water in Soils

Caractéristiques:

- Émission en bande-L
- 1.4 GHz (21 cm)
- Polarisation: H & V
- Taille: 3 m
- Poids: 200 kg
- Régulation thermique
 47.20 ± 0.02 °C
- Résolution: 0.1-0.2 K
- Stabilité: 0.05 K

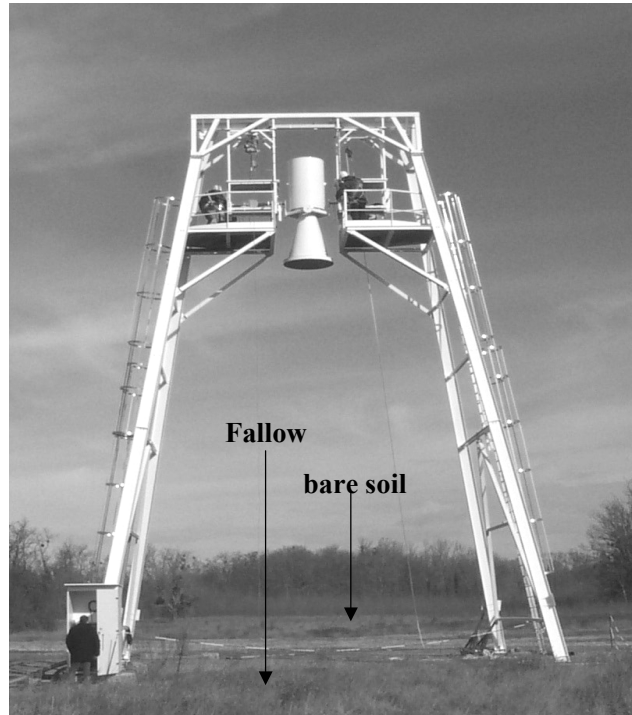


FIG. 3.7: Le radiomètre LEWIS installé sur le site de SMOSREX depuis février 2003 et ses principales caractéristiques.

3.4.2 Réflectances

Deux luminancemètres CIMEL installés sur le portique de mesure de SMOSREX (fig. 3.8) permettent d'estimer la réflectance de la jachère dans cinq bandes spectrales (les fenêtres spectrales sont indiquées entre parenthèses) :

1. blue : 450 nm (430-470) ;
2. vert : 549 nm (506.5-591.5) ;
3. rouge : 648 nm (621.5-674.5) ;
4. proche infrarouge (NIR) : 837.5 nm (792-883) ;
5. moyen infrarouge (MIR) : 1640 nm (1557.7-1722.5).

L'un des deux luminancemètres mesure la luminance L provenant de la jachère, l'autre mesure l'éclairement incident I . Le rapport entre ces deux quantités va nous fournir une estimation de la

réflectance sur jachère. Bien que ces mesures ne soient pas affectées par l'atmosphère à cause de la proximité des instruments avec la surface, elles doivent, cependant, être corrigées de quelques effets (voir Annexe D).

CIMEL Reflectancemètres



Heitronics Pyromètres KT15.D Bandwidth: 9600 – 11500 nm

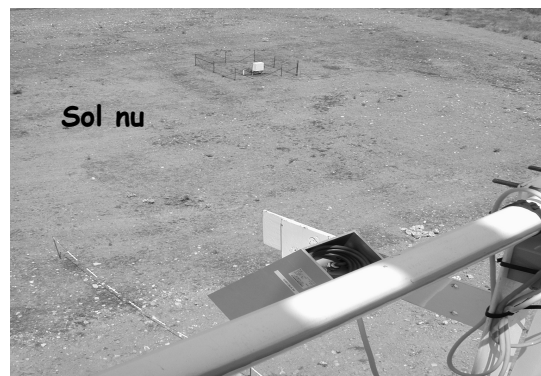


FIG. 3.8: Instrumentation dans les spectres visible, proche-infrarouge, moyen infrarouge et infrarouge thermique sur SMOSREX. A gauche : reflectancemètre CIMEL pour mesurer la réflectance de la jachère. A droite : pyromètres Heitronics pour mesurer la température infrarouge sur jachère et sur sol nu.

3.4.2.1 Réflectances journalières sur SMOSREX

La fig. 3.9 montre les réflectances spectrales journalières depuis juillet 2003 (début de ces mesures) et jusqu'à la fin de 2004. Pour avoir une mesure significative de la réflectance, il faut des conditions de ciel clair, puisque sa valeur peut être fortement réduite si des nuages s'interposent entre la direction du rayonnement solaire direct et les capteurs. Le filtrage a été fait à partir d'une simple analyse visuelle de l'évolution des courbes journalières d'éclairement solaire. Des discontinuités significatives sur cette courbe sont un signe du passage de nuages. La réflectance

montre aussi un cycle diurne très marqué. Les courbes de réflectances présentées à la fig. 3.9 ont été obtenues au midi solaire, car c'est à cette heure que la réflectance atteint sa valeur maximum et que son évolution temporelle est la plus lente.

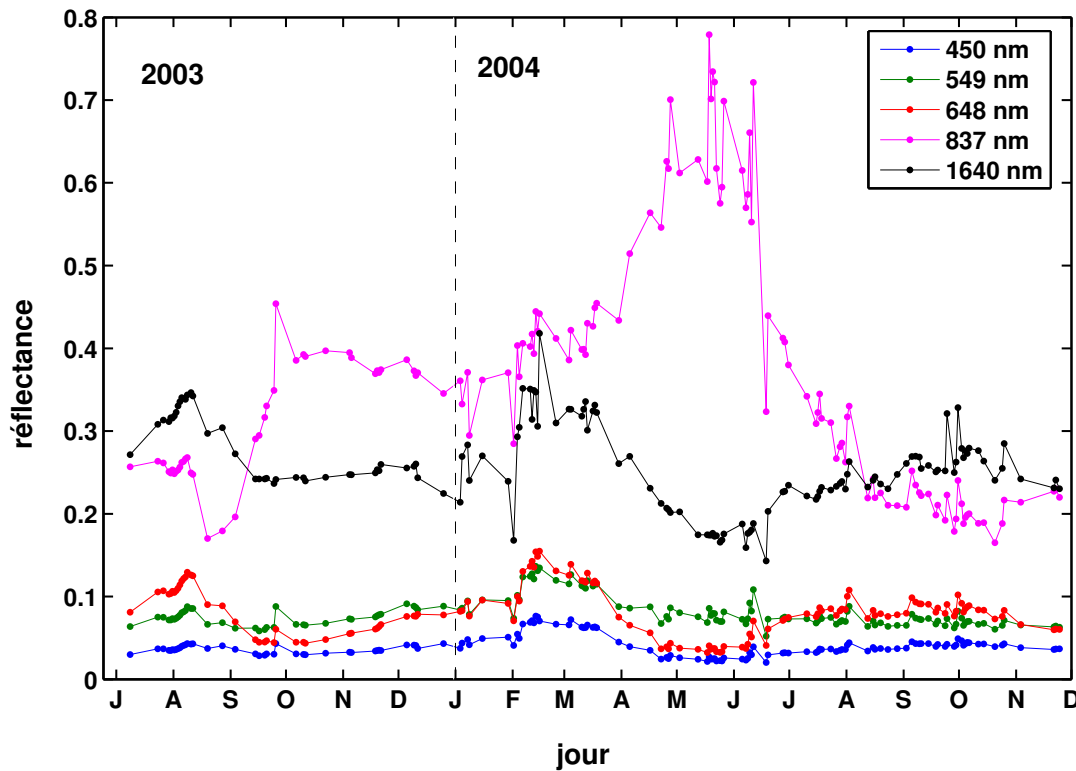


FIG. 3.9: Réflectances spectrales mesurées sur la jachère de SMOSREX en 2003 et 2004, à midi, à un angle d'incidence de 40° , pour les bandes spectrales CIMEL dans les domaines visible (450, 549, 648 nm), proche-infrarouge (837 nm), et moyen-infrarouge (1640 nm).

On observe, pendant les périodes de sécheresse de 2003 et 2004, que la réflectance dans le vert (549 nm) devient inférieure à la réflectance dans le rouge (648 nm). En effet, la bande rouge est sensible au contenu en chlorophylle. L'activité photosynthétique étant réduite pendant les périodes de sécheresse, l'absorption du rouge diminue, ce qui se traduit par une plus grande réflectance. De même, la réflectance est plus grande dans le moyen infrarouge (1640 nm) pendant les périodes sèches que dans le proche infrarouge (837 nm), car elle est plus sensible au contenu en eau des feuilles.

3.4.2.2 Indices de végétation

Les indices de végétation, calculés comme une combinaison des réflectances à différentes longueurs d'onde, permettent de suivre la dynamique de la végétation et sa productivité. Les cinq bandes spectrales du réflectancemètre CIMEL permettent de construire des indices sensibles

à l'état de la végétation. Deux de ces indices ont été très utilisés durant les dernières décennies : le " Simple Ratio " (SR, Jordan 1969) et le " normalized difference vegetation index " (NDVI, Deering 1978). Ils sont définis par les combinaisons suivantes des réflectances spectrales :

$$SR = \frac{\rho_{NIR}}{\rho_{RED}} \quad (3.1)$$

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (3.2)$$

Le principe physique donnant leur pertinence à ces indices est l'absorption de la lumière, dans les longueurs d'onde correspondant au rouge, par les pigments de chlorophylle des feuilles. Les réflectances dans le rouge sont inversement proportionnelles à la quantité de chlorophylle présente dans la végétation. D'autre part, les réflectances NIR et MIR sont influencées par la structure interne de la feuille, et plusieurs couches de feuilles peuvent même influencer la réflectance (Nemani et Running, 1989). La fig. 3.10 montre l'évolution temporelle de ces deux indices de végétation sur la jachère de SMOSREX depuis que les mesures de réflectance sont disponibles, obtenues à partir des réflectances de la fig. 3.9. Bien que l'évolution temporelle du SR présente une dynamique plus marquée que celle du NDVI, les deux indices montrent le même comportement. Quand la végétation s'assèche, l'absorption par la chlorophylle est plus faible dans le rouge et la réflectance dans le rouge augmente, ce qui fait diminuer la valeur du SR et du NDVI. Inversement, ces deux indices augmentent parallèlement à la croissance de la végétation. Ces courbes sont assez bien corrélées avec les courbes de LAI (fig. 3.5) et de la biomasse de la végétation (fig. 3.6), ce qui est consistant avec leur utilisation comme indicateurs de l'état de la végétation. Une multitude d'indices peuvent être construits en combinant les réflectances à plusieurs longueurs d'onde, ce qui permet de suivre différents aspects de la végétation. Par exemple, un indice construit à partir des bandes du NIR et du MIR, peut fournir une indication de la quantité d'eau présente dans la végétation.

3.4.3 Température infrarouge de la surface

Deux pyromètres installés sur le portique de SMOSREX permettent de mesurer la température infrarouge de la surface à un angle d'incidence de 40 degrés sur les deux parcelles de sol nu et de jachère. Ces mesures peuvent être utilisées pour estimer la température de la surface (T_S) et pour valider les simulations de T_S fournies par un LSM. On peut également les assimiler dans un LSM ou construire des indicateurs pour évaluer l'état hydrique du sol. La fig. 3.11 montre la corrélation existant entre la température infrarouge et la température de surface simulée par ISBA-A-gs sur la jachère entre juillet 2003 et fin décembre de l'année 2004 (les mesures ne sont

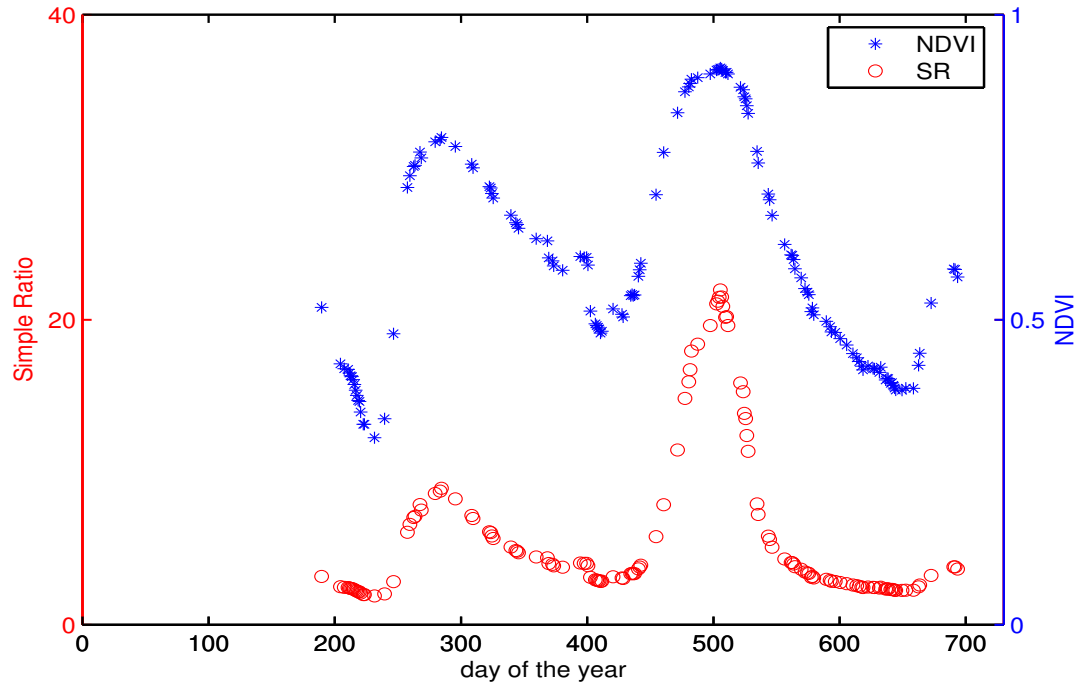


FIG. 3.10: Comparaison entre le NDVI et le SR pour des journées de ciel clair, pour la période de 2003-2004.

disponibles qu'à partir de juillet 2003). La fig. 3.11, à gauche, représente l'évolution temporelle de T_S sur une période d'une dizaine de jours uniquement, pour plus de lisibilité. La figure de droite représente la corrélation entre les deux températures sur toute la période. On peut constater une très bonne corrélation entre les deux températures, avec un coefficient de corrélation quadratique (r^2) de 0.95, un RMSE de 2.65 K et un biais moyen de -0.30 K. La période d'étude s'achevant à la fin de 2004, on dispose d'un an et demi d'observations de la température infrarouge avec le pyromètre.

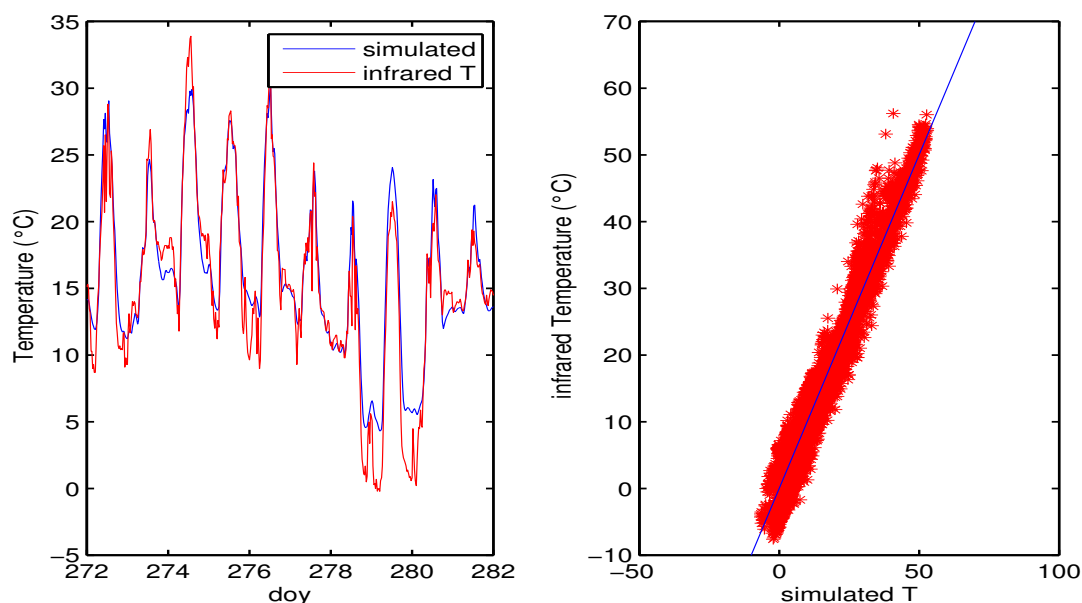


FIG. 3.11: Comparaison entre la température infrarouge mesurée sur la jachère de SMOSREX, et simulée par ISBA-A-gs. A gauche : évolution sur une dizaine de jours en 2003. A droite : corrélation entre les deux températures entre Juillet 2003 et Décembre de l'année 2004.

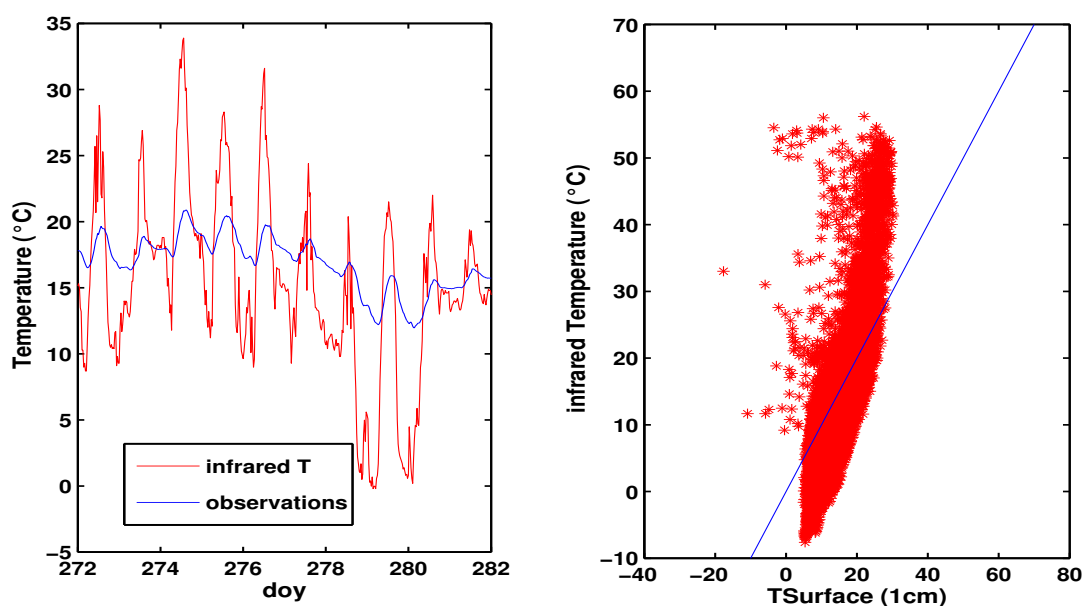


FIG. 3.12: Comparaison entre la température infrarouge mesurée sur la jachère de SMOSREX et la température mesurée à 1 cm de profondeur. A gauche, évolution sur une dizaine de jours en 2003. A droite, corrélation entre les deux températures entre Juillet 2003 et décembre de l'année 2004.

Deux possibilités existent pour compléter la série, de janvier 2001 à juillet 2003 : utiliser la température mesurée à la surface du sol (-1 cm) ou les mesures du bilan radiatif de la jachère.

La fig. 3.12 montre que la corrélation entre la température infrarouge et la température mesurée à 1 cm de profondeur n'est pas très bonne, avec r^2 de 0.57, RMSE de 8.13 K et un biais moyen de 0.46 K. En effet, les cycles diurnes de la température à 1 cm sont beaucoup plus faibles car ils sont fortement atténués par la couche de végétation de la jachère.

Cependant, si on utilise les mesures des composantes du bilan de rayonnement thermique à la surface (eq. (3.3)), on peut étendre la série à partir de 2001 :

$$R_A^\uparrow = \epsilon_s \sigma T_s^4 + (1 - \epsilon_s) R_A^\downarrow \quad (3.3)$$

où ϵ_s est l'émissivité de la surface supposée égale à 0.97 et σ la constante de Steffan-Boltzmann.

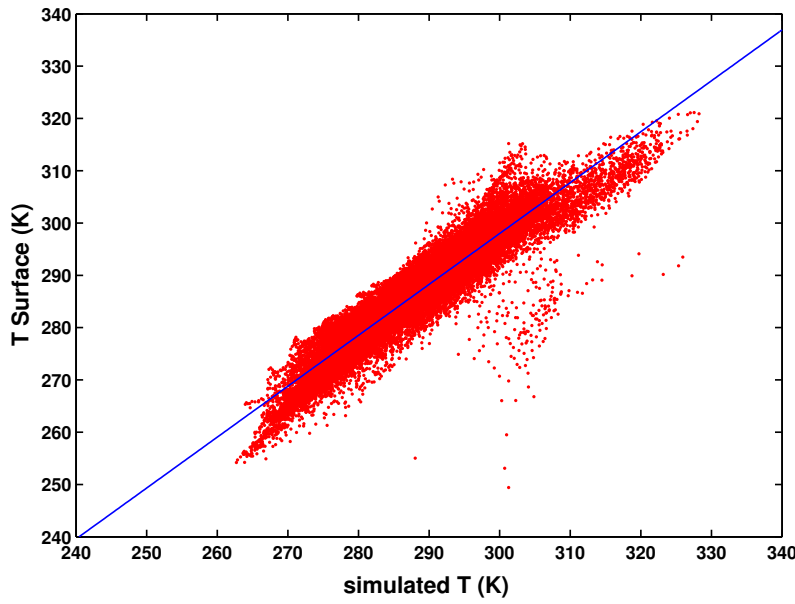


FIG. 3.13: Comparaison entre la température de la surface simulée par ISBA-A- g_s et celle issue de bilan radiatif thermique à la surface, pour la période de l'année 2001 jusqu'à Juillet de l'année 2003.

Etant donné que la température de la surface issue de l'équation (3.3) est une mesure indirecte, on peut être amené à supprimer un biais. En regard de la bonne corrélation existant entre T_S et T_{ISBA} , on peut chercher une relation linéaire entre T_S de l'équation (3.3) et T_{ISBA} pour supprimer le biais :

$$T_S = A \cdot T_{ISBA} + B \quad (3.4)$$

où A et B sont deux constantes qui sont estimées par régression linéaire. Les résultats de la régression attribuent aux constantes A et B les valeurs de 0.97 et 4.9, respectivement. Le r^2 de la régression est de 0.92 et la RMSE = 3.46 K. La fig. 3.13 montre la corrélation existant entre la température de la surface issue de l'équation 3.3, corrigée du biais, et la température simulée par ISBA-A- g_s , pour la période allant de janvier 2001 jusqu'à juillet 2003. Cette série temporelle des

estimations de T_S a été ajoutée aux mesures dans l'infrarouge, pour former une série 2001-2004 permettant de dériver des taux de réchauffement diurne ou " heating rates " (section 3.5), qui sont des indicateurs possibles de l'état hydrique du sol, susceptibles d'être assimilés dans un LSM.

3.5 Heating Rates

A la section 3.4.3 on a présenté une série temporelle de températures de la surface (T_S) dans le domaine infrarouge, sur jachère, obtenue grâce aux mesures du pyromètre sur la jachère de SMOSREX. La bonne corrélation qu'elle montre avec la température de surface simulée par ISBA-A-g_s permet de penser que les variations temporelles de cette température sont susceptibles d'être utilisées pour analyser w_2 grâce à une méthode d'assimilation. C'est ce que l'on appelle la méthode des " heating rates " (Wetzel *et al.*, 1984, McNider *et al.*, 1994). Cette méthode est basée sur le fait que la variation de T_S , surtout le matin entre 0800 et 1000 (heure solaire locale ou HSL), est sensible à l'état hydrique du sol. Après le levé du soleil, une partie du rayonnement solaire incident est absorbé par le sol et la végétation, ce qui va provoquer une augmentation de T_S . Si le sol est sec, une plus grande quantité d'énergie est libérée sous forme de chaleur sensible, contrairement au cas d'un sol humide où le flux de chaleur latente est plus important et, ainsi, la T_S variera plus ou moins rapidement suivant le contenu en eau du réservoir du sol.

Les heating rates (Υ) sont définis comme :

$$\Upsilon = \frac{\partial T_S}{\partial \cos \xi} \quad (3.5)$$

où T_S est la température de la surface et ξ l'angle zénithal solaire. L'avantage d'utiliser Υ au lieu de $\partial T_S / \partial t$ est, en première approximation, sa linéarité (Van Den Hurk et The, 2002) et, en conséquence, la période de temps disponible pour calculer Υ est moins critique que pour $\partial T_S / \partial t$. En effet, $\partial T_S / \partial t$ tend à diminuer pendant la matinée. Suivant l'algorithme de Van Den Hurk et The (2002) on a considéré les observations de T_S entre deux horaires : après la levée du soleil et 1100 HSL. Ensuite, Υ est calculé par régression linéaire entre T_S et $\cos \xi$. L'effet des nuages sur les variations de T_S est très important. L'évaluation des Υ doit donc être réalisée en conditions de ciel clair et cette période de ciel clair doit être suffisamment longue. Deux méthodes pour le filtrage de mesures sans nuages ont été utilisées :

1. Sélection de journées strictement sans nuages, qui consiste à éliminer les journées dont l'évolution temporelle de l'éclairement solaire est marquée par des discontinuités (qui indiquent le passage de nuages). Le problème avec cette méthode est que l'on perd beaucoup

de données susceptibles de fournir des estimations de Υ , surtout pour les journées où il y a alternance de passages nuageux et de ciel clair. Par exemple, sur les 163 jours traités de juillet à décembre 2003, même si tous les Υ calculés par régression linéaire sont estimés avec un r^2 supérieur à 0.96, seuls 28 valeurs ont été obtenues (fig. 3.14).

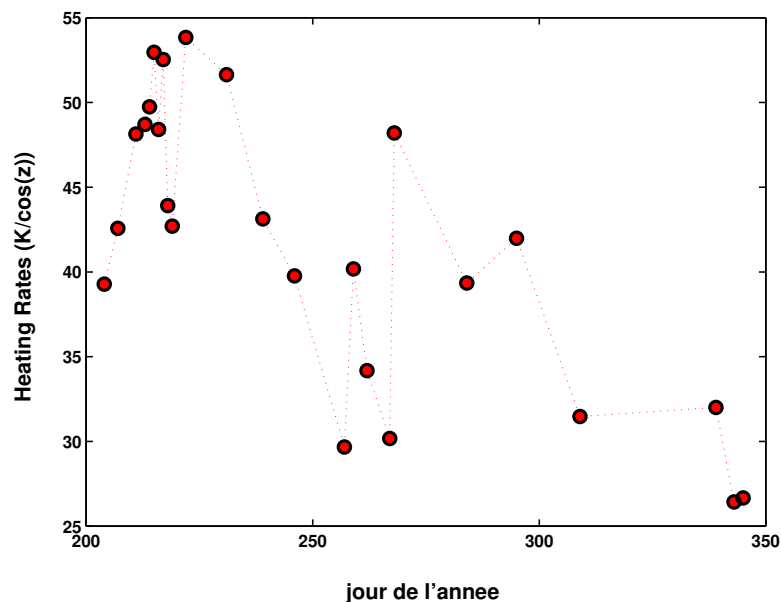


FIG. 3.14: Heating rates pour des journées sans nuages entre juillet et fin décembre 2003.

2. Utilisation de cycles diurnes de maximum d'éclairement mensuel. On considère qu'une mesure de T_S n'est pas affectée par les nuages lorsque l'éclairement est supérieur à un certain pourcentage (à déterminer) d'éclairement maximum mensuel à l'heure considérée (fig. 3.15). Seules les observations entre 0800 et 1100 HSL ont été utilisées. On a fixé deux seuils : observations de T_S dont l'éclairement correspondant est plus grand que 70% du maximum mensuel et un autre à 80%. Pour estimer Υ par régression linéaire, on a besoin, au moins, de deux points consécutifs qui satisfont au critère précédent. On a considéré entre 2 et 5 points. Les résultats sur le nombre de Υ obtenus sont présentés au tableau 3.1.

Les contraintes pour la recherche des Υ sont les suivantes :

- Si on utilise comme critère pour le filtrage des nuages la courbe d'évolution de l'éclairement journalier, un nombre insuffisant de Υ est obtenu ;
- Si on utilise comme critère pour le filtrage des nuages les courbes d'éclairement maximum mensuel, le nombre de Υ obtenu dépend du seuil imposé pour la détection de nuages. Dans ce cas, si on utilise comme critère 70% d'éclairement maximum, logiquement davantage d'estimations sont obtenues qu'avec le critère de 80%, mais l'incertitude des estimations

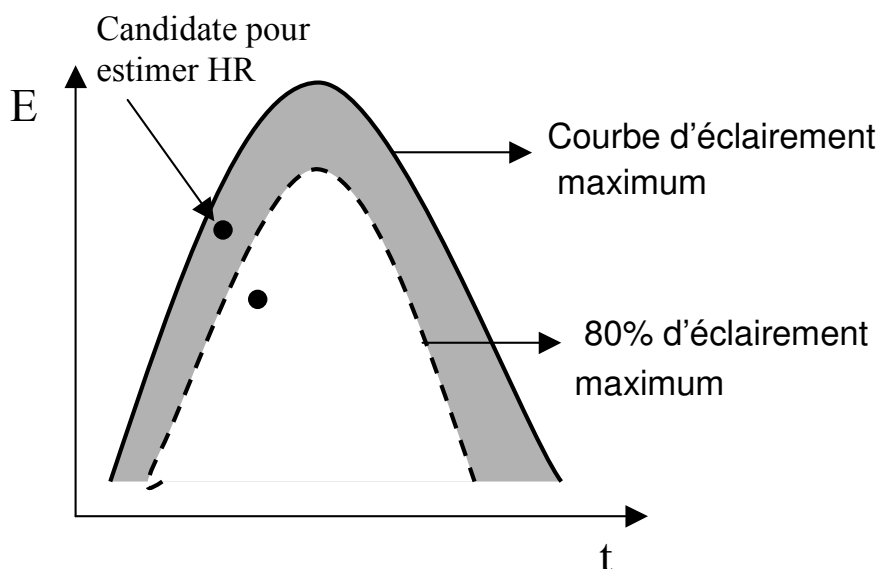


FIG. 3.15: Filtrage des mesures influencées par les nuages avec la méthode d'éclairement maximum mensuel.

TAB. 3.1: Nombre total de " Heating rates " (Υ) obtenus de 2001 à 2004 par régression linéaire avec 2, 3, 4 ou 5 observations consécutives (obs) de la température infrarouge de la surface. Deux seuils sont imposés pour filtrer observations sans l'effet des nuages : éclairement supérieur à 70% ou 80% de l'éclairement maximum mensuel. Dans chaque cas, le nombre total d'observations disponibles pour calculer les " Heating rates " est aussi indiqué (" candidates ").

70%			80%		
obs	candidates	Υ	obs	candidates	Υ
2	4510	800	2	3140	635
3	4009	699	3	2905	503
4	3904	543	4	2737	388
5	3510	462	5	2602	289

est plus forte ;

- Le nombre de Υ obtenu est plus grand lorsque le nombre de mesures consécutives utilisées pour son estimation est plus faible. Cependant, puisque les Υ sont dérivés à partir d'une régression linéaire, la qualité de la régression diminue lorsque peu de mesures sont utilisées.

La fig. 3.16 montre la relation entre les Υ et w_2 entre 2001 et 2004. Les données de T_S entre 2001 et juillet 2003 sont obtenues à partir de l'équation du bilan de rayonnement thermique à la surface (eq. (3.3)). Comme critère pour le filtrage des nuages on a utilisé les observations de T_S avec un éclairement supérieur à 80% de l'éclairement maximum et, pour l'évaluation de Υ , quatre points consécutifs ont été utilisés. De cette façon, on a un nombre suffisant de Υ , en

moyenne 1 tous les 3,7 jours pendant toute la période d'étude, et le bruit est réduit. Compte tenu des résultats montrés à la fig. 3.16, la relation entre l'humidité du sol et les Υ n'est pas très concluante. Bien que l'on observe une tendance de Υ à augmenter au cours du temps lors des périodes sèches (trait rouge), on observe aussi, au voisinage de la capacité au champ des valeurs de Υ aussi grandes que lors des périodes sèches. De plus, la distribution des Υ autour de la ligne de tendance est très dispersée. Les journées claires se produisent le plus fréquemment pendant la période de printemps-été, ce qui se traduit par une plus forte concentration des valeurs de Υ . Au contraire, pendant l'automne et surtout, pendant l'hiver, moins de points sont disponibles pour les comparaisons avec w_2 .

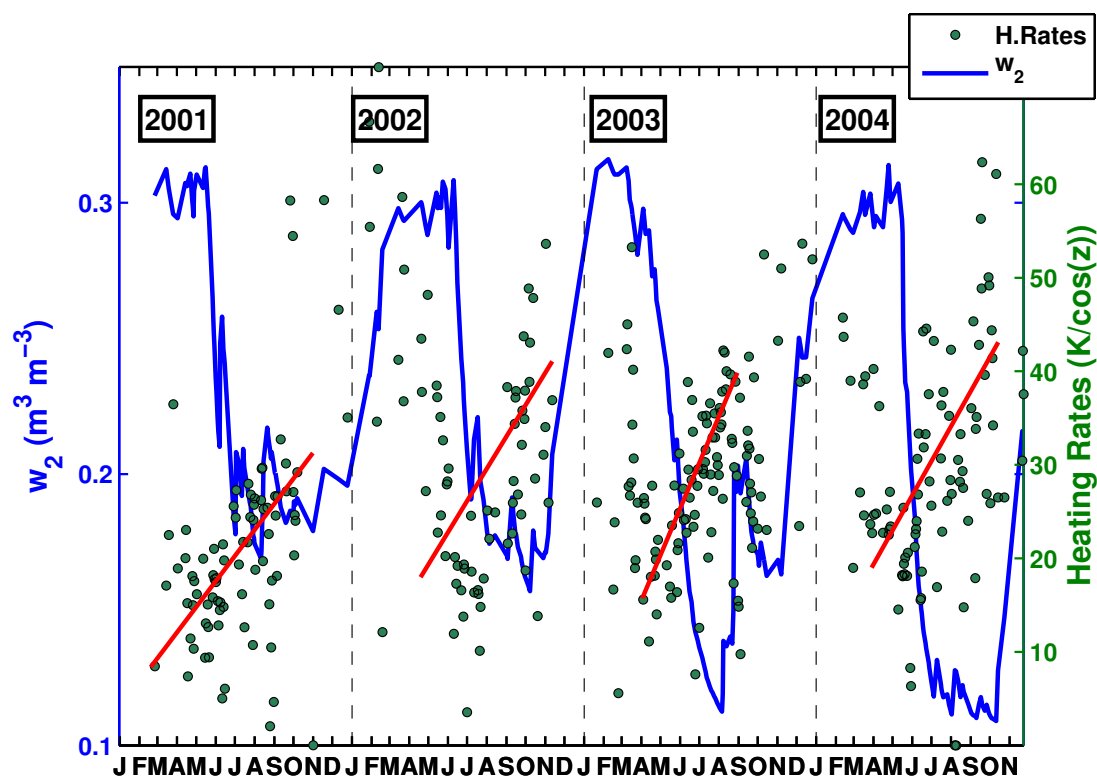


FIG. 3.16: corrélation entre les heating rates et l'humidité de la zone racinaire.

On a vu que l'estimation des Υ est assez indirecte et sa qualité dépend du critère de filtrage des nuages employé ainsi que de la quantité de mesures de T_S utilisées chaque jour. De plus, dans cette étude, les mesures de T_S employées entre 2001 et juillet 2003 sont obtenues de façon indirecte. Tous ces facteurs font que les estimations des Υ sont associées à de fortes incertitudes. Physiquement, l'estimation de Υ ne dépend pas seulement que de w_2 , d'autres facteurs l'influencent également. L'inertie thermique de la végétation peut produire un certain

retard dans la croissance de T_S par rapport au rayonnement solaire incident. De plus, le couplage aérodynamique entre le sol et l'atmosphère peut aussi influencer fortement ces estimations. Ainsi, en conditions de vent faible, les échanges turbulents entre l'atmosphère et le sol sont réduits et l'augmentation de T_S sera plus grande qu'en conditions de vent fort. Si l'effet du vent est important, cela pourrait expliquer une partie de la grande dispersion des estimations de Υ . Dans ce cas, la seule information contenue dans les T_S ne serait pas suffisante pour caractériser l'état hydrique du sol. Pour essayer de visualiser cet effet, la relation entre l'humidité du sol et les Υ est montrée à la fig. 3.17 pour différentes classes de vitesse moyenne du vent pendant la période d'estimation :

- classe 1** : Υ pour lesquels la vitesse du vent est $< 1 m \cdot s^{-1}$;
- classe 2** : Υ pour lesquels la vitesse du vent est entre $1 m \cdot s^{-1}$ et $2 m \cdot s^{-1}$;
- classe 3** : Υ pour lesquels la vitesse du vent est entre $2 m \cdot s^{-1}$ et $3 m \cdot s^{-1}$;
- classe 4** : Υ pour lesquels la vitesse du vent est entre $3 m \cdot s^{-1}$ et $6 m \cdot s^{-1}$;
- classe 5** : Υ pour lesquels la vitesse du vent est $> 6 m \cdot s^{-1}$.

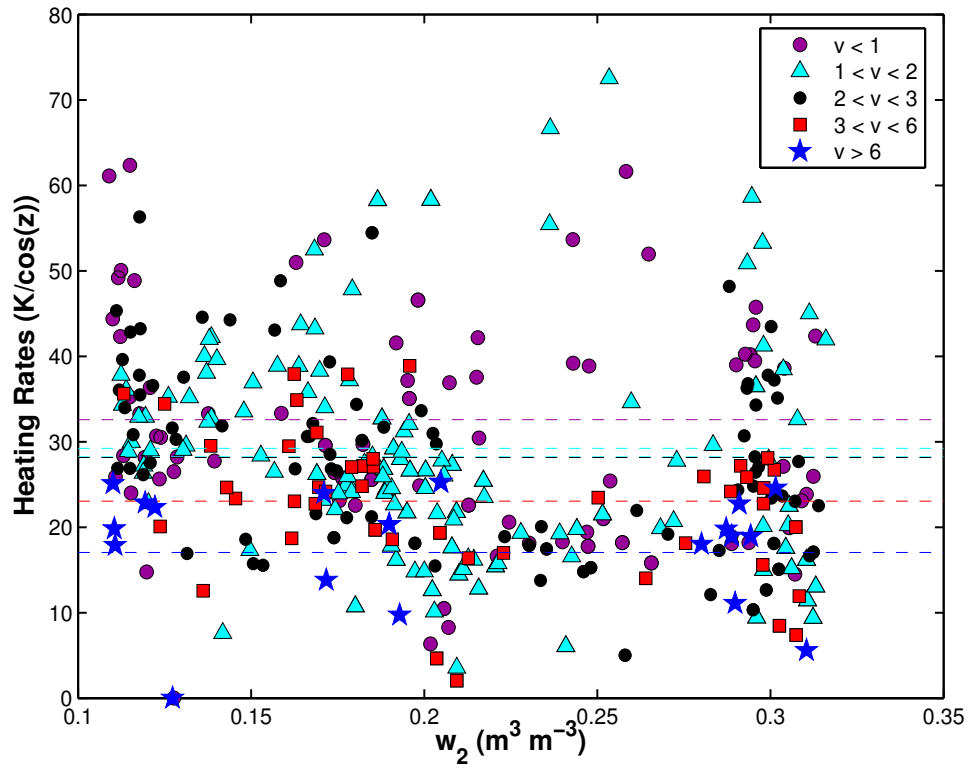


FIG. 3.17: Relation entre Υ et w_2 pour différents vitesses du vent v (en $m \cdot s^{-1}$). Les lignes discontinues horizontales représentent la moyenne des Υ sur toute la période de 2001 à 2004, pour chaque intervalle de vitesse du vent.

On constate que, bien que la dispersion des estimations des Υ associés à des vitesses du vent situées dans le même intervalle est grande, en moyenne sur la période de quatre ans, il existe une certaine corrélation avec la vitesse du vent. Les Υ associés aux fortes vitesses du vent ($> 6 m \cdot s^{-1}$) sont logiquement moins nombreux, mais ils montrent les valeurs les plus basses, tant en conditions de sol humide que de sol sec, avec une moyenne inférieure à $20 K/\cos(\xi)$. Au contraire, les journées où les échanges turbulents entre l'atmosphère et la surface du sol sont plus faibles, à cause de faibles vitesses du vent ou de conditions de forte stabilité atmosphérique, l'incrément de T_S et la valeur de Υ sont plus forts. Ainsi, pour des vitesses du vent inférieur à $1 m \cdot s^{-1}$, la moyenne des Υ (supérieure à $30 K/\cos(\xi)$) est la plus grande.

En conclusion, on observe qu'il existe une forte dépendance des Υ à la vitesse du vent, qui peut être même plus forte que la dépendance à l'humidité du sol. Ceci, ajouté au fait que les estimations de Υ sont assez indirectes et affectées par une grande incertitude, limite fortement l'intérêt de cette technique pour caractériser l'état hydrique du sol. Dans les chapitres suivants, on s'intéressera plutôt à l'utilisation d'observations directes de l'humidité superficielle du sol.

3.6 Le modèle de surface

L'objectif de cette section est de fournir une description des caractéristiques du schéma de surface ISBA-A- g_s qui sont significatives pour cet étude. Une description plus détaillée de la paramétrisation d'ISBA-A- g_s peut être consultée dans les références citées dans ce paragraphe et à la section " methodology " dans les articles aux pages 64 et 100.

3.6.1 ISBA-A- g_s

Le modèle d'Interaction entre la Surface, Biosphere et Atmosphère, ISBA, décrit les échanges d'eau et de chaleur à l'interface sol-végétation-atmosphère. Il a pour origine les travaux de Noilhan et Planton (1989), et est basé sur cinq équations pronostiques pour déterminer l'évolution de cinq variables de surface : T_S , la température du sol (T_2) à proximité de la surface, w_g , w_2 et le réservoir d'interception d'eau par les plantes (W_r).

Dans sa version de base, ISBA utilise une paramétrisation simple de la conductance stomatique g_s pour l'évaluation de la transpiration des plantes. Cette paramétrisation est basée sur l'approche de Jarvis (1976). Pour passer à l'échelle de la canopée, g_s est multiplié par une estimation du LAI imposée par l'utilisateur du modèle. Cependant, en pratique, ces estimations de

LAI sont tirées de climatologies et ne prennent pas en compte de possibles changements rapides de la végétation, par exemple, lors de sécheresses exceptionnelles.

Une bonne estimation de la biomasse de la végétation sur le site de SMOSREX est importante, car elle a une forte influence sur le contenu en eau du sol (par extraction d'eau par les racines) et, en même temps, conditionne fortement le flux d'évapotranspiration. ISBA-A- g_s (Calvet *et al.*, 1998b, Gibelin *et al.*, 2006) est une nouvelle version du modèle ISBA, où la photosynthèse de la plante est paramétrisée par le modèle de Jacobs *et al.* (1996). En complément du modèle de Jacobs, plusieurs représentations de la réponse au stress hydrique ont été introduites par : Calvet *et al.* (1998b), Calvet (2000a), et Calvet *et al.* (2004). Le modèle de Jacobs calcule le taux d'assimilation nette de CO_2 en fonction de plusieurs facteurs environnementaux, comme la température de la feuille, l'humidité de l'air et le rayonnement global incident, mais aussi du type de végétation. Dans ISBA-A- g_s , l'assimilation nette peut être utilisée pour simuler le LAI et la biomasse de la végétation. Ainsi, l'évolution de la biomasse verte (B) résulte d'un bilan fondé sur deux termes : Ainsi, l'évolution de la biomasse verte (B) résulte d'un bilan fondé sur

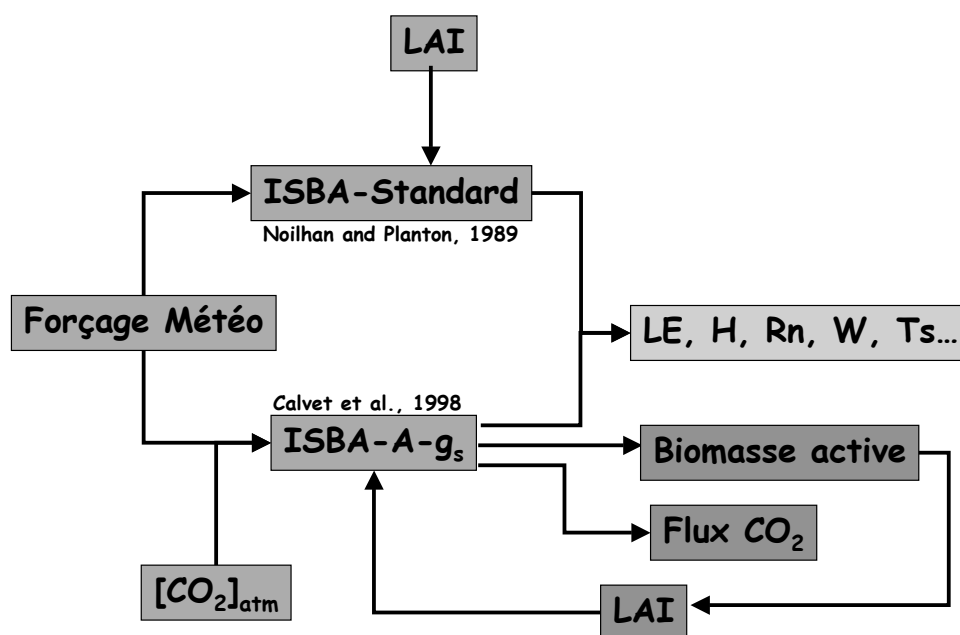


FIG. 3.18: Différences entre la première version d'ISBA et ISBA-A- g_s , par rapport aux variables nécessaires d'entrée et les variables modélisées.

deux termes :

- un terme de croissance (ΔB^+), qui est un terme dépendant du flux d'assimilation nette calculé par le module de photosynthèse ;
- un terme de mortalité (ΔB^-), modulé par une espérance de vie des feuilles qui est diminuée par les effets limitant la photosynthèse (stress hydrique, température trop haute ou trop

basse, rayonnement trop faible, etc.) et qui ne peut pas dépasser une valeur maximale (τ_M).

La biomasse B au temps $t + \Delta t$ est en conséquence calculée comme : $B(t + \Delta t) = B(t) + \Delta B^+ - \Delta B^-$. La biomasse ne peut pas être inférieure à un seuil, relié à une valeur minimale du LAI (LAI_{min}), ce qui permet à la plante de recommencer à assimiler du carbone lorsque les conditions atmosphériques deviennent plus favorables à la photosynthèse.

Le site de SMOSREX n'est pas irrigué, et la jachère du site est sensible aux périodes de déficit hydrique. Dans le cas où la végétation est en conditions de stress hydrique, deux stratégies de réponse de la plante sont possibles (Calvet, 2000a) :

1. Stratégie offensive : elle correspond à une stratégie de tolérance du stress, c'est à dire, pas ou peu d'augmentation de l'efficacité d'utilisation de l'eau par la plante en réponse au stress ;
2. Stratégie défensive : elle correspond à une stratégie d'évitement. Dans ce cas, la régulation stomatique est renforcée par le stress et l'efficacité d'utilisation de l'eau augmente.

En conclusion, ISBA-A- g_s est une amélioration de la version initiale d'ISBA. Cette nouvelle version permet de calculer les flux de CO_2 ainsi que le LAI et la biomasse de la végétation. Comme la version initiale d'ISBA, ISBA-A- g_s est un modèle relativement simple, utilisant un faible nombre de paramètres et peu coûteux en temps de calcul, ce qui facilite l'assimilation de données pour l'analyse de l'humidité du sol et de la biomasse de la végétation.

3.6.2 Simulations d'ISBA-A- g_s

Dans cette section, les simulations initiales de ISBA-A- g_s de w_2 et de LAI pour la jachère de SMOSREX sont présentées (fig. 3.19). Pour estimer les paramètres du modèle (sol et végétation) on a utilisé les mesures in situ (texture du sol par exemple) ou bien repris les paramètres de la végétation déjà utilisés dans les simulations du modèle ISBA-A- g_s sur MUREX (Calvet *et al.*, 1999), tableau 1 . Les deux stratégies de réponse au stress hydrique sont utilisées. Pour comparaison, les données de validation in situ sont superposées aux simulations. On observe que, en général, ISBA-A- g_s simule assez bien l'humidité de la zone racinaire. Les principaux problèmes sont observés pendant les périodes de sécheresse de 2003 et de 2004, lorsque la zone racinaire est plus sèche que le point de flétrissement utilisé par le modèle. Le point de flétrissement est le contenu en eau minimal dans le sol pour que la plante puisse réaliser de la photosynthèse. Dans ce cas il se produit de forts désaccords entre la simulation et les observations. En tout cas, compte tenu des résultats de la fig. 3.19, l'humidité de SMOSREX semble être mieux modélisée si la jachère suit une stratégie de tolérance au stress hydrique (" stratégie offensive "). Ceci est

consistant avec les résultats de Calvet et Noilhan (2000b) portant sur la jachère de MUREX. Les différences entre les deux stratégies apparaissent lors du printemps de l'année 2002 et pendant la deuxième moitié de l'année 2003, et la simulation de w_2 se compare mieux aux observations avec la stratégie offensive.

Dans la suite de l'étude, c'est la stratégie offensive qui est utilisée.

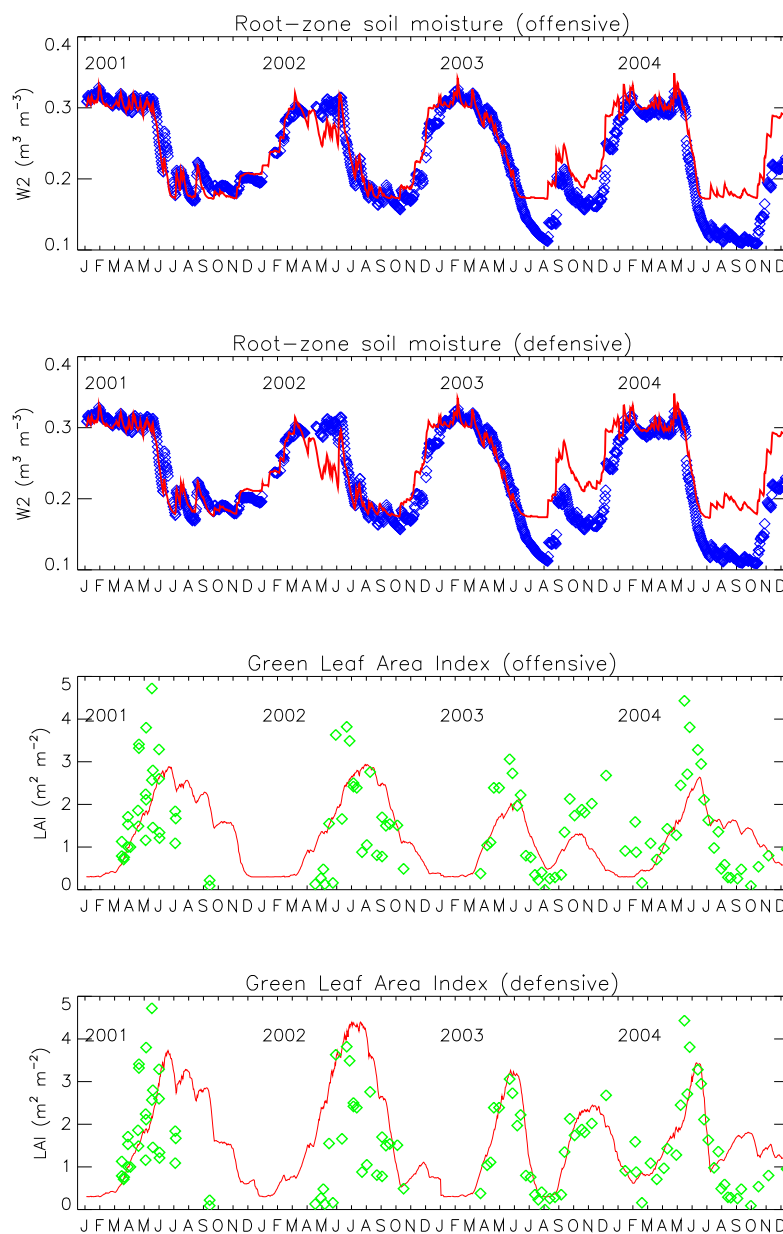


FIG. 3.19: Simulation du contenu intégré en eau du sol et du LAI avec ISBA-A- g_s , en utilisant l'une ou l'autre des stratégies de réponse à la sécheresse (tolérance -stratégie offensive- ou évitement -stratégie défensive-), et comparaison avec les observations.

Chapitre 4

Méthodes d'assimilation des observations de w_g dans ISBA-A-g_s

Sommaire

4.1	Introduction	59
4.2	"From near-surface to root-zone soil moisture using different assimilation techniques."	61
4.2.1	Résumé	61
4.2.2	Article	62
4.3	Une méthode dérivée de l'EnKF	85

4.1 Introduction

Les chapitres précédents ont été consacrés à la description des éléments nécessaires pour tester l'assimilation de données de télédétection sur le site de SMOSREX : le site, le jeu de données expérimentales et le schéma de surface. L'application à l'assimilation de données a été réalisée en deux étapes décrites dans les chapitres 4 et 5. Dans ce chapitre 4, il s'agit d'identifier une méthode d'assimilation de données de télédétection performante pour une implémentation pré-opérationnelle dans le schéma de surface ISBA-A-g_s de Météo-France. Dans ce but, plusieurs méthodes ont été comparées sur le critère de l'analyse de w_2 par assimilation des observations de w_g obtenues sur le site de SMOSREX. Parmi un grand nombre de méthodes existant dans la littérature, quatre méthodes ont été considérées qui satisfont plusieurs critères :

- Relative simplicité pour une implémentation opérationnelle ;
- Temps de calcul limité ;

- Maturité (présence significative dans la littérature).

Les quatre méthodes retenues dans l'étude d'intercomparaison sont :

- Filtre de Kalman étendu (EKF) ;
- Filtre de Kalman Ensemble (EnKF) ;
- Méthode variationnelle simplifiée (1D-VAR) ;
- Méthode variationnelle par réglage simple (T-VAR).

Deux de ces méthodes découlent de la méthode de Kalman standard. Elles sont bien connues dans la communauté scientifique : EKF (Gelb, 1974) et EnKF (Evensen, 1994, Burgers *et al.*, 1998). Dans l'article inclus à la section 4.2 le fonctionnement de ces méthodes est expliqué. Elles représentent une application des filtres de Kalman standard au cas de systèmes non-linéaires, mais avec deux approches très différentes. Les deux autres méthodes sont issues des méthodes variationnelles, même si elles présentent des analogies avec le filtre de Kalman standard (Lorenz, 1986). Elles permettent d'éviter le calcul du gradient de la fonction coût générale $\mathcal{J}$ qui utilise des modèles adjoints. Cela évite le codage d'un modèle adjoint et sa maintenance. Le 1D-VAR simplifié (Balsamo *et al.*, 2004) linéarise le modèle en utilisant une méthode par différences finies et on ne fait pas de minimisation pour résoudre le gradient de $\mathcal{J}$. Le T-VAR (Calvet *et al.*, 1998a) est une méthode simple, démonstrative, qui n'utilise aucune information à priori et aucune spécification des erreurs d'observation. Ces deux méthodes variationnelles utilisent, simultanément, plusieurs observations réparties sur une fenêtre d'assimilation, tandis que l'EKF et l'EnKF corrigent l'ébauche du système d'une façon séquentielle.

L'échelle locale de notre site expérimental est favorable pour tester toutes ces méthodes de manière approfondie sur une longue période de temps (4 ans) avec un temps de calcul raisonnable. De plus, les études à échelle locale, où les paramètres du sol, de la végétation et du forçage atmosphérique sont mieux connus, sont nécessaires pour étudier la réponse des systèmes d'assimilation et effectuer les calages indispensables avant de considérer leur extension spatiale.

Ce chapitre est divisé en deux parties :

- Dans la première section, un article publié dans "Journal of Hydrometeorology" est présenté. Dans cet article est abordé le sujet de l'assimilation des observations de w_g pour l'analyse de w_2 avec les quatre méthodes d'assimilation citées plus haut ;
- La section 4.3 discute les résultats obtenus à partir d'une méthode dérivée de l'EnKF.

4.2 "From near-surface to root-zone soil moisture using different assimilation techniques."

4.2.1 Résumé

L'objectif principal de cet article est la comparaison de quatre méthodes d'assimilation d'observations de w_g dans le modèle ISBA-A-g_s, pour la ré-initialisation de la réserve utile du sol. Bien que la version d'ISBA utilisée dans cet article soit capable de simuler la biomasse de la végétation, le modèle est forcé avec les mesures de LAI réalisées sur le site en jachère de SMOS-REX. La période d'analyse comprend quatre années, de 2001 à 2004, et inclut des périodes de forte sécheresse. Les quatre méthodes testées sont basées sur des hypothèses contrastées et incluent des approches séquentielles et variationnelles.

On veut identifier quelle est la meilleure méthode en terme de performance des analyses et en coût numérique de l'assimilation. Les paramètres de l'assimilation sont discutés dans la partie décrivant l'implémentation. La méthode 1D-VAR retenue sera appliquée à une échelle régionale pour une possible implémentation dans une configuration opérationnelle dans le schéma de surface de Météo-France.

La comparaison entre les différentes méthodes montre les avantages du 1D-VAR et du EnKF par rapport au EKF et T-VAR, pour l'analyse du contenu en eau du sol. De plus, le temps de calcul inférieur obtenu grâce au 1D-VAR et son bon comportement face à différents niveaux d'erreur de background et des observations, font de lui la méthode la plus adaptée pour être implémentée de façon opérationnelle.

FROM NEAR-SURFACE TO ROOT-ZONE SOIL MOISTURE USING DIFFERENT ASSIMILATION TECHNIQUES.

Joaquín Muñoz Sabater, Lionel Jarlan, Jean-Christophe Calvet, François Bouyssel,

Météo-France / CNRM, Toulouse, France

and Patricia De Rosnay

Centre d'Etudes Spatiales de la Biosphère, Toulouse, France

ABSTRACT

Root zone soil moisture constitutes an important variable for hydrological and for weather forecast models. Microwave radiometers like the L-band instrument onboard the ESA's future Soil Moisture and Ocean Salinity (SMOS) mission are being designed to provide estimates of near surface soil moisture (0-5 cm). This quantity is physically related to root-zone soil moisture through diffusion processes, and both surface and root-zone soil layers are commonly simulated by Land Surface Models (LSM). Observed time series of surface soil moisture may be used to analyse the root-zone soil moisture using data assimilation systems. In this paper, various assimilation techniques derived from Kalman filters (KF) and variational methods (VAR) are implemented and tested. The objective is to correct the modeled root-zone soil moisture deficiencies of the ISBA-A-gs LSM, using the observations of the surface soil moisture of the SMOSREX experiment over a four year period (2001-2004). This time period includes contrasting climatic conditions. Among the different algorithms, the EnKF and a simplified 1D-VAR show the best performances. The lower computational cost of the 1D-VAR is an advantage for operational root zone soil moisture analysis based on remotely sensed surface soil moisture observations at a global scale.

1. Introduction

The accuracy of short term to seasonal weather predictions depends on a good initialization of several surface variables of slow variation in the coupled land surface-atmosphere system. Among these variables, root-zone soil moisture is of prime importance.

Root-zone soil moisture plays a vital role in the regulation of water and energy budgets at the soil-vegetation-atmosphere interface through evaporation processes of the uppermost surface soil layer and plant transpiration (Shukla and Mintz, 1982). If the initialization of this variable is not accurate, significant drifts of the temporal

evolution of the surface state variables may develop and may consequently cause a degradation of the weather forecast (Beljaars et al.1996, Dirmeyer 2000, Koster and Suarez 2003).

Land surface models (LSM) aim to describe the continental lower boundary conditions for Numerical Weather Prediction (NWP) models (i.e., water and energy exchanges). They are now able to simulate the main processes of the surface functioning (e.g. soil water dynamics, vegetation-hydrology interaction, water and energy fluxes) but are still limited by several constraints: (1) the need of a high amount of input data (soil and vegetation characteristics) that cannot be provided

accurately at large scales, (2) the incompatibility between the relatively low spatial scale ($\sim 0.1\text{-}1\text{ km}$) of surface and hydrological processes (in particular run-off, sub-surface flow), and the grid scale of NWP models ($\sim 10\text{-}100\text{ km}$) and (3) the meteorological forcing errors, especially for rainfall, which has the most significant influence on soil moisture variability. These constraints have an effect on the simulation of soil moisture evolution within the LSM and may adversely affect the quality of the weather predictions.

A possible solution to improve simulations of LSM is to assimilate observations sensitive to soil moisture by using data assimilation schemes. Operational optimal interpolation systems for NWP models have been developed (Giard and Bazile, 2000, Douville et al., 2000) with the aim to analyze soil moisture by incorporating air temperature and humidity observations at a height of 2 meters over the soil surface. Within the framework of the ELDAS (European Land Data Assimilation System) project, Balsamo et al. (2004) have tested a simplified variational system. Nevertheless, air temperature and humidity are quite indirectly linked to soil moisture. A more direct source of information is provided by L-band microwave remote sensing, which links the observed brightness temperature (T_B) to the surface soil moisture (top 0-5 cm). These observations show a large sensitivity to soil moisture variations (Eagleman and Lin 1976, Wigneron et al. 2002) and, they could be included in NWP systems by assimilating T_B directly, which requires a radiative transfer model, or by assimilating derived soil moisture products. The potential of the analysis of root-zone soil moisture (w_2) from surface soil moisture observations (w_g) was highlighted by Calvet et al. (1998a) and Calvet and Noilhan (2000). Several authors have already conducted the analysis of w_2 at a local scale using observations of microwave brightness temperatures (Houser et al., 1998), synthetic observations (Reichle et al., 2002) or soil surface moisture retrievals from the

Synthetic Aperture Radar on board the ERS satellites (François et al., 2003). Global scale w_g products also exist like those provided by the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) sensor (Njoku et al. 2003), onboard NASA's AQUA satellite and those derived from the ERS-1 and ERS-2 Scatterometer (Wagner et al., 2003). The future Soil Moisture and Ocean Salinity (SMOS) satellite of the European Space Agency (ESA), planned to be launched in 2007 (Kerr et al., 2001), will provide an estimation of the soil moisture in L-band at a global scale, with a sampling time of around 3 days at the equator and with a spatial resolution compatible with NWP models. If we want to take advantage of this vast amount of available data, assimilation systems have to be developed and integrated within the NWP models (Seuffert et al. 2003, Balsamo et al. 2006). Within this context, this study aims to analyze w_2 by assimilating w_g measurements. The data set used in the present study comes from the SMOSREX (Surface Monitoring of the Soil Reservoir EXperiment) campaign (De Rosnay et al., 2006) over a fallow ground area, which is similar to that used by Calvet et al. (1999) (MUREX: monitoring the usable soil reservoir experimentally). The data set comprises four years of measurements (2001-2004), during which the area underwent very contrasting climatic conditions. In particular the severe drought of the summer of 2003 over Western Europe is well represented. This study follows the work of Calvet and Noilhan (2000) by adding a comparison between several assimilation methods.

Following this introduction, section 2 presents the experimental site, the data set, the LSM ISBA-A-gs and the assimilation methods employed in this study (Extended Kalman Filter -EKF-, Ensemble Kalman Filter -EnKF-, simplified 1-D variational -1D-VAR- and tuning variational -T-VAR-). Furthermore, the methodology to estimate the model error is presented, and the implementation of the four assimilation schemes is described. In section 3, the results

for each assimilation method are shown and a sensitivity study (to model and observation error) is carried out for the best performing method. Finally, section 4 summarizes the main conclusions and prospects.

2. Methodology

a. Experimental site and data set

The SMOSREX site is situated within the ONERA (French National Aerospace Research Establishment) center of Fauga-Mauzac, located 40 km at the South of Toulouse (43°23'N, 1°17'E, 188 m altitude). SMOSREX is a field scale experiment, operative since 2001 with measurements similar to those of MUREX (Calvet et al. 1999). The experimental data set is described below:

1) AUTOMATIC MEASUREMENTS

A meteorological station is providing continuous measurements, every 30 minutes, of precipitation, atmospheric pressure, wind speed and direction, air humidity, air temperature and incident and emitted solar and infrared radiation. Deep and surface soil temperature and soil moisture are monitored continuously on a half-hourly basis. In order to obtain a representative estimate of w_g , four probes (ThetaProbe, Delta T Devices) are vertically installed at different locations within the area, providing a measurement over the top 6 cm of the soil layer. Daily mean values w_g^j are obtained by averaging the four probe measurements. The uncertainty in w_g^j is given by the standard deviation of these measurements. w_2 is obtained by calculating an average bulk soil water content from these surface probes and three ThetaProbe profiles set up at the same locations (1 to 3 m apart). These profiles consist of soil moisture sensors installed vertically at the surface (0-6 cm) and horizontally at depths of 10, 20, 30, 40, 50, 60, 70, 80 and 90 cm. Our goal, here, is to use the information provided by the

experimental data as much as possible in order to define prescribed observation error statistics coherent with the field observations. For lack of a sufficient sampling of the plot, the spatial averaging is replaced by a temporal one (ergodicity principle). From the individual measurements, a mean and a standard deviation are computed on a daily time step. The daily standard deviation averaged over the year 2001 is assumed to be equal to the observation error: $\sigma(w_g^{OBS}) = 0.03 \text{ m}^3\text{m}^{-3}$ and $\sigma(w_2^{OBS}) = 0.02 \text{ m}^3\text{m}^{-3}$. These errors are attributed to the subsequent years (2002-2004).

2) MANUAL MEASUREMENTS

Measurements of the vegetation characteristics (leaf area index (LAI), green and dry biomass and height of the canopy) were carried out every two weeks from spring to autumn. Figure 1 shows the in-situ observations of the LAI , the root-zone soil moisture and the monthly accumulated precipitation for the four years (2001-2004). It can be observed that 2003 was a particularly dry year, with a yearly accumulated precipitation of less than 600 mm. Unlike the other years, 2003 shows an atypical double cycle of LAI , with a first maximum in spring and another one at the beginning of the winter season. Precipitation is quite irregularly distributed during 2004, with a wet spring and a very dry summer. That causes a rapid growth of the vegetation and a marked senescence during the dry period, with w_2 reaching values below wilting point (w_p) during all the summer season and part of autumn.

In Table 1, a list of the most relevant characteristics of the soil over the SMOSREX site for ISBA-A-gs is provided. The soil is a loam characterized by its texture and density which were determined in the laboratory. The wilting point and the field capacity parameters were derived from the clay content observations, by using the relationships given by Noilhan and Mahfouf (1996).

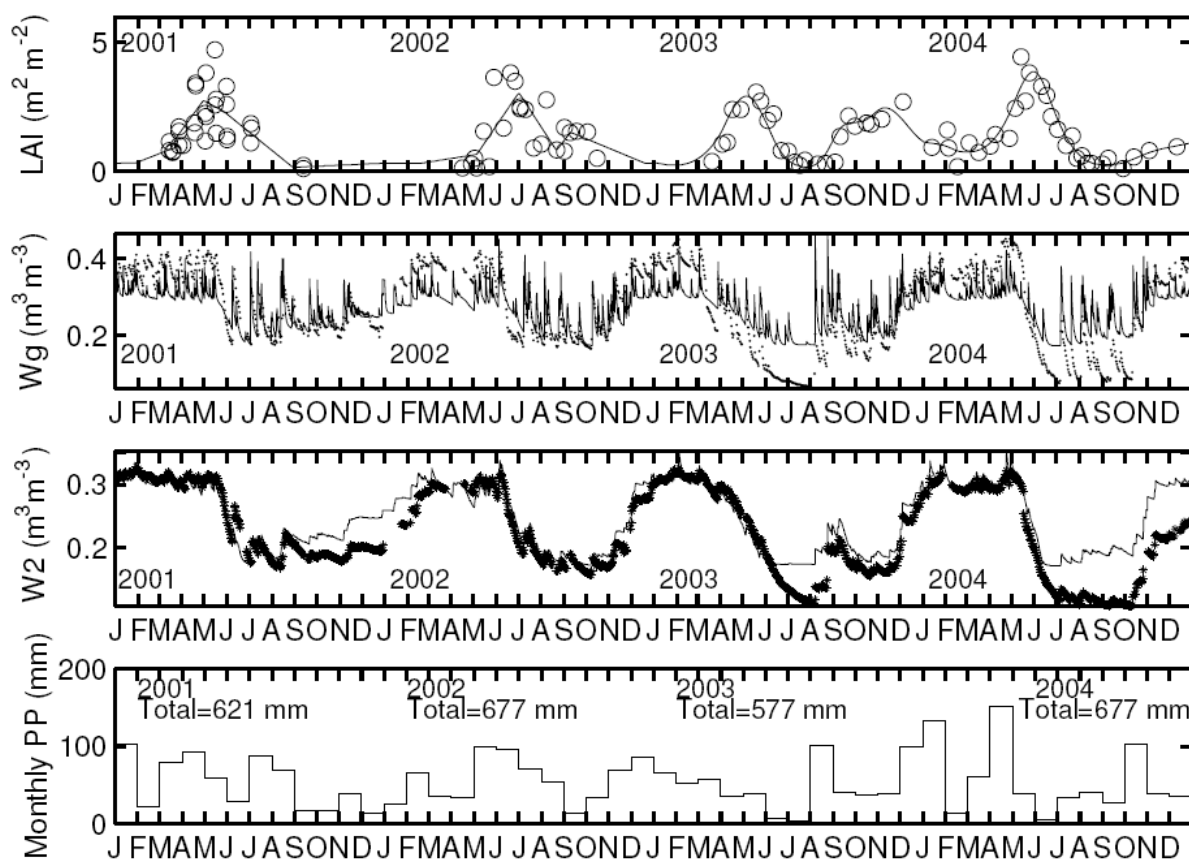


Figure 1.- LAI , w_g , w_2 and monthly precipitation (from top to bottom) for 2001 to 2004 as measured over the SMOSREX site. The interpolated LAI (solid line), is used to drive the ISBA-A-gs simulations and is superimposed on the observations (triangles). Note that the LAI interpolation technique may change from one year to another, depending on data quality and frequency. The w_g and w_2 simulations of ISBA-A-gs (solid line) are superimposed on the observations (dots).

b. Land surface model: ISBA-A-gs

The Interaction between Soil, Biosphere and Atmosphere scheme (ISBA) was firstly developed by Noilhan and Planton (1989) and further improved by Noilhan and Mahfouf (1996) to describe the surface processes in weather and climate models.

The ISBA model uses the equations of the force-restore method (Deardoff 1977, , 1978) to describe the evolution of five surface state variables: surface temperature

(T_s), mean surface temperature (T_2), surface soil volumetric moisture (w_g), total soil volumetric moisture (w_2) and canopy interception reservoir (W_s), together with the surface energy fluxes (LE , H , G). The model was modified in order to account for the effect of the atmospheric carbon dioxide concentration on the stomatal aperture. This new version of ISBA was called ISBA-A-gs (Calvet et al. 1998b, Gibelin et al. 2006). The net assimilation of CO_2 is used to predict the vegetation biomass and the LAI .

TABLE 1.

Main soil and vegetation parameters used in the ISBA-A-gs model over the SMOSREX site.

Soil parameters			
parameter	symbol	Unit	Value
Soil root depth	d_2	cm	95
Sand Content	SAND	%	32.0
Clay content	CLAY	%	22.8
Field capacity	w_{fc}	$m^3 \cdot m^{-3}$	0.30
Wilting point	w_{wilt}	$m^3 \cdot m^{-3}$	0.17
Vegetation parameters			
parameter	symbol	Unit	Value
Mesophyll conductance	g_m	$mm \cdot s^{-1}$	0.56
Critical extractable soil moisture	θ_c	%	50
Plant response to water stress	-	-	drought-tolerant

However, a study of the impact of using an interactive LAI on the w_2 analyses is beyond the scope of this paper and, in this study, the LAI is prescribed from measurements. In Figure 1, the ISBA-A-gs control simulation of w_2 with the prescribed interpolated LAI (solid lines) is superimposed over the observations. The photosynthesis parameters (see Table 1) are the same as those prescribed during the MUREX campaign (Calvet and Soussana, 2001). In section 3 these results are discussed.

c. Assimilation methodologies

1) DERIVED FROM KALMAN FILTERS (KF)

The well-known Kalman filter (Kalman, 1960) is an optimal sequential assimilation method extensively used in various environmental problems. It is based on a minimum variance analysis scheme (see for example Teunissen, 2000). In such a scheme the best linear unbiased estimators (BLUE) yield at time i of the measurements for the variables to be analyzed (w_g and w_2 in

this study, for the KF case, hereafter embedded in the state vector $\mathbf{x}_i$) and for the associated prognostic state variance-covariance matrix $\mathbf{P}_i$ the following equations:

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i [\mathbf{y}_i - \mathbf{H} \mathbf{x}_i^f] \quad [1]$$

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i \mathbf{H}) \mathbf{P}_i^f \quad [2]$$

with

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}^T [\mathbf{H} \mathbf{P}_i^f \mathbf{H}^T + \mathbf{R}_i]^{-1} \quad [3]$$

where the superscripts $(^f)$ and $(^a)$ refer to the point in time just before and after the analysis, respectively; $\mathbf{y}_i$ is the observation vector at time i (w_g^{obs} in this study); $\mathbf{R}_i$ the associated variance-covariance error matrix; $\mathbf{I}$ the identity matrix; and $\mathbf{K}_i$ is called the Kalman gain. In the standard KF, a linear relationship is assumed as follows:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i + u_i \quad [4]$$

where $\mathbf{H}$ is the observation operator, and u_i is a function accounting for the uncertainties of the measurements and the observation model, given the variance-covariance $\mathbf{R}_i$.

The analyzed state (Eq. [1]) is equal to the forecasted state plus the analysis increment which is equal to the innovation vector (the difference between observations $\mathbf{y}_i$ and the associated simulations in the observation space $\mathbf{H}\mathbf{x}_i^f$) multiplied by the Kalman gain $\mathbf{K}_i$. $\mathbf{K}_i$ accounts for the errors in the observations and the prognostic state (the correction will be higher as more confidence is given to the observations). During the propagation step the system evolves according to the linear dynamics of the system:

$$\mathbf{x}_{i+1}^f = \mathbf{M} \mathbf{x}_i^a + w_i \quad [5]$$

where $\mathbf{M}$ is the prognostic model operator and w_i groups all modeled uncertainties (assumed normal distributed with zero mean and covariance equal to $\mathbf{Q}_i$). Finally, for the propagation law of variances, the forecast error covariance matrix $\mathbf{P}$ will evolve according to:

$$\mathbf{P}_{i+1}^f = \mathbf{M}\mathbf{P}_i^a\mathbf{M}^T + \mathbf{Q}_i \quad [6]$$

(i) Extended Kalman filter (EKF)

Equations [1] to [6] describe the system propagation and the optimal updating completely, for linear problems. However, in

most cases, physical systems are non-linear and, in contrast to Eq. [5], the model system propagates in time using the full non-linear dynamics of the system:

$$\mathbf{x}_{i+1}^f = M(\mathbf{x}_i^a) + w_i, \quad [7]$$

where $M()$ is a non linear operator which groups all the model equations. The EKF equations differ from those of the standard KF in the way that the system is locally linearized around the forecasted vector $\mathbf{x}_i^f$ at the time i of the observations. $\mathbf{M}$ of Eq. [6] becomes the Jacobian of the prognostic model:

$$\mathbf{M}_i = \left. \frac{\partial M(\mathbf{x}_i)}{\partial \mathbf{x}_i} \right|_{\mathbf{x}_i = \mathbf{x}_i^f} \quad [8]$$

(ii) Ensemble Kalman Filter (EnKF)

One of the main drawbacks of the EKF is the time consuming process of propagating the variance-covariance matrix $\mathbf{P}_i$ when dealing with systems with a large number of state variables such as the one of NWP models (typically around 10^6). The EnKF (Evensen 1994, Burgers et al., 1998) circumvents this problem by using an ensemble of j state vectors, each of which represents a potential model trajectory.

Equations are those of the standard KF (Eq. [1] to [6]), but now the state variance-covariance matrix $\mathbf{P}_i$ is estimated throughout the statistics of the ensemble:

$$\mathbf{P}_i^f \approx \mathbf{P}_{e,i}^f = \frac{1}{N-1} \sum_{j=1}^N (\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_i^f})^T (\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_i^f}) \quad [9]$$

where $\overline{\mathbf{x}_i^f}$ denotes the average of the ensemble. Moreover, observations are considered as random variables too, and an ensemble of observations is created at each observation time step by artificially perturbing them with Gaussian noise, following Burgers et al. (1998).

The state covariance matrix is implicitly propagated by the ensemble and, unlike the EKF, no linear approximation is

involved. The mean of the ensemble is considered to be the most probable assimilated state and the dispersion of the ensemble will be an approximation of second moment of the model potential trajectory distribution. When the size of the ensemble tends to infinite, the ensemble $\mathbf{P}_{e,i}^f$ matrix will converge to $\mathbf{P}_i^f$ (Evensen, 2003).

2) DERIVED FROM VARIATIONAL METHODS

In the variational method, instead of updating the system each time an observation is available, an assimilation window with a predefined length in time is used. The simulations are adjusted to the observations which are available within the assimilation window (at observations times) by minimizing a cost function J , with respect to a background information $\mathbf{x}^b$. Both the initial state and the model trajectory within the assimilation window are updated. The general form of J is given by:

$$\begin{aligned} J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \frac{\mathbf{B}^{-1}}{2} (\mathbf{x} - \mathbf{x}^b) + \\ &+ (\mathbf{y} - H(\mathbf{x}))^T \frac{\mathbf{R}^{-1}}{2} (\mathbf{y} - H(\mathbf{x})) \\ &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \end{aligned} \quad [10]$$

The cost function in Eq. [10] has two terms: the background term $J_b(\mathbf{x})$ which measures the distance between the state vector $\mathbf{x}$ and the a priori state $\mathbf{x}^b$ (weighted by the background error matrix $\mathbf{B}$), and the observation term $J_o(\mathbf{x})$ which accounts for the distance between the vector of observations during the assimilation window, $\mathbf{y}$, and the simulations weighted by the observation error matrix $\mathbf{R}$. The subscript i has been omitted in Eq. [10] as, in contrast to sequential methods, all the observations available within the assimilation window are considered for variational methods. The projection of the state vector in the observation space is done through the observation operator $H()$ which

is often non linear and includes the integration over time through the model operator $M()$. The minimization of J is generally computed by applying the descent gradient method for which the adjoint and the tangent linear models are needed. Building these models is usually a time consuming task. In this study, as described below, a numerical linearization is used in order to avoid the use of the adjoint and tangent linear models.

(i) Simplified 1D-VAR

The simplified 1D-VAR is an assimilation method developed by Balsamo et al. (2004) and adapted to our study. The observation operator is numerically linearized by perturbing the initial conditions (i.e. the state vector at the beginning of the assimilation window). Under this tangent linear hypothesis the observation operator $H()$ can be developed by a first-order expansion:

$$H(\mathbf{x} + \Delta\mathbf{x}) \approx H(\mathbf{x}) + H(\Delta\mathbf{x}) \quad [11]$$

The minimum of the cost function is given by $\nabla J = 0$ and, with the hypothesis that errors follow a normal distribution, it takes the following form:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K} (\mathbf{y} - H(\mathbf{x}^b)) \quad [12]$$

with

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T [\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad [13]$$

Note that the 1D-VAR analyses in this study only concern w_2 (1×1 control vector), whereas the KF analyses concern w_2 and w_g (2×1 control vector). Indeed, a control vector of two variables does not imply higher computational costs for the KF, whereas the inclusion of the w_g analysis in the 1D-VAR would require an extra run.

(ii) *Variational tuning method (T-VAR)*

T-VAR is a simplified sub-optimal variational method introduced by Calvet and Noilhan (2000). It has the ability to retrieve soil moisture estimates of the deep reservoir by using a window of 10 days with four independent observations (if no missing values) of w_g . They are globally adjusted to the model estimates by a systematic exploration of all the potential initial values of the root-zone soil moisture. Here, the control vector is composed of w_2 , only (w_g is not analyzed). The retrieved value of w_2 corresponds to the minimum of the root mean square error (RMSE), which it is in fact the cost function $\mathbf{J}$ to be minimized, without a background term and with $\mathbf{R} = \mathbf{I}$. Although a minimization is performed by systematic search of model initial state that best-fit the observations, there is no optimal use of error statistics.

3) *METHODOLOGICAL DISCUSSION*

In this study, three methods over four rely on the linear least square theory: EKF, EnKF and 1DVAR. Besides this apparent similarity, differences in the analysis calculation exist. First, EKF and 1DVAR rely on the local linearization of the model equations whereas the full non linear system dynamics is accounted for by the EnKF. The linearization of the model is valid if the time

step between two observations is smaller than the correlation length of the state variables. This is not always the case, particularly for the rapid variations affecting surface soil moisture. In the case of highly non-linear models, the filter diverges from optimality and may become unstable (Miller et al., 1999). For example, in our case the linear hypothesis may not work when strong precipitation or evaporation rates take place. A decoupling between both variables may then lead to inadequate corrections.

For Kalman filters, the background error covariance is sequentially updated. The information stored in the covariance matrix is propagated in time and thus, extends the coherence of the assimilation beyond the assimilation time-window. The propagation of the covariance information is done through the linear model for EKF whereas it is implicit thanks to a stochastic sampling of the *a priori* space for EnKF. In contrast to the Kalman type methods, a fixed background error matrix is assumed for the 1D-Var.

A fourth method is tested, in this study, which mimics the case where no *a priori* information is available. This method is often used on extended time-windows, e.g. one year, to tune parameters related to soil moisture, like field capacity. Although it is clearly inferior to the other methods, this simple algorithm permits to assess to what extent w_2 can be analyzed when both the quality of the observations is not known and *a priori* estimates of w_2 are unavailable.

d. Implementation of the assimilation methods

In this subsection the practical implementation of the assimilation algorithms employed in this study and the requirement for working in a normalized space are described. A description of the model forecast and background covariance error matrices (**P** and **B**) and the observation error matrix **R** is presented for the four assimilation methods in Table 2. Note that we have set the observation error **R** to twice the uncertainty of the observations ($\sigma(w_g^{OBS})$). This step was taken, because the experimental setup does not permit to quantify the spatial representativeness error. To take this effect into account the error in the observations was inflated empirically by a factor of 2.

1) NORMALIZATION OF THE STATE VARIABLES

Calvet and Noilhan (2000) pointed out the need to normalize soil moisture before any data assimilation is undertaken, because of an existing bias in the simulated w_g . Indeed, w_g is a model-dependent variable. In particular, peak values within a year of w_g are likely to vary from one model to another. In Figure 2 the comparison between the w_g observations and the ISBA-A-gs simulations are plotted for 2001. It is shown that their relationship is far from the 1:1 line, which would be the case if the model were perfect. Therefore, all soil moisture observations for the period of 2001-2004 were normalized using the maximum and minimum values observed in 2001, which was chosen as the calibration year. Similarly, all soil moisture

TABLE 2.

Definition of the background error matrix **P** and the observation error matrix **R**, for four assimilation schemes: Ensemble Kalman Filter (EnKF), Extended Kalman Filter (EKF), a simplified 1D-VAR and a tuning variational method (T-VAR).

	Background error matrix P	Observation error matrix R
EnKF	$\mathbf{P}_i \approx \frac{(\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_{j,i}^f})^T (\mathbf{x}_{j,i}^f - \overline{\mathbf{x}_{j,i}^f})}{N-1}$	$\mathbf{R}_i = \frac{(\mathbf{y}_{j,i} - \overline{\mathbf{y}_{j,i}})^T (\mathbf{y}_{j,i} - \overline{\mathbf{y}_{j,i}})}{N-1}$
1D-VAR	$\mathbf{B} = \mathbf{Q}^{\text{var}}$	$\mathbf{R} = (2\sigma(w_g^{obs}))^2$
EKF	$\mathbf{P}_i^f = \mathbf{M}_i \mathbf{P}_{i-1}^a \mathbf{M}_i^T + \mathbf{Q}_i$	$\mathbf{R}_i = (2\sigma(w_g^{obs}))^2$
T-VAR	no background term	$\mathbf{R} = \mathbf{I}$

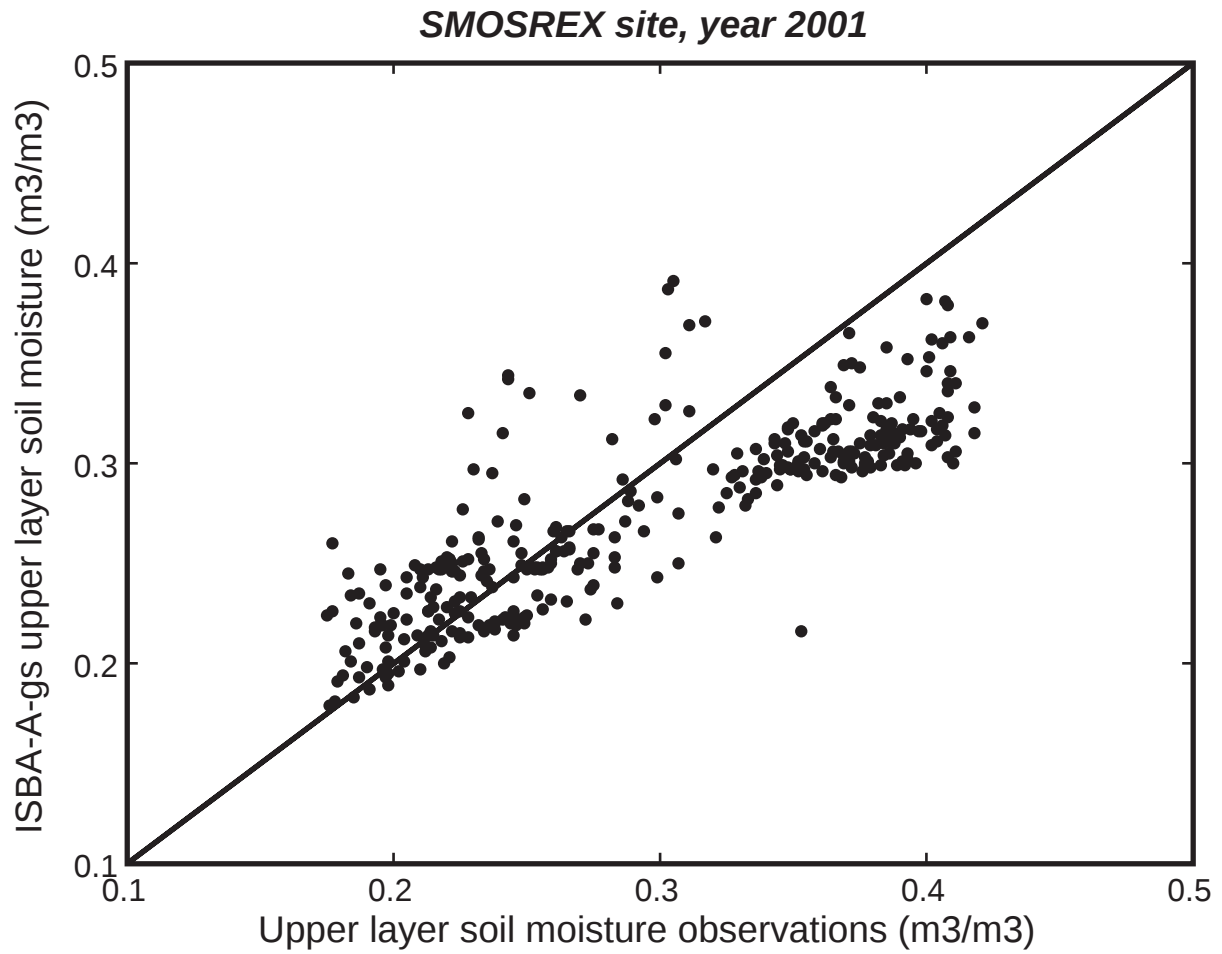


Figure 2.- Observed versus modeled surface soil moisture (w_g) for 2001. The scatter plot is compared with the 1:1 relationship (solid line).

simulations were normalized between the maximum and minimum values estimated from the model simulations for the year 2001. In this way observations and simulations can be compared in a normalized space (w_g and w_2 ranging from 0 to 1).

3) IMPLEMENTATION

In this subsection, the technique to estimate the model error is described, as well as the approach to apply the assimilation schemes.

(i) Model error

An accurate estimation of the model error is likely to be the most difficult task in the errors prescription (Bouttier 1994, Reichle et al. 2002, François et al. 2003). An approximation of the model imperfections

was obtained as follows: at each observation time, an ensemble of initial conditions was created for the two state variables (w_g and w_2) from the observation values. The model was then run for each member of the ensemble, every three days (which mimics the frequency of satellite derived w_g observations) to estimate the forecasted error for the sequential methods and every 10 days, i.e. the assimilation window duration, to estimate the background error for the 1D-VAR. At the end of each time window (3 or 10 days), the dispersion of the residuals (difference between an ensemble member and the observation value) was calculated. This value is considered as the q_{x_i} term of the $\mathbf{Q}^{type}$ (*type* being sequential *-seq-* or variational *-var-*) matrix at time i , and an annual evolution of this term is obtained.

2001 was chosen as the calibration year because of its characteristics in terms of atmospheric forcing, which resembled an average year. If the model error is stationary, the temporal evolution is close to constant. This is the case for the state variable w_2 and to a lesser extent for w_g , due to its shorter temporal correlation length. Nevertheless, from here onwards, the estimated forecast and background error are defined as follows:

$$\mathbf{Q}^{seq} = \begin{bmatrix} 0.05^2 (m^3 m^{-3})^2 & 0 \\ 0 & 0.02^2 (m^3 m^{-3})^2 \end{bmatrix}$$

$$\mathbf{Q}^{var} = \begin{bmatrix} 0.06^2 (m^3 m^{-3})^2 & 0 \\ 0 & 0.01^2 (m^3 m^{-3})^2 \end{bmatrix} \quad [14]$$

In Eq. [14], it is assumed that there is no correlation between the model error on w_g and the model error on w_2 , by setting the non-diagonal terms to zero. Indeed, preliminary calculations (not shown) of the cross-correlation terms produced negligible values. The introduction of these values into the covariance matrix had only minor effects on the results of the assimilation.

(ii) EKF

The assimilation of remote sensing data into LSMs usually constitutes a low dimensional problem in comparison with the assimilation of observations in atmospheric or oceanic models. Therefore the propagation of the model error covariance matrix is rather straightforward and methods like the EKF can be tested easily.

An initial background error covariance matrix $\mathbf{P}_0$ is constructed using the uncertainty of the observations and assuming no initial correlation between the state variables, hence a block diagonal covariance matrix. In order to propagate $\mathbf{P}$ between observations and apply the tangent linear hypothesis, a perturbation of the initial state vector (composed in the present case of w_g and w_2) is carried out, yielding the

numerical linear matrix $\mathbf{M}_i$ (which is substituted into Eq. [6]):

$$\mathbf{M}_i = \begin{bmatrix} \frac{M_i(w_g^{(i-1)} + \Delta w_g^{(i-1)}) - M_i(w_g^{(i-1)})}{\Delta w_g^{(i-1)}} \\ \frac{M_i(w_2^{(i-1)} + \Delta w_2^{(i-1)}) - M_i(w_2^{(i-1)})}{\Delta w_2^{(i-1)}} \\ \frac{M_i(w_g^{(i-1)} + \Delta w_g^{(i-1)}) - M_i(w_g^{(i-1)})}{\Delta w_2^{(i-1)}} \\ \frac{M_i(w_2^{(i-1)} + \Delta w_2^{(i-1)}) - M_i(w_2^{(i-1)})}{\Delta w_g^{(i-1)}} \end{bmatrix} \quad [15]$$

where $M_i()$ is the non-linear full operator at time i , and Δw_g and Δw_2 are the perturbations of the updated state variables w_g and w_2 at the precedent time $i-1$, respectively. The size of the perturbation of the initial state is critical. Theoretically, an infinitesimal perturbation in the neighborhood of the initial state vector would ensure that the linear hypothesis is fulfilled. However it may cause an adverse effect due to numerical errors. Large perturbations may also produce errors when non-linear effects are predominant. In this study a value of $0.05 m^3 m^{-3}$ was chosen as perturbation of the initial state and the error variance-covariance matrix of the forecasted state is integrated in time using Eq. [6]. Since the state variable w_g is directly observed, the observation operator $\mathbf{H}$ is in this case: $\mathbf{H} = [1 \ 0]$. Thus, developing Eq. [3] and combining it with Eq. [1], the corrections of the forecasted state variables at the time i of the observation are given by:

$$w_{g,i}^a = w_{g,i}^f + \frac{\mathbf{P}_i(1,1)}{\mathbf{P}_i(1,1) + \mathbf{R}_i} (w_{g,i}^{obs} - w_{g,i}^f) \quad [16]$$

and

$$w_{2,i}^a = w_{2,i}^f + \frac{\mathbf{P}_i(1,2)}{\mathbf{P}_i(1,1) + \mathbf{R}_i} (w_{g,i}^{obs} - w_{g,i}^f) \quad [17]$$

where the $\mathbf{P}_i(1,1)$ and $\mathbf{P}_i(1,2)$ terms are elements of the 2×2 $\mathbf{P}_i$ matrix.

(iii) EnKF

Samples of the initial ensemble are created assuming a normal distribution with a mean equal to the first observation and a variance-covariance matrix equal to $\mathbf{R}$. An ensemble of $N=100$ members is used following Evensen (2003). A rapid convergence of the forecasted w_2 is observed (not shown) and the analysis ensemble tends to collapse. Physically, the explanation is that the water loss by evaporation modulate the root zone soil moisture: the members of the ensemble starting with a wet soil are loosing more water than the drier soils. This tends to make w_2 converge to the same value on average over a year. In order to prevent the collapse of the ensemble, the new ensembles are built by multiplying the variance of the updated ensemble by an inflation factor. This approach is equivalent to the covariance inflation described in Anderson and Anderson (1999). Moreover the atmospheric forcing is perturbed at each model step by adding random Gaussian noise. The inflation factor was empirically calibrated by minimizing the RMSE between the analyzed and the observed w_2 for the year 2001. A value of 1.35 was found and kept constant for the other years. For the perturbation of the atmospheric forcing, the following standard deviations were used: 60 Wm^{-2} , 35 Wm^{-2} , 50% relative difference, 10 K, 1 ms^{-1} and 1000 Pa, for shortwave and longwave incident radiation, precipitation, air temperature, wind speed and surface pressure, respectively. As for the EKF, the observation operator remains $\mathbf{H} = [1 \ 0]$. Finally, since information about the observation error is obtained from the ThetaProbe measurements, an ensemble of normally distributed observations is created, with σ equal to twice the observation error of w_g .

(iv) Simplified 1D-VAR

In this simplified version of the 1D-VAR the $\mathbf{B}$ and $\mathbf{R}$ variance-covariance matrices are estimated once and remain unchanged for the rest of the assimilation period. After the updating step, the assimilation window slides in time until a new observation is found (minimum of three days). Since observations are assimilated more than once, this method departs from optimality in theory. Nevertheless, in this context, rather than searching for the optimality, our objective is the comparison of this assimilation approach with the T-VAR in the fairest way. For an operational application, it is recommended using sequential assimilation windows and suppressing the first observation within the assimilation window. In this study, this does not adversely affect the analyses (not shown). The background matrix has been left fixed and equal to $\mathbf{Q}^{\text{var}}$ (Eq.[14]). On the other hand, in consistency with the sequential methods, the variance of the observations has been set to twice the uncertainty of the observations of w_g . The linearization of the model is done through the observation operator $\mathbf{H}$ by perturbing the initial state of w_2 . Finally, the magnitude of the perturbation has been set to the same value as for EKF.

(v) T-VAR

In this simple variational method no initial prescription of variance-covariance error matrices, nor background information, are needed. At each analysis step a systematic exploration of the potential initial conditions is conducted, by initialising the model with values between $0.10 \text{ m}^3 \text{ m}^{-3} (w_{2,\text{min}})$ and $0.40 \text{ m}^3 \text{ m}^{-3} (w_{2,\text{max}})$, incremented by steps of $0.015 \text{ m}^3 \text{ m}^{-3}$. The analysis of w_2 is undertaken by minimizing the cost function

between observations w_g^{obs} and model estimations w_g^{sim} :

$$J(w_2) = \sqrt{\frac{1}{n} \sum_{i=1}^n (w_{i,g}^{sim} - w_{i,g}^{obs})^2} \quad [18]$$

with n the number of measurements within the assimilation window. According to Calvet and Noilhan (2000), if the w_g^{obs} are available every 3 days, a 10 days assimilation window yields the best results for the MUREX site (i.e. using four observations).

3. Results and discussion

a. Root zone soil moisture simulation (2001-2004)

In Figure 1 the ISBA-A-gs simulations of w_2 and w_g are compared with the observations during the period 2001-2004. Error statistics are given in Table 3 (RMSE,

bias, and skill score). The skill score E is defined as:

$$E = 1 - \frac{\sum_i (x_i^{obs} - x_i^{mod/ana})^2}{\sum_i (x_i^{obs} - \bar{x}^{obs})^2} \quad [19]$$

where x refers to the soil moisture variables, either observed (*obs*), simulated (*mod*) or analysed (*ana*).

In general, the agreement is good as long as the observed w_2 is above the w_p . The model overestimates w_2 from September 2001 to March 2002. This may be due to the lack of regular *LAI* measurements during this period and to the underestimation of *LAI* by the linear interpolation employed. Low values of *LAI* tend to decrease the root water extraction and transpiration rate, leading to an overestimation of the soil moisture with regard to the observations during this period. The model overestimates w_2 also during the droughts of the summers of 2003 and 2004. In this case, the modelled w_2 reaches the prescribed w_p value and

TABLE 3.

Surface and root-zone soil moisture yearly scores of the control simulation and for the whole 2001-2004 period (RMSE ($\text{m}^3 \text{m}^{-3}$), bias ($\text{m}^3 \text{m}^{-3}$) and skill score E).

	2001			2002			2003			2004			2001-2004		
	RMSE	bias	E	RMSE	bias	E	RMSE	bias	E	RMSE	bias	E	RMSE	bias	E
w_g	0.05	-0.020	0.56	0.06	-0.023	0.50	0.08	0.027	0.51	0.08	0.006	0.60	0.07	-0.03	0.59
w_2	0.02	0.010	0.82	0.02	0.009	0.92	0.03	0.015	0.83	0.05	0.044	0.51	0.03	0.020	0.73

root extraction stops, whereas in reality, evaporation may continue even with a soil moisture below the prescribed wilting point. Our goal is to investigate to what extent the assimilation schemes used here are able to improve the model simulation.

b. Analysis of the root-zone soil moisture: EKF and T-VAR.

Figure 3 shows the results of the EKF and T-VAR analysis of the root-zone soil moisture. The analyses, the observations and the free model simulations are plotted together for comparison purposes (hereafter, all the soil moisture results are given in absolute units of m^3m^{-3}). By using the EKF, an enhancement is achieved with regard to the control model simulation ($E=0.85$ against $E=0.73$). Small improvements are achieved for the years 2003 ($E=0.93$) and 2004 ($E=0.66$), where w_2 goes below the wilting point. Here, the non-linear effects observed during the dry periods trigger large Kalman gains, and a significant correction of w_2 . For 2001 and 2002, the EKF tends to degrade the model estimations (Table 4). This is a consequence of the lack of sensitivity of the state variables to a single perturbation of the surface soil moisture between two observations (3 days). This means that, except for periods of large recharge or high evaporation rates, the system behaves stable to perturbations of w_g . As a consequence, the product $\mathbf{MPM}^T$ of Eq. [6] often acquires a very low value. Hence, the forecasted $\mathbf{P}_i$ matrix is mainly controlled by the estimated model covariance error matrix $\mathbf{Q}_i$ (see Eq. [6]). In a first approximation, the correction of $w_{2,i}^f$ (Eq. [17]) is proportional to the $\mathbf{Q}_{12}$ term of the model error matrix, which was assumed to be zero (Eq. [14]). Therefore, by using this method, it is expected that only small corrections of the $w_{2,i}^f$ a priori estimate are undertaken. With the EKF, the control simulations are virtually unchanged during the first half of 2003, and since the control model fits the observations well, a very good skill score is obtained.

Therefore, even though an apparently good performance of the EKF is obtained ($E=0.85$ for the whole period), the corrections are small for the major part of the period. This is attributed to the force-restore scheme of ISBA which produces a low sensitivity of the surface soil moisture to a perturbation of the root-zone soil moisture, during the days following the perturbation. Concerning T-VAR, a high scattering of the retrieved w_2 is observed (Figure 3), which deteriorates the skill score with regard to the model estimation. In general this method is able to reproduce the overall shape of the evolution of the w_2 observations, but also the limitations are obvious, since no background information is used. Locally, retrieved points are found within the uncertainty of the observations, and either during the drought of 2003 or 2004, the analyzed w_2 is below the wilting point, which confirms the potential of this method to retrieve information on w_2 in spite of the lack of any background information.

c. Analysis of the root-zone soil moisture: EnKF and simplified 1D-VAR.

In Figure 4 the results obtained with the EnKF and 1D-VAR are shown. For both methods, a significant overall improvement of the control model simulations is achieved (see Tables 3 and 4), in particular the model overestimation at the end of 2001 and at the beginning of 2002 is corrected. The 1D-VAR shows a slightly better performance during periods where the simulations are limited by the prescribed wilting point, i.e., the summers of 2003 and 2004. In our study case, the EnKF outperforms 1D-VAR for 2001, which is the calibration year for the inflation factor, but the 1D-VAR is better, on average, for the whole 2001-2004 period ($RMSE = 0.02 m^3 m^{-3}$ and $E = 0.86$, compared to $RMSE = 0.03 m^3 m^{-3}$ and $E = 0.78$ for an EnKF with 100 members). Furthermore, the 1D-VAR analyses are smoother than those of the EnKF.

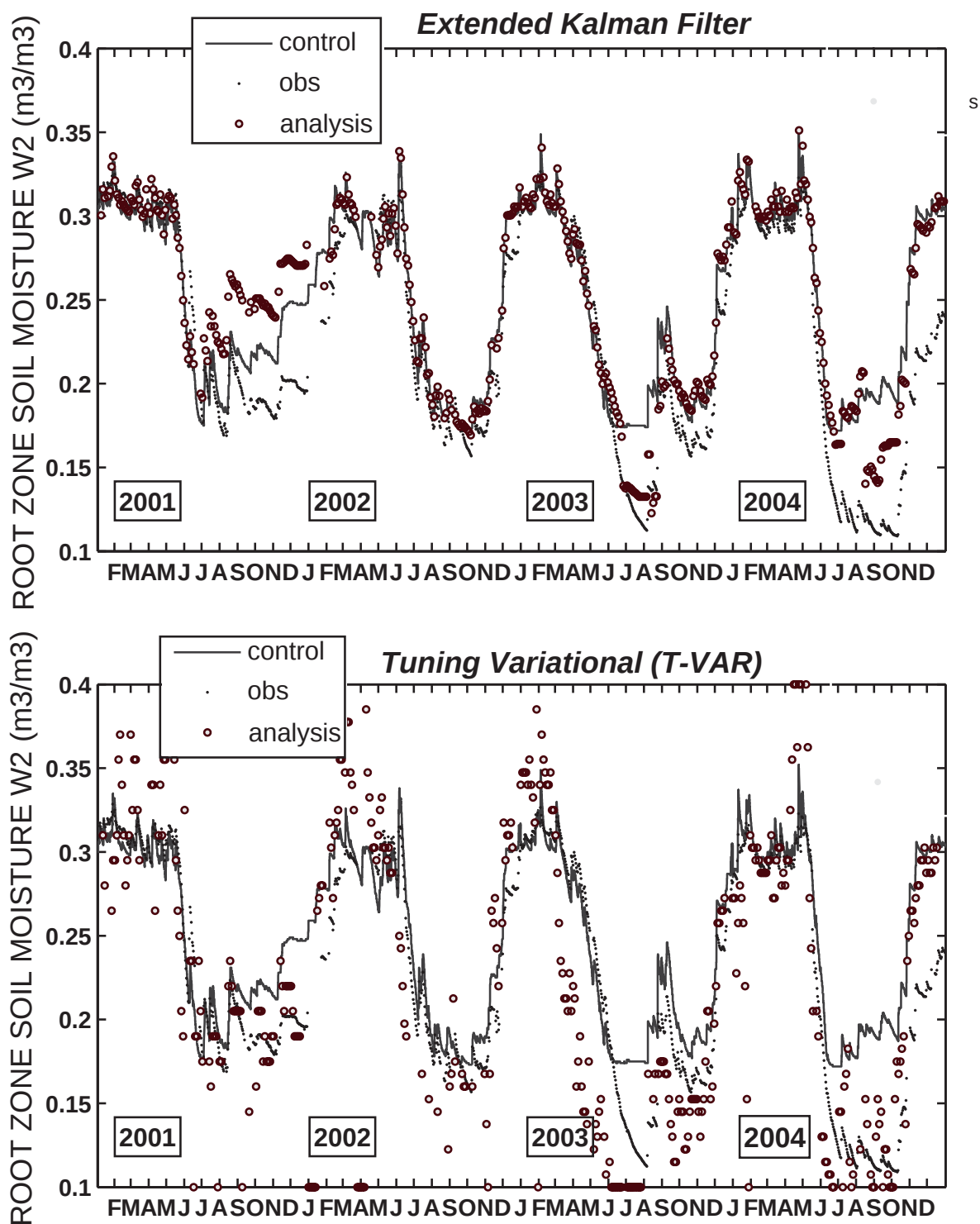


Figure 3.- Analysis of the root-zone soil moisture from surface observations using an Extended Kalman Filter (top) and a tuning variational method (bottom), for 2001-2004 over the SMOSREX experimental site: model control simulation (solid line), observations (dots) and analysis (circles).

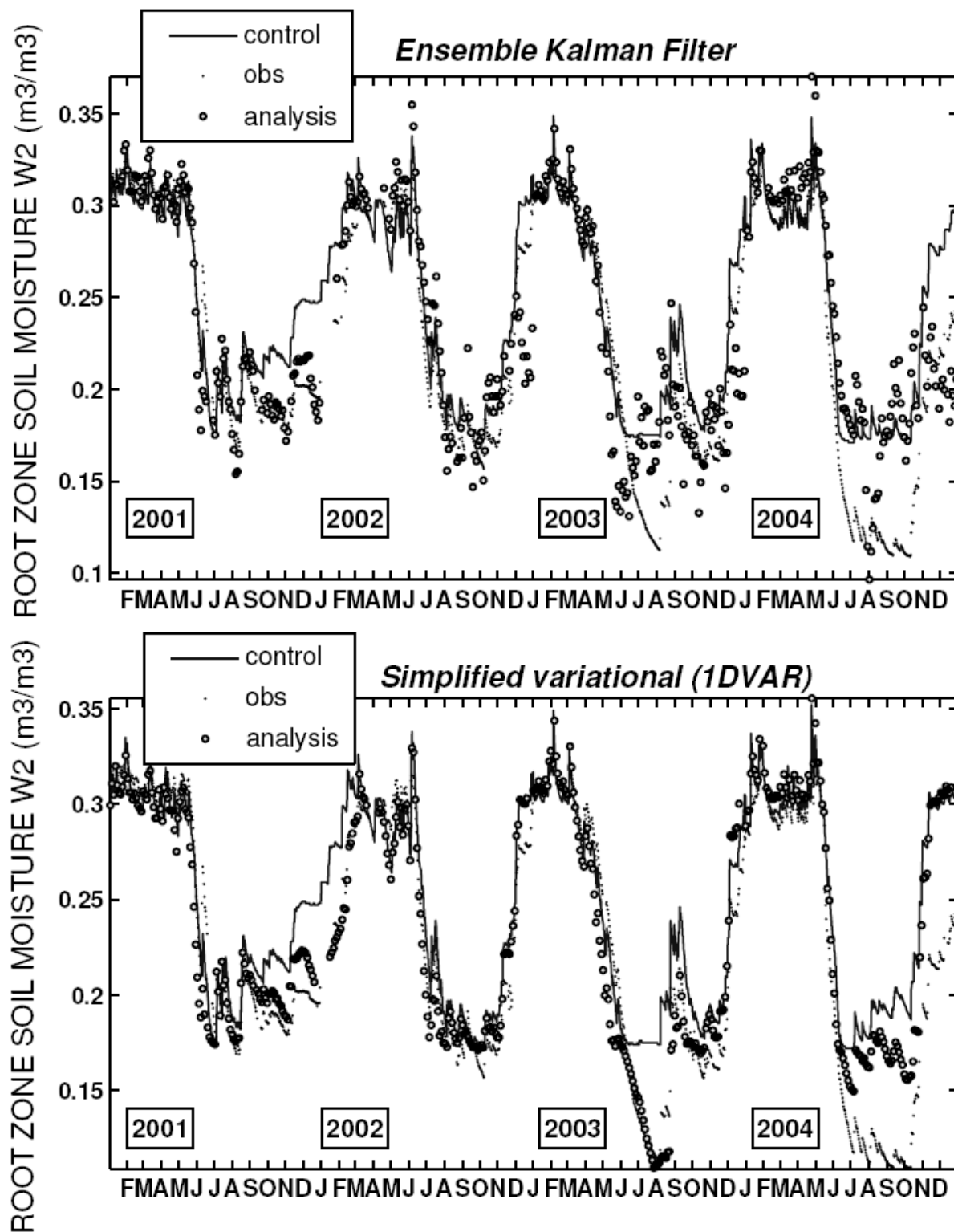


Figure 4.- Analysis of the root-zone soil moisture from surface observations using an Ensemble Kalman Filter (top) and a simplified 1D-VAR method (bottom), for 2001-2004 over the SMOSREX experimental site: model control simulation (solid line), observations (dots) and analysis (circles).

TABLE 4.

Root-zone soil moisture analysis yearly scores and for the whole 2001-2004 period (RMSE ($\text{m}^3 \text{m}^{-3}$), bias ($\text{m}^3 \text{m}^{-3}$) and skill score E), using an Ensemble Kalman Filter (EnKF) with N=10, 20, 50, 100 and 200 members, an Extended Kalman Filter (EKF), a simplified 1D-VAR and a tuning variational method (T-VAR).

	2001			2002			2003			2004			2001-2004		
	RMS E	bias	E	RMS E	bias	E	RMS E	bias	E	RMS E	Bias	E	RMS E	bias	E
EnKF (N=10)	0.02	-0.001	0.91	0.05	0.043	0.15	0.07	0.052	0.12	0.08	0.064	0.05	0.06	0.040	0.25
(N=20)	0.02	0.001	0.91	0.05	0.044	0.18	0.04	-0.008	0.71	0.07	0.065	0.15	0.05	0.024	0.43
(N=50)	0.02	-0.001	0.92	0.03	0.004	0.62	0.03	-0.005	0.77	0.06	0.065	0.19	0.04	0.016	0.58
(N=100)	0.01	-0.001	0.94	0.03	0.006	0.76	0.03	-0.005	0.74	0.04	0.029	0.68	0.03	0.009	0.78
(N=200)	0.01	-0.001	0.94	0.03	0.007	0.67	0.03	0.004	0.82	0.04	0.028	0.76	0.03	0.010	0.81
1D-VAR	0.02	-0.011	0.90	0.01	0.001	0.93	0.02	-0.009	0.89	0.04	0.029	0.79	0.02	0.002	0.86
EKF	0.04	0.026	0.44	0.02	0.014	0.81	0.02	0.010	0.93	0.05	0.037	0.66	0.03	0.011	0.85
T-VAR	0.04	-0.056	0.50	0.04	-0.006	0.37	0.06	-0.026	0.32	0.05	0.009	0.58	0.05	0.014	0.53

The main shortcoming of the EnKF is observed close to wilting point. Indeed, the main difference between the two methods consists in the EnKF background error covariance matrix propagated by the ensemble, as opposed to the 1D-VAR fixed background error. Close to wilting point, the spread of the ensemble broadens significantly (not shown) and the analysis does not match the observations. These results suggest that, over the SMOSREX site, the analyses are more stable and accurate by using a fixed background error. A possible explanation is that the normalization of w_g , performed for 2001, does not eliminate all the seasonal biases.

The assimilation in the transition period between 2001 and 2002 and the drought of 2003 shows a much better performance of the 1D-VAR, resulting in better yearly skill scores for 2002, 2003 and 2004 (Table 4). On the other hand, the EnKF performs better than the 1D-VAR in 2004. To understand how the model estimations are corrected by the 1D-VAR method, Figure 5 shows an example of the temporal

evolution of w_2 along with the gain components for the year 2003 (Figure 5a and 5b, respectively). In our case the gain of the 1D-VAR is a vector of four components (one for each observation within the assimilation window). The beginning of the assimilation time window coincides with the first available observation (see section 2-d-2-iv), therefore, the first component of the gain is negligible with regard to the other three. Then in Figure 5 (bottom) only the last three components of the gain are considered. It is observed that with the soil moisture at field capacity the greatest term of the gain is generally the second term, i.e. the one corresponding to the difference between the second observation (within the assimilation window) and the model estimate. For the following observations, small corrections of w_2 were undertaken, of the order of 5%. Non-linearities are more significant during the rest of the hydrological cycle, when no privileged gain component is found. In that case, the tangent linear model may depart from the real model trajectory and important deviations with regard to the observations

are found. Moreover the innovation term is also larger, which indicates a decoupling between w_g and w_2 , leading sometimes to inaccurate corrections.

d. Sensitivity to different levels of prescribed errors

The performance of an assimilation method is very dependent on accurate prescription of the error statistics. Over the SMOSREX experimental site, continuous observations of both w_g and w_2 , are available. We have taken advantage of all this source of information to define errors, coherent with the terrain observations. A sensitivity study permits to assess, how the system behaves with regard to background and observation errors. Figure 6 shows the contour lines of regions of the same performance for

different values of the observation and forecasted/background errors for the simplified 1D-VAR. A maximum of efficiency is obtained for an observation error of $0.07 \text{ m}^3 \text{ m}^{-3}$ and a forecasted error equal to the estimated model error Q^{var} with a skill score close to 0.90 and a RMSE of $0.02 \text{ m}^3 \text{ m}^{-3}$. Even though this sensitivity study is specific to this experimental site, it is important to note that around this peak of efficiency a broad region is found where the skill score is higher than 0.8 and the RMSE is lower than $0.03 \text{ m}^3 \text{ m}^{-3}$, confirming thus the skill of a method which uses several observations (for each analysis) to generate correct gains. However at the boundaries, i.e. very small observation errors or large forecasted state errors, a much larger RMSE and a sharp drop in performance is found.

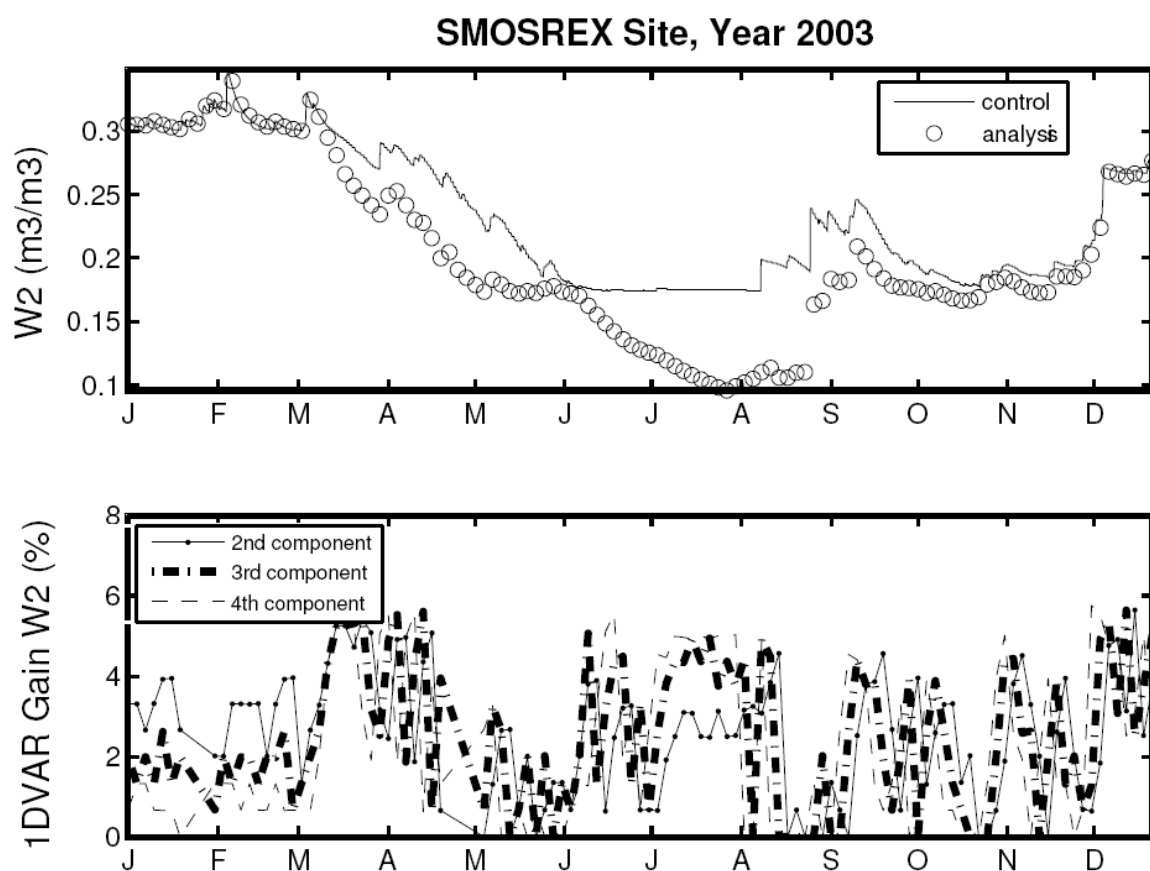


Figure 5.- Gain components for the simplified 1D-VAR method compared with the yearly evolution of w_2 in 2003. Top: the control simulation (solid line) is superimposed on the analyses (circles). Bottom: the evolution of the gain components (multiplied by 100) at each analysis step.

Monte-Carlo based methods, like the EnKF, could partially correct and counteract large deviations from the true state by using an ensemble of model trajectories.

Furthermore, within the range of the expected SMOS observation errors, of about $0.04 \text{ m}^3 \text{ m}^{-3}$, a good performance (around or higher than 0.8) of the system is observed for a wide range of background model error. This shows the potential use of the SMOS data to obtain a spatialized information on the root zone soil moisture.

e. Processing time

To finish this analysis the importance of computational time of an assimilation algorithm in an operational system has to be emphasized, in particular when run over

large areas. In Table 5, the total processing times for a whole year and for the four assimilation algorithms are compared. The runs were performed on the same platform, an Intel Pentium IV processor with a 2.40 GHz CPU. The simplified 1D-VAR appears to be a good trade-off between computing time and the quality of the results. It can be seen (Table 5) that by using an ensemble of around 10 members the EnKF and the simplified 1D-VAR are comparable in terms of computing time. Nevertheless, the statistics of the ensemble would be of lesser quality and, consequently, the quality of the retrievals of w_2 . In order to get closer to a performance similar to the 1D-VAR, an ensemble of around 200 members is necessary for the EnKF, thus increasing the EnKF processing time.

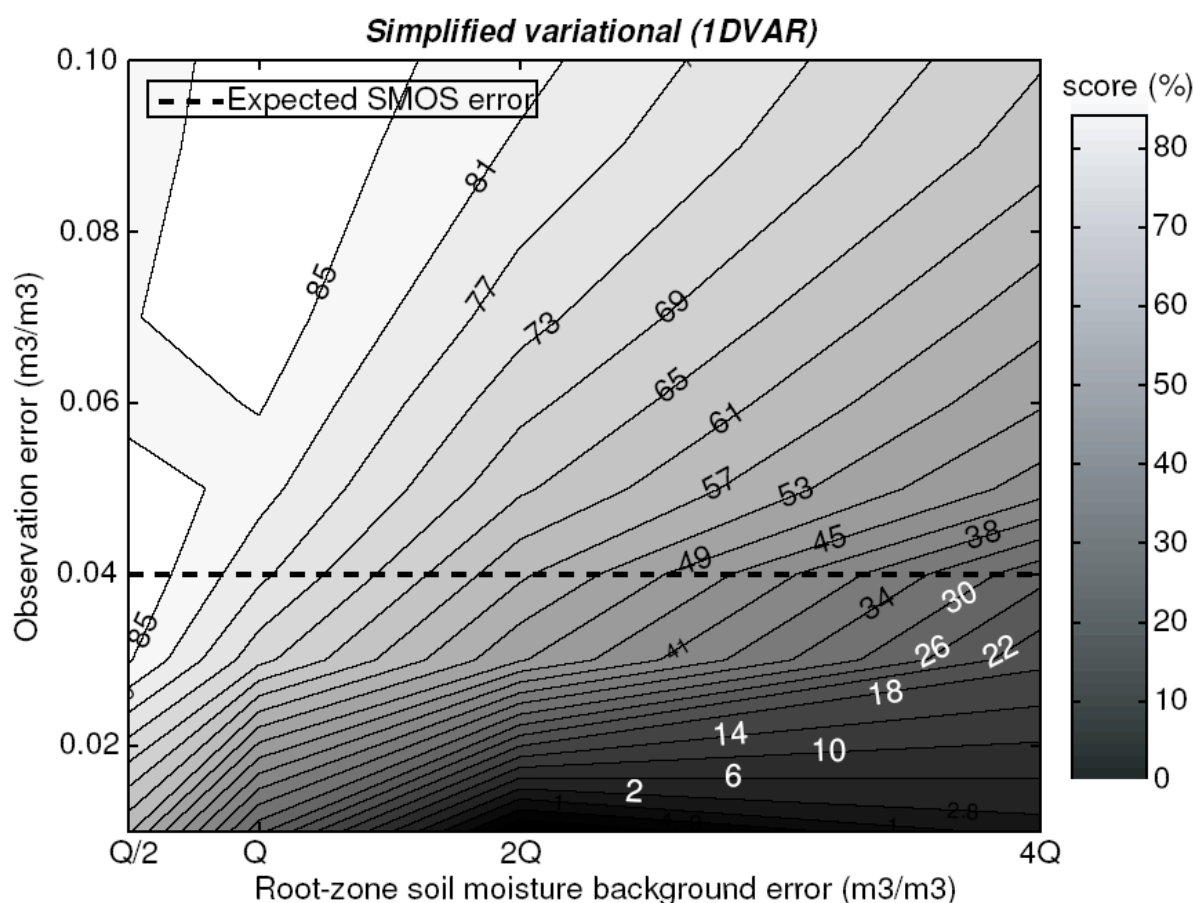


Figure 6.- Skill score for different error levels in observations (R) and forecasted state variables (P) for the simplified 1D-VAR. Note that the Q values in the abscises axis are those in Eq. [14]. The dashed-line superimposed on the figures corresponds to the performance with an observation error corresponding to the SMOS satellite specification.

TABLE 5.

Computer processing time for a whole year (in seconds) for the Ensemble Kalman Filter (EnKF), Extended Kalman Filter (EKF), Simplified 1D-VAR and a tuning variational (T-VAR). The computing time for the EnKF is shown for ensembles of $N=10, 20, 50, 100$ and 200 members.

EnKF					EKF	1D-VAR	T-VAR
10	20	50	100	200			
81	146	353	738	1358	36	76	300

4. Summary and conclusions

Four assimilation methods were implemented over an experimental site in southwestern France (SMOSREX) and the analysed soil moisture results were compared. Three methods were based on least squares principles (EKF, EnKF and 1DVAR) and one was a simple tuning method (T-VAR). The assimilation approaches and their practical implementation in the ISBA-A-gs land surface model were described and discussed. The multi-year SMOSREX data set (2001-2004) allowed assessing the performance of the assimilation methods in contrasting conditions. In particular, marked droughts were observed during the summers of 2003 and 2004, for which the observed root-zone soil moisture was lower than the wilting point of the control simulation of ISBA-A-gs. The difficulty of the model to adequately reproduce the drought in 2003 and 2004 offered a good test for the assimilation schemes. In general, the four methods provided satisfactory results. The best performance was shown by the 1D-VAR, with a skill score of 0.86, improving the control simulations (skill score of 0.73). Finally, a sensitivity study of the 1D-VAR performance to different levels of background and observation errors was conducted.

The analysis results over SMOSREX show that:

- The w_2 analyses were improved by using a background/a priori information: the 1D-VAR outperformed the T-VAR, which did not use any background term.
- The EnKF, which propagates the covariance through the sampling of several model trajectories, was more efficient than the EKF which relies on the tangent linearization of the model equations to propagate the covariance information.
- The EnKF is a promising technique to deal with high non linear systems but, over the SMOSREX site, it was outperformed by the 1D-VAR. This result was attributed to: (1) the limited non linearity of the system which could have prevented the expression of the added value of the EnKF; (2) the difficulty in “tuning” the algorithmic parameters of the EnKF, such as the inflation factor which, in this study, was required to prevent the ensemble from collapsing.

It is also important to remark that, in this study, the limited performance of the EKF (despite its apparently good behaviour) could be related to the functioning of the ISBA-A-gs model (and models relying on

the force restore scheme) rather than to the assimilation method itself. Indeed, the force restore approach presents a low sensitivity of the KF state variables to a perturbation of the surface soil moisture for the days following the date of the perturbation.

A sensitivity study showed that the 1D-VAR method leads to good performance for a large range of background and observation errors. Moreover, using $0.04 \text{ m}^3 \text{ m}^{-3}$ as a prescribed error in the w_g observations, i.e. the error level which is expected from the SMOS satellite (shown as a dashed-line in Figure 6), good results are also obtained provided that variances in the forecasted state variables do not exceed the estimated model error.

Finally, with the lower computing time, the 1D-VAR is a good alternative to the EnKF for the development of an operational data assimilation system aiming to analyse root zone soil moisture from surface soil moisture

observations. Nevertheless, this promising method needs to be tested at other experimental sites representing different geoclimatic environments and further research is needed before the implementation of a full 2D application, in particular concerning the spatial correlation of background errors (Reichle and Koster 2003).

Acknowledgment

This study was co-funded by the European Commission within the GMES initiative in FP6, in the framework of the geoland integrated GMES project on land cover and vegetation. The authors would like to thank Dr. Sébastien Masart (CERFACS, Toulouse), Dr. Christoph Rüdiger (CNRM, Toulouse), Dr. Gianpaolo Balsamo (ECMWF, Reading) for fruitful discussions, as well as the anonymous reviewers, for their helpful comments

References

- Anderson, J.L., and S.L. Anderson, 1999: A Monte Carlo Implementation of the Non linear Filtering problem to produce ensemble assimilations and forecast, *Mon. Wea. Rev.*, **127**, 2741-2758.
- Balsamo, G., F. Bouyssel, and J. Noilhan, 2004: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Q. J. R. Met. Soc.*, **130A**, 895-915.
- , J.-F. Mahfouf, S. Belair, and G. Deblonde, 2006: A global root-zone soil moisture analysis in NWP using simulated L-band brightness temperature in preparation for the Hydros satellite mission. *J. Hydrometeor.* In press.
- Beljaars, A.C.M., P. Viterbo, M.J. Miller, and A.K. Betts, 1996: The anomalous rainfall over the USA during 1993: Sensitivity to land surface parameterization and soil moisture anomalies. *Mon. Wea. Rev.*, **124**, 362-383.
- Bouttier, F., 1994: Sur la prévision de la qualité des prévisions météorologiques. Ph.D dissertation. Univ. Paul Sabatier III, 240 pp. [Available from Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France]
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.
- Calvet, J.-C., J. Noilhan, and P. Bessemoulin, 1998a: Retrieving the Root-Zone Soil Moisture from Surface Soil Moisture or Temperature Estimates: A Feasibility Study Based on Field Measurements. *J. Appl. Meteor.*, **37**, 371-386.
- , J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A.

- Oliosio, J.-P. Wigneron, 1998b: An interactive vegetation SVAT model tested against data from six contrasting sites. *Agric. For. Meteor.*, **92**, 73-95.
- , and coauthors, 1999: MUREX: A land-surface field experiment to study the annual cycle of the energy and water budgets. *Ann. Geophys.*, **17**, 838-854.
- , and J. Noilhan, 2000: From near-surface to root-zone soil moisture using year-round data. *J. Hydrometeor.*, **1**, 393-411.
- , and J.F. Soussana, 2001: Modelling CO₂ enrichment effects using an interactive vegetation SVAT scheme. *Agric. For. Meteor.*, **108**, 129-152.
- De Rosnay, P., and Coauthors, 2006: SMOSREX: A long term field campaign experiment for soil moisture and land surface processes remote sensing. *Remote Sens. Env.*, **102**, 377-389.
- Deardorff, J. W., 1977: A parameterization of ground-surface moisture content for use in atmospheric prediction models. *J. Appl. Meteor.*, **16**, 1182-1185.
- , 1978: Efficient prediction of ground temperature and moisture with inclusion of a layer of vegetation. *J. Geophys. Res.*, **83**, 1889-1903.
- Dirmeyer, P.A., 2000: Using a global soil wetness dataset to improve seasonal climate simulation. *J. Climate*, **13**, 2900-2922.
- Douville, H., P. Viterbo, J.-F. Mahfouf, and A. C. M. Beljaars, 2000: Evaluation of optimal interpolation and nudging techniques for soil moisture analysis using FIFE data. *Mon. Wea. Rev.*, **128**, 1733-1756.
- Eagleman, J.R., and W.C. Lin, 1976: Remote sensing of soil moisture by a 21 cm passive radiometer, *J. Geophys. Res.*, **81**, 3660-3666.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.
- , 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343-367.
- François, C., A. Quesney, and C. Ottlé, 2003 : Sequential Assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an Extended Kalman Filter. *J. Hydrometeor.*, **4**, 473-487.
- Giard, D. and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997-1015.
- Gibelin, A.-L., J.-C. Calvet, J.-L. Roujean, L. Jarlan, S. Los, 2006: Validation of global leaf area index simulated by the land surface model ISBA-A-gs through comparison with satellites products. *J. Geophys. Res.*, in press.
- Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich, 1998: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resour. Res.*, **34**, 3405-3420.
- Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *Trans. of the ASME-Journal of Basic Engineering*, **82**, 35-45.
- Kerr, Y.K., P. Waldteufel, J.-P. Wigneron, J.M. Martinuzzi, J. Font, and M. Berger, 2001: Soil moisture retrieval from space : The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sensing*, **39**, 1729-1735.
- Koster, R.D., and M.J. Suarez, 2003: Impact of Land Surface initialization on seasonal precipitation and temperature prediction. *J. Hydrometeor.*, **4**, 408-423.
- Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167-194.
- Njoku, E.G., T.J. Jackson, V. Lakshmi, T.K. Chan, and S.V. Nghiem, 2003: Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosc. Remote Sens.*, **41**, 215 – 229.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface

processes for meteorological models.
Mon. Wea. Rev., **117**, 536-549.

- , J., and J.-F., Mahfouf, 1996: The ISBA land surface parameterisation scheme. *Global Planet. Changes*, **13**, 145-159.
- Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser, 2002: Extended versus ensemble Kalman filtering for land data assimilation. *J. Hydrometeor.*, **3**, 728-740.
- Reichle, R. H., and R. D. Koster, 2003: Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. *J. Hydrometeor.*, **4**, 1229-1242.
- Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet, 2003: Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data. *Geophys. Res. Lett.*, **30**, 1498, doi:10.1029/2003GL017128.
- Shukla, J., and Y. Mintz, 1982: Influence of Land-Surface Evapotranspiration on the Earth's Climate. *Science*, **215**, 1498-1501.
- Teunissen, P.J.G., 2000: *Adjustment theory: an introduction*. Delft University Press, 194 pp.
- Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf, 2003: Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *J. Geophys. Res.*, **108**, 4611-4626.
- Wigneron J.-P., A. Chanzy, J.-C. Calvet, A. Olioso, Y. Kerr, 2002: Modeling approaches to assimilating L-band passive microwave observations over land surfaces. *J. Geophys. Res.*, **107**, D14, (DOI10.1029/2001JD000958).

Pour compléter les résultats montrés dans l'article précédent, le tableau 4.6 présente le score des analyses de w_2 , la RMSE, le biais moyen, le pourcentage des incréments non-nuls effectués sur les estimations du modèle et le temps de calcul, des quatres méthodes d'assimilation. Au vu des résultats de ce tableau on identifie clairement les faiblesses des différentes méthodes d'assimilation par rapport aux analyses de w_2 . En particulier, l'EKF n'ajuste l'humidité du sol simulé par ISBA-A-g_s que seulement sur un 23 % de cas, contrairement à plus de 90 % pour les méthodes EnKF et le 1D-VAR. La bonne performance de l'EnKF est associée à un ensemble constitué d'un grand nombre de membres, ce qui en contrepartie pénalise le coût numérique de la méthode. On voit directement le bon compromis obtenu par la méthode 1D-VAR entre performance des analyses et le temps de calcul.

TAB. 4.6: Résultats de la comparaison entre les quatres méthodes d'assimilation et le schéma de surface ISBA-A-g_s pour le score des analyses de w_2 (E), le RMSE, le pourcentage des incréments non-nuls et le coût numérique

	E	RMSE ($m^3 \cdot m^{-3}$)	$\Delta \neq 0$ (%)	temps calcul (s/an)
EKF	0.85	0.03	23.2	36
EnKF	0.81	0.03	93.5	1358
1DVAR	0.86	0.02	97.6	76
T-VAR	0.53	0.05	-	300
ISBA-A-g _s	0.70	0.03	-	20

4.3 Une méthode dérivée de l'EnKF

L'un des principaux problèmes rencontrés pour l'implémentation de l'EnKF a été l'effondrement rapide de l'ensemble. En effet, l'évolution temporelle de l'humidité du sol dans ISBA est pilotée par les équations de la méthode " force-restore " (Deardorff, 1977, Deardorff, 1978). En l'absence de précipitations la valeur de l'humidité superficielle du sol tend rapidement vers une valeur d'équilibre voisine de w_2 . Dans ces conditions, même si on perturbe la valeur initiale de l'humidité (membres de l'ensemble), le modèle tend à faire converger tout les membres vers la même valeur d'équilibre, ce qui réduit la dispersion de l'ensemble avec le temps. C'est ce que l'on voit à la fig. 4.7a. Ici, après l'étape qui corrige l'ébauche du modèle et la matrice de covariance associée, chaque membre du nouvel ensemble est obtenu en ajoutant une erreur à la valeur analysée (valeur moyenne de l'ensemble). Cette erreur est estimée en employant une distribution gaussienne avec une variance égale à celle de l'ensemble mis à jour (Evensen, 2003).

En suivant strictement cette approche, l'ensemble de 100 membres de notre système s'effondre après quelque dizaines de jours. Les points rouges de la fig.4.7b montrent comment l'écart type de la distribution diminue rapidement avec le temps. La solution proposée dans l'article de la section précédente est de multiplier la variance mise à jour de l'ensemble par un facteur d'inflation estimé égal à 1.35 (paragraphe "Implementation ; EnKF" de l'article à la section 4.2), ce qui évite la diminution rapide de la dispersion de l'ensemble (voir points noirs de la fig.4.7b).

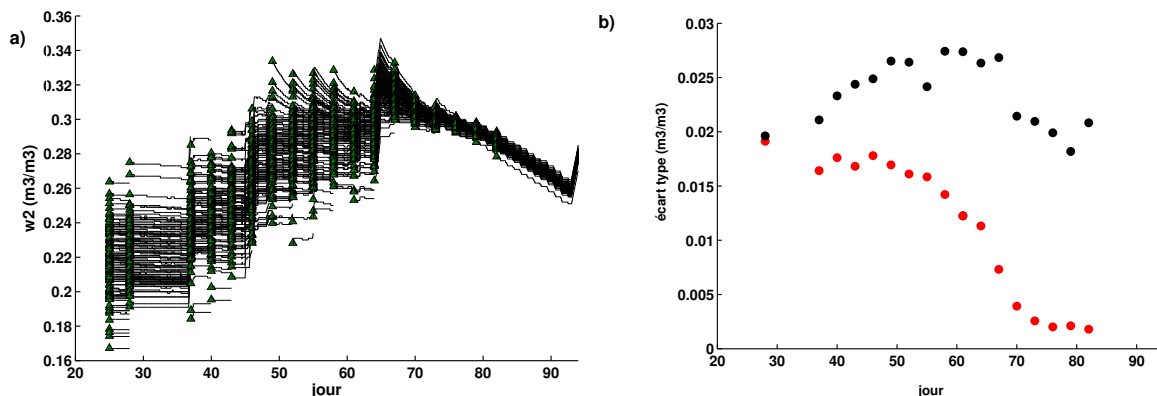


FIG. 4.7: a) Evolution temporelle et effondrement d'un ensemble de 100 membres avec la méthode EnKF "pure".

A chaque observation, les membres de l'ensemble corrigé sont dessinés en triangles verts. La trajectoire de chaque ensemble entre deux observations est représentée par une ligne continue noire. b) Ecart type de l'ensemble avec la méthode EnKF "pure" (points rouges) et avec l'EnKF avec un facteur d'inflation (points noirs).

Un autre test a été conduit qui n'est pas présenté dans l'article. Ce test porte sur la façon dont le nouvel ensemble de l'EnKF est créé : au lieu de multiplier la variance de l'ensemble par un facteur d'inflation, on a considéré que cette variance est fixe et égale à une valeur prédéfinie. Cela nous ramène à une méthode d'interpolation optimale (OI), à la différence près que, pour l'OI, une seule réalisation du modèle est conduite, tandis qu'ici on en fait une pour chaque membre de l'ensemble. Le désavantage de cette variante de l'EnKF est que l'on perd l'information provenant de la propagation de la matrice de covariance d'état, qui utilise un ensemble de trajectoires sans faire d'hypothèse de linéarité. Pour que cette méthode fonctionne correctement, on s'assure que la matrice de covariance $\mathbf{P}$ (constante au début de chaque pas d'assimilation) est construite à partir d'un ensemble suffisamment dispersé. La fig. 4.8 compare les analyses de w_2 avec les observations et la simulation de contrôle en suivant cette approche. Les périodes sèches de 2003 et 2004 sont bien corrigées et, en général, les caractéristiques des analyses montrent des résultats proches de ceux du 1D-VAR pour un ensemble de 100 éléments (tableau 4.7). Le fait de créer des ensembles ayant la même dispersion à chaque pas d'assimilation conduit à un plus grand contrôle sur la taille des corrections, surtout durant des périodes de fortes précipitations ou

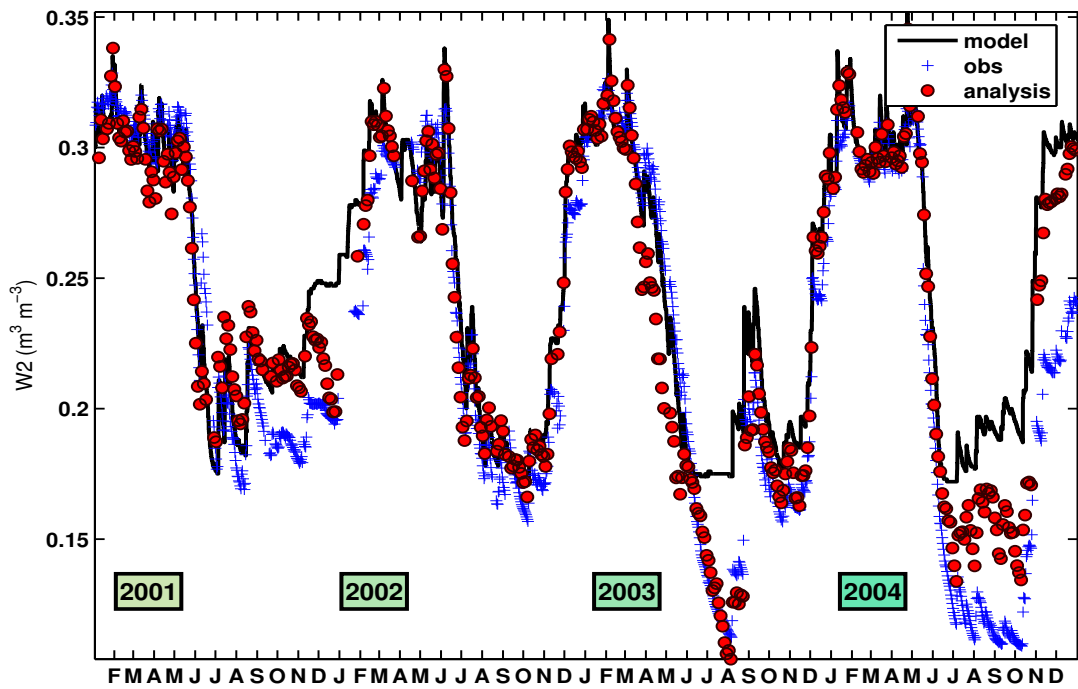


FIG. 4.8: Analyses de w_2 comparées aux observations et à la simulation de contrôle, avec un ensemble de 100 membres.

d'évaporation intense où l'ensemble se disperse rapidement. En conséquence, les analyses sont plus lissées et restent plus stables que pour l'EnKF présenté dans le papier. Cela peut donner l'impression d'être un avantage, mais ne le serait peut-être pas pour des modèles moins dissipatifs pour les variables pronostiques. Avec une méthode "pure" d'EnKF on attend que l'incertitude de l'état du système soit réduite entre deux observations. C'est un grand avantage de l'EnKF. Par contre, pour obtenir de bonnes analyses avec la méthode décrite ici, l'incertitude de l'état du système doit varier très peu (en analogie avec la covariance fixe du 1D-VAR) pour (1) éviter un effondrement de l'ensemble et (2) contrôler la taille de la correction au travers de la matrice de covariance $\mathbf{P}$.

TAB. 4.7: RMSE, biais moyen et critère de Nash annuel, de 2001 à 2004, et pour l'ensemble des quatre années, entre les observations de w_2 et les analyses.

	2001			2002			2003			2004			2001-2004		
	RMSE	biais	E	RMSE	biais	E	RMSE	biais	E	RMSE	biais	E	RMSE	biais	E
N=10	0.02	0.010	0.82	0.02	0.012	0.84	0.03	-0.012	0.82	0.04	0.027	0.76	0.03	0.084	0.80
N=20	0.02	0.006	0.87	0.02	0.010	0.87	0.03	-0.007	0.83	0.03	0.026	0.80	0.03	0.012	0.81
N=50	0.02	0.005	0.87	0.02	0.010	0.88	0.02	-0.008	0.88	0.03	0.026	0.81	0.03	0.010	0.85
N=100	0.02	0.006	0.88	0.01	0.007	0.92	0.02	-0.007	0.88	0.03	0.026	0.83	0.02	0.008	0.87
N=200	0.02	0.001	0.88	0.02	0.010	0.91	0.02	-0.005	0.89	0.03	0.027	0.82	0.02	0.010	0.87
1D-VAR	0.02	-0.011	0.90	0.01	0.001	0.93	0.02	-0.009	0.89	0.04	0.029	0.79	0.02	0.002	0.86

Chapitre 5

Assimilation des observations de w_g et LAI dans ISBA-A-g_s en mode interactif

Sommaire

5.1	Introduction	90
5.2	Assimilation des w_g avec LAI interactif	90
5.2.1	Stratégie défensive	90
5.2.2	Stratégie offensive	93
5.2.3	Fenêtre séquentielle vs fenêtre glissant pour le 1D-VAR simplifié	94
5.2.4	Conclusion	95
5.3	Assimilation d'observations de LAI	96
5.4	"Joint assimilation of surface soil moisture and LAI observations using a simplified 1D-VAR : The SMOSREX case study."	98
5.4.1	Résumé	98
5.4.2	Article	99
5.5	Assimilation d'observations des instruments de télédétection	119
5.5.1	Températures de brillance	119
5.5.2	Réfectances	125
5.5.3	Erreur des pseudo-observations	126
5.5.4	assimilation des produits de télédétection	128
5.5.5	Conclusion	129

5.1 Introduction

Au chapitre 4 il a été montré, pour l'analyse de w_2 , qu'une méthode variationnelle simplifiée, 1D-VAR, présente des avantages en terme de performance et de coût numérique en comparaison à d'autres méthodes, dont l'EnKF et l'EKF. Cette méthode 1D-VAR est utilisée dans ce chapitre et son domaine d'application est étendu à l'analyse de la biomasse de la végétation.

Ce chapitre est divisé en quatre parties :

- Dans la section 5.2 on analyse w_2 , mais en utilisant la configuration de végétation interactive du schéma de surface (c'est à dire que le LAI et la biomasse sont simulés par le modèle au lieu d'utiliser une valeur prescrite de LAI dérivée des observations sur le site). Les deux stratégies possibles de réponse de la plante au stress hydrique sont testées dans cette section ;
- Dans la section 5.3 on n'assimile que les observations de LAI et on étudie l'effet que la ré-initialisation de la biomasse de la végétation a sur les simulations de w_2 ;
- Ces deux approches convergent à la section 5.4, où les observations de w_g et de LAI sont assimilées simultanément. Un article soumis au journal " Agricultural and Forest Meteorology " est présenté ;
- Dans la section 5.5 on présente une première approche de l'assimilation des observations fournies par les instruments de télédétection présents sur le site de SMOSREX, dont le but est de simuler les observations de capteurs satellitaires.

5.2 Assimilation des w_g avec LAI interactif

Cette section a pour but de tester l'analyse de w_2 mais, à différence du chapitre précédent, les variables pronostiques de la végétation sont simulées par ISBA-A- g_s . Les tests menés dans cette section vont nous permettre de quantifier l'impact sur les analyses de w_2 d'un LAI calculé par le modèle. En même temps, étant donné la disponibilité d'un jeu d'observations pour la validation du LAI, on évaluera le LAI simulé après ré-initialisation de l'humidité du sol. Les méthodes 1D-VAR simplifié et EnKF avec facteur d'inflation sont appliqués aux observations de w_g dans ISBA-A- g_s avec les deux stratégies possibles de réponse de la plante au stress hydrique : tolérance à la sécheresse (offensive) et évitement de la sécheresse (défensive).

5.2.1 Stratégie défensive

La fig. 5.1 montre les analyses de w_2 et le LAI estimé à partir des nouveaux états initiaux de l'humidité du sol, pour le 1D-VAR simplifié. La partie supérieure de cette figure compare, de la

même façon que les figs. 4.3 et 4.4, les analyses de w_2 (cercles), avec la simulation de référence (ligne continue) et les observations (points bleus). La partie inférieure de la figure montre le LAI simulé par ISBA-A-gs à chaque pas d'assimilation (cercles), comparé au LAI interpolé à partir des observations disponibles sur le site expérimental (points bleus) ainsi qu'à la solution du modèle sans assimilation (trait noir). Des différences importantes de LAI apparaissent. Pour les années 2001 et 2002, le LAI après ré-initialisation de l'humidité du sol est largement surestimé pendant une bonne partie de l'année en comparaison aux observations. Ceci est une conséquence de l'utilisation de la stratégie défensive, pour laquelle la plante accentue son efficacité d'utilisation de l'eau. La dynamique et le niveau de LAI obtenus ici sont cohérents avec la simulation de référence. Cependant, ici on trouve des valeurs maximales de LAI plus élevées. La raison est qu'à certaines périodes, les corrections effectuées sur w_2 ajoutent de l'eau par rapport à la simulation de référence, et en conséquence plus d'eau est disponible pour la croissance de la végétation. Les nouvelles estimations du LAI ont aussi un impact sur l'évolution de w_2 . En effet, la zone racinaire s'assèche d'avantage quand on force les simulations avec des valeurs de LAI plus fortes que les observations. On observe faiblement cet effet à l'automne 2001 et plus clairement au printemps de l'année 2002, quand on le compare avec la courbe de w_2 de référence. Au contraire, la période de sécheresse de l'année de 2003 est marquée par une correction plus faible, probablement à cause d'une extraction racinaire plus basse (le LAI est en général, durant cette période, inférieur aux observations). On observe aussi une mortalité de la végétation très forte aux périodes où w_2 est au dessous du point de flétrissement permanent (" wilting point "). Ce problème est discuté dans la section 5.4.

Les résultats pour l'EnKF avec facteur d'inflation montrent aussi que le LAI après ré-initialisation de w_2 n'est pas bien reproduit, avec un décalage du début de la sénescence de la végétation pour 2001 et des surestimations pour 2001, 2002 et 2004. Par contre, le plus remarquable durant toute la période d'étude sont les chutes soudaines des valeurs du LAI de 2002 à 2004. Pour ces trois années, à cause du caractère stochastique de la méthode, des états initiaux de w_2 apparaissent fréquemment au dessous du wilting point. A ce moment là, le modèle supprime le flux d'évapotranspiration et la croissance de la végétation est stoppée. Cet aspect est traité dans la section 5.4.

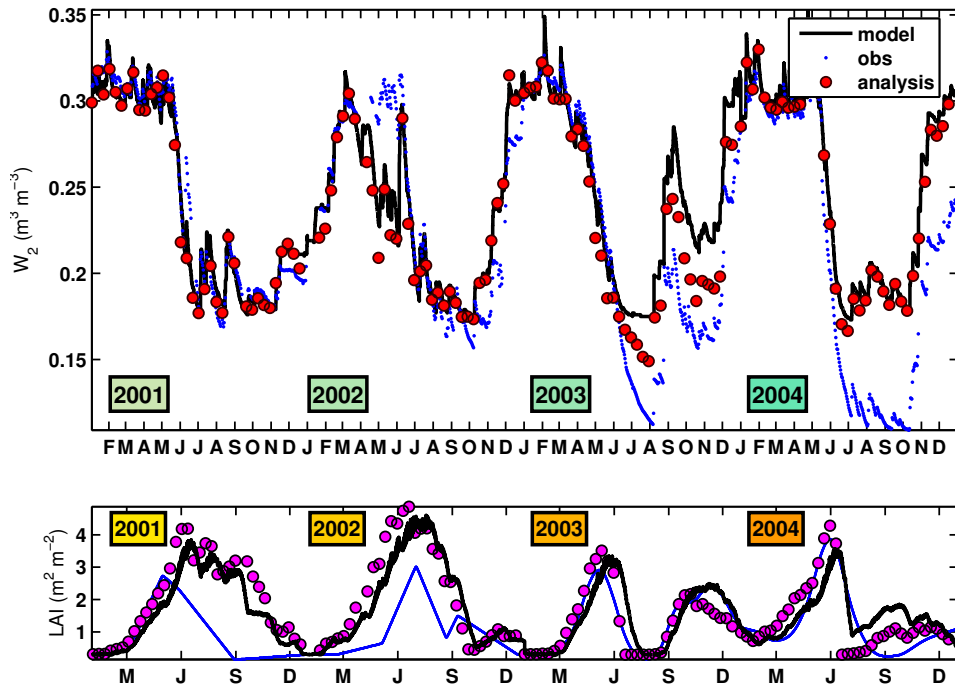


FIG. 5.1: Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode 1D-VAR simplifiée. En haut, les analyses de w_2 (cercles rouges) sont comparées aux observations (points bleus) et à la simulation de contrôle (trait continu noir). En bas, le LAI estimé par le modèle après assimilation (points violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).

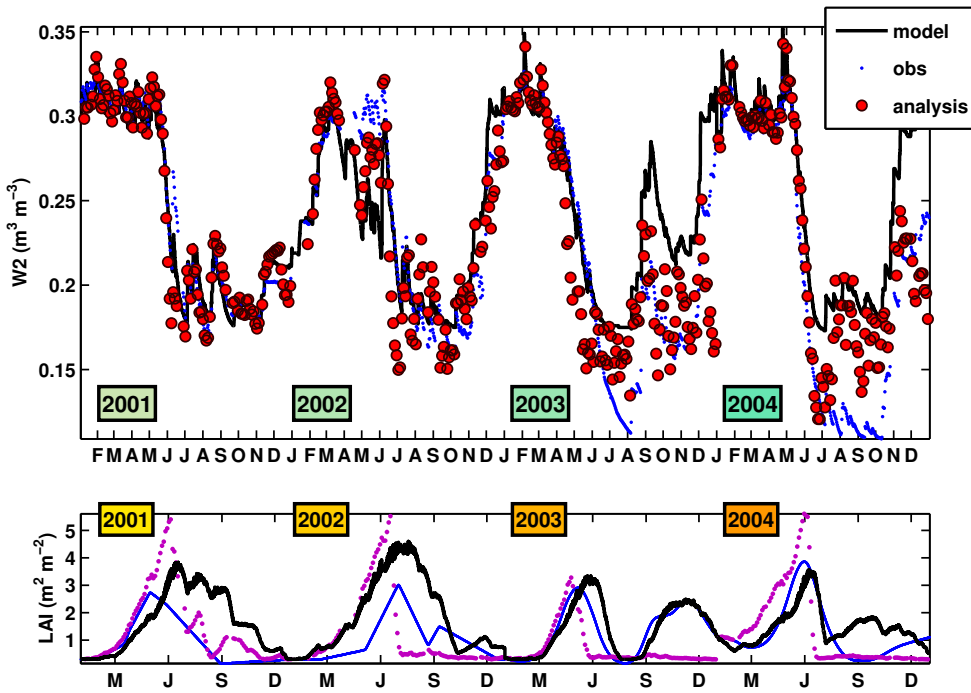


FIG. 5.2: Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode EnKF. En haut, les analyses de w_2 (cercles rouges) sont comparées aux observations (points bleus) et à la simulation de contrôle (trait continu noir). En bas, le LAI estimé par le modèle après assimilation (points violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).

5.2.2 Stratégie offensive

Quand en conditions de stress hydrique la végétation est modélisée en suivant une stratégie offensive, la principale différence que l'on observe pour le 1D-VAR (fig. 5.3) et pour l'EnKF (fig. 5.4) par rapport aux figs. 5.1 et 5.2 est le niveau plus faible de LAI estimé. Cependant, les désaccords avec les observations de LAI sont encore plus grands. Pour ce qui est des valeurs analysées de w_2 on n'obtient pas de différence significative en utilisant l'une ou l'autre des deux stratégies.

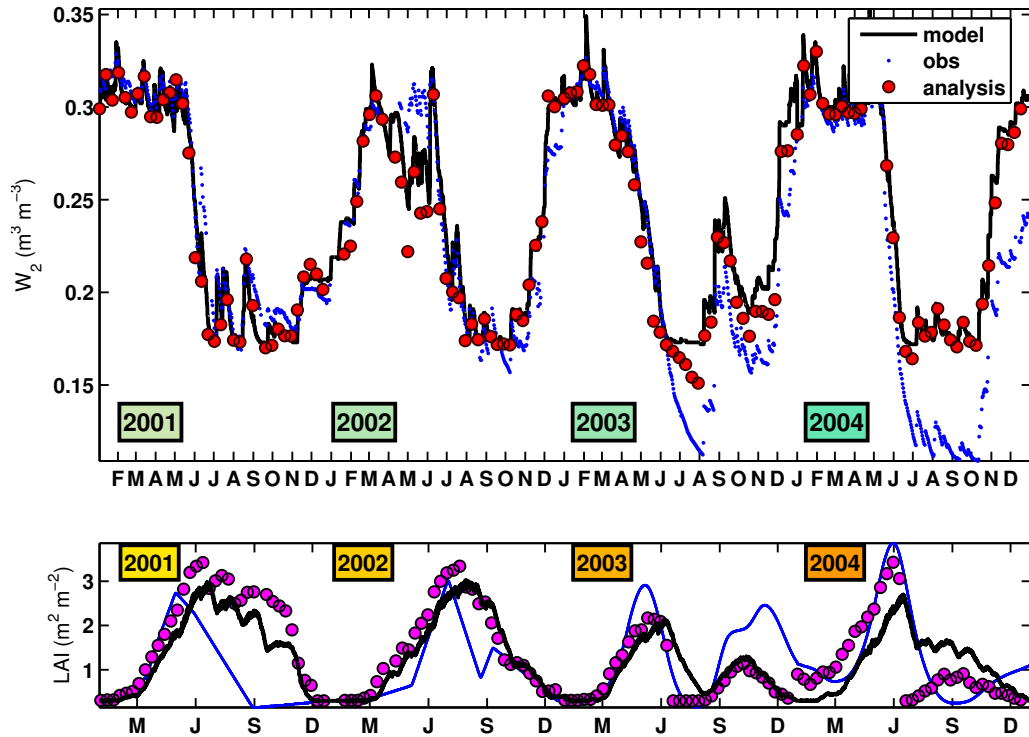


FIG. 5.3: Analyses de w_2 et LAI dans ISBA-A-gs en stratégie offensive avec la méthode 1D-VAR simplifiée. En haut, les analyses de w_2 (cercles rouges) sont comparés aux observations (points bleus) et à la simulation de contrôle (trait noir). En bas, le LAI estimé par le modèle après assimilation (point violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).

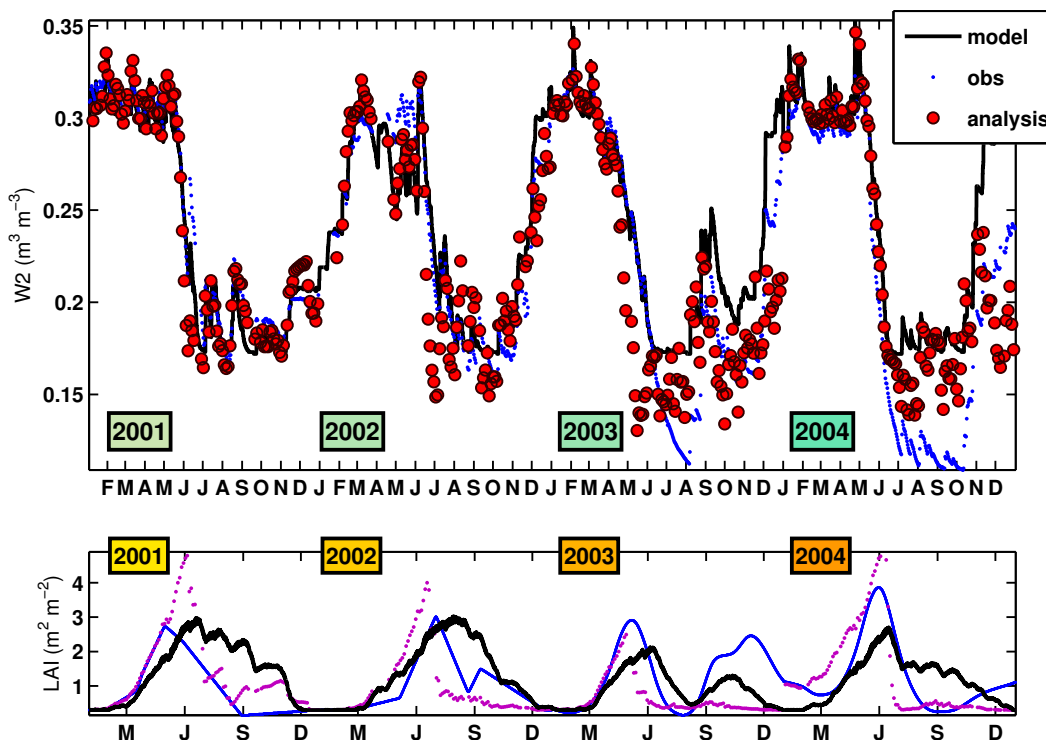


FIG. 5.4: Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode EnKF. En haut, les analyses de w_2 (cercles rouges) sont comparées aux observations (points bleus) et à la simulation de contrôle (trait noir). En bas, le LAI estimé par le modèle après assimilation est montré (point violets) et comparé à la simulation de référence (trait noir) et aux observations (points bleus).

5.2.3 Fenêtre séquentielle vs fenêtre glissant pour le 1D-VAR simplifié

Dans cette section, les résultats montrés pour le 1D-VAR sont obtenus tous les 10 jours en utilisant une fenêtre d'assimilation séquentielle de 10 jours (correspondant à la périodicité typique des estimations de LAI), de façon à ce que les observations ne soient assimilées qu'une seule fois (pas de chevauchement des fenêtres). Dans le chapitre précédent, au contraire, les analyses avec le 1D-VAR simplifié utilisent une fenêtre d'assimilation glissante, et les observations sont assimilées plusieurs fois. En effet, il s'agissait d'un exercice d'intercomparaison et on a essayé d'être le plus cohérent possible avec les méthodes séquentielles. De plus, le LAI était forcé, et le choix du type de fenêtre a beaucoup moins d'influence que dans le cas où les variables de la végétation sont calculées par le modèle. Pour illustrer l'effet du choix de l'une ou l'autre configuration dans le cas où le LAI est calculé par le modèle, la fig. 5.5 montre, à titre d'exemple, les simulations de w_2 et de LAI pour les 50 premiers jours de 2002, période hivernale toutefois marquée par un stress hydrique : au LAI simulé par le modèle après ré-initialisation de w_2 on a superposé le LAI correspondant quand une fenêtre d'assimilation glissante est utilisée. On

observe que, dans ce dernier cas, pendant les premiers 35 jours on ajoute davantage d'eau et plus fréquemment. Cela provoque une augmentation importante de la croissance de la végétation, qui va atteindre un maximum de $7.7 \text{ m}^2 \cdot \text{m}^{-2}$, à comparer à $5.1 \text{ m}^2 \cdot \text{m}^{-2}$ dans le cas d'une fenêtre séquentielle de 10 jours.

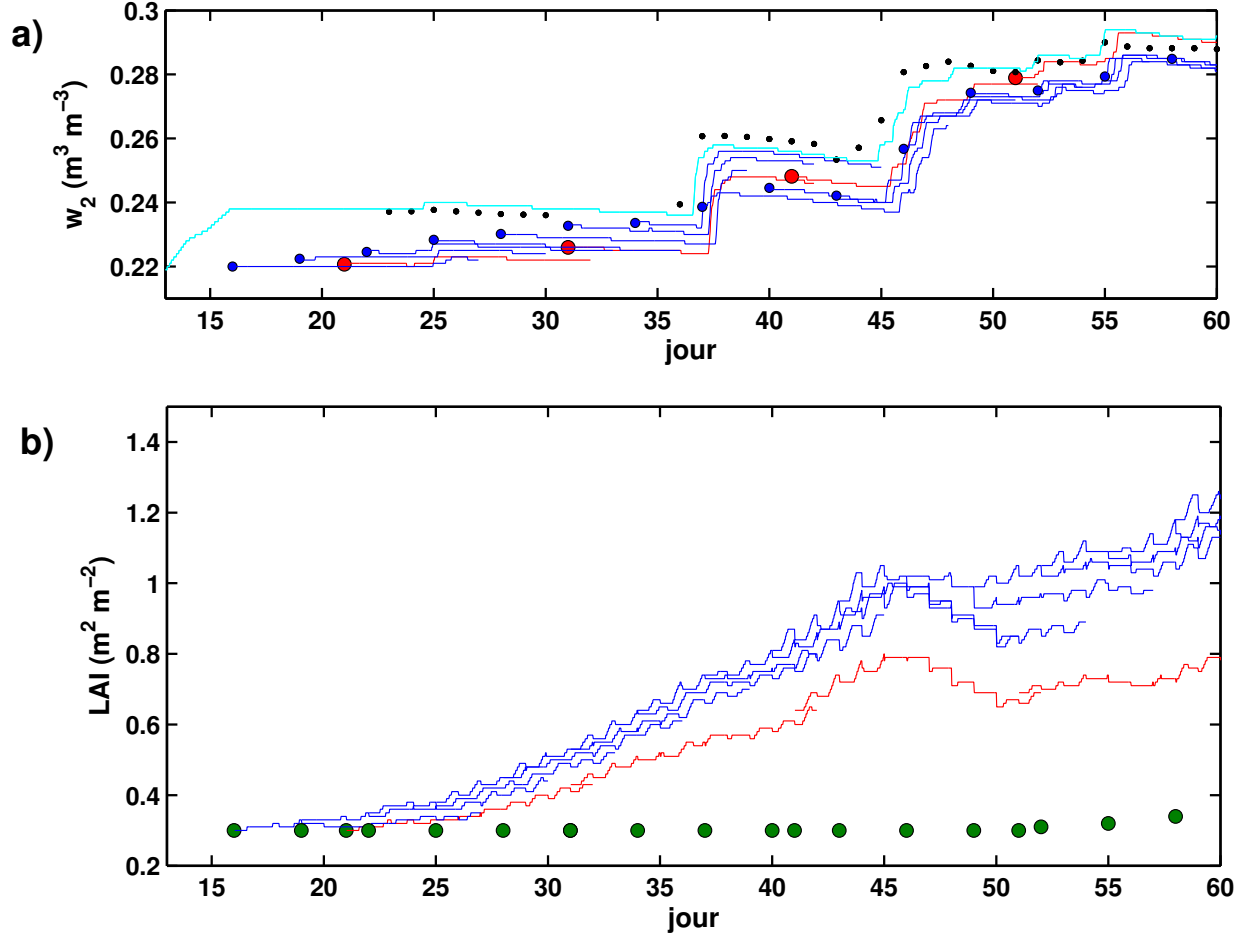


FIG. 5.5: a) Analyses et évolution de w_2 avec le 1D-VAR en configuration de fenêtre glissante (points et lignes bleus foncés, respectivement) et en configuration de fenêtre séquentielle (points et lignes rouges). Les observations de w_2 et la simulation de contrôle d'ISBA-A-gs sont représentées par des points noirs et par une trait de couleur bleu ciel, respectivement. b) Les observations de LAI (points verts) sont superposées à l'évolution du LAI après réinitialisation de w_2 avec le 1D-VAR fenêtre glissante (ligne bleu) et fenêtre séquentielle (ligne rouge).

5.2.4 Conclusion

L'assimilation d'observations de w_g pour l'analyse de w_2 a été testée quand les variables de végétation sont calculées par ISBA-A-gs pour les deux stratégies de réponse de la plante au stress hydrique. Il a été observé que (1) la dynamique annuelle du LAI avec les deux stratégies

est la même que celle obtenue pour le modèle sans assimilation, mais avec des niveaux maximum différents à cause des ré-initialisations de w_2 , (2) l'assimilation des seules données de w_g permet d'analyser w_2 mais n'améliore pas le LAI, (3) l'utilisation d'une fenêtre glissante pour le 1D-VAR produit une forte surestimation du LAI.

L'assimilation d'observations sensibles à la biomasse de la végétation, en complément de w_g , est nécessaire pour améliorer les deux variables.

5.3 Assimilation d'observations de LAI

L'objectif de cette section est d'étudier si l'assimilation des seules observations de LAI permet de suivre les variables de la végétation observées sur SMOSREX, ainsi que l'humidité du sol. En effet, ces deux variables sont physiquement liées, et on peut penser qu'une information sur la dynamique de la biomasse peut aider à suivre l'état hydrique du sol. La méthode testée est le 1D-VAR avec ISBA-A- g_s en configuration de stratégie offensive. Pour imiter le temps de répétition des mesures de LAI obtenues avec les capteurs satellitaires, les observations sont considérées avec une fréquence de 10 jours. Ce jeu d'observations de LAI est obtenu à partir de l'interpolation des mesures faites sur le site de SMOSREX. Comme on l'a déjà vu à la fig. 3.5, la manque d'observations pendant certaines périodes limite parfois la qualité de l'interpolation. Les paramètres de l'assimilation sont les mêmes que ceux présentés dans l'article à la section 5.4. La fig. 5.6 montre les résultats obtenus : les analyses de la biomasse sont comparées à la simulation de contrôle et aux observations ; le LAI avant l'assimilation (c.à.d., l'estimation du modèle) et après l'assimilation des observations in situ du LAI. On observe une amélioration de l'estimation de ces deux variables, grâce à l'analyse des données in situ. En particulier, la phase de sénescence est améliorée, en réduisant la biomasse simulée. L'humidité de la zone racinaire avant et après assimilation est également comparée aux observations et à la simulation de référence. On n'observe pas un impact positif sur le contenu en eau du sol pour toute la période. Au contraire, l'estimation de w_2 est parfois dégradée par rapport au modèle, comme on peut observer, par exemple, pour les périodes de sécheresse de 2003 et 2004. Le nouvel état de la végétation modifie aussi le taux d'extraction racinaire et le flux d'évapotranspiration. Par exemple, une valeur de la biomasse plus basse en été, en 2001 et en 2002, s'accompagne d'une humidification du sol.

En conclusion, on a vu que l'assimilation d'une seule variable, w_g ou LAI, améliore l'estimation par le modèle du contenu total en eau du sol ou de la biomasse de la végétation,

respectivement. Néanmoins, l'assimilation d'une seule variable n'est pas suffisant pour obtenir une amélioration simultanée des deux variables analysées. Ce résultat nous amène à tester l'assimilation simultanée des deux variables, w_g et LAI. Dans l'article présenté dans la section suivant, ce sujet est traité en détail.

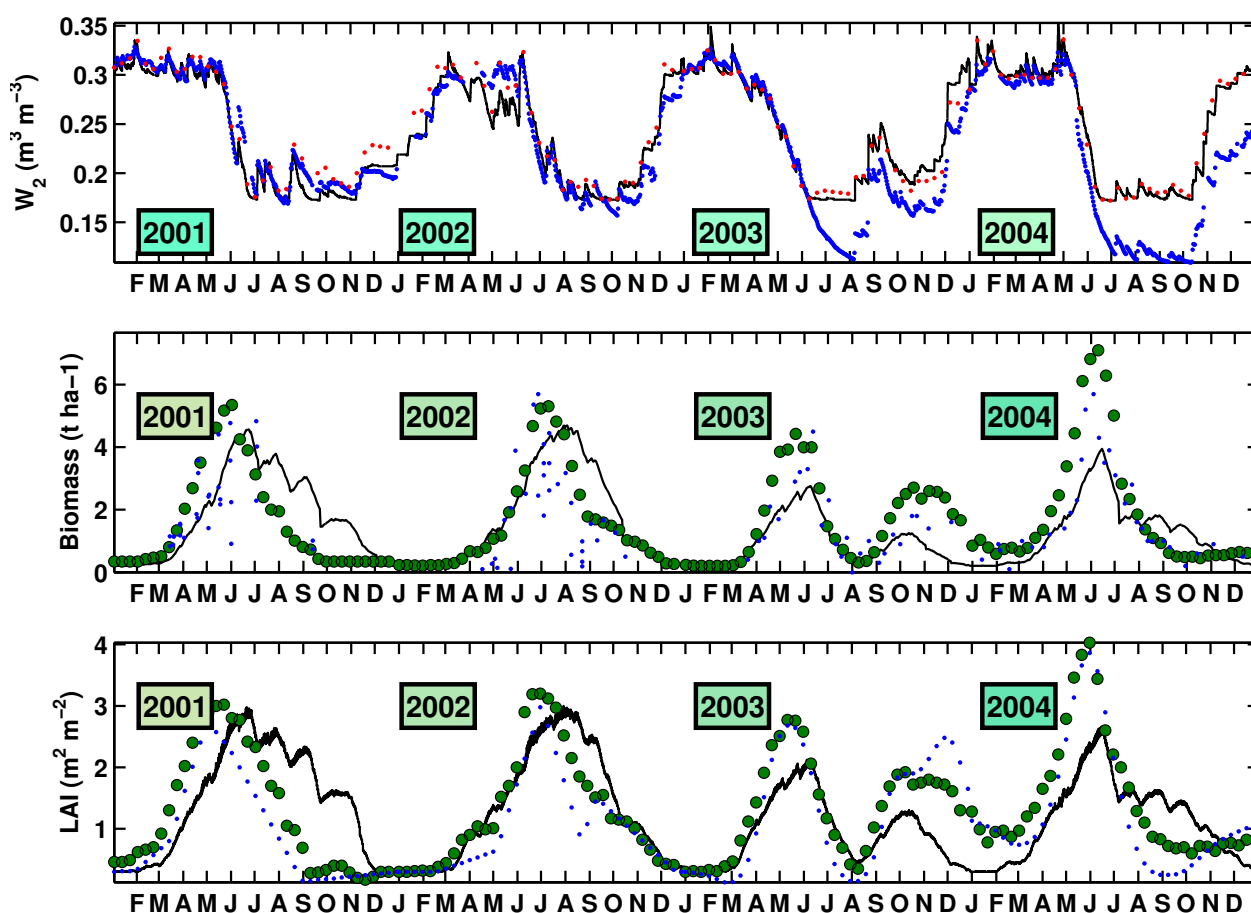


FIG. 5.6: En haut, l'estimation de w_2 après l'assimilation des observations de LAI (cercles rouges) est comparée au modèle (trait noir) et aux observations (points bleus). La figure au milieu montre les analyses de la biomasse (cercles verts) en comparaison à l'estimation du modèle (trait noir) et aux observations (points bleus). En bas, le LAI après assimilation (cercles verts) est aussi superposé à la référence (trait noir) et aux observations in-situ (points bleus).

5.4 ”Joint assimilation of surface soil moisture and LAI observations using a simplified 1D-VAR : The SMOSREX case study.”

5.4.1 Résumé

Observations de l’humidité superficielle du sol et LAI sont assimilés simultanément dans le modèle ISBA-A- g_s dans la configuration du modèle où la biomasse de la végétation est simulée (et pas prescrite). L’objectif est d’analyser et d’améliorer conjointement l’humidité de la zone racinaire et la biomasse de la végétation. L’amélioration de ces deux variables permettra d’estimer plus précisément les flux de chaleur et d’eau. La méthode d’assimilation utilisée est le 1D-VAR simplifié, dont on a montré plus haut qu’il améliore les estimations de l’humidité de la zone racinaire sur la jachère de SMOSREX. La période d’étude est longue et s’étend de 2001 à 2004. Son coût numérique relativement faible fait du 1D-VAR une méthode compétitive, pouvant potentiellement être implémentée dans une configuration opérationnelle. Plusieurs tests basés sur une méthode de Monte-Carlo permettent de fixer les paramètres de l’assimilation. Une solution est proposée pour résoudre le problème de la reprise de la végétation à la suite d’une période où l’assimilation impose un w_2 en dessous du wilting point. Finalement, les bénéfices de l’assimilation sont examinés dans des conditions où le forçage atmosphérique du modèle est affecté par de fortes erreurs, en particulier, en fixant les précipitations à zéro pendant toute la période.

5.4.2 Article (Soumis à Agricultural and Forest Meteorology)

JOINT ASSIMILATION OF SURFACE SOIL MOISTURE AND LAI OBSERVATIONS USING A SIMPLIFIED 1D-VAR: THE SMOSREX CASE STUDY

Joaquín Muñoz Sabater, Christoph Rüdiger, Jean-Christophe Calvet

Météo-France - CNRM, Toulouse, France

Lionel Jarlan

and Yann Kerr

CESBIO, Toulouse, France

ABSTRACT

Land surface models (LSM) offer a description of land surface processes and set the lower boundary conditions for meteorological models. In particular the accurate description of those surface variables which display a slow response in time, like root zone soil moisture or vegetation biomass, is of great importance. Errors in their estimation yield significant inaccuracies in the estimation of heat and water fluxes in Numerical Weather Prediction (NWP) models. In the present study, the ISBA-A-g_s LSM is used decoupled from the atmosphere. In this configuration, the model is able to simulate the vegetation growth, and LAI is estimated. A simplified 1D-VAR assimilation method is applied to observed surface soil moisture and LAI observations of the SMOSREX site near Toulouse, in south-western France, from 2001 to 2004. This period includes severe droughts in 2003 and 2004. The data are jointly assimilated into ISBA-A-g_s in order to analyse the root zone soil moisture and the vegetation biomass. It is shown that the 1D-VAR improves the model results. The score is increased from 0.79 to 0.86 for root-zone soil moisture and from 0.17 to 0.23 for vegetation biomass.

1. Introduction

The Land Surface Models (LSMs) used in meteorology have been developed in order to simulate continuous land surface processes, such as plant transpiration, soil evaporation and the evolution of soil moisture and surface temperature at regional and global scales. A number of CO₂-responsive LSMs, like the ISBA-A-g_s model of Météo-France (Calvet et al. 1998, Calvet et al. 2004, Gibelin et al. 2006), are able to simulate photosynthesis and plant growth. In particular, the vegetation biomass and the leaf area index (LAI) evolve dynamically in response to climate conditions. These models

allow the assimilation of soil moisture and LAI observations.

In the present study, two variables with significant impact on the heat and water fluxes are considered: root zone soil moisture (w_2) and vegetation biomass. Soil moisture regulates the partitioning between latent and sensible heat fluxes, which has a significant influence on the amount of cloud formation, temperature and humidity, among others (Segal et al., 1995, Shaw et al., 1997, Seuffert et al., 2002). At the same time, vegetation biomass plays a very important role in the exchange of water vapor and CO₂ between the vegetation canopy and the atmosphere. All aforementioned variables

are key input variables into Global Circulation Models (GCM) (Chase et al., 1996, Bounoua et al., 2000). However, as a consequence of different model complexities and surface parameterizations, there is a range of uncertainty in the simulation of these variables by LSMs (Henderson-Sellers et al., 1993, Henderson-Sellers et al., 1995, International GEWEX Project Office, 1995). Data assimilation systems allow to integrate the rich information provided by remotely sensed surface variables into LSMs, in order to improve the model predictions. To achieve this, LSMs are coupled with Radiative Transfer Models (RTM) to relate remotely sensed quantities to land-surface variables. For example, the L-Band Microwave Emission of the Biosphere (L-MEB) model (Wigneron et al., 2005) relates L-band brightness temperatures to surface soil moisture (w_g) variations. w_g products are already available in X-band at global scale through instruments such as the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) sensor (Njoku et al., 2003) or those that will be provided in L-band by the future Soil Moisture and Ocean Satellite (SMOS) mission of the European Space Agency (ESA) (Kerr et al., 2001). These observations can be used to analyze w_2 (Entekhabi et al., 1994, Reichle et al., 2001). Additionally, a large quantity of satellite-derived LAI products is available, such as those from the MODerate-Resolution Imaging Spectroradiometer (MODIS) at 1 km spatial resolution (Tian et al. 2002a, Tian et al. 2002b, Tan et al. 2005) or the POLarization and Directionality of Earth Reflectances (POLDER) sensor (Bicheron and Leroy, 1999, Roujean and Lacaze, 2002). LAI observations have been shown to be useful for the analysis of the vegetation biomass (Cayrol et al., 2000).

This paper is a continuation of the study of Muñoz Sabater et al. (2007) (henceforth called MU07). In MU07, w_g observations were assimilated using four different assimilation approaches. The simplified 1D-VAR method demonstrated to be the most suitable for an implementation in

an operational configuration. Although the results for the w_2 analyses were generally satisfactory, the LSM was forced with a prescribed LAI obtained from in-situ measurements. In the present paper, a step forward is made, and the LAI is simulated by the surface scheme using the parameterization for the plant photosynthesis and the vegetation growing and senescence phases of ISBA-A-gs (Gibelin et al. 2006). The main objective is the analysis of w_2 and vegetation biomass through the joint assimilation of LAI and w_g observations. The experimental data originates from the SMOSREX experimental site (De Rosnay et al., 2006). The period under investigation extends over four years from 2001 to 2004, during which very contrasting meteorological conditions were observed.

In section 2, the experimental site, the data set, the ISBA-A-gs LSM, and the data assimilation scheme are described. Also, the choice of the assimilation parameters is discussed through sensitivity and Monte-Carlo based methods. In section 3, the results are presented for the 4-year period and discussed for two configurations of the assimilation (fixed and time-variant wilting point) and for different accuracies of the atmospheric forcing. Finally, section 4 discusses and summarizes the main conclusions.

2. Methodology

2.1. Experimental site and data set

The SMOSREX site is situated within the ONERA (French National Aerospace Research Establishment) centre of Fauga-Mauzac, located 40 km at the South-West of Toulouse (43°23'N, 1°17'E, 188 m altitude). SMOSREX (De Rosnay et al., 2006) is a field scale experiment, operative since 2001. In this study, the southern part of the site, covered by vegetation (natural fallow) is studied. A representative estimation of w_g is obtained through the

measurements provided by groups of four probes (ThetaProbe, Delta T Devices) vertically installed at the surface at different locations within the experimental field site. They provide continuous observations over the first 6 cm of the soil surface layer. Daily mean w_g^j values are obtained by averaging the measurements of the four probes and their standard deviation permits to estimate the uncertainty of the observations. The volumetric soil moisture is obtained by calculating an average bulk soil water content from the surface probes and additional soil moisture sensors horizontally installed at depths of 10, 20, 30, 40, 50, 60, 70, 80 and 90 cm. From the individual measurements, a mean and a standard deviation are computed on a daily basis, and the daily standard deviation averaged over the year 2001 is assumed to represent the observation error: $\sigma(w_g^{OBS}) = 0.03 \text{ m}^3 \text{ m}^{-3}$ and $\sigma(w_2^{OBS}) = 0.02 \text{ m}^3 \text{ m}^{-3}$. Due to a lack of sufficient spatial sampling at the field site, the spatial averaging is replaced by a temporal one (ergodicity principle). These uncertainties are attributed to the moisture observations in subsequent years (2002-2004). The atmospheric forcing (precipitation, atmospheric pressure, wind speed and direction, air humidity, air temperature, and incident solar and infrared radiation) is obtained from a meteorological station at the field site performing continuous measurements at 30 minute intervals.

Manual measurements of the vegetation characteristics (LAI, green and dry biomass and height of the canopy) were generally carried out every two weeks. Fig. 1 shows the in-situ observations of the LAI, the surface and root-zone soil moisture and the monthly accumulated precipitation for the four years (2001-2004). It is shown that 2003 was a particularly dry year, with an annual cumulative precipitation of less than 600 mm. Unlike the other years, 2003 shows an atypical double cycle of LAI, with a first peak in spring and a second one at the

beginning of the winter season. Precipitation is quite irregularly distributed during 2004, with a wet spring and a very dry summer. This causes a rapid growth of the vegetation and a marked senescence during the dry period, with moisture observations reaching values below the wilting point (w_{wilt}) throughout the summer and part of autumn.

In Table 1 a list of the most relevant characteristics of the soil at the SMOSREX site for ISBA-A- g_s is provided. The soil is a loam characterized by its texture and density which were determined in the laboratory. The w_{wilt} and the field capacity parameters were derived from the clay content observations, by using the relationships given by Noilhan and Mahfouf (1996).

2.2. Land surface model: ISBA-A- g_s

The surface scheme used in this study is the ISBA-A- g_s model (Calvet et al., 1998, Calvet et al., 2004, Gibelin et al., 2006). It is an improved version of the Interaction between the Soil, Biosphere and Atmosphere (ISBA) model, first developed by Noilhan and Planton (1989) and further improved by Noilhan and Mahfouf (1996), to describe the land surface processes in weather and climate models.

The version of the ISBA model used in this study is based on the equations of the force-restore method (Deardorff, 1977, 1978), and the soil is represented by a single layer. Together with the surface fluxes (LE , H , G), five surface state variables are described: surface temperature (T_s), mean surface temperature (T_2), surface soil volumetric moisture (w_g), total soil volumetric moisture (w_2) and canopy interception reservoir (W_s). The new version (ISBA-A- g_s) accounts for the effect of the atmospheric CO_2 concentration on the stomatal aperture. The net assimilation of CO_2 is used to predict the vegetation biomass and the LAI. The ecosystem respiration is calculated. However, the parameterization does not account for the

TABLE 1.

Main soil and vegetation parameters used in the ISBA-A-g_s simulations over the SMOSREX site.

Soil parameters			
parameter	symbol	Unit	Value
Soil root depth	d_2	cm	95
Sand Content	SAND	%	32.0
Clay content	CLAY	%	22.8
Field capacity	w_{fc}	$m^3 \cdot m^{-3}$	0.30
Wilting point	w_{wilt}	$m^3 \cdot m^{-3}$	0.17
Vegetation parameters			
parameter	symbol	Unit	Value
Mesophyll conductance	g_m	$mm \cdot s^{-1}$	0.56
Critical extractable soil moisture	θ_c	%	50
Plant response to water stress	-	-	drought-tolerant
Potential leaf life expectancy	τ	days	80
LAI minimal	LAI_{min}	$m^2 \cdot m^{-2}$	0.3
Cuticular conductance	g_c	$mm \cdot s^{-1}$	0
Nitrogen plasticity parameter (slope)	e	$m^2 \cdot kg^{-1} \cdot \%^{-1}$	5.84
Nitrogen plasticity parameter(intercept)	f	$m^2 \cdot kg^{-1}$	6.32
Leaf nitrogen concentration	N_l	%	1.4
Respiration du sol a 25 °C	Re25	$mg \cdot CO_2 \cdot m^2 \cdot s^{-1}$	0.103

effect of the soil moisture variations. Two strategies are represented for the response of the plant to the water stress: drought-avoiding and drought-tolerant (Calvet 2000, Calvet et al. 2004). The vegetation parameters of ISBA-A-g_s are presented in

Table 1. They are derived from the simulation of Calvet (2000) for the MUREX test site (Calvet et al., 1999). Following their work, the drought-tolerant strategy is used in this paper.

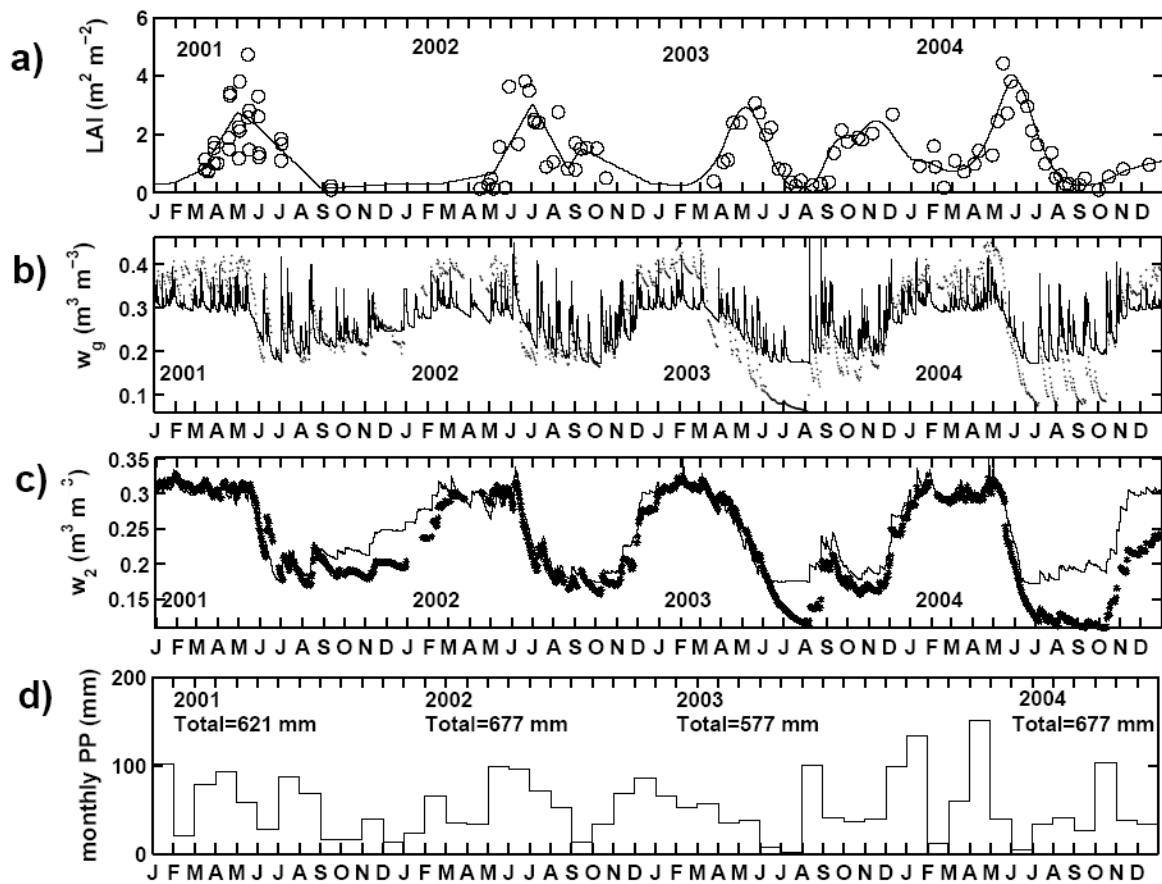


Figure 1.- a) LAI , b) w_g , c) w_2 and d) monthly precipitation from 2001 to 2004 as measured over the SMOSREX site. On top the interpolated LAI (solid line) is superimposed to the observations (circles). Note that the LAI interpolation technique may change from one year to another, depending on data quality and frequency. The w_g and w_2 simulations of ISBA-A-g, (solid line) using the offensive strategy as response to water stress (drought tolerant) are superimposed to the observations (dots for w_g and stars for w_2).

2.3. Assimilation method: Simplified 1D-VAR

The choice of the assimilation scheme is based on the results obtained in MU07. Of the four assimilation methods compared in that study, the Ensemble Kalman Filter (EnKF) and a simplified 1D-VAR showed the best performance. However, the variational scheme presented an important advantage, as the computational cost was significantly lower, making it more suitable for operational purposes. Therefore,

a simplified 1D-VAR was employed in this study.

2.3.1. Formalism

Variational methods use an assimilation window of predefined length, usually containing several observations. The initial value of the model state vector (the variables to be analysed) is perturbed and the optimal combination of model states is then found by minimizing a cost function J (at observation times within the assimilation window). In our case, the state vector is

composed of two variables (w_2 and vegetation biomass (Bio)). The cost functions for the two state variables are:

$$J(Bio) = \frac{1}{2} (Bio - Bio^b)^T \mathbf{B}_{BIO}^{-1} (Bio - Bio^b) + \frac{1}{2} (LAI^{OBS} - H(Bio))^T \mathbf{R}_{LAI}^{-1} (LAI^{OBS} - H(Bio)) = J_{Bio}^b + J_{LAI}^{OBS} \quad [9]$$

$$J(w_2) = \frac{1}{2} (w_2 - w_2^b)^T \mathbf{B}_{w_2}^{-1} (w_2 - w_2^b) + \frac{1}{2} (w_g^{OBS} - H(w_g))^T \mathbf{R}_{wg}^{-1} (w_g^{OBS} - H(w_g)) = J_{w_2}^b + J_{wg}^{OBS} \quad [10]$$

where the superscripts b and OBS denote the initial values of the state variables within the assimilation window (background) and the observed variables, respectively; and H is the non-linear observation operator which relates observations and simulations projected in the space of observations. The “best” analyses are the result of minimizing the distance between the state variables and the first-guess weighted by the background covariance matrix $\mathbf{B}$ (J_{Bio}^b and $J_{w_2}^b$ terms) and the distance between the observations and the analogous in the space of the observations (J_{LAI}^{OBS} and J_{wg}^{OBS} terms) within the assimilation window, weighted by the covariance matrix of the observations $\mathbf{R}$. In the case of a linear observation operator and a normal distribution of the errors, the expressions minimizing the cost functions [9] and [10] are, respectively:

$$Bio = Bio^b + \mathbf{K}_{BIO} (LAI^{OBS} - \mathbf{H}_{BIO} Bio^b); \quad \mathbf{K}_{BIO} = \mathbf{B}_{BIO} \mathbf{H}_{BIO}^T [\mathbf{H}_{BIO} \mathbf{B}_{BIO} \mathbf{H}_{BIO}^T + \mathbf{R}_{LAI}]^{-1} \quad [11]$$

$$w_2 = w_2^b + \mathbf{K}_{w_2} (w_g^{OBS} - \mathbf{H}_{w_2} w_2^b); \quad \mathbf{K}_{w_2} = \mathbf{B}_{w_2} \mathbf{H}_{w_2}^T [\mathbf{H}_{w_2} \mathbf{B}_{w_2} \mathbf{H}_{w_2}^T + \mathbf{R}_{wg}]^{-1} \quad [12]$$

In the non-linear case, the minimum of the cost functions [9] and [10] is evaluated by deriving adjoint or tangent linear models. However, this is an often difficult and time consuming task (Reichle et al., 2002). The simplified 1D-VAR (Balsamo et al., 2004) circumvents this problem by using a numerical linear approximation of the observation operator. This approximation consists of a perturbation of the initial value of the state variables ($w_2' = w_2^b + \delta w_2^b$ for w_2 and $Bio' = Bio^b + \delta Bio^b$ for the vegetation biomass) and studying the impact of this perturbation on the predicted variables at each observation time step i within the assimilation window (with n observations). In this way, linearised observation operators

of the type: $\mathbf{H} = \left[\frac{\partial x^{OBS}}{\partial x} \Big|_{t=1}, \dots, \frac{\partial x^{OBS}}{\partial x} \Big|_{t=n} \right]$ are

obtained, where x represents a state variable and x^{OBS} an observed variable. This operator is then substituted in [11] and [12].

2.3.2. Setting of the main 1D-VAR parameters

The most difficult task in the implementation of an assimilation scheme is the description of the error matrices. It is a key aspect, because the correction of the system state depends on the background and observation error prescription (see Eqs. [11] and [12]). The size of the initial perturbation and the length of the assimilation window is also discussed in this section.

(i) Assimilation window length

The choice of the assimilation window length for a variational assimilation scheme is particularly important when the scheme is applied to large regions. In this study, a 10-day assimilation window, which is close to the sampling time of many LAI products (e.g. a 8-day MODIS product is available), was chosen because shorter assimilation windows would regularly exclude LAI observations altogether. Longer windows resulted in a degradation of the

model prediction. For example, a test with a 30-day assimilation window showed that the

$$\text{score } E \left(E = 1 - \frac{\sum_i (x_i^{\text{obs}} - x_i^{\text{mod/ana}})^2}{\sum_i (x_i^{\text{obs}} - \overline{x^{\text{obs}}})^2} \right) \text{ of } w_2$$

and the vegetation biomass decreased to 0.79 and became negative, respectively.

(ii) Observational errors

In this study, observed time series of surface soil moisture (w_g) and LAI were used for the assimilation into the model. For w_g , the same method as in MU07 was applied, i.e. the error derived from the dispersion of in-situ observations was doubled and therefore set to $\mathbf{R}_{wg} = 0.06 \text{ m}^3 \text{m}^{-3}$. For LAI, an observational error is difficult to obtain from the experimental procedure followed in SMOSREX. Moreover, the LAI observations exhibited quite different cycles (in terms of shape and maxima) in 2001 and 2002. It was decided to use an empirical value of $\mathbf{R}_{LAI} = 1 \text{ m}^2 \text{m}^{-2}$, which is expected to account for all sources of uncertainties (observational system and representativeness error). Both values were assumed to be constant throughout the whole period 2001-2004.

(iii) Background errors

For the description of the error in the estimation of w_2 , an ensemble of simulations was created between two consecutive observations by using different initial states of w_2 and by perturbing the atmospheric forcing. For each assimilation window the dispersion of the residuals (difference between an observation and the model estimation) was assessed. During the calibration year (2001) this quantity was shown to have a constant value of $0.02 \text{ m}^3 \text{m}^{-3}$. The same approach was used to estimate a model covariance error for the vegetation biomass over the 10-days of the individual assimilation windows. It was observed that the evolution of the standard deviation of the residuals does not vary much over the year

2001. However, a slight seasonal effect is observed (not shown). Departure from the mean value is about 0.01 kg m^{-2} on average. Therefore, the background error for the vegetation biomass was assumed to be constant and was set equal to the average of the standard deviation of the residuals in 2001 (0.05 kg m^{-2}).

(iv) Linear hypothesis

A further important parameter in the implementation of this simplified variational assimilation scheme is the choice of the perturbation size of the state variables, because it determines the quality of the linear approximation through the observation operator $\mathbf{H}$. To evaluate the “optimal” size of this perturbation, the following test was conducted: an ensemble of linear operators $\mathbf{H}$ was obtained by finite differences (by a random Gaussian perturbation of the initial value of w_2 and/or the vegetation biomass) for a 10-day period (four components for $\mathbf{H}_{w_2}$ and one for $\mathbf{H}_{Bio}$) and their standard deviation (σ) was calculated. Several values of the initial perturbation of the state variables were tested. The smallest perturbations of w_2 (between 0.005 and $0.015 \text{ m}^3 \text{m}^{-3}$) produced a strong sensitivity of σ to the four linearized components of $\mathbf{H}_{w_2}$ during the calibration year (not shown). This is due to the numerical noise produced by infinitesimal perturbations in the vicinity of the initial value. The most suitable perturbation should result in a low sensitivity of $\mathbf{H}_{w_2}$. $\mathbf{H}_{w_2}$ was found to be less sensitive for perturbations close to $0.05 \text{ m}^3 \text{m}^{-3}$ and this value was used in the assimilation algorithm.

Fig. 2 shows similar results for the sole component of $\mathbf{H}_{Bio}$ in a 10-day assimilation window. The optimal perturbation is located between 0.07 and 0.1 kg m^{-2} and it was finally set to 0.07 kg m^{-2} . Larger intervals of the perturbation were found to produce larger errors in $\mathbf{H}_{Bio}$ (not shown).

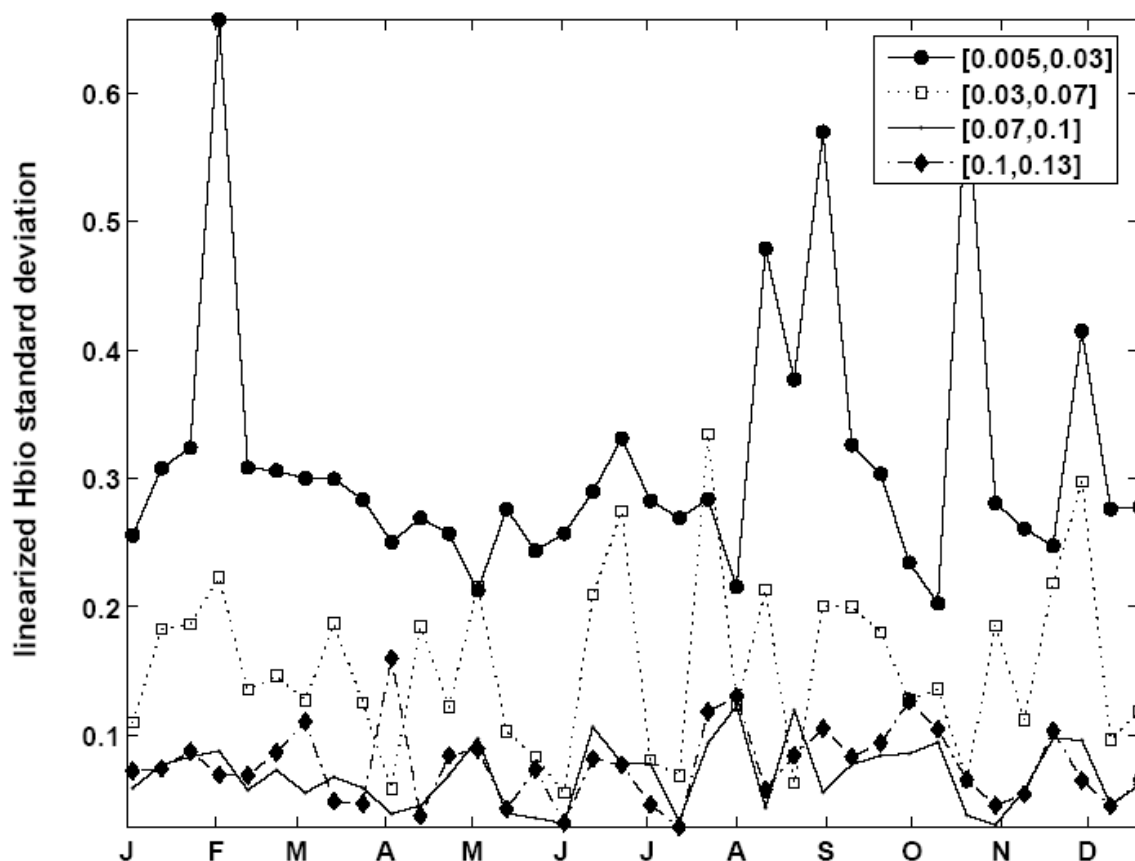


Figure 2.- Temporal evolution of the linearized H_{bio} standard deviation. H_{bio} is computed by an initial perturbation of the vegetation biomass. Four different intervals for the initial perturbation of the biomass are tested. Perturbations are given in $m^2 Kg^{-1}$.

2.3.3. Dynamical correction of the wilting point

As mentioned before, w_{wilt} is estimated by the LSM and then treated as a constant parameter set to $0.17 m^3 m^{-3}$. However, the assimilation scheme is allowed to initialise the model with w_2 states below w_{wilt} , following the assimilation of w_g . This occurred during the drought periods of the years 2003 and 2004 (MU07). In this case, ISBA-A- g_s did not allow any further evapotranspiration and soil water extraction through the roots, even after significant rainfall events. This led to a strong vegetation mortality rate. However, the SMOSREX field observations performed during the summer 2003 show that the vegetation growth in response to rainfall may be rapid (Fig. 1). The performance of the

vegetation analyses was also affected by this constraint. The following solution is proposed: for each assimilation step with an analysed value of w_2 below the fixed w_{wilt} , the model w_{wilt} is re-set by substituting its value with the analysed value of w_2 . This simple modification to the model allows plant growth in response to rainfall to occur after a drought.

3. Results

3.1. w_2 and biomass analyses (2001-2004)

In Fig. 3, the 1D-VAR analysis of w_2 and the vegetation biomass are presented for the years 2001 to 2004 (Fig. 3a and 3b, respectively) with a non-stationary w_{wilt} (see section 2c). The predicted LAI (control and

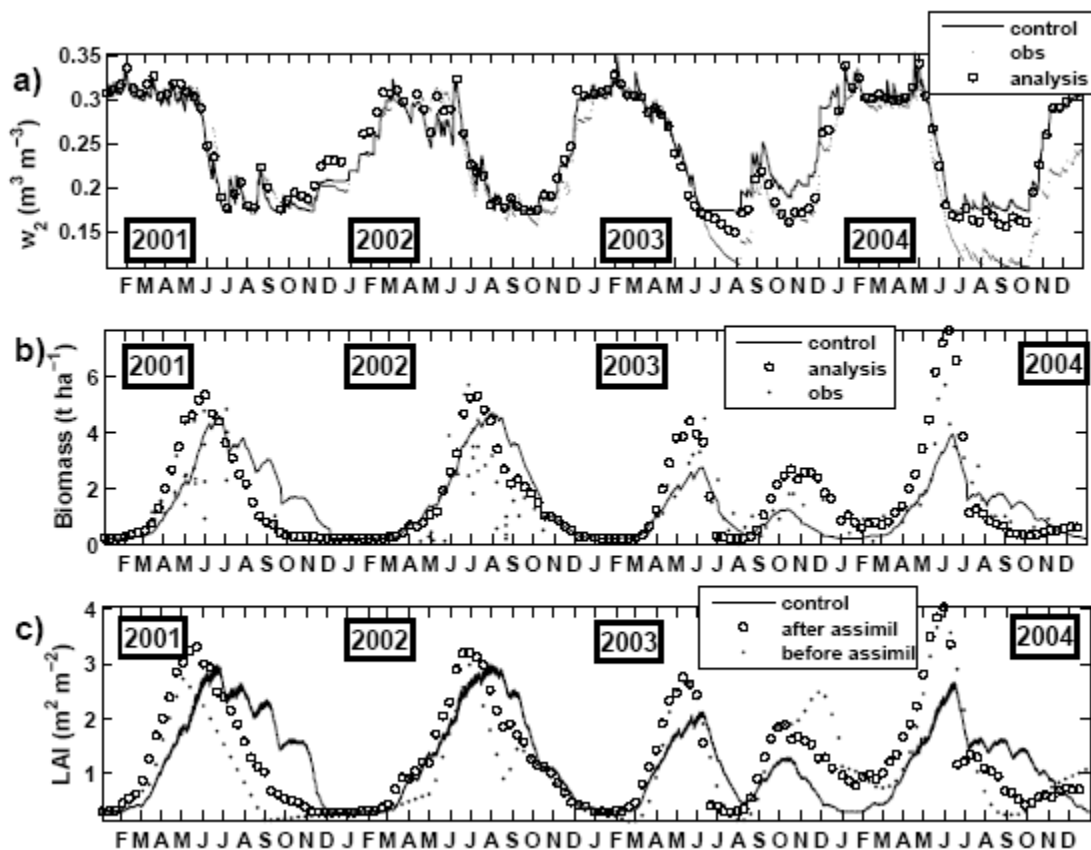


Figure 3.- Analysis of a) the root zone soil moisture (circles) and b) vegetation biomass, using a simplified 1D-VAR method from 2001 to 2004 over the SMOSREX experimental site. c) LAI before and after the assimilation. For comparison purposes, analysed values are superimposed over the in-situ observations (points) and the model basic estimations (solid line).

TABLE 2.

Yearly and global RMSE (in $\text{m}^3 \text{m}^{-3}$), mean bias (in $\text{m}^3 \text{m}^{-3}$) and score (E), for the root zone-soil moisture, LAI and biomass between the observations and the control model simulation.

	2001			2002			2003			2004			2001-2004		
	RMSE	mb	E	RMSE	mb	E	RMSE	mb	E	RMSE	mb	E	RMSE	mb	E
w_2	0.01	-.005	0.97	0.02	-.004	0.89	0.03	0.021	0.74	0.05	0.04	0.58	0.03	0.015	0.79
LAI	0.97	0.47	0.40	0.60	0.32	0.35	0.79	-0.33	0.25	0.81	-.12	0.33	0.80	0.09	0.17
Biomass	1.17	-.07	0.15	1.90	1.19	-.47	0.64	-.28	0.53	0.77	-.20	0.62	1.26	0.20	0.17

analysis) is also shown on Fig. 3c. The assimilated observations are the surface soil moisture (one observation every three days) and the LAI (one observation every 10 days). Analysis, observations and the control model reference are superposed for comparison purposes. It is shown that w_2 is generally well reproduced during the whole period from 2001 to 2004, with a score E equal to 0.86 after the assimilation in contrast to 0.79 for the open-loop simulation (Tables 2 and 3). The added value of the assimilation is particularly noticeable during the six months following June 2003, for which the model without assimilation is not able to descend below the prescribed w_{wilt} and consequently overestimates the moisture during the rest of that year. When comparing Fig.3a of this study with Fig.4 of MU07, two main differences are observed with regard to the analyses of w_2 : first, a degradation at the end of 2001 is observed, second, the correction over the dry period of 2004 is less accurate. This is a consequence of analysing the vegetation biomass in addition to soil moisture. From October 2001 to February 2002 no LAI observations were available, which resulted in interpolated values of less than $0.5 \text{ m}^2 \text{ m}^{-2}$ during this period. In this study, these estimates were assimilated and the LAI retrievals after assimilation were close to these observations. This produced lower transpiration and water extraction rates, hence a slower drying of the water reservoir and as a consequence an overestimation of the w_2 analyses.

On Fig. 3c, the impact of the joint w_2 and biomass analysis is shown for the LAI predictions. The predicted values of LAI are quite close to the observations, improving the performance from $E = 0.17$ for the control run to $E = 0.72$, and RMSE from $0.80 \text{ m}^2 \text{ m}^{-2}$ to $0.47 \text{ m}^2 \text{ m}^{-2}$ after the assimilation (Tables 2 and 3). The main disagreements are found during the senescence periods of the vegetation when the model clearly overestimates the LAI (except for 2003, where an anomalous double cycle of LAI was observed). Since the biomass observations are sparse and very

scattered for 2001 and 2002, it is not clear whether the assimilation improves the open-loop model simulation for these years. It is observed that during the second cycle of vegetation in 2003, the biomass is slightly overestimated after the analysis, which is in agreement with the higher assimilated LAI observations. The RMSE and the score for the years 2003 and 2004 were not improved, which may be explained by the limited number of available observations of biomass. However, for the whole study period, the analysis of the biomass improved the score from 0.17 to 0.23.

On Fig. 4, the impact of setting a time-invariant w_{wilt} is only shown for 2003 and 2004, because w_{wilt} was modified during these two years, only. During the periods with initial states of w_2 below w_{wilt} an increased vegetation mortality rate is observed. The need for re-setting w_{wilt} is particularly apparent between July and October 2004, when the biomass analyses were close to zero. During this period, each time w_2 analyses were below w_{wilt} ISBA-A-gs stopped plant transpiration and photosynthesis, which resulted in a high vegetation mortality rate. Consequently, the performance of the biomass analyses for 2003 and 2004 was low (Table 3) compared to the case of modifying w_{wilt} . This effect is less evident in 2003, since the drying period is shorter than in 2004. For the complete four years, the score of the vegetation biomass decreases (compared with using a non-stationary w_{wilt}) from 0.23 to 0.12. The lower biomass during the dry periods also had an impact on the soil water reservoir, which produced higher moisture levels. The score for w_2 decreased from 0.86 to 0.82 (Table 3).

3.2. CO_2 fluxes

It is important to assess the impact the analyses have on the prediction of CO_2 and water fluxes. The carbon balance estimation is fundamental for a good simulation of the plant functioning (growing

TABLE 3.

Yearly and global RMSE (in $\text{m}^3 \text{m}^{-3}$), mean bias (in $\text{m}^3 \text{m}^{-3}$) and score (E) for the root zone-soil moisture and vegetation biomass, between analysis and observations for four different model configurations. For the LAI, statistics are calculated between observations and model simulations after data assimilation.

		2001			2002			2003			2004			2001-2004		
		RMS E	mb	E	RMS E	mb	E	RMS E	mb	E	RMS E	mb	E	RMS E	mb	E
Fixed w_{wilt}	w_2	0.01	0.03	0.94	0.02	0.01	0.90	0.02	0.01	0.90	0.05	0.04	0.60	0.03	0.01	0.82
	LAI	0.41	0.38	0.65	0.46	0.32	0.62	0.45	- 0.07	0.75	0.50	- 0.02	0.73	0.48	0.16	0.70
	Biomass	1.26	0.15	0.02	1.63	1.16	0.08	0.93	0.31	0.26	1.22	0.37	0.18	1.41	0.32	0.12
Non-fixed w_{wilt}	w_2	0.01	0.03	0.94	0.02	0.01	0.90	0.02	0.01	0.94	0.04	0.02	0.72	0.03	0.01	0.86
	LAI	0.41	0.38	0.65	0.46	0.32	0.62	0.41	- 0.07	0.76	0.45	0.05	0.79	0.47	0.19	0.72
	Biomass	1.26	0.15	- 0.21	1.63	1.16	- 0.08	0.91	0.30	0.48	1.15	0.39	0.39	1.43	0.78	0.23
PP=0, fixed w_{wilt}	w_2	0.03	- 0.02	0.68	0.05	- 0.04	0.09	0.04	- 0.03	0.65	0.05	- 0.02	0.66	0.04	- 0.03	0.57
	LAI	0.31	- 0.02	0.81	0.73	- 0.46	0.02	1.16	- 0.72	0.63	0.85	- 0.49	0.27	0.82	- 0.46	0.14
	Biomass	1.25	- 0.70	0.03	1.84	- 1.31	0.38	1.46	- 0.83	0.94	1.08	- 0.42	0.47	1.44	- 0.82	- 0.06
PP=0, non-fixed w_{wilt}	w_2	0.03	- 0.02	0.68	0.05	- 0.04	0.09	0.05	- 0.04	0.09	0.06	- 0.03	0.51	0.05	- 0.03	0.47
	LAI	0.30	- 0.02	0.86	0.73	- 0.46	0.02	0.73	- 0.46	0.02	0.73	- 0.36	0.46	0.64	- 0.29	0.48
	Biomass	1.25	- 0.70	0.03	1.84	- 1.31	0.38	0.96	0.34	0.15	0.87	- 0.31	0.66	1.26	0.40	0.13

and senescence phases), which indirectly has a significant impact on the water reservoir evolution. In Fig. 5, the gross primary production (GPP), the ecosystem respiration (Reco), and the balance between both fluxes, the Net Ecosystem Exchange (NEE), are presented, before and after data assimilation. In order to reduce noise due to daily variations, the model outputs (1 value every 30 min) are smoothed using a boxcar average of 10 days. The reduction of the

GPP and NEE after the data assimilation at the end of 2001 and 2004 is consistent with the lower values of LAI and biomass. Despite the reduction of the net assimilation, the larger vegetation biomass obtained during the start of the growing phase produced an increase of the annual GPP. Indeed, an increase in GPP is observed for the four years (Table 4). The annual NEE is negative only for 2003, ($-60.2 \text{ g} \cdot \text{C} \cdot \text{m}^{-2} \cdot \text{year}^{-1}$ after data assimilation), transforming the

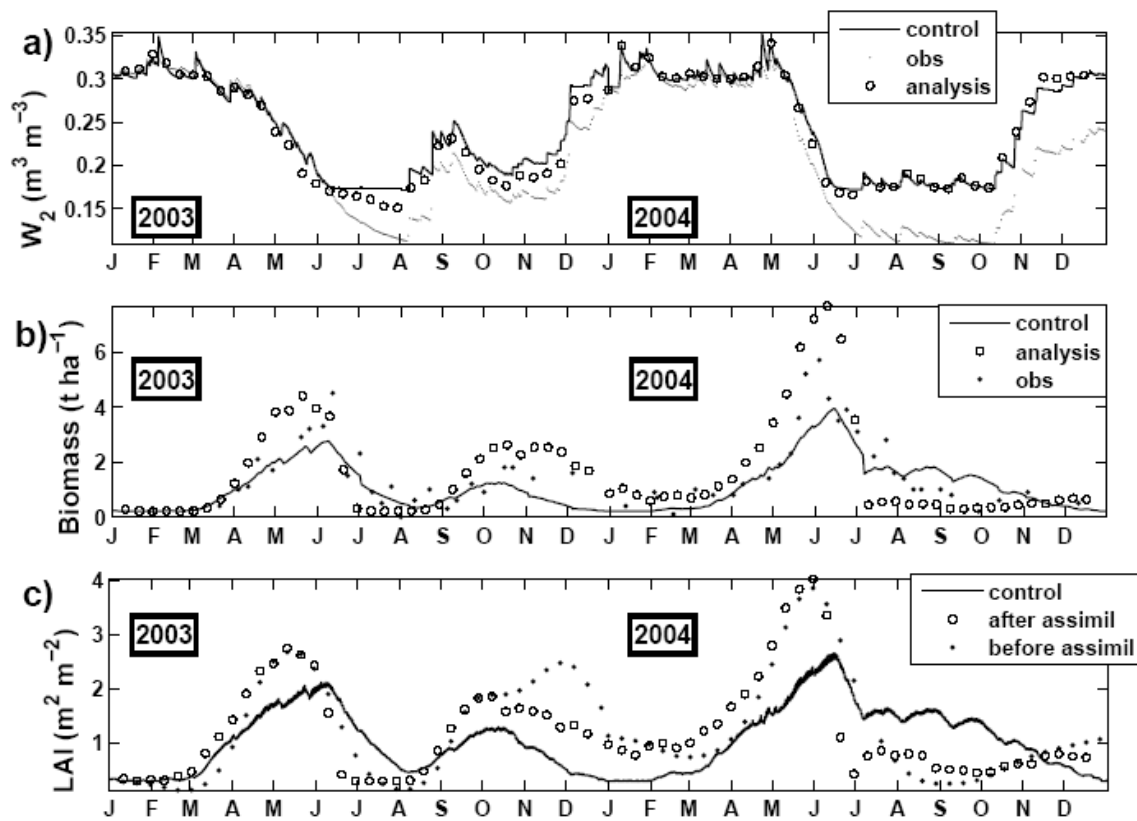


Figure 4.- Same as Fig. 3, but using a fixed wilting point. The results for the years 2001 and 2002 are omitted as being similar to those of Fig. 3.

fallow of the chosen site into a CO₂ source, as it happened in many parts of Europe for 2003 (Ciais et al., 2005). However, it must be noted that in this study, the ecosystem respiration does not account for soil moisture and is probably overestimated in 2003. This may explain why it is not significantly affected by the data assimilation.

In Fig. 5d, the evolution of evapotranspiration is shown before and after data assimilation. It is observed that, as a result of using lower initial values of w_{wilt} , plant transpiration and GPP occurred in July 2003, whereas a time-invariant w_{wilt} produced flux values close to zero (not shown). The lower values of the analysed biomass from July 2004 onwards, result in a decrease of the evapotranspiration. This lower vegetation biomass resulted in small reductions of the yearly total cumulated evapotranspiration flux for 2003 and 2004,

respectively, in comparison to the control (Table 4). By using a fixed w_{wilt} the GPP and NEE fluxes are likely to be underestimated, because the vegetation mortality is not well reproduced, which results in lower GPP fluxes as compared to predictions with a non-stationary w_{wilt} (difference of 13.9 and 31.9 g·C·m⁻²·year⁻¹ for 2003 and 2004, respectively).

3.3. Zero precipitation.

The simulations at SMOSREX are forced with good quality measurements of meteorological variables, representative for the study area. However, at larger scales the information about meteorological variables is less accurate. In order to test the robustness of the 1D-VAR, even when forcing data are subjected to large errors, the precipitation was set to zero. Precipitation

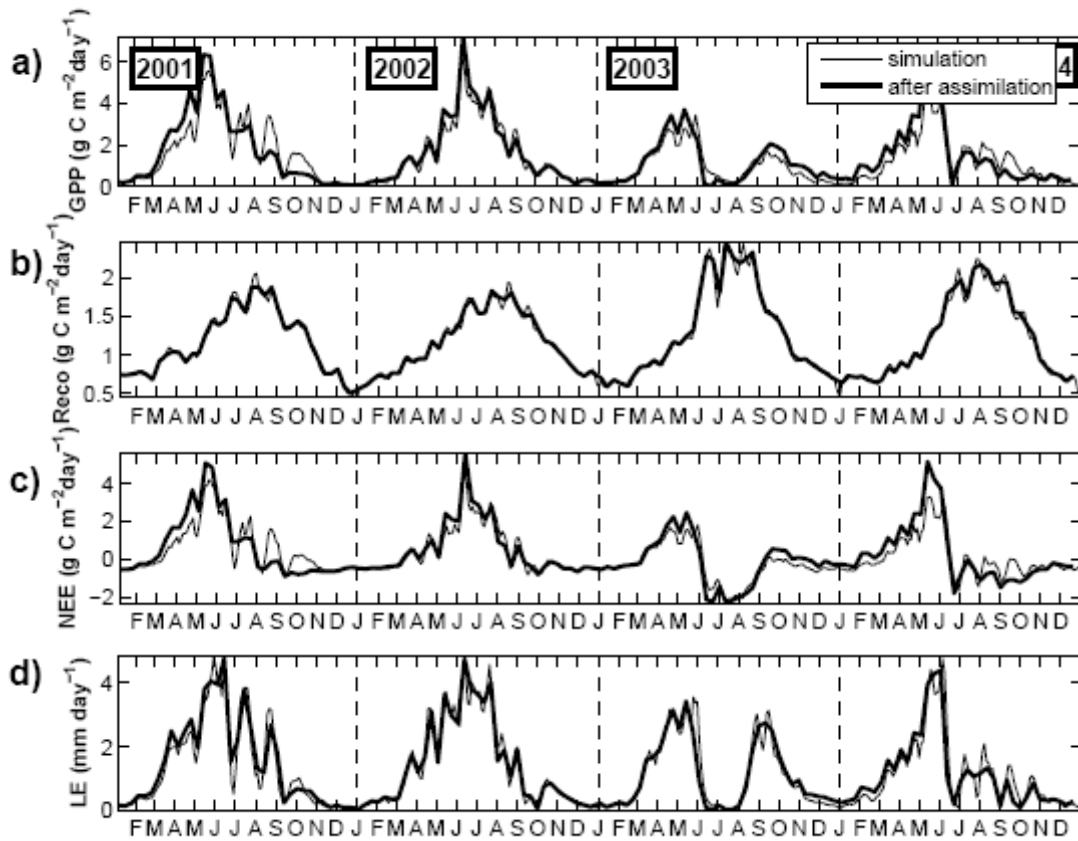


Figure 5.- Yearly evolution of a) the Gross Primary Production (GPP), b) the ecosystem respiration (Reco), c) the Net Ecosystem Exchange (NEE), and d) the evapotranspiration flux (LE), before (thick line) and after (fine line) the assimilation of surface soil moisture and LAI observations. The data for the model has been smoothed using a 10-day boxcar average window.

is a key variable within atmospheric forcing data sets and is particularly important for the water reservoir evolution. For this experiment, the assimilation scheme was run again for the whole 4-year period under investigation. Fig. 6 shows the same plots as Fig. 3, but with the new, degraded forcing data. Although the new analyses produced a drier soil, the general shape of the curve of the analyzed w_2 was maintained. It is now much smoother, since the information about the precipitation events is lacking. It is important to remark that the increase in soil volumetric moisture at the beginning of the growing period is mainly due to the assimilation scheme. Otherwise, given the absence of precipitation, a growing season would not be simulated. The LSM, with

precipitation set to zero, tends to decrease the soil volumetric moisture between two observations. However, the 1D-VAR generates positive increments which correct the water deficiency in order to match the w_g observations. The advantage over an open-loop simulation is considerable. Under the precipitation constriction, the open-loop simulation would rapidly fall to $0.17 \text{ m}^3 \text{ m}^{-3}$ (constant w_{wilt}), and since the model would not receive precipitation, the moisture reservoir would remain at this lower boundary. The benefits of the data assimilation are also clear for the vegetation: although the biomass analyses are obtained with less accuracy (in particular, lower peak values are obtained), especially during the

TABLE 4.

CO₂ and evapotranspiration (LE, in mm year⁻¹) yearly cumulated fluxes. The assimilated CO₂ by the vegetation (GPP, in g·C·m⁻²·year⁻¹), the CO₂ emission by soil and plants (Reco, in g·C·m⁻²·year⁻¹) and the balance between both terms (NEE, in g·C·m⁻²·year⁻¹) after the assimilation of LAI and w_g observations for four different model configurations, is compared to the control simulation.

		2001	2002	2003	2004
Control simulation	GPP	621.8	583.6	363.3	473.2
	Reco	421.6	422.8	467.7	443.1
	NEE	200.2	160.8	-104.4	30.1
	LE	505.8	523.8	407.1	398.8
fixed w_{wilt}	GPP	641.9	614.0	393.8	497.1
	Reco	419.5	419.0	467.6	441.5
	NEE	222.4	195	-73.8	55.6
	LE	535.8	520.5	378.5	375.4
Non-fixed w_{wilt}	GPP	654.3	614.0	405.7	529.0
	Reco	419.2	419.1	465.9	440.2
	NEE	235.1	194.9	-60.2	88.8
	LE	543.8	520.4	400.0	395.2
fixed w_{wilt} , PP = 0	GPP	342.7	185.4	125.5	197.8
	Reco	436.5	439.1	482.9	453.1
	NEE	-93.8	-253.7	-357.5	-255.3
	LE	230.8	144.7	113.4	156.0
Non-fixed w_{wilt} , PP = 0	GPP	342.7	185.4	283.8	272.6
	Reco	436.5	439.1	479.5	451.7
	NEE	-93.8	-253.7	-195.7	-179.1
	LE	230.8	144.7	171.1	174.4

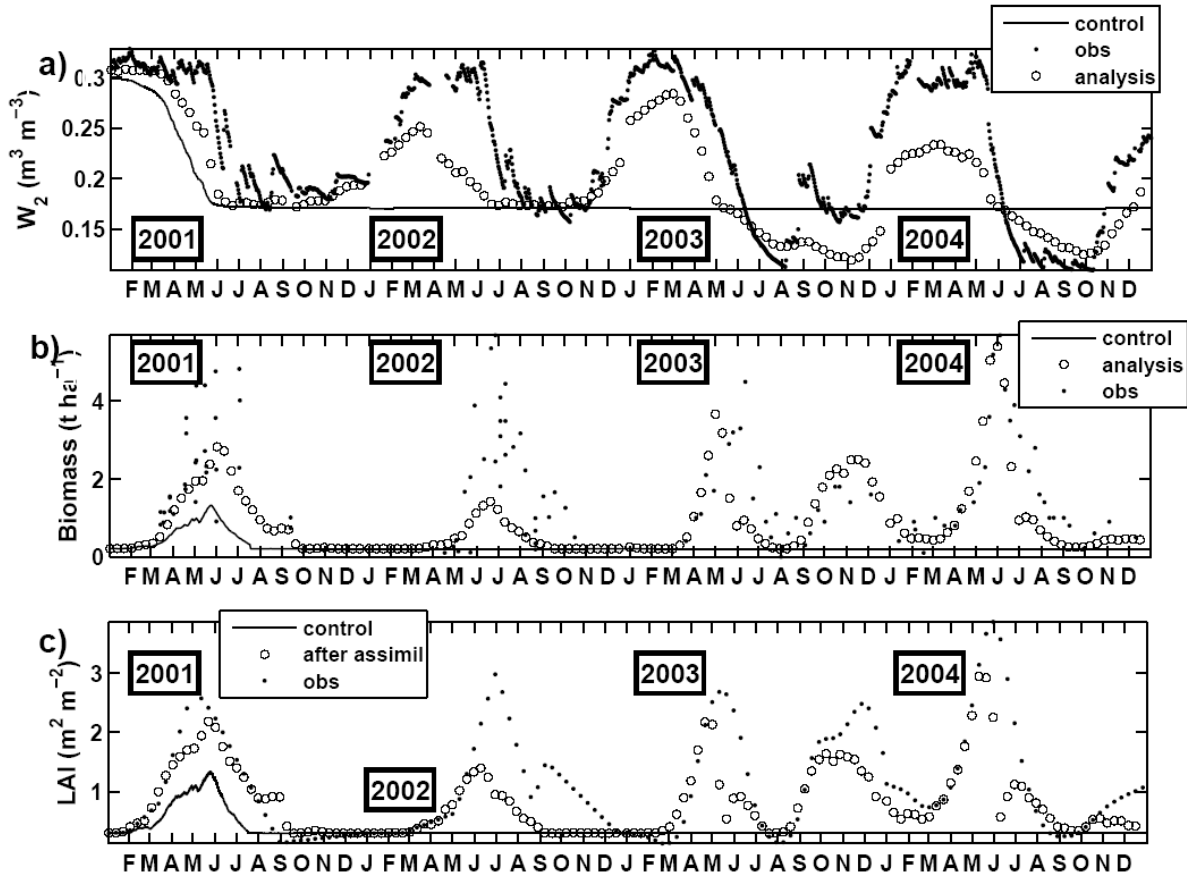


Figure 6.- Analysis of a) the root zone soil moisture (circles) and, b) vegetation biomass, using a simplified 1D-VAR method from 2001 to 2004, with precipitation set to zero. c) LAI before and after the assimilation. Analysed values are superimposed over the in-situ observations (points) and the model basic estimations (solid line) for comparison purposes.

wet years 2001 and 2002, the start of the growing season is still well predicted. The need for a non-stationary w_{wilt} is highlighted when Fig. 6 is compared to Fig. 7. In the latter, the vegetation biomass reaches a value of zero in those periods where w_2 is below w_{wilt} , which now have a longer duration, due to the lack of precipitation. The score for the biomass becomes negative and for the LAI is reduced from 0.48 (non-stationary w_{wilt}) to 0.14 (stationary w_{wilt}). In Fig. 7, the w_2 analyses for the summer period are overestimated in the years 2003 and 2004, because of a very low level of vegetation,

while better retrievals are obtained by re-setting w_{wilt} .

The fluxes of CO_2 and evapotranspiration are strongly reduced without precipitation, since the simulation of the vegetation biomass underestimates the real conditions. With this new scenario, the yearly cumulated NEE is negative for each year, i.e. the vegetation emits more CO_2 than with normal precipitation. The decrease in the fluxes is even stronger when using a constant w_{wilt} , with annual net assimilation of CO_2 reaching levels which are less than half of the values obtained with a complete atmospheric forcing (Table 4).

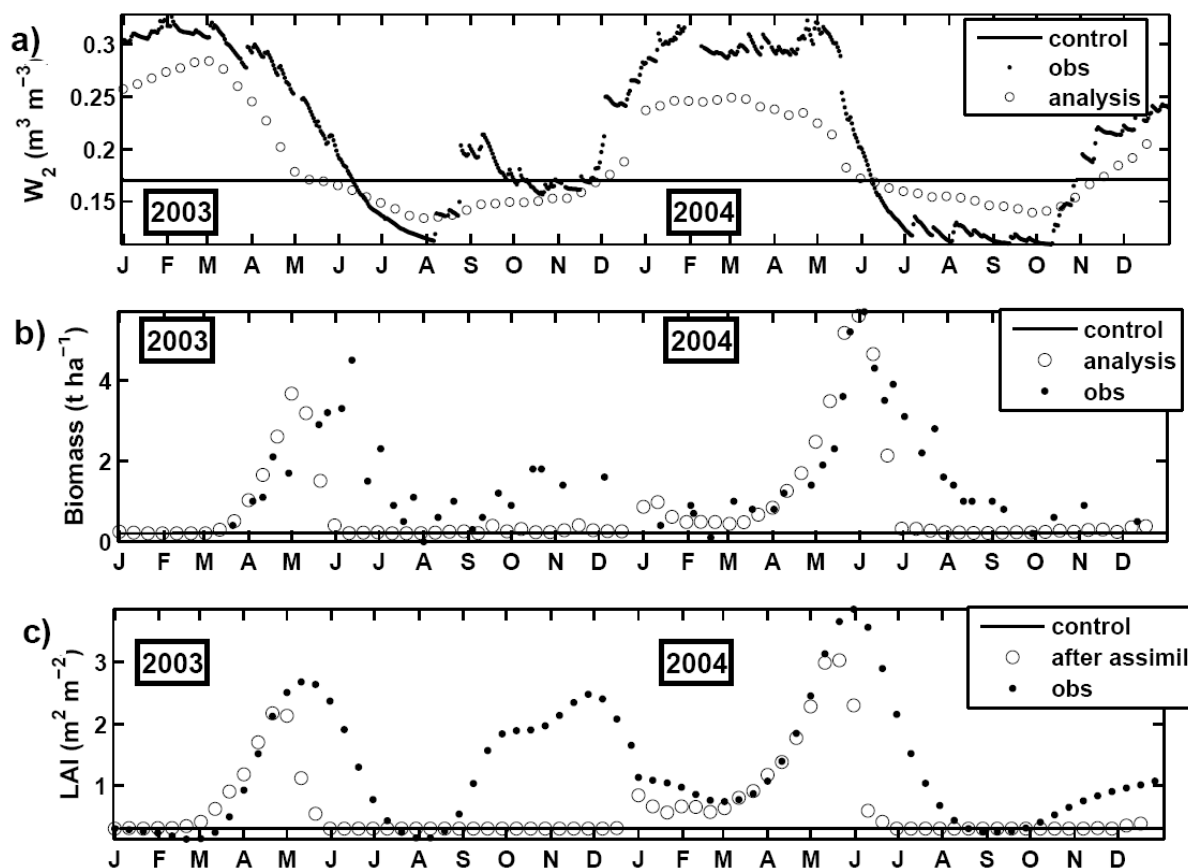


Figure 7.- Same as Fig.6, but using a fixed wilting point. The results for the years 2001 and 2002 are omitted as being similar as in Fig. 6.

4. Summary and Discussion

The present study is a continuation of the work initiated in MU07. There, a simplified 1D-VAR assimilation scheme was compared with other methods, and it was shown that it offered the best performance in view of a future implementation in a regional experiment. However, the LAI was forced and the vegetation biomass was not corrected for, as only w_g observations were assimilated. Nevertheless, vegetation biomass is also an important surface variable, vital to estimate the carbon and water vapour fluxes, and its analysis is also essential. In this study, field observations of w_g and LAI were assimilated at the same time into the ISBA-A- g_s LSM.

The key finding of the present study is that for the SMOSREX fallow site, the joint assimilation of w_g and LAI is able to improve w_2 and the vegetation biomass.

However, in order to generalize this conclusion for the application of the assimilation scheme to other types of land cover, further research has to be undertaken.

The implementation of the simplified assimilation scheme can be undertaken in different ways. In this study, w_2 and vegetation biomass were perturbed at the beginning of each assimilation window, in order to determine the linearized observation operator. This configuration implied 3 runs for each assimilation window (one for each perturbed variable and one for the control run). As an alternative, a four-step configuration can be used. First, a control run is obtained with the initial states of w_2 and vegetation biomass. Second, only the vegetation biomass within the 10-day assimilation window is analysed. Third, the updated state of the vegetation biomass is used to obtain a new reference simulation, and fourth, w_2 is analysed. However, this

procedure requires and additional run for each assimilation window, and therefore more computational cost. Moreover, the differences in the score between the three- and four-step method were minimal (not shown). Consequently, preference was given to the method with a reduced computational cost.

No flux validation data was available and the positive or negative impact of the assimilation on the fluxes could not be quantified. On a daily basis, the simulated CO_2 and evapotranspiration fluxes were not fundamentally changed by the assimilation. However, the new annual cycles of the fluxes were more consistent with the available observations of biomass and LAI, which showed that the LAI produced by the control simulation was underestimated during the start of the growing season and overestimated during the senescence phase.

The w_{wilt} limitation imposed by the single soil layer of the model used in this study tends to trigger a high vegetation mortality rate during marked droughts, and to underestimate the impact (on vegetation growth and on the water and carbon fluxes) of significant rainfall events after the droughts. A simple solution consists in resetting w_{wilt} dynamically during anomalously dry periods. The introduction of a non-stationary w_{wilt} was shown to improve the moisture and vegetation predictions as compared to an open-loop simulation with a stationary w_{wilt} . This led to an improvement of the score of the analysed w_2 from 0.79 to 0.86. Furthermore, it was slightly improved from 0.17 to 0.23 for the biomass analysis. The final value of 0.23 is rather low, partly because the high level of scatter in the observations found for 2001 and 2002 prevents the score of reaching higher values.

Solutions potentially suitable to overcome the w_{wilt} problem, include the attribution of more weight to the LAI observations by reducing the observational error when the initial w_2 is found below the prescribed w_{wilt} . However, this poses a problem when LAI observations are not reliable. The use of a multi-soil layer

scheme, which describes more accurately the evolution of the prognostic variables during dry periods than a single layer model with an averaged soil moisture content, may lead to a better prediction of soil moisture and LAI. In particular, the impact on the vegetation of rewetting the top soil layer is better accounted for by a multilayer model. However, the inclusion of several soil layers would considerably augment the computational burden of the assimilation. It is shown that when precipitation observations contain large errors, the joint assimilation of w_g and LAI is capable of restoring the annual trend of w_2 and vegetation biomass, but large errors may occur in the absolute value of these quantities. An attempt was made to deactivate the plant growth module of the model and to analyse w_2 only. In this case, the observed LAI was prescribed to the model instead of being assimilated. The soil moisture evolution was better predicted, because the root-plant water extraction and evapotranspiration rates were more realistic. In case of large precipitation errors, imposing LAI from external observations may be an advantage for the retrieval of soil moisture. However, as stated above, a large error in the LAI observations may prevent the assimilation scheme to lead to a significant improvement.

5. Conclusion

In the SMOSREX case study, the joint assimilation of w_g and LAI observations by using a simplified 1D-VAR variational scheme over a fallow, has demonstrated that:

- The temporal evolution of both w_2 and the vegetation biomass are improved by the assimilation. The assimilation scheme generally reduces the model underestimation of biomass during the growing phase and, especially, the overestimation of biomass during the senescence phase of the vegetation.

- The use of a non-stationary w_{wilt} is required for dry years during which soil moisture observations are found below the set w_{wilt} . This change limits the vegetation mortality during (and after) very dry conditions and the underestimation of the CO_2 and evapotranspiration fluxes following an inaccurate simulation of the biomass.
- In the presence of significant errors in the atmospheric forcing, the leaf onset and the seasonal dynamics

of the moisture are still retrieved well even, for example, if precipitation is set to zero. However, the error on the analysed variables (w_2 and the vegetation biomass) is large. In this case, the analysis of w_2 is more efficient if the vegetation biomass is not analysed, i.e. if good quality LAI observations are prescribed to the model, instead of simulating the vegetation growth.

References

- Balsamo, G., Bouyssel, F., Noilhan, J., 2004. A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Q. J. R. Met. Soc.*, **130A**, 895-915.
- Bicheron, P., Leroy, M., 1999. A method of biophysical parameter retrieval at global scale by inversion of a vegetation reflectance model. *Remote Sens. Environ.*, **67**, 251-266.
- Bounoua, G. J., Collartz, G. J., Los, S. O., Sellers, P. J., Dazlich, D. A., Tucker, C. J., Randall, D., 2000. Sensitivity of climate to changes in NDVI. *J. Clim.*, **13**, 2277-2292.
- Calvet, J.-C., Noilhan, J., Roujean, J.-L., Bessemoulin, P., Cabelguenne, M., Olioso, A., Wigneron, J.-P., 1998. An interactive vegetation SVAT model tested against data from six contrasting sites. *Agric. For. Meteorol.*, **92**, 73-95.
- Calvet, J.-C., and coauthors, 1999. MUREX: A land-surface field experiment to study the annual cycle of the energy and water budgets. *Ann. Geophys.*, **17**, 838-854.
- Calvet, J.-C., 2000. Investigating soil and atmospheric plant water stress using physiological and micrometeorological data. *Agric. For. Meteorol.*, **103**, 229-247.
- Calvet, J.-C., Soussana, J.F., 2001. Modelling CO_2 enrichment effects using an interactive vegetation SVAT scheme. *Agric. For. Meteorol.*, **108**, 129-152.
- Calvet, J.-C., Rivalland, V., Picon-Cochard, C., Guehl, J.-M., 2004. Modelling forest transpiration and CO_2 fluxes - response to soil moisture stress, *Agric. For. Meteorol.*, **124**(3-4), 143-156, doi: 10.1016/j.agrformet.2004.01.007.
- Cayrol, P., Chehbouni, A., Kergoat, L., Dedieu, G., Mordelet, P., Nouvellon, Y., 2000. Grassland modeling and monitoring with SPOT-4 VEGETATION instrument during the 1997-1999 SALSA experiment, *Agric. For. Meteorol.*, **105**, 91-115.
- Chase, T. N., Pielke, R. A., Kittel, T. G. F., Nemani, R., Running, S. W., 1996. The sensitivity of a general circulation model to global changes in leaf area index. *J. Geophys. Res.*, **101**, 7393-7408.
- Ciais, P., and coauthors, 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, **437**, 529-533.
- De Rosnay, P., Calvet, J.-C., Kerr, Y., Wigneron, J.-P., Lemaître, F., Escorihuela, M.J., Muñoz Sabater, J., Saleh, K., Barrie, J., Coret, L., Cherel, G., Dedieu G., Durbe, R., Fritz, N.E.D., Froissard, F., Kruszwski, A., Lavenu, F., Suquia, D., Waldteufel, P., 2006. SMOSREX: A long term field campaign experiment for soil moisture and land

- surface processes remote sensing. *Remote Sens. Environ.*, **102**, 377-389.
- Deardorff, J. W., 1977. A parameterization of ground-surface moisture content for use in atmospheric prediction models. *J. Appl. Meteor.*, **16**, 1182-1185.
- Deardorff, J. W., 1978. Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation. *J. Geophys. Res.*, **20**, 1889-1903.
- Entekhabi, D., Nakamura, H., Njoku, E.G., 1994. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Trans. Geos. Remot. Sens.*, **32**(2), 438-448.
- Gibelin, A.-L., Calvet, J.-C., Roujean, J.-L., Jarlan, L., Los S., 2006. Ability of the land surface model ISBA-A-g_s to simulate leaf area index at the global scale: comparison with satellites products, *J. Geophys. Res.*, **111**, D18102, doi:10.1029/2005JD006691.
- Henderson-Sellers, A., Yang, Z.-L., Dickinson, R. E., 1993. The Project for Intercomparison of Land-surface Parameterization Schemes. *Bull. Am. Meteorol. Soc.*, **74**, 1335-1349.
- Henderson-Sellers, A., Pitman, A. J., Love, P. K., Irannejad, P., Chen, T. H., 1995. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Am. Meteorol. Soc.*, **76**, 489-503.
- International GEWEX Project Office., 1995. Global Soil Wetness Project, version 1.0. 47 pp. [Available from International GEWEX Project Office, 1100 Wayne Avenue, Suite 1210, Silver Spring, MD 20910.].
- Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., Berger, M., 2001. Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geos. Remot. Sens.*, **39**, 1729-1735.
- Muñoz Sabater, J., Jarlan, L., Calvet, J.-C., Bouyssel, F., De Rosnay, P., 2007. From near surface to root zone soil moisture using different assimilation techniques. *J. Hydrometeor.*, **8**(2), 194-206.
- Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V., 2003. Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosc. Remote Sens.*, **41**(2), 215 – 229.
- Noilhan, J., Planton, S., 1989. A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536-549.
- Noilhan, J., Mahfouf, J.-F., 1996. The ISBA land surface parameterisation scheme. *Global Planet. Changes*, **13**, 145-159.
- Reichle, R. H., McLaughlin, D. B., Entekhabi, D., 2001. Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1708-1718.
- Reichle, R.H., Walker, J.P., Koster, R.D., and P.R. Houser, 2002. Extended versus Ensemble Kalman Filtering for land data assimilation. *J. Hydromet.*, **3**, 728-740.
- Roujean, J.-L., Lacaze, R., 2002. Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions: a pragmatic method and validation. *J. Geophys. Res.*, **107**, 4150.
- Segal, M., Arritt, R. W., Clark, C., Rabin, R., Brown, J., 1995. Scaling evaluation of the effect of surface characteristics on potential for deep convection over uniform terrain. *Mon. Weather Rev.*, **123**, 383-400.
- Seuffert, G., Gross, P., Simmer, C., Wood, E.F., 2002. The influence of hydrologic modelling on the predicted local weather: Two-way coupling of a mesoscale weather prediction model and a land surface hydrologic model. *J. Hydrometeor.*, **3**, 505-523.
- Shaw, B. L., Pielke, R. A., Ziegler, C. L., 1997. A three-dimensional numerical simulation of a Great Plains dryline. *Mon. Weather Rev.*, **125**, 1489-1506.

- Tian, Y., Woodcock, C. E., Wang, Y., Privette, J. L., Shabanov, N. V., Zhou, L., Buermann, W., Dong, J., Veikkanen, B., Hame, T., Ozdogan M., Knyazikhin Y., Myneni, R. B., 2002a. Multiscale Analysis and Validation of the MODIS LAI Product I. Uncertainty assessment. *Remote Sens. Environ.* **83**, 414-430.
- Tian, Y., Woodcock, C. E., Wang, Y., Privette, J. L., Shabanov, N. V., Zhou, L., Buermann, W., Dong, J., Veikkanen, B., Hame, T., Ozdogan M., Knyazikhin Y., and Myneni, R. B., 2002b. Multiscale Analysis and Validation of the MODIS LAI Product II. Sampling strategy. *Remote Sens. Environ.* **83**, 431-441.
- Tan, B., Hu, J., Huang, D., Yang, W., Zhang, P., Shabanov, N. V., Knyazikhin, Y., Nemani, R. R., Myneni, R. B., 2005. Assessment of the broadleaf crops leaf area index product from the Terra MODIS instrument, *Agric. For. Meteorol.*, **135**, 124-134.
- Wigneron, J.-P., Pellarin, T., Calvet, J. -C., de Rosnay, P., Saleh, K., Kerr, Y., 2005. L-MEB: A simple model at L-Band for the continental areas – application to the simulation of a half-degree resolution and global scale data set. *Radiative transfer models for microwave radiometry*, Matzler C., ed, Stevenage, UK: Institution of Electrical Engineers

5.5 Assimilation d'observations des instruments de télédétection

Les capteurs de télédétection satellitaires offrent l'avantage de mesurer des données sensibles à la variation spatio-temporelle des variables de surface à une échelle globale, ce qui est particulièrement important pour les zones où les mesures in-situ sont peu nombreuses ou inexistantes. De plus, des méthodes de désagrégation spatiale permettent d'utiliser les observations à faible résolution spatiale (comme SMOS par exemple) pour être exploitées dans des LSMs à maille plus fine. L'assimilation de données de télédétection dans un LSM peut être conduite de plusieurs manières. Premièrement, les données de télédétection peuvent être assimilées directement. Dans cet objectif, un modèle de transfert radiatif est couplé avec un schéma de surface pour convertir, avec une précision suffisante, des variables de surface en quantités directement observables par les capteurs des satellites. Deuxièmement, des algorithmes d'inversion ou des relations statistiques peuvent être utilisés pour obtenir des produits de télédétection corrélés aux variables pronostiques des LSM, et qui sont donc susceptibles d'être assimilés. Dans cet étude, cette dernière approche est employée pour convertir des observations de températures de brillance (T_B) en bande L et de réflectances sur jachère en humidités de surface et produits de LAI, respectivement.

5.5.1 Températures de brillance

5.5.1.1 Méthode directe

Le modèle d'émission en bande L, L-MEB (Pellarin *et al.*, 2003), permet de simuler les T_B à partir de l'humidité superficielle du sol et d'autres caractéristiques du sol et de la végétation. Ce modèle peut être couplé avec ISBA-A-gs pour assimiler directement les observations satellitaires et corriger l'humidité de la couche racinaire à partir des différences entre les observations de T_B (provenant dans notre cas du radiomètre LEWIS) et celles estimées par L-MEB couplé avec ISBA-A-gs. La fig. 5.14 montre, schématiquement, comment les estimations de T_B sur jachère peuvent être obtenues à partir du système couplé ISBA-A-gs / L-MEB. Comme entrées, L-MEB utilise les simulations de T_s , T_2 et w_g d'ISBA-A-gs, une estimation du LAI (provenant ou non du modèle), ainsi que la température de l'air et quelques paramètres constants du sol et de la végétation.

La fig. 5.15 présente l'évolution des T_B , à 40 degrés d'angle d'incidence, observés en 2003 sur la jachère de SMOSREX et simulés par L-MEB, et la corrélation entre les deux quantités. Les problèmes de la modélisation sur jachère apparaissent quand on examine la corrélation entre les deux températures (fig. 5.15, en bas). Dans les deux polarisations, le modèle sous-estime

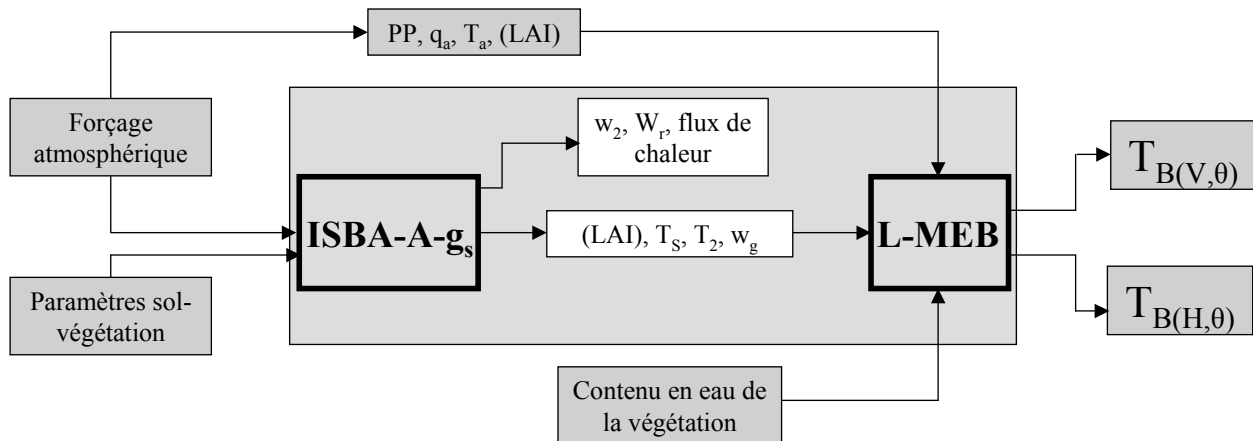


FIG. 5.14: Système couplé ISBA-A- g_s / L-MEB.

T_B presque systématiquement, et l'erreur augmente avec l'humidité du sol. Constatant que les paramètres et variables fournis à L-MEB ont des valeurs raisonnables, il est probable que des aspects mal pris en compte par la modélisation sont à l'origine de ce désaccord. D'autres études menées avec le même jeu de données (Saleh *et al.*, 2006a) montrent qu'il y a trois raisons principales à ce désaccord :

- Les effets de la rugosité du sol sont modulés par l'humidité du sol et cet effet n'est pas pris en compte dans la version actuelle de L-MEB (Pellarin *et al.* 2003). Même si cet effet est surtout sensible en conditions de sol nu, il contribue à dégrader les simulations sur la jachère ;
- L'interception de l'eau de pluie et de la rosée par les feuilles des plantes constituant la jachère (et formant une couche très dense), ainsi que par la litière, joue un rôle important sur l'émission micro-onde de la jachère ;
- Le fort effet masquant de la litière influence l'émission de la jachère en bande-L, et cet effet n'est pas correctement pris en compte dans la modélisation.

On peut conclure que l'interception de l'eau par la végétation et la litière couvrant le sol, ont un effet important sur l'émission de la jachère en bande L. Ceci affecte fortement la qualité des estimations de T_B fournies par L-MEB (version Pellarin *et al.*, 2003), et en particulier, leur dépendance à l'humidité du sol. Davantage de recherches est nécessaire pour une meilleure modélisation des processus qui influencent l'émission en bande L sur jachère, mais cela n'est pas l'objectif de la présente étude. En conséquence, une stratégie différente a été suivie pour l'assimilation des T_B dans ISBA-A- g_s .

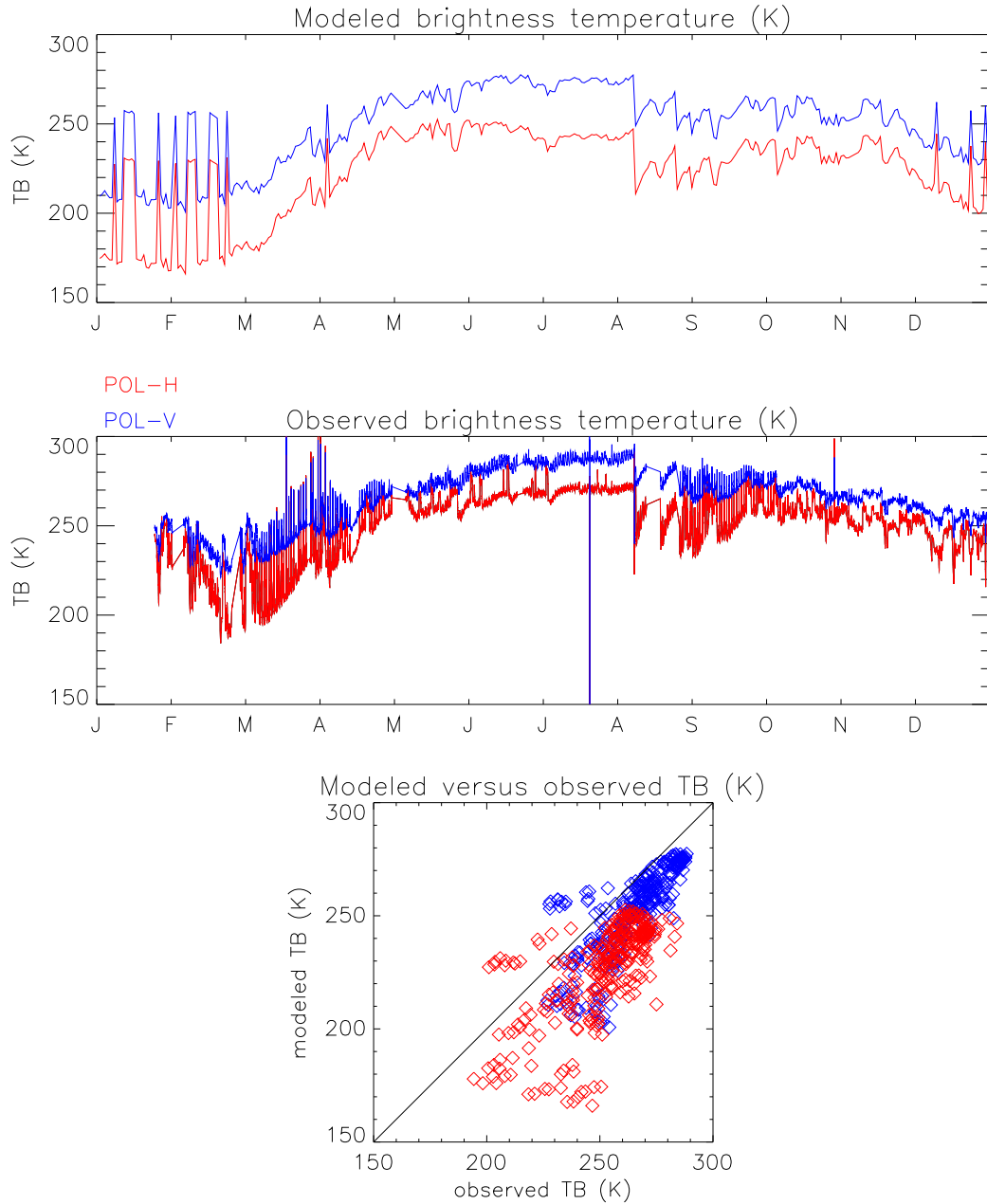


FIG. 5.15: Observations (radiomètre LEWIS) et simulations de L-MEB de la T_B sur la jachère de SMOSREX pour l'année 2003 et pour les polarisations horizontale (en bleu) et verticale (en rouge), à un angle d'incidence de 40° . En bas, les deux valeurs sont comparées et superposées à la ligne 1 : 1.

5.5.1.2 Inversion des T_B

On a vu que la paramétrisation de la litière reste encore une question ouverte et limite la performance du modèle d'émission L-MEB à l'échelle locale de la jachère de SMOSREX, alors que L-MEB a montré de bonnes performances sur d'autres types de couvert (Parde *et al.*, 2004, Wigneron *et al.*, 2004). L'approche suivie ici pour l'assimilation des observations de T_B dans ISBA-A- g_s est la suivante : des régressions statistiques qui relient entre elles les variables w_g et T_B sont utilisées. L'inconvénient de cette méthode est son manque de généralité puisque les conditions locales du site sont intégrées dans les relations statistiques, qui sont spécifiques au site.

Des relations statistiques empiriques ont été obtenues par Saleh *et al.* (2006b) à partir des paramètres du sol et de la végétation observés sur le site de SMOSREX. Dans cette thèse est utilisée une configuration utilisant les deux polarisations V et H de T_B , à un angle d'incidence de 50° :

$$w_g = \exp(A \cdot \log(1 - e_{50,H}) + B \cdot \log(1 - e_{50,V}) + C) \quad (5.1)$$

où $e_{50,H}$ et $e_{50,V}$ sont les émissivités effectives (il s'agit, pour simplifier, de T_B normalisé par la température du sol) en bande L de la jachère à un angle d'incidence de 50° , pour les polarisations horizontale et verticale. A, B et C sont des paramètres de calibration qui dépendent du type de sol, de la polarisation et de l'angle d'incidence. De plus, Saleh *et al.* (2006b) ont développé des seuils d'interception (définis comme une combinaison d'indices micro-ondes) pour filtrer les mesures influencées par l'interception de l'eau par la biomasse aérienne. Les deux indices dérivés des T_B utilisés pour définir les combinaisons de seuils (ou " flags ") sont les suivants :

$$PR_{50} = \frac{T_{B,(50,V)} - T_{B,(50,H)}}{T_{B,(50,V)} + T_{B,(50,H)}} \quad (5.2)$$

$$ADR_{V,20,50} = \frac{T_{B,(50,V)} - T_{B,(20,V)}}{T_{B,(50,V)} + T_{B,(20,V)}} \quad (5.3)$$

Nous avons utilisé trois flags, correspondant à ceux pour lesquels les coefficients A, B et C étaient disponibles au moment de notre étude. Le tableau 5.5 présente les définitions des trois flags ainsi que la valeur numérique des coefficients de l'équation 5.1. Ainsi, une inversion du modèle flaggé 0 signifie que toutes les mesures disponibles sont considérées dans l'inversion statistique, tandis que le flag 4 est le plus restrictif.

La fig. 5.16 montre les résultats de cette inversion en appliquant l'eq. 5.1. On observe que l'allure générale de l'évolution de w_g est bien reconstituée. Cependant, des défauts de la régression statistique apparaissent : les pics d'humidification du sol à la suite de précipitations importantes ne sont pas bien reproduits et, dans la série reconstituée, l'assèchement du sol en 2004 com-

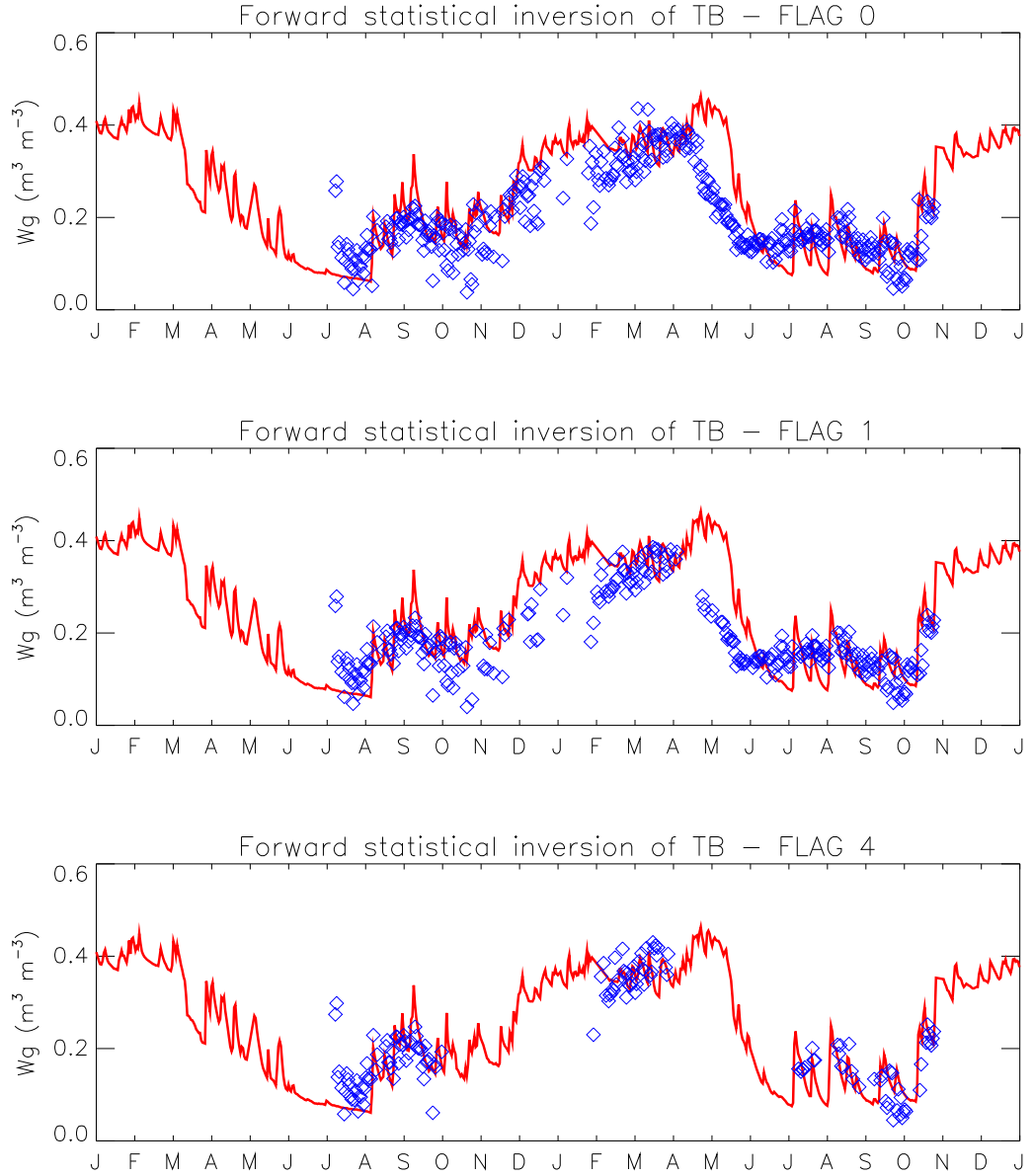


FIG. 5.16: Humidités de la surface de la jachère de SMOSREX en 2003 et 2004 obtenues en appliquant l'équation 5.1 aux observations de T_B de LEWIS (cercles bleus). Les données sont filtrées selon les conditions imposées pour la définition de plusieurs "flags". A ces estimations, les observations de w_g sur SMOSREX sont superposées (trait rouge).

TAB. 5.5: Définition de chaque "flag" appliquée à la série temporelle de T_B du radiomètre LEWIS. Les trois dernières colonnes correspondent à la valeur des coefficients de la relation empirique 5.1 pour chaque flag.

Flag	définition	A	B	C
0	toutes les données	-0.707	1.276	0.596
1	$PR_{50} > 0.031$	-0.557	1.202	0.673
4	$PR_{50} > 0.05$ and $ADR_{V,20,50} > 0$	-0.524	1.280	0.978

mence avec un mois d'avance par rapport aux observations in situ directes. L'erreur quadratique moyenne (RMSE), le biais moyen et le score (critère de Nash) du w_g reconstitué (en utilisant comme référence les moyennes journalières des observations directes de w_g par les sondes ThetaProbe) sont fournis dans le tableau 5.6. Comme attendu, les estimations sont plus précises en utilisant un flag plus restrictif, car l'effet de la litière et de l'interception de l'eau pour la végétation est davantage filtré, en particulier lors des périodes octobre-janvier 2003 et mai-juin 2004, où l'effet de la végétation sur l'émission micro-ondes semble être plus important. Cependant, le nombre d'estimations décroît rapidement, seules 214 estimations sont disponibles en appliquant le flag 4, à comparer aux 508 valeurs obtenues si on n'utilise aucun filtre.

TAB. 5.6: Statistiques correspondant à l'inversion des observations de T_B pour chaque flag employé. Les deux dernières colonnes correspondent au nombre d'estimations obtenues pour chaque flag et le pourcentage sur le nombre initial d'observations.

	RMSE	mb	E	Estim.	% Estim
flag 0	0.07	-0.02	0.64	508	69.4
flag 1	0.07	-0.01	0.67	391	53.4
flag 4	0.05	0.006	0.82	214	29.2

Les trois jeux d'observations de T_B ont été assimilés dans ISBA-A- g_s avec la méthode 1D-VAR simplifiée, en analysant seulement w_2 . Pour spécifier les erreurs d'observations, on a utilisé les RMSE correspondant aux inversions statistiques du tableau 5.6. Les résultats quantitatifs sont présentés dans le tableau 5.7. Bien qu'en utilisant le flag 0 l'effet de la litière soit présent et que certaines estimations puissent être biaisées, la performance globale des analyses en terme de scores est la même que si on utilise le flag 1, dont 32 observations sont filtrées. Ce résultat suggère que l'information contenue dans un plus grand nombre d'observations assimilées (129) est un avantage par rapport à un jeu d'observations de meilleure qualité, mais dont l'effectif

est réduit. Une forte décroissance du nombre d'observations conduit à une analyse plus proche des simulations sans assimilation(flag 4). En conséquence, pour cette étude, la meilleure solution pour analyser w_2 consiste à utiliser toutes les mesures de T_B disponibles. Ces observations seront utilisées pour être assimilées avec le 1D-VAR dans la section 5.5.4.

TAB. 5.7: Statistiques correspondant à l'assimilation avec la méthode 1D-VAR des produits de w_g obtenus à partir des observations de T_B .

	RMSE	mb	E	obs
flag 0	0.04	0.03	0.70	129
flag 1	0.04	0.03	0.70	97
flag 4	0.04	0.04	0.66	47
référence	0.05	0.04	0.58	173

5.5.2 Réflectances

De la même façon que pour les observations de T_B , les réflectances sur jachère ont été d'abord inversées pour obtenir une variable de la surface, le LAI, qui ensuite est assimilée dans ISBA-A-gs en utilisant la méthode 1D-VAR. Plusieurs méthodes existent pour relier directement les mesures de la réflectance de la végétation au LAI. Ici, l'algorithme de Roujean et Lacaze (2002) a été retenu, car c'est une méthode compétitive et efficace en temps de calcul permettant une implémentation opérationnelle. Cet algorithme est basé sur le lien existant entre la Fonction de Distribution de la Réflectance Bi-Directionnelle (BRDF) et plusieurs variables biophysiques de la surface par le biais de l'indice de végétation par différence (DVI). Dans une première étape, la BRDF de Roujean *et al.* (1992) est estimée pour toute bande spectrale :

$$\rho(\theta_S, \theta_V, \phi) = K_0 + K_1 \cdot f_1(\theta_S, \theta_V, \phi) + K_2 \cdot f_2(\theta_S, \theta_V, \phi) \quad (5.4)$$

où f_1 et f_2 sont deux fonctions angulaires prenant en compte les effets géométriques et de dispersion volumique, respectivement, θ_s l'angle solaire zénital, θ_V l'angle de vue et ϕ l'azimut relatif entre le soleil et les directions de vue. Le calcul de la BRDF est effectué en utilisant les équations du Filtre de Kalman et nécessite des mesures multiangulaires de la réflectance du sol ($\rho(\theta_S, \theta_V, \phi)$), lesquelles sont disponibles toutes les 15 minutes à SMOSREX grâce au réflectancemètre CIMEL. Cependant, en pratique, une mesure sur 4 a été utilisée (une par heure) de façon à limiter le coût numérique de l'inversion de la matrice de covariance des observations. Le résultat pour chaque journée est un jeu de coefficients angulaires K_i ($i=0,1,2$) pour chaque

bande spectrale, qui peuvent être reliés à plusieurs variables biophysiques, telle que le LAI ou la Fraction de Couverture Végétale (FVC). Roujean *et al.* (2002) ont proposé pour le LAI la relation suivante :

$$LAI = -\frac{\log(1 - FVC)}{b \cdot G \cdot \Omega} \quad (5.5)$$

où b est un paramètre qui prend en compte l'albédo des feuilles, G un facteur de projection des feuilles et Ω l'indice de groupage (ou "clumping index"). Le FVC est relié au coefficient K_0 par :

$$FVC = c_1 \cdot (K_{0,NIR} - K_{0,RED}) + c_2 \quad (5.6)$$

où c_1 et c_2 sont des coefficients empiriques qui varient selon différents seuils du DVI ($DVI = K_{0,NIR} - K_{0,RED}$). Ces coefficients ont été calculés pour les bandes spectrales de l'instrument POLDER (Roujean et Lacaze, 2002). Une normalisation spectrale est donc nécessaire pour adapter les coefficients de POLDER aux bandes spectrales du CIMEL (Samain, 2006). Etant donné que le calcul des coefficients K_i ($i=0,1,2$) de l'équation 5.4 est basé sur l'hypothèse que le rayonnement incident provient de la direction solaire, seules les périodes de ciel clair ont été considérées. Pour filtrer les mesures qui sont influencées par la présence de nuages, les observations d'éclairement ont été comparées aux courbes d'éclairement maximum mensuel (voir section 3.5). Si pour une journée toutes les observations de réflectance sont filtrées, le LAI est fixé à la valeur du jour précédant. La fig. 5.17 montre le LAI issu de l'application de l'algorithme décrit dans cette section au jeu d'observations de réflectance depuis juillet 2003. Pour toute la période d'étude, l'estimation de LAI est assez satisfaisante. Bien que le maximum de LAI estimé soit cohérent avec les observations, une avance d'environ un mois du maximum de LAI obtenu à partir des réflectances est observé. Il semble aussi que le LAI de l'hiver 2003-2004 soit surestimé de 0.5 à 1 $m^2 \cdot m^{-2}$ environ.

5.5.3 Erreur des pseudo-observations

Pour assimiler les variables de surface estimées à partir des données de télédétection, une erreur d'observation doit leur être attribuée. Pour les estimations de w_g obtenues en appliquant eq. 5.1 aux observations de T_B , la RMSE montrée au Tableau 5.6 ($0.07 m^3 \cdot m^{-3}$) a été assignée comme son erreur d'observation. Pour le LAI, une estimation de son erreur d'observation est obtenue en différentiant l'équation (5.5) :

$$\sigma_{LAI} = \frac{c_1 \cdot \sqrt{\sigma^2(K_{0,NIR}) + \sigma^2(K_{0,RED})}}{b \cdot \Omega \cdot G \cdot (1 - FVC)} \quad (5.7)$$

où les variances associées aux coefficients $K_{0,(NIR,RED)}$ dépendent du nombre d'observations journalières et du bruit de mesure et du modèle. En conséquence, une erreur qui varie avec

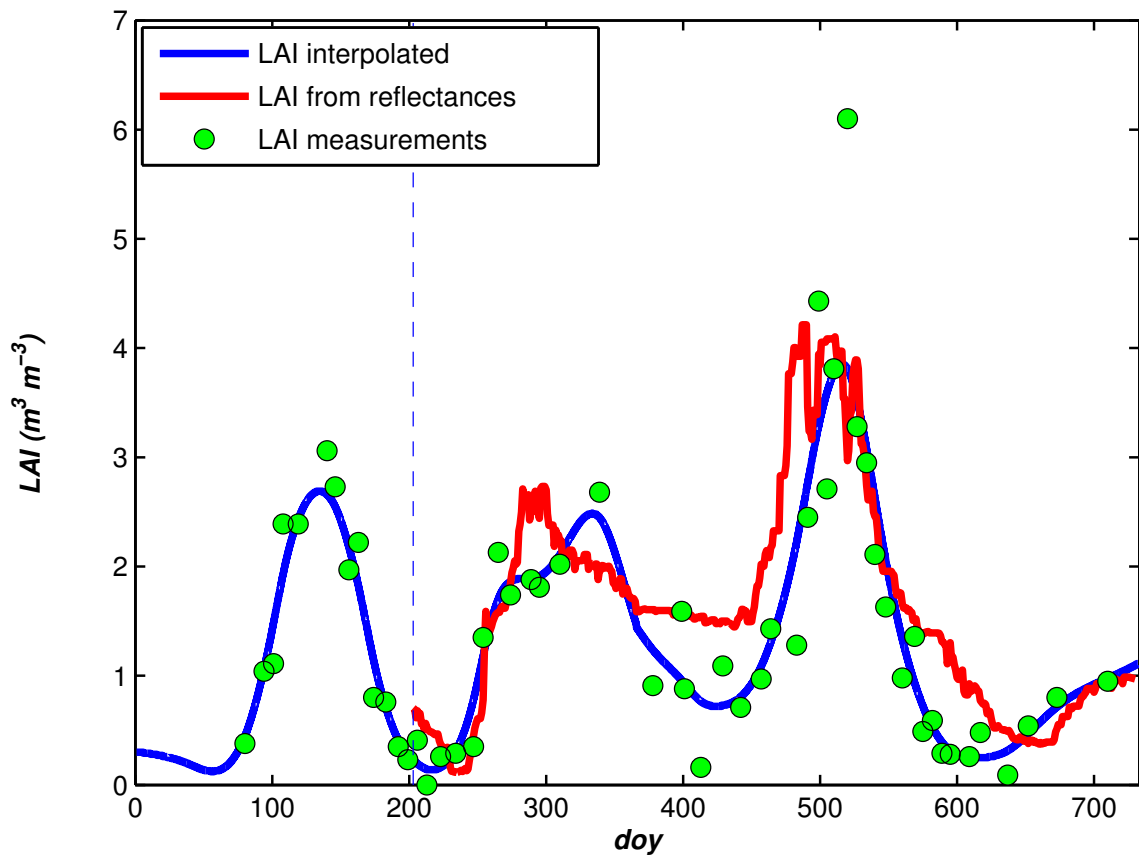


FIG. 5.17: Comparaison entre le LAI interpolé (ligne solide en bleu) à partir des observations sur le terrain (point verts), et le LAI issue de l'application de la méthode de Roujean et Lacaze (2002) au jeu d'observations de réflectances mesurées avec le reflectancemètre CIMEL (ligne solide rouge), en 2003 et 2004.

le temps est obtenue. Par rapport aux erreurs du modèle sur les variables analysées (w_2 et biomasse de la végétation), les mêmes études de sensibilité conduites à l'article de la section 5.4 sont valables ici, car le modèle de surface appliqué est le même. En conséquence, l'erreur de background attribuée à w_2 est $0.02 \text{ m}^3 \cdot \text{m}^{-3}$ et $0.05 \text{ kg} \cdot \text{m}^{-2}$ pour la biomasse de la végétation.

5.5.4 assimilation des produits de télédétection

Ici, on applique le jeu d'observations décrit précédemment à la méthode 1D-VAR. Les deux types d'observations sont : (1) le LAI dérivé de l'application de l'algorithme de Roujean et Lacaze (2002) (section 5.5.2) aux réflectances mesurées sur la jachère en utilisant une observation tous les dix jours et (2) les w_g obtenus par l'inversion statistique (eq. 5.1, section 5.5.1.2) appliquée à toutes les observations de T_B , en utilisant une observation tous les trois jours. Les deux produits de télédétection sont assimilés simultanément dans le schéma 1D-VAR en utilisant une fenêtre d'assimilation de 10 jours. Les mesures sont disponibles pour la période allant de juillet 2003 à décembre 2004. Dans la fig. 5.18 les variables analysées (w_2 et la biomasse de végétation) et le LAI estimé par le modèle après ré-initialisation sont présentés. De même que pour les figures de l'article présenté plus haut, les analyses, observations et simulations de référence sont superposées pour être comparées. Les analyses de w_2 sont, en général, comparables à celles de la fig.5.10 de l'article de la section 5.4. L'erreur plus importante attribuée aux observations de w_g issues des T_B empêche les analyses de w_2 d'améliorer la simulation de contrôle pendant l'été 2004. La principale différence par rapport à la fig. 5.10 de l'article est la surestimation de la biomasse pendant la phase de croissance de la végétation en 2004. Le résultat est relié à la valeur plus grande des valeurs de LAI assimilées et à leur erreur associée. Cette erreur n'est pas fixée, mais a été évaluée en appliquant l'eq. 5.7. En moyenne, des erreurs inférieures à $0.4 \text{ m}^2 \cdot \text{m}^{-2}$ sont associées aux valeurs de LAI inférieures à $2 \text{ m}^2 \cdot \text{m}^{-2}$, et l'erreur augmente avec la valeur de LAI. Dans une situation où la biomasse est faible, une erreur absolue de $0.05 \text{ kg} \cdot \text{m}^{-2}$ est relativement grande. En même temps, l'erreur absolue du LAI peut-être relativement petite, ce qui a pour conséquence de relacher plus rapidement l'assimilation vers les observations du LAI. En revanche, si la valeur de la biomasse est importante, son erreur relative devient plus faible et les incréments sur l'estimation du modèle seront moins importants, et en conséquence l'analyse de la biomasse sera principalement contrôlée par le modèle. Ce résultat suggère que de meilleures analyses pourraient être obtenues si au lieu d'utiliser des erreurs d'observation et du modèle fixes, une fonction d'erreur dépendante de la période de l'année était utilisée.

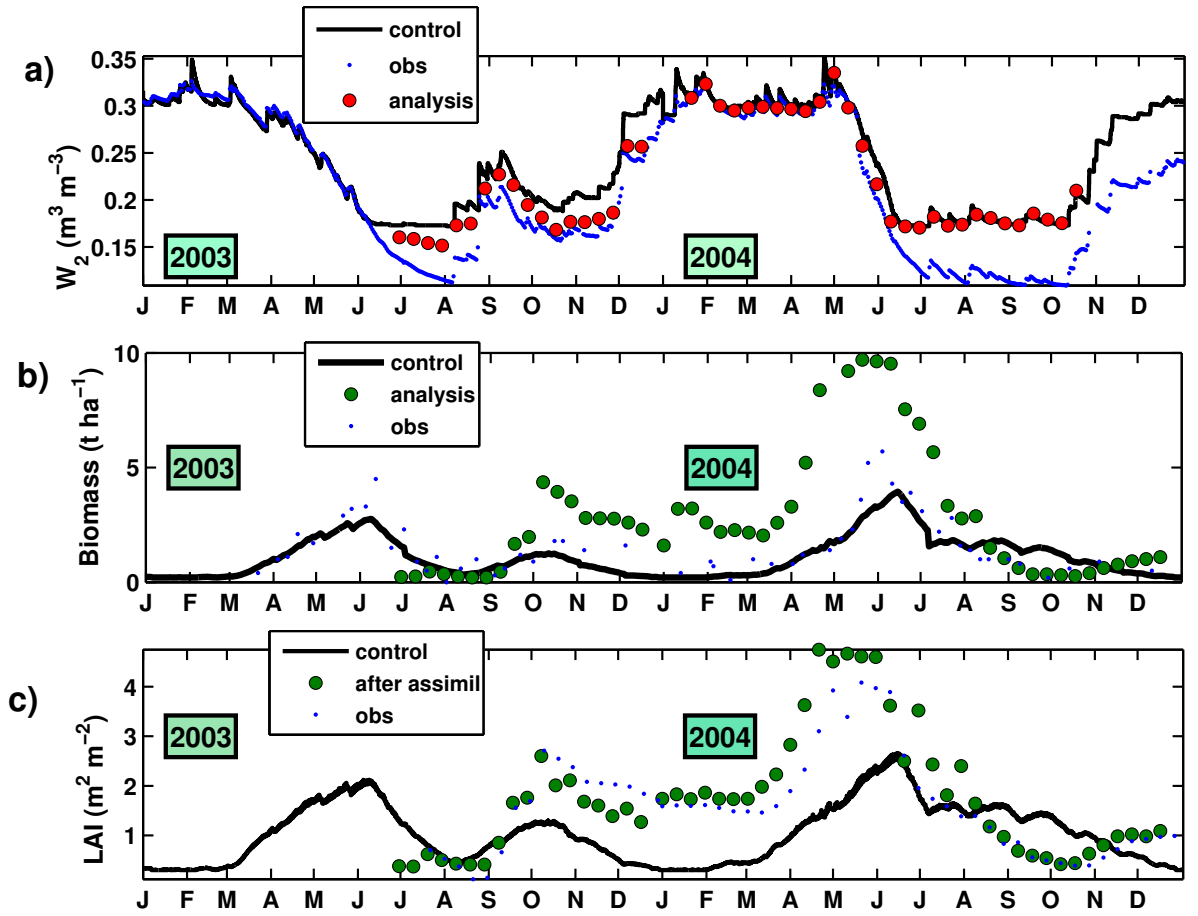


FIG. 5.18: Analyses de a) w_2 (cercles rouges) et b) biomasse de la végétation (cercles verts), en appliquant la méthode 1D-VAR simplifiée. Les observations sont w_g et LAI dérivés des observations de T_B et réflectance, respectivement. La période de mesure s'étend de juillet 2003 jusqu'à la fin de 2004. Les analyses sont superposées aux observations in-situ (points bleus) et à la simulation de contrôle (trait noir). c) Le LAI simulé après ré-initialisation de w_2 et la biomasse de la végétation (cercles verts) est superposé aux observations (point bleus) et à la simulation de contrôle (ligne solide noire).

5.5.5 Conclusion

Des observations réelles de télédétection ont été utilisées dans le schéma d'assimilation 1D-VAR (w_g et LAI dérivés des T_B et des réflectances, respectivement). Les observations de LAI obtenues à partir de l'estimation de la BRDF (eq. 5.4) sur jachère, montrent en général, pour SMOSREX, un bon accord avec les observations in-situ. Cependant, la méthode surestime le LAI pendant l'hiver 2003-2004. Cela peut provenir du fait que le nombre d'observations utilisées pour l'inversion de la BRDF pendant l'hiver 2003-2004 est inférieur aux autres périodes. Toutes les observations de T_B ont été utilisés, car pour SMOSREX l'utilisation de flags n'a pas montré avoir un impact positif sur les analyses de w_2 . Cette conclusion serait probablement différente

si les erreurs du modèle étaient importantes par rapport à l'incertitude des observations. Nos résultats sur les variables analysées suggèrent également que le modèle et les erreurs d'observation pourraient être formulés comme une fonction du temps au lieu d'utiliser des moyennes annuelles fixes, car le poids relatif attribué aux observations et aux variables pronostiques serait mieux contrôlé. Cependant, un des objectifs de cette étude était autant que possible d'utiliser toutes les variables d'observation pour estimer soit les erreurs d'observation soit les erreurs de modèle, au lieu de les ajuster artificiellement.

Chapitre 6

Conclusions

Sommaire

6.1	Conclusions	131
6.2	Perspectives	134

6.1 Conclusions

Le travail entrepris dans cette thèse porté sur l'assimilation des données de télédétection pour le suivi des variables des surfaces continentales, en particulier, l'état hydrique du sol et la biomasse de la végétation. Pendant une période de quatre ans, deux méthodes d'assimilation séquentielle (EKF et EnKF) et deux méthodes basées sur une approche variationnelle (1D-VAR simplifié et TVAR), ont été appliquées aux observations de l'humidité de la surface pour corriger les erreurs des estimations du réservoir d'eau du sol avec le modèle de surface ISBA-A-g_s. Le EKF, EnKF et 1D-VAR corrigent l'ébauche du modèle de surface à partir de la même équation optimale issue d'une théorie d'estimation linéaire.

Une quatrième méthode d'assimilation, le T-VAR, une méthode simple et démonstrative, a été aussi testée. Elle est utile pour évaluer les situations où aucune information sur l'incertitude des observations ou des variables modélisées n'est disponible. Cette méthode est moins performante que les autres car elle s'appuie sur une minimisation simple, non optimale (elle n'utilise pas la mémoire et les incertitudes des variables du système).

Le premier objectif de ce travail était de sélectionner une méthode d'assimilation, permettant d'améliorer les estimations de w_2 et de la biomasse de la végétation du modèle de surface ISBA-

A- g_s . Pour répondre à cette question, une étude d'intercomparaison entre les quatre méthodes d'assimilation citées précédemment a été conduite et appliquée à des observations de l'humidité de la couche superficielle du sol. Le but était d'analyser le contenu en eau total du sol. Cette étude permet de dégager plusieurs conclusions sur l'utilisation de ces méthodes :

- La méthode 1D-VAR simplifié offre le meilleur compromis entre performance des analyses de l'humidité de la zone racinaire et temps de calcul. En effet, seulement trois intégrations indépendantes d'ISBA-A- g_s sont nécessaires à chaque pas d'assimilation pour améliorer w_2 sur toute la période d'étude. Pour obtenir des performances proches de celles du 1D-VAR, une méthode d'EnKF devrait utiliser au moins un ensemble de 200 membres, demandant un temps de calcul important ;
- L'utilisation d'un schéma de surface, sur jachère, dont l'humidité de la couche racinaire est peu sensible aux perturbations initiales de l'humidité en surface pendant une période de trois jours, (c'est le cas d'un système basé sur les équations de force-restore comme ISBA-A- g_s), fait que les méthodes d'assimilation dont la correction de l'humidité de la zone racinaire dépend de cette sensibilité (EKF) soient peu performantes ;
- Ce caractère convergent de la méthode force-restore utilisée dans ISBA-A- g_s pour l'humidité du sol à sa valeur d'équilibre, fait que l'ensemble des trajectoires possibles de la méthode EnKF finit par s'effondrer à sa valeur d'équilibre. En conséquence, une inflation artificielle de la dispersion de l'ensemble est nécessaire.

Un autre objectif était de déterminer dans quelle mesure l'assimilation d'une seule variable (w_g ou LAI) suffit pour avoir une bonne estimation des deux variables à analyser (w_2 et biomasse). Plusieurs tests ont été conduits au chapitre 5 avec le modèle ISBA-A- g_s dans sa configuration LAI interactif. Les principales conclusions sont les suivantes :

- L'assimilation avec la méthode 1D-VAR simplifiée des seules observations de w_g améliore l'estimation de w_2 , en particulier durant les périodes de sécheresse de 2003 et 2004. Cependant, elle ne permet pas au modèle de reconstituer l'évolution saisonnière des variables de la végétation. L'analyse de l'état hydrique du sol, chaque trois jours, a un impact qui est plutôt négatif sur les estimations du LAI, avec des maxima annuels supérieurs aux observations ;
- L'assimilation des seules observations de LAI montre une amélioration globale de l'estimation de la biomasse de la végétation. Cependant, une dégradation de w_2 est observée sur plusieurs périodes ;

Ces résultats nous sont amené à proposer l'assimilation simultanée des observations de w_g et de LAI dans le but d'améliorer le suivi des deux variables à analyser (w_2 et biomasse). Les conclusions principales sont les suivantes :

- L'assimilation conjointe des observations de l'humidité de la surface et du LAI sur jachère avec le 1D-VAR simplifié, améliorent simultanément l'humidité du sol et la biomasse de la végétation par rapport à une simulation libre de ISBA-A-g_s ;
- Si de fortes incertitudes sur les variables de forçage atmosphérique sont présentes (en particulier, sur la précipitation), la méthode 1D-VAR permet de restituer en partie la dynamique du cycle annuel de l'humidité du sol et de la biomasse de la végétation.
- Avec un modèle de sol monocouche, l'introduction d'un " wilting point " dynamique en parallèle à l'assimilation est nécessaire pour préserver la cohérence entre les valeurs analysées de la biomasse et de l'état hydrique du sol.

Finalement, les mesures directement fournies par les instruments de télédétection de SMOS-REX (réflectance, température de brillance et température infrarouge de la surface de la jachère) ont été employées pour réaliser (1) le suivi temporel de l'état de la végétation et (2) effectuer un premier pas vers l'assimilation de produits issus de la télédétection. L'utilisation des réflectances dans différentes bandes spectrales d'un réflectancemètre a permis de construire des indices de végétation, relativement bien corrélés aux mesures in situ de biomasse et de LAI. L'utilisation des données de télédétection pour leur assimilation dans ISBA-A-g_s a montré que :

- l'assimilation de " heating rates " obtenus à partir d'observations de la température infrarouge de la surface est rendue difficile par une forte dépendance de ces observations à la vitesse du vent (qui n'est pas toujours connue avec précision dans un contexte opérationnel) ;
- l'effet de la litière semble être à l'origine de la difficulté d'utilisation du modèle L-MEB pour estimer les températures de brillance en bande L sur jachère. Les relations statistiques appliquées sur SMOSREX aux mesures des T_B ont montré un accord satisfaisant entre les observations et estimations des w_g . Néanmoins, l'utilisation de seuils pour filtrer l'effet de l'interception de l'eau n'apporte pas d'amélioration du suivi de l'état hydrique de la jachère, et il est préférable d'utiliser toutes les mesures de T_B dans l'assimilation ;
- Finalement, ces produits de télédétection ont été assimilés dans ISBA-A-g_s en appliquant la méthode 1D-VAR simplifié. L'assimilation des produits de w_g et LAI sur la période de juillet 2003 jusqu'à décembre 2004, n'amène qu'une faible amélioration de w_2 et provoque une surestimation de la biomasse par rapport à la simulation de contrôle. Il est sans doute nécessaire d'affiner la description des erreurs d'observation et du modèle dynamique,

notamment en fonction du temps afin d'améliorer le contrôle sur le poids attribué aux observations et au modèle.

6.2 Perspectives

Le dispositif expérimental mis en place à SMOSREX est unique. Les données de télédétection n'ayant été analysées que sur 18 mois dans cette étude il sera nécessaire de poursuivre l'analyse de l'assimilation des données de télédétection sur plusieurs cycles annuels.

A terme, la méthode identifiée dans cette étude (1D-VAR) devra être mise en œuvre sur la France ou l'Europe. Il est envisagé d'appliquer ce système, dans un premier temps, au Sud-Ouest de la France pour la période 1999-2005. Les observations de LAI viendront des capteurs SPOT/VEGETATION ou MODIS, et seront complétées avec les observations de l'humidité du sol provenant du satellite SMOS (à partir de 2008) et/ou du diffusiomètre vent ASCAT (sur METOP). La validation croisée des variables de la végétation et de l'état hydrique du sol sur le Sud-Ouest de la France sera possible grâce aux observations d'humidité du sol du réseau SMOSMANIA (" Soil Moisture Observing System - Meteorological Automatic Network Integrated Application ") du CNRM, combinées avec les mesures de biomasse et de flux réalisées sur les sites CarboEurope (INRA et CESBIO). Les données SMOSMANIA permettront de valider les estimations de l'état hydrique du sol provenant du satellite SMOS et d'ASCAT et de valider les sorties opérationnelles du modèle hydrologique SIM de Météo-France. Etant donné que les données de SMOSMANIA sont locales, la modélisation sera utilisée pour réaliser l'upscaling des observations pour une utilisation à l'échelle régionale. Enfin, ces données in situ (éventuellement associées à des mesures aéroportées - projet CAROLS piloté par le CETP) permettront de valider l'assimilation des données SMOS dans ISBA-A-gs, sur le Sud-Ouest de la France.

Bibliographie

Balsamo, G., F. Bouyssel, and J. Noilhan, 2004 : A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Q.J.R.Meteorological Society*, **130**, 895–915.

Baret, F., C. Bacour, M. Weiss, K. Pavageau, D. Béal, V. Bruniquel, P. Regner, J. Moreno, C. Gonzalez, , and J. Chen, 2004 : Canopy biophysical variables estimation from meris observations based on neural networks and radiative transfer modelling : principles and validation. *Q.J.R.Meteorological Society*, **130**, 895–915.

Bouttier, F. and P. Courtier, 1999 : Data assimilation concepts and methods. Technical report, European Centre for Medium-Range Weather Forecasts.

Bratseth, A., 1986 : Statistical interpolation by means of successive corrections. *Tellus*, **38**(A), 439–447.

Brown, R. and P. Hwang, 1997 : *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, New York, NY, 3rd ed.

Burgers, G., P. V. Leeuwen, and G. Evensen, 1998 : Analyses scheme in the ensemble kalman filter. *Monthly Weather Review*, **126**, 1719–1724.

Calvet, J.-C., 2000a : Investigating soil and atmospheric plant water stress using physiological and micrometeorological data. *Agricultural and Forest Meteorology*, **103**, 229–247.

Calvet, J.-C., P. Bessemoulin, J. Noilhan, C. Berne, I. Braud, D. Courault, N. Fritz, E. Gonzalez-Sosa, J. P. Goutorbe, R. Haverkamp, G. Jaubert, L. Kergoat, G. Lachaud, J. P. Laurent, P. Mordelet, A. Oliso, P. Péris, J. L. Roujean, J.-L. Thony, C. Tosca, M. Vauclin, and D. Vignes, 1999 : Murex : a land-surface field experiment to study the annual cycle of the energy and water budgets. *Annales Geophysicae.*, **17**, 838–854.

Calvet, J.-C. and J. Noilhan, 2000b : From near-surface to root-zone soil moisture using year-round data. *Journal of Hydrometeorology*, **1**(5), 393–411.

- Calvet, J.-C., J. Noilhan, and P. Bessemoulin, 1998a : Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates : A feasibility study based on field measurements. *Journal of Applied Meteorology*, **37**(4), 371–386.
- Calvet, J.-C., J. Noilhan, J. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J. Wigneron, 1998b : An interactive vegetation SVAT model tested against data from six contrasting sites. *Agricultural and Forest Meteorology*, **92**, 73–95.
- Calvet, J.-C., V. Rivalland, C. Picon-Cochard, and J.-M. Guehl, 2004 : Modelling forest transpiration and CO₂ fluxes - response to soil moisture stress. *Agricultural and Forest Meteorology*, **124**, 143–156.
- Cayrol, P., A. Chehbouni, L. Kergoat, G. Dedieu, P. Mordelet, and Y. Nouvellon, 2000 : Grassland modeling and monitoring with spot-4 vegetation instrument during the 1997-1999 salsa experiment. *Agricultural and Forest Meteorology*, **105**, 91–115.
- Cohn, S., 1997 : An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, **75**(1B).
- Deardorff, J., 1977 : A parameterization of ground-surface moisture content for use in atmospheric prediction models. *Journal of Applied Meteorology*, **16**, 1182, 1185.
- Deardorff, J., 1978 : Efficient prediction of ground temperature and moisture with inclusion of a layer of vegetation. *Journal of Geophysical Research*, **83**, 1889–1903.
- Deering, D., 1978 : *Rangeland reflectance characteristics measured by aircraft and spacecraft sensors*. PhD thesis, Texas A&M University, 338 pp.
- DeRosnay, P., J. Calvet, Y. Kerr, J. Wigneron, F. Lemaître, M. Escorihuela, J. M. Sabater, and et al., 2006 : SMOSREX : A long term field campaign experiment for soil moisture and land surface processes remote sensing. *Remote Sensing of Environment*, **102**, 377–389.
- Desroziers, G., 2001 : Optimal estimation in meteorology. Technical report, ALATNET seminar.
- Entekhabi, D., H. Nakamura, and E. Njoku, 1995 : *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*. VPS publisher, 485-498.
- Evensen, G., 1994 : Sequential data assimilation with a non linear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10143–10162.
- Evensen, G., 2003 : The ensemble kalman filter : theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Francois, C., A. Quesney, and C. Ottlé, 2003 : Sequential assimilation of ERS-1 and ERS-2 data into a coupled land surface-hydrological model using an extended kalman filter. *Journal of Hydrometeorology*, **4**, 473–486.

Gelb, A., 1974 : *Applied Optimal Estimation*. Analytic Science Corporation, 382 pp.

Gibelin, A.-L., J.-C. Calvet, J.-L. Roujean, L. Jarlan, and S. Los, 2006 : Ability of the land surface model ISBA-A-g_s to simulate leaf area index at the global scale : Comparison with satellites products. *Journal of Geophysical Research*, **111**(D18102, doi :10.1029/2005JD006691).

Isaacson, E. and H. B. Keller, 1994 : *Analysis of numerical methods*. Courier Dover Publications.

Jacobs, C., B. Van-Den-Hurk, and H. de Bruin, 1996 : Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions. *Agricultural and Forest Meteorology*, **80**, 111–134.

Jarvis, P., 1976 : The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London*, **273**(927), 593–610.

Jordan, C., 1969 : Derivation of leaf area index from quality of light on the forest floor. *Ecology*, **50**, 663–666.

Journel, A., 1977 : Kriging in terms of projections. *Mathematical Geology*, **9**, 563–586.

Julier, S. and J. Uhlmann, 2004 : Unscented filtering and nonlinear estimation. In *IEEE Review* **92**(3), 401–422.

Kerr, Y., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger, 2001 : Soil moisture retrieval from space : The soil moisture and ocean salinity (SMOS) mission. *IEEE Trans. Geosc. Remote Sens.*, **39**(8), 1729–1735.

Lemaitre, F., J. Poussière, Y. Kerr, M. Dejus, R. Durbe, and P. D. R. et al., 2004 : Design and test of the ground based L-band radiometer for estimating water in soils (LEWIS). *IEEE Trans. Geosc. Remote Sens.*, **42**(8), 1666–1676.

Lewis, F., 1986 : *Optimal Estimation, with an Introduction to Stochastic Control Theory*. John Wiley & Sons, 400 pp.

Lorenc, A., 1986 : Analysis methods for numerical weather prediction. *Q.J.R.Meteorological Society*, **112**, 1177–1194.

Mahfouf, J.-F., 1991 : Analysis of soil moisture from near-surface parameters : a feasibility study. *Journal of Applied Meteorology*, **30**, 1534–1547.

- McNider, R., A. Song, D. Casey, P. Wetzel, W. Crosson, and R. Rabin, 1994 : Toward a dynamic-thermodynamic assimilation of satellite surface temperature in numerical atmospheric models. *Monthly Weather Review*, **122**, 2784–2803.
- Nemani, R. and S. Running, 1989 : Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied Meteorology*, **28**, 276–284.
- Ni-Meister, W., P. Houser, and J. Walker, 2006 : Soil moisture initialization for climate prediction : Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model. *Journal of Geophysical Research*, **111**(D20102), doi :10.1029/2006JD007190.
- Noilhan, J. and S. Planton, 1989 : A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, **117**, 536–549.
- Ostdiek, V. and D. Bord, 2005 : *Inquiry into physics*. West Publishing Company.
- Pardé, M., J.-P. Wigneron, P. Waldteufel, Y. Kerr, Y. Chanzy, and S. Sobjaerg, 2004 : N-parameter retrievals from l-band microwave observations acquired over a variety of crop fields. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(6), 1168–1178.
- Pellarin, T., J.-P. Wigneron, J.-C. Calvet, and P. Waldteufel, 2003 : Global soil moisture retrieval from a synthetic l-band brightness temperature data set. *Journal of Geophysical Research*, **108**(D12), ACL :9.1–9.15.
- Reed, B., J. Brown, D. V. der Zee, T. Loveland, and J. Merchant, 1994 : Vegetation effects on the microwave emission of soils. *Journal of Vegetation Science*, **5**, 703–714.
- Reichle, R., D. McLaughlin, and D. Entekhabi, 2001 : Variational data assimilation of microwave radiobrightness observations for land surface hydrologic applications. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(8), 1708–1718.
- Roujean, J. and R. Lacaze, 2002 : Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions : A pragmatic method and its validation. *Journal of Geophysical Research*, **107**(D12), 4150, doi :10.1029/2001JD000751.
- Roujean, J., M. Leroy, and P. Deschamps, 1992 : A bidirectional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, **97**, 20455–20468.
- Sabater, J. M., 2007 : *Assimilation de données de télédétection pour le suivi des surfaces continentales : mise en œuvre sur un site expérimental*. PhD thesis, Université Toulouse III, 163 pp.

-
- Saleh, K., J.-P. Wigneron, P. D. Rosnay, J.-C. Calvet, M. J. Escorihuela, Y. Kerr, and P. Waldteufel, 2006a : Impact of rain interception by vegetation and mulch on the L-band emission of natural grass. *Remote Sensing of Environment*, **101**(1), 127–139.
- Saleh, K., J.-P. Wigneron, P. D. Rosnay, J.-C. Calvet, and Y. Kerr, 2006b : Semi-empirical regressions at L-band applied to surface soil moisture retrievals over grass. *Remote Sensing of Environment*, **101**(3), 415–426.
- Samain, O., 2006 : *Fusion multi-capteurs de données satellitaires optiques pour la restitution de variables biophysiques de surface*. PhD thesis, Université Toulouse III, 190 pp.
- Schmugge, T., 1983 : Remote sensing of soil moisture : Recent advances. *IEEE Trans. Geosc. Remote Sens.*, **GE21**, 334–344.
- Staley, D. and G. Jurica, 1972 : Effective atmospheric emissivity under clear skies. *Journal of Applied Meteorology*, **11**, 349–356.
- Talagrand, O., 1997 : Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, **75**(1B), 191–209.
- Teunissen, P., 1999 : *Adjustment Theory*. Delft University Press, 193 pp.
- Thiessen, A., 1911 : Precipitation averages for large areas. *Monthly Weather Review*, **39**, 1082–1089.
- Tian, Y., C. Woodcock, Y. Wang, J. Privette, N. Shabanov, L. Zhou, W. Buermann, J. Dong, B. Veikkanen, T. Hame, M. Ozdogan, Y. Knyazikhin, and R. Myneni, 2002a : Multiscale analysis and validation of the MODIS LAI product I. Uncertainty assessment. *Remote Sensing of Environment*, **83**, 414–430.
- Tian, Y., C. Woodcock, Y. Wang, J. Privette, N. Shabanov, L. Zhou, W. Buermann, J. Dong, B. Veikkanen, T. Hame, M. Ozdogan, Y. Knyazikhin, and R. Myneni, 2002b : Multiscale analysis and validation of the MODIS LAI product II. Sampling strategy. *Remote Sensing of Environment*, **83**, 431–441.
- Tobler, W., 1979 : Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association*, **74**(367), 519–536.
- Van Den Hurk, B. and H. The, 2002 : Assimilation of satellite derived surface heating rates in a numerical weather prediction model. Technical report, KNMI, The Netherlands.
- Wagner, W., K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf, 2003 : Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *Journal of Geophysical Research*, **108**, 4611–4626.

Walker, J., 1999 : *Estimating soil moisture profile dynamics from near surface soil moisture measurements and standard meteorological data*. PhD thesis, Dept. of Civil, Surveying and Environmental Engineering, Univ. of Newcastle, Callaghan, Australia.

Walker, J. and P. Houser, 2001 : A methodology for initializing soil moisture in a global climate model : Assimilation of near-surface soil moisture. *Journal of Geophysical Research*, **106**(D11), 11761–11774.

Wetzel, P., D. Atlas, and R. Woodward, 1984 : Determining soil moisture from geosynchronous satellite infrared data : A feasibility study. *Climate Applied Meteorology*, **108**(23), 375–391.

Wigneron, J.-P., M. Pardé, P. Waldteufel, A. Chanzy, Y. Kerr, and S. Schmidl, 2004 : Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(2), 416–425.

Annexe A

Forçage Atmosphérique 2001-2004

Sur SMOSREX un forçage atmosphérique est créé grâce aux données enregistrées par une station météorologique située sur la jachère (fig. 3.1). La fréquence temporelle des mesures est de 30 minutes. Les mesures automatiques sont stockées dans deux stations que l'on appelle Campbell-1 et Campbell-2. Elles enregistrent l'information suivante :

- Campbell-I
 - Humidité de la couche superficielle, fournie par quatre Theta-sondes (mV) ;
 - Humidité de la zone racinaire à 10, 20, 30, 40 et 50 cm de profondeur fournie par 3 Theta-sondes et, l'humidité à 60, 70, 80 et 90 cm de profondeur fournie par 2 sondes (mV) ;
 - Résistivité de la rosée (Ωm) ;
 - Fraction du temps pendant lequel la rosée est présente (s) ;
 - Vitesse du vent à 2 m d'hauteur (m/s) ;
 - Humidité de l'air à 2 m d'hauteur (%) ;
 - Hauteur de précipitation (mm).
- Campbell-II
 - Information sur la date des mesures ;
 - Rayonnement total descendant (W/m^2) ;
 - Rayonnement total montant (W/m^2) ;
 - Température du corps noir (K) ;
 - Rayonnement global descendant (W/m^2) ;
 - Rayonnement global montant (W/m^2) ;
 - Température de la couche de surface ($^{\circ}C$) ;
 - Température du sol à 5, 20, 50 et 90 cm ($^{\circ}C$) ;

- Température de l'air à 2 m d'hauteur ($^{\circ}C$);
- Température dans le boîtier de la centrale ($^{\circ}C$);

Les mesures manuelles de végétation (biomasse sèche, biomasse frais, LAI, etc.) sont rassemblées dans un fichier construit par le CESBIO (Centre d'Études Spatiales de la BIOSphère) de Toulouse.

Avec l'ensemble de ces mesures on prétend construire un forçage atmosphérique qui contient les variables suivants :

1. Rayonnement solaire incident (W/m^2);
2. Rayonnement thermique incident (W/m^2);
3. Précipitation (mm/s);
4. Température de l'air (K);
5. Vitesse de l'air (m/s);
6. Pression atmosphérique (Pa);
7. Humidité spécifique de l'air ($\frac{kg \text{ de } H_2O}{kg \text{ d'air}}$);
8. Température du sol à 50 cm ($^{\circ}C$);
9. hauteur de la couverture végétal (m);
10. Indice foliaire de la végétation ($m^2 \cdot m^{-2}$).

À ces mesures est associé un indice de qualité journalière. Il prend en compte le nombre de variables manquantes, de façon que lorsque cet indice de qualité est égal à zéro ça veut dire que toutes les variables du forçage sont ce jour là disponibles.

Les stations de mesure Campbell-I et II ne fournissent pas directement la valeur directe de quelque variable du forçage. Il faut, donc, au préalable quelques manipulations. Ainsi, aux mesures brutes du rayonnement solaire incident et montant il faut l'ajouter la contribution du corps noir selon :

$$R_T^{\downarrow} = R_T'^{\downarrow} + \sigma T_{CN}^4 \quad (A.1)$$

$$R_T^{\uparrow} = R_T'^{\uparrow} + \sigma T_{CN}^4 \quad (A.2)$$

ou $\sigma = 5.67 \cdot 10^{-8} W \cdot m^{-2} K^4$ est la constante de Steffan-Boltzmann et $R_T'^{\downarrow}$ et $R_T'^{\uparrow}$ les mesures brutes des pyranomètres.

L'humidité spécifique de l'air (exprimée en kg de H_2O par kg d'air humide) est estimée à partir de l'équation suivant :

$$q_h = \frac{1}{(1 + RM)} \quad (A.3)$$

avec

$$RM = \frac{0.622}{\frac{P}{\text{tvap}} - 1}$$

et

$$\text{tvap} = 100 \cdot 6.107 \cdot \left(1 + \sqrt{2} \cdot \sin \pi \cdot \frac{T}{540}\right)^8 \cdot 827 \cdot \frac{H}{100}$$

T est la température de l'air mesurée en degrés centigrades, P la pression atmosphérique en Pascals et H l'humidité relative de l'air en %.

Le rayonnement incident de grande longueur d'onde est calculé comme la différence entre le total et le solaire ($R_G^\downarrow$) :

$$R_A^\downarrow = R_T^\downarrow - R_G^\downarrow \quad (\text{A.4})$$

Annexe B

Variables manquants du forçage atmosphérique

Une condition nécessaire pour faire tourner les modèles de surface est de disposer un forçage continu dans le temps. Dans la réalité un instrument peut tomber en panne, être décalibré ou encore le dispositif d'acquisition des données peut être défaillant. Il nous faut dans ce cas compléter les trous d'information. Le tableau B.1 est un exemple d'information relative à l'année 2003. Il montre comment les données manquantes sont remplacées par des données venant de stations secondaires. La première colonne montre les variables du forçage atmosphérique. Les colonnes suivantes sont cochées lorsqu'on utilise l'information de la station de mesure qui est en tête de la colonne correspondante : nous disposons de deux principales, Campbell-I et Campbell-II, une station secondaire que se trouve à quelque mètres de distance de notre station principale, appelée "hydromet", et une quatrième appelé *LHERM*. Cette dernière est située dans un autre emplacement que l'on peut considérer aussi représentatif que notre site expérimental au cas où l'on ne dispose pas de mesures disponibles sur le site. Finalement, la dernière colonne fait référence au fichier où les mesures de biomasse sont stockées. Par exemple, pour la radiation solaire incidente on a utilisé toutes les stations disponibles pour les mesures manquantes (pour 2003 ils sont de 3.65% pour la station Campbell-I, 8.71% pour Campbell-II et 2.36% pour le rayonnement). La proximité des stations secondaires à SMOSREX nous assure un remplissage de trous d'information de façon qu'elles soient encore représentatives du site expérimental. Une étude de comparaison (pas montrée ici) montrent, qu'en moyenne, il y a un bon accord entre les variables mesurées dans les différentes stations.

Lorsque des données de $R_A^{\downarrow}$ sont absentes on a choisi d'utiliser l'estimation fournie par la

	Campbell-1	Campbell-2	hydromet	LHERM	Biom03
Solar incident radiation (RGD)	-	$\sqrt{(1)}$	$\sqrt{(3)}$	$\sqrt{(2)}$	-
Thermal incident radiation (RA)	$\sqrt{(1)}$	$\sqrt{(1)}$	-	$\sqrt{(1)}$	-
Precipitation (PP)	$\sqrt{(1)}$	-	$\sqrt{(3)}$	$\sqrt{(2)}$	-
Air temperature (T)	-	$\sqrt{(1)}$	-	$\sqrt{(1)}$	-
Wind speed (U)	$\sqrt{(1)}$	-	-	$\sqrt{(1)}$	-
Atmospheric pressure (PA)	-	-	-	-	-
Air Specific humidity (H)	$\sqrt{(1)}$	-	-	$\sqrt{(1)}$	-
Soil Temperature at 50 cm (T50)	-	$\sqrt{(1)}$	-	-	-
Canopy height (higveg)	-	-	-	-	$\sqrt{(1)}$
Green Leaf Area Index (LAIveg)	-	-	-	-	$\sqrt{(1)}$
Deep Water Inflow (excm)	-	-	-	-	-

(1) : using data from the whole year

(2) : from 01-01-2003 until 02-09-2003

(3) : from 02-09-2003 until 31-12-2003

FIG. B.1: Création du forçage atmosphérique pour l'année 2003.

relation B.1 pour les reconstituer (1972) :

$$R_A = 0.67 \cdot ((1670 \cdot q_h)^{0.08}) \cdot (\sigma \cdot (T + 273.15)^4) \quad (\text{B.1})$$

Annexe C

Conversion des unités de l'humidité du sol

L'humidité du sol est mesurée dans SMOSREX avec des dispositifs Theta-Probes. Ils sont sensibles au contenu volumétrique de l'humidité du sol (rapport entre le volume d'eau présent au sol et le volume total de l'échantillon du sol) . Ils font les mesures grâce à un ensemble d'électrodes d'acier inoxydable à l'extrémité de la sonde implantée directement dans le sol. Par conséquent, le contenu en eau du sol va être mesuré sous forme de tension électrique V (comme le rapport entre une énergie émise et une énergie réfléchie). Des relations empiriques, fonction de la profondeur de la mesure dans le sol, vont permettre de convertir cette tension en contenu volumétrique de l'eau. Pour SMOSREX, ces relations calculées au CESBIO sont les suivantes :

- Humidités de la couche plus superficielle : $w_g(\%) = \frac{4.7 \cdot V^3 - 6.4 \cdot V^2 + 6.6 \cdot V - 0.23}{7.7}$;
- Humidité jusqu'à 5 centimètres d'épaisseur du sol : $w_g(\%) = 0.57472 \cdot V - 0.062877$;
- Jusqu'à 10 cm : $w_g(\%) = 0.56659 \cdot V - 0.088614$;
- jusqu'à 20 cm : $w_g(\%) = 0.50037 \cdot V - 0.09782$;
- 30 cm et plus : $w_g(\%) = -0.58376 \cdot V - 0.17406$.

Annexe D

Correction des éclairements et luminances sur SMOSREX

- **Irradiance (I)** : Le radiomètre mesurant le rayonnement incident doit suivre une loi Lambertienne parfaite (loi de cosinus). Cependant, le capteur dévie légèrement de cette loi et doit être corrigée. Cette correction est faite à partir de polynômes dépendants du cosinus de l'angle zénithal. Elle a été entreprise par le laboratoire de l'Institut Scientifique de Recherche Agronomique (INRA) d'Avignon, qui a fourni les polynômes suivants :

▷ Canaux dans l'optique et proche infrarouge :

$$P(\theta) = 1.3377 \cdot \cos^4 \theta - 3.1519 \cdot \cos^3 \theta + 2.6942 \cdot \cos^2 \theta + 0.1253 \cdot \cos \theta \quad (\text{D.1})$$

▷ Canal MIR :

$$P(\theta) = 1.0153 \cdot \cos^5 \theta - 1.0598 \cdot \cos^4 \theta - 0.5171 \cdot \cos^3 \theta + 1.0707 \cdot \cos^2 \theta + 0.0983 \cdot \cos \theta \quad (\text{D.2})$$

L'irradiance corrigée est :

$$I = I_r \cdot \frac{\cos(\theta)}{P(\theta)} \quad (\text{D.3})$$

où I_r est la valeur brute de l'irradiance et θ l'angle zénital. Cet angle est calculé par l'équation d'estimation de la hauteur h du soleil :

$$\sin(h) = \sin(\phi) \cdot \sin(\zeta) + \cos(\phi) \cdot \cos(\zeta) \cdot \cos(HA) \quad (\text{D.4})$$

avec ϕ la latitude, ζ la déclinaison et HA l'angle horaire.

- **Radiance (L_0)** : Le capteur mesurant la radiance réfléchie est calibré en comparant les mesures de réflectance enregistrées avec la mesure effectuée sur une surface de référence

dont la réflectance est connue. La valeur brute de la radiance est ensuite multipliée par un coefficient de calibration, qui a également été estimé par le laboratoire de l'INRA d'Avignon. Les valeurs numériques de ces coefficients pour les cinq canaux sont :

- ▷ $K_{c,BLUE} = 0.102$;
- ▷ $K_{c,VERT} = 0.134$;
- ▷ $K_{c,ROUGE} = 0.165$;
- ▷ $K_{c,NIR} = 0.214$;
- ▷ $K_{c,MIR} = 0.153$.

Les radiances corrigées sont, en conséquence :

$$L_{\lambda} = L_{0,\lambda} \cdot K_{C(\lambda)} \quad (D.5)$$

avec $L_{0,\lambda}$ la radiance spectrale brute et λ qui est blue, vert, rouge, NIR ou MIR.

Glossaire

- ALADIN:** Aire Limitée Adaptation Dynamique développement InterNational (*Modèle de prévision atmosphérique de Météo-France à l'échelle régionale*)
- ARPEGE:** Action de Recherche Petite Echelle Grande Echelle (*modèle de prévision atmosphérique de Météo-France à l'échelle globale*)
- ASCAT:** Advanced SCATterometer onboard of MetOp
- BRDF:** Bi-directional Reflectivity Distribution Function (*Fonction de Distribution Bi-Directional de la Réflectance*)
- CEPMMT:** Centre Europeen Pour les prévisions Météorologiques à Moyen Terme (*European Centre for Medium-Range Weather Forecasts*)
- CarboEurope:** Integrated Project CarboEurope-IP Assessment of the European Terrestrial Carbon Balance
- CAROLS:** Combined Airborne Radio-instruments for Ocean and Land Studies
- CESBIO:** Centre d'Études Spatiales de la Biosphère
- CETP:** Centre d'Etude des Environnements Terrestre et Planétaires
- CYCLOPES:** Carbon cYcle and Change in Land Observational Products from an Ensemble of Satellites
- CNRM:** Centre National de Recherches Météorologiques
- GMES:** Global Monitoring for Environment and Security
- EKF:** Extended Kalman Filter (*Filtre de Kalman Étendu*)
- EnKF:** Ensemble Kalman Filter (*Filtre de Kalman Ensemble*)
- ERS:** European Remote Sensing
- FVC:** Fractional Vegetation Cover (*Fraction de couvert végétale*)
- HSL:** Heure Solaire Locale
- INRA:** Institut Scientifique de Recherche Agronomique publique français
- ISBA:** Interactions between Soil-Biosphere-Atmosphere model
- ISBA-A-g_s:** Interactions between Soil-Biosphere-Atmosphere, CO₂-reactive model

- LAI:** Leaf Area Index (*Indexe de Surface Foliaire*)
- LEWIS:** L-band Radiometer for Estimating Water in Soils
- L-MEB:** L-band Microwave Emission of the Biosphere model
- LSM:** Land Surface Model (*Modèle de Surface*)
- MetOp:** Meteorological Operational satellite programme
- MERIS:** Medium REsolution Imaging Spectrometer
- MIR:** Mid InfraRed (*Moyen infrarouge*)
- MODIS:** MODerate-resolution Imaging Spectroradiometer
- MUREX:** Monitoring the Usable Soil Reservoir Experimentally
- NIR:** Near InfraRed (*proche infrarouge*)
- NDVI:** Normalized Difference Vegetation Index
- NWP:** Numerical Weather Prediction model (*Modèle de Prediction Numérique*)
- OI:** Optimal Interpolation (*Interpolation Optimale*)
- ONERA:** Office National d'Études et de Recherches Aérospatiales
- PIRRENE:** Programme Interdisciplinaire de Recherche sur la Radiométrie en Environnement Extérieur
- POLDER:** POLarization and Directionality of the Earth's Reflectance
- PRCD:** Programme Cadre de Recherche et Développement
- SAFRAN:** Système d'Analyse Fournissant des Renseignements Atmosphériques pour la Nivologie
- SIM:** SAFRAN - ISBA - MODCOU
- SMOS:** Soil Moisture and Ocean Salinity
- SMOSMANIA:** Soil Moisture Observing System - Meteorological Automatic Network Integrated Application
- SMOSREX:** Surface Monitoring Of the Soil Reservoir Experiment
- SPOT:** Satellite Pour l'Observation de la Terre
- SR:** Simple Ratio
- SWVI:** Soil Wetness Variation Index

Table des figures

1.1	Principe de l'assimilation de données de télédétection pour la réinitialisation des schémas de surface.	4
1.2	Structure du projet GEOLAND (projet intégré GMES du 6 ^{ième} PCRD portant sur la végétation). L'observatoire des flux naturels de carbone est indiqué.	6
3.1	Instrumentation de l'expérience SMOSREX. Les instruments de télédétection (le radiomètre en bande L, LEWIS, le luminancemètre CIMEL et les pyromètres Heitronics) sont installés sur le portique de mesure.	35
3.2	Evolution temporelle des variables du forçage atmosphérique de SMOSREX, pendant la période 2001-2004. Les barres verticales indiquent les valeurs maximum et minimum enregistrées pour chaque mois.	37
3.3	Précipitations cumulées mensuelles et annuelles sur le site de SMOSREX de 2001 à 2004.	38
3.4	Evolution temporelle de (du haut vers le bas) l'humidité de la couche superficielle du sol (0 à -6 cm), du contenu intégré en eau du sol, du profil d'humidité de la zone racinaire.	39
3.5	Dynamique de l'évolution du LAI et de la hauteur de la jachère sur le site de SMOSREX, pendant la période 2001-2004.	40
3.6	Biomasse aérienne mesurée sur la jachère de SMOSREX.	41
3.7	Le radiomètre LEWIS installé sur le site de SMOSREX depuis février 2003 et ses principales caractéristiques.	42
3.8	Instrumentation dans les spectres visible, proche-infrarouge, moyen infrarouge et infrarouge thermique sur SMOSREX. A gauche : reflectancemètre CIMEL pour mesurer la réflectance de la jachère. A droite : pyromètres Heitronics pour mesurer la température infrarouge sur jachère et sur sol nu.	43

3.9	Réfectances spectrales mesurées sur la jachère de SMOSREX en 2003 et 2004, à midi, à un angle d'incidence de 40° , pour les bandes spectrales CIMEL dans les domaines visible (450, 549, 648 nm), proche-infrarouge (837 nm), et moyen-infrarouge (1640 nm).	44
3.10	Comparaison entre le NDVI et le SR pour des journées de ciel clair, pour la période de 2003-2004.	46
3.11	Comparaison entre la température infrarouge mesurée sur la jachère de SMOSREX, et simulée par ISBA-A-gs. A gauche : évolution sur une dizaine de jours en 2003. A droite : corrélation entre les deux températures entre Juillet 2003 et Décembre de l'année 2004.	47
3.12	Comparaison entre la température infrarouge mesurée sur la jachère de SMOSREX et la température mesurée à 1 cm de profondeur. A gauche, évolution sur une dizaine de jours en 2003. A droite, corrélation entre les deux températures entre Juillet 2003 et décembre de l'année 2004.	47
3.13	Comparaison entre la température de la surface simulée par ISBA-A-gs et celle issue de bilan radiatif thermique à la surface, pour la période de l'année 2001 jusqu'à Juillet de l'année 2003.	48
3.14	Heating rates pour des journées sans nuages entre juillet et fin décembre 2003. .	50
3.15	Filtrage des mesures influencées par les nuages avec la méthode d'éclairement maximum mensuel.	51
3.16	corrélation entre les heating rates et l'humidité de la zone racinaire.	52
3.17	Relation entre Υ et w_2 pour différents vitesses du vent v (en $m \cdot s^{-1}$). Les lignes discontinues horizontales représentent la moyenne des Υ sur toute la période de 2001 à 2004, pour chaque intervalle de vitesse du vent.	53
3.18	Différences entre la première version d'ISBA et ISBA-A-gs, par rapport aux variables nécessaires d'entrée et les variables modélisées.	55
3.19	Simulation du contenu intégré en eau du sol et du LAI avec ISBA-A-gs, en utilisant l'une ou l'autre des stratégies de réponse à la sécheresse (tolérance -stratégie offensive- ou évitement -stratégie défensive-), et comparaison avec les observations. .	57

4.1	LAI, w_g , w_2 and monthly precipitation (from top to bottom) for 2001 to 2004 as measured over the SMOSREX site. The interpolated LAI (solid line), is used to drive the ISBA-A-g _s simulations and is superimposed on the observations (triangles). Note that the LAI interpolation technique may change from one year to another, depending on data quality and frequency. The w_g and w_2 simulations of ISBA-A-g _s (solid line) are superimposed on the observations (dots).	65
4.2	Observed versus modeled surface soil moisture (w_g) for 2001. The scatter plot is compared with the 1 :1 relationship (solid line).	71
4.3	Analysis of the root-zone soil moisture from surface observations using an Extended Kalman Filter (top) and a tuning variational method (bottom), for 2001-2004 over the SMOSREX experimental site : model control simulation (solid line), observations (dots) and analysis (circles).	76
4.4	Analysis of the root-zone soil moisture from surface observations using an Ensemble Kalman Filter (top) and a simplified 1D-VAR method (bottom), for 2001-2004 over the SMOSREX experimental site : model control simulation (solid line), observations (dots) and analysis (circles).	77
4.5	Gain components for the simplified 1D-VAR method compared with the yearly evolution of w_2 in 2003. Top : the control simulation (solid line) is superimposed on the analyses (circles). Bottom : the evolution of the gain components (multiplied by 100) at each analysis step.	79
4.6	Skill score for different error levels in observations (R) and forecasted state variables (P) for the simplified 1D-VAR. Note that the Q values in the abscises axis are those in eq. (14). The dashed-line superimposed on the figures corresponds to the performance with an observation error corresponding to the SMOS satellite specification.	80
4.7	a) Evolution temporelle et effondrement d'un ensemble de 100 membres avec la méthode EnKF "pure". A chaque observation, les membres de l'ensemble corrigé sont dessinés en triangles verts. La trajectoire de chaque ensemble entre deux observations est représentée par une ligne continue noire. b) Ecart type de l'ensemble avec la méthode EnKF "pure" (points rouges) et avec l'EnKF avec un facteur d'inflation (points noirs).	86
4.8	Analyses de w_2 comparées aux observations et à la simulation de contrôle, avec un ensemble de 100 membres.	87

5.1	Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode 1D-VAR simplifiée. En haut, les analyses de w_2 (cercles rouges) sont comparées aux observations (points bleus) et à la simulation de contrôle (trait continu noir). En bas, le LAI estimé par le modèle après assimilation (points violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).	92
5.2	Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode EnKF. En haut, les analyses de w_2 (cercles rouges) sont comparées aux observations (points bleus) et à la simulation de contrôle (trait continu noir). En bas, le LAI estimé par le modèle après assimilation (points violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).	92
5.3	Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie offensive avec la méthode 1D-VAR simplifiée. En haut, les analyses de w_2 (cercles rouges) sont comparés aux observations (points bleus) et à la simulation de contrôle (trait noir). En bas, le LAI estimé par le modèle après assimilation (point violets) est montré et comparé à la simulation de référence (trait noir) et aux observations (points bleus).	93
5.4	Analyses de w_2 et LAI dans ISBA-A- g_s en stratégie défensive avec la méthode EnKF. En haut, les analyses de w_2 (cercles rouges) sont comparés aux observations (points bleus) et à la simulation de contrôle (trait noir). En bas, le LAI estimé par le modèle après assimilation est montré (point violets) et comparé à la simulation de référence (trait noir) et aux observations (points bleus).	94
5.5	a) Analyses et évolution de w_2 avec le 1D-VAR en configuration de fenêtre glissante (points et lignes bleus foncés, respectivement) et en configuration de fenêtre séquentielle (points et lignes rouges). Les observations de w_2 et la simulation de contrôle d'ISBA-A- g_s sont représentées par des points noirs et par une trait de couleur bleu ciel, respectivement. b) Les observations de LAI (points verts) sont superposées à l'évolution du LAI après réinitialisation de w_2 avec le 1D-VAR fenêtre glissante (ligne bleu) et fenêtre séquentielle (ligne rouge).	95
5.6	En haut, l'estimation de w_2 après l'assimilation des observations de LAI (cercles rouges) est comparée au modèle (trait noir) et aux observations (points bleus). La figure au milieu montre les analyses de la biomasse (cercles verts) en comparaison à l'estimation du modèle (trait noir) et aux observations (points bleus). En bas, le LAI après assimilation (cercles verts) est aussi superposé à la référence (trait noir) et aux observations in-situ (points bleus).	97

5.7	a) LAI, b) w_g , c) w_2 and d) monthly precipitation from 2001 to 2004 as measured over the SMOSREX site. On top the interpolated LAI (solid line) is superimposed to the observations (circles). Note that the LAI interpolation technique may change from one year to another, depending on data quality and frequency. The w_g and w_2 simulations of ISBA-A-gs (solid line) using the offensive strategy as response to water stress (drought tolerant) are superimposed to the observations (dots for w_g and stars for w_2).	103
5.8	Temporal evolution of the linearized H_{bio} standard deviation. H_{bio} is computed by an initial perturbation of the vegetation biomass. Four different intervals for the initial perturbation of the biomass are tested. Perturbations are given in $m^2 \cdot kg^{-1}$.	106
5.9	Analysis of a) the root zone soil moisture (circles) and b) vegetation biomass, using a simplified 1D-VAR method from 2001 to 2004 over the SMOSREX experimental site. c) LAI before and after the assimilation. For comparison purposes, analysed values are superimposed over the in-situ observations (points) and the model basic estimations (solid line).	107
5.10	Same as Fig. 5.9, but using a fixed wilting point. The results for the years 2001 and 2002 are omitted as being similar to those of Fig. 5.9.	110
5.11	Yearly evolution of a) the Gross Primary Production (GPP), b) the ecosystem respiration (Reco), c) the Net Ecosystem Exchange (NEE), and d) the evapotranspiration flux (LE), before (thick line) and after (fine line) the assimilation of surface soil moisture and LAI observations. The data for the model has been smoothed using a 10-day boxcar average window.	111
5.12	Analysis of a) the root zone soil moisture (circles) and, b) vegetation biomass, using a simplified 1D-VAR method from 2001 to 2004, with precipitation set to zero. c) LAI before and after the assimilation. Analysed values are superimposed over the in-situ observations (points) and the model basic estimations (solid line) for comparison purposes.	113
5.13	Same as Fig.5.12, but using a fixed wilting point. The results for the years 2001 and 2002 are omitted as being similar as in Fig. 5.12.	114
5.14	Système couplé ISBA-A-gs / L-MEB.	120
5.15	Observations (radiomètre LEWIS) et simulations de L-MEB de la T_B sur la jachère de SMOSREX pour l'année 2003 et pour les polarisations horizontale (en bleu) et verticale (en rouge), à un angle d'incidence de 40° . En bas, les deux valeurs sont comparées et superposées à la ligne 1 :1.	121

5.16	Humidités de la surface de la jachère de SMOSREX en 2003 et 2004 obtenues en appliquant l'équation 5.1 aux observations de T_B de LEWIS (cercles bleus). Les données sont filtrées selon les conditions imposées pour la définition de plusieurs "flags". A ces estimations, les observations de w_g sur SMOSREX sont superposées (trait rouge).	123
5.17	Comparaison entre le LAI interpolé (ligne solide en bleu) à partir des observations sur le terrain (point verts), et le LAI issue de l'application de la méthode de Roujean et Lacaze (2002) au jeu d'observations de réflectances mesurées avec le reflectancemètre CIMEL (ligne solide rouge), en 2003 et 2004.	127
5.18	Analyses de a) w_2 (cercles rouges) et b) biomasse de la végétation (cercles verts), en appliquant la méthode 1D-VAR simplifiée. Les observations sont w_g et LAI dérivés des observations de T_B et réflectance, respectivement. La période de mesure s'étend de juillet 2003 jusqu'à la fin de 2004. Les analyses sont superposées aux observations in-situ (points bleus) et à la simulation de contrôle (trait noir). c) Le LAI simulé après ré-initialisation de w_2 et la biomasse de la végétation (cercles vertes) est superposé aux observations (point bleus) et à la simulation de contrôle (ligne solide noir).	129
B.1	Création du forçage atmosphérique pour l'année 2003.	146

Liste des tableaux

2.1	Présentation synthétique des 3 méthodes d'estimation linéaire.	21
3.1	Nombre total de " Heating rates " (Υ) obtenus de 2001 à 2004 par régression linéaire avec 2, 3, 4 ou 5 observations consécutives (obs) de la température infrarouge de la surface. Deux seuils sont imposés pour filtrer observations sans l'effet des nuages : éclaircissement supérieur à 70% ou 80% de l'éclaircissement maximum mensuel. Dans chaque cas, le nombre total d'observations disponibles pour calculer les " Heating rates " est aussi indiqué (" candidates ").	51
4.1	Main soil and vegetation parameters used in the ISBA-A- g_s model over the SMOS-REX site.	66
4.2	Definition of the background error matrix P and the observation error matrix R, for four assimilation schemes : Ensemble Kalman Filter (EnKF), Extended Kalman Filter (EKF), a simplified 1D-VAR and a tuning variational method (T-VAR).	70
4.3	Surface and root-zone soil moisture yearly scores of the control simulation and for the whole 2001-2004 period (RMSE ($m^3 \cdot m^{-3}$), bias ($m^3 \cdot m^{-3}$) and skill score E).	74
4.4	Root-zone soil moisture analysis yearly scores and for the whole 2001-2004 period (RMSE ($m^3 \cdot m^{-3}$), bias ($m^3 \cdot m^{-3}$) and skill score E), using an Ensemble Kalman Filter (EnKF) with N=10, 20, 50, 100 and 200 members, an Extended Kalman Filter (EKF), a simplified 1D-VAR and a tuning variational method (T-VAR).	78
4.5	Computer processing time for a whole year (in seconds) for the Ensemble Kalman Filter (EnKF), Extended Kalman Filter (EKF), Simplified 1D-VAR and a tuning variational (T-VAR). The computing time for the EnKF is shown for ensembles of N=10, 20, 50, 100 and 200 members.	81

4.6	Résultats de la comparaison entre les quatres méthodes d'assimilation et le schéma de surface ISBA-A-g _s pour le score des analyses de w_2 (E), le RMSE, le pourcentage des increments non-nuls et le coût numérique	85
4.7	RMSE, biais moyen et critère de Nash annuel, de 2001 à 2004, et pour l'ensemble des quatre années, entre les observations de w_2 et les analyses.	87
5.1	Main soil and vegetation parameters used in the ISBA-A-g _s simulations over the SMOSREX site.	102
5.2	Yearly and global RMSE (in $m^3 \cdot m^{-3}$), mean bias (in $m^3 \cdot m^{-3}$) and score (E), for the root zone-soil moisture, LAI and biomass between the observations and the control model simulation.	107
5.3	Yearly and global RMSE (in $m^3 \cdot m^{-3}$), mean bias (in $m^3 \cdot m^{-3}$) and score (E) for the root zone-soil moisture and vegetation biomass, between analysis and observations for four different model configurations. For the LAI, statistics are calculated between observations and model simulations after data assimilation. . .	109
5.4	CO ₂ and evapotranspiration (LE, in $mm \cdot year^{-1}$) yearly cumulated fluxes. The assimilated CO ₂ by the vegetation (GPP, in $g \cdot C \cdot m^{-2} \cdot year^{-1}$), the CO ₂ emission by soil and plants (Reco, in $g \cdot C \cdot m^{-2} \cdot year^{-1}$) and the balance between both terms (NEE, in $g \cdot C \cdot m^{-2} \cdot year^{-1}$) after the assimilation of LAI and w_g observations for four different model configurations, is compared to the control simulation. . .	112
5.5	Définition de chaque "flag" appliquée à la série temporelle de T_B du radiomètre LEWIS. Les trois dernières colonnes correspondent à la valeur des coefficients de la relation empirique 5.1 pour chaque flag.	124
5.6	Statistiques correspondant à l'inversion des observations de T_B pour chaque flag employé. Les deux dernières colonnes correspondent au nombre d'estimations obtenues pour chaque flag et le pourcentage sur le nombre initial d'observations. . .	124
5.7	Statistiques correspondant à l'assimilation avec la méthode 1D-VAR des produits de w_g obtenus à partir des observations de T_B	125

REMOTE SENSING DATA ASSIMILATION FOR THE CONTINENTAL SURFACES MONITORING : IMPLEMENTATION OVER AN EXPERIMENTAL SITE

Thèse de l'Université Toulouse III - Paul Sabatier

Discipline : Télédétection spatiale et modélisation de la biosphère

Auteur : Joaquín MUÑOZ SABATER

Directeur de thèse : Jean-Christophe CALVET

Soutenue le 13 avril 2007 au

Centre National de recherches météorologiques, Météo-France,
42 avenue G. Coriolis, 31057 TOULOUSE Cedex 1, France

Abstract :

The research undertaken in this thesis focuses on the assimilation of remotely sensed data for continental surface monitoring. The analysed variables are root-zone soil moisture content and above-ground vegetation biomass. The data cover the period 2001 to 2004 and were collected over fallow at the SMOSREX experimental site, south-west of Toulouse, France. Four assimilation techniques (EKF, EnKF, 1D-VAR and T-VAR) were implemented into the Météo-France SVAT model ISBA-A-g_s and their performances compared to each other. It was found that the 1D-VAR is the most effective technique, for both quality of the analyses and lower computational cost. This particular technique was applied in a joint assimilation of surface soil moisture and leaf area index observations, as well as of remotely sensed measurements collected over the site. The assimilation improves the control simulation, if a dynamic wilting point is introduced (this is particularly useful during strong drought periods). The impact of a high level of uncertainty in the precipitation data is also evaluated.

Keywords : data assimilation, remote sensing, soil moisture, vegetation biomass

ASSIMILATION DE DONNÉES DE TÉLÉDÉTECTION POUR LE SUIVI DES SURFACES CONTINENTALES : MISE EN ŒUVRE SUR UN SITE EXPÉRIMENTAL

Thèse de l'Université Toulouse III - Paul Sabatier

Discipline : Télédétection spatiale et modélisation de la biosphère

Auteur : **Joaquín MUÑOZ SABATER**

Directeur de thèse : **Jean-Christophe CALVET**

Soutenue le 13 avril 2007 au

Centre National de recherches météorologiques, Météo-France,
42 avenue G. Coriolis, 31057 TOULOUSE Cedex 1, France

Résumé :

Le travail réalisé dans le cadre de cette thèse porte sur l'assimilation de données de télédétection pour le suivi des surfaces continentales. Les variables analysées sont l'état hydrique du sol et la biomasse de la végétation, sur la jachère du site expérimental SMOS-REX, de 2001 à 2004, au sud-ouest de Toulouse. Quatre méthodes d'assimilation (EKF, EnKF, 1D-VAR et T-VAR) ont été mises en œuvre dans le modèle ISBA-A-g_s de Météo-France, et comparées. La méthode 1D-VAR est la plus performante, aussi bien pour la qualité des analyses qu'en temps de calcul. Cette méthode a été appliquée à l'assimilation simultanée des observations d'humidité de la surface et de l'indice foliaire, ainsi qu'aux mesures de télédétection réalisées sur le site. L'assimilation améliore la simulation de contrôle, à condition d'introduire un point de flétrissement dynamique (cela est particulièrement utile pendant les périodes de forte sécheresse). Les effets d'une incertitude sur les précipitations sont évalués.

Mots Clés : assimilation de données, télédétection, humidité du sol, biomasse de la végétation