



**HAL**  
open science

# Désagrégation spatiale de températures Météosat par une méthode d'assimilation de données (lisseur particulière) dans un modèle de surface continentale

Rihab Mechri

► **To cite this version:**

Rihab Mechri. Désagrégation spatiale de températures Météosat par une méthode d'assimilation de données (lisseur particulière) dans un modèle de surface continentale. Climatologie. Université de Versailles-Saint Quentin en Yvelines, 2014. Français. NNT : 2014VERS0036 . tel-01238472

**HAL Id: tel-01238472**

**<https://theses.hal.science/tel-01238472>**

Submitted on 5 Dec 2015

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

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

Université de Versailles Saint-Quentin-en-Yvelines

**THÈSE DE DOCTORAT**

Spécialité : Météorologie, Océanographie physique et Physique de  
l'Environnement

---

**DÉSAGRÉGATION SPATIALE DE TEMPÉRATURES  
MÉTÉOSAT PAR UNE MÉTHODE D'ASSIMILATION  
DE DONNÉES (LISSEUR PARTICULAIRE) DANS  
UN MODÈLE DE SURFACE CONTINENTALE**

---

par **Rihab MECHRI**

Présentée et soutenue publiquement le 4 décembre 2014, devant le jury  
composé de :

Pr. Philippe BOUSQUET ..... Président du jury  
Dr. Selma CHERCHALI ..... Examineur  
Dr. Thomas CORPETTI ..... Rapporteur  
Dr. Dominique COURAULT ..... Rapporteur  
Dr. Abdelaziz KALLEL ..... Examineur  
Dr. Carole NAHUM ..... Examineur  
Dr. Catherine OTTLÉ ..... Directeur de thèse  
Dr. Olivier PANNEKOUCKE ..... Co-directeur de thèse





*Je dédie ce travail à mes chers  
parents Amina et Badi et à mon  
cher époux Taha.*

---



*“Que chacun juge d’après son  
opinion personnelle en se basant  
sur ses propres lectures et non pas  
d’après ce que lui disent les  
autres.”*

---

Albert Einstein, Comment je vois  
le monde (1934)

---

Cher lecteur, avant de vous infliger ces cent cinquante pages garnies d'informations scientifiques, je souhaite d'abord adresser un clin d'œil à des personnes merveilleuses qui ont fait la richesse de ces trois ans de thèse.

En premier, je souhaite remercier ma directrice de thèse Mme Catherine Ottlé, pour m'avoir accueilli à mon arrivée en France, pour avoir eu confiance en moi malgré mes modestes connaissances dans les domaines d'assimilation de données et de physique de l'environnement, pour avoir piloté mon travail de thèse avec beaucoup de patience, pour m'avoir transmis un peu de son sens de critique et pour m'avoir appris à me remettre en question. Catherine, merci pour ta disponibilité inconditionnelle et pour tout ce que tu m'as appris. J'ai été extrêmement ravie de travailler avec toi.

J'adresse de chaleureux remerciements à mon codirecteur de thèse, Mr Olivier Pannekoucke pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de cette thèse. Olivier, merci de m'avoir emmené dans ton univers plein de "X", "Y", "variance", "particules", ..., de m'avoir fait redécouvrir cet univers que j'ai abandonné après mes deux ans de préparatoire. Je n'oublierai sans doute pas nos petites prises de tête et nos fructueuses discussions au CNRM. Ce fut un grand plaisir de travailler avec toi.

Un grand merci à Mr Abdelaziz Kallel, qui m'a accordé cette chance de rencontrer ma directrice de thèse, qui m'a offert un encadrement de qualité lors de mon projet de fin d'étude à l'école nationale d'ingénieur en Tunisie ainsi que mon projet de fin de master de recherche et m'a permis de me plonger un peu dans le domaine de traitement de signal et d'images satellitaires. Abdelaziz, merci infiniment pour tes conseils, pour ton écoute et pour la confiance que tu m'as accordée tout au long de ce petit parcours scientifique.

Je voudrais aussi remercier les rapporteurs de cette thèse, Mme Dominique Couraut et Mr Thomas Corpetti pour l'intérêt qu'ils ont porté à mon travail de thèse.

J'associe à ce remerciement, mes examinateurs de thèse Mme Carole Nahum, coordinatrice à la DGA, Mme Selma Cherchali, coordinatrice au CNES, et Mr Philippe Bousquet, président de mon Jury de thèse et directeur de l'école doctorale ED-129. Merci pour l'intérêt que vous avez porté à mon travail de thèse.

Je remercie particulièrement, la Direction Générale de l'Armement (DGA) et le Centre National des Etudes Spatiales (CNES) pour avoir financé mon travail de thèse et m'avoir permis de finir cette thèse dans les meilleures conditions.

Je souhaite adresser un grand merci aux directeurs du Laboratoire des Sciences des Climat et de l'environnement (LSCE), Mr Cyril Moulin et Mme Elsa Cortijo, pour leur accueil et leur disponibilité.

Je remercie aussi tout le Groupe Administratif (GA) du LSCE et en particulier Mme Laurance Bayard, Mme Florence Aptel, Mme Cessile Tessier et Mme Catherine Bertrand.

Je souhaite aussi remercier tout les membres du bureau des étrangers du Commissariat à l'Énergie Atomique (CEA) et en particulier Mme Laure Reuter pour sa compétence incomparable et son accompagnement pour toutes les démarches administratives concernant mon séjour en tant que scientifique étranger en France. Merci Laure pour tous tes conseils concernant la préfecture et les démarches nécessaires qui me sont très précieuses.

J'adresse un grand remerciement à tous les membres de mon équipe de recherche, anciennement appelé INVSAT et actuellement appelé MOSAIC, pour cette ambiance de travail agréable. Je remercie particulièrement, Mr Frédéric Chevallier ancien directeur de l'équipe INVSAT et Mr Philippe Peylin directeur de l'équipe MOSAIC pour avoir toujours répondu à mes requêtes concernant les serveurs de calcul. Je remercie aussi Mme Fabienne Maignan, qui m'a beaucoup aidé pour m'initier à débayer en Fortran, à Mr Jean-Yves Peterschmitt pour les petites gourmandises et les conseils avisés en langage Python, à Mr Pascal Maugis pour toutes les discussions et tous les commentaires

---

constructifs qu'on a partagé autour de mon travail, ainsi qu'à Mme Natasha MacBean pour avoir relu et corrigé mes modestes expressions en anglais scientifique.

J'aimerais aussi remercier toute l'équipe informatique du LSCE et particulièrement Mr Julien Bruna. Julien merci beaucoup pour m'avoir aidé à résoudre mes problèmes avec mon PC, pour avoir toujours répondu à mes demandes d'installation de logiciels, mises à jour, coup de fil et e-mail à temps. Sans ton aide et ta disponibilité, ma thèse n'aurait pas pu se dérouler dans d'aussi bonnes conditions de travail.

Voilà que j'arrive à mes amis du LSCE, je ne sais donc pas quels mots seraient les plus pertinents pour vous transmettre mes remerciements. Je commence par M. Robin Locatelli, mon "BTE" (Best Team-mate Ever), merci pour m'avoir écoutée quand j'avais le plus besoin de parler et merci de m'avoir donné l'envie de courir dehors ! Merci aussi pour les petites prise de têtes même si je regrette un peu le temps qu'on a passé sans se parler... Bref merci d'être mon BTE ! Mme Cindy Cressot et Mme Irène Mappe, merci les filles pour votre écoute et votre présence dans les moments de joie et de galère, votre générosité et votre amitié. Merci pour toutes les pauses café/thé, toutes les discussions de filles et les discussions scientifiques aussi ! Je remercie aussi, le génie Mr Sébastien Léonard. Sébastien merci pour tous tes conseils, tes astuces et ton génie ! Je remercie très chaleureusement mes collègues, Mr Pradeebane Vaittinada, Mr Benjamin Quesada, Mr Cyrille André, Mr Fan Li, Mr Emmanuel Renault, Mme Constansa Comté, Mr Cedric Bacour, Mme Lisbeth Ron, Mme Xiaoxia Chang, Mme Mai-Lan Boytard, Mr Felix Voguel, Mme Martha Voguel (Marty Bird), Mme Lyzia Benshil, Mme Mayssa Ben Sassi, Mme Sarah Dantec, Mr Nicholas Roux, Mr Tao Wang, Mr Yue Chao, Mr Luis Moulina et Mr Pierre Brender. Sans vous ces trois ans de thèse n'auraient jamais pu être aussi agréables ! Merci pour tout ce qu'on a partagé, discussions scientifiques à la cantines, Ultimate (Frisbee), course à pied, etc.

J'arrive ainsi à la famille, je remercie d'abord mon cher époux, Mr Taha Ouali, qui m'a soutenue tout au long de ces trois ans de thèse. Ton amour, ta patience et ton soutien sont derrière ma réussite et sans toi je n'aurai jamais pu tenir jusqu'au bout. Merci mon amour. Maman (Mme Amina Garbout), Papa (Mr Badî Mechri) et ma chère grand-mère (Mme Aycha Bayard), merci pour tout l'amour et tous les sacrifices que vous avez fait pour moi afin que je puisse progresser dans ma carrière. Vous m'avez accompagnée jusqu'au bout avec beaucoup de patience et d'amour, tous les mots de remerciement ne peuvent guère décrire ma gratitude, je me contente donc de vous dire merci pour tout !

Ma chère sœur Folla (FliL', Flifil, Falfoula), mon cher petit frère Mehdi (Hdouch, Mhidich, Mehouch) merci pour votre amour et votre confiance en moi ! Merci pour votre soutien. J'ai de la chance de vous avoir comme frère et sœur et je vous aime tellement. Mes meilleures amies Mme Bouthaina Hammami (Bothboth, Bathbouthti) et Mme Rawia Mhiri Hakim (Rawitta, Raw, Riwti), merci d'avoir toujours été là pour me soutenir et m'encourager dans tout ce que je fais, de m'avoir accepté avec tous mes défauts et de m'avoir aimé sans conditions ! Votre amitié est le plus précieux trésor ! Merci les filles.

J'aimerais aussi remercier mes beaux parents, Mme Serra Ouali et Mr Mohammed Ouali ainsi que ma belle sœur Mme Asma Ouali (Sayma) et ma cousine Mme Basma Cheffi (Bassoumti) pour leur soutien et leurs encouragements.

Un remerciement particulier à un couple merveilleux, si accueillant et si chaleureux, à qui je dois beaucoup de respect et d'amour Mr Michel Ricart et Mme Catherine Ricart. Merci pour tout ce que vous m'avez appris et fait découvrir de la France et à tous les moments de bonheur qu'on a partagé.

Je souhaite clôturer ces remerciements par un clin d'œil à certaines personnes que j'ai rencontré durant cette thèse et qui m'ont marquées, dont Mr Mahmoud Al Az-zam, Mr Safouene Chibani, Mr Mayssara Moukhirar, Mr Michail D. Vrettas, Mr Denis



---

Dreano, Mr Mohammed Gharamti, Mme Amal Elakkraoui, Mme Divya Surendran ainsi qu' à tous les amis de l'école d'été d'assimilation de données "Les Houches". Ce fut un grand plaisir de vous rencontrer et merci pour vos précieux conseils et agréable esprit scientifique.

**RÉSUMÉ** La température des surfaces continentales (LST) est une variable météorologique très importante car elle permet l'accès aux bilans d'énergie et d'eau du continuum Biosphère-Atmosphère. Sa haute variabilité spatio-temporelle nécessite des mesures à haute résolution spatiale (HRS) et temporelle (HRT) pour suivre au mieux les états hydriques du sol et des végétations.

La télédétection infrarouge thermique (IRT) permet d'estimer la LST à différentes résolutions spatio-temporelles. Toutefois, les mesures les plus fréquentes sont souvent à basse résolution spatiale (BRS). Il faut donc développer des méthodes pour estimer la LST à HRS à partir des mesures IRT à BRS/HRT. Cette solution est connue sous le nom de désagrégation et fait l'objet de cette thèse.

Ainsi, une nouvelle approche de désagrégation basée sur l'assimilation de données (AD) est proposée. Il s'agit de contraindre la dynamique des LSTs HRS/HRT simulées par un modèle en minimisant l'écart entre les LST agrégées et les données IRT à BRS/HRT, sous l'hypothèse d'homogénéité de la LST par type d'occupation des sols à l'échelle du pixel BRS. La méthode d'AD choisie est un lisseur particulière qui a été implémenté dans le modèle de surface SETHYS (Suivi de l'Etat Hydrique du Sol).

L'approche a été évaluée dans une première étape sur des données synthétiques et validée ensuite sur des données réelles de télédétection sur une petite région au Sud-Est de la France. Des séries de températures Météosat à 5 km de résolution spatiale ont été désagrégées à 90m et validées sur une journée à l'aide de données ASTER. Les résultats encourageants nous ont conduit à élargir la région d'étude et la période d'assimilation à sept mois. La désagrégation des produits Météosat a été validée quantitativement à 1km à l'aide de données MODIS et qualitativement à 30m à l'aide de données Landsat7. Les résultats montrent de bonnes performances avec des erreurs inférieures à 2.5K sur les températures désagrégées à 1km.

**Mots-Clés** : Désagrégation, Filtre Particulaire, Lisseur Particulaire, Météosat, Assimilation de données, Température des surfaces continentales, Infrarouge Thermique, Télédétection.

**ABSTRACT** Land surface temperature (LST) is one of the most important meteorological variables giving access to water and energy budgets governing the Biosphere-Atmosphere continuum. To better monitor vegetation and energy states, we need high temporal and spatial resolution measures of LST because its high variability in space and time.

Despite the growing availability of Thermal Infra-Red (TIR) remote sensing LST products, at different spatial and temporal resolutions, both high spatial resolution (HSR) and high temporal resolution (HTR) TIR data is still not possible because of satellite resolutions trade-off : the most frequent LST products being low spatial resolution (LSR) ones.

It is therefore necessary to develop methods to estimate HSR/HTR LST from available TIR LSR/HTR ones. This solution is known as "downscaling" and the present thesis proposes a new approach for downscaling LST based on Data Assimilation (DA) methods. The basic idea is to constrain HSR/HTR LST dynamics, simulated by a dynamical model, through the minimization of their respective aggregated LSTs discrepancy toward LSR observations, assuming that LST is homogeneous at the land cover type scale inside the LSR pixel.

Our method uses a particle smoother DA method implemented in a land surface model : SETHYS model (Suivie de l'Etat Hydrique de Sol). The proposed approach has been firstly evaluated in a synthetic framework then validated using actual TIR LST over a small area in South-East of France. Meteosat LST time series were downscaled from 5km to 90m and validated with ASTER HSR LST over one day. The encouraging results conducted us to expand the study area and consider a larger assimilation period of seven months. The downscaled Meteosat LSTs were quantitatively validated at 1km of spatial resolution (SR) with MODIS data and qualitatively at 30m of SR with Landsat7 data. The results demonstrated good performances with downscaling errors less than 2.5K at MODIS scale (1km of SR).

**Keywords** : Downscaling, Particle Filter, Particle Smoother, Meteosat, Data Assimilation, Land Surface Temperature, Thermal Infrared, Remote sensing.



---

# Table des matières

Table des figures	xiii
Liste des tableaux	xv
Liste des abréviations	xvii
<b>1 Introduction</b>	<b>1</b>
1.1 Hypothèses générales	3
1.2 État de l'art sur la désagrégation de la LST par les méthodes de régression	4
1.3 État de l'art sur la désagrégation de la LST par les méthodes d'assimilation/inversion de données	6
1.4 Motivations pour le choix de l'approche de désagrégation	10
<b>2 Assimilation de données</b>	<b>13</b>
2.1 Introduction à l'assimilation de données	13
2.2 Problème de filtrage et formalisme Bayésien	14
2.3 Dynamique linéaire et statistiques gaussiennes : le filtre linéaire de Kalman	16
2.4 Cadre général de la dynamique non linéaire	18
2.4.1 Le filtre de Kalman d'ensemble ( <i>EnKF</i> )	18
2.4.2 Le filtre particulière ( <i>PF</i> )	20
Conclusion	24
<b>3 Application au modèle SETHYS</b>	<b>25</b>
3.1 Le modèle SETHYS	25
3.1.1 Bilan d'énergie au niveau de l'interface	26
3.1.2 Bilan hydrique et contenu en eau du sol	32
3.1.3 La température de brillance de la surface	34
3.1.4 Description générale du fonctionnement du modèle SETHYS	35
3.2 Analyse de sensibilité de la température de brillance vis-à-vis des paramètres du modèle SETHYS sur notre site d'étude	37
3.2.1 Introduction à l'Analyse de Sensibilité (AS)	37
3.2.2 Analyse de Sobol	39
3.2.3 Calcul des indices de Sobol $S_i$ et $S_{T_i}$ pour le modèle SETHYS	42
3.2.4 Résultats de l'analyse de Sobol	44

Conclusion . . . . .	45
<b>4 Région d'étude et base de données Crau-Camargue pour les années 2006 et 2009</b>	<b>53</b>
4.1 Région d'étude : Crau-Camargue . . . . .	53
4.2 Occupation des sols . . . . .	54
4.3 Données de forçage du modèle SETHYS . . . . .	58
4.3.1 Données atmosphériques . . . . .	58
4.3.2 Caractéristiques de la végétation . . . . .	59
4.4 Données de NDVI . . . . .	59
4.5 Données de température de surface . . . . .	61
4.5.1 Données d'observation . . . . .	61
4.5.2 Données de validation . . . . .	62
4.6 Inter-étalonnage des données d'observation et de validation . . . . .	63
4.6.1 Inter-étalonnage des données pour l'année 2006 . . . . .	64
4.6.2 Inter-étalonnage des données pour l'année 2009 . . . . .	65
<b>5 Application de la méthode de désagrégation par lissage particulière sur des données synthétiques et réelles</b>	<b>73</b>
5.1 Introduction . . . . .	73
5.2 Méthode . . . . .	73
5.3 Article 1 : Genetic particle filter application to land surface temperature downscaling . . . . .	77
5.4 Résultats et conclusions . . . . .	94
<b>6 Désagrégation de séries temporelles Météosat à plus grande échelle</b>	<b>97</b>
6.1 Introduction . . . . .	97
6.2 Article 2 (version <i>draft</i> ) : Downscaling of Meteosat Land Surface Temperature over an heterogeneous landscape using a Data Assimilation approach . . . . .	99
6.2.1 Introduction . . . . .	99
6.2.2 Data and methods . . . . .	100
6.2.3 Results and discussion . . . . .	105
6.2.4 Conclusions . . . . .	108
6.3 Résultats et conclusions . . . . .	119
<b>7 Conclusions et perspectives</b>	<b>121</b>
7.1 Principaux résultats . . . . .	122
7.2 Limites et difficultés . . . . .	124
7.3 Perspectives . . . . .	125

---

## Table des figures

1.1	Schématisation du problème de désagrégation à l'échelle d'un pixel à basse résolution spatiale. . . . .	7
2.1	Schéma descriptif de l'assimilation de données séquentielle. . . . .	14
2.2	Génération d'évènements : les barres en gris présentent les probabilités relatives à chaque évènement $i_{q_1 \geq q \leq 5}$ , la courbe en noir correspond à la CDF et le trait en pointillé présente la transformation du nombre uniforme $u_k$ en un évènement (Figure extraite de <a href="#">Hol (2004)</a> ). . . . .	21
2.3	Description du principe de filtrage particulière. . . . .	22
3.1	Schéma du bilan radiatif global (figure extraite de <a href="#">Saux Picart (2007)</a> ) . . .	26
3.2	Schéma du bilan d'énergie pour une surface infinie : $\alpha$ , $\epsilon$ et $T$ désignent respectivement l'albédo, l'émissivité et la température du sol. . . . .	27
3.3	Schéma du bilan d'énergie pour le modèle SETHYS (figure extraite de <a href="#">Saux Picart (2007)</a> ). . . . .	28
3.4	Schéma résistif représentant les échanges de surface (figure extraite de la thèse de <a href="#">Saux Picart (2007)</a> ) . . . . .	31
3.5	Schéma du bilan hydrique dans le modèle SETHYS (figure extraite de <a href="#">Saux Picart (2007)</a> ). . . . .	34
3.6	Schéma descriptif du fonctionnement global du modèle SETHYS. . . . .	36
3.7	Synthèse détaillée des méthodes d'AS (Figure extraite de <a href="#">Iooss (2011)</a> ). . .	38
3.8	Diagramme de décision pour le choix de la méthode d'AS appropriée (Figure extraite de <a href="#">de Rocquigny et al. (2008)</a> ). . . . .	39
3.9	Schéma de présentation du modèle en boîte noire . . . . .	40
3.10	Comparaison d'un échantillon issu d'un générateur de nombres aléatoires classique de la loi uniforme (couleur bleue) et d'un échantillon issu d'un générateur LD (couleur rouge) : les tailles des échantillons sont égales à $N = 800$ . . . . .	43
3.11	Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/02/2009 au 07/02/2009. . . . .	49
3.12	Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/05/2009 au 07/05/2009. . . . .	50
3.13	Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/09/2009 au 07/09/2009. . . . .	51

4.1	Région de Crau-Camargue : zones définies par les cartes d'occupation du sol et sites de mesures des forçages micro-météorologiques, et de végétation. . . . .	54
4.2	Carte d'occupation des sols initiale à 12 classes : produit ASTER à 90m de résolution spatiale pour l'année 2006. . . . .	55
4.3	Nouvelle carte d'occupation des sols à 7 classes (année 2006). . . . .	56
4.4	Carte d'occupation des sols initiale à 8 classes : produit Landsat7 à 30m de résolution spatiale pour l'année 2009. . . . .	57
4.5	Nouvelle carte d'occupation des sols à 7 classes (année 2009). . . . .	57
4.6	Cartes de <i>NDVI</i> Landsat 7 disponibles pour l'année 2009 . . . . .	60
4.7	LAI moyens et écarts-types pour les classes de végétation de la carte d'occupation des sols Landsat7 de l'année 2009 . . . . .	61
4.8	Résultat de la correction des températures Météosat acquises sur la journée 26/07/2006 à 10h45 par la méthode RANSAC . . . . .	65
4.9	Fonctions de distribution cumulée des températures Météosat et des températures Landsat7 (à HRS et agrégées à la résolution spatiale de Météosat) : les courbes en rose présentent les CDF des températures HRS Landsat7(CDF LANDSAT7 HSR), les courbes en vert présentent les CDF des températures Landsat7 agrégées (CDF LANDSAT7 BRS)et les courbes en bleu présentent les CDF des températures Météosat (CDF SEVIRI BRS). . . . .	66
4.10	CDFs des températures MODIS (à HRS et agrégées à la résolution spatiale de Météosat ) versus CDFs Météosat : les courbes en rose présentent les CDF des températures MODIS à 1K (CDF MODIS HSR), les courbes en vert présentent les CDF des températures MODIS agrégées (CDF MODIS BRS) et les courbes en bleu présentent les CDF des températures Météosat (CDF SEVIRI BRS). . . . .	67
4.11	Séries temporelles des températures MODIS agrégées à l'échelle de Météosat et des températures Météosat pour le pixel le plus homogène de sol nu et le pixel le plus homogène de couvert arboré sur les 7 mois de données de l'année 2009. . . . .	70
4.12	Séries temporelles des températures MODIS agrégées à l'échelle de Météosat et des températures Météosat pour le pixel le plus homogène d'eau et un pixel Météosat mixte sur les 7 mois de données de l'année 2009. . . . .	71
4.13	Histogrammes des biais moyens pour les quatre classes majoritaires à la RS de Météosat. . . . .	72
4.14	Variation des biais moyen des quatre classes majoritaires à la RS de Météosat en fonction de la date et de l'heure d'acquisition. . . . .	72
5.1	Problème de désagrégation dans le cas d'un pixel BRS à $F$ classes . . . . .	73
5.2	Schéma descriptif de l'opérateur d'observation $\mathcal{H}$ . . . . .	75
5.3	Schéma descriptif de la procédure de désagrégation par lissage particulière. . . . .	76
6.1	Study area : Crau-Camargue Region . . . . .	109
6.2	Landsat 7 7-class land cover map (30m $\times$ 30m spatial resolution). . . . .	111
6.3	Cumulative Distribution Functions (CDF) estimated for Meteosat and Landsat 7 LSTs for the seven validation days over the year 2009 : magenta and green curves represent respectively 30m and LSR Landsat 7 LST CDFs however blue curves represent Meteosat LST CDFs. . . . .	112
6.4	Cumulative Distribution Functions (CDF) estimated for Meteosat and MODIS LSTs for the six of Landsat 7 validation dates over the year 2009 : magenta and green curves represent respectively 1km and LSR MODIS LST CDFs however blue curves represent Meteosat LST CDFs. . . . .	113

6.5	Histograms of bias evaluated for the classes of the most homogeneous Meteosat pixels. . . . .	114
6.6	Bias evaluated for the classes corresponding to the most homogeneous Meteosat pixels. . . . .	115
6.7	Error statistics map versus Meteosat spatial resolution cover fraction rates for bare soil class : the colored scale corresponds to the error statistic value in [K] and the gray scale corresponds to the cover fraction of the bare soil class in each Meteosat pixel. . . . .	116
6.8	HSR GPS downscaling results versus HSR Landsat 7 data and LSR Meteosat data over the Landsat 7 extraction zone. . . . .	117




---

## Liste des tableaux

2.1	<b>Algorithme du filtre de Kalman d'Ensemble . . . . .</b>	<b>19</b>
2.2	<b>Algorithme du filtre particulaire . . . . .</b>	<b>23</b>
3.1	Paramètres du modèle SETHYS . . . . .	46
3.2	Paramètres communs aux différents types d'occupation des sols . . . . .	47
3.3	Paramètres variables selon le type d'occupation des sols . . . . .	47
3.4	Résultats de l'analyse de Sobol effectuée sur la région Crau-Camargue avec les données de forçage de l'année 2009 . . . . .	48
4.1	Températures moyennes et écarts-types des classes à regrouper (en ° C) . . .	55
4.2	Températures moyennes et écarts-types des classes après la nouvelle classification (en ° C). . . . .	55
4.3	Températures moyennes et écarts-types des classes à regrouper (en ° C) pour les différentes dates d'acquisition Landsat7 en 2009. . . . .	56
4.4	Résumé de l'ensemble des données de validation disponibles sur la région d'étude en 2006 et en 2009 . . . . .	62
6.1	Description of the selected parameters for the considered land cover types .	110
6.2	Experimental settings for the different experiments . . . . .	111
6.3	Prior and GPS downscaling results averaged over the whole region and all the validation dates. . . . .	111
6.4	Prior results by class (averaged over MODIS homogeneous pixels and all the validation dates). . . . .	112
6.5	GPS downscaling results by class (averaged over MODIS homogeneous pixels and all the validation dates). . . . .	118







---

# Liste des abréviations

AS	Analyse de Sensibilité
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BRS	Basse Résolution Spatiale
BRT	Basse Résolution Temporelle
CDF	Cumulative Density/Distribution Function
DA	Data Assimilation
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
GPF	Genetic Particle Filter
GPS	Genetic Particle Smoother
HRS	Haute Résolution Spatiale
HRT	Haute Résolution Temporelle
HSR	High Spatial Resolution
HTR	High Temporal Resolution
LAI	Leaf Area Index
LSR	Low Spatial Resolution
LST	Land Surface Temperature
LTR	Low Temporal Resolution
MAE	Mean Absolute Error
MODIS	Moderate Resolution Imaging Spectroradiometer
MODTRAN	MODerate resolution atmospheric TRANsmission
MSG	Meteosat Second Generation
NDVI	Normalized Difference Vegetation Index
PDF	Probabiliy Density/Distribution Function
PF	Particle Filter
PS	Particle Smoother
RANSAC	RANdom SAmples Consensus
RMSE	Root Mean Square Error
SETHYS	Suivi de l'ETat HYdrique des Sol
SEVIRI	Spinning Enhanced Visible and Infrared Imager
TIR	Thermal InfraRed
TsHARP	Temperature sHARPening
TU	Temps Universel



---

# Introduction

LA TEMPÉRATURE DE SURFACES CONTINENTALES (*Land Surface Temperature LST*) encore connue sous le nom de “Température de peau” (*Skin Temperature*) est une variable météorologique très importante car elle permet l'accès aux bilans d'énergie et d'eau régissant les interactions Biosphère-Atmosphère.

Grâce aux développements technologiques dans le domaine de la télédétection infrarouge, il est possible d'obtenir une mesure de la température radiative de la surface de la Terre depuis l'espace. En effet, un radiomètre embarqué sur un satellite mesure le rayonnement thermique de la Terre au sommet de l'atmosphère qui, après correction des effets atmosphériques, permet de calculer la luminance<sup>1</sup> au niveau du sol. La diffusion par les aérosols étant négligeable en infrarouge thermique, les interactions du rayonnement avec l'atmosphère se résument à de l'absorption due principalement à la présence de la vapeur d'eau dans l'atmosphère.

Ces effets peuvent toutefois être minimisés en se plaçant dans des régions spectrales où l'atmosphère est particulièrement transparente au rayonnement infrarouge thermique (IRT) (par exemple les bandes spectrales 11 et 12 $\mu\text{m}$  où l'absorption due à la vapeur d'eau est particulièrement faible). Après correction des effets atmosphériques, la température radiométrique de la surface se déduit de la luminance au niveau du sol par l'intermédiaire de l'émissivité et de la fonction de Planck à la longueur d'onde considérée.

Bien que le principe de la mesure par télédétection IRT semble simple à réaliser, l'interprétation des mesures de LST est loin de l'être. En effet, la surface bidimensionnelle observée par un satellite sous un angle donné, peut contenir en réalité des structures 3D, dues à la présence de plusieurs types de couverts végétaux (forêt, sol nu, blé, etc.) produisant des interactions d'énergies émises par les différents éléments de la surface observée (effets d'ombres non pris en compte).

---

1. La luminance énergétique (radiance),  $L$ , d'une surface est égale à la quantité d'énergie électromagnétique rayonnée par cette surface (par réflexion et/ou par émission propre), par unité de temps, par unité de surface perpendiculaire à la direction d'observation et par unité d'angle solide. L'unité est donc le Watt par mètre carré par stéradian ( $\text{w}/\text{m}^2/\text{st}$ ). La luminance d'une surface peut être calculée pour une longueur d'onde  $\lambda$  dans une bande spectrale  $d\lambda$  comme étant la quantité d'énergie électromagnétique rayonnée par cette surface dans la bande spectrale considérée et dont l'unité est le Watt par mètre carré par stéradian par micromètre ( $\text{w}/\text{m}^2/\text{st}/\mu\text{m}$ ). Dans ce cas on parle de luminance spectrale (Becker (2011)).

Le signal mesuré par le satellite est le plus souvent un signal composite des signaux émis par les différents éléments de la surface observée et dépend donc de plusieurs facteurs (angle de visée, nature des couverts, altitude, etc.). Il est donc nécessaire de séparer ces différents signaux afin d'obtenir une mesure représentative de chaque écosystème présent dans la surface en question. A l'exception d'un milieu isotherme, homogène et en équilibre thermodynamique, le terme "température" est aussi mal défini que le terme "surface" à laquelle il est censé se rapporter (Becker and Li (1995); François (1995)).

Une estimation exacte de la LST est un défi qui relève de la gageure vu la multiplicité et la complexité des processus pouvant se produire lors de son acquisition. Au cours des trente dernières années, de nombreuses recherches ont été menées pour interpréter les mesures de la LST IRT et pour l'utiliser dans le domaine de la modélisation de la biosphère continentale. En effet, la LST est une variable importante car elle permet de contrôler les échanges de flux de chaleur latente et de flux de chaleur sensible entre la biosphère et l'atmosphère (Aires et al. (2001); Sun (2003); Anderson et al. (2008); Coudert et al. (2008)) et d'accéder aux processus hydrologiques (Sandholt et al. (2002); Crow and Wood (2003)). Les applications utilisant les données de LST sont le plus souvent liées au suivi des bilans d'énergie et d'eau et des variables de surface. On peut citer à titre d'exemple l'utilisation de la LST pour le suivi de l'humidité du sol via l'estimation de l'inertie thermique (Aires et al. (2001)) et l'estimation de l'évapotranspiration (Kalma et al. (2008)). Toutefois, la LST peut être utilisée dans d'autres applications telles que le suivi du phénomène d'îlot de chaleur urbaine (Voogt and Oke (2003); Weng et al. (2004)) ou la détection des feux de forêt (Dennison and Matheson (2011)).

Compte tenu de la haute variabilité spatiale et temporelle de la LST, liée principalement à l'hétérogénéité des surfaces observées et aux effets d'absorption atmosphérique, il est important d'avoir accès à cette variable à des résolutions spatiales et temporelles assez fines afin d'assurer un meilleur suivi des différents processus régissant le continuum Biosphère-Atmosphère. Malheureusement, les radiomètres actuellement disponibles permettent d'acquérir la LST soit à haute résolution spatiale mais à de faibles fréquences de passage soit à haute résolution temporelle mais à des résolutions spatiales kilométriques. De ce fait, deux solutions peuvent se présenter. La première solution consiste à utiliser les données HRS/BRT et à interpoler la variation temporelle de la LST. La deuxième solution consiste plutôt à désagréger les données très fréquentes et abondantes de cette variable disponibles à BRS/HRT. Compte tenu de la faible disponibilité de données HRS, la première solution est difficile à mettre en œuvre. La solution la plus séduisante est donc la désagrégation des séries temporelles de températures BRS/HRT.

La désagrégation de la LST (*disaggregation*), encore connue sous les appellations descente en échelle (*downscaling*), et raffinement thermique (*thermal sharpening*) a fait l'objet de plusieurs travaux de recherches qui peuvent être classés en deux grandes familles : la désagrégation basée sur les méthodes de régression et la désagrégation basée sur les méthodes d'assimilation de données. Dans la première famille de solutions on cherche à développer/exploiter des relations (physiques/empiriques) existant entre les données BRS/HRT de la LST et des données HRS d'autres variables. Dans la deuxième famille de solutions, qui fait appel aux méthodes d'assimilation de données, on vise à assimiler les signaux de températures BRS/HRT dans des modèles (physiques/numériques) capables de reproduire la variabilité de la température à HRS.

Nous présentons par la suite, un bref état de l’art des méthodes développées pour chaque famille et nos motivations dans le choix de la méthode de désagrégation développée au cours de cette thèse.

## 1.1 Hypothèses générales

Avant de passer à la présentation des hypothèses générales du problème de désagrégation, il est important de bien préciser la définition du terme “élément HRS” qui sera utilisé dans ce qui suit. En effet, un élément HRS désigne l’entité de surface supposée homogène. Selon [Zhan et al. \(2013\)](#) quatre hypothèses doivent être prises en considération pour parler de désagrégation :

- la surface du pixel à désagréger est assez grande pour contenir plus qu’un élément HRS. Aussi triviale qu’elle soit, cette hypothèse est très importante car sinon le problème de désagrégation n’a aucun sens.
- la surface d’un élément contenu dans un pixel BRS est supposée isotherme. En réalité, cette hypothèse n’est pas vraie à cause de la haute variabilité spatiale de la température même à la résolution finale de la désagrégation. En effet, à l’échelle de l’élément HRS, la surface présente toujours des hétérogénéités qui font que la température n’est pas tout à fait la même en tout point de la surface de l’élément HRS. Cependant, cette hypothèse est très importante car elle permet de réduire la complexité du problème de désagrégation vu qu’elle permet de séparer les différents éléments HRS inclus dans le pixel BRS. Cette hypothèse peut toutefois être partiellement levée en prescrivant une distribution à la LST à l’échelle du pixel HRS. Selon [Zhan et al. \(2013\)](#) cette hypothèse est connue sous le nom d’hypothèse de “*séparabilité*”.
- la disponibilité des données auxiliaires à HRS. Il est extrêmement important de disposer de données auxiliaires à HRS car elles permettent de réduire le nombre d’inconnues du problème de désagrégation. Ces données peuvent être par exemple, des cartes d’occupation des sols permettant d’identifier le nombre d’éléments HRS au sein du pixel BRS et/ou les types de couverts présents, etc. Cette hypothèse est connue sous le nom d’hypothèse de “*connectivité*”.
- l’additivité de l’agrégation. Cette hypothèse suppose que l’agrégation est linéaire et qu’il n’y a pas d’interaction d’énergie entre les différents éléments HRS au sein du pixel BRS. Cette hypothèse n’est pas toujours vérifiée. En effet, si les surfaces présentes dans le pixel mixte BRS possèdent des structures 3D non prises en compte à la longueur d’onde de mesure, l’agrégation n’est plus linéaire à cause des effets d’ombres. Cette hypothèse est connue comme l’hypothèse d’ “*additivité*”.

## 1.2 État de l'art sur la désagrégation de la LST par les méthodes de régression

L'approche la plus connue pour le raffinement d'échelle par les méthodes de régression est celle qui exploite la relation négative entre la LST et la densité de végétation. Celle-ci étant généralement estimée à partir des données de la différence normalisée de l'indice foliaire connue sous le nom de NDVI (Normalized Difference Vegetation Index) et/ou la fraction de couvert  $f_{cover}$  (Choudhury et al. (1994)) dont les expressions sont décrites respectivement par les équations 1.1 et 1.2.

$$NDVI = \frac{\rho_{PIR} - \rho_R}{\rho_{PIR} + \rho_R}, \quad (1.1)$$

$$f_{cover} = 1 - \left( \frac{NDVI_{max} - NDVI}{NDVI_{max} - NDVI_{min}} \right)^{0.625}, \quad (1.2)$$

avec  $\rho_R$  et  $\rho_{PIR}$  désignent respectivement les réflectances dans la bande spectrale rouge (0.62-0.78  $\mu\text{m}$ ) et dans la bande spectrale proche infrarouge (0,78-3  $\mu\text{m}$ ) et  $NDVI_{min}$  et  $NDVI_{max}$  désignent respectivement la valeur minimale et la valeur maximale du NDVI estimées sur l'image considérée.

Dans le cas où  $NDVI_{min} = 0$  et  $NDVI_{max} = 1$  cette relation se réduit à :

$$f_{cS} = 1 - (NDVI)^{0.625}, \quad (1.3)$$

avec  $f_{cS}$  désigne le  $f_{cover}$  simplifié. Ces indices sont d'autant plus importants que la végétation est dense et photosynthétiquement active.

En effet, quand une surface est couverte de végétation dense et photosynthétiquement active, l'exposition de cette surface à l'énergie solaire augmente le phénomène d'évapotranspiration. La transpiration des plantes refroidit ainsi la température de cette surface contrairement à une surface de sol nu où la température serait plus élevée (faible évapotranspiration). Il apparaît ainsi une relation négative entre les indices de végétation et la LST : plus la végétation est importante et photosynthétiquement active moins la température est élevée et inversement.

Le premier modèle qui exploite la relation entre le NDVI et la LST est le modèle DisTrad développé par Kustas et al. (2003). Ce modèle suppose que la relation entre la température et le NDVI est unique et peut se ramener à une régression polynomiale d'ordre 2. Le principe est simple et suppose qu'à la résolution de la température radiative, la relation liant le NDVI (agrégé à la résolution spatiale de la température) et la température est parfaitement décrite par une régression des moindres carrés tel que :

$$T_R^*(NDVI_{BRS}) = f(NDVI_{BRS}), \quad (1.4)$$

où  $T_R^*$  est la température estimée par la régression des moindres carrés et  $NDVI_{BRS}$  est le NDVI à BRS.

Cette relation est ensuite appliquée au NDVI à sa résolution initiale ( $NDVI_{HRS}$ ) pour estimer la température à fine résolution spatiale. Cette version du modèle a été améliorée par Agam et al. (2007b) qui utilise l'indice  $f_{cover}$  à la place du NDVI. La nouvelle version est connue sous le nom de  $TsHARP$ .

Selon [Agam et al. \(2007b\)](#), la version de *TsHARP* basée sur la relation linéaire entre le  $f_{cS}$  et la température est la plus recommandée. Selon eux, ceci s'explique en grande partie par la linéarité de la relation entre le  $f_{cS}$  et la température radiative (Eq. 1.3).

Pour améliorer encore l'algorithme de *TsHARP*, [Merlin et al. \(2010\)](#) proposent de différencier les végétations photosynthétiquement actives et non photosynthétiquement actives dans la régression de la température. En effet, les végétations non vertes ne transpirent pas comme les végétations vertes et la relation liant la température et l'indice de végétation est forcément différente selon l'activité photosynthétique. La différenciation a permis d'améliorer l'estimation de la température radiative par rapport à la version classique de *TsHARP*. Cependant, comparées aux données de validations HRS, l'erreur de désagrégation est toujours importante (de l'ordre de 3K).

Même si les algorithmes basés sur la relation entre la LST et le NDVI/ $f_{cover}$  ont montré de bonnes performances sur certains sites caractérisés par des couverts assez homogènes et denses, ces performances vont vite se dégrader sur des sites plus étendus et dont les couverts sont épars et hétérogènes ([Karnieli et al. \(2006\)](#); [Sun and Kafatos \(2007\)](#); [Inamdar and French \(2009\)](#); [Jeganathan et al. \(2011\)](#); [Chen et al. \(2011\)](#)). Selon [Inamdar et al. \(2008\)](#), ceci serait dû à une mauvaise interprétation des valeurs de NDVI sur les régions où la canopée est très éparse et hétérogène car les mesures ont une signature composite du sol et des végétations et la combinaison des deux signatures est souvent non-linéaire.

A cela s'ajoute le fait que la corrélation négative entre la LST et le NDVI/ $f_{cover}$  n'est pas toujours vérifiée, surtout en présence de structures 3D assez importantes (forêt, montagnes, etc) et en dehors de la période de l'après-midi où la corrélation entre le NDVI et la LST devient très faible ([Sun and Kafatos \(2007\)](#)). Une autre source d'erreur serait, selon [Kustas et al. \(2003\)](#); [Agam et al. \(2007b\)](#) et [Merlin et al. \(2010\)](#), la conséquence de la variabilité spatiale de la température radiative liée à la variation de l'humidité du sol non prise en compte dans les différentes versions de *TsHARP*. D'autre part, la relation entre la LST et le NDVI est mal définie pour les paysages de type urbain limitant ainsi le domaine d'application des approches basées sur cette relation aux régions de type agricole ([Dominguez et al. \(2011\)](#)).

Il apparaît ainsi que la relation entre la LST et les indices de végétation n'est pas suffisante et d'autres variables doivent être considérées pour améliorer l'estimation de la LST à l'aide des données auxiliaires à HRS.

Ainsi, [Inamdar and French \(2009\)](#) ont introduit l'émissivité des couverts végétaux comme étant une variable alternative pour désagréger les températures GOES (Geostationary Operational Environmental Satellite) sur la région Sud-Est des États Unis. L'émissivité étant directement liée à la LST dans l'équation de transfert radiatif, cette variable possède une influence directe sur la température de brillance observée par les satellites en IRT. De ce fait, l'estimation de la LST à partir des données d'émissivité présente une approche plus "physique" que la relation empirique LST-NDVI/ $f_{cover}$ . La comparaison entre les résultats obtenus pour la désagrégation des températures des cycles journaliers en moyennes mensuelles (à partir des données GOES à 5km) sur la région Sud-Est des États Unis en utilisant des données MODIS d'émissivité à 1km a montré de meilleures performances que la méthode basée sur la relation empirique entre la LST et le NDVI avec une amélioration des biais de  $-4K$  à  $-0.6K$  et des RMSE de  $2.4K$  à  $0.8K$  ([Inamdar and French \(2009\)](#)).



Toujours dans le cadre de l'amélioration de la régression entre la LST et des variables auxiliaires, [Merlin et al. \(2012\)](#) proposent d'introduire en plus de la différenciation entre les fractions des végétations vertes et sénescentes, les fractions de surface d'eau libre (lac, bassin, etc.) et d'une estimation de l'état hydrique du sol fournie par des mesures micro-ondes. Le principe de la méthode repose sur une linéarisation de l'équation de transfert radiatif qui distingue la température du sol nu, la température des végétations et la température de l'eau. Cette nouvelle approche a permis de désagréger des températures ASTER agrégées à l'échelle de données MODIS (1km) à 250m de résolution spatiale sur une zone de surface 160km<sup>2</sup> caractérisée par une forte activité d'irrigation. Les températures ont été estimées avec une erreur de 1.16K comparées aux températures ASTER de résolution spatiale 250m. Cependant, la validité de l'approche est limitée à des régions fortement irriguées où le stress hydrique des végétations est négligeable. Dans le cas contraire, le problème de désagrégation serait partiellement indéterminé à cause de la non disponibilité des données qui permettent de différencier les végétations en état de stress hydrique de celles qui ne le sont pas ([Merlin et al. \(2012\)](#)). D'autre part, cette approche reste très liée à la disponibilité des mesures fréquentes des variables auxiliaires en particulier les données micro-ondes aéroportées.

Ainsi, les méthodes exploitant la corrélation entre la LST et les indices de végétation, l'émissivité et d'autres variables auxiliaires ont permis de résoudre le problème de désagrégation pour plusieurs types de couverts et à différentes échelles (échelle de la parcelle agricole, échelle régionale, etc.). Cependant, ces méthodes présentent des limitations liées soit à la non prise en compte des états hydriques des végétations, soit à la non-linéarité de la relation en présence de fortes hétérogénéités spatiales et/ou des structures 3D (signature composite de plusieurs signaux IRT) soit à la disponibilité des données auxiliaires HRS (suffisantes et de qualité). Nous verrons par la suite que cette dernière limitation, liée à l'hypothèse de connectivité est légèrement levée par les méthodes de désagrégation basées sur l'assimilation/inversion des données dont la description fait l'objet de la section suivante.

### 1.3 État de l'art sur la désagrégation de la LST par les méthodes d'assimilation/inversion de données

Afin d'être moins dépendant de la disponibilité des données HRS pour réaliser la désagrégation de la température et résoudre le problème de l'interpolation temporelle, une solution consiste à utiliser un modèle permettant de simuler les températures et leurs dynamiques temporelles à la résolution spatiale de la désagrégation.

Les modèles de Transfert-Sol-Végétation-Atmosphère TSVA présentent un cadre adéquat pour décrire les bilans d'énergie et d'eau régissant le continuum Biosphère-Atmosphère. Ces modèles décrivent l'évolution temporelle des flux d'énergie et des variables du sol en particulier l'humidité du sol et la LST. Dans ce contexte, des méthodes de désagrégation ont été proposées basées sur l'inversion statistique ou l'assimilation des températures BRS dans des modèles TSVA.

L'un des premiers avantages de ce type de méthode est sa faible dépendance aux données auxiliaires (la variabilité spatiale et temporelle étant décrite dans les équations du modèle). Il suffit alors d'avoir une carte d'occupation du sol à la résolution spatiale désirée pour réaliser la désagrégation des températures BRS. Le principe est assez simple et peut être décrit comme suit.

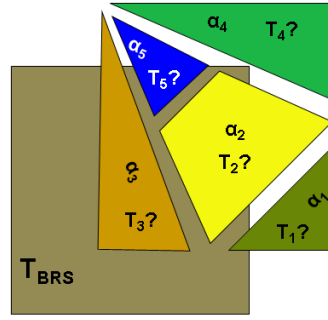


FIGURE 1.1 – Schématisation du problème de désagrégation à l'échelle d'un pixel à basse résolution spatiale.

Supposons que l'élément HRS est la classe de végétation et que l'on désire désagréger la température  $T_{BRS}$  d'un pixel BRS dont l'occupation des sols est définie par les fractions  $\alpha_i$  des différentes classes de végétation présentes dans ce pixel (voir Figure 1.1). Le problème de désagrégation se rapporte donc à l'estimation des températures sous-pixelliques  $T_i$  relatives à chaque type de couvert présent dans ce pixel mixte. D'après la loi de Stefan-Boltzmann et compte-tenu des hypothèses de séparabilité et d'additivité, il est possible d'écrire que :

$$M = \epsilon \sigma T_{BRS}^4 = \frac{\sum_i \alpha_i M_i}{\sum_i \alpha_i}, \quad (1.5)$$

tel que  $\sigma$  désigne la constante de Stefan-Boltzmann,  $\{\epsilon, M\}$  désignent respectivement l'émissivité et l'énergie émise par la surface du pixel mixte et  $\{\epsilon_i, M_i\}$  désignent respectivement l'émissivité et l'énergie émise par un élément de surface HRS  $i$ . Etant donné que la somme des différentes fractions  $\alpha_i$  pour un pixel BRS est égale à l'unité, il est possible d'écrire :

$$\epsilon T_{BRS} = \left( \sum_i \alpha_i \epsilon_i T_i^4 \right)^{\frac{1}{4}}. \quad (1.6)$$

Selon Liu et al. (2006), l'équation 1.6 peut être simplifiée comme suit :

$$T_{BRS} = \sum_i \alpha_i T_i + \zeta, \quad (1.7)$$

où l'incertitude  $\zeta$  est supposée être inférieure à 0.1K. Cette représentation du problème de désagrégation est équivalente à résoudre un problème d'inversion statistique de la température BRS. Cette approche a été en premier introduite par Otlé et al. (2008) où  $\zeta$  a été remplacé par un bruit blanc gaussien. Selon Otlé et al. (2008), le problème d'inversion peut être réalisé à l'échelle du pixel dans le cadre d'une approche bayésienne connaissant des solutions *a priori* des températures sous-pixelliques  $T_i$  fournies par un modèle TSVA. En supposant que ces températures sous-pixelliques peuvent être décrites par des variables aléatoires gaussiennes, résoudre le problème d'inversion revient à minimiser une fonction de coût au moyen d'un estimateur de maximum de vraisemblance qui est la suivante :

$$J = \frac{1}{\sigma_b^2} \sum_i (T_i - T_{bi})^2 + \frac{1}{\sigma_T^2} \left( \sum_i \alpha_i T_i - T_{BRS} \right)^2, \quad (1.8)$$

où  $T_{bi}$  sont les températures *a priori* des différents éléments du pixel BRS simulées par le modèle TSVA et  $\sigma_b$  et  $\sigma_T$  sont les écarts-types des incertitudes liées respectivement

aux solutions *a priori* et aux observations. La résolution du problème de désagrégation revient ainsi à la résolution du système d'équations linéaires obtenu par la dérivation de la fonction de coût  $J$ .

Les performances de l'approche proposée dépendent strictement de l'amplitude des incertitudes sur les observations et les températures *a priori*. Une manière de réduire les incertitudes liées aux observations serait de considérer tous les pixels BRS de la région qu'on souhaite désagréger. Cela suppose donc que les températures sous-pixelliques soient stationnaires en espace. De cette manière la fonction  $J$  peut être écrite comme suit :

$$J = \frac{1}{\sigma_b^2} \sum_i (T_i - T_{bi})^2 + \sum_{k=1}^K \frac{1}{\sigma_T^2} \left( \sum_i \alpha_i(k) T_i - T_{BRS}(k) \right)^2, \quad (1.9)$$

où  $K$  désigne le nombre total des pixels BRS à désagréger.

L'application de la méthode sur une base de données synthétiques a montré de bonnes performances pour l'estimation des températures sous-pixelliques. Cependant, ces performances sont très sensibles à l'amplitude des incertitudes sur les observations et au nombre de sous-pixels présents dans le pixel BRS. Une autre limitation de cette approche est l'hypothèse de stationnarité spatiale qui n'est pas vraie à cause de la haute variabilité spatiale de la température liée aux changements des conditions hydriques et phénologiques qui varient d'un pixel BRS à l'autre (Ottlé et al. (2008); Kallel et al. (2013)).

Afin de lever partiellement cette hypothèse de stationnarité, Kallel et al. (2013) proposent d'utiliser les champs aléatoires de Markov (*Markov Random Fields*, MRF) pour décrire la distribution des données BRS de la LST en supposant que son évolution temporelle est approchée par une chaîne aléatoire de Markov et que l'image BRS est soit un champ stationnaire, soit un MRF. Ceci revient à rajouter deux termes de régularisation dans la fonction de coût (Eq. 1.8), liés à la prise en compte des corrélations spatiales et temporelles de la température.

Ces hypothèses se justifient selon Kallel et al. (2013) par le fait que les caractéristiques physiques de l'atmosphère, des végétations et du sol varient lentement à l'échelle du pixel BRS. Cette approche a été comparée à l'approche d'inversion proposée par Ottlé et al. (2008) sur une base de données synthétiques et a montré de meilleures performances en terme de réduction de l'erreur quadratique moyenne  $RMSE$  par rapport à une simulation de référence. L'application de cette méthode pour la désagrégation d'une série temporelle Meteosat 8 (3km  $\times$  5km de résolution spatiale) à la résolution spatiale de ASTER (90m) sur une zone de la région de Crau-Camargue au Sud-Est de la France a montré de meilleures performances comparées à la version classique de TsHARP et à l'inversion simple proposée par Ottlé et al. (2008) en terme de réduction du RMSE par rapport à l'image ASTER de validation :  $RMSE_{TsHARP} = 2.5$  K,  $RMSE_{Inversion} = 1.81$  K et  $RMSE_{MRF} = 1.63$  K.

Même si l'inversion par MRF proposée par Kallel et al. (2013) a permis de lever partiellement l'hypothèse de stationnarité spatiale, elle suppose d'une manière ou d'une autre que la variabilité spatiale de la température est faible entre les différents pixels BRS. Cette hypothèse forte n'est pas toujours valide. De plus, si on désire prendre en compte plus de pixels pour mieux contraindre la variabilité spatiale de la température pour des distances de corrélation plus grandes (plus de pixels pris en compte pour décrire la corrélation spatiale et temporelle) on aboutit à des inversions de matrices très

grandes qui peuvent vite devenir ingérables. D'autre part, cette approche suppose qu'il existe une solution unique au problème de désagrégation. Cependant, tel qu'il est posé, il s'agit plutôt d'un problème d'équi-finalité car plusieurs valeurs des températures sous-pixelliques peuvent être solutions au problème de désagrégation. Il serait donc intéressant de proposer une approche permettant d'éliminer complètement l'hypothèse de stationnarité spatiale et de prendre en compte toutes les solutions possibles de désagrégation.

L'assimilation de données présente un cadre mathématique approprié pour résoudre ce type de problème. Selon [Montzka et al. \(2012\)](#), le problème de désagrégation pourrait se rapporter à un problème d'assimilation de données uni-variable/multi-résolutions UVMS (UniVariate MultiScale data assimilation) et les méthodes d'ensemble présentent un cadre approprié pour résoudre ce genre de problèmes. C'est dans ce cadre que se situe ce travail de thèse.

Il s'agit donc de **développer une nouvelle méthode de descente en échelle spatiale** basée sur l'assimilation des observations BRS de la LST dans un modèle de surface continentale TSVA (Sol-Végétation-Atmosphère), **afin de déterminer les températures des différents types d'occupation des sols contenus dans un pixel BRS et d'obtenir un jeu de paramètres optimisé pour chacun des types d'occupation des sols considérés.**

### 1.4 Motivations pour le choix de l'approche de désagrégation

Comme présenté précédemment, les problèmes rencontrés par les différentes approches de désagrégation relèvent principalement de deux sources :

- la non-linéarité de la relation entre la température et la/les variables auxiliaires (méthodes de régression).
- l'hypothèse de stationnarité ou l'hypothèse de la faible variabilité spatiale des températures sous-pixelliques à BRS (inversion statistique).

La méthode proposée dans cette thèse est basée sur l'assimilation des observations BRS dans un modèle TSVA plus précisément le modèle SETHYS (Suivi de l'Etat Hydrique du Sol) qui a déjà été utilisé pour la désagrégation dans les travaux de [Ottlé et al. \(2008\)](#) et [Kallel et al. \(2013\)](#) et dont les performances ont été testées dans plusieurs travaux sur différents types de climats et de couverts ([Coudert \(2006\)](#); [Coudert et al. \(2006\)](#); [Coudert and Ottlé \(2007\)](#); [Coudert et al. \(2008\)](#); [Saux-Picart et al. \(2009a,b\)](#)).

Cette approche permet de se libérer des hypothèses de stationnarité locale, souvent utilisées pour tirer parti de l'information observée dans les pixels voisins du pixel considéré. En réalité, ces hypothèses ne sont pas nécessairement justifiées en raison de l'hétérogénéité forte de la température d'un pixel à l'autre qui est généralement liée à l'hétérogénéité des propriétés optiques et hydriques du couvert végétal d'un pixel à un autre voir même au sein d'un même pixel ([Ottlé et al. \(2008\)](#); [Kallel et al. \(2013\)](#)). Un autre avantage de notre approche est qu'elle permet d'avoir non pas une solution unique mais plutôt un ensemble de solutions possibles grâce au cadre de l'assimilation de données d'ensemble. Les méthodes d'assimilation de données d'ensemble les plus connues sont le filtre de Kalman d'Ensemble (EnKF) et le filtre particulaire (PF). Notre choix de la méthode a été confirmé par les travaux de [Hill et al. \(2011\)](#) qui a utilisé le PF pour la désagrégation des données de LAI et qui a montré de bonnes performances. Il aurait été possible d'envisager un EnKF, mais nous avons choisi d'innover en testant

une méthode de lissage particulière.

L'approche que l'on propose utilise SETHYS pour simuler un ensemble de températures sous-pixelliques des éléments de surface présents dans un pixel BRS en faisant varier un ensemble de paramètres de SETHYS. Le PF va donc sélectionner les températures sous-pixelliques qui, agrégées à l'échelle du pixel BRS, sont les plus proches des observations. Le résultat est donc l'ensemble formé par les températures sous-pixelliques sélectionnées par le PF et les valeurs des paramètres correspondant à ces températures. Un dernier avantage est la production de données de température continues, interpolées par le modèle physique et donc prenant en compte implicitement les corrélations temporelles du signal thermique. De plus, la méthode donne l'accès à d'autres variables de sorties liées à la température comme par exemple l'évapotranspiration.

Le deuxième chapitre présente en détails le cadre mathématique de l'assimilation de données d'ensemble en décrivant le principe de l'approche bayésienne d'assimilation de données et les équations des filtres EnKF et le PF. Le troisième chapitre détaille les équations du modèle SETHYS et l'analyse de sensibilité réalisée sur ce modèle pour choisir les paramètres à étalonner dans la procédure de désagrégation. Le quatrième chapitre, présente la région d'étude et les données de forçage, d'observation et de validation utilisées dans le cadre de mon travail de thèse. Le chapitre 5 présente l'application de la méthode de désagrégation sur des pseudo-observations pour évaluer dans un cadre synthétique les performances de la méthode proposée et ensuite sur des données réelles de télédétection acquises sur la région d'étude. Les résultats encourageants de cette première étude ont conduit à élargir la procédure de désagrégation sur une région d'étude plus grande et sur une période d'assimilation plus longue. Les performances de la méthode de désagrégation proposée ont été évaluées en termes de réduction des écarts par rapport aux données de validation. Le chapitre 7 résume les principaux résultats obtenus au cours de cette thèse, accompagné d'une discussion générale et de suggestions quant à l'extension des présents travaux dans de futurs axes de recherche.

---

# Assimilation de données

CE CHAPITRE présente les différentes composantes méthodologiques du travail effectué au cours de cette thèse. Une brève introduction à l'assimilation de données est d'abord présentée en Section 2.1. Le formalisme de l'assimilation de données est ensuite décrit en section 2.2 en particulier le problème de filtrage qui s'inscrit dans le cadre d'un formalisme bayésien. Les sections 2.3 et 2.4 présentent respectivement des solutions du problème de filtrage dans le cas d'un modèle linéaire et non linéaire. La première partie de la section 2.4 présente une approximation pragmatique du problème de filtrage non linéaire dans le cadre d'hypothèse d'erreurs d'observation gaussiennes, le filtre de Kalman d'ensemble. La deuxième partie de la section 2.4 décrit l'approximation particulière du filtrage non linéaire adapté aux dynamiques non linéaires, avec en fin de section une description de la méthode d'étalonnage/désagrégation par lissage particulière développée au cours de cette thèse.

## 2.1 Introduction à l'assimilation de données

L'assimilation de données est "*l'ensemble des techniques statistiques qui permettent d'améliorer notre connaissance (passée, présente ou future) de l'état d'un système physique à partir de la donnée concomitante d'observations expérimentales sur le système et de sa connaissance théorique (ou a priori)*"<sup>1</sup>. Une première approche consiste à faire évoluer l'information dont on dispose sur le modèle à chaque nouvelle observation. Cette approche est connue sous le nom de **filtrage séquentiel** et sa formulation dynamique a été donnée par Kalman (1960) dans le cadre linéaire et gaussien. La figure 2.1, illustre de manière schématique cette dynamique où les observations permettent de forcer un modèle, dont la tendance naturelle serait de s'écarter, à long terme, de la réalité qu'il doit décrire.

En pratique, les observations assimilées de manière séquentielle sont rarement prises en compte à l'instant même de leur réalisation. Le plus souvent, on regroupe ces observations sur une période/fenêtre temporelle sous l'hypothèse qu'elles soient toutes représentatives du système étudié. Ceci nous conduit alors à la deuxième approche qui consiste à construire une trajectoire du modèle, la plus proche de l'ensemble des observations prise sur une fenêtre temporelle donnée : il s'agit donc d'un **lissage**.

---

1. "M. Bocquet *Introduction aux principes et méthodes de l'assimilation de données en géophysique, cours de master, 2004-201*.

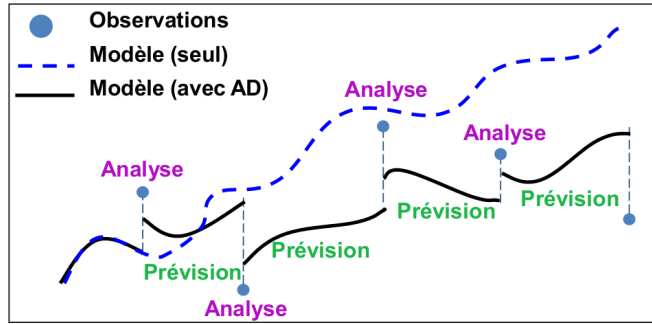


FIGURE 2.1 – Schéma descriptif de l'assimilation de données séquentielle.

L'avantage de cette technique est qu'elle permet de documenter une partie de l'information non-échantillonnée par les observations en s'appuyant sur la dynamique du modèle (Thompson (1961)). En pratique, c'est l'approche mise en oeuvre dans l'assimilation variationnelle de type 4D-Var ((Le Dimet and Talagrand, 1986; Courtier et al., 1994)). L'approche Bayésienne permet de plonger ces deux approches dans un même formalisme (Jazwinski (1970)).

Le formalisme de l'assimilation de données est maintenant introduit, avec en premier la notion de filtrage non linéaire vue dans le cadre d'une approche Bayésienne. Le filtre de Kalman apparaît comme un cas particulier de filtrage non linéaire sous l'hypothèse de gaussianité des statistiques considérées et de linéarité de la dynamique des modèles de prévision. Ensuite le cadre non linéaire des dynamiques décrites par le formalisme de l'assimilation de données est introduit par le filtre de Kalman d'ensemble et le filtre particulaire. Enfin, le lissage particulaire est détaillé. Ce dernier sera par la suite appliqué dans le cadre de l'estimation des températures sous-pixelliques, encore connu sous le nom de désagrégation des températures et présente le cœur de la méthode développée et mise en oeuvre au cours de cette thèse.

## 2.2 Problème de filtrage et formalisme Bayésien

Le formalisme présenté ici est très général et s'inscrit dans le cadre des processus de filtrage non linéaire. Le problème de filtrage tel que défini précédemment, consiste à estimer l'état d'un système en évolution au cours du temps. L'état du système à l'instant  $q$ , est représenté par un vecteur  $x_q$  de dimension  $N_x$ . La modélisation numérique de l'évolution temporelle du système de l'instant  $q - 1$  à l'instant  $q$  est réalisée à l'aide d'un opérateur de transition  $\mathcal{M}$ , tel que :

$$x_q = \mathcal{M}(x_{q-1}) + W_q, \quad (2.1)$$

où  $W_q$  désigne l'erreur modèle, conceptualisée sous la forme d'un vecteur aléatoire de dimension  $N_x$ . Le suivi temporel du système est réalisé à l'aide d'observations prises à l'instant  $q$ , représentées sous la forme d'un vecteur  $y_q$  de dimension  $N_y$ .  $y_q$  est lié à l'état du système par un opérateur d'observation  $\mathcal{H}$  tel que :

$$y_q = \mathcal{H}(x_q) + V_q, \quad (2.2)$$

où  $V_q$  désigne l'erreur d'observation, conceptualisée sous la forme d'un vecteur aléatoire de dimension  $N_y$ .

L'erreur d'observation regroupe les erreurs de mesures et celles de représentativités associées, par exemple, aux défauts d'interpolation dans l'opérateur  $\mathcal{H}$ . L'objectif de l'assimilation de données est d'estimer  $x_q$  connaissant les observations  $y_q$ . D'un point de vue probabiliste, cette information est caractérisée par la densité de probabilité  $p(x_q|y_q)$ . La formule de Bayes permet d'écrire cette distribution sous la forme :

$$p(x_q|y_q) \propto p(y_q|x_q) p(x_q|y_{q-1}), \quad (2.3)$$

où  $p(x_q|y_{q-1})$  désigne l'information qui provient du passé et déduite de la relation de Chapman-Kolmogorov :  $p(x_q|y_{q-1}) = \int_{x_{q-1}} p(x_q|x_{q-1}) p(x_{q-1}|y_{q-1}) dx_{q-1}$ , où  $p(x_q|x_{q-1})$  correspond à la probabilité de transition pour passer de l'état  $x_{q-1}$  à l'état  $x_q$  entre l'instant  $q-1$  et l'instant  $q$ . Dans cette relation, le symbole  $\propto$  désigne que la distribution est définie à une constante de normalisation près. Il apparaît ainsi une récursivité pour l'estimation de l'état à l'instant  $q$  à partir de l'estimation à l'instant  $q-1$ . Cette récursivité est traduite par un cycle d'analyse et de prévision, l'information des observations aux instants antérieurs étant contenue dans la distribution de l'estimateur. Il est possible de montrer que l'analyse optimale (celle qui minimise l'erreur quadratique moyenne) du processus au temps  $q$  connaissant les observations jusqu'au temps  $q$  est donnée par :

$$x_q^a = \mathbf{E}[x_q|y_q]. \quad (2.4)$$

De même, le meilleur estimateur du processus cherché au temps  $q$  connaissant les observations jusqu'au temps  $q-1$  est donné par (Van Leeuwen and Evensen (1996)) :

$$x_q^b = \mathbf{E}[x_q|y_{q-1}], \quad (2.5)$$

une quantité qui correspond en pratique à ce qu'on nomme l'ébauche.

Il faut alors bien distinguer quatre processus : le processus réel ( $x_q$ ) (noté d'ordinaire  $x^t$  pour "true"), l'observation associée ( $y_q$ ), l'ébauche correspondante ( $x_q^b$ ), et enfin l'état analysé ( $x_q^a$ ). Le schéma global d'un cycle de prévision et d'assimilation de données est composé de deux étapes. Une première étape de prédiction qui permet de déterminer l'ébauche  $x_q^b$  à partir de l'état analysé au temps précédent  $x_{q-1}^a$  et du propagateur associé au modèle. Une seconde étape d'analyse qui permet de réaliser le filtrage à l'aide de la formule de Bayes ; c'est à dire évaluer la quantité  $x_q^a$  à partir de l'ébauche  $x_q^b$  et de l'observation  $y_q$ . Cet enchaînement de prévision et d'analyse peut alors s'écrire

$$\mathbf{p}(\mathbf{x}_q|y_{q-1}) \xrightarrow{\text{Analyse}} \mathbf{p}(\mathbf{x}_q|y_q) \xrightarrow{\text{Prédiction}} \mathbf{p}(\mathbf{x}_{q+1}|y_q) \xrightarrow{\text{Analyse}} \dots$$

En général, il n'y a pas de résolution analytique pour cette dynamique de l'information. Cependant, dans le cas où la dynamique est linéaire et les statistiques sont gaussiennes, il est possible de résoudre explicitement cette dynamique. Il s'agit alors des équations du filtre de Kalman (1960), noté **KF** (*Kalman Filter*). Pour tenter de conserver le bénéfice de cette solution exacte, on peut chercher à étendre le formalisme du filtre de Kalman dans le cas où la dynamique des erreurs est linéaire. Dans ce cas, les états évoluent de manière non linéaire, mais les erreurs se propagent à l'aide du modèle linéaire tangent le long de la trajectoire non linéaire. Il s'agit du filtre de Kalman étendu (Haykin (2001)), noté **EKF** (*Extended Kalman Filter*). En pratique, le coût d'un filtre de Kalman étendu est assez important. En effet, d'après une étude réalisée par Bourgois et al. (2011), les EKF sont de complexité cubique  $O(n^3)$  avec  $n$  la dimension du vecteur d'état pour déterminer les matrices de covariance de l'erreur de prédiction et d'analyse. La mise en œuvre de ce type de filtre devient très lourde pour un vecteur d'état de grande dimension. Pour pallier à ce problème, Evensen (1994a,b) a introduit une méthode sous-optimale de nature stochastique qui décrit la propagation des incertitudes en utilisant un **ensemble de prévisions** comme support : il s'agit du filtre de Kalman d'ensemble (*Ensemble Kalman Filter*, **EnKF**).



Si ces deux algorithmes, *EKF* et *EnKF*, ont pour mérite de proposer un formalisme simple et pragmatique, ils ne correspondent pas exactement à la dynamique de l'information décrite plus haut. Pour parvenir à reproduire précisément cette dynamique, il faut avoir recours à une discrétisation des distributions de probabilité ; il s'agit alors du formalisme de filtrage particulaire *PF* (*Particle Filter*) (Arulampalam et al. (2002)). Afin de préciser ces différentes approches, nous revenons sur leurs formalismes, en commençant par le filtre de Kalman.

### 2.3 Dynamique linéaire et statistiques gaussiennes : le filtre linéaire de Kalman

Avant de préciser le formalisme du filtre de Kalman, essayons de comprendre l'origine de ces hypothèses et comment elles peuvent permettre de retrouver une expression analytique du cycle d'analyse et de prévision ; La grande simplification apportée par l'hypothèse de gaussianité est que l'information probabiliste de  $p(x_{q-1}|y_{q-1})$  se réduit à celle de la connaissance de sa moyenne  $x_{q-1}^a$  et de sa matrice de covariance, notée  $\mathbf{A}_{q-1}$ . Ensuite, le rôle joué par la linéarité de la dynamique est de permettre la conservation du caractère gaussien, ainsi  $p(x_q|y_{q-1})$  est entièrement déterminée par sa moyenne  $x_{q+1}^b$  et sa matrice de covariance  $\mathbf{B}_q$ . À nouveau, l'hypothèse de gaussianité de l'erreur d'observation et de la linéarité de l'opérateur d'observation, permettent de réduire la connaissance de  $p(x_q|y_q)$  à celle de sa moyenne  $x_q^a$  et de sa matrice de covariance  $\mathbf{A}_q$ .

Dans le cas d'une dynamique linéaire et pour des statistiques d'erreurs gaussiennes, la dynamique de  $x_q$  s'écrit :

$$x_q = \mathbf{M}_q x_{q-1} + \mathbf{W}_q, \quad (2.6)$$

où  $\mathbf{M}_q$  est l'opérateur linéaire de prévision à l'instant  $q$  et  $\mathbf{W}_q$  désigne l'erreur modèle représentée sous la forme d'un vecteur aléatoire gaussien de moyenne nulle et de matrice de covariance  $\mathbf{Q}_q$ , ce qu'on écrit pour simplifier  $\mathbf{W}_q \sim \mathcal{N}(0, \mathbf{Q}_q)$ . Le lien entre l'état  $x_q$  et les observations  $y_q$  est décrit par

$$y_q = \mathbf{H}_q x_q + \mathbf{V}_q, \quad (2.7)$$

avec  $\mathbf{H}_q$  l'opérateur d'observation linéaire à l'instant  $q$  et  $\mathbf{V}_q \sim \mathcal{N}(0, \mathbf{R}_q)$  l'erreur d'observation gaussienne supposée de moyenne nulle et de matrice de covariance  $\mathbf{R}_q$ . En supposant que l'erreur modèle  $\mathbf{W}_q$  et d'observation  $\mathbf{V}_q$  sont indépendantes, et compte-tenu de la gaussianité des distributions,

$$p(x_q|y_{q-1}) \propto \exp\left(-\frac{1}{2}(x_q - x_q^b)^T \mathbf{B}_q^{-1}(x_q - x_q^b)\right), \quad (2.8)$$

$$p(y_q|x_q) \propto \exp\left(-\frac{1}{2}(y_q - \mathbf{H}_q x_q)^T \mathbf{R}_q^{-1}(y_q - \mathbf{H}_q x_q)\right). \quad (2.9)$$

On peut écrire l'équation d'analyse (2.3) :

$$p(x_q|y_q) \propto \exp\left(-\frac{1}{2}J^a(x_q)\right), \quad (2.10)$$

avec

$$J^a(x_q) = (x_q - x_q^b)^T \mathbf{B}_q^{-1}(x_q - x_q^b) + (y_q - \mathbf{H}_q x_q)^T \mathbf{R}_q^{-1}(y_q - \mathbf{H}_q x_q). \quad (2.11)$$

Il est clair que la fonction  $J^a(x_q)$  est quadratique en  $x_q$ , ce qui est explicite en écrivant que

$$J^a(x_q) = J^b(\mathbf{x}_q) + J^o(\mathbf{x}_q) = \frac{1}{2} \|x_q - x_q^b\|_{\mathbf{B}_q^{-1}}^2 + \frac{1}{2} \|y_q - \mathbf{H}_q x_q\|_{\mathbf{R}^{-1}}^2, \quad (2.12)$$

avec  $\|\bullet\|_{F^{-1}}$  la norme telle que  $\|x\|_{F^{-1}}^2 = x^T F x$ .  $J^b$  correspond à la partie de la log-vraisemblance reliée à l'ébauche et  $J^o$  la partie reliée à l'observation. Le caractère quadratique de  $J^a$ , avec  $\mathbf{B}_q$  et  $\mathbf{A}_q$  des matrices symétriques définies positives, permet d'en déduire que  $J^a$  est convexe, caractérisée (à une constante près) par un unique minimum  $x_q^a$  et sa hessienne  $\mathbf{A}_q^{-1} = \mathbf{B}_q^{-1} + \mathbf{H}_q^T \mathbf{R}_q^{-1} \mathbf{H}_q$ , tel que

$$J^a(x_q) = (x_q - x_q^a)^T \mathbf{A}_q^{-1} (x_q - x_q^a) + c_1, \quad (2.13)$$

où  $c_1$  désigne une constante.

Le calcul permet d'obtenir pour le minimum l'expression

$$x_q^a = x_q^b + \mathbf{A}_q \mathbf{H}_q^T \mathbf{R}_q^{-1} (y_q - \mathbf{H}_q x_q), \quad (2.14)$$

qui correspond à l'état optimal : celui qui minimise la variance d'erreur. La matrice qui multiplie la différence entre l'observation et la prévision est appelée la matrice de gain, elle est notée  $\mathbf{K}_q$ . D'après l'expression de la matrice  $\mathbf{A}_q$  on peut déduire une autre écriture du gain  $\mathbf{K}_q$  à l'aide de la formule de Sherman-Morrison-Woodbury tel que :

$$\mathbf{K}_q = \mathbf{B}_q \mathbf{H}_q^T (\mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q)^{-1}. \quad (2.15)$$

On calcule à présent l'expression de la matrice de covariance d'erreur d'ébauche  $\mathbf{B}_{q+1}$  à l'instant  $q+1$ . Les propriétés de transformation d'une gaussienne par une dynamique linéaire permettent d'écrire que  $x_{q+1}^b = \mathbf{M}_{q+1} x_q^a$ . En faisant intervenir l'erreur d'analyse  $e_q^a = x_q^a - x_q$ , il vient que la matrice  $\mathbf{A}_q$  s'écrit alors

$$\mathbf{A}_q = \mathbf{E} \left[ \begin{pmatrix} e_q^a \\ e_q^a \end{pmatrix} \begin{pmatrix} e_q^a \\ e_q^a \end{pmatrix}^T \right]. \quad (2.16)$$

On en déduit l'expression de l'erreur d'ébauche  $e_{q+1}^b$

$$e_{q+1}^b = x_{q+1}^b - x_{q+1}, = \mathbf{M}_{q+1} (x_q^a - x_q) - (x_{q+1} - \mathbf{M}_{q+1} x_q), = \mathbf{M}_{q+1} e_q^a - W_q. \quad (2.17)$$

puis celle de la matrice de covariance d'erreur d'ébauche  $\mathbf{B}_{q+1}$  donnée par

$$\begin{aligned} B_{q+1} &= \mathbf{E} \left[ \begin{pmatrix} e_{q+1}^b \\ e_{q+1}^b \end{pmatrix} \begin{pmatrix} e_{q+1}^b \\ e_{q+1}^b \end{pmatrix}^T \right], \\ &= \mathbf{E} \left[ \begin{pmatrix} \mathbf{M}_{q+1} e_q^a - W_q \\ \mathbf{M}_{q+1} e_q^a - W_q \end{pmatrix} \begin{pmatrix} \mathbf{M}_{q+1} e_q^a - W_q \\ \mathbf{M}_{q+1} e_q^a - W_q \end{pmatrix}^T \right], \\ &= \mathbf{M}_{q+1} \mathbf{E} \left[ \begin{pmatrix} e_q^a \\ e_q^a \end{pmatrix} \begin{pmatrix} e_q^a \\ e_q^a \end{pmatrix}^T \right] \mathbf{M}_{q+1}^T + \mathbf{E} \left[ \begin{pmatrix} W_q \\ W_q \end{pmatrix} \begin{pmatrix} W_q \\ W_q \end{pmatrix}^T \right]. \end{aligned}$$

soit

$$\mathbf{B}_{q+1} = \mathbf{M}_{q+1} \mathbf{A}_q \mathbf{M}_{q+1}^T + \mathbf{Q}_q, \quad (2.18)$$

En résumé, il est alors possible d'écrire les équations du filtre de Kalman avec :

- une étape d'analyse comporte trois sous-étapes : le calcul de la matrice de gain  $\mathbf{K}_q$  (celle-ci étant aussi la matrice de gain **optimale**), le calcul de l'état analysé  $\mathbf{x}_q^a$  et le calcul de la matrice de covariance des erreurs d'analyse  $\mathbf{A}_q$ .

$$\mathbf{K}_q = \mathbf{B}_q \mathbf{H}_q^T \left( \mathbf{H}_q \mathbf{B}_q \mathbf{H}_q^T + \mathbf{R}_q \right)^{-1}, \quad (2.19)$$

$$\mathbf{x}_q^a = \mathbf{x}_q^b + \mathbf{K}_q \left( \mathbf{y}_q - \mathbf{H}_q \mathbf{x}_q^b \right), \quad (2.20)$$

$$\mathbf{A}_q = (\mathbf{I} - \mathbf{K}_q \mathbf{H}_q) \mathbf{B}_q, \quad (2.21)$$

où  $\mathbf{I}$  est la matrice identité.

- une étape de prédiction composée de deux équations qui permettent de faire évoluer l'état optimal et la matrice de covariance associée par le modèle :

$$\mathbf{x}_{q+1}^b = \mathbf{M}_{q+1} \mathbf{x}_q^a, \quad (2.22)$$

$$\mathbf{B}_{q+1} = \mathbf{M}_{q+1} \mathbf{A}_q \mathbf{M}_{q+1}^T + \mathbf{Q}_q, \quad (2.23)$$

où  $\mathbf{Q}_q$  représente la matrice de covariance d'erreur modèle.

## 2.4 Cadre général de la dynamique non linéaire

D'un point de vu général, quand les dynamiques introduites par les opérateurs dans le formalisme de l'assimilation de données sont non linéaires, l'hypothèse de gaussianité des statistiques ne peut être maintenue par le système d'Analyse-Prévision}. Dans ce cas les solutions de ce problème sont plus compliquées et nécessitent d'autres implémentations. Les méthodes d'ensemble sont à même de résoudre ce problème et ci-après deux solutions sont présentées : le filtre de Kalman d'ensemble et le filtre particulière.

### 2.4.1 Le filtre de Kalman d'ensemble (EnKF)

Dans le cadre des systèmes faiblement non linéaires une première tentative était d'utiliser le filtre de Kalman étendu EKF (Extended Kalman Filter) qui permet de remplacer le modèle non linéaire par le modèle tangent linéaire tel que  $\mathbf{M} = \frac{\partial \mathbf{M}}{\partial \mathbf{x}}$ . Cependant, dans le cadre des systèmes de grande taille et non linéaires, le filtre EKF rencontre deux principaux problèmes qui sont : le coût très important de calcul pour les matrices de covariances d'erreurs pour le modèle tangent et la mauvaise réponse en présence de fortes non linéarités.

Le filtre de Kalman d'ensemble a été proposé par **Geir Evensen** en 1994 pour plus de détails, il est possible de se référer à [Evensen \(1994b,a\)](#) et [Evensen \(2003\)](#). Cette méthode a d'abord été présentée comme une alternative stochastique au filtre de Kalman étendu déterministe. L'utilisation d'une méthode de Monte Carlo a été imaginée pour résoudre les deux problèmes rencontrés par le filtre EKF. Le filtre de Kalman d'ensemble est conceptuellement très simple et sa mise en œuvre est aisée vu qu'on n'a besoin ni de dériver des opérateurs tangent-linéaires ou des équations adjointes, ni de propager explicitement la matrice de covariance d'erreur d'analyse. En effet, au lieu de propager une matrice de covariance, les erreurs sont représentées statistiquement par un nuage de points propagés par le modèle d'évolution, sans aucune linéarisation. L'étape d'analyse est ensuite celle d'un filtre de Kalman standard.

Depuis son apparition, de nombreuses variantes du filtre EnKF ont été développées. Parmi les variantes du filtre EnKF on retient le filtre de Kalman d'ensemble *stochastique* ou encore à *observations perturbées* (Burgers et al. (1998); Mitchell et al. (1998)). Pour ce type de filtre, les ensembles d'assimilation sont construits par perturbation des observations, qui perturbent en retour les analyses lors de l'étape de correction. Les observations sont perturbées en utilisant une loi normale de moyenne nulle et d'écart-type correspondant à l'écart-type des observations.

L'algorithme du filtre de Kalman d'Ensemble est décrit par la table 2.1. Cependant, il faut toujours garder en mémoire que cet algorithme ne peut être implémenté de cette façon que dans le cas où on est face à des problèmes à faible nombre de degrés de liberté ( $n = \mathcal{O}(10^2)$ ). Pour les systèmes de grande dimension (tels que les systèmes de prévision météorologique, avec  $n = \mathcal{O}(10^7)$ ) d'autres algorithmes sont utilisés pour éviter d'inverser explicitement les matrices de covariance, et limiter la divergence du filtre (Houtekamer and Mitchell (1998)).

TABLE 2.1 – Algorithme du filtre de Kalman d'Ensemble

**Initialisation**

État du système  $x_0^b$  et matrice de covariance d'erreur  $B_0$ .

**Pour tout instant**  $q \geq 1$

**Observation**

+Création d'un jeu d'observations perturbées  $i = 1, \dots, N$ , avec  $V_i \sim \mathcal{N}(0, R_q)$  :

$$\longrightarrow y_{i,q} = y_q + V_{i,q},$$

**Analyse**

+ Calcul de la moyenne de l'ensemble

$$\longrightarrow x_q^b = \frac{1}{N} \sum_{i=1}^N x_{i,q}^b$$

+ Calcul de la matrice de covariance  $B_q$

$$\longrightarrow B_q = \frac{1}{N-1} \sum_{i=1}^N (x_{i,q}^b - x_q^b) (x_{i,q}^b - x_q^b)^T$$

+ Calcul du gain

$$\longrightarrow \mathbf{K}_q^* = B_q H_q^T (H_q B_q H_q^T + R_q)^{-1}$$

+ Calcul des analyses perturbées

$$\longrightarrow \forall i = 1, \dots, N, x_{i,q}^a = x_{i,q}^b + \mathbf{K}_q^* [y_{i,q} - H_q x_{i,q}^b]$$

et de leur moyenne

$$\longrightarrow x_q^a = \frac{1}{N} \sum_{i=1}^N x_{i,q}^a$$

+ calcul de la matrice de covariance  $A$

$$\longrightarrow A = \frac{1}{N-1} \sum_{i=1}^N (x_{i,q}^a - x_q^a) (x_{i,q}^a - x_q^a)^T$$

**Prévision**

+Calcul de l'estimé de prévision pour  $i = 1, \dots, N$  :  $x_{i,q+1}^b = \mathcal{M}_q(x_{i,q}^a)$

### 2.4.2 Le filtre particulaire (PF)

Comme pour le filtre de Kalman d'ensemble, le filtre à particules emprunte son principe de l'approche Monte Carlo. En effet, pour représenter la *pdf* (fonction densité de probabilité) du modèle, le PF utilise un ensemble d'échantillons aléatoires dits "*particules*". Il s'agit alors d'une discrétisation de la (*pdf*) sous forme de somme de distributions de *Dirac* centrées sur les particules. A chaque particule est associé un poids dont l'importance varie selon la ressemblance de la particule aux observations. Les poids sont calculés de façon récursive. Plusieurs algorithmes de filtrage particulaire ont été développés et les travaux de Doucet et al. (2000); Dahia (2005); Sheng and Hu (2005); Douc et al. (2009); Hlinka et al. (2010); Van Leeuwen (2010); Snyder (2011); Hoteit et al. (2012) et Ades and van Leeuwen (2013) présentent des exemples d'application du filtre particulaire. Nous avons choisi d'utiliser l'algorithme du filtre particulaire à ré-échantillonnage multinomial avec sélection génétique connu souvent comme le filtre particulaire génétique (*Genetic Particle Filter, GPF*) et utilisé dans plusieurs travaux dont les travaux de Baehr and Pannekoucke (2010); Rémy et al. (2012); Del Moral (2004) et Bartoli and Del Moral (2011). Dans ce qui suit on présente le principe de filtre particulaire ainsi que le principe de ré-échantillonnage multinomial.

#### 2.4.2.1 Présentation du Filtre

Pour mieux comprendre le fonctionnement du filtre *PF* revenons aux équations 2.2 et 2.1. On supposera que  $x_q$  suit un processus markovien de premier ordre (c'est à dire que à un instant  $q$ ,  $x_q$  ne dépend que de  $x_{q-1}$ . Soit  $(\mathbf{x}_{i,q})$ ,  $i = [1, N]$ , un ensemble formé de  $N$  variables aléatoires *iid* (indépendantes identiquement distribuées) . On suppose que cet ensemble représente  $p(x_q)$  à travers la somme discrète :

$$p(x_q) = \frac{1}{N} \sum_{i=1}^N \delta(x_q = x_{i,q}), \quad (2.24)$$

avec  $\delta$  représente la fonction de Dirac. En effet cette approximation est vraie si la taille de notre ensemble est très grande car d'après le théorème des grands nombres :

$$E_{p_e}(f) = \frac{1}{N} \sum_{i=1}^{i=N} f(x_i) \xrightarrow{N \rightarrow +\infty} E_p(f), \quad (2.25)$$

où  $E_{p_e}$  désigne la moyenne empirique (arithmétique) et  $E_p(f)$  désigne l'espérance mathématique. L'application de la formule de Bayes conduit à écrire que

$$p(x_q|y_q) = \frac{p(y_q|x_q)p(x_q)}{p(y_q)}. \quad (2.26)$$

Posons par  $G_i$  la fonction potentiel tel que :

$$G_{i,q} = p(y_q|x_{i,q}). \quad (2.27)$$

Cette fonction permet d'évaluer la vraisemblance entre les observations et les particules. Ainsi on peut écrire que :

$$p(x_q|y_q) = \sum_{j=1}^N w_{j,q} \delta(x_q = x_{j,q}), \quad (2.28)$$

où  $w_{i,q}$  sont les poids donnés par l'expression suivante :

$$w_{i,q} = \frac{G_{i,q}}{\sum_{i=1}^N G_{i,q}}. \quad (2.29)$$

Dans l'expression 2.29, le dénominateur est un facteur de normalisation et ne dépend pas de l'état  $x_q$ . Essayons alors d'évaluer la fonction  $G_{i,q}$ . On suppose que cette fonction est connue, ainsi cette fonction dépend fortement de la loi de probabilité qui régit les erreurs d'observation. Si de plus, on fait l'hypothèse que les erreurs d'observation sont gaussiennes, les poids  $w_{i,q}$  sont par conséquent des fonctions gaussiennes, tel que le potentiel s'écrit :

$$G_{i,q} = \exp \frac{-1}{2} (y_q - \mathbf{H}x_{i,q})^T \mathbf{R}^{-1} (y_q - \mathbf{H}x_{i,q}) . \quad (2.30)$$

L'estimateur optimal est alors donné par la moyenne  $x_a = \sum_{i=1}^N w_{i,q} x_{i,q}$  (Van Leeuwen and Evensen (1996)) qui n'est autre que la somme pondérée des particules sélectionnées. Une fois les poids calculés, on procède à l'étape de sélection et ré-échantillonnage qui sera détaillée dans ce qui suit.

### 2.4.2.2 Ré-échantillonnage

L'étape de ré-échantillonnage (ou sélection et ré-échantillonnage) repose sur le principe de la sélection de  $N$  particules avec un échantillonnage multinomial. Le principe de sélection multinomiale est de sélectionner avec remplacement de  $N$  particules  $x_j^*$  à partir de l'ensemble des particules initiales  $\{x_i\}$  tel que  $\mathbf{P}(x_j^* = x_i) = w_i$ ,  $i, j = 1 \dots N$ .

En effet, pour générer un évènement  $I_j$  tel que  $P(I_j = i) = w_i$  il suffit de :

- Tirer au sort selon une loi normale un nombre  $u_j \sim \mathbf{U}[0, 1)$  .
- Attribuer à  $I_k$  la valeur de  $i$  selon que  $Q_{i-1} \leq u_k \leq Q_i$ , avec  $Q_i = \sum_{l=1}^i w_l$ .

On remarque que cette méthode de génération d'évènement implique l'utilisation de l'inverse généralisée de la fonction de distribution/densité cumulée (*Cumulative Distribution Function, CDF*) pour créer les évènement  $I_j$  à partir des nombres  $u_j$ .

La figure 2.2 illustre comment est utilisée l'inverse de la (*CDF*) pour générer des évènements.

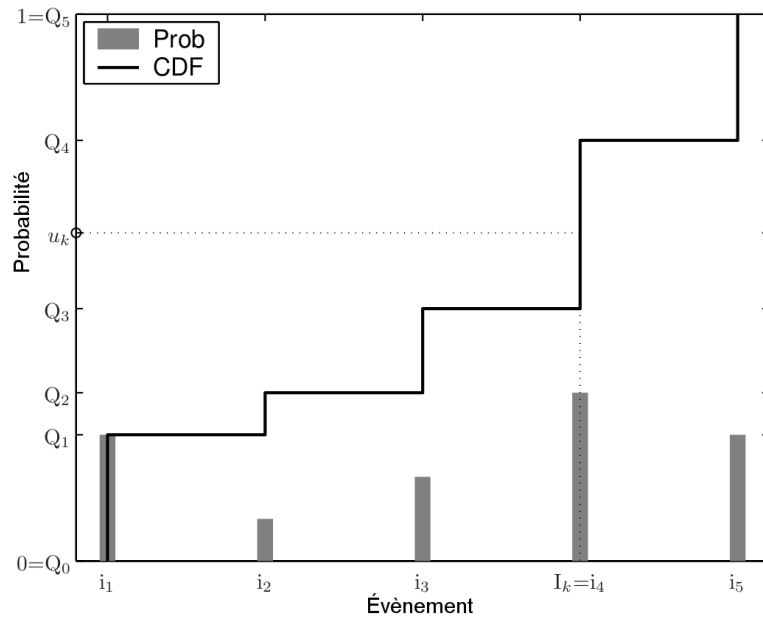


FIGURE 2.2 – Génération d'évènements : les barres en gris présentent les probabilités relatives à chaque évènement  $i_{q_1 \geq q \leq 5}$ , la courbe en noir correspond à la CDF et le trait en pointillé présente la transformation du nombre uniforme  $u_k$  en un évènement (Figure extraite de Hol (2004)).

Les évènements générés par cette méthode peuvent facilement être utilisés pour déterminer les particules ré-échantillonnées en les considérant comme les indices des particules initiales.

Suite à cette étape, la densité estimée est égale à la densité d'origine (densité issue de l'étape d'analyse décrite par les poids de l'ensemble des particules) à une erreur de d'échantillonnage près :

$$\sum_{i=1}^N w_i \delta(x - x_i) \approx \sum_{i=1}^N \frac{n_i}{N} \delta(x - x_i), \quad (2.31)$$

où  $n_i$  désigne le nombre de copies de la particule  $i$  issue de la sélection multinomiale. La mesure  $\mathbf{E}[(n_i - Nw_i)^2]$  désigne la distance entre ces deux densités pour le  $i$ -ème pique. Intuitivement, si ces distances peuvent être réduites pour tous les piques les deux distributions seront d'autant plus proches et la qualité de ré-échantillonnage sera meilleure. On peut très bien remarquer que cette mesure de distance est identique à  $\mathbf{Var}(n_i)$  si tout les  $n_i$  sont non biaisées (c'est à dire  $(\mathbf{E}n_i = Nw_i)$ ). Étant donné que les algorithmes de ré-échantillonnage utilisés dans la littérature sont tous non biaisés, le choix sera basé sur la minimisation de la variance des  $n_i$ .

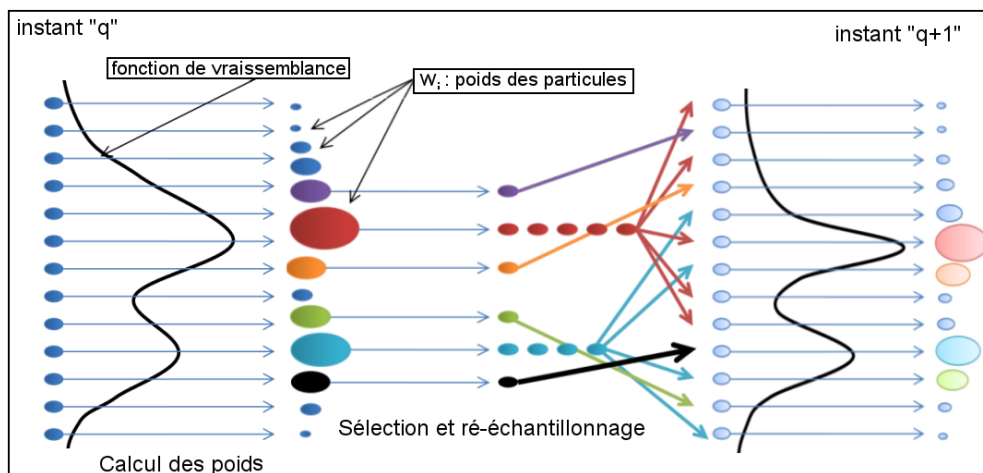


FIGURE 2.3 – Description du principe de filtrage particulaire.

La figure 2.3 présente une description schématique de l'étape de sélection et ré-échantillonnage et l'algorithme général du filtre particulaire est décrit par la table 2.2. Le filtrage particulaire tel qu'on l'a présenté montre que les particules, initialement supposées indépendantes, sont maintenant en interaction. Il n'est donc plus possible de justifier à l'aide de la loi des grands nombres la convergence vers le filtrage non linéaire 2.25. Cependant, en réalité même si les particules interagissent, si le nombre de particule est assez grand alors l'indépendance se propage. Il s'agit de la propagation du chaos (Sznitman (1991)).

On trouve dans la littérature différentes stratégies d'implémentation de l'algorithme de ré-échantillonnage comme par exemple le ré-échantillonnage aléatoire simple (*Simple Random Resampling*), ré-échantillonnage stratifié (*Stratified Resampling*), le ré-échantillonnage systématique (*Systematic Resampling*) et le ré-échantillonnage résiduel (*Residual Resampling*) (voir Arulampalam et al. (2002); Hol (2004) et Hol et al. (2006)).

TABLE 2.2 – Algorithme du filtre particulaire

<p><b>Initialisation</b></p> <p><math>p(x_0) = p_0</math>, (exemple : <math>p_0 \sim \mathcal{U}</math>).</p> <p><b>Pour tout instant</b> <math>q \geq 1</math></p> <p><b>Analyse</b></p> <p>Pour tout <math>i = 1, \dots, N</math></p> <p>+ Calcul des fonctions potentiel <math>G_{i,q}</math> relatives aux différentes particules</p> <p>+ Calcul des différents poids</p> <p>→ <math>w_{i,q} = \frac{G_{i,q}}{\sum_{i=1}^N G_{i,q}}</math></p> <p>avec, <math>p(x_q y_q) = \sum_{i=1}^N w_{i,q} \delta(x_q = x_{i,q}^*)</math></p> <p>+ Sélection et ré-échantillonnage des particules</p> <p><b>Prévision</b></p> <p>+ Calcul de <math>p(x_{i,q+1} y_{i,q}) = \frac{1}{N} \sum_{i=1}^N \delta_M^{q+1}(x_{i,q}^*)</math></p> <p>avec, <math>\delta_M^{q+1}(x_{i,q}^*) = \delta(x_{q+1} = \mathbf{M}_{q+1}(x_{i,q}^*))</math></p>
---

### 2.4.2.3 Le lisseur particulaire : Particle Smoother (PS)

Pour tirer parti de la dynamique on souhaite étendre le formalisme du filtrage à celui d'un filtrage des trajectoires ce qui conduit au formalisme du lisseur. Ainsi, on se met dans le cadre où on a une séquence d'observations entre deux analyses. Ainsi les observations vont être assimilées sur une fenêtre temporelle et non pas à un instant donné. D'après Rémy et al. (2012), si une famille d'observations  $(y_q^m)$ ,  $m = 1, \dots, M$  est disponible aux instants  $m$  entre les instants d'analyse  $q-1$  et  $q+1$  alors pour chaque observation  $y_q^m$  et chaque particule  $x_{i,m}$  la fonction "potentiel" est calculée comme dans l'équation 2.27.

Ainsi :

$$G_q(x_{i,m}) = p(y_q^m | x_{i,m}). \quad (2.32)$$

En effet, la fonction "potentiel" de la  $i$ -ème particule sur la fenêtre temporelle associée à la  $q$ -ième analyse est le produit de toutes les fonctions "potentiel" calculées aux instants  $m$  (étant donné que l'on suppose les différentes observations indépendantes) :

$$\mathcal{G}_q(x_i) = \prod_{m=1}^M G_q(x_{i,m}), \quad (2.33)$$

$$\mathcal{W}_{i,q} = \frac{\mathcal{G}_q(x_i)}{\sum_{j=1}^N \mathcal{G}_q(x_j)}. \quad (2.34)$$

De cette manière on obtient ce qu'on appelle un lisseur particulaire vu que la correction prend en compte toutes les observations disponibles sur la fenêtre d'assimilation. Le choix des trajectoires/particules "optimales" est par conséquent contraint par toutes les observations.



### **Conclusion**

Ce chapitre a fait l'objet d'une introduction assez concise à l'assimilation de données en particulier l'assimilation de données d'ensemble dans le cadre d'une assimilation de données séquentielle. Deux approches ont été présentées qui sont le EnKF et le PF. La dernière approche a été adaptée pour réaliser une assimilation de données sur une fenêtre temporelle aboutissant à ce qu'on appelle un lisseur particulaire PS. C'est ce lisseur qui sera utilisé par la suite dans notre approche de désagrégation pour estimer les températures sous-pixelliques à travers l'étalonnage/l'optimisation d'un jeu de paramètres d'un modèle de Transfert-Sol-Végétation-Atmosphère. Le chapitre suivant présentera donc le modèle utilisé au cours de cette thèse.

---

## Application au modèle SETHYS

CE CHAPITRE présente le modèle de surface que nous avons utilisé pour simuler la température d'un pixel satellite et son implémentation sur notre région d'étude. La description de ce modèle (SETHYS) a été présentée dans un certain nombre de travaux auxquels on peut se référer pour plus d'informations ([Bernard et al. \(1986\)](#); [Taconet et al. \(1986\)](#); [Ben Mehrez et al. \(1992\)](#); [Ottlé and Vidal-Madjar \(1994\)](#); [Coudert \(2006\)](#); [Coudert and Ottlé \(2007\)](#); [Coudert et al. \(2008\)](#); [Saux-Picart et al. \(2009a,b\)](#)). Dans ce qui suit on reprend l'essentiel des descriptions qui figurent dans la thèse de [Saux-Picart et al. \(2009b\)](#). Le modèle SETHYS présente un certain nombre de paramètres qu'il faut étalonner avant de pouvoir lancer une simulation. Une analyse de sensibilité sur les paramètres a permis d'identifier ceux qui devront être étalonnés par le biais de l'assimilation. Les résultats de cette analyse menée sur les différents types d'occupation du sol présents dans notre région sont présentés à la suite de la présentation du modèle SETHYS.

### 3.1 Le modèle SETHYS

Le modèle SETHYS, pour « Suivi de l'Etat Hydrique du Sol », ([Bernard et al. \(1986\)](#); [Taconet et al. \(1986\)](#); [Ben Mehrez et al. \(1992\)](#); [Ottlé and Vidal-Madjar \(1994\)](#); [Coudert \(2006\)](#); [Coudert and Ottlé \(2007\)](#); [Coudert et al. \(2008\)](#); [Saux-Picart et al. \(2009a,b\)](#)) est un modèle de Transfert Sol-Végétation-Atmosphère (TSVA). En fait un modèle TSVA est un modèle qui décrit les différents échanges de flux de chaleur, d'énergie et d'eau qui régissent le continuum {Biosphère-Atmosphère}. SETHYS a été développé pour suivre le bilan hydrique et les états des couverts agricoles homogènes. Ce modèle est couplé à un modèle de transfert radiatif permettant la distribution de l'énergie incidente (solaire et atmosphérique) entre la végétation et le sol. Le modèle de transfert radiatif prend en compte les interactions multiples entre le sol et la végétation. Le modèle SETHYS est un modèle multi-sources, il considère que la plupart des couverts végétaux peut être représentée par deux couches : une première pour les végétations basses (prairie, blé, herbes, etc.) et une deuxième pour les végétations hautes (arbres). Le calcul des bilans d'énergie et d'eau est donc assuré pour chacune de ces couches. Dans ce qui suit on présente les détails des calculs des bilans d'énergie et d'eau dans SETHYS.

### 3.1.1 Bilan d'énergie au niveau de l'interface

Le rayonnement solaire incident au sommet de l'atmosphère, est la variable principale qui gère les différents processus d'échange d'énergie et d'eau à la surface terrestre. La quantité d'énergie disponible au sommet de l'atmosphère est appelée la constante solaire  $I_0$  et vaut  $1370 \text{ W.m}^2$ . Généralement on ramène cette quantité d'énergie, au sommet de l'atmosphère, à une moyenne annuelle sur la surface du globe terrestre d'environ  $350 \text{ W.m}^2$  (Figure 3.1).

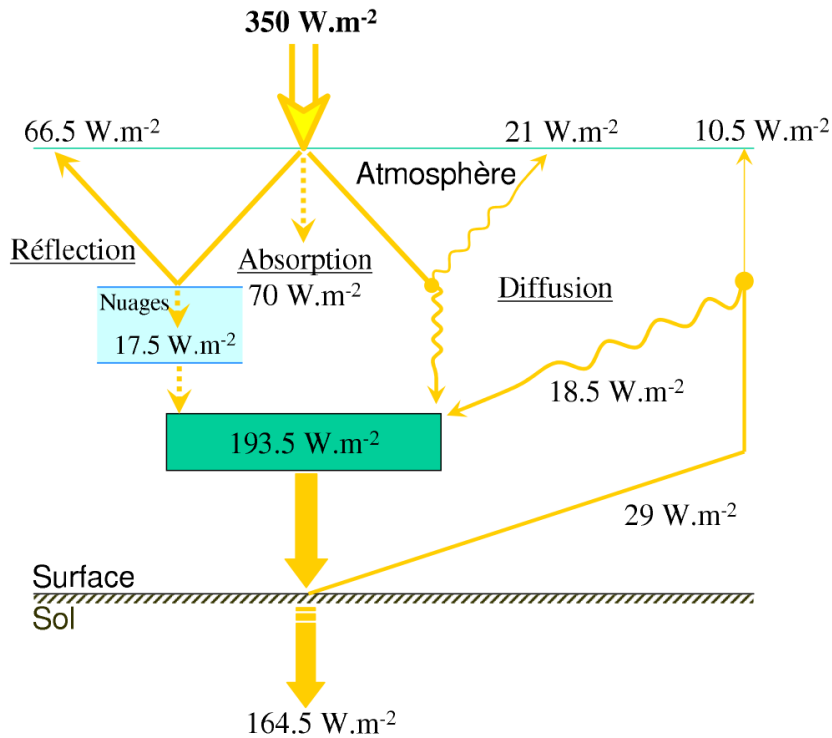


FIGURE 3.1 – Schéma du bilan radiatif global (figure extraite de Saux Picart (2007))

Ce flux incident subit différents phénomènes d'absorption, diffusion et réflexion lors de sa traversée de l'atmosphère, en particulier en interagissant avec les différents composants de l'atmosphère : gaz carbonique, hydrogène, oxygène, ozone, vapeur d'eau, etc. Au niveau de la surface, ce flux se divise en deux parties : rayonnement solaire direct et rayonnement solaire diffus. Ces deux termes forment le forçage radiatif de courtes longueurs d'ondes (de  $0.3$  à  $3\mu\text{m}$ ), il est égal à  $193.5 \text{ W.m}^2$  en moyenne annuelle intégrée spectralement sur la surface totale du globe. Un deuxième terme s'ajoute au forçage radiatif solaire, le rayonnement thermique de grandes longueurs d'onde émis par l'atmosphère (entre  $3$  à  $300\mu\text{m}$ ). La somme de ces deux quantités est à l'origine des échanges d'énergie et d'eau à l'interface SVA. Le rayonnement net est la résultante du bilan radiatif de surface, il est égal à la somme des rayonnement émis et réfléchis au niveau de la surface et s'écrit comme :

$$R_n = R_{CL\downarrow} - R_{CL\uparrow} + R_{GL\downarrow} - R_{GL\uparrow} \quad (3.1)$$

où les indices  $CL$  et  $GL$  se réfèrent respectivement aux contributions courtes longueurs et grandes longueurs d'onde, et les flèches  $\uparrow$  et  $\downarrow$  à la direction de propagation ascendante et descendante. L'unité de chaque terme est le  $\text{W.m}^{-2}$ . Les parties suivantes présentent plus de détails sur le calcul du bilan radiatif et des différents termes de courtes et grandes longueurs d'onde.

### 3.1.1.1 Le bilan radiatif

Le formalisme qui décrit le bilan radiatif s'appuie sur les lois de l'électromagnétisme, et son expression dépend des caractéristiques des éléments considérés et de l'atmosphère. Pour décrire l'interaction onde-matière, trois phénomènes physiques interviennent : la réflexion, l'absorption et la transmission, auxquels sont associés respectivement les coefficients de réflectivité, d'absorptivité et de transmissivité. Suivant le principe de conservation de l'énergie, la somme des trois coefficients vaut 1.

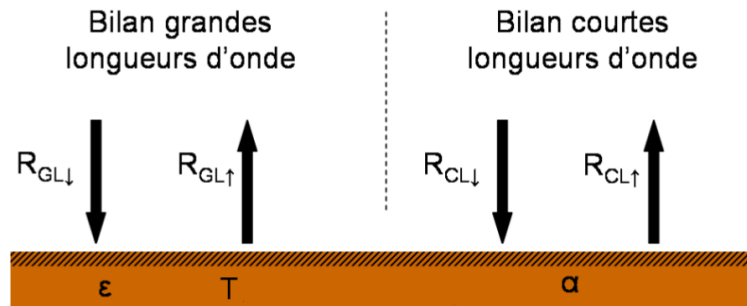


FIGURE 3.2 – Schéma du bilan d'énergie pour une surface infinie :  $\alpha$ ,  $\epsilon$  et  $T$  désignent respectivement l'albédo, l'émissivité et la température du sol.

Dans le cas d'une surface de sol nu homogène, le bilan d'énergie peut être représenté de la façon suivante (voir Figure 3.2) : L'équation 3.1 décrit le calcul du rayonnement.

- $\mathbf{R}_{CL\downarrow}$  est le rayonnement solaire incident correspondant à  $R_g$  dit rayonnement global.
- $\mathbf{R}_{CL\uparrow}$  est le rayonnement solaire réfléchi par la surface :  $\mathbf{R}_{CL\uparrow} = \alpha_s \mathbf{R}_g$  où  $\alpha_s$  désigne l'albédo<sup>1</sup> de la surface.
- $\mathbf{R}_{GL\downarrow}$  est le rayonnement atmosphérique descendant. C'est la somme des émissions des différentes couches de l'atmosphère. On considère généralement que ce terme est l'émission d'une atmosphère à la température équivalente  $T_a$  et d'émissivité  $\epsilon_a$  sous l'hypothèse suivante : les couches les plus basses de l'atmosphère sont celles qui contribuent le plus car elles sont les plus chaudes. D'après la loi de Stefan-Boltzmann<sup>2</sup> on déduit que  $\mathbf{R}_{GL\downarrow} = \epsilon_a \sigma \mathbf{T}_a^4$  avec  $\sigma$ <sup>3</sup> la constante de Stefan-Boltzmann.
- $\mathbf{R}_{GL\uparrow}$  est la somme de la réflexion du rayonnement atmosphérique et de l'émissivité propre de la surface.  $\mathbf{R}_{GL\uparrow} = (1 - \epsilon_s) \epsilon_a \sigma \mathbf{T}_a^4 + \epsilon_s \sigma \mathbf{T}_s^4$ .

1. L'albédo, grandeur sans dimension, est le rapport de l'énergie solaire réfléchie par une surface à l'énergie solaire incidente. C'est une notion comparable à la réflectivité. Par définition, l'albédo est une valeur comprise entre 0 et 1 : un corps noir parfait, qui absorberait toutes les ondes électromagnétiques sans en réfléchir aucune, aurait un albédo nul, tandis qu'un miroir parfait, qui réfléchirait toutes les ondes électromagnétiques sans en absorber une seule, aurait un albédo égal à 1.

2. La loi de Stefan-Boltzmann (du nom des physiciens Jožef Stefan et Ludwig Boltzmann) établit que la puissance totale rayonnée par unité de surface dans le demi-espace libre du corps noir (exitance énergétique du corps noir) s'exprime par la formule :  $\mathbf{M} = \sigma \mathbf{T}^4$  avec  $\sigma$  la constante de Stefan-Boltzmann.

3. La constante de Stefan-Boltzmann  $\sigma$  est une constante (non fondamentale) de la physique. Elle intervient dans la loi de Stefan-Boltzmann pour exprimer la relation entre la température et l'énergie rayonnée par un corps noir et vaut :  $\sigma = 5.670400 \times 10^{-8} \text{Wm}^{-2} \text{K}^{-4}$

Ainsi, l'expression du bilan d'énergie pour une surface peut être écrit de la manière suivante :

$$\mathbf{R}_n = (1 - \epsilon_s)\mathbf{R}_g + \sigma\epsilon_s(\epsilon_a\mathbf{T}_a^4 - \mathbf{T}_s^4) \quad (3.2)$$

Dans le modèle SETHYS, on représente le système SVA par 2 ou 3 sources : le sol et deux couches superposées de couvert végétal de hauteurs différentes permettant de simuler un couvert arboré associé à une strate herbacée. Le bilan radiatif peut être calculé en considérant la végétation comme un milieu semi-transparent et le sol comme un milieu opaque. Des interactions entre les différentes strates vont donc intervenir. La partition de l'énergie radiative est définie par le facteur d'écran  $\sigma_f$  défini pour chacune des courtes et grandes longueurs d'onde. Ce facteur d'écran peut être estimé à partir du LAI (pour "Leaf Area Index") qui désigne la surface totale des feuilles des végétaux d'une parcelle rapportée à sa surface. Une paramétrisation de ce facteur d'écran a été proposée par Rosenthal et al. (1977) et Deardorff (1978) dans le domaine solaire. François (1995) a estimé ces facteurs dans le domaine thermique. Ainsi :

$$\begin{cases} \sigma_f = 1 - \exp(-0.825LAI), & \text{pour les grandes longueurs d'onde} \\ \sigma_f = 1 - \exp(-0.5LAI), & \text{pour les courtes longueurs d'onde} \end{cases} \quad (3.3)$$

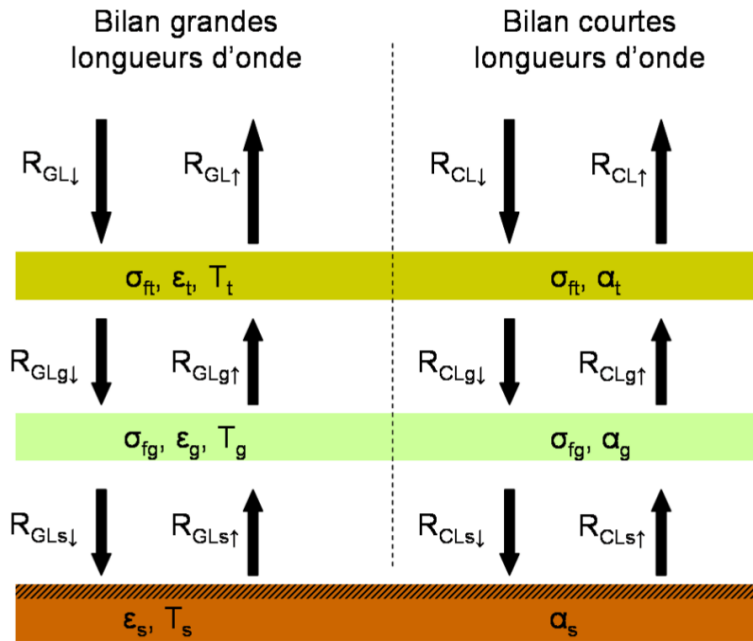


FIGURE 3.3 – Schéma du bilan d'énergie pour le modèle SETHYS (figure extraite de Saux Picart (2007)).

La figure 3.3 montre le schéma des bilans radiatifs du système à 3 sources modélisé dans SETHYS. Pour chaque source on calcule le bilan des flux entrants et sortants.

En grandes longueurs d'onde :

$$\begin{cases} R_{GLg\downarrow} = (1 - \sigma_{ft})R_{GL\downarrow} + \epsilon_t\sigma_{ft}\sigma T_t^4 + (1 - \epsilon_t)\sigma_{ft}R_{GLg\uparrow} \\ R_{GLg\uparrow} = (1 - \sigma_{fg})R_{GLs\uparrow} + \epsilon_g\sigma_{fg}\sigma T_g^4 + (1 - \epsilon_g)\sigma_{fg}R_{GLg\downarrow} \\ R_{GLs\downarrow} = (1 - \sigma_{fg})R_{GLg\downarrow} + \epsilon_g\sigma_{fg}\sigma T_g^4 + (1 - \epsilon_g)\sigma_{fg}R_{GLs\uparrow} \\ R_{GLs\uparrow} = \epsilon_s\sigma T_s^4 + (1 - \epsilon_s)R_{GLs\downarrow} \\ R_{GL\uparrow} = (1 - \epsilon_s)\sigma_{ft}R_{GL\downarrow} \end{cases} \quad (3.4)$$

En courtes longueurs d'onde :

$$\begin{cases} R_{CLg\downarrow} = (1 - \sigma_{f_t})R_{CL\downarrow} + \alpha_t\sigma_{f_t}R_{CLg\uparrow} \\ R_{CLg\uparrow} = \alpha_g\sigma_{f_g}R_{CLg\downarrow} + \alpha_s(1 - \sigma_{f_g})R_{GLs\downarrow} \\ R_{CLs\downarrow} = (1 - \sigma_{f_g})R_{CLg\downarrow} + \alpha_g\sigma_{f_g}R_{CLs\uparrow} \\ R_{GLs\uparrow} = \alpha_sR_{CLs\downarrow} \\ R_{GL\uparrow} = \alpha_t\sigma_{f_t}R_{CL\downarrow} + (1 - \sigma_{f_t})R_{CLg\uparrow} \end{cases} \quad (3.5)$$

Avec,  $\sigma$  la constante de Stefan-Boltzmann,  $T_t$ ,  $T_g$  et  $T_s$  sont respectivement les températures thermodynamiques des arbres, de l'herbe (ou plus généralement de la végétation basse) et du sol. On peut ainsi calculer le rayonnement net à chacun des niveaux dans le système sol-végétation.

### 3.1.1.2 Les échanges conductifs

La conduction constitue le mode de transfert de chaleur principal dans le sol (milieu solide). Ce mode de propagation de l'énergie correspond à la redistribution/transfert d'énergie cinétique entre les différentes molécules formant le substrat. Pour les milieux fluides comme l'air les échanges par conduction sont souvent négligés par rapport aux échanges par convection (principal mode de transfert de chaleur pour les milieux fluides). On considère uniquement le flux vertical dans le sol, son expression est donnée par la loi de Fourier sous forme de relation flux gradient comme suit :

$$\mathbf{G}(\mathbf{z}, \mathbf{t}) = -\lambda_s(\mathbf{z}) \frac{\partial \mathbf{T}_{\text{sol}}(\mathbf{z}, \mathbf{t})}{\partial \mathbf{z}} \quad (3.6)$$

$$\rho_s \mathbf{C}_s(\mathbf{z}) = \frac{\partial \mathbf{T}_{\text{sol}}(\mathbf{z}, \mathbf{t})}{\partial \mathbf{t}} = -\frac{\partial \mathbf{G}(\mathbf{z}, \mathbf{t})}{\partial \mathbf{z}} \quad (3.7)$$

$T_{\text{sol}}(z, t)$  représente la température du sol à la profondeur  $z$  ( $T$  est exprimée en  $K$ ),  $\lambda_s(z)$  est la conductivité thermique du sol (en  $W.m^{-1}.K^{-1}$ ) et  $\rho_s C_s(z)$  est la capacité calorifique du sol à la profondeur  $z$ .

### 3.1.1.3 Les échanges convectifs

Les échanges convectifs diffèrent des échanges de chaleur décrits précédemment par le fait qu'il caractérisent les transferts de chaleur et de masse par déplacement de matière. On trouve différentes formes de convection :

- *La convection libre* qui résulte des gradients de densité issus des gradients de température dans l'air.
- *La convection forcée* qui est provoquée par la turbulence induite par le champ de vent.
- *La convection mixte* qui est rencontrée dans des conditions de vent faible. Les mouvements de l'air sont alors produits par l'action combinée de la convection libre et de la convection forcée.

L'expression des flux de quantité de mouvement, de chaleur sensible et de chaleur latente est liée aux gradients verticaux à travers les coefficients de diffusivité turbulente<sup>4</sup> (K-Theory) par analogie avec la diffusion moléculaire. C'est le principe de fermeture au premier ordre du système d'équations qui permet d'écrire :

- le flux de quantité de mouvement :

$$\tau = -\rho_a K_m(z) \frac{\partial u(z)}{\partial z} \quad (3.8)$$

- le flux de chaleur sensible :

$$H = -\rho_a C_p K_h(z) \frac{\partial T(z)}{\partial z} \quad (3.9)$$

- le flux de chaleur latente :

$$LE = -L \rho_a K_\nu(z) \frac{\partial q(z)}{\partial z} = -\frac{\rho_a C_p}{\gamma} K_\nu(z) \frac{\partial e(z)}{\partial z} \quad (3.10)$$

avec ;  $\rho_a$  est la masse volumique de l'air,  $C_p$  la chaleur massique isobare de l'air ( $J.kg^{-1}.K^{-1}$ ),  $L$  la chaleur latente de vaporisation de l'eau ( $J.kg^{-1}$ ),  $\gamma$  la constante psychométrique ( $0.66hPa.K^{-1}$ ),  $u$  le module du vent ( $m.s^{-1}$ ),  $T$  la température de l'air (K),  $q$  l'humidité spécifique de l'air ( $kg.kg^{-1}$ ),  $e$  la pression partielle de vapeur d'eau dans l'air au niveau  $z$  et  $K_i$  ( $i = \{m, h, \nu\}$ ) sont les coefficients de diffusivité turbulente. Seuls les flux  $H$  et  $LE$  interviennent explicitement dans le bilan d'énergie. Le calcul de ces derniers peut être réalisé en intégrant leurs équations respectives entre deux niveaux atmosphériques  $z_1$  et  $z_2$ . Pour cela il faut d'abord formuler l'hypothèse que les diffusivités turbulentes, selon la théorie de similarité (Monin and Obukhov (1954); Thom (1972)), pour les flux  $H$  et  $LE$  sont identiques à celles de la quantité de mouvement ( $K_m = K_h = K_\nu$ ). On obtient ainsi :

$$H = -\rho_a C_p \frac{T_2 - T_1}{\int_{z_1}^{z_2} \frac{dz}{K_m(z)}} \quad (3.11)$$

$$LE = -\frac{\rho_a C_p}{\gamma} \frac{e_2 - e_1}{\int_{z_1}^{z_2} \frac{dz}{K_m(z)}} \quad (3.12)$$

L'expression des flux, par analogie à la loi d'Ohm, est décrite par un gradient de température pondéré par une résistance aérodynamique aux transferts turbulent de chaleur  $r_{a(z_1 \rightarrow z_2)}$  (voir Figure 3.4).

$$r_{a(z_1 \rightarrow z_2)} = \int_{z_1}^{z_2} \frac{dz}{K_m(z)} \quad (3.13)$$

4. Hypothèse de modélisation permettant d'écrire de façon approchée le flux turbulent de quantité de chaleur. Ce dernier est supposé aligné avec le gradient de température moyenne. L'opposé de la constante de proportionnalité est appelée diffusivité turbulente et on la note  $K_T [m^2.s^{-1}]$

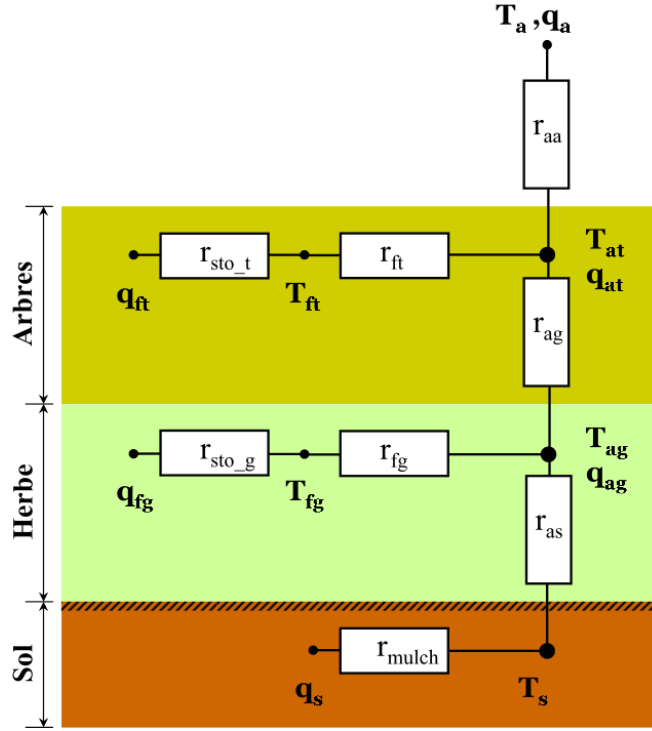


FIGURE 3.4 – Schéma résistif représentant les échanges de surface (figure extraite de la thèse de [Saux Picart \(2007\)](#))

D'autre part, les flux de chaleur rencontrent, en plus de la résistance aérodynamique, une résistance de couche limite foliaire, qui caractérise l'écoulement laminaire autour des feuilles (cas de diffusion moléculaire). Son expression est donnée (par unité de surface foliaire) par [Jones \(1983\)](#) :

$$r_{clt}(z) = A' \left( \frac{l_{ft}}{u(z)} \right)^{\frac{1}{2}} \quad (3.14)$$

$$r_{clg}(z) = A' \left( \frac{l_{fg}}{u(z)} \right)^{\frac{1}{2}} \quad (3.15)$$

où  $A'$  est une constante intrinsèque aux propriétés moléculaires de l'air (prise égale à  $100\text{m}^{-1}.\text{s}^{\frac{1}{2}}$ ),  $l_{ft}$  et  $l_{fg}$  sont respectivement les dimensions de la feuille d'arbre et d'herbe dans la direction du vent (approchée dans la modélisation par la largeur de la feuille). Les résistances  $r_{clt}$  et  $r_{clg}$  sont données par unité de surface foliaire, il est donc nécessaire de les calculer pour l'ensemble du couvert en les intégrant comme une multitude de résistance en parallèle sur la totalité de l'indice de surface foliaire ( $LAI$ ). En supposant une répartition uniforme du  $LAI$  sur la hauteur du couvert végétal on obtient :

$$r_{ft} = \frac{A'n(LAI_t)}{2LAI_t \left( 1 - \exp\left(\frac{-n(LAI_t)}{2}\right) \right)} \left( \frac{l_{ft}}{u(z_t)} \right) \quad (3.16)$$

$$r_{fg} = \frac{A'n(LAI_g)}{2LAI_g \left( 1 - \exp\left(\frac{-n(LAI_g)}{2}\right) \right)} \left( \frac{l_{fg}}{u(z_g)} \right) \quad (3.17)$$

Les flux de chaleur latente sont affectés par deux autres résistances : la résistance du couvert au niveau de la feuille et la résistance de sol au niveau de la surface du sol. La *résistance du couvert* intègre la résistance stomatique à la transpiration et la résistance à l'évaporation de l'eau libre sur les feuilles.



- la résistance stomatique permet de donner un rôle direct à la végétation. La conductance stomatique est sensible à plusieurs conditions environnementales tels que la lumière, l’humidité de l’air, la concentration en dioxyde de carbone, etc. Le modèle stomatique utilisé dans *SETHYS* est basé sur le modèle de [Ball \(1988\)](#) :

$$r_{sto} = g_s^{-1} = \left( m \frac{A_n}{c_s} h h_s p + b \right)^{-1} \quad (3.18)$$

où  $A_n$  est le taux d’assimilation nette (modèle de [Farquhar et al. \(1980\)](#)),  $c_s$  et  $h_s$  sont respectivement la pression partielle de  $CO_2$  et l’humidité relative à la surface de la feuille,  $p$  est la pression atmosphérique et  $m$  et  $b$  des facteurs empiriques dépendants du type de végétation, du *LAI* et de l’humidité de la zone racinaire.

- la résistance à l’évaporation de l’eau libre est définie en fonction de la proportion d’eau libre sur les feuilles selon [Monteith and Szicz \(1961\)](#).

La *résistance du sol* limite l’évaporation du sol avec la diminution du contenu en eau (écart à la saturation). Cette résistance dépend du type de sol et de l’épaisseur de la couche totalement sèche, c’est à dire à l’humidité résiduelle  $w_{resid}$ , qui se forme en surface du sol à laquelle on donne le nom de “mulch” (voir partie [3.1.2](#)). La résistance du sol est directement proportionnelle à l’épaisseur du mulch ([Perrier \(1975\)](#)) :

$$r_{mulch} = \frac{\tau}{D_v} z_{mulch} \quad (3.19)$$

où  $z_{mulch}$  représente l’épaisseur du mulch,  $D_v$  est la diffusivité de la vapeur d’eau dans le sol et  $\tau$  est la tortuosité du sol (qui représente le rapport entre une épaisseur de sol et la longueur du chemin effectivement parcouru par une particule de vapeur d’eau pour traverser cette épaisseur). Ces deux derniers paramètres sont à ajuster en fonction de la texture du sol.

### 3.1.2 Bilan hydrique et contenu en eau du sol

Dans *SETHYS* le sol est représenté par deux couches superposées correspondant aux zones racinaires explorées par les racines des deux types de végétation (haute et basse). Pour mieux représenter l’évaporation de l’eau du sol, le concept de mulch a été adopté. Il consiste à faire évoluer dans le temps, une couche sèche en surface dont la hauteur évolue en fonction de l’évaporation et des pluies. L’eau du sol contenue dans les pores est exprimée par le contenu en eau volumique  $w(m^3.m^{-3})$  du sol. La porosité est souvent exprimée par le contenu en eau volumique à saturation ( $w_{sat}$ ), définie comme la quantité maximale d’eau qu’un sol peut contenir. D’autre part, on définit ( $w_{resid}$ ) qui représente par contre la quantité minimale d’eau non évaporable qu’un sol peut contenir (eau résiduelle). Dans *SETHYS* l’humidité de la couche mulch est toujours égale à  $w_{resid}$ .

*SETHYS*, ne résout pas les équations de diffusion de Richards pour calculer et modéliser les transferts verticaux de l’eau dans le sol mais adopte plutôt une approche conceptuelle. La représentation du sol est volontairement simple pour limiter le nombre de paramètres. Seuls les mouvements verticaux à travers des réservoirs sont considérés (l’eau perdue par ruissellement latéral est calculée mais ne peut pas se ré-infiltrer).

### 3.1.2.1 Courbe de rétention et conductivité hydraulique

Les transferts d'eau dans le sol nécessitent la définition de deux fonctions : la courbe de rétention qui relie le potentiel hydrique au contenu en eau du sol et la relation liant la conductivité hydraulique au contenu en eau du sol. Pour une couche de sol  $i$ , les expressions du potentiel hydrique<sup>5</sup> et de la conductivité hydraulique sont les suivantes :

$$\Psi(w_i) = \Psi_{sat} \left( \frac{w_i}{w_{sat}} \right)^{-b} \quad (3.20)$$

$$K(w_i) = K_{sat} \left( \frac{w_i}{w_{sat}} \right)^{2b+3} \quad (3.21)$$

Avec  $\{b, w_i\}$  désignent respectivement la pente de la courbe de rétention en eau et l'humidité volumique de la  $i$ -ème couche du sol,  $K_{sat}[m.s^{-1}]$  est la conductivité hydraulique et  $\Psi_{sat}$  est le potentiel hydrique.

### 3.1.2.2 Les termes de diffusion, drainage et ruissellement

- La diffusion et le drainage sont les processus majeurs des mouvements verticaux d'eau dans le sol. La diffusion de l'eau dépend du gradient de potentiel hydrique entre les deux couches de sol représentées dans SETHYS. Son signe varie en fonction du gradient d'humidité entre ces deux couches (l'équation 3.20 montre bien que le potentiel hydrique est directement lié à l'humidité volumique). Ainsi on peut écrire la diffusivité comme suit :

$$D(w) = K(w) \frac{\partial \Psi(w)}{\partial w} = \frac{-b \Psi_{sat} K_{sat}}{w_{sat}} \left( \frac{w}{w_{sat}} \right)^{b+2} \quad (3.22)$$

- Le drainage correspond à l'écoulement gravitationnel de l'eau dans le sol. On le considère directement proportionnel à la conductivité hydraulique (d'après l'équation 3.21). Son expression s'écrit alors :

$$D_r(w) = K_{sat} \left( \frac{w - w_{resid}}{w_{sat} - w_{resid}} \right)^{2b+3} \quad (3.23)$$

- Le ruissellement de surface est un phénomène important à prendre en compte dans le bilan hydrique d'un sol. Dans SETHYS deux types de ruissellement sont pris en compte : le ruissellement dit de Horton qui caractérise les zones arides (le ruissellement par dépassement de la capacité d'infiltration d'un sol dépend de l'intensité des pluies) et le ruissellement de Dunne (ruissellement par débordement du réservoir indépendant de l'intensité de la pluie). Le ruissellement est strictement lié à l'infiltrabilité  $I(w)$  qui représente la quantité maximale d'eau qu'un sol peut absorber, elle est exprimée en  $mm.s^{-1}$ . Généralement  $I(w)$  est calculée pour une épaisseur du sol de 1 à 10cm.

Decharme and Douville (2006) ont proposé une expression de l'infiltrabilité déduite de l'intégration des équations 3.20 et 3.21 et qui s'écrit comme suit :

$$I(w) = k_{sat} \left[ \frac{b \Psi_{sat}}{\Delta z} \left( \frac{w}{w_{sat}} - 1 \right) + 1 \right] \quad (3.24)$$

5. Le potentiel hydrique est défini comme étant la somme de deux quantités facilement mesurables : la pression hydrostatique et la pression osmotique. L'eau se déplace toujours d'une région de potentiel hydrique élevé vers une région de potentiel hydrique faible.

où  $\Delta z$  représente la profondeur que l'on choisit de prendre en compte pour calculer l'infiltrabilité, et  $w$  l'humidité moyenne de la couche d'épaisseur  $\Delta z$ .

### 3.1.2.3 Le bilan hydrique

La description du bilan d'eau dans le modèle SETHYS est représentée par la figure 3.5. Le mulch agit sur l'évaporation, et le ruissellement est très sensible au choix de la paramétrisation de cette croûte. Le mulch ne se développe que dans la couche superficielle du sol. La croissance du mulch est dirigée par les pertes d'eau par évaporation. Lorsqu'une pluie survient, le mulch est humidifié à l'humidité volumique correspondante à la capacité du champ, et une certaine épaisseur de mulch peut être "emprisonnée" entre deux couches et dans ce cas un deuxième mulch peut se développer en surface.

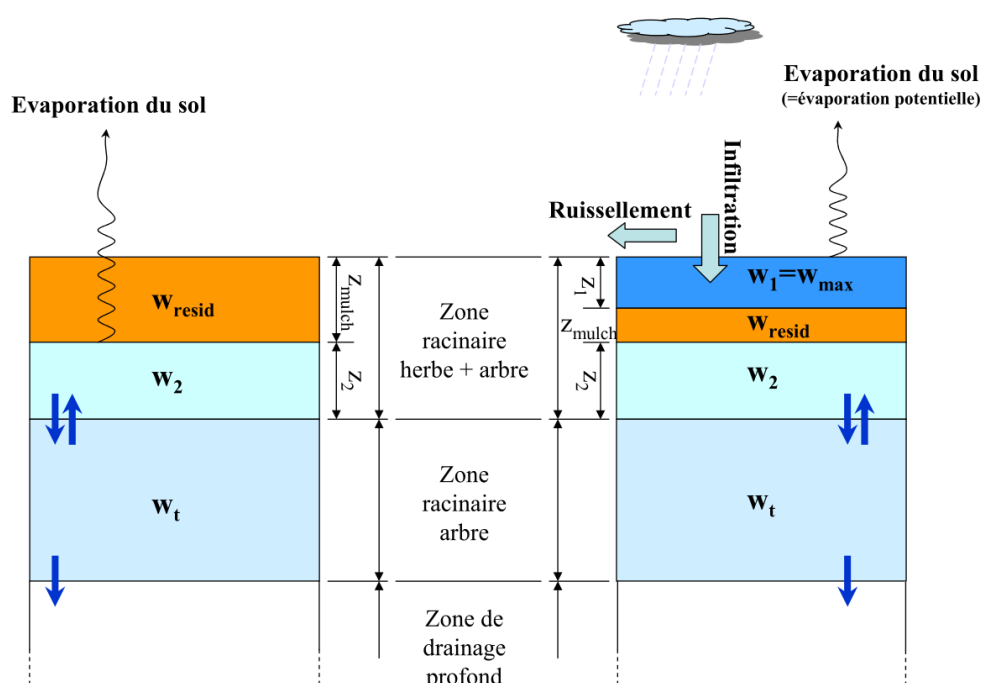


FIGURE 3.5 – Schéma du bilan hydrique dans le modèle SETHYS (figure extraite de Saux Picart (2007)).

### 3.1.3 La température de brillance de la surface

Le calcul de la température de surface dans le modèle SETHYS est assuré par un modèle de transfert radiatif. Ce modèle permet de simuler la température radiative ainsi que l'émissivité totale de surface, respectivement  $T_{rad,i}$  et  $\epsilon_i$ , dans une direction donnée pour une classe de végétation (entité)  $i$  et dans ce cas on parle de température radiative et d'émissivité directionnelle ou de manière intégrée sur la demi-sphère et on parle dans ce cas de température radiative et d'émissivité hémisphérique.

Ainsi, pour une classe de végétation  $i$ , si on fait l'hypothèse du corps gris et en intégrant spectralement sur toutes les longueurs d'onde, la luminance (flux radiatif émis par unité de surface vers l'atmosphère, supposé isotrope sur tout l'hémisphère supérieur) peut être exprimée comme suit :

$$L_i = \sigma T_{B_i}^4 = \sigma \epsilon_i T_{rad,i}^4, \quad (3.25)$$

où  $\sigma$  désigne la constante de Stefan-Boltzmann et  $T_{B_i}$  est la température de brillance (ou encore température radiative apparente  $T_B$ ) de la classe  $i$ .  $T_B$  correspond à la température d'un corps noir délivrant la même luminance que la surface étudiée. En effet, sur une bande spectrale  $[\lambda_1, \lambda_2]$  on a :

$$\int_{\lambda_1}^{\lambda_2} B_\lambda(T_B) d\lambda = \int_{\lambda_1}^{\lambda_2} \epsilon_\lambda B_\lambda(T) d\lambda \quad (3.26)$$

Avec ;

$B_\lambda$  est la luminance du corps noir

$\epsilon_\lambda$  est l'émissivité spectrale de la surface

La température de brillance varie selon les bandes spectrales des satellites utilisés. Dans SETHYS, l'angle de visée et d'ouverture du radiomètre peuvent être précisés pour le calcul d'une température de brillance directionnelle  $T_B(\theta)$  définie par [François \(1995\)](#) à partir des rayonnements émis par la surface.

### 3.1.4 Description générale du fonctionnement du modèle SETHYS

L'efficacité d'un modèle TSVA à simuler les variables des différents bilans (d'énergie et d'eau) à la surface est fortement contrainte par la représentation du sol (une description verticale du contenu en eau du sol) et de la végétation (type de végétation, LAI, hauteur de végétation, etc.), les incertitudes sur les conditions de forçages météorologiques ainsi que la complexité du modèle.

La description des différents phénomènes physiques pris en compte par SETHYS nécessite un grand nombre de paramétrisations. Il en résulte un nombre important de paramètres dont la valeur n'est pas souvent facile à définir et qui dépend du cadre de l'application du modèle (exemple : climat dominant, caractéristiques du sol de la région d'étude, type de végétation dominante, etc.). Ces paramètres doivent être optimisés afin d'assurer une meilleure estimation des variables de sortie du modèle.

Dans SETHYS, l'ensemble des conditions initiales (forçages météorologiques, états du sol : épaisseur du mulch, contenu en eau initial du sol) et des paramétrisations sont assurés par un certain nombre de fichiers dit fichiers d'entrée "de forçage" et de "paramètres". La prescription des valeurs des paramètres dans ces fichiers d'entrée permet le calcul des différentes variables des bilans d'énergies et de masse présentés précédemment. Il s'agit principalement de trois fichiers :

- Le fichier de paramètres : Ce fichier contient l'ensemble des valeurs des paramètres du sol et de la végétation ainsi que les paramètres d'initialisation (jours de début et de fin des simulations, pas de temps du modèle, conditions de visée, ...). Une liste exhaustive de ces paramètres ainsi que les processus qu'ils contrôlent et leur définition est donnée par le tableau 3.1. Les tableaux 3.3 et 3.2 présentent les valeurs de référence des paramètres ainsi que les intervalles de variation respectifs à chacun. Ces valeurs ont été déterminés à partir des paramétrisations d'un modèle d'agronomie (STICS [Brisson et al. \(2004\)](#)) appliqué à certains types de couverts de la région d'étude, de communications d'experts sur les propriétés des sols et de la végétation dans la région d'étude et des travaux précédents de [Saux Picart \(2007\)](#) et de [Coudert \(2006\)](#).

- Le fichier de forçage de la végétation : Ce fichier contient les valeurs journalières de LAI ainsi que les hauteurs des végétations.
- Le fichier de forçage atmosphérique : Ce fichier contient l'ensemble des conditions météorologiques acquises en continu sur la période d'étude (pas de temps de fonctionnement du modèle). Ces données comprennent : la température de l'air ( $T_a$ ), l'humidité spécifique de l'air ( $q_a$ ), la vitesse et la direction du vent ( $u_a$ ), le rayonnement atmosphérique incident ( $R_a$ ), le rayonnement solaire incident ( $R_g$ ) et les précipitations.

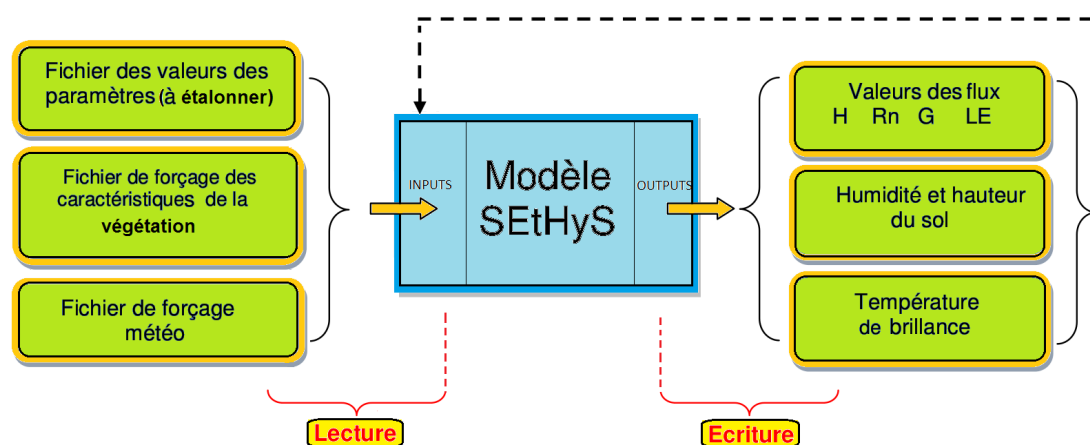


FIGURE 3.6 – Schéma descriptif du fonctionnement global du modèle SETHYS.

Le schéma de la figure 3.6 décrit de façon grossière les interactions entre le modèle SETHYS et les différents fichiers d'entrée/sortie. Ainsi, pour lancer une simulation SETHYS sur un type d'occupation de sol donné il faut impérativement :

- disposer de mesures journalières de LAI et de hauteur de couvert,
- disposer de mesures journalières des différentes variables, atmosphériques prescrites dans les fichiers de forçage atmosphériques et qui soient interpolées au pas de fonctionnement de SETHYS,
- définir le pas de calcul (fonctionnement) de SETHYS,
- prescrire les valeurs des paramètres de sol et de végétation nécessaires aux calculs des bilans radiatifs et hydriques présentés dans les parties précédentes.

Cependant, comme on vient de le dire précédemment, les valeurs des paramètres ne sont pas connues avec certitude et sont susceptibles de changer selon le type d'occupation du sol, selon le type de sol, selon le climat et les conditions météorologiques, etc. L'optimisation des ces paramètres a d'ailleurs fait l'objet de plusieurs travaux, dont les travaux de [Saux Picart \(2007\)](#) qui ont montré que certains paramètres peuvent même avoir une sensibilité saisonnière et journalière pour un type d'occupation du sol donné. Étant donnée que notre région d'étude, région Crau-Camargue au Sud-Est de la France, est très différente de la région étudiée par [Saux Picart \(2007\)](#), région du Sahel en Afrique de l'Ouest, il est donc nécessaire d'identifier les paramètres les plus sensibles (ou pas) afin de les étalonner et par conséquent avoir une meilleure estimation des variables de sortie en particulier la température de surface. D'où l'intérêt de réaliser une analyse de sensibilité des différents paramètres définis par le fichier de paramètres sur la région d'étude.

## 3.2 Analyse de sensibilité de la température de brillance vis-à-vis des paramètres du modèle SETHYS sur notre site d'étude

Notre but est, nous le rappelons, d'optimiser la température agrégée (à l'échelle des observations) en faisant varier les paramètres du modèle pour les différents types d'occupation du sol présents dans la zone d'intérêt de la région d'étude. Cependant, une meilleure estimation de la température n'est possible que si une "bonne" paramétrisation du modèle a lieu. Ceci nous conduit donc à la nécessité de bien étalonner ces paramètres suivant le type d'occupation du sol et les conditions météorologiques présentes sur la région d'étude. Au vu du nombre non négligeable de paramètres à l'entrée de SETHYS (22 paramètres, voir tableau 3.1) il est plus raisonnable d'effectuer une analyse de sensibilité pour identifier ceux qui ont le plus d'impact sur la sortie température du modèle. Ceci permettra de réduire la liste des paramètres à étalonner et réduire les risques d'optimisation mal conditionnée de cette sortie.

### 3.2.1 Introduction à l'Analyse de Sensibilité (AS)

Les méthodes d'analyse de sensibilité sont des outils d'une grande importance en modélisation numérique. Leur rôle étant de déterminer, pour un ensemble de sorties du modèle, les variables d'entrées les plus influentes, elles permettent ainsi de simplifier le modèle en fixant les entrées non influentes et de mieux l'appréhender en identifiant les relations entre les variables d'entrée.

D'après [Sorooshian S. and Gupta V.K. \(1995\)](#), on distingue deux grandes familles d'analyse de sensibilité :

**les méthodes d'analyse de sensibilité locales** : qui suivent une approche déterministe mettant en œuvre des hypothèses de linéarité et de normalité de la sortie du modèle par rapport à une variation locale des variables d'entrée en question.

**les méthodes d'analyse de sensibilité globales** : qui suivent une approche statistique et qui, par opposition aux méthodes locales, s'intéressent à l'ensemble du domaine de variation des variables d'entrée.

Le choix de la méthode d'analyse de sensibilité est conditionné par plusieurs facteurs entre autre le type du modèle (linéaire ou non-linéaire), le nombre de paramètres/variables d'entrée, le temps de calcul du modèle (durée d'une simulation), etc. Il est évident que les méthodes globales présentent plus de liberté par rapport aux méthodes locales, au niveau de la modélisation (pas d'hypothèse ni de restriction sur le domaine de variation des variables d'entrées). Cependant, vu le nombre important de méthodes d'analyse globale (les heuristiques, les plans d'expériences, les méthodes de type Monte Carlo, etc.), le choix le plus dur est donc le choix de la méthode à adopter. Dans notre cas, celle-ci doit remplir deux conditions : la facilité de la mise en œuvre et la facilité de l'identification des paramètres sensibles avec un minimum de coût de calcul. Dans sa revue sur les méthodes d'analyse de sensibilité, [Iooss \(2011\)](#) fait la synthèse des principales méthodes relatives à la variabilité de la sortie d'un modèle numérique présenté par la figure 3.7 en plus du diagramme de décision de [de Rocquigny et al. \(2008\)](#) (figure 3.8). Ce qu'il faut retenir des méthodes de sensibilité globale est qu'il existe deux types de méthodes : les méthodes basées sur le criblage qui permettent de réaliser une analyse "qualitative" des entrées en question en classant ces derniers par ordre d'importance et les méthodes basées sur la décomposition de la variance qui permettent de réaliser une analyse quantitative en estimant la contribution de chaque paramètre à la variance de la sortie étudiée et qui permettent en plus d'étudier les interactions entre les paramètres d'entrée.

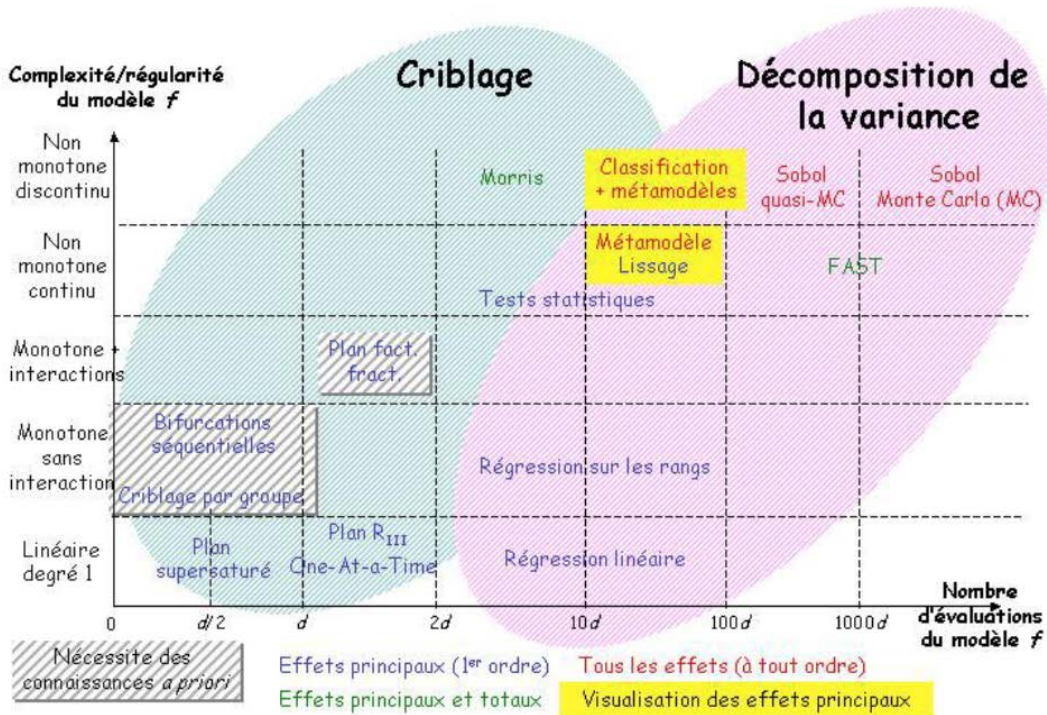


FIGURE 3.7 – Synthèse détaillée des méthodes d’AS (Figure extraite de Iooss (2011)).

Notre choix s’est plutôt dirigé vers les méthodes de décomposition de la variance et en particulier les méthodes stochastiques du type Monte Carlo. La méthode d’AS qu’on a adopté est connue sous le nom d’analyse de Sobol. C’est une méthode très répandue en AS pour traiter les systèmes de faible nombre d’entrées (une dizaine d’entrées) avec un coût de calcul intéressant (voir Figure 3.8). L’avantage d’utiliser une méthode stochastique basée sur un échantillonnage Monte Carlo est de pouvoir explorer le plus de combinaisons possibles des valeurs plausibles des entrées du modèle et de converger vers un espace de solutions, et non pas une solution unique, avec une incertitude qui est d’autant plus faible que l’échantillonnage de l’espace des entrées du modèle est fin. L’analyse de Sobol permet en plus d’identifier les éventuelles corrélations entre certaines entrées du modèle par le simple calcul de deux indices dits indice “Total” et indice de “premier ordre” que l’on détaillera par la suite.

Le modèle SETHYS, a déjà fait l’objet de deux analyses de sensibilité sur deux autres régions. La première analyse a été faite sur une ancienne version de SETHYS (version qui représente la biosphère en deux sources uniquement) par la méthode MCIP (*Multi-objective Calibration Iterative Process*) dans les travaux Demarty et al. (2005); Coudert (2006). La région d’étude était une parcelle de type agricole dont les couverts sont principalement du blé et de l’orge et qui est située dans les Alpilles en France. Mais, comme la modélisation de la végétation dans le modèle a beaucoup changé entre la version utilisée dans Coudert (2006) et la version actuelle, les paramètres sensibles peuvent aussi changer.

La deuxième analyse de sensibilité a été réalisée par Saux Picart (2007) sur la version actuelle du modèle (version 3 sources) avec la méthode MCIP. La région d’étude était

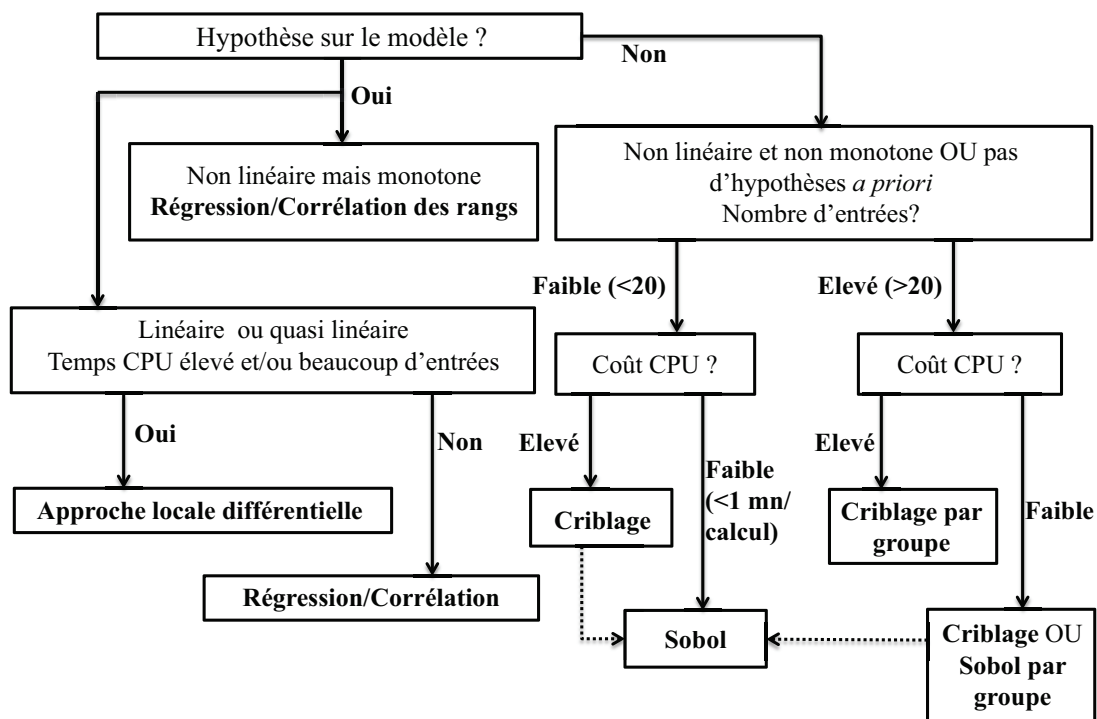


FIGURE 3.8 – Diagramme de décision pour le choix de la méthode d'AS appropriée (Figure extraite de de Rocquigny et al. (2008)).

cependant une région de climat aride au Sahel au Niger dont le climat et les types de couverts (principalement du Mil et de la jachère) sont très différents de ceux de notre région d'étude (climat méditerranéen et couvert de type agricole principalement). D'où l'intérêt de réaliser une nouvelle analyse de sensibilité sur la région de Crau-Camargue. La méthode MCIP fait aussi partie des méthodes d'analyse de sensibilité globale de type MC cependant l'avantage de choisir la méthode de Sobol est de pouvoir identifier les éventuelles corrélations et quantifier les interactions entre les entrées du modèle SETHYS.

### 3.2.2 Analyse de Sobol

L'analyse de Sobol, encore connue sous le nom "ANOVA" (ANalysis Of VAriance) (Chan et al. (1997); Sobol (2001); Saltelli (2002); Christopher Frey and Patil (2002); Sobol et al. (2007); Saltelli et al. (2010)), est comme l'indique son deuxième nom, une analyse de la variance. Elle cherche à établir le lien entre la variation introduite à une variable d'entrée d'un modèle donné sur la variance de la sortie de ce dernier. Le principe de l'analyse de Sobol est très simple et se base sur le calcul de certains indices dits "indices de Sobol" compris entre "0" et "1" qui présentent la sensibilité de la sortie du modèle par rapport à une ou plusieurs variables d'entrée.

Pour calculer ces indices on suppose que l'on dispose d'un modèle déterministe  $M$  (physique, mathématique, économique, etc.) dont l'expression analytique est très complexe. Soient  $(X_i)_{i=1,\dots,d}$  les paramètres/variables d'entrée  $M$  et  $Y$  la sortie ( $Y$  est un scalaire).



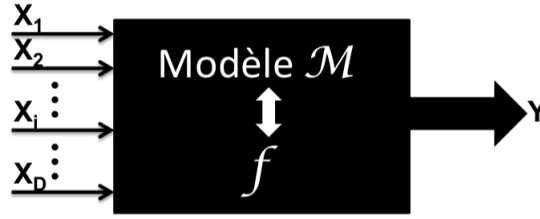


FIGURE 3.9 – Schéma de présentation du modèle en boîte noire

Ce modèle peut être vu comme étant une boîte noire ayant pour entrées/sorties les entrées/sorties de  $M$  (Figure 3.9). Il est ainsi possible d'imaginer une fonction  $f$  tel que :

$$Y = f(X_1, X_2, \dots, X_d) \quad (3.27)$$

Si les variables  $(X_i)_{i=1,\dots,d}$  sont indépendantes et identiquement distribuées (iid) sur  $[0, 1]$ , et si  $f$  est carrée-intégrable sur  $[0, 1]$  (il est toujours possible de définir les variables d'entrée d'un modèle sur  $[0, 1]$  en normalisant leurs intervalles de définition respectifs sous réserve que ces derniers soient bornés) alors selon [Hoeffding \(1948\)](#), il existe une unique décomposition de  $f$  en fonctions orthogonales telle que :

$$Y = f(X) = f_0 + \sum_{i=1}^d f_i(X_i) + \sum_{i=1}^d \sum_{j>i}^d f_{ij}(X_i, X_j) + \dots + f_{12\dots i\dots d} \quad (3.28)$$

tel que :

$$f_0 = E(Y), \quad (3.29)$$

$$f_i = E_{\sim X_i}(Y|X_i) - f_0, \quad (3.30)$$

$$f_{i,j} = E_{\sim X_{ij}}(Y|X_i, X_j) - f_i - f_j - f_0, \quad (3.31)$$

et ainsi de suite pour les ordres d'indices supérieurs, avec  $\sim X_i$  présente la matrice des valeurs possibles des variables d'entrée du modèle en fixant la  $i$ -ème variable. En intégrant l'équation 3.28 sur  $(0, 1)$ , on obtient la décomposition de la variance totale en variances partielles introduites par [Sobol \(1993\)](#) tel que :

$$V(Y) = \sum_{i=1}^d V_i + \sum_{i=1}^d \sum_{j>i}^d V_{i,j} + \dots + V_{12\dots i\dots d}, \quad (3.32)$$

avec :

$$V_i = V(f_i) = V_{X_i}(E_{\sim X_i}[Y|X_i]) \quad (3.33)$$

$$V_{ij} = V(f_{ij}) = V_{X_i X_j}(E_{\sim X_{ij}}[Y|X_i, X_j]) - V_i - V_j, \quad (3.34)$$

et ainsi de suite pour les ordres d'indices supérieurs.  $V_{X_i}$  présente la variance de la sortie calculée sur toutes les valeurs possibles prises par la  $i$ -ème variable dans la matrice  $\sim X_i$ . A partir de l'équation 3.32, on remarque que la variance totale est une somme de variances élémentaires de la sortie par rapport aux variables d'entrée variant indépendamment ( $V_i$ ), deux à deux ( $V_{ij}$ ), trois à trois ( $V_{ijk}$ ), et ainsi de suite.

Si on divise de part et d'autre l'équation 3.32 par  $V(Y)$  on obtient :

$$\sum_{i=1}^d S_i + \sum_{i=1}^d \sum_{j>i} S_{i,j} + \dots + S_{12\dots i\dots d} = 1, \quad (3.35)$$

où  $S_i$  désignent les indices de Sobol de premier ordre,  $S_{ij}$  les indices de second ordre, et ainsi de suite pour les ordres supérieurs. L'expression de  $S_i$  est donc la suivante :

$$S_i = \frac{V_{X_i}(E_{\sim X_i}[Y|X_i])}{V(Y)} \quad (3.36)$$

D'après Mood et al. (1974) :

$$V(Y) = V_{X_i}(E_{\sim X_i}[Y|X_i]) + E_{X_i}[V_{\sim X_i}(Y|X_i)], \quad (3.37)$$

où  $V_{X_i}(E_{\sim X_i}[Y|X_i])$  présente la réduction de la variance totale attendue (*expected*) obtenue en fixant la variable  $X_i$  (ce qu'on appelle aussi l'effet de premier ordre de la variable  $X_i$  sur la sortie  $Y$ ) et  $E_{X_i}[V_{\sim X_i}(Y|X_i)]$  présente la variance attendue restante en fixant toutes les variables sauf  $X_i$  (appelé souvent résiduel).

En divisant l'équation 3.37 par  $V(Y)$  on retrouve l'expression de l'indice "Total" de Sobol introduit en premier par Homma and Saltelli (1996) comme suit :

$$S_{T_i} = \frac{E_{X_i}[V_{\sim X_i}(Y|X_i)]}{V(Y)} = 1 - \frac{V_{X_i}(E_{\sim X_i}[Y|X_i])}{V(Y)} \quad (3.38)$$

En effet si l'indice de premier ordre exprime l'effet **individuel** de la variation de la  $i$ -ème variable sur la variance totale du modèle, l'indice "Total" quant à lui, indique l'effet de la variation de cette variable en interactions avec les autres variables d'entrée. De cette manière il est clair que si l'indice total et l'indice d'ordre supérieur sont très proches, cela suppose que les indices d'ordres supérieurs sont tous négligeables car la somme des indices de Sobol est égale à 1 et qu'il sont tous positifs.

Pour réaliser une analyse de Sobol complète (calcul de tous les indices de Sobol) il faut calculer  $2^{d-1}$  indices. Il devient donc compliqué de calculer ces indices dès que le nombre d'entrées dépasse une dizaine. Cependant, il existe un moyen de réduire ce nombre à  $2 \times d$ . En effet, l'idée consiste à commencer par le calcul des indices  $S_i$  et  $S_{T_i}$ , et si ces derniers sont très proches alors il n'est plus nécessaire de calculer les indices d'ordres supérieurs.

Saltelli et al. (2010) ont proposé une méthode optimisée permettant de calculer d'un seul coup les indices  $S_i$  et  $S_{T_i}$ .

L'idée est la suivante : supposons que l'on dispose de deux matrices  $A$  et  $B$  avec  $a_{ij}$  et  $b_{ij}$  leurs termes génériques respectifs, l'indice  $i$  désigne la variable d'entrée et varie de 1 à  $d$  et l'indice  $j$  désigne le nombre de simulations et varie de 1 à  $N$  ( $N$  est la taille de l'ensemble de tirage Monte Carlo des variables d'entrée). Soit alors la matrice  $A_B^{(i)}$  ( $B_A^{(i)}$ ) construite à partir des colonnes de la matrice  $A$  ( $B$ ) à l'exception de la  $i$ -ème colonne qui correspond à la  $i$ -ème colonne de la matrice  $B$  ( $A$ ).

Selon Saltelli et al. (2010)  $S_i$  pourrait être calculée avec l'un ou l'autre des couples de matrice  $(A, B_A^{(i)})$  et  $(B, A_B^{(i)})$  comme suit :

$$S_i \propto V_{X_i}(E_{X \sim i}[Y|X_i]) = \frac{1}{N} \sum_{j=1}^N f(A)_j f(B_A^{(i)})_j - f_0^2. \quad (3.39)$$

À partir des équations 3.37 et 3.35 que :

$$S_{T_i} \propto V_{X \sim i}(E_{X \sim i}[Y|X_i]) = \frac{1}{N} \sum_{j=1}^N f(A)_j f(A_B^{(i)})_j - f_0^2. \quad (3.40)$$

Les expressions de  $S_i$  et de  $S_{T_i}$  présentées ci-dessus découlent de la définition de la variance donnée par :  $V(Y) = E[Y^2] - (E[Y])^2$  (Mood et al. (1974)). La démonstration des équations 3.39 et 3.40 a été présentée en détails dans Saltelli et al. (2010).

De cette manière il est clair que le nombre de simulations total est  $N \times (d + 2)$  (à la place de  $2 \times N \times d$ ) dans le cas où les indices sont calculés de manière indépendante). En pratique, une taille d'échantillon de l'ordre de 10000 est suffisante pour estimer la sensibilité d'un modèle à une dizaine de variables d'entrée.

### 3.2.3 Calcul des indices de Sobol $S_i$ et $S_{T_i}$ pour le modèle SETHYS

Une fois le principe et les équations présentés, on peut passer à l'application de cette analyse de sensibilité sur le modèle SETHYS. SETHYS possède plusieurs sorties (Saux Picart (2007); Coudert (2006)), cependant dans tout ce qui suit on s'intéresse uniquement à la sortie température de brillance de SETHYS. Les forçages atmosphériques et les forçages de végétation utilisés au cours de cette analyse ont été récupérés par l'équipe EMMAH (Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes, <https://www6.paca.inra.fr/emmah>) sur la région d'étude, "Crau-Camargue" au cours de l'année 2009. Les 7 types d'occupations de sol (principalement de type agricole) ont été considérés à compter : le sol nu, l'eau, la prairie, des cultures variées (sorgho, maïs, tournesol), la forêt, le blé et le riz. Les variables d'entrée de l'analyse de Sobol sont décrites par les tableaux 3.2 et 3.3 présentant respectivement l'ensemble des paramètres communs et spécifiques à chaque type d'occupation du sol. Certaines valeurs des paramètres de sol et de végétation sont prises en se basant sur une paramétrisation du modèle de culture *STICS* (Brisson et al. (1998, 2003, 2004)), d'autres ont été définies en se basant sur des connaissances générales de certaines propriétés des types de végétation sur la région "Crau-Camargue". Les intervalles de variation (incertitudes) ont été définis en se basant sur les travaux de Saux Picart (2007) et Coudert (2006).

Au vu du nombre important de paramètres de SETHYS (22 paramètres décrits par le tableau 3.1), la taille de notre échantillon doit être très grande pour avoir des indices significatifs à l'issue de l'analyse de Sobol. Afin de diminuer la taille de l'échantillon, on a réduit le nombre de paramètres à l'entrée de l'analyse de Sobol en éliminant certains paramètres connus comme étant non sensibles à la sortie température de SETHYS. En effet, on réalise une analyse préliminaire de la variance de la sortie température de SETHYS pour les différents types d'occupation du sol en faisant varier à chaque fois un des paramètres décrits par le tableau 3.1. Cette analyse a permis d'identifier les paramètres qui ne sont pas sensibles car la variance de la sortie obtenue était quasi-nulle pour ces paramètres. Les paramètres décrits par les tableaux 3.2 et 3.3 présentent ceux qui ont donné des variances non nulles. De cette manière on a pu réduire le nombre d'entrée à 13 paramètres (à peu près une dizaine d'entrées). En pratique pour une dizaine d'entrées il est recommandé d'avoir une taille de l'échantillon supérieure ou égale à  $N = 10000$ . Dans notre cas on a utilisé un échantillon de plus petite taille ( $N = 800$ ) pour les 13 variables d'entrée. En effet le choix d'une taille assez grande de l'échantillon permet d'échantillonner le plus finement possible l'espace des paramètres. Notre idée était de tirer profit des méthodes sophistiquées de ré-échantillonnage quasi-aléatoires ou encore dites quasi-Monte Carlo basées sur des séquences dites à faible écart (*low discrepancy, LD*) qui permettent de réaliser un échantillonnage plus performant qu'un échantillonnage issu d'un générateur de nombres aléatoires classique.

L'algorithme utilisé est celui proposé par Owen (1998) et Faure and Tezuka (2002). L'intérêt d'utiliser ce type de ré-échantillonnage revient en grande partie à l'uniformité de l'échantillon qu'il génère et aux très faibles erreurs de ré-échantillonnage introduites comparé à un générateur de nombres aléatoires classique (Owen (1998), Faure and Tezuka (2002)).

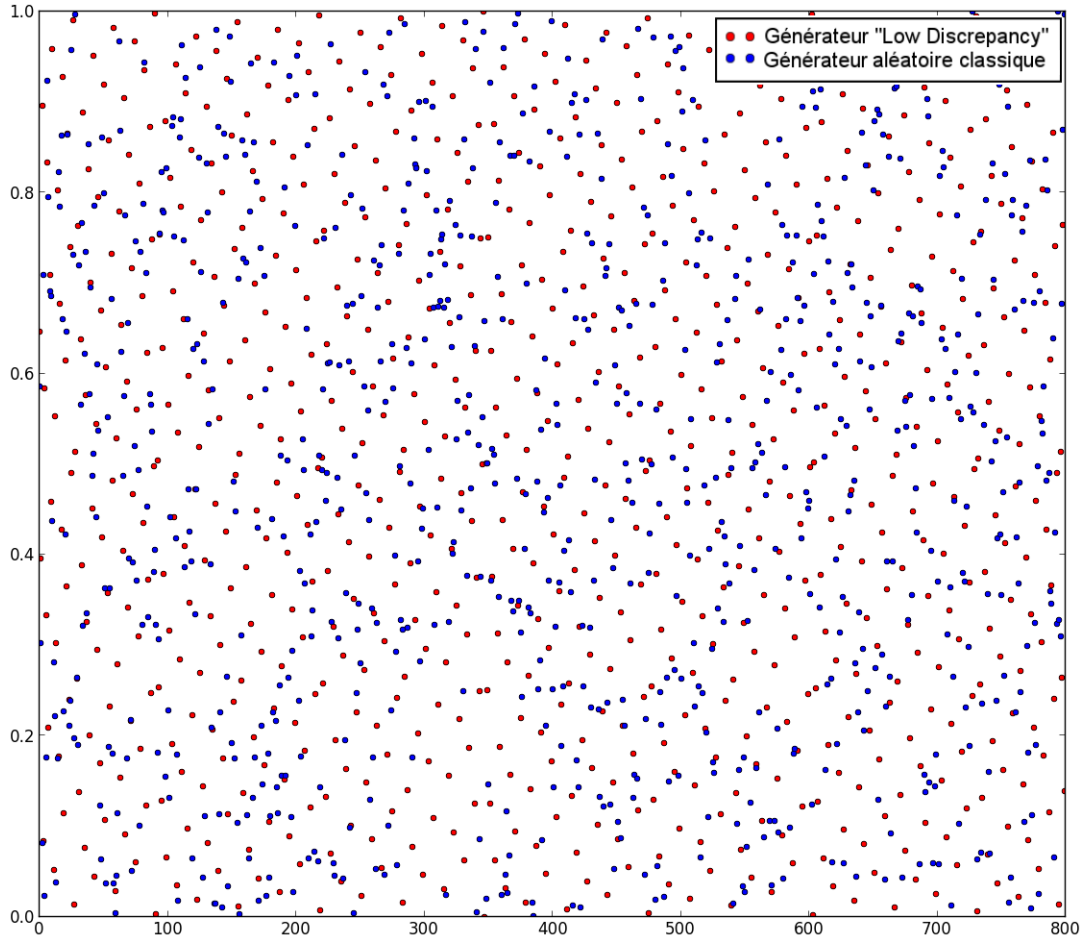


FIGURE 3.10 – Comparaison d'un échantillon issu d'un générateur de nombres aléatoires classique de la loi uniforme (couleur bleue) et d'un échantillon issu d'un générateur LD (couleur rouge) : les tailles des échantillons sont égales à  $N = 800$ .

La figure 3.10 présente la superposition d'un échantillon issu d'un générateur aléatoire classique de la distribution uniforme (couleur bleue) et d'un échantillon issue d'un générateur LD (en rouge) de tailles égales ( $N = 800$ ). Bien que le générateur de nombre aléatoire classique et celui utilisant un algorithme LD produisent des séquences réparties uniformément, ils sont très différents quant au principe. De plus, cette illustration montre que le générateur LD a donné un échantillonnage plus uniforme de l'hypercube unité. Ceci pourrait s'expliquer comme suit. Supposons que l'on dispose d'un générateur aléatoire uniforme classique sur  $[0, 1)$ . Ce dernier va produire des points de sorte que chaque tirage ait la même probabilité de tirer un point dans l'un des sous-intervalles égaux répartis sur l'intervalle  $[0, 1)$ . Par exemple, si les sous-intervalles égaux sont  $[0, \frac{1}{2})$  et  $[\frac{1}{2}, 1)$ , il est possible que  $(N - 1)$  tirages se trouvent tous dans la première moitié de l'intervalle, tandis que le  $N$ -ième tirage se trouve dans l'autre moitié, ce qui crée une erreur de ré-échantillonnage du fait que la probabilité du point  $N$  est égale à  $\frac{1}{2}$  dans ce cas particulier.

L'avantage des séquences quasi-aléatoires revient au fait que la génération des points est contrainte par une condition de "faible écart" entre ces derniers d'où le nom *low discrepancy*. C'est comme si chaque point *connaissait* les places occupées par les autres points qui l'ont précédé (forte corrélation) et choisissait donc sa place de manière à optimiser son écart par rapport aux autres points occupants l'hypercube unité. En utilisant ce type d'algorithme de ré-échantillonnage on explore de manière plus efficace l'espace des paramètres sans pour autant utiliser des échantillons de très grandes tailles.

L'analyse a été menée sur 3 périodes différentes de l'année 2009. La première période ( $P_1$ ) s'étend du 1<sup>er</sup> au 7 Février de la saison hivernale. Pour cette période, les végétations n'ont pas encore poussé et les paramètres du sol sont normalement les plus sensibles. La deuxième période ( $P_2$ ) s'étend du 1<sup>er</sup> au 7 Mai durant le printemps lors de la croissance des végétations. Durant cette période les paramètres de végétation sont supposés les plus sensibles. Et la dernière période ( $P_3$ ) est choisie en fin d'été du 1<sup>er</sup> au 7 Septembre pour couvrir une partie de la période de sénescence des végétations.

Le choix de faire l'analyse de Sobol sur des périodes différentes de l'année est en grande partie basé sur les résultats de [Coudert \(2006\)](#) et de [Saux Picart \(2007\)](#) qui ont montré que la sensibilité des paramètres varie à différentes échelles temporelles : selon la saison, et dans la journée (en particulier les paramètres liés aux processus dominants sur la température tel que l'évapotranspiration). D'où notre motivation à décomposer l'analyse sur des parties différentes de l'année et sur plusieurs parties de la journée. Pour cela on a défini une fenêtre temporelle de 4h sur laquelle l'analyse sera faite à chaque fois. Les fenêtres temporelles sont définies comme suit : 00h00 → 04h00, 04h00 → 08h00, 08h00 → 12h00, 12h00 → 16h00, 16h00 → 20h00 et 20h00 → 24h00. Ce choix d'analyse par partie permet de prendre en compte la variabilité diurne de certains paramètres.

### 3.2.4 Résultats de l'analyse de Sobol

Le tableau 3.4 résume les résultats de l'analyse de Sobol pour tous les types de couvert considérés. D'après le tableau, on a constaté que pour ( $P_1$ ) seuls les paramètres du sol ( $fact_{therm}$  et  $alb_{sec}$ ) sont apparus sensibles quelque soit la période de la journée. Durant la période  $P_2$  on a remarqué que les paramètres de végétation sont devenus sensibles et que leur sensibilité est plus importante en milieu de journée lorsque l'évapotranspiration est maximale. On reconnaît alors les paramètres liés à l'évapotranspiration (exemple :  $Vmax0$ ,  $alb_g$ ,  $e_g$ ). La période  $P_3$ , montre une chute de la sensibilité des paramètres de végétation pour la plupart des couverts végétaux (à l'exception de la classe de riz qui a un cycle de croissance décalé vu qu'il s'agit d'une culture d'été et en début de septembre on est toujours en période de croissance).

Les figures 3.11, 3.12 et 3.13 présentent les résultats de l'analyse de Sobol pour la classe blé pour les trois périodes considérées. La décision sur la sensibilité ou non d'un paramètre est faite comme suit : si l'indice de sensibilité estimé est supérieur à 0.01, le paramètre est retenu comme étant sensible. En comparant les valeurs obtenues pour les indices totaux et les indices de premier ordre on constate que les paramètres sont indépendants car les valeurs des deux indices sont égales et qu'il n'est donc pas nécessaire de calculer les indices d'ordre supérieurs. L'analyse de Sobol nous a permis d'identifier 5 paramètres sensibles par type d'occupation du sol.

Les résultats obtenus avec notre analyse de sensibilité rejoignent ceux trouvés dans les travaux de [Coudert \(2006\)](#) et de [Saux Picart \(2007\)](#). Les paramètres retenus sensibles pour chaque classe seront étalonnés au cours de la procédure de désagrégation par fil-

trage particulaire. Le chapitre 5 présente plus de détails sur l'implémentation du filtre particulaire sur le modèle SETHYS ainsi que les premiers résultats de l'application du filtre sur des pseudo-observations et sur des données réelles de télédétection.

## Conclusion

Cette partie a permis de donner un aperçu sur les différents mécanismes et phénomènes physiques gérés par le modèle TSVA en particulier le modèle SETHYS. L'étude de sensibilité a permis par la suite de sélectionner les paramètres qui ont le plus d'impact sur la variation de la température radiative à la sortie du modèle. Les résultats obtenus à l'issue de cette analyse de sensibilité seront utilisés dans le chapitre 6 pour définir les paramètres à étalonner pour les différents types d'occupation du sol.

Paramètres	Description	Unité	Processus
$fact_{therm}$	Facteur thermique corrigeant l'inertie thermique du sol	$J.m^{-3}.K^{-1}$	Calcul du flux thermique dans le sol.
$e_s$	Émissivité du sol nu	–	Processus radiatifs thermiques (Équations 3.2 3.26).
$alb_{sec}$	Albédo sec du sol nu	–	Processus radiatifs solaires.
$w_{sat}$ et $w_{resid}$	Humidité volumique à saturation et résiduelle	$m^3.m^{-3}$	Transferts hydriques dans le sol : rétention en eau, conductivité hydraulique, drainage, ruissellement et diffusion (Équations 3.21, 3.20, 3.22, 3.23 et 3.24).
$k_{sat}$ et $k_{sat}(1)$	Conductivité hydraulique à saturation (>10 cm) et conductivité hydraulique à saturation de surface (>10cm)	$m.s^{-1}$	Transferts hydriques dans le sol : rétention en eau, conductivité hydraulique, drainage, ruissellement et diffusion (Équations 3.20, 3.22, 3.23 et 3.24).
$\tau$	Tortuosité du sol	–	Calcul de l'évaporation du sol (Équation 3.19).
$D_v$	Diffusivité de la vapeur d'eau dans le sol	$m^2.s^{-1}$	Calcul de l'évaporation du sol (Équation 3.19).
$b$	Pente de la courbe de rétention	–	Calcul de l'état hydrique du sol (Équation 3.21).
$\psi_{sat}$	Potentiel hydrique du sol à saturation	m	Calcul d'état hydrique du sol (Équation 3.21).
$Z_{sol1}$	Épaisseur de la couche "mulch" de surface	mm	Calcul de l'évaporation du sol (Équation 3.19).
$e_g$ et $e_t$	Émissivité de la végétation haute et de la végétation basse	–	Processus radiatifs thermiques au niveau de la végétation (Équations 3.2 3.26).
$alb_g$ et $alb_t$	Albédo de la végétation basse et haute	–	Processus radiatifs solaires au niveau de la végétation.
$lf_g$ et $lf_t$	Largeur des feuilles de la végétation basse et haute	m	Transpiration des deux types de végétation : calcul des résistances de couche limite et totales du couvert (Équations 3.14, 3.15, 3.17 et 3.16).
$V_{max0}(g, t)$	Capacité photosynthétique des feuilles	–	Transpiration des deux types de végétation (végétation basse "g", et haute "t").
$P_{hc}(g, t)$	Potentiel hydrique semi-critique de la végétation	m	Transpiration des deux types de végétation.

TABLE 3.1 – Paramètres du modèle SETHYS

3.2. Analyse de sensibilité de la température de brillance vis-à-vis des paramètres du modèle SETHYS sur notre site d'étude

Paramètres	Intervalle de variation	Valeur de référence
$e_s$	[0.93, 0.97]	0.94
$\tau$	[1.00, 3.00]	2.00
$e_g$	[0.96, 1.00]	0.97
$e_t$	[0.96, 1.00]	0.99
$alb_g$	[0.10, 26.00]	0.18
$alb_t$	[0.15, 0.33]	0.24
$lf_t$	[5.00, 15.00] $10^{-3}$	$10^{-2}$
$V_{max_0}(1)$	[20.00, 80.00]	50.00
$V_{max_0}(2)$	[50.00, 150.00]	100.00

TABLE 3.2 – Paramètres communs aux différents types d'occupation des sols

Classes	$fact_{term}$	$alb_{sec}$	$w_{sat}$	$lf_g$
Sol nu	[1.20, 2.20], 1.70	[0.25, 0.35], 0.3	[0.20, 0.30], 0.25	–, –
Eau	[2.20, 3.20], 2.70	[0.25, 0.35], 0.3	[0.20, 0.30], 0.25	–, –
Prairie	[1.10, 2.10], 1.60	[0.20, 0.30], 0.25	[0.25, 0.35], 0.30	[5.00, 20.00] $10^{-3}$ , $10 \cdot 10^{-3}$
Cultures	[1.25, 2.25], 1.50	[0.20, 0.30], 0.25	[0.25, 0.35], 0.26	[9.00, 100.00] $10^{-3}$ , $48 \cdot 10^{-3}$
Forêt	[2.03, 3.03], 2.76	[0.25, 0.35], 0.3	[0.20, 0.30], 0.25	[5.00, 15.00] $10^{-3}$ , $10 \cdot 10^{-3}$
Blé	[1.25, 2.25], 1.50	[0.20, 0.30], 0.25	[0.25, 0.35], 0.26	[9.00, 100.00] $10^{-3}$ , $48 \cdot 10^{-3}$
Riz	[2.25, 3.25], 2.75	[0.20, 0.30], 0.25	[0.40, 0.50], 0.45	[5.00, 15.00] $10^{-3}$ , $10 \cdot 10^{-3}$

TABLE 3.3 – Paramètres variables selon le type d'occupation des sols



TABLE 3.4 – Résultats de l'analyse de Sobol effectuée sur la région Crau-Camargue avec les données de forçage de l'année 2009

Heure/Période	Sol nu	Eau	Prairie	Cultures	Forêt	Blé	Riz
0 - 4h	$P_1$	$fact_{therm}$	$fact_{therm}, w_{sat}$	$fact_{therm}$	$fact_{therm}$	$fact_{therm}, e_g, e_t$	$fact_{therm}, w_{sat}$
	$P_2$	$fact_{therm}$	$fact_{therm}, w_{sat}, e_s$	$fact_{therm}, e_g$	$fact_{therm}$	$fact_{therm}, e_g, e_t, lft$	$fact_{therm}, w_{sat}, e_s$
	$P_3$	$fact_{therm}$	$fact_{therm}, e_s, w_{sat}$	$fact_{therm}, e_g$	$fact_{therm}$	$fact_{therm}, e_g, e_t, lft$	$fact_{therm}, e_g$
4 - 8h	$P_1$	$fact_{therm}$	$fact_{therm}, w_{sat}$	$fact_{therm}$	$fact_{therm}$	$fact_{therm}, e_g, e_t$	$fact_{therm}, w_{sat}$
	$P_2$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}, V_{max}, alb_g, e_g$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_g, e_t, alb_t$	$fact_{therm}, alb_{sec}, V_{max}, e_g$
	$P_3$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_g, alb_{sec}, V_{max}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_g, e_t, lfg, lft, alb_t$	$fact_{therm}, e_g, V_{max}$
8 - 12h	$P_1$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_g, e_t, lfg$	$fact_{therm}, alb_{sec}$
	$P_2$	$alb_{sec}$	$alb_{sec}$	$fact_{therm}, alb_{sec}, V_{max}, alb_g, e_g$	$alb_{sec}$	$alb_t, e_g, e_t, lfg, lft$	$fact_{therm}, alb_{sec}, V_{max}, e_g$
	$P_3$	$fact_{therm}, alb_{sec}$	$alb_{sec}$	$fact_{therm}, e_g, alb_{sec}, V_{max}$	$fact_{therm}, alb_{sec}$	$alb_{sec}, alb_t, e_g, e_t, lfg, lft$	$alb_{sec}, e_g, V_{max}$
12 - 16h	$P_1$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_g, e_t, lfg, alb_g$	$fact_{therm}, alb_{sec}$
	$P_2$	$alb_{sec}$	$alb_{sec}$	$alb_{sec}, V_{max}, alb_g, e_g$	$alb_{sec}$	$alb_{sec}, alb_t, e_g, e_t, lft$	$alb_{sec}, V_{max}, e_g$
	$P_3$	$fact_{therm}, alb_{sec}$	$alb_{sec}$	$fact_{therm}, e_g, alb_{sec}, V_{max}$	$alb_{sec}$	$alb_{sec}, alb_t, e_g, e_t, lfg, lft$	$alb_{sec}, e_g, V_{max}$
16 - 20h	$P_1$	$fact_{therm}$	$fact_{therm}, w_{sat}$	$fact_{therm}$	$fact_{therm}$	$fact_{therm}, e_g, e_t$	$fact_{therm}, w_{sat}$
	$P_2$	$fact_{therm}, alb_{sec}$	$fact_{therm}, alb_{sec}, e_s, w_{sat}$	$fact_{therm}, alb_{sec}, e_g$	$fact_{therm}, alb_{sec}$	$alb_t, e_g, e_t, lft$	$fact_{therm}, alb_{sec}, e_g$
	$P_3$	$fact_{therm}, alb_{sec}$	$fact_{therm}, e_s, w_{sat}, alb_{sec}$	$fact_{therm}, e_g, alb_{sec}$	$fact_{therm}, alb_{sec}$	$alb_{sec}, alb_t, e_g, e_t, lfg, lft$	$fact_{therm}, alb_{sec}, e_g$
20 - 0h	$P_1$	$fact_{therm}$	$fact_{therm}, w_{sat}$	$fact_{therm}$	$fact_{therm}$	$fact_{therm}, e_g, e_t$	$fact_{therm}, w_{sat}$
	$P_2$	$fact_{therm}$	$fact_{therm}, alb_{sec}, e_s, w_{sat}$	$fact_{therm}, e_g$	$fact_{therm}$	$fact_{therm}, e_g, e_t, lfg, lft$	$fact_{therm}, alb_{sec}, w_{sat}, e_s$
	$P_3$	$fact_{therm}$	$fact_{therm}, e_s, w_{sat}$	$fact_{therm}, e_g$	$fact_{therm}$	$fact_{therm}, e_g, e_t, lfg, lft$	$fact_{therm}, alb_{sec}, e_g$

### 3.2. Analyse de sensibilité de la température de brillance vis-à-vis des paramètres du modèle SETHYS sur notre site d'étude

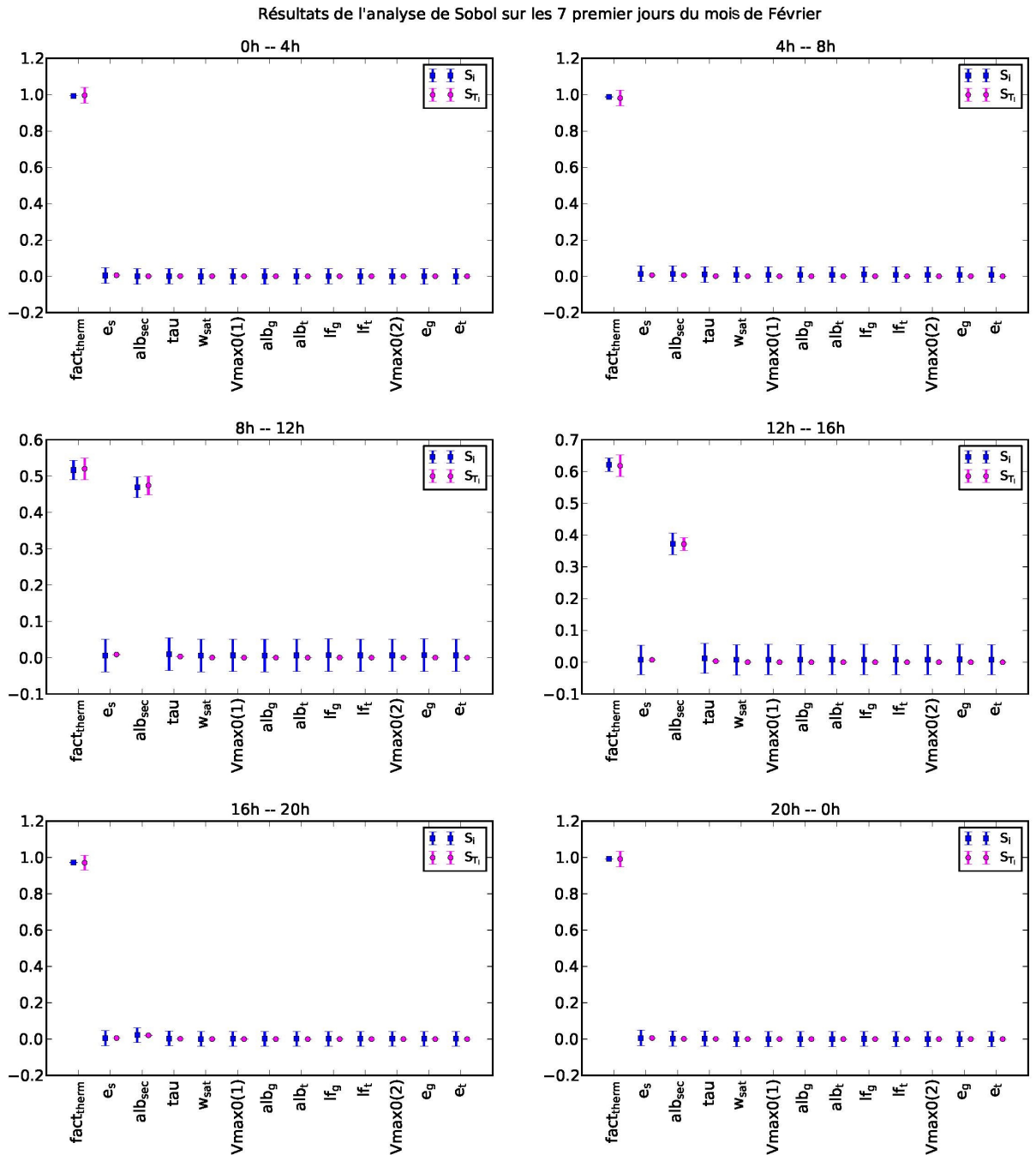


FIGURE 3.11 – Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/02/2009 au 07/02/2009.

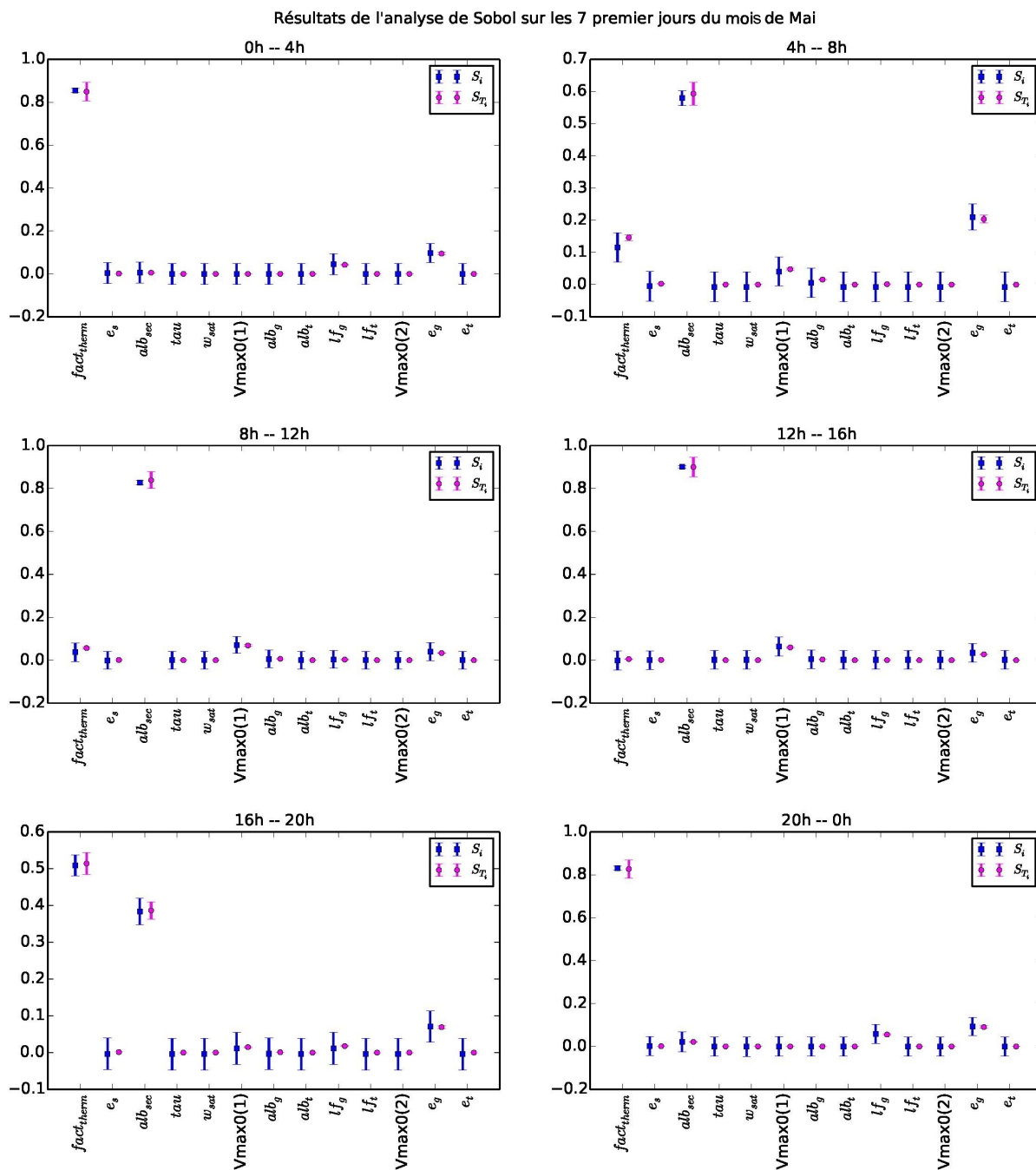


FIGURE 3.12 – Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/05/2009 au 07/05/2009.

### 3.2. Analyse de sensibilité de la température de brillance vis-à-vis des paramètres du modèle SETHYS sur notre site d'étude

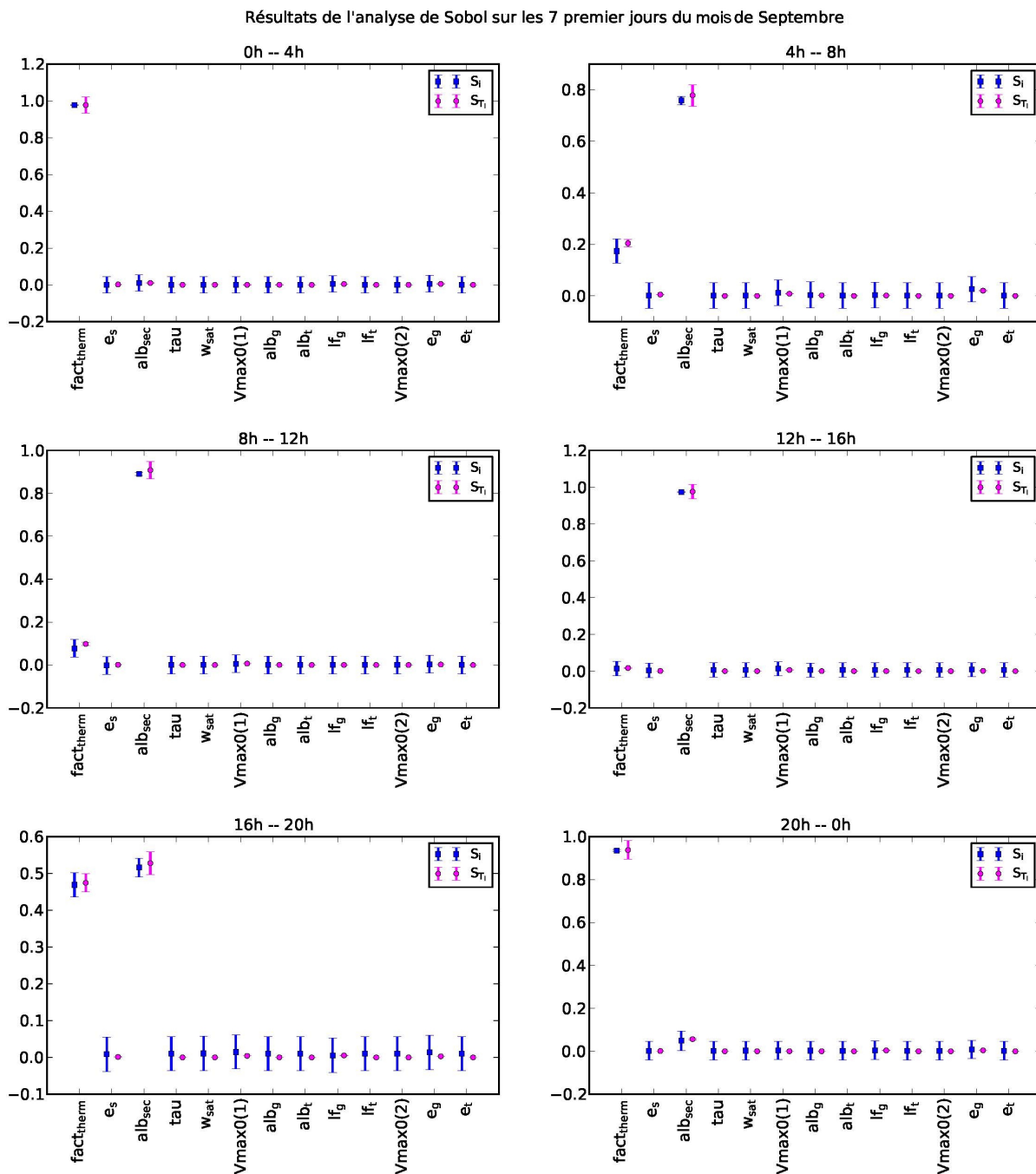


FIGURE 3.13 – Résultats de l'analyse de sensibilité sur la parcelle de blé pour la période allant du 01/09/2009 au 07/09/2009.



---

## Région d'étude et base de données Crau-Camargue pour les années 2006 et 2009

CE CHAPITRE présente la région d'étude ainsi que l'ensemble des données qui ont été utilisées dans le cadre de notre travail et qui étaient disponibles sur la région d'étude pour les années 2006 et 2009. Ces données se divisent en trois catégories : les données d'entrée du modèle SETHYS (données météorologiques et données relatives aux propriétés des sols et des végétations), les données d'observation à grande échelle utilisées pour contraindre le modèle et les données de validation à HRS utilisées pour évaluer les performances de notre approche de désagrégation par lissage particulière par rapport à des données thermiques à haute et moyenne résolution spatiale.

### 4.1 Région d'étude : Crau-Camargue

La zone d'étude est la région de la Crau-Camargue au Sud-Est de la France. C'est une région au climat typiquement méditerranéen, avec un été aride à la fois chaud et sec, des précipitations faibles concentrées sur l'automne, un hiver doux et un ensoleillement important.

Compte tenu de la situation géographique de la Crau, au débouché de la vallée du Rhône, et de la platitude de ses reliefs, cette région est connue pour la fréquence et la violence du vent Mistral qui tend à accentuer la sécheresse, à augmenter l'ensoleillement, et à rafraîchir les températures hivernales. Cette région est aussi connue pour la diversité de son patrimoine agricole. On distingue, les prairies irriguées (qui fournissent le foin de la Crau) et sèches (steppe de Crau ou Coussoul), les parcelles de céréales d'hiver (blé) et de printemps (maïs) et les rizières de Camargue.

Les différentes pratiques d'irrigation présentes dans la région font de cette région, une terre très contrastée en termes d'humidité du sol en passant des régions très sèches (coussoul - plaine steppique de la Crau) à des régions très humides (prairie de fauche, réserve naturelle de la Camargue, étang de Berre et rizières). La région présente donc une mosaïque de couverts végétaux très hétérogènes en ce qui concerne les types de végétation et de conditions hydriques. Cette hétérogénéité conduit à des contrastes de températures très importants entre les régions sèches et les régions humides, et en font un site idéal pour les études thermiques, comme l'atteste l'abondante bibliographie et les nombreuses expérimentations conduites dans la région dans le domaine de l'infrarouge thermique (Olioso et al. (1996); Seguin et al. (1999); Courault et al. (2008); Bsaibes et al. (2009); Courault et al. (2010); Kallel et al. (2013), etc.).

#### 4. RÉGION D'ÉTUDE ET BASE DE DONNÉES CRAU-CAMARGUE POUR LES ANNÉES 2006 ET 2009



FIGURE 4.1 – Région de Crau-Camargue : zones définies par les cartes d'occupation du sol et sites de mesures des forçages micro-météorologiques, et de végétation.

Au cours de cette thèse nous avons bénéficié de deux bases de données acquises en 2006 et en 2009 issues de deux campagnes de mesures menées par l'UMR EMMAH (<http://www.umr-emmah.fr/>) (Courault et al. (2008); Bsaibes et al. (2009); Courault et al. (2010)) sur deux zones de la région de Crau-Camargue. La première zone, délimitée par le rectangle rouge sur la figure 4.1, correspond à la zone de l'expérimentation de 2006, et les contours sont définis par une carte d'occupation des sols issue d'une image ASTER acquise en 2006. La seconde zone, délimitée par le quadrilatère blanc, présente la zone d'étude de la campagne de 2009, et correspond aux contours de la carte d'occupation de sol fournie par des images Landsat7 pour l'année 2009. Pour ces deux années, les données de forçages micro-météorologiques ainsi que certaines caractéristiques des couverts végétaux (indice foliaire, et hauteur de couvert) nous ont été fournies.

### 4.2 Occupation des sols

L'occupation des sol a été étudiée par l'UMR EMMAH sur les deux zones précédemment décrites. Deux cartes sont ainsi disponibles pour les années 2006 et 2009 :

- pour l'année 2006 l'occupation des sols a été estimée à partir d'images ASTER de résolution 90m dans le domaine du visible. Les coordonnées latitude et longitude des quatre coins de la carte sont les suivants :
  - coin en haut à gauche : (43 ° 42' 22.77'' N, 4 ° 31' 40.40'' E)
  - coin en haut à droite : (43 ° 42' 18.17'' N, 4 ° 35' 34.44'' E)
  - coin en bas à gauche : (43 ° 40' 52.50'' N, 4 ° 31' 39.94'' E)
  - coin en bas à droite : (43 ° 40' 47.90'' N, 4 ° 35' 12.27'' E)

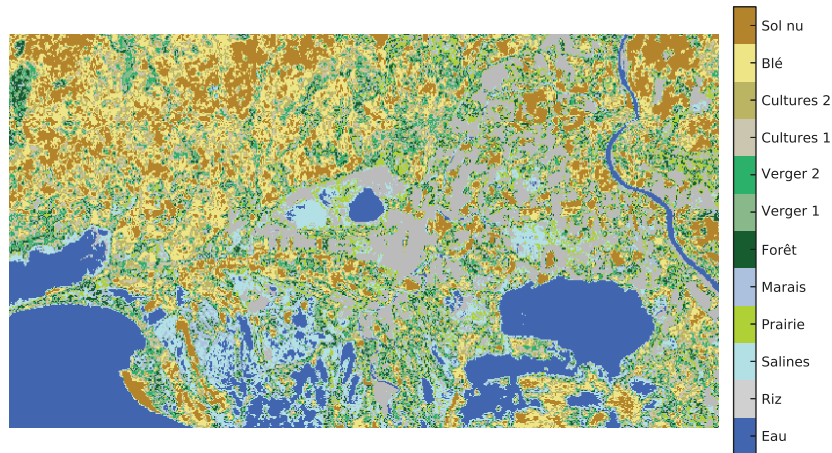


FIGURE 4.2 – Carte d’occupation des sols initiale à 12 classes : produit ASTER à 90m de résolution spatiale pour l’année 2006.

TABLE 4.1 – Températures moyennes et écarts-types des classes à regrouper (en ° C)

Classe	Salines	Marais	Forêt	Verger1	Verger2	Cultures1	Cultures2	Blé
$T_{Mean}$	34.99	38.03	39.43	40.79	42.18	43.51	44.97	46.93
$T_{SDV}$	0.87	0.48	0.42	0.44	0.47	0.45	0.47	0.72

La carte d’occupation des sols décrivait initialement 12 classes de végétation comme le montre la figure 4.2. Comme on ne dispose pas de toutes les données de forçages du modèle pour ces 12 classes, certaines classes dont la variabilité spatiale de la température et les conditions hydriques sont très proches ont été regroupées aboutissant au final à une carte d’occupation des sols à 7 classes (voir Figure 4.3). Pour cela, on a commencé par identifier les classes pouvant être regroupées sous une même classification. La table 4.1 présente la liste des classes à regrouper ainsi que les températures moyennes et les écart-types respectifs calculés à partir de l’image ASTER acquise sur la même région le 26/07/2006 à 10h20. Les températures moyennes du blé ( $\approx 47^\circ\text{C}$ ) et des deux classes de cultures ( $\approx 43.5^\circ\text{C}$ ,  $\approx 45^\circ\text{C}$ ) ainsi que leurs variabilités spatiales étant très proches, nous avons choisi de les regrouper sous une seule classe que l’on a appelé “blé”. Après regroupement, la température moyenne du blé est égale à  $45.5^\circ\text{C}$  avec un écart type de  $1.5^\circ\text{C}$ . Idem pour les classes “Marais” et “Salines” regroupées sous une classe nommée “Marais & Salines” et les classes “Forêt”, “Verger1” et “Verger2” regroupées sous une classe nommée “Vergers”. Comparées aux classes

TABLE 4.2 – Températures moyennes et écarts-types des classes après la nouvelle classification (en ° C).

Classes	Marais & Salines	Vergers	Blé	Prairie	Riz	Sol nu	Eau
$T_{Mean}$	36.23	40.89	45.54	36.66	32.67	50.18	30.48
$T_{SDV}$	1.67	1.22	1.52	0.63	1.35	1.68	1.61

“originales”, les variabilités spatiales des classes reconstituées sont très proches des autres classes ( $T_{SDV} \approx 1.6^\circ\text{C}$ ).



La figure 4.3 présente la nouvelle carte d'occupation à 7 classes comprenant : l'eau, le sol nu, la prairie irriguée, le blé, le riz, les vergers et les marais & salines. La table 4.2 présente les températures moyennes et les écart-types issus de la nouvelle carte d'occupation des sols.

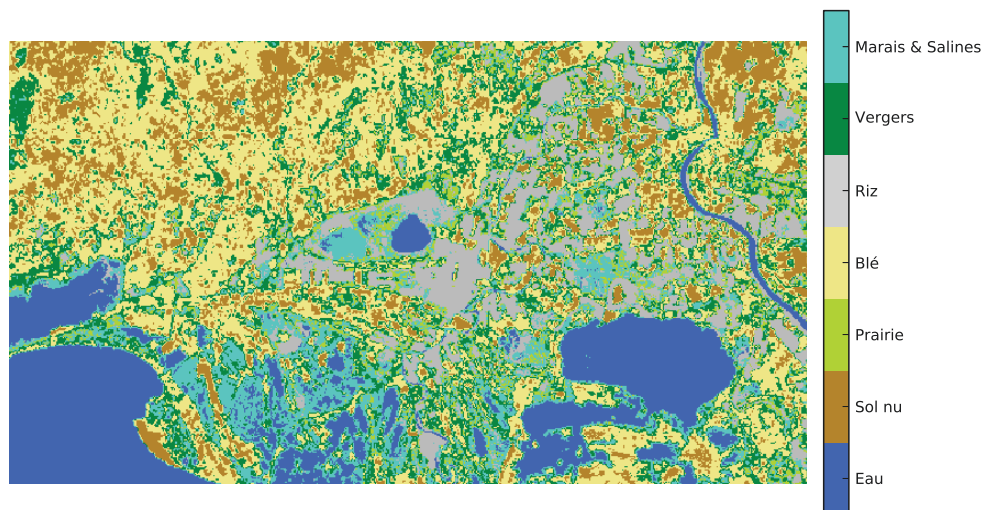


FIGURE 4.3 – Nouvelle carte d'occupation des sols à 7 classes (année 2006).

- Pour l'année 2009, l'occupation des sols a été estimée à partir d'images Landsat7 de résolution 30m. Les coordonnées latitude et longitude des quatre coins de la carte d'occupation de sol sont :
  - coin en haut à gauche : (44 ° 1' 40.96'' N, 4 ° 48' 12.23'' E)
  - coin en haut à droite : (43 ° 56' 26.78'' N, 5 ° 24' 28.92'' E)
  - coin en bas à gauche : (43 ° 10' 13.42'' N, 4 ° 30' 50.73'' E)
  - coin en bas à droite : (43 ° 9' 35.70'' N, 5 ° 8' 0.97'' E)

TABLE 4.3 – Températures moyennes et écarts-types des classes à regrouper (en ° C) pour les différentes dates d'acquisition Landsat7 en 2009.

Date \ Classe	21-03	06-04	22-04	25-06	27-07	28-08	29-09
Forêt	11.86 ±1.27	18.49 ±2.45	20.92 ±1.64	25.15 ±5.63	30.67 ±2.86	31.78 ±2.68	23.22 ±1.90
Vergers	13.25 ±1.31	20.33 ±2.17	22.27 ±1.32	27.93 ±5.92	32.94 ±3.18	34.19 ±2.99	24.99 ±1.72

La carte comprenait initialement 8 classes de végétation comme le décrit la figure 4.4). En analysant la carte d'occupation des sols on a remarqué que la classe vergers est très faiblement représentée sur la carte et que la classe forêt est la classe la plus dominante. La comparaison des températures moyennes et des écart-types respectifs de ces deux classes pour les différentes dates d'acquisition Landsat7 en 2009 montre un très bon accord (voir Table 4.3). On a donc regroupé les classes vergers et forêt sous une même classe nommée "couvert arboré" pour aboutir à la nouvelle carte d'occupation des sols à 7 classes présentée par la Figure 4.5. On reconnaît ainsi les classes eau, sol nu, prairie irriguée, cultures (ex : sorgho, maïs, tournesol etc...), couvert arboré, blé et riz.

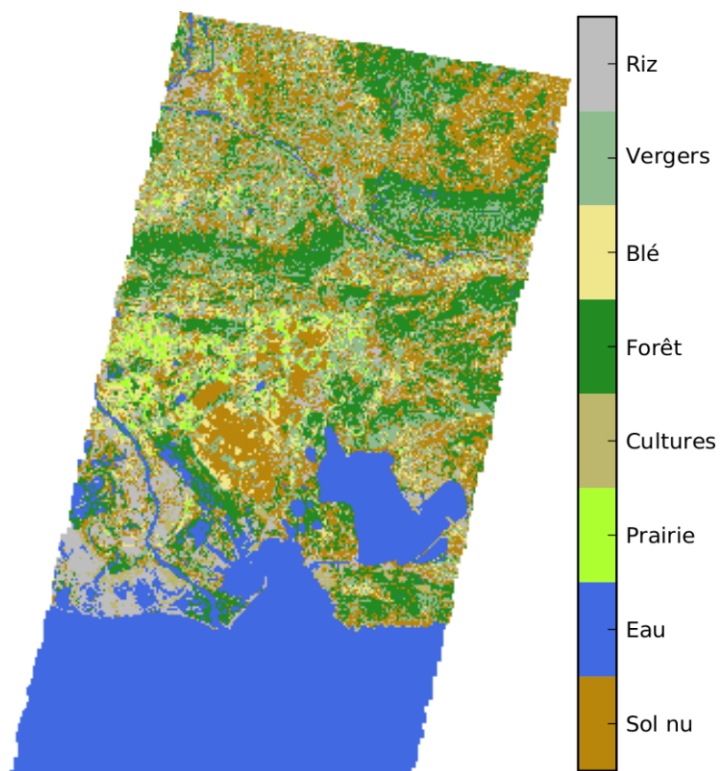


FIGURE 4.4 – Carte d'occupation des sols initiale à 8 classes : produit Landsat7 à 30m de résolution spatiale pour l'année 2009.

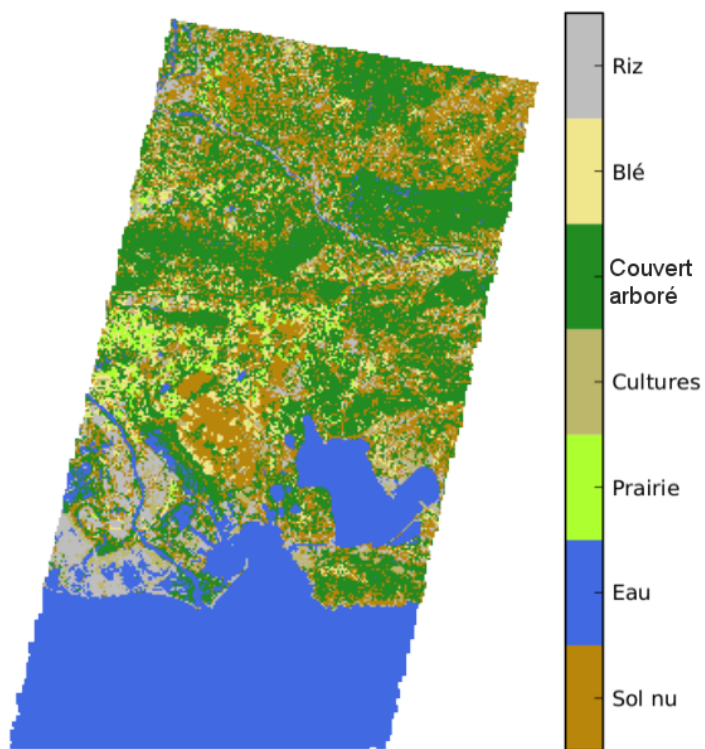


FIGURE 4.5 – Nouvelle carte d'occupation des sols à 7 classes (année 2009).

### 4.3 Données de forçage du modèle SETHYS

Les données de forçages nécessaires au fonctionnement du modèle SETHYS concernent les données météorologiques et les données de LAI et de hauteur de la végétation (voir section 3.1.4).

Deux bases de données de forçages étaient disponibles pour les années 2006 et 2009 sur 5 parcelles de la zone d'étude présentées sur la figure 4.1.

#### 4.3.1 Données atmosphériques

Les données de forçage atmosphériques nécessaires pour lancer le modèle SETHYS sur un type d'occupation de sol donné sont les suivantes : l'année, le jour (DOY Day Of Year), l'heure, les minutes, le rayonnement solaire global, le rayonnement diffus, le rayonnement atmosphérique, la température de l'air, l'humidité relative de l'air, la vitesse de vent et les précipitations/irrigations.

En 2006, toutes les données de forçages, à l'exception du rayonnement atmosphérique et des précipitations/irrigations étaient disponibles sur cinq parcelles : une parcelle de blé, une prairie, une parcelle de maïs, une parcelle de sol nu et une parcelle de riz. Pour les précipitations, ces données étaient disponibles uniquement pour la parcelle de riz. Étant donné que les parcelles sont toutes proches, on a supposé que le forçage de la pluie est le même pour toutes les parcelles et les données de précipitation de la parcelle du riz ont été utilisées pour forcer les autres types d'occupation du sol. Pour l'irrigation, comme le blé n'est pas irrigué en général sur la région de la Crau, ces données ont été fournies uniquement pour la parcelle de prairie. Concernant le rayonnement atmosphérique, ce dernier a été calculé à partir de la formule décrite par l'équation 4.1 en utilisant les données de forçage de la température de l'air et de l'humidité relative de l'air.

Pour certaines parcelles, les données de forçage n'étaient disponibles que sur une partie de l'année. On a donc considéré uniquement la partie de l'année 2006 où les données étaient disponibles pour toutes les parcelles, c'est à dire la période qui s'étend du 28 avril au 13 septembre 2006.

La carte d'occupation des sols estimée pour l'année 2006 (voir Figure 4.3) présente certaines classes dont les données atmosphériques n'ont pas été fournies, notamment les données des classes : "Vergers", "Eau" et "Marais & Salines". Pour combler cette absence de données, on a choisi d'utiliser le forçage atmosphérique de la parcelle de riz pour l'eau et les marais et salines et le forçage du blé pour la classe qui regroupe forêt et verger. Notre choix n'était pas arbitraire mais guidé par la position géographique et les conditions hydriques des parcelles. Ainsi, étant donné que les classes eau et marais et salines sont des classes "humides", la température de l'air du riz (qui est aussi une classe humide) est plus proche de celles de ces dernières. D'autre part, étant donné que la parcelle de blé est située dans le nord du domaine étudié dans la zone où les forêts et les vergers sont majoritaires, les données atmosphériques correspondantes ont été affectées à la classe (forêt/vergers).

En 2009, les données atmosphériques, à part le rayonnement atmosphérique, étaient disponibles sur toute l'année et pour les parcelles de sol nu, de prairie, de blé et de riz. Le rayonnement atmosphérique a donc été calculé en utilisant la formule proposé par Idso (1981) et décrite par l'équation 4.1.

Pour l'année 2006, les données atmosphériques n'étant pas disponibles pour les classes cultures, eau et forêt, on a choisi le forçage du blé pour les classes cultures et forêt et le forçage du riz pour la classe "eau".

$$R_{atm_{diffus}} = 0.179 \left( q_a^{\frac{1}{7}} \right) \exp \left( \frac{350}{T_a} \right) \sigma T_a^4, \quad (4.1)$$

avec,  $\sigma$  est la constante de Stefan-Boltzmann qui est égale à  $5.67 \cdot 10^{-8}$ ,  $T_a$  est la température de l'air en K et  $q_a$  est l'humidité relative de l'air en hPa.

### 4.3.2 Caractéristiques de la végétation

Les données de végétation consistent en des mesures in situ de l'indice foliaire et de la hauteur du couvert faites pour certaines dates des années 2006 et 2009. Les parcelles pour lesquelles ces mesures sont disponibles en 2006 sont celles de la prairie, du riz, du blé et des vergers. En effet pour les vergers, 3 sites de mesures différents ont été instrumentés dont deux sont des vergers de pommiers et un est un verger de nectarines. Ces sites apparaissent sur la carte de la figure 4.1 avec des pointeurs de position de couleur bleue claire. Pour ces trois sites de vergers, une seule mesure a été réalisée au cours de l'année 2006 pour une journée en fin de printemps (lorsque l'arbre fruitier commence à avoir des fruits). Les mesures d'indice foliaire et de hauteur de végétation pour les parcelles de vergers sont faites pour les arbres et pour les herbes qui poussent sous les arbres et en inter-rangs. Les mesures d'indice foliaire et de hauteur de couvert ont été interpolées sur l'année 2006 afin de créer des données de forçages de végétation journalière.

Pour l'année 2009, seules des mesures de la hauteur de couvert de la prairie étaient disponibles. Les données d'indice foliaire (LAI : Leaf Area Index) n'ont pas été fournies en 2009. Pour combler ce manque de données, on a estimé le LAI à partir des données de NDVI issues d'images Landsat7 disponibles sur la région pour 7 dates : 21/03/09, 06/04/09, 22/04/09, 25/06/09, 27/07/09 et 28/08/09. Ces données ont été par la suite interpolées sur l'année 2009 (fréquence d'une valeur par jour). Plus de détails sur la méthode d'estimation du LAI et son interpolation temporelle à partir des données de NDVI sont présentés dans la section 4.4 de ce chapitre. Pour les données de hauteur de couvert, on a utilisé les données mesurées en 2006 pour les classes de blé et de riz après avoir vérifié qu'elles étaient bien en phase avec les données de LAI interpolées à partir des données de NDVI sur l'année 2009.

## 4.4 Données de NDVI

Des images de NDVI (Normalized Difference Vegetation Index) issues de l'imagerie Landsat7, de résolution spatiale 30m, nous ont été fournies sur la région d'étude pour l'année 2009 (voir la figure 4.6). Ces cartes ont servi à estimer des LAI moyens pour chacune des classes de végétation dont les forçages de l'indice foliaire n'étaient pas disponibles en 2009.

Le NDVI est indicateur de biomasse. Il s'agit d'une mesure normalisée de la différence d'énergie réfléchiée par le couvert végétal dans la bande rouge du domaine spectral du visible et l'énergie réfléchiée par celle-ci dans la bande spectrale du proche infrarouge divisée par la somme des deux énergies (voir Eq 1.1 du chapitre 1).

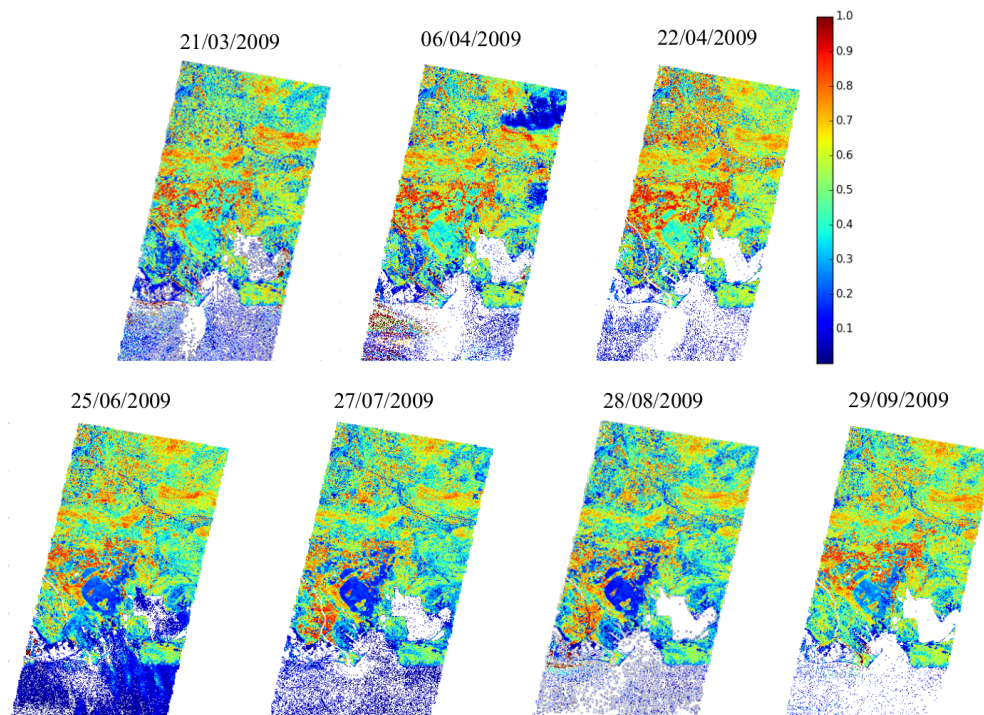


FIGURE 4.6 – Cartes de *NDVI* Landsat 7 disponibles pour l'année 2009

Le *NDVI* est donc un indice sans dimension dont la valeur est comprise entre 0 et 1, où des valeurs inférieures ou égale à 0.1 correspondraient aux étendues d'eau et aux sols nus et des valeurs plus élevées indiqueraient une forte activité de photosynthèse comme des couverts arbustifs, des forêts tempérées et des couverts agricoles en pleine croissance.

Il vient ainsi que l'indice foliaire (le *LAI*), et le *NDVI* sont forcément corrélés vu que plus la végétation est dense (*LAI* élevé), plus les valeurs de *NDVI* sont élevées. [Asrar et al. \(1984\)](#), ont établi qu'il est possible d'estimer le *LAI* à partir du *NDVI* par la relation suivante :

$$LAI = - \left( \frac{1}{K_{LAI}} \right) \times \ln \left( \frac{NDVI - NDVI_{\infty}}{NDVI_s - NDVI_{\infty}} \right) \quad (4.2)$$

où,  $NDVI_{\infty}$  représente la valeur asymptotique de *NDVI* quand le *LAI* tend vers sa valeur maximale,  $NDVI_s$  représente le *NDVI* du sol nu et  $K_{LAI}$  est un coefficient d'extinction. Les valeurs des constantes  $NDVI_{\infty}$ ,  $NDVI_s$  et  $K_{LAI}$  ont été estimées par [Bsaibes et al. \(2009\)](#) à partir des données de *NDVI* de l'année 2006 sur la région de Crau-Camargue. Comme il s'agit de la même région et des mêmes types d'occupation du sol, on a gardé les mêmes valeurs proposées par [Bsaibes et al. \(2009\)](#) :  $NDVI_{\infty} = 0.89$ ,  $NDVI_s = 0.1$  et  $K_{LAI} = 0.71$ .

La figure 4.7 présente les valeurs moyennes (carrés colorés) et les écart-types (barres d'erreur verticales) des *LAI* estimés à partir des cartes de *NDVI* pour chaque classe de végétation et pour toutes les dates d'acquisition des cartes de *NDVI*.

On remarque que les valeurs des *LAI*s moyens estimées pour les classes de cultures et de riz sont très basses ( $LAI_{Cultures} < 1$ ,  $LAI_{Riz} \leq 2$ ). D'autre part, comparées aux valeurs mesurées sur le riz en 2006, les valeurs de *LAI* apparaissent trop faibles.

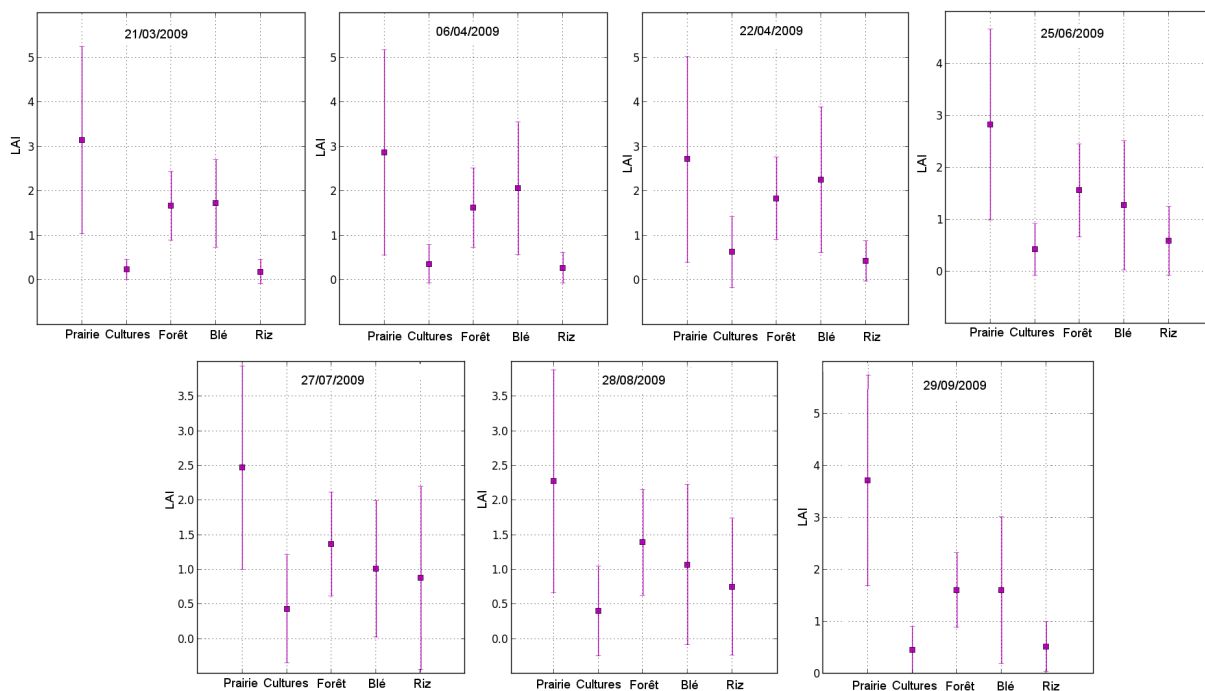


FIGURE 4.7 – LAI moyens et écarts-types pour les classes de végétation de la carte d'occupation des sols Landsat7 de l'année 2009

Pour cette classe, nous avons préféré reprendre les données de forçages de végétation de l'année 2006. En comparant les valeurs de LAI et de la hauteur de couvert mesurées en 2006 et les valeurs de LAI estimées par le NDVI sur le blé, on a remarqué que les résultats sont plutôt cohérents et bien en phase. On a donc repris uniquement le forçage de la hauteur de couvert de 2006 pour la classe blé en gardant les valeurs de LAI estimées à partir du NDVI. Pour la classe cultures, comme on n'avait pas de mesures de LAI en 2006 pour cette classe, on a gardé les valeurs estimées par la relation de [Bsaibes et al. \(2009\)](#). Concernant les classes prairie et couvert arboré, les valeurs de LAI étant plutôt raisonnables, elles ont été interpolées sur l'année 2009 pour créer les données de forçage de LAI.

## 4.5 Données de température de surface

### 4.5.1 Données d'observation

Les données d'observation utilisées, sont des séries temporelles de températures de surface mesurées par le radiomètre imageur SEVIRI (**S**pinning **E**nhanced **V**isible and **I**nfraRed **I**mager) à bord du satellite Météosat 8 appelé aussi MSG (**M**eteosat **S**econd **G**eneration).

MSG est un satellite géostationnaire dont la charge utile comporte le radiomètre SEVIRI. Cet instrument mesure le rayonnement solaire réfléchi et le rayonnement thermique émis par la surface de la Terre dans 12 canaux du visible, de l'infrarouge moyen et de l'infrarouge thermique toute les 15 minutes avec une résolution spatiale de 3km au nadir.

Dans l'infrarouge thermique, SEVIRI enregistre une énergie correspondant au rayonnement émis par l'ensemble du système Terre-atmosphère. Un étalonnage permet de convertir la quantité enregistrée en luminance reliée directement à la température de brillance. La température de brillance (radiative) d'une surface correspond à la température d'un corps noir délivrant la même luminance que la surface en question.

Les produits utilisés pour les données d'observation sont des températures de surface fournies par le LSA SAF (*Land Surface Analysis Satellite Applications Facility*, <http://landsaf.meteo.pt/>). Il s'agit directement d'une température de surface produite à l'aide d'un algorithme de Split Window qui permet à partir des températures radiatives mesurées à au moins 2 fréquences distinctes (généralement 10.8 et 12 $\mu$ m) d'inverser la température de surface en éliminant les effets d'absorption de l'atmosphère (Becker et al. (1990); Faysash and Smith (1999)).

Nous avons ainsi commandé et récupéré des températures de surface sur notre région d'étude pour 3 jours de l'année 2006 autour du 26/07/2006 (date de l'image ASTER disponible) et pour 7 mois de l'année 2009 du /01/03/2009 au 30/09/2009. Dans tout ce qui suit, les données du capteur SEVIRI seront désignées par "températures Météosat".

#### 4.5.2 Données de validation

Pour la validation de notre méthode de désagrégation, on a eu recours à des données de températures de surface à haute résolution spatiale. Pour cela, trois types de données ont été récupérées sur la région d'étude pour les années 2006 et 2009.

La table 4.4 résume l'ensemble des données de validations disponibles en 2006 et 2009 sur la région d'étude.

TABLE 4.4 – Résumé de l'ensemble des données de validation disponibles sur la région d'étude en 2006 et en 2009

Satellite/Capteur	Résolution spatiale	Fréquence de passage	Disponibilité
MODIS/ASTER	90 [m]	1 image/16 jours	1 seule image le 26/07/06 à 10 :47TU
LANDSAT7/ETM+	30 [m]	1 image/16 jours	8 images : 21/03/09, 06/04/09, 22/04/09, 25/06/09, 27/07/09, 28/08/09,29/09/09 et 02/12/09 à 10 :20TU
MODIS/ Aqua & Terra	1000 [m]	4 images/jour (2 Aqua et 2 Terra)	Tous les jours du 01/03/09 au 30/09/09

#### Données Landsat7

Le capteur ETM+ (Enhanced Thematic Mapper Plus) à bord du satellite Landsat7 acquiert des images de la Terre depuis juillet 1999 avec une fréquence de revisite de 16 jours. Landsat ETM+ possède 8 bandes spectrales : les bandes 1 à 7 possèdent une résolution spatiale de 30 mètres et la huitième bande pan-chromatique a une résolution égale à 15 mètres. La bande 6 correspondant au domaine spectral thermique (10.40 à 12.50 $\mu$ m) a une résolution spatiale égale à 60m. A partir du 25/02/2010, les données ont été ré-échantillonnées à la résolution de 30m. Ainsi, 29 images Landsat7 ont été récupérées sur la région d'étude entre 2007 et 2010. L'heure d'acquisition est 10 : 20 TU et l'angle de visée est égale à 15 ° 39'. Les images ont été récupérées sur le site de l'USGS <http://edcsns17.cr.usgs.gov/NewEarthExplorer/>.

En 2009, 8 cartes de températures de surface (à 30m) Landsat7 nous ont été fournies sur la région d'étude comme décrit dans la table 4.4. Plus de détails sur les données Landsat7 récupérées sur la région d'étude sont disponibles sur le site <http://www.cesbio.ups-tlse.fr/multitemp/?p=3825>.

### Données MODIS Aqua/Terra

Les capteurs MODIS (Moderate Resolution Imaging Spectroradiometer) à bord des plateformes Terra et Aqua, acquièrent une scène de la surface de la Terre ( $4 \times /jour$ ). MODIS collecte les données sur 36 bandes spectrales, allant de  $0.4\mu\text{m}$  (violet) à  $14.4\mu\text{m}$  (infrarouge) avec une résolution spatiale allant de 250m (au nadir) à 1km.

Les satellites Aqua et Terra ont des orbites héliosynchrones, ils passent donc au dessus d'un point toujours à la même heure locale. Le satellite Terra a une orbite programmée de telle manière qu'il passe au dessus de l'équateur dans la direction Nord-Sud le matin (10h) et dans la direction Sud-Nord le soir (22h), alors que le satellite Aqua passe au dessus de l'équateur dans la direction Sud-Nord l'après-midi (13h) et dans la direction Nord-Sud la nuit (1h).

Les produits de températures de surface Terra et Aqua sont respectivement *MOD11A1.004* et *MYD11A1.004* : deux images le matin (vers 10h00 UTC et vers 1h00 UTC, respectivement pour MODIS/Aqua et MODIS/Terra) et deux images l'après-midi (vers 22h00 UTC et vers 13h00 UTC, respectivement pour MODIS/Aqua et MODIS/Terra) chaque jour.

L'EOS délivre les données sur des "tuiles" de ( $10^\circ \times 10^\circ$ ) projetées sur une grille sinusoïdale à 1km de résolution au format *HDF – EOS*. Ces produits sont livrés avec les indices de qualité associés. Seules les données présentant le meilleur indice de qualité ont été sélectionnées (pas de nuage, pas de problème de correction atmosphérique, ni de problème d'acquisition).

Les données de température de surface sur la région d'étude correspondant aux dates comprises entre le 01/03/2009 et le 30/09/2009 ont été directement récupérées à partir du site [http://reverb.echo.nasa.gov/reverb/#utf8=%E2%9C%93&spatial\\_map=satellite&spatial\\_type=rectangle&keywords=MOD11A1&selected=C108956779-LPDAAC\\_ECS](http://reverb.echo.nasa.gov/reverb/#utf8=%E2%9C%93&spatial_map=satellite&spatial_type=rectangle&keywords=MOD11A1&selected=C108956779-LPDAAC_ECS).

## 4.6 Inter-étalonnage des données d'observation et de validation

Comme les données de validation et les données d'observation proviennent de différents capteurs, il est fort probable qu'elles présentent des écarts liés à :

- des erreurs de mise en correspondance (*co-registration*) entre les capteurs pouvant générer des décalages entre les pixels des deux images.
- des problèmes d'inter-étalonnage entre les capteurs liés à l'utilisation de différents algorithmes pour convertir la radiance mesurée au sommet de l'atmosphère en une température de surface (différents critères d'élimination des nuages, différentes valeurs d'émissivité, etc).

Il est donc important de procéder à un inter-étalonnage des données *a posteriori* pour identifier l'origine de ces écarts et éventuellement corriger ces problèmes avant de passer au traitement des données. Dans cette partie, on présente les différentes démarches adoptées pour réaliser l'inter-étalonnage des données d'observation et de validation précédemment décrites.



#### 4.6.1 Inter-étalonnage des données pour l'année 2006

L'application de notre méthode de descente en échelle sur une base de données réelles a été faite, dans un premier temps, sur des données utilisées pour la validation de l'approche de désagrégation de la température proposée par [Kallel et al. \(2013\)](#). Pour ce jeu de données nous avons procédé aux mêmes traitements à titre vérificatif. La méthode d'inter-étalonnage utilisée est l'algorithme RANSAC (RANdom SAMple Consensus) introduit en premier par [Fischler and Bolles \(1981\)](#). Il s'agit d'une méthode permettant d'étalonner des données expérimentales/d'observation par rapport à un modèle mathématique donné. L'algorithme RANSAC permet en effet de lisser un ensemble de données/d'observations en éliminant les données affectées par des erreurs grossières tel que les erreurs de mesures. Il présente ainsi un outil très important pour le traitement et l'analyse d'images.

Le principe de l'algorithme est simple et consiste à estimer de manière itérative les paramètres d'un modèle mathématique (une droite de régression linéaire d'un nuage de points par exemple) à partir d'un ensemble de données observées contenant un ensemble de valeurs aberrantes souvent nommées "outliers". Les données conservées à la fin de l'algorithme sont dites "inliers". C'est un algorithme non-déterministe dans le sens où il produit un résultat correct avec une certaine probabilité, celle-ci augmentant à mesure que le nombre d'itérations est grand.

La recherche des solutions par l'algorithme RANSAC nécessite la prescription de certains paramètres qui sont l'incertitude que l'on souhaite avoir sur les solutions, et le nombre maximum de boucles de recherches que l'on désire avoir. Le premier paramètre permet de préciser le seuil à partir duquel une observation est considérée comme un "outlier". Le deuxième paramètre est un critère d'arrêt. En effet, si au bout de  $M$  itérations on n'a toujours pas convergé, la solution issue du RANSAC correspond à la solution de la  $M^{\text{ième}}$  itération. Plus de détails sur l'algorithme RANSAC sont disponibles dans [Fischler and Bolles \(1981\)](#). En 2006, les données disponibles sur la région sont les suivantes :

1. une image ASTER pour la validation à haute résolution spatiale acquise à 10h47 le 26/07/2006,
2. une série temporelle d'observations Météosat à basse résolution spatiale pour toute la journée du 26/07/2006.

Les données ASTER ont été utilisées comme données de références par rapport aux quelles les données Météosat seront corrigées (suivant [Kallel et al. \(2013\)](#)). Dans un premier temps, l'image ASTER a été agrégée à l'échelle de Météosat. L'inter-étalonnage est réalisé entre l'image Météosat à 10h45min et l'image ASTER agrégée à 10h47min. Le seuil d'incertitude est choisi égal à  $1.5K$  et le seuil d'arrêt est de 1000 itérations RANSAC. La droite de régression linéaire optimale estimée à partir de 1000 itérations ( $a = 1$ ,  $b = -3$ , et  $r = 0.96$ ) a été appliquée à l'image Météosat à 10h45 après élimination des "outliers" (voir Figure 4.8). Cette correction a permis de réduire le biais initialement égal à  $3K$  à seulement  $1K$ . Étant donné que l'on dispose d'une seule image ASTER, la correction estimée à 10h45 a été appliquée à toute la série temporelle Météosat.

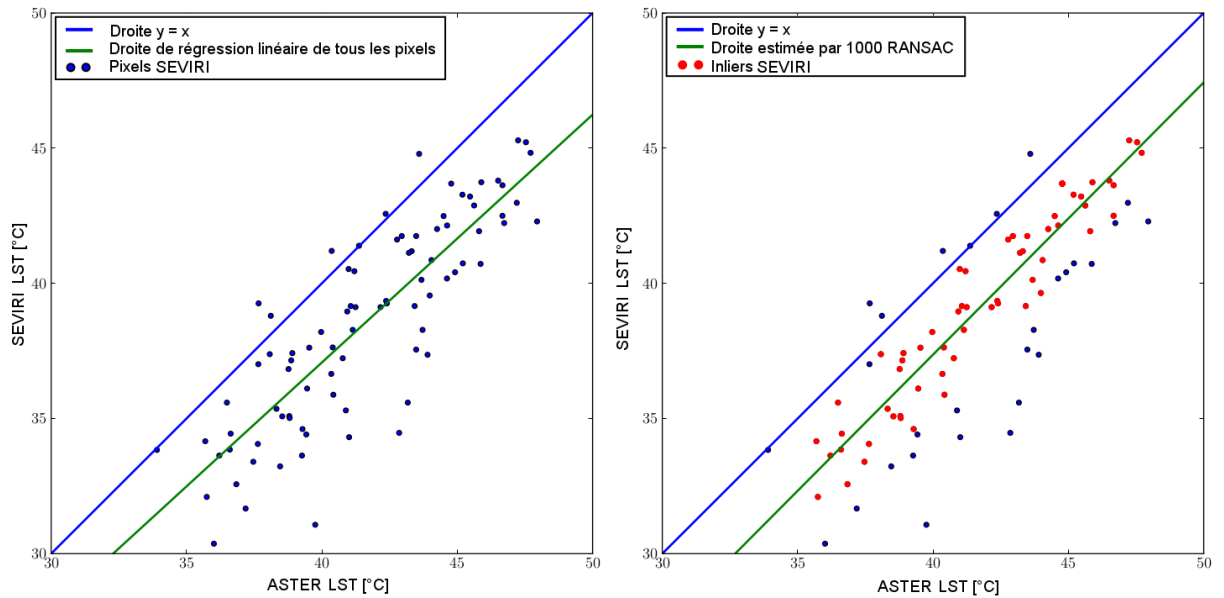


FIGURE 4.8 – Résultat de la correction des températures Météosat acquises sur la journée 26/07/2006 à 10h45 par la méthode RANSAC

#### 4.6.2 Inter-étalonnage des données pour l'année 2009

En 2009, plusieurs données étaient disponibles sur la région d'étude dont 7 mois d'observations Météosat (du premier Mars au 30 Septembre), 7 images de validation HRS Landsat7 étalées sur la période d'acquisition des données d'observation et 7 mois de données MODIS sur la même période. Dans un premier temps on a commencé par vérifier si les données BRS et les données HRS (Landsat7 et MODIS) nécessitaient d'être inter-étalonnées.

##### 4.6.2.1 Températures Landsat7 versus températures Météosat

Comme on vient de le préciser, les données de validation disponibles sur l'année 2009 sont 7 images Landsat7 de températures de surface sur la région d'étude. Étant donné que l'on n'a pas de moyens pour corriger les données Météosat puisque les données à HRS ne sont pas assez fréquentes pour pouvoir les interpoler dans le temps, on choisit d'étalonner les données Landsat7 par rapport aux données d'observations Météosat. Pour cela nous avons calculé, pour les différentes dates d'acquisition des données Landsat7, les fonctions de distribution cumulées (CDF : Cumulative Distribution Function) pour les températures Météosat, les températures Landsat7 et les températures Landsat7 agrégés à la résolution spatiale de Météosat.

Les graphes de la figure 6.3 présentent les CDF des températures Landsat7 et Météosat pour les 7 dates d'acquisition des données Landsat7 sur la période d'observation Météosat. En examinant les distributions de la température de surface, on remarque que les températures Météosat et Landsat7 agrégées à l'échelle de Météosat présentent des biais importants pour toutes les dates de validations et ces biais sont soit positifs, soit négatifs selon la date d'acquisition. La comparaison de la CDF des températures Landsat7 à la résolution spatiale initiale (30m) aux CDF des températures Landsat7 et Météosat à la résolution spatiale de Météosat montre que ces dernières ont des distributions beaucoup plus lisses comparées à la distribution des températures à HRS. En d'autres termes, les variances des températures sont beaucoup plus élevées à HRS.

#### 4. RÉGION D'ÉTUDE ET BASE DE DONNÉES CRAU-CAMARGUE POUR LES ANNÉES 2006 ET 2009

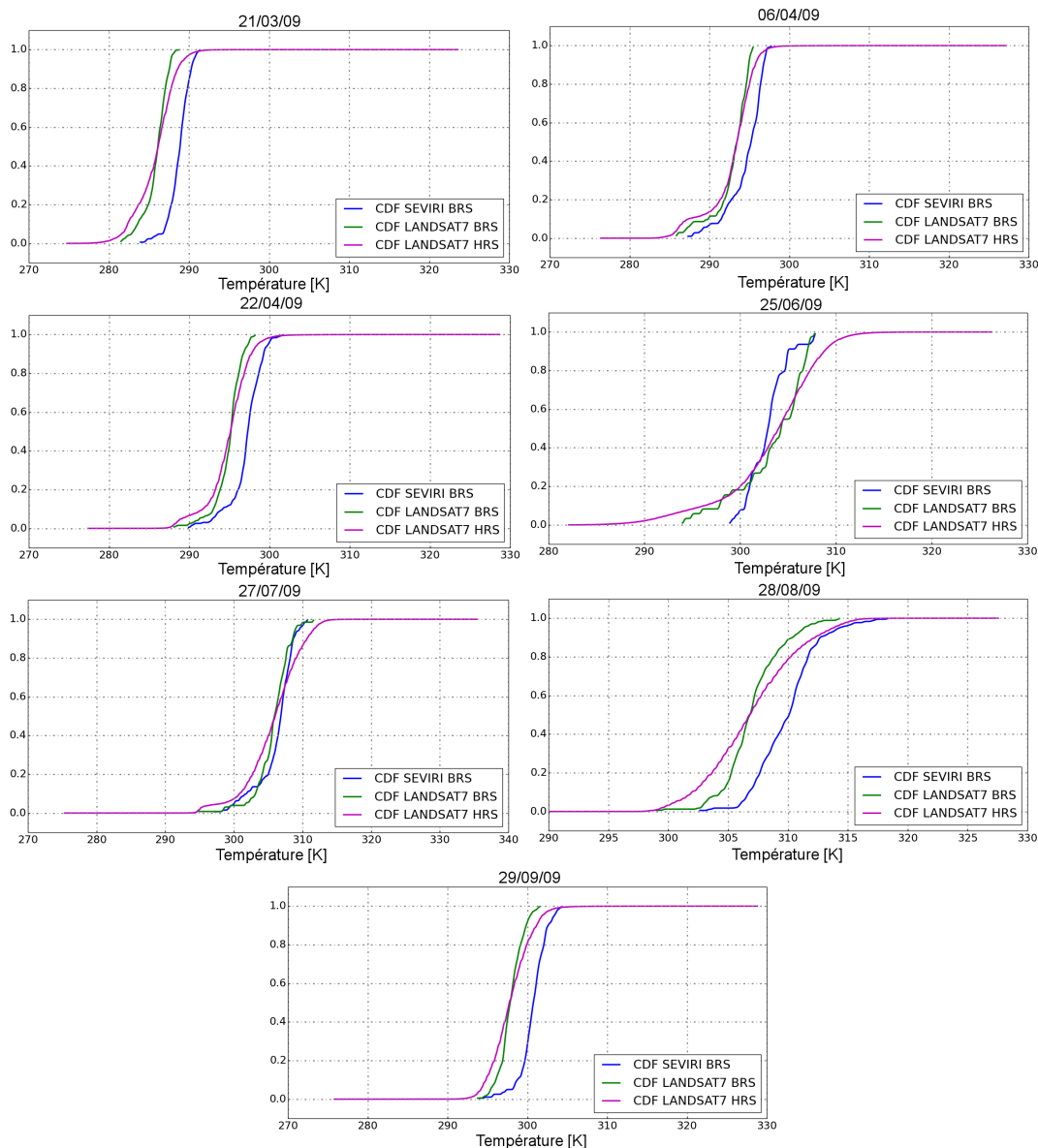


FIGURE 4.9 – Fonctions de distribution cumulée des températures Météosat et des températures Landsat7 (à HRS et agrégées à la résolution spatiale de Météosat) : les courbes en rose présentent les CDF des températures HRS Landsat7(CDF LANDSAT7 HSR), les courbes en vert présentent les CDF des températures Landsat7 agrégées (CDF LANDSAT7 BRS)et les courbes en bleu présentent les CDF des températures Météosat (CDF SEVIRI BRS).

Des méthodes d'ajustement des CDF peuvent permettre de corriger les biais et les variances à l'échelle de Météosat. Cependant, cette correction dépendant de l'échelle, elle ne peut pas être appliquée pour corriger les images Landsat7 à HRS. Il est donc impossible à ce niveau de faire l'inter-étalonnage des données Landsat7 et Météosat. La validation par rapport à ces données (données Landsat7) sera limitée à une validation "qualitative" des résultats de désagrégation obtenus par lissage particulière.

#### 4.6.2.2 Températures MODIS versus températures Météosat

Afin de vérifier si les données MODIS et Météosat sont bien cohérentes on a commencé par agréger les données MODIS à la résolution spatiale de Météosat et on a tracé pour les dates de disponibilité des données Landsat7 les CDF des températures MODIS, les CDF des températures MODIS agrégées à la résolution spatiale de Météosat et les CDF des températures Météosat (voir Figure. 6.4). La comparaison de la CDF des températures MODIS à la résolution spatiale initiale (1km) aux CDF des températures MODIS et Météosat à la résolution spatiale de Météosat montre que ces dernières ont des distributions assez proches de la distribution des températures à MODIS à la résolution initiale. En effet comme l'écart entre les résolutions des données de validation et d'observation n'est pas très important ( $1\text{km} \times 1\text{km}$  vs  $3 \times 5\text{km}$ ), la variance de la température était moins importante que celle rencontré avec les données Landsat7. Dans ce cas, il est possible de faire la correction des données MODIS HRS en utilisant les données Météosat. Pour explorer plus finement l'origine des biais entre les données MODIS agrégées et les données Météosat, on a tracé les séries temporelles pour les pixels Météosat les plus homogènes quant à l'occupation des sols et d'un pixel Météosat mixte (contenant les 7 classes décrites par la cartes d'occupation des sols de l'année 2009) afin de vérifier si l'origine des biais est liée à une classe de végétation en particulier.

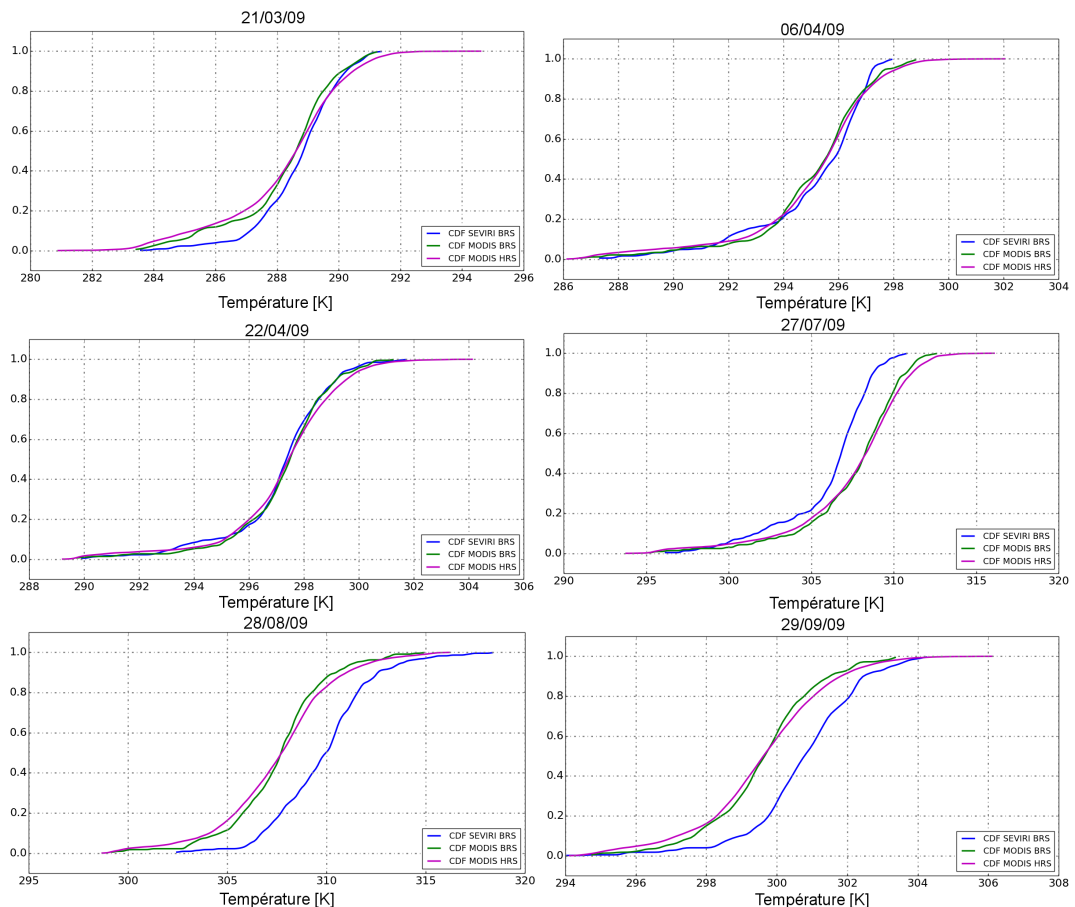


FIGURE 4.10 – CDFs des températures MODIS (à HRS et agrégées à la résolution spatiale de Météosat) versus CDFs Météosat : les courbes en rose présentent les CDF des températures MODIS à 1K (CDF MODIS HRS), les courbes en vert présentent les CDF des températures MODIS agrégées (CDF MODIS BRS) et les courbes en bleu présentent les CDF des températures Météosat (CDF SEVIRI BRS).

Les graphes des figures 4.11 et 4.12 présentent respectivement les séries temporelles pour les pixels les plus homogènes de sol nu et de couvert arboré et les séries temporelles du pixel le plus homogène d'eau et d'un pixel Météosat mixte. D'après ces figures, on remarque qu'il existe un biais positif entre les températures Météosat et MODIS pour les pixels les plus homogènes d'eau, de sol nu et de couvert arboré pendant la journée. L'amplitude de ce biais varie selon le type de couvert (plus importante pour l'eau). Pour le pixel mixte, les températures MODIS sont plutôt en accord avec les températures Météosat ce qui laisse à penser que les biais sont plutôt dépendants de la classe de végétation et se manifestent donc pour les pixels les plus homogènes et pendant la journée. Afin de mieux explorer ces derniers propos, on a choisi de filtrer les images MODIS en ne gardant que les images acquises durant la journée (entre 10h et 14h) et les moins affectées par des problèmes de nuages. Ce filtrage a permis de conserver 29 images MODIS pour la validation avec un biais moyen égal à 0.76K. Pour ces 29 images, on a estimé les biais à partir des pixels majoritaires définis comme étant des pixels contenant plus que 55% d'une classe de végétation donnée. Cette définition a permis d'identifier 3 classes majoritaires : le sol nu avec 9 pixels majoritaires, le couvert arboré avec 68 pixels majoritaires et l'eau avec 9 pixels majoritaires. On a aussi retenu un pixel à 50% de riz comme pixel majoritaire pour calculer le biais entre MODIS et Météosat pour la classe de riz.

Les graphes des figures 4.13 et 4.14 présentent respectivement les histogrammes des biais par classe et leurs valeurs en fonction de la date et de l'heure d'acquisition des données MODIS. Les biais moyens estimés sur l'ensemble des pixels homogènes par classe et pour les 29 dates varient entre 1K (biais estimé pour la classe "Sol nu") et 3.5K (biais estimé pour la classe "Eau"). Les histogrammes de la figure 4.13 confirment bien que les températures Météosat sont généralement plus chaudes que les températures MODIS durant la journée avec une distribution du biais qui varie d'un type d'occupation de sol à l'autre. En examinant de plus près la variation du biais en fonction de la date et de l'heure d'acquisition (Figure 4.14) on constate aussi que ces biais sont plus importants en milieu de journée (vers 13h et 13h30).

Pour les classes de sol nu, de riz et de couvert arboré, un tel comportement du biais pourrait s'expliquer par les effets directionnels de la température. Les deux instruments MODIS et Météosat étant situés sur des orbites différentes, ils ne visent pas la région d'étude sous le même angle. Météosat étant sur une orbite géostationnaire, il observe la région d'étude depuis la direction Sud. MODIS situé sur une orbite polaire vise la région d'étude depuis l'Ouest en orbite descendante et depuis l'Est en orbite montante. Ces différentes configurations font que les fractions des zones ombrées/éclairées du sol et des végétations, qui présentent des températures différentes, varient dans le temps d'un capteur à un autre. Les températures des pixels étant des températures composites des différentes fractions de sol et de végétations, elles sont donc différentes selon la configuration de chacun des capteurs Météosat et MODIS. Selon [Ermida et al. \(2014\)](#) et [Guillevic et al. \(2013\)](#), cet effet d'ombre est d'autant plus important que la fraction d'arbre est importante.

Pour la classe "Eau", ce biais non négligeable (3.5K) pourrait s'expliquer par les différentes valeurs d'émissivités utilisées et les différentes corrections atmosphériques effectuées ([Ottlé and Stoll \(1993\)](#)). En effet, contrairement à MODIS qui mesure à la fois la température de surface de la mer et des surfaces continentales, Météosat utilise des configurations différentes pour chaque produit. Étant donné qu'il s'agit de mesurer la température de surface continentale, Météosat considère que les pixels d'eau sont des pixels terrestres. Il est donc possible que la valeur de l'émissivité prescrite par

Météosat soit inférieure à celle de MODIS.

Cette sous-estimation de la valeur d'émissivité pour les pixels d'eau expliquerait ainsi les valeurs plus élevées des températures estimées par Météosat comparées à celles estimées par MODIS pour une même valeur de la radiance ce qui expliquerait le biais positif observé entre les deux capteurs.

Ainsi, pour inter-étalonner les capteurs MODIS et Météosat, comme on n'a pas les moyens d'interpoler la correction estimée sur les 29 jours sur les séries temporelles Météosat, on a choisi de corriger les température MODIS (à la résolution spatiale de MODIS) et de garder les températures Météosat inchangées. La correction du biais est faite en ajoutant un biais composite estimé pour chaque date de validation (29 dates). Ce biais est dit composite car il est calculé en pondérant les biais estimés pour les quatre classes majoritaires précédemment présentées, pondérés par leur fractions respectives à l'échelle du pixel MODIS comme le décrit l'équation 4.3

$$Bias_{Mi,j}^d = \sum_{k=1}^4 \alpha_{i,j}^k Bias_{Sk}^d, \quad (4.3)$$

avec,  $d$  présente la date de validation,  $i$  et  $j$  présentent respectivement les coordonnées latitude et longitude du pixel MODIS,  $k$  désigne le type d'occupation de sol et les indices  $M$  et  $S$  indiquent la résolution spatiale à laquelle les biais sont calculés ( $M$  pour MODIS et  $S$  pour Météosat (SEVIRI)). Les données MODIS ainsi dé-biaisées nous serviront par la suite pour valider les résultats de désagrégations des températures sous-pixelliques à 1K de résolution (voir Chapitre 6).

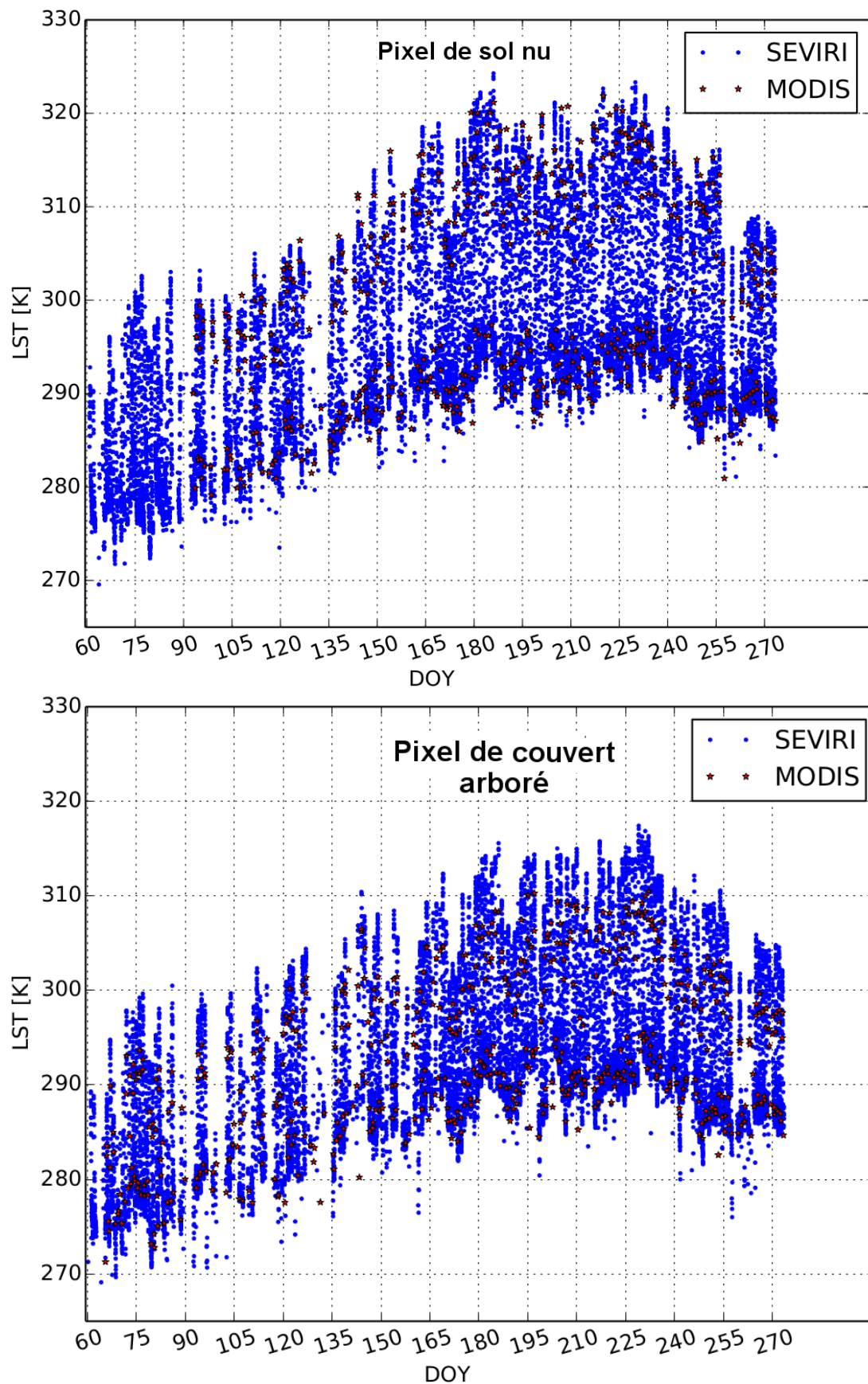


FIGURE 4.11 – Séries temporelles des températures MODIS agrégées à l'échelle de Météosat et des températures Météosat pour le pixel le plus homogène de sol nu et le pixel le plus homogène de couvert arboré sur les 7 mois de données de l'année 2009.

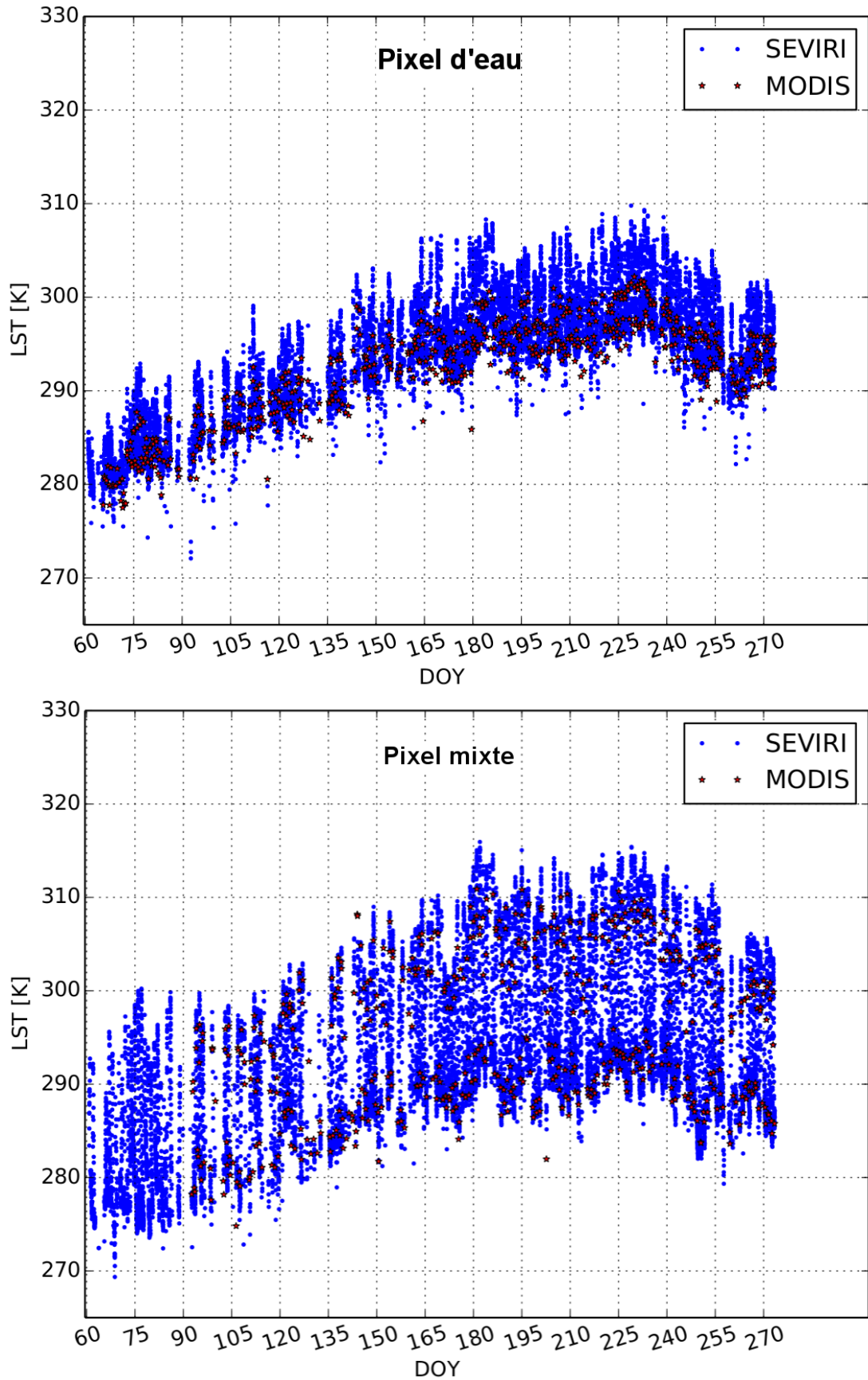


FIGURE 4.12 – Séries temporelles des températures MODIS agrégées à l'échelle de Météosat et des températures Météosat pour le pixel le plus homogène d'eau et un pixel Météosat mixte sur les 7 mois de données de l'année 2009.



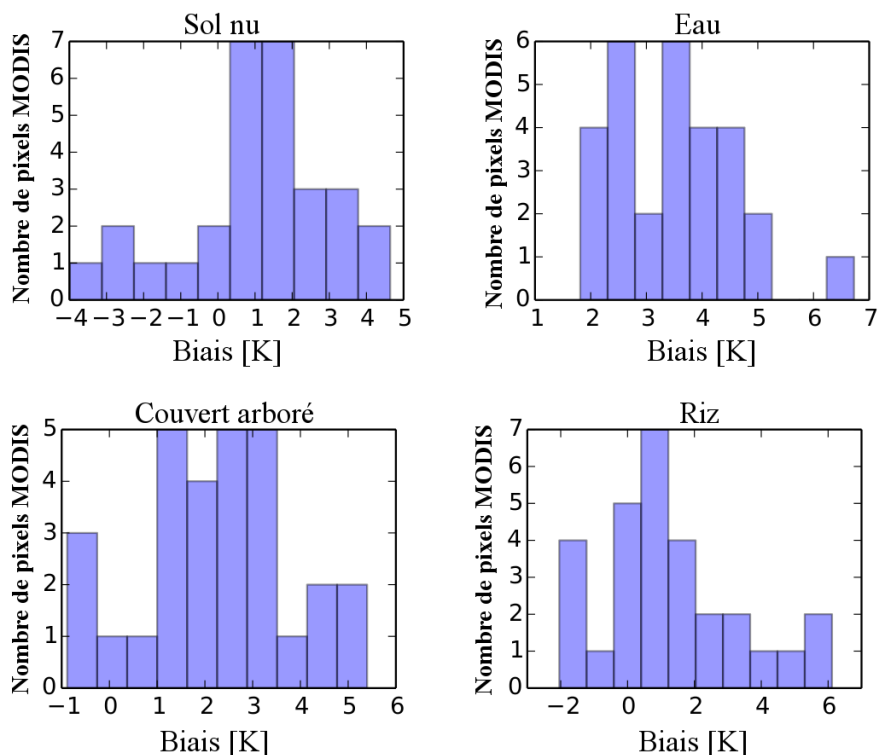


FIGURE 4.13 – Histogrammes des biais moyens pour les quatre classes majoritaires à la RS de Météosat.

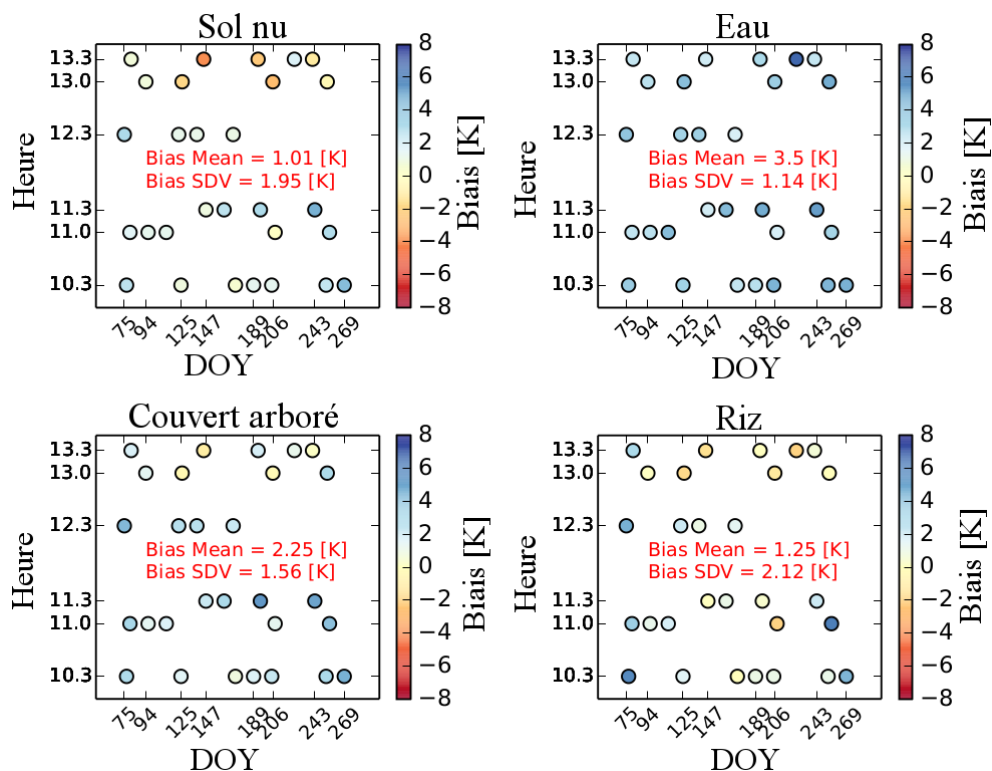


FIGURE 4.14 – Variation des biais moyen des quatre classes majoritaires à la RS de Météosat en fonction de la date et de l'heure d'acquisition.

# Application de la méthode de désagrégation par lissage particulière sur des données synthétiques et réelles

## 5.1 Introduction

Ce chapitre présente la méthode de désagrégation proposée dans ce travail de thèse et se divise en deux parties. La première partie est consacrée à la présentation du principe de la méthode de désagrégation par lissage particulière dans un cadre simplifié. La deuxième partie présente les résultats de l'application de cette méthode sur des pseudo-observations dans le cadre d'une expérience synthétique et sur des données réelles de télédétection pour désagréger des séries temporelles Météosat ( $3 \times 5\text{km}$ ) à la résolution spatiale de données ASTER (90m) sur une zone de la région de Crau-Camargue au Sud-Est de la France.

## 5.2 Méthode

Pour comprendre le principe de la méthode de désagrégation proposée, mettons nous dans le cas le plus simple où l'on cherche à estimer les températures à l'échelle de la classe de végétation pour un seul pixel BRS d'une image IRT.

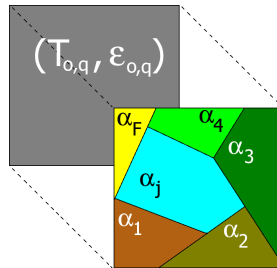


FIGURE 5.1 – Problème de désagrégation dans le cas d'un pixel BRS à  $F$  classes

Supposons que pour ce pixel BRS, l'on ait une observation de sa température  $T_{o,q}$  (l'indice "o" désigne les données d'observation) à chaque instant "q" avec une erreur d'observation  $\zeta_{o,q}$  qui suit la loi  $\mathcal{N}(0, \sigma_{o,q})$  les observations sont supposées non biaisées). Soit  $F$  le nombre de types de végétation présents dans ce pixel et  $(\alpha_i)_{i=1,F}$  leurs fractions respectives fournies par une carte d'occupation des sols supposée parfaitement connue (pas d'incertitude sur les  $(\alpha_i)_{i=1,F}$ ) et constante pour toute la période d'assimilation de données. La figure 5.1 présente une illustration de ce qu'on vient de décrire.

Résoudre le problème de désagrégation revient donc à estimer les températures respectives de chaque type de couvert présent dans le pixel IRT mixte. Pour cela, il nous faut disposer d'un modèle qui permet de simuler la variabilité spatiale et temporelle des différentes températures sous-pixelliques (températures à l'échelle de la classe de végétation). Dans notre cas, ceci est assuré par le modèle SETHYS qui permet, à partir des forçages atmosphériques et des caractéristiques de la végétation, supposées disponibles pour les différentes classes considérées, de simuler des températures à la résolution temporelle des observations et à l'échelle de la classe de végétation.

Pour simuler toutes les valeurs possibles de température des classes *a priori*, on a commencé par identifier les paramètres de SETHYS les plus sensibles à la température. Ainsi, simuler différentes valeurs de température d'une classe donnée revient à échantillonner l'espace des paramètres sensibles définis à l'aval de l'analyse de sensibilité (voir Section 3.2).

Soit alors  $(P_1^i, \dots, P_{D^i}^i)_{i \in [1, F]}$  la liste des paramètres sélectionnés, tel que  $D^i$  désigne le nombre de paramètres sensibles pour une classe  $i$ . Pour plus de simplicité, supposons que le nombre de paramètres sensibles est le même pour les  $F$  classes tel que :  $(D^i)_{i \in [1, F]} = D$ .

Pour coupler le lisseur particulaire et le modèle SETHYS il faut définir ce que représente une particule. Dans notre cas, une particule désigne l'ensemble des valeurs prises par les paramètres sensibles définis pour chacune des classes considérées. Ainsi, à l'instant  $q$  (ou pour la " $q$ -*ème*" fenêtre d'assimilation), la particule s'écrit alors comme suit :

$$x_q = \left( P_{1,q}^1, \dots, P_{D,q}^1, \dots, P_{1,q}^i, \dots, P_{D,q}^i, \dots, P_{1,q}^F, \dots, P_{D,q}^F \right) \quad (5.1)$$

Générer un ensemble de  $N$  particules, n'est autre que générer  $N$  combinaisons des paramètres sensibles de SETHYS pour les différentes classes considérées.

Ainsi, pour chaque particule, SETHYS simule un F-uplet de températures sous-pixelliques tel que :

$$\left( x_q^k \right) \longrightarrow \text{SETHYS} \longrightarrow \left[ T_{1,q}^k, \dots, T_{i,q}^k, \dots, T_{F,q}^k \right], \forall i \in [1, F], \forall k \in [1, N] \quad (5.2)$$

avec  $T_{i,q}^k$  désigne la température de la classe  $i$  à l'instant  $q$ .

L'agrégation des  $N$  F-uplets de températures donne un ensemble de  $N$  températures  $(T_{k,q})_{k \in [1, N]}$  comparables aux observations BRS  $T_{o,q}$  tel que :

$$T_{k,q} = \left[ \frac{\sum_{i=1}^4 \alpha_i \epsilon_i (T_{i,q}^k)^4}{\sum_{i=1}^4 \alpha_i \epsilon_i} \right]^{\frac{1}{4}}, \quad (5.3)$$

avec  $\epsilon_i, \alpha_i$  désignent respectivement l'émissivité et la fraction de couvert de la classe  $i$ ,  $\sum_{i=1}^F \alpha_i = 1$

Le rôle du lisseur particulaire est de sélectionner les particules qui donnent des températures agrégées  $T_{k,q}$  les plus "*proches*" des observations  $T_{o,q}$ .

La mise en équations du problème de désagrégation sous la forme classique d'un problème d'assimilation de données est donc la suivante :

$$y_q = \mathcal{H}(x_q) + \zeta_{o,q}, \quad (5.4)$$

$$x_{q+1} = \mathcal{M}(x_q) + \zeta_q, \quad (5.5)$$

avec  $y_q$  désigne la température du pixel IRT ( $T_{o,q}$ ),  $\mathcal{H}$  représente l'opérateur d'observation qui permet de passer de l'espace des paramètres aux températures  $T_{o,q}$  et représente donc la combinaison du modèle SETHYS et de l'opérateur d'agrégation (voir Figure 5.2) et  $\mathcal{M}$  désigne le modèle qui fait propager les particules avec  $\zeta_q$  est l'erreur modèle.

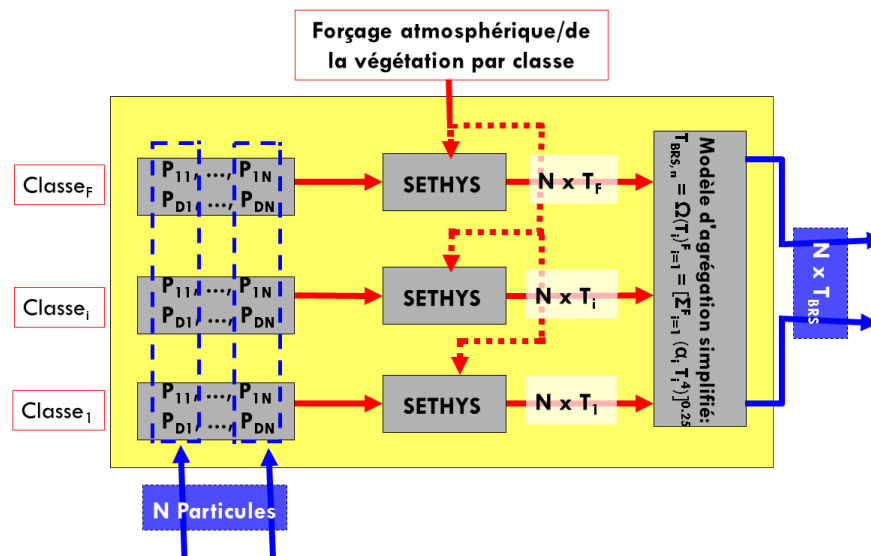


FIGURE 5.2 – Schéma descriptif de l'opérateur d'observation  $\mathcal{H}$ .

Etant donné que les particules ne sont autre que les paramètres de SETHYS, supposés constants durant une fenêtre d'assimilation,  $\mathcal{M}$  présente dans ce cas la fonction identité et l'erreur modèle est donc le bruit rajouté au particules après ré-échantillonnage permettant de mieux explorer l'espace des paramètres.

Une fois les opérateurs et les variables du problème définis, la procédure de désagrégation pourrait être décrite par l'algorithme suivant :

### 1. Initialisation

Cette étape est une étape d'initialisation et elle a lieu lors de la première fenêtre d'assimilation.

- Génération des  $N$  particules ( $N$  tirages aléatoires des valeurs des paramètres sensibles définis pour chaque classe).

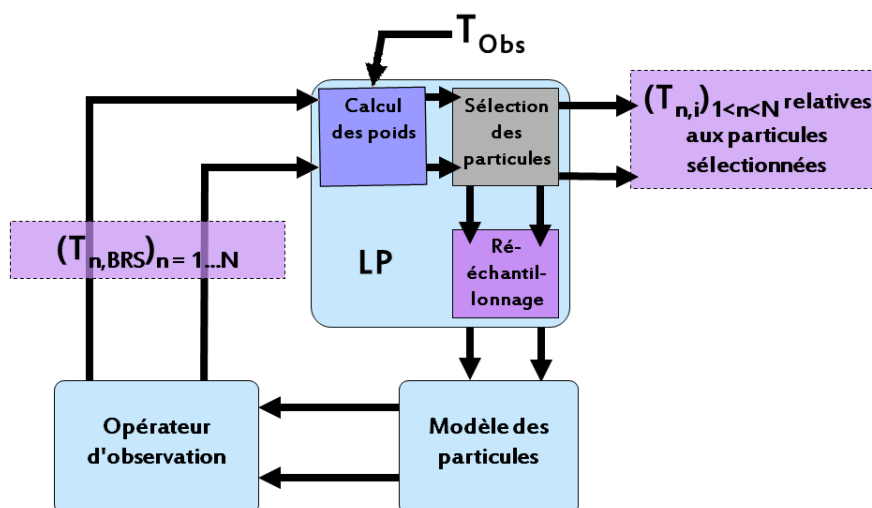


FIGURE 5.3 – Schéma descriptif de la procédure de désagrégation par lissage particulaire.

## 2. Analyse

- Simulation des  $N$   $F$ -uplets de températures sous-pixelles par le modèle SE-THYS.
- Agrégation des  $N$   $F$ -uplets de températures et calcul des  $N$  températures comparables aux observations.
- Calcul des distances relatives à chaque particule par rapport à la température observée pour le pixel BRS : il s’agit de calculer les distances entre les températures agrégées issues de chaque particule et la température IRT du pixel mixte.
- Sélection des particules dont les températures agrégées sont les plus “proches” des observations (étape de sélection).
- Conservation des  $F$ -uplets de températures correspondants aux particules sélectionnées : la moyenne de chaque ensemble de températures de classe est considérée comme la solution de désagrégation.

## 3. Prévision

- Ré-échantillonnage des particules sélectionnées et ajout de bruit aux copies des particules pour mieux explorer l’espace des paramètres du modèle autour des valeurs sélectionnées (Equation 5.5).

## 4. Itérer sur les étapes 2 et 3 pour les fenêtres d’assimilation suivantes.

La figure 5.3 résume le processus de désagrégation par lissage particulaire qu’on vient de décrire ci-dessus.

Supposons maintenant que l'on désire appliquer notre méthode non pas sur un pixel unique mais sur une carte de  $P$  pixels BRS. Pour résoudre ce problème on fait l'hypothèse que les corrélations spatiales entre les différents pixels BRS sont négligeables. Compte tenu de cette hypothèse, désagréger l'image composée de  $P$  pixels BRS revient à reproduire les étapes décrites précédemment de manière indépendante à chaque pixel de l'image IRT.

Le couplage du filtre ou du lisseur particulaire avec le modèle SETHYS pour la désagrégation des températures à l'échelle de la classe de végétation, et son application sont détaillées dans l'article qui suit [Mechri et al. \(2014\)](#) publié dans la revue *Journal of Geophysical Research*.

La première partie de cet article présente l'application de la méthode dans un cadre synthétique et vise à évaluer l'apport de l'assimilation des observations BRS pour réduire l'incertitude sur l'estimation des températures des classes comparée au modèle SETHYS sans assimilation de données. Cette étude a considéré différents scénarios en faisant varier la fréquence des observations, les fenêtres d'observation et les erreurs d'observation.

La deuxième partie présente l'application de notre approche pour la désagrégation des séries temporelles Météosat de résolution spatiale kilométrique à des séries temporelles de résolution spatiale décimétrique sur une image contenant 108 pixels BRS.

### **5.3 Article 1 : Genetic particle filter application to land surface temperature downscaling**

## RESEARCH ARTICLE

10.1002/2013JD020354

## Key Points:

- The use of genetic particle filter to downscale land surface temperature
- Good performances of particle filter downscaling method for a synthetic case
- Application to actual data showed promising performances

## Correspondence to:

R. Mechri,  
rihab.mechri@lsce.ipsl.fr

## Citation:

Mechri, R., C. Ottlé, O. Pannekoucke, and A. Kallel (2014), Genetic particle filter application to land surface temperature downscaling, *J. Geophys. Res. Atmos.*, *119*, 2131–2146, doi:10.1002/2013JD020354.

Received 11 JUN 2013

Accepted 5 FEB 2014

Accepted article online 9 FEB 2014

Published online 6 MAR 2014

## Genetic particle filter application to land surface temperature downscaling

Rihab Mechri<sup>1</sup>, Catherine Ottlé<sup>1</sup>, Olivier Pannekoucke<sup>2</sup>, and Abdelaziz Kallel<sup>3</sup>

<sup>1</sup>LSCE, UMR 8212, CNRS-CEA, Gif-sur-Yvette, France, <sup>2</sup>CNRM-GAME, UMR 3589, Météo-France-CNRS, Toulouse, France, <sup>3</sup>ENET'Com, ATMS, Cite Ons, Sfax, Tunisia

**Abstract** Thermal infrared data are widely used for surface flux estimation giving the possibility to assess water and energy budgets through land surface temperature (LST). Many applications require both high spatial resolution (HSR) and high temporal resolution (HTR), which are not presently available from space. It is therefore necessary to develop methodologies to use the coarse spatial/high temporal resolutions LST remote-sensing products for a better monitoring of fluxes at appropriate scales. For that purpose, a data assimilation method was developed to downscale LST based on particle filtering. The basic tenet of our approach is to constrain LST dynamics simulated at both HSR and HTR, through the optimization of aggregated temperatures at the coarse observation scale. Thus, a genetic particle filter (GPF) data assimilation scheme was implemented and applied to a land surface model which simulates prior subpixel temperatures. First, the GPF downscaling scheme was tested on pseudoobservations generated in the framework of the study area landscape (Crau-Camargue, France) and climate for the year 2006. The GPF performances were evaluated against observation errors and temporal sampling. Results show that GPF outperforms prior model estimations. Finally, the GPF method was applied on Spinning Enhanced Visible and InfraRed Imager time series and evaluated against HSR data provided by an Advanced Spaceborne Thermal Emission and Reflection Radiometer image acquired on 26 July 2006. The temperatures of seven land cover classes present in the study area were estimated with root-mean-square errors less than 2.4 K which is a very promising result for downscaling LST satellite products.

## 1. Introduction

Land surface temperature (LST) is one of the most important variable giving access to surface water and energy states. For many applications—agrometeorology, urban climatology, water balance studies, to cite but a few—LST high-resolution monitoring in space and time is required because of its high spatiotemporal variability. Nevertheless, despite the growing availability of LST satellite products, the available satellite sensors offer either a high spatial resolution (HSR) or a high temporal one. For instance, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) has a spatial resolution of 90 m in the thermal infrared (TIR) spectral domain but a bimonthly sampling frequency. The Spinning Enhanced Visible and InfraRed Imager (SEVIRI) radiometer onboard Meteosat Second Generation (MSG) satellite has a sampling frequency of 15 min with a spatial kilometeric resolution ranging from 3 km to 5 km. Such coarse spatial resolution (CSR) is not sufficient to monitor surface fluxes and variables on heterogeneous landscapes. Thus, the development of methods to overcome these limitations becomes essential in order to track LST evolution at high spatiotemporal resolutions (HSTR). Different approaches have been so far developed to downscale LST, ranging from linear regression schemes to data assimilation ones; see Zhan *et al.* [2013] for a review.

The most common approach is based on linear regressions [Kustas *et al.*, 2003; Agam *et al.*, 2007a, 2007b; Liu and Pu, 2008], which exploit the negative relationship between LST and vegetation density, generally assessed with the Normalized Difference Vegetation Index (NDVI) (or the vegetation cover fraction  $f_{\text{cover}}$ ), both increasing with the amount of biomass. In satisfactory soil moisture conditions, the canopy evapotranspiration is generally higher than the bare soil evaporation (because the root system allows access to deeper water content, combined to the higher surface roughness. For more details, see Olioso *et al.* [2002]; Douville *et al.* [2012]). Consequently, vegetated surfaces present generally lower temperatures than barren ones, leading to a negative correlation between LST and NDVI or  $f_{\text{cover}}$ . Some studies have focused on

this empirical covariation of LST and NDVI to develop sharpening schemes known as TsHARP (the classical sharpening approach/model) [see *Kustas et al.*, 2003; *Agam et al.*, 2007a, 2007b; *Liu and Pu*, 2008]. TsHARP approach assumes that LST has a unique relationship to photosynthetically active biomass which can be approached with shortwave reflected radiation ( $\text{NDVI}/f_{\text{cover}}$ ) across a given satellite image scene. To enhance CSR LST, TsHARP uses an inverse linear covariation between  $\text{NDVI}/f_{\text{cover}}$  maps acquired at HSR and CSR LST maps. This assumption is valid for homogeneous vegetated areas, and good results have been found over rainfed agricultural areas in *Agam et al.* [2007b]. Yet the validity of  $\text{LST}-\text{NDVI}/f_{\text{cover}}$  relationship remains questionable when working on complex heterogeneous landscapes. *Jeganathan et al.* [2011] investigated four TsHARP variants over a mixed agricultural landscape in India and concluded that TsHARP is efficient only when locally applied over relatively homogeneous fields and not for regional scales. To address this issue, *Inamdar et al.* [2008] and *Inamdar and French* [2009] introduced emissivity data to enhance Geostationary Operational Environmental Satellite LST over the U.S. Southwest region and better results have been found compared to the classical TsHARP. *Merlin et al.* [2010] proposed to extend the validity domain of the classical TsHARP by introducing the temperature difference between photosynthetically active and nonphotosynthetically active vegetation. Statistical results showed improvements in terms of robustness and accuracy, for the disaggregated LST, and better agreement with validation data, compared to the classical model. Furthermore, *Merlin et al.* [2012] introduced a new approach by integrating the main surface parameters involved in the surface energy budget which are open water and soil evaporative efficiency in addition to the senescent/green  $f_{\text{cover}}$ . The method uses a linearized radiative transfer to relate four satellite products inverted from different shortwave bands at different resolutions to CSR LST. The introduction of open water and soil evaporative efficiency fractions improved the disaggregation results, but the operational application of the method is still not possible because soil evaporative efficiency data are currently not available at fine resolution over large areas. In the case of artificial surfaces, *Dominguez et al.* [2011] found that the LST-NDVI relationship is not consistent for urban areas like the city of *Puerto Rico* and proposed a bivariational model including albedo in addition to NDVI variable in their high-resolution urban thermal sharpener model. This yielded to smaller mean absolute error (MAE) and higher correlation coefficient compared to the classical TsHARP. To extend the validity of sharpening schemes over more complex areas and at regional scales a data mining sharpener (DMS) has been introduced by *Gao et al.* [2012]. The DMS technique builds regression trees between TIR brightness temperatures and shortwave spectral reflectances based on intrinsic characteristics. A comparison between TsHARP and DMS showed that DMS outperforms TsHARP in all the cases (homogeneous areas as well as complex heterogeneous areas). All these semiempirical schemes suppose that the LST relationship to the different observed variables could be reproduced with a linear combination or a polynomial decomposition. However, the relationship between LST and shortwave band signals is much more complicated because of the complexity of the interacting biophysical processes.

Other statistical approaches were proposed to solve the downscaling problem based on data assimilation strategies. For example, *Ottlé et al.* [2008] proposed the inversion of subpixel variables by multilinear regressions constrained by prior temperatures provided by a physical land surface model (LSM). Two spatial stationary assumptions were considered to solve the ill-posed problem. First, the mixed pixels are assumed to be composed of a few fundamental components called end-members. Second, the CSR LST is modeled as a linear combination of end-member temperatures weighted by the proportion of each end-member within the mixed pixel area. The results show rapid decrease of the performances when too much heterogeneous landscapes are considered. Then, *Kallel et al.* [2013] proposed a more adapted inversion scheme introducing *Markovrandomfield* modeling to represent LST in space and time. The stationary assumptions were relaxed through sequential inversions of the end-member temperatures by using the maximum a posteriori criterion. Compared to the  $\text{NDVI}/f_{\text{cover}}$  downscaling approach, better correlations were shown as the physical process interactions are accounted for in the LSM [*Kallel et al.*, 2013].

Even though LSM approaches are more difficult to implement operationally and rely on data and models subject to various sources of uncertainties, they allow to provide prior estimations of the spatial and temporal variability of the surface temperature which can be used to solve the downscaling issue. In this work, we propose a new approach based on an ensemble data assimilation (DA) method. The approach is based on Sequential Monte Carlo filter, also referred to as particle filter (PF) (or particle smoother if observations are taken over a time window), which has been successfully implemented to estimate parameters or states in nonlinear models [see *Doucet et al.*, 2000; *Arulampalam et al.*, 2002; *Del Moral et al.*, 2010;



Van Leeuwen, 2010; Snyder, 2011]. The idea here is to drive a sharpening algorithm at pixel scale with no spatial or temporal stationary assumptions, using a dynamic model able to simulate the end-member (or subpixel) temperature variability, and to assimilate the CSR observations. Given the fact that LST is mainly determined by land cover, meteorological forcing and soil moisture conditions, the implementation of an LSM on each type of vegetation (and hydric conditions) present in the CSR pixel should be sufficient for the first step (the end-members are the different land cover classes). In this work, the Suivi de l'Etat Hydrique des Sols (SEtHyS) LSM [Coudert, 2006] was used to simulate the surface temperature of the different end-members and to assimilate the coarse resolution observations. A LSM is a dynamic model representing the energy and mass transfers governing the soil-vegetation-atmosphere continuum as well as the surface variables in particular soil moisture and surface temperature. The originality of our approach is the use of a DA method to simultaneously downscale CSR data and calibrate the LSM parameters for all the end-member temperatures making up the CSR pixel grid cell. Our paper is organized as follows: the DA approach and its application on a synthetic pixel composed of four end-members is presented in section 2. Section 3 gives the details on the experiment setup and the resulting downscaled temperatures and discusses the performances of the method considering the observation sampling and the errors statistics. The application of our approach on actual data is presented in section 4. A summary and conclusions are given in section 5.

## 2. Methodology

In the following work, we assume that the subpixel (end-member) temperature variability is mainly attributed to land cover heterogeneity. Then, the objectives are to estimate the respective temperatures of each land cover class inside a mixed pixel composed of fractions of these different vegetation types. This assumption is valid at regional scale (few kilometers size) where the atmospheric forcing can be supposed homogeneous and when flat landscapes are considered. In our case, the prior knowledge of the vegetation classes fractions within the CSR pixel is needed and can be provided by a HSR land cover mapping. The problem can be addressed by constraining a dynamic LSM able to simulate prior subtemperatures and pixel-aggregated temperature given the fractions of each vegetation type, with the assimilation of CSR. The following section presents the DA developments and their implementation in the SEtHyS LSM.

### 2.1. The Particle Filter

Data assimilation aims to provide the best estimation of the state of a system or its unknown parameters,  $x_q$ , using observations,  $y_q$ . The system is assumed to evolve according to a Markov process featured by the transition probabilities  $p(x_q|x_{q-1})$  and represented by the propagator  $x_{q+1} = \mathcal{M}(x_q)$ . The observations  $y_q$  are related to the state  $x_q$ , thanks to the nonlinear observational operator  $\mathcal{H}$  according to  $y_q = H(x_q) + \varepsilon_q^o$  where  $\varepsilon_q^o$  denotes an error modeled as a random centered Gaussian vector of covariance matrix  $R_q$ . The general estimation Bayesian framework is given by the nonlinear filtering theory [Jazwinski, 1970]. It describes the time evolution of the full probability density function  $p(x_q)$  conditioned by the dynamics and the observations. The probability dynamics can be expanded in two steps: a forecast step and an analysis step. The forecast step is obtained from the Chapman-Kolmogorov rule

$$p(x_q|y_{q-1}) = \int p(x_q|x_{q-1})p(x_{q-1}|y_{q-1})dx_{q-1} \quad (1)$$

while the analysis step is deduced from the Bayes rule as

$$p(x_q|y_q) \propto p(y_q|x_q)p(x_q|y_{q-1}) \quad (2)$$

(to within a normalization term). An analytic solution of the distribution evolution can be obtained for linear dynamics and Gaussian distributions, leading to the Kalman filter equations [Kalman, 1960]. However, for practical applications where distributions are not Gaussian and their dynamics not linear, this approach requires additional simplifications where a sampling distribution is considered in place of full distributions. This leads to the ensemble Kalman filter algorithm [Kalman, 1960] for the nonlinear ensemble-based

extension of the Kalman filter, or the particle filter approach [Doucet et al., 2000; Del Moral, 2004] for the discretization of the nonlinear filter. It follows that the density  $p(x_q)$  is approximated as the empirical density

$$p(x_q) \approx p_N(x_q) = \frac{1}{N} \sum_k \delta_{x_q^k} \tag{3}$$

where  $x_q^k$  is an ensemble of  $N$  independent identically distributed samples of the law  $p(x_q)$  and where  $\delta$  denotes the Dirac's distribution located in the sample  $x_q^k$ . From this discrete point of view, the analysis step is equivalent to

$$p(x_q|y_q) \approx p_N(x_q|y_q) = \sum_k w_q^k \delta_{x_q^k} \tag{4}$$

where

$$w_q^k = \frac{p(y_q|x_q^k)}{\sum_i p(y_q|x_q^i)} \tag{5}$$

At each analysis step, a new ensemble  $\hat{x}_q^k$  is obtained from the sampling of the distribution  $p_N(x_q|y_q)$  so that

$$p(x_q|y_q) \approx \frac{1}{N} \sum_k \delta_{\hat{x}_q^k} \tag{6}$$

The forecast step is now written as follows:

$$p(x_{q+1}) \approx \frac{1}{N} \sum_k \delta_{x_{q+1}^k} \tag{7}$$

where  $x_{q+1}^k = \mathcal{M}(\hat{x}_q^k)$  denotes the ensemble of forecast deduced from the states  $\hat{x}_q^k$ . The filter can be extended into a smoother considering observations over an assimilation window [see Del Moral, 2004; Rémy et al., 2012]. In this case, we assume the availability of  $M$  observations between two analysis times  $q$  and  $q+1$ :  $(y_q^m)_{m=1,\dots,M}$ . The potential function  $g_q^m(x_q^k) = p(y_q^m|x_q^k)$ , which designs the likelihood function in filtering problems, allows to decide whether a particle  $x_q$  is killed (with low potential) or kept (with high potential) and duplicated into several offsprings. The weights can be expressed as follows:

$$w_m^k = \frac{g_m(x_q^k)}{\sum_{j=1}^N g_m(x_q^j)} \tag{8}$$

Finally, we have

$$G_q(x_q^k) = \prod_{m=1}^M g_q^m(x_q^k)$$

$$W_q^k = \prod_{m=1}^M w_m^k = \frac{G_q(x_q^k)}{\sum_{j=1}^N G_q(x_q^j)} \tag{9}$$

**2.2. Genetic Selection-Multinomial Resampling**

One of the major drawbacks of PF schemes is the degeneracy phenomenon. In fact, after several iterations, all but one particle will have negligible weight. A suitable measure of degeneracy of the PF algorithm is the effective ensemble size  $N_{\text{eff}}$  [Bergman, 1999; Arulampalam et al., 2002]. In our case, the estimation of  $N_{\text{eff}}$  can be obtained by

$$N_{\text{eff}} = \left[ \sum_{k=1}^N (W_q^k)^2 \right]^{-1} \tag{10}$$

where  $N$  is the ensemble size and  $W_q^k$  is the normalized weights defined in equation (9). According to the  $N_{\text{eff}}$  definition in equation (10), we have  $N_{\text{eff}} \leq N$ . Small  $N_{\text{eff}}$  indicates severe degeneracy. Such problem could be avoided either by choosing very large ensemble sizes, which is often impractical, or by resampling the particles. A review of particle filter resampling schemes is presented in Arulampalam et al. [2002] and Douc and Cappe [2005].

In our case, we choose to work with the basic genetic particle filter (GPF) with multinomial resampling, which belongs to the genetic algorithm (GA) family. In fact, the particle filtering problem could be studied from a GA perspective if we focus on the selection-resampling step. GA could be defined as a stochastic searching algorithm (a function optimizer) ensuing from Darwin's evolution theory, simulating the well-known *survival of the fittest* individuals evolution. The fitness of an individual could be presented as the closeness of a chromosome from the optimal solution. Fitter individuals crossbreed and produce better offsprings to improve the whole population fitness. Mutation could also happen during offspring production. This definition highlighted two main similarities between GA and PF: the first one is the selection of the "best" offsprings according to their closeness to the optimal solution (selection of particles with higher potential  $G_q$ ), and the second one is the crossover and mutation to produce the new offsprings that promote the whole population (resampling particle from the selected particles distribution). More details on the GPF implementation is available in *Uosaki et al.* [2004] and *Kwok and Zhou* [2005].

Once particles are selected, a multinomial resampling [*Rémy et al.*, 2012] is performed to produce the new population from the selected particles (mutation). The multinomial algorithm replicates particles with higher weights, and a noise is added to copied particles. This noise is chosen large enough to differentiate copied particles from seed ones (to allow a better exploration of the solution space) and keep coherent with model uncertainties (to avoid filter collapse when the noise is too large). More details on the genetic selection and multinomial resampling implementation of PF algorithm could be found in *Kwok and Zhou* [2005] and *Rémy et al.* [2012].

### 2.3. Downscaling-Calibration Approach

As already mentioned in the introduction, the SETHyS LSM [*Coudert*, 2006] has been used to assimilate the coarse observations and to estimate the subpixel temperatures. SETHyS predicts the surface temperature evolution by solving the hydric and energy budgets at the land surface. It is forced with micrometeorological parameters such as air temperature and humidity, downwelling shortwave and longwave radiations, irrigation/precipitation, and wind speed. Two sources are used to represent the soil-vegetation-atmosphere system: the soil and the aboveground vegetation. An energy budget is computed for both sources, and consequently, both temperatures are determined. A detailed description of the radiative transfer in the solar and thermal infrared domains allows to compute the directional surface reflectances and radiative temperature above the canopy [*Coudert et al.*, 2008]. The model has been already used to test LST downscaling approaches in previous studies [*Ottlé et al.*, 2008; *Kallel et al.*, 2013; *Guillevic et al.*, 2012], and particularly on the "Crau-Camargue" agricultural site in southeastern France (43.53°N, 4.66°E). The case study chosen for this work consists of a single pixel composed of four end-members equally distributed in order to represent a typical agricultural area. The climate of the study region is Mediterranean, with irregular rainfall, long dry periods in spring and summer, and strong winds. The micrometeorological forcing database for the year 2006 [*Courault et al.*, 2008] was used in our test case. The four end-members composing the CSR pixel are agricultural. During the spring/summer period, which will be used in the assimilation tests, bare soil, prairie, wheat, and rice fields can coexist. In the following, we design by the subscripts "1," "2," "3," and "4" the dominant land cover classes "Bare Soil," "Prairie," "Wheat," and "Rice." Their respective fractions were set equal to 0.25 and assumed constant during the study period. For each of these classes, reference soil and vegetation parameters were assigned and were used to generate reference temperatures over the year 2006. These temperatures will be further considered as reference subpixel (or end-member) temperatures and aggregated to generate CSR pseudo-observations that will be used in the assimilation process. A white Gaussian noise  $\epsilon_q^o$  was added after aggregation to generate noisy observations. The observation error amplitude was initially set to the standard value of 2 K similarly to *Kallel et al.* [2013]. The impact of the observation error amplitude on the assimilation performances has been studied in a second step. In our case, the particles represent a set of SETHyS parameters corresponding to the different classes. More generally, the particle can be represented as follows:

$$x_q = \left[ P_{1,q}^1, \dots, P_{D^1,q}^1, P_{1,q}^2, \dots, P_{D^2,q}^2, P_{1,q}^3, \dots, P_{D^3,q}^3, P_{1,q}^4, \dots, P_{D^4,q}^4 \right], \quad (11)$$

where  $(D^i)_{i \in [1:4]}$  refers to the number of selected parameters for the land cover class  $i$ . The model  $\mathcal{M}$ , as represented in the subsection 2.1, is the particle propagator. Given that particles (SETHyS parameters) remain static during an assimilation window,  $\mathcal{M}$  is, in our case, the identity function to which a white Gaussian noise with a standard deviation (SD) equal to  $\sigma_\beta = 0.1$  is added after every resampling step:  $x_{q+1}^k = \mathcal{M}x_q^k = x_q^k + \xi_q$

**Table 1.** SEtHyS Parameters Used for the Calibration

Land Cover Class	Selected Parameters and Prior	Variation Range	Reference Values
Bare soil	$P_1^1$ : Factor controlling the volumetric heat soil capacity	[0.5, 3.0]	2.5
	$P_2^1$ : Bare soil emissivity	[0.93, 0.97]	0.935
Prairie	$P_1^2$ : Factor controlling the volumetric heat soil capacity	[0.5, 3.0]	2.5
	$P_2^2$ : Vegetation emissivity	[0.96, 1.0]	0.965
Wheat	$P_1^3$ : Factor controlling the volumetric heat soil capacity	[0.5, 3.0]	2.5
	$P_2^3$ : Vegetation emissivity	[0.96, 1.0]	0.965
Rice	$P_1^4$ : Factor controlling the volumetric heat soil capacity	[0.5, 3.0]	2.5
	$P_2^4$ : Vegetation emissivity	[0.96, 1.0]	0.965

where  $k$  refers to the  $k$ th particle ( $k \in [1, \dots, M]$ ),  $\xi_q = \mathcal{N}(0, \beta_q)$  and  $\beta_q = \sigma_\beta^2 \times l$ . Indeed, the nonlinearity of land surface processes and the rapid changes of boundary conditions like precipitation or irrigation events and various agricultural practices can lead to rapid changes in surface states and parameter values and impact the estimation of subtemperatures. Such a problem could be overcome with the resampling step. In fact, resampling particles allows to explore globally the parameter space and to find optimal temperatures far from the local prior values without being drawn in false directions. A previous temperature sensitivity analysis based on a screening approach [Frey and Patil, 2002] permitted the selection of the parameters to calibrate. The results show that on the study period, two parameters for each class have larger sensitivities. A detailed list of these parameters is available in Table 1. Once the parameters are defined, a first-guess particles ensemble is generated from  $N$  random samples of the parameter space ( $N$  random draws from a uniform distribution). Every particle leads to a quadruplet of subpixel prior temperatures  $[T_1^k, T_2^k, T_3^k, T_4^k]_{k \in [1, \dots, M]}$ , as output of SEtHyS model (the subpixel temperatures are completely determined once the parameters fixed). Then, these  $N$  subpixel temperatures are aggregated and compared to the synthetic observations using the GPF. Then, a weight is attributed to each particle according to the discrepancy between the observed temperature  $T_o$  and the simulated aggregated temperatures  $T_s$ . At the end of the analysis step, the mean of the subpixel temperatures corresponding to the selected particles represents the solution of the downscaling problem. The kept particles are resampled to create the new set of calibrated SEtHyS parameters. The definition of the observation operator  $\mathcal{H}$  is described by equation (12), and the distance between the observation “ $Y_q$ ” and a particle “ $x_q^k$ ” is given by equation (13).

$$\begin{aligned} \mathcal{H}(x_q^k) &= \left[ \frac{\sum_{i \in [1:4]} \alpha_i \epsilon_i \{G(x_q^k)\}^4}{\sum_{i \in [1:4]} \alpha_i \epsilon_i} \right]^{\frac{1}{4}} \\ &= \left[ \frac{\sum_{i \in [1:4]} \alpha_i \epsilon_i T_{i,q}^4}{\sum_{i \in [1:4]} \alpha_i \epsilon_i} \right]^{\frac{1}{4}} \end{aligned} \tag{12}$$

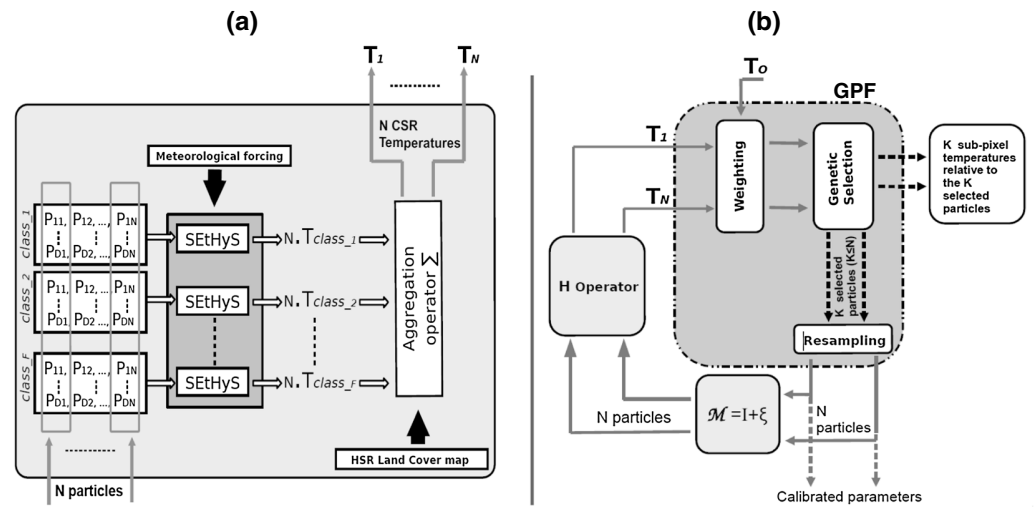
$$\| \mathcal{H}(x_q^k) - Y_q \| = \sqrt{\frac{\left[ \left( \frac{\sum_{i=1}^4 \alpha_i \epsilon_i (T_{i,q}^k)^4}{\sum_{i=1}^4 \alpha_i \epsilon_i} \right)^{\frac{1}{4}} - T_{o,q} \right]^2}{\sigma_o^2}} \tag{13}$$

where  $G$  designs SEtHyS model, superscripts  $k$  and  $q$  refer, respectively, to the  $k$ th particle and the  $q$ th time window,  $\epsilon_i$  is the surface emissivity of the  $i$ th class, and  $\alpha_i$  is the corresponding cover fraction.

The charts of Figure 1 give an overview of the downscaling-calibration process. The first chart (Figure 1a) presents in details the observation operator  $H$ , and the second one (Figure 1b) presents the downscaling-calibration loop.

### 3. Results and Discussion

In this section, we present the results obtained by the application of the particle filter on the synthetic pixel previously described in section 2. The performance of the downscaling approach is discussed in terms of efficiency, in relation with the amplitude of the observation errors and the observation sampling.



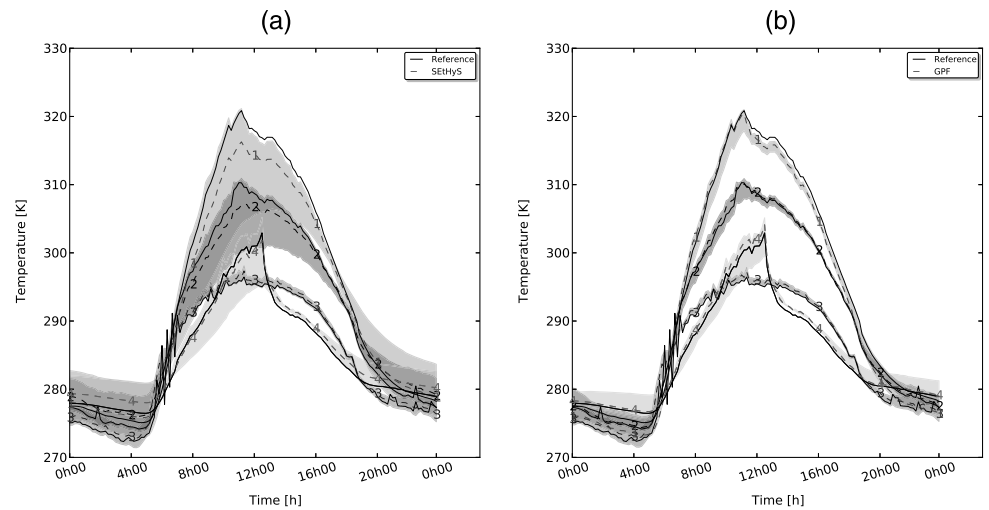
**Figure 1.** (a) The different parts of the  $H$  operator diagram. The subscripts  $F$ ,  $D$ , and  $N$  design, respectively, the number of classes present in the CSR pixel, the number of parameters to calibrate for each class, and the number of particles.  $(T_{class_j})_{j \in \{1, \dots, N\}}$  present the  $N$  CSR temperatures corresponding to the  $N$  particles (SEThyS outputs). (b) An overview of the different steps inside the GPF downscaling-calibration process. The subscript  $K$ , refers to the number of kept particles after the genetic selection.

### 3.1. Experimental Settings

For clarity, in the following, we will omit the time subscript " $q$ ". For each land cover class, the synthetic reference temperatures and the pseudo-observations were generated on a 124 day period, starting from 28 April 2006 at a 20 min time step. Observations are assumed unbiased. The standard deviation (SD) of the observation error  $\sigma_o$  was set to a value of 2 K (to be comparable to the previous work of Kallel *et al.* [2013]). The particle ensemble size was set to  $N = 200$  after a series of tests to avoid collapse [Rémy *et al.*, 2012]. At the beginning of the assimilation period, wheat, rice, prairie, and bare soil coexist. The wheat class is, however, in a senescent phenological phase and was harvested on 20 May. Thus, the whole assimilation period was divided in two periods. In the first one, ( $P1$ ), from 28 April to 20 May, the CSR pixel contains four distinct classes and in the second one, ( $P2$ ), from 21 May to 29 August, only three classes remain in the CSR pixel (bare soil, prairie, and rice). The micrometeorological forcing data were independently acquired for the three vegetation classes and assumed equal for bare soil and wheat, because the hydric states of these two land cover classes are very close. Indeed, unlike prairie and rice classes which are irrigated, wheat and bare soil are very dry. *Leaf area index (LAI)*, vegetation height, and soil texture were prescribed according to field measurements.

### 3.2. Downscaling Results and Impact of Observation Error Amplitude

In all the synthetic experiments, the assimilation is performed continuously with an assimilation window of 1 day with a maximum number of 72 observations/assimilation window. Only the results obtained for period ( $P1$ ) are shown here. Actually, ( $P1$ ) is the period when the four classes present the most different properties since wheat was harvested at the beginning of period ( $P2$ ) and comes down as bare soil. The GPF performances are evaluated by comparing prior and posterior downscaling results over the assimilation period. From the start of the assimilation process, a significant reduction of the ensemble variance was observed. Figure 2 shows the prior (Figure 2a) and posterior (Figure 2b) downscaling results for the 6th assimilation window. The assimilation of CSR observations has significantly improved the downscaling results. The downscaled temperatures of the four classes are closer to the reference values ( $RMSE_{prior} = 1.2K$ ,  $RMSE_{posterior} = 0.3K$ ). Figure 3 presents the mean average diurnal cycle of surface temperature over the period ( $P1$ ). Comparing the prior (Figure 3a) and posterior (Figure 3b) diurnal cycles, the spread of solutions is largely reduced (40% reduction) and the classes are properly separated. However, the improvement is less significant than for the 6th time window. In fact, the filter performances are fairly sensitive to meteorological conditions such as windy or nonlimited soil moisture conditions and to observation noise. Since the observations are noised randomly, some days will be more or less noisy than others. Indeed, meteorological changes occurring along the assimilation period such as strong winds characterizing the rice

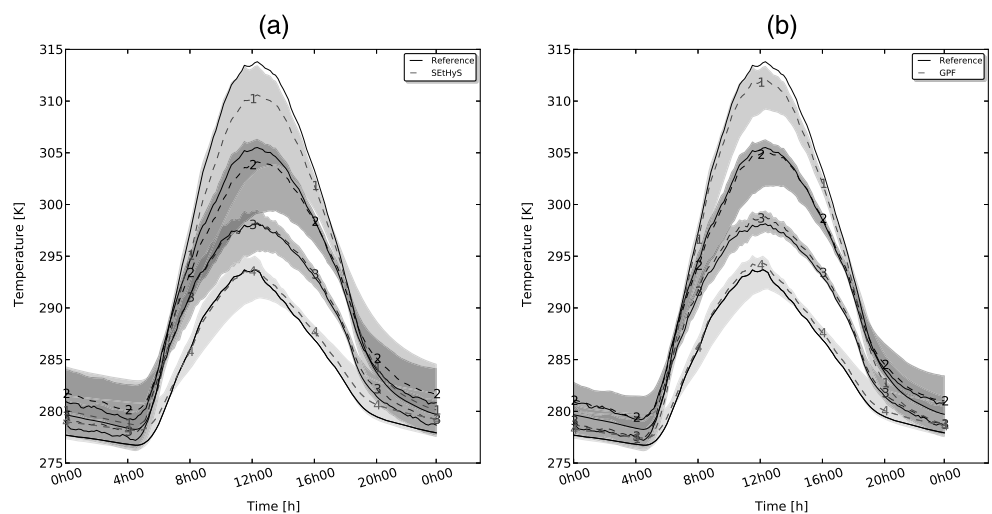


**Figure 2.** Downscaling temperatures for Day 6 (3 May 2006). (a) The prior downscaled temperatures compared to the reference temperatures. (b) The downscaled temperatures after assimilating CSR observations. The line markers 1, 2, 3, and 4 correspond to bare soil, prairie, wheat, and rice classes, respectively. The continuous lines correspond to the reference subpixel temperatures.

forcing data and irrigation practices of the prairie class lead to a decrease of the simulated temperatures (two irrigations at day of year (DOY) 122 and DOY 137 and two precipitation events at DOY 128 and DOY 133 occurred during period ( $P_1$ )). In these cases, the subtemperatures tend toward the air temperature ( $T_i \approx T_{air}$ ,  $i \in \{2, 4\}$ ) assumed to be exact. The prior error tends therefore to zero, and the assimilation of CSR observations does not really improve the downscaling results. To illustrate the contribution of the CSR observations, the efficiency rate  $I$  (see equation (14)) has been calculated.

$$I_i[\%] = \left( 1 - \frac{RMSE_{posterior}^k}{RMSE_{prior}^i} \right)_{i \in [1:4]} \times 100 \quad (14)$$

The mean of the efficiency rates on the first period ( $P_1$ ) are shown in the second column of Table 2. In fact, the more efficient the GPF, the closest  $I_{i \in [0:4]}$  to 100%. The results averaged over 100 realizations on the



**Figure 3.** Downscaling temperatures for the period ( $P_1$ ). (a) The prior downscaled temperatures compared to the reference temperatures. (b) The downscaled temperatures after assimilating CSR observations. The line markers 1, 2, 3, and 4 correspond to bare soil, prairie, wheat, and rice classes, respectively. The continuous lines correspond to the reference subpixel temperatures.

**Table 2.** Efficiency Rates for Different Amplitudes of the Observation Error SD  $\sigma_O$

Land Cover Class	$I_{\sigma_O=0.5K}^k$	$I_{\sigma_O=2.0K}^k$	$I_{\sigma_O=4.0K}^k$
Bare Soil	54%	56%	53%
Prairie	33%	30%	26%
Wheat	60%	59%	54%
Rice	48%	46%	43%

period ( $P_1$ ) showed that the downscaling is more efficient for the wheat and the bare soil land cover classes. The respective efficiency rates are 59% and 56%. Lower rates were found for rice (46%) and prairie (30%). This skill reduction is explained by the hydric conditions of these two classes. Indeed, rice and prairie are well irrigated and evapotranspiration is close to potential rates. Therefore, their respective temperatures are close to the air temperature and the prior model error is lower compared to bare soil and wheat classes. The contribution of the observations is then lower. These synthetic experiments have been carried out for 2 other values of the observation error (SD), 0.5 K and 4 K, in order to evaluate the impact of the observation error amplitude on GPF-downscaling method performances (respectively, columns 1 and 3 of Table 2). Assimilating CSR observations has significantly improved the estimation of the subpixel temperatures for the different classes ( $I_i \geq 25\%, \forall i \in [1 : 4]$ ). The results show also that GPF performances decrease with the amplitude of SD as expected (the larger the SD value, the less confidence we have in the observations).

### 3.3. Impact of Observation Sampling on GPF Performances

The effect of observation sampling on GPF skills was investigated and the contribution of nighttime observations as well. The observation error standard deviation was still set to 2 K, and the observations were provided for different periods of the day. Four scenarios have been tested. In the first three scenarios, the time step of the observations was set to 20 min, but the day time observation period was different. The different scenarios are described below: (i) Scenario 1: all the observations are available (i.e., 72 observations per day), (ii) Scenario 2: the observation window is daytime only ([10:00 → 18:00], i.e., 25 observations), (iii) Scenario 3: the observation window is 4 h around noon ([10:00 → 14:00], i.e., 13 observations), and (iv) Scenario 4: only one observation is available at noon (12:00). The objective of these experiments is to assess the impact of the observation time period on the performances of the GPF-downscaling approach and the contribution of nighttime observations (before 10:00 and after 18:00) on the downscaling results. A number of 100 realizations for each of these scenarios were performed, and the efficiency rates were calculated. The results illustrated in Table 3 were obtained by averaging the efficiency rates over the 20 day assimilation window of the period ( $P_1$ ) and 100 realizations. At first sight, a slight decrease of GPF performances is noticed when the number of observations decreases. In fact, observations retrieved before 10:00 and after 14:00 do not really improve downscaling results (nonsignificant differences between the efficiency rates which are less than 10%). Thus, we can conclude that the contribution of nighttime observations is much lower than daytime ones and that the most important observations are those acquired during the daytime period when the surface temperature deviation from the air temperature is the largest (see Figures 2 and 3). In Scenario 4, despite the few number of observations (only one observation per day), the efficiency rates are still positive and above 20%. Such result is very interesting and proves the added value of the GPF even with a limited number of observations. To better assess this point and the impact of the observation time, various scenarios similar to Scenario 4 have been performed varying the observation time during the daytime period of the assimilation window (since nighttime observation contribution is not important, only observations between 6 A.M. and 6 P.M. were considered). Table 4 presents the different scenarios performed and their respective efficiency rates (averaged over 100 realizations). Results show that for all land cover types the best efficiency rates are obtained when the observation is available in the morning (maximum efficiency rates are at 10 A.M. and 12 A.M.). However, for wet classes such as prairie and rice, the efficiency rates are less sensitive to the observation time (nearly the same value for all the observation times). This result confirms our first conclusion about the contribution of noontime observations in the case of a single observation

**Table 3.** Efficiency Rates for Different Observation Times/Frequencies

Land Cover Class	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Bare Soil	56%	55%	54%	39%
Prairie	30%	26%	26%	20%
Wheat	59%	56%	55%	44%
Rice	46%	42%	40%	34%

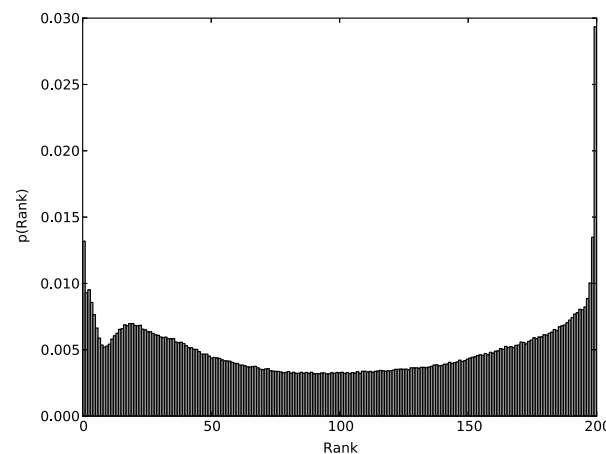
**Table 4.** Scenario 4: Efficiency Rates for Different Observation Times

Observation Time	6 A.M.	8 A.M.	10 A.M.	12 A.M.	2 P.M.	4 P.M.	6 P.M.
$I_1$	33%	36%	40%	39%	36%	33%	33%
$I_2$	18%	20%	21%	20%	19%	18%	18%
$I_3$	39%	41%	44%	44%	41%	39%	39%
$I_4$	31%	32%	34%	34%	31%	30%	31%

availability. Finally, we can draw that the most valuable observations are noontime observations and the best efficiency rates are obtained when more observations are considered over the assimilation window.

### 3.4. Rank Histogram

To assess GPF performances for downscaling CSR LSTs, rank histograms have been calculated over the whole period ( $P1$ ). In fact, rank histograms, also known as “Talagrand” diagrams, are interesting tools to evaluate ensemble methods forecasts/solutions by determining their reliability and diagnosing errors in mean and spread. Rank histograms are generated by repeatedly tallying the rank of the verification variable (observations) relatively to ensemble members values sorted in increasing order. In other words, they are a measure of the statistical indistinguishability of an observation from the predicted ensemble values assuming that they are independent realizations of a same predicted distribution function. The flatness of the rank histogram implies that observations fall with equal probability in each of the different intervals defined by the ensemble prediction values. Thus, the flatness is a measure of the reliability of the prediction system. More details on the implementation of the rank histogram are available in *Anderson* [1996], *Talagrand et al.* [1997], and *Hamill* [1997, 2001]. In our case, the reference CSR rank histogram has been drawn from 1000 repetitions of the Scenario 1 (see section 3.3). A bias of 0.25 K has been considered to get a more realistic observation error representation. The reference temperatures are the same than for Scenario 1. The assimilation period is ( $P1$ ), the ensemble size is  $N = 200$ , and the observation frequency is 20 min. After the selection step, the number of selected temperatures is generally less than  $N = 200$ . A white Gaussian noise with a SD equal to 0.1 K has been added to copied particle temperatures to be able to draw the reference rank histogram. Figure 4 presents the true rank histogram which shows a slight U shape with a shallow inflation located at the low ranks. The U shape reflects that the solution ensemble spread is not large enough to include all reference temperatures. Thus, the probability that reference temperatures falls out the solution interval is large. When comparing the extreme ranks, we notice that  $\text{rank}^N$  frequency is greater than  $\text{rank}^0$  one ( $\text{freq}(N) = 2.9\% > \text{freq}(0) = 1.8\%$ ). This means that reference temperatures tend to be greater than the maximum selected temperatures. Such behavior was expected since reference temperatures are generated using parameter values that are very close to parameter ranges maximum values. The presence of the shallow inflation in the histogram is the result of the bias prescribed in the observations as confirmed in a similar experience with no biased observations (not shown here). This interpretation was also found in *Hamill's*



**Figure 4.** Reference LST rank histogram.

[2001] and *Wilks's* [2006] works, who noted that the deviation from flatness of a rank histogram can be interpreted as ensemble overdispersion/underdispersion and/or unconditional biases in forecast/observations. We should also note that the miss-flatness in the histogram is not huge when comparing the probabilities of the extreme ranks 0 and  $N$  to the other ranks (the misfit is less than 3%). Such result confirms that GPF-downscaling solutions are quite reliable even when biased CSR observations are considered. Furthermore, it is true that the flatness of rank histogram tells about ensemble consistency; however, the converse is not true in all the cases [see *Hamill*, 2001; *Wilks*, 2011].



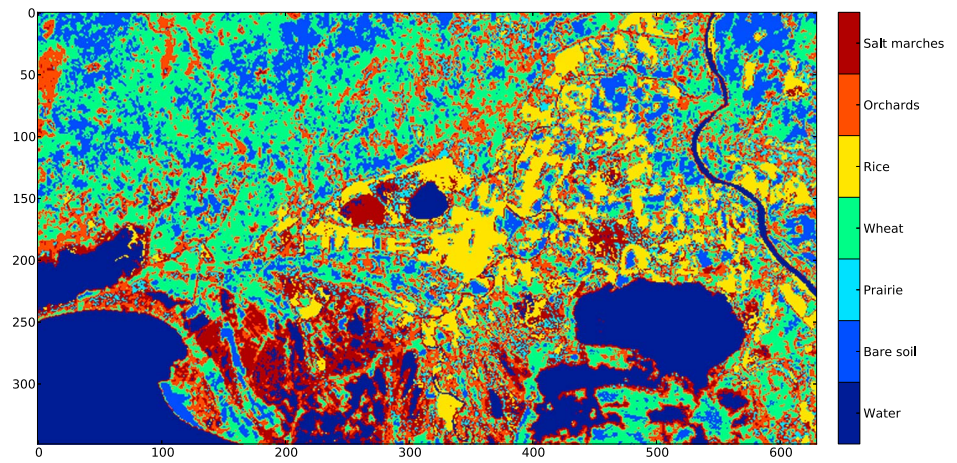


Figure 5. The 90 m × 90 m land cover map of the Crau-Camargue area with seven classes.

#### 4. Application of GPF on Actual Data

In this section, the GPF approach is evaluated against actual data acquired over the Crau-Camargue region in southeastern France. An ASTER image was acquired during period (P2) on 26 July at 10:47 A.M. Two ASTER products were generated: the land cover map with 15 m × 15 m spatial resolution in the visible band and the LST map with a 90 m × 90 m spatial resolution in the thermal band. The land cover map has been aggregated to the LST map resolution. The initial land cover map classification provided by Courault *et al.* [2008] separates 12 classes. It has been simplified to get a seven class land cover map by combining some classes with similar soil and vegetation properties (i.e., forest and orchards, salt marches and wetlands, etc.). The

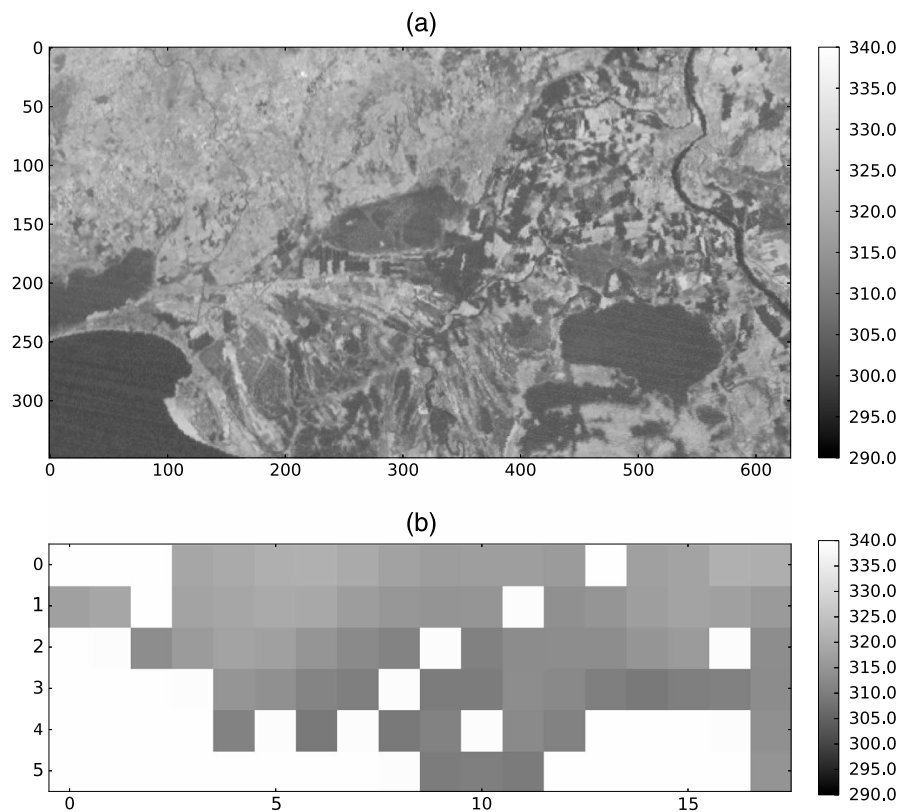
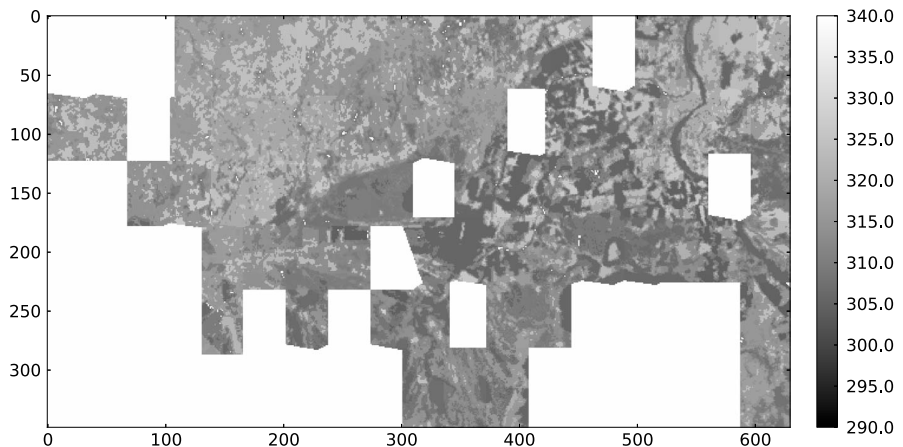


Figure 6. (a) ASTER and (b) SEVIRI LST images over the Crau-Camargue area at 10:47 A.M. and 10:45 A.M., respectively.

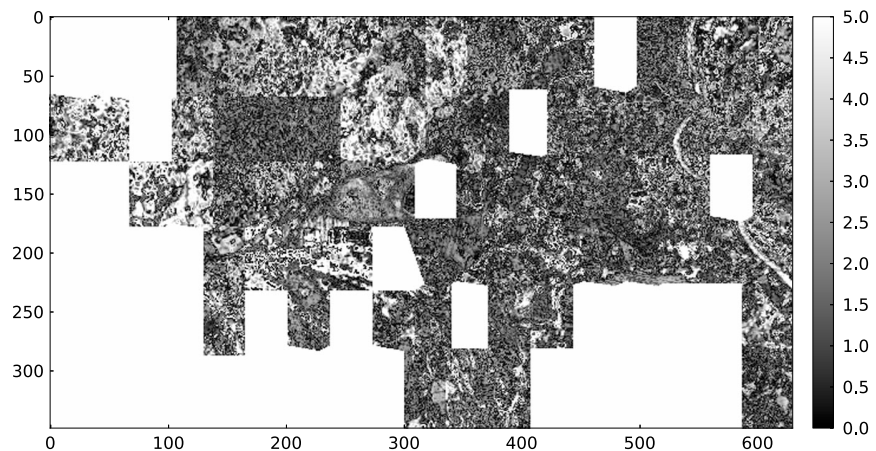
**Table 5.** SEtHyS Parameters Used for the Calibration: Actual Data Case

Land Cover Class	Selected Parameters and Prior	Variation Range
Bare soil	$P_1^1$ : Factor controlling the volumetric heat soil capacity	[0.7, 1.7]
	$P_2^2$ : Bare soil albedo	[0.25, 0.35]
	$P_3^3$ : Mulch thickness	[0.0, 0.4] (m)
Prairie	$P_2^1$ : Factor controlling the volumetric heat soil capacity	[0.6, 1.6]
	$P_2^2$ : Mulch thickness	[0.0, 0.6] (m)
	$P_3^3$ : Green vegetation albedo	[0.10, 0.26]
Wheat	$P_3^1$ : Factor controlling the volumetric heat soil capacity	[0.75, 1.75]
	$P_3^2$ : Bare soil albedo	[0.2, 0.3]
	$P_3^3$ : Mulch thickness	[0.2, 0.6] (m)
Rice	$P_4^1$ : Factor controlling the volumetric heat soil capacity	[1.75, 2.75]
	$P_4^2$ : Bare soil emissivity	[0.93, 0.97]
	$P_4^3$ : Mulch thickness	[0.0, 0.3] (m)
Orchards	$P_5^1$ : Factor controlling the volumetric heat soil capacity	[1.53, 2.53]
	$P_5^2$ : Mulch thickness	[0.0, 0.4] (m)
	$P_5^3$ : Tree albedo	[0.15, 0.33]
Water	$P_6^1$ : Factor controlling the volumetric heat soil capacity	[0.75, 1.75]
	$P_6^2$ : Bare soil emissivity	[0.93, 0.97]
	$P_6^3$ : Bare soil albedo	[0.25, 0.35]
Salt marches	$P_7^1$ : Factor controlling the volumetric heat soil capacity	[1.7, 2.7]
	$P_7^2$ : Bare soil emissivity	[0.93, 0.97]
	$P_7^3$ : Bare soil albedo	[0.25, 0.35]

final classes distinguish water, bare soil, wheat, prairie, rice, orchards, and salt marches. Figure 5 presents the new 90 m × 90 m land cover map. CSR LST time series from SEVIRI were also available for the same date over the study area with spatial and temporal resolutions equal to 3 km × 5 km and 15 min, respectively. The SEVIRI LST time series will be used as CSR observations and the ASTER LST as “reference” (“ground truth”) to validate the GPF approach. Since ASTER and SEVIRI sensors have different Earth projection systems (cylindrical coordinate system universal transverse Mercator for ASTER and nominal geostationary projection for SEVIRI), image coregistration was performed, the ASTER image taken as reference. The two LST products have also been intercalibrated using the Random Sample Consensus (RANSAC) process (more details on this process are available in *Kallel et al.* [2013] and *Fischler and Bolles* [1981]). After RANSAC calibration, several SEVIRI pixels, which are not aligned with the best linear regression line (calculated in the RANSAC process), have been removed. Figure 6 presents both ASTER and recalibrated SEVIRI images acquired at about the same time (2 min delay between ASTER and SEVIRI). The comparison of the aggregated ASTER image and the calibrated SEVIRI one showed no bias and better correlation (before correction, bias = −3.02 K, after correction, bias = 0 K,  $r^2 = 0.86$ ) and the RMS difference (RMSD) between both images is equal to 1.56 K. This value of RMSD will be used as the SD of the observation error. The ensemble size was set to  $N = 200$ . New



**Figure 7.** The 90 m × 90 m downscaled SEVIRI-MSG LST map.



**Figure 8.** Temperature differences between ASTER and downscaled MSG-SEVIRI LST map.

parameters ranges have been defined for the different classes ensuing from a variance-based sensitivity analysis [Marzban, 2013] performed for the seven classes and using the actual forcing data of the Crau-Camargue region. The micrometeorological and vegetation forcing data have been provided for the different classes according to ground truth measurements. A full description of the selected parameters for the different classes and their boundary values is given in Table 5. Three parameters per class were selected, and their respective variation intervals have been adjusted for the different land cover types according to prior knowledge of the soil characteristics of the case study region. The same variation ranges have been considered for the different classes (mean value  $\pm 50\%$ ). The albedo mean values of the different end-members have been adjusted according to FORMOSAT albedo products available on the region for the same year [Courault et al., 2008]. The LAI values have been estimated from 90 m resolution NDVI ASTER data for the same day (26 July 2006) following Bsaibes et al. [2009]. Prior end-member temperatures (SEtHyS subtemperatures) have been estimated from 1000 parameter sets. The GPF-downscaling method was applied independently over the different CSR SEVIRI pixels remaining after RANSAC calibration. The assimilation window corresponds to DOY 207 (26 July 2006). Only observations between 06:00 A.M. and 12:00 A.M. were considered because the SEVIRI time series were affected by clouds after 12 A.M. Figures 7 and 8 present, respectively, the downscaled SEVIRI LST image and the absolute difference (AD) map between ASTER LST image and the resulting downscaled image. The results show that more than 50% of the pixels present AD values less than 2 K (dominant dark gray color). The global RMSE, MAE, and Bias between the downscaled SEVIRI image and the ASTER image as well as between the SEtHyS prior estimate image and the ASTER image are presented in Table 6. The results show that the prior RMSE, calculated over the remaining CSR SEVIRI pixels, is greater than the GPF one (3 K compared to 2.4 K). The GPF correction may appear not very important compared to other works using the same actual data: in Kallel et al. [2013], the reduction is about 1.5 K but this result is still promising because the GPF permitted to reduce the prior error with only one run over the DOY 207 assimilation window. We should also note the interesting reduction of the MAE (0.4 K) and of the bias (1.5 K). Table 7 presents for each, land cover type the mean temperatures and their SD for ASTER, SEtHyS, and GPF high-resolution images as well as the RMSE, the MAE, and the bias of the downscaled images compared to the ASTER reference. The results show that for all the classes (except bare soil and wheat), the GPF RMSE are less or equal to the prior ones. The best results are obtained for the prairie and rice classes for which all the statistics were improved. The RMSE have been strongly reduced (from 3.6 K to 1.3 K for prairie and from 6.1 K to 1.8 K for rice) as well as the MAE (reduction of 2.4 K for prairie and of 4.4 K for rice) and the biases (reduction of 3.4 K for prairie and of 5.2 K for rice). The statistics have been also slightly improved for salt marches and orchards and remain unchanged for the water class. On the opposite, the results are worse for the bare soil and the wheat classes, whatever the statistical indices. The posterior RMSE, MAE, and biases are larger than the prior ones. This is explained considering that these two classes

**Table 6.** Global RMSE, MAE, and Bias for the GPF Downscaled SEVIRI and SEtHyS (Prior) Estimated Image Compared to ASTER Image

	SEtHyS Results	GPF Results
RMSE (K)	3.0	2.4
MAE (K)	2.3	1.9
Bias (K)	1.9	-0.5

to 1.8 K for rice) as well as the MAE (reduction of 2.4 K for prairie and of 4.4 K for rice) and the biases (reduction of 3.4 K for prairie and of 5.2 K for rice). The statistics have been also slightly improved for salt marches and orchards and remain unchanged for the water class. On the opposite, the results are worse for the bare soil and the wheat classes, whatever the statistical indices. The posterior RMSE, MAE, and biases are larger than the prior ones. This is explained considering that these two classes

**Table 7.** Reference Data and Downscaling Results for the Different End-Members

	Water	Bare Soil	Prairie	Wheat	Rice	Orchards	Salt Marches
ASTER LST (°C)	30.5 ± 2.6	50.2 ± 2.8	36.7 ± 0.4	45.5 ± 2.3	32.7 ± 1.8	40.9 ± 1.5	36.2 ± 2.8
SEtHyS LST (°C)	33.3 ± 0.5	50.5 ± 2.8	40.2 ± 2.1	47.0 ± 2.6	38.6 ± 1.0	42.2 ± 1.1	37.5 ± 1.2
GPF LST (°C)	33.3 ± 0.2	51.3 ± 1.5	36.6 ± 1.2	43.7 ± 1.9	34.1 ± 1.9	40.1 ± 1.1	37.2 ± 0.6
SEtHyS RMSE (K)	2.2	1.7	3.6	2.1	6.1	1.8	2.1
GPF RMSE (K)	2.2	2.4	1.3	3.1	1.8	1.8	1.8
SEtHyS MAE (K)	1.8	1.4	3.5	1.7	5.9	1.4	1.8
GPF MAE (K)	1.8	2.0	1.1	2.6	1.5	1.5	1.5
SEtHyS bias (K)	1.7	0.2	3.5	1.4	5.9	1.3	1.3
GPF bias (K)	1.8	1.0	-0.1	-1.9	0.7	-0.9	0.5

were the one presenting the lowest prior errors, comparable to the observation error (prescribed to 1.56 K as already noted). In this case, a single analysis step is not enough to constrain the model parameters and more particles are needed to better sample the parameter space in order to obtain errors less or equal than the prior errors. However, these preliminary results are quite satisfactory and demonstrate the potentialities of the GPF approach for downscaling surface temperatures. Compared to current semiempirical approaches, the GPF methodology requires supplementary meteorological data to force the LSM and larger computing resources. However, GPF-downscaling approach allows to monitor dynamically the subtemperatures. Concerning computing requirements, the downscaling of a MSG pixel containing seven classes requires 3.51 s of processing time on personal computer. For our MSG image of 68 pixels (remaining after RANSAC from an initial 108 pixel image), 1 min and 2 s is the processing time needed for a sequential implementation of the downscaling algorithm (a loop over the different pixels of the MSG image). A parallel implementation of the GPF-downscaling algorithm could be easily realized and will reduce significantly the processing time. Such requirements allow to envisage operational applications at regional scale.

## 5. Conclusions

This study investigated the downscaling of land surface temperature through a data assimilation system based on particle filtering. The method uses a dynamic LSM (SEtHyS) able to simulate prior estimations of the temperature time evolution and the spatial variability inside the observed pixel. The subpixel variability accounts for land cover, soil, and atmospheric forcing characteristics, and the pixel is represented by fractions of different vegetation classes. The GPF assimilation technique consists of generating an ensemble of candidate solutions for the subpixel temperatures and selecting the ones which minimize the discrepancy between prior aggregated temperatures and the observations at pixel scale and over an assimilation period. The mean of the selected temperatures at the end of the assimilation period is kept as the solution of the downscaled temperatures. In our study, a synthetic pixel composed of four different classes was studied and the assimilation of pseudo-observations was performed on a 124 day period. The particle resampling and selection processes were performed using a daily time window. The results show that the GPF is suitable for LST downscaling because of its easy implementation and its adaptive capability. Indeed, the nonlinearity of land surface processes and the rapid changes of boundary conditions, like precipitation or irrigation events and various agricultural practices, can lead to rapid changes in surface states and parameter values. The particle resampling step allows a fuller exploration of the parameter space and to find the optimal temperatures far from the local prior values. Therefore, crucial steps like rainfall events or vegetation phenological changes can be overcome without getting trapped in local minima. In our synthetic experiments, the LSM was implemented on four classes of vegetation equally represented, forced with their respective atmospheric and surface conditions provided at smaller scale than the observations. These conditions were prescribed and assumed perfectly known, but uncertainties in these data could be introduced further.

In our synthetic experiments, we have tested the assimilation of observations as they could be provided by various space instruments onboard polar or geostationary platforms. As a first step, the observation error statistics were assumed to be unbiased, Gaussian and additive (no time or spatial correlations). The results show that the assimilation improved the estimation of the temperature of the four different land cover classes in all the cases and that the improvement varies with the observation time and with the boundary conditions. The influence of the assimilation is minor when background model errors are the smallest, i.e., when the prior subtemperatures are close to the air temperature (assumed certain), in windy or in nonlimited soil moisture conditions. In such cases, the assimilation of noisy observations could even degrade the

downscaling results. The contribution of nighttime observations, in our synthetic case, appears also not significant, the observations around noon being the most valuable. This is also explained by the lower first guess errors at night compared to daily values. The assimilation methodology was also tested at larger scale on actual data acquired in the framework of the Crau-Camargue experimental database which collects thermal infrared data acquired at different scales. The work was performed for 1 day for which SEVIRI LST time series at 15 min frequency and one ASTER image were available. The results show that the subpixel temperatures may be estimated with a root-mean-square error lower than 2.4 K after only a single assimilation period of half a day, compared to prior errors reaching for some classes 6 K. These preliminary results prove that GPF is a promising approach for land surface temperature downscaling even if more work is required to better account for land cover map uncertainties as well as observations and model error correlations. The assimilation of multiple instrument (and consequently, multiresolution) data could be also a promising research axis to optimally exploit the irregular observations provided by the existing instruments.

#### Acknowledgments

This research is supported by CNES (Centre National d'Etudes Spatiales) through the TOSCA program and by the French National INSU-PNTS program. The French Defense Minister (DGA) is also acknowledged for supporting part of R. Mechri's PhD grant. The authors also thank D. Courault for having made available the Crau-Camargue database and her precious advice in the data processing (jointly with F. Maignan), C. Snyder and D. Carrer for interesting discussions, as well as P. Guillevic, N. MacBean, P. Maugis, and the anonymous reviewers for their careful reading of the manuscript and their helpful comments to improve the paper.

#### References

- Agam, N., W. P. Kustas, M. C. Anderson, F. Li, and P. D. Colaizzi (2007a), Utility of thermal sharpening over Texas high plains irrigated agricultural fields, *J. Geophys. Res.*, *112*, D19110, doi:10.1029/2007JD008407.
- Agam, N., W. P. Kustas, M. C. Anderson, F. Li, and C. M. Neale (2007b), A vegetation index based technique for spatial sharpening of thermal imagery, *Remote Sens. Environ.*, *107*(4), 545–558, doi:10.1016/j.rse.2006.10.006.
- Anderson, J. L. (1996), A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *J. Clim.*, *9*(7), 1518–1530, doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.
- Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Trans. Signal Process.*, *50*(2), 174–188, doi:10.1109/78.978374.
- Bergman, N. (1999), Recursive Bayesian estimation: Navigation and tracking applications, PhD dissertation, Linköping Univ., Linköping, Sweden.
- Bsaibes, A., et al. (2009), Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring, *Remote Sens. Environ.*, *113*(4), 716–729, doi:10.1016/j.rse.2008.11.014.
- Coudert, B. (2006), Apport des mesures de température de surface par télédétection infrarouge thermisue pour la modélisation des échanges d'énergie et d'eau à l'interface sol végétation atmosphère, PhD thesis.
- Coudert, B., C. Ottlé, and X. Briottet (2008), Monitoring land surface processes with thermal infrared data: Calibration of SVAT parameters based on the optimisation of diurnal surface temperature cycling features, *Remote Sens. Environ.*, *112*(3), 872–887, doi:10.1016/j.rse.2007.06.024.
- Courault, D., A. Bsaibes, E. Kpemlie, R. Hadria, O. Hagolle, O. Marloie, J.-F. Hanocq, A. Olioso, N. Bertrand, and V. Desfonds (2008), Assessing the potentialities of FORMOSAT-2 data for water and crop monitoring at small regional scale in South-Eastern France, *Sensors*, *8*(5), 3460–3481, doi:10.3390/s8053460.
- Del Moral, P. (2004), *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems With Applications*, Probability and Its Applications, Springer, New York.
- Del Moral, P., A. Doucet, and S. Singh (2010), Forward smoothing using sequential Monte Carlo, *Tech. Rep.*
- Dominguez, A., J. Kleissl, J. C. Luvall, and D. L. Rickman (2011), High-resolution urban thermal sharpener (HUTS), *Remote Sens. Environ.*, *115*(7), 1772–1780, doi:10.1016/j.rse.2011.03.008.
- Douc, R., and O. Cappe (2005), Comparison of resampling schemes for particle filtering, *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pp. 64–69, Zagreb, Croatia, doi:10.1109/ISPA.2005.195385.
- Doucet, A., S. Godsill, and C. Andrieu (2000), On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.*, *10*(3), 197–208, doi:10.1023/A:1008935410038.
- Douville, H., B. Decharme, A. Ribes, R. Alkama, and J. Sheffield (2012), Anthropogenic influence on multi-decadal changes in reconstructed global evapotranspiration, *Nat. Clim. Change*, *3*, 59–62, doi:10.1038/NCLIMATE1632.
- Fischler, M. A., and R. C. Bolles (1981), Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, *24*, 381–395, doi:10.1145/358669.358692.
- Frey, C. H., and S. R. Patil (2002), Identification and review of sensitivity analysis methods, *Risk Anal.*, *22*(3), 553–578, doi:10.1111/0272-4332.00039.
- Gao, F., W. Kustas, and M. Anderson (2012), A data mining approach for sharpening thermal satellite imagery over land, *Remote Sens.*, *4*(12), 3287–3319, doi:10.3390/rs4113287.
- Guillevic, P. C., J. L. Privette, B. Coudert, M. A. Palecki, J. Demarty, C. Ottlé, and J. A. Augustine (2012), Land surface temperature product validation using NOAA's surface climate observation networks—Scaling methodology for the Visible Infrared Imager Radiometer Suite (VIIRS), *Remote Sens. Environ.*, *124*, 282–298, doi:10.1016/j.rse.2012.05.004.
- Hamill, T. M. (1997), Reliability diagrams for multicategory probabilistic forecasts, *Weather Forecast.*, *12*(4), 736–741, doi:10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2.
- Hamill, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, *129*(3), 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.
- Inamdar, A. K., and A. French (2009), Disaggregation of GOES land surface temperatures using surface emissivity, *Geophys. Res. Lett.*, *36*, L02408, doi:10.1029/2008GL036544.
- Inamdar, A. K., A. French, S. Hook, G. Vaughan, and W. Lueckert (2008), Land surface temperature retrieval at high spatial and temporal resolutions over the southwestern United States, *J. Geophys. Res.*, *113*, D07107, doi:10.1029/2007JD009048.
- Jazwinski, A. (1970), *Stochastic Processes and Filtering Theory*, Mathematics in Science and Engineering, Elsevier Science, London, U. K.
- Jeganathan, C., N. Hamm, S. Mukherjee, P. Atkinson, P. Raju, and V. Dadhwal (2011), Evaluating a thermal image sharpening model over a mixed agricultural landscape in India, *Int. J. Appl. Earth Obs. Geoinf.*, *13*(2), 178–191, doi:10.1016/j.jag.2010.11.001.
- Kallel, A., C. Ottlé, S. Le Hégarat-Masclé, F. Maignan, and D. Courault (2013), Surface temperature downscaling from multiresolution instruments based on Markov models, *IEEE Trans. Geosci. Remote Sens.*, *51*(3), 1588–1612, doi:10.1109/TGRS.2012.2207461.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *82*(1), 35–45, doi:10.1115/1.3662552.

- Kustas, W. P., J. M. Norman, M. C. Anderson, and A. N. French (2003), Estimating sub-pixel surface temperatures and energy fluxes from the vegetation index radiometric temperature relationship, *Remote Sens. Environ.*, *85*(4), 429–440, doi:10.1016/S0034-4257(03)00036-1.
- Kwok, N., and W. Zhou (2005), Evolutionary particle filter: re-sampling from the genetic algorithm perspective, in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2935–2940, IEEE, Edmonton Alberta, Canada, doi:10.1109/IROS.2005.1545119, (to appear in print).
- Liu, D., and R. Pu (2008), Downscaling thermal infrared radiance for sub-pixel land surface temperature retrieval, *Sensors*, *8*(4), 2695–2706, doi:10.3390/s8042695.
- Marzban, C. (2013), Variance-based sensitivity analysis: An illustration on the Lorenz'63 model, *Mon. Weather Rev.*, *141*(11), 4069–4079, doi:10.1175/MWR-D-13-00032.1.
- Merlin, O., B. Duchemin, O. Hagolle, F. Jacob, B. Coudert, G. Chehbouni, G. Dedieu, J. Garatuza, and Y. Kerr (2010), Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images, *Remote Sens. Environ.*, *114*(11), 2500–2512, doi:10.1016/j.rse.2010.05.025.
- Merlin, O., F. Jacob, J.-P. Wigneron, J. Walker, and G. Chehbouni (2012), Multidimensional disaggregation of land surface temperature using high-resolution red, near-infrared, shortwave-infrared, and microwave-L bands, *IEEE Trans. Geosci. Remote Sens.*, *50*(5), 1864–1880, doi:10.1109/TGRS.2011.2169802.
- Olioso, A., et al. (2002), Monitoring energy and mass transfers during the Alpilles-ReSEDA experiment, *Agronomie*, *22*, 597–611.
- Ottlé, C., A. Kallel, G. Monteil, S. Le Hégarat-Masclé, and B. Coudert (2008), Subpixel temperature estimation from low resolution thermal infrared remote sensing, *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, *3*(1), 403–406, doi:10.1109/IGARSS.2008.4779369.
- Rémy, S., O. Pannekoucke, T. Bergot, and C. Baehr (2012), Adaptation of a particle filtering method for data assimilation in a 1D numerical model used for fog forecasting, *Q. J. R. Meteorolog. Soc.*, *138*(663), 536–551, doi:10.1002/qj.915.
- Snyder, C. (2011), Particle filters, the optimal proposal and high-dimensional systems, *ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, pp. 6–9, Reading, U. K.
- Talagrand, O., R. Vautard, and B. Strauss (1997), Evaluation of probabilistic prediction systems, in *Proceedings, ECMWF Workshop on Predictability*, pp. 1–25, ECMWF, Reading, U. K.
- Uosaki, K., Y. Kimura, and T. Hatanaka (2004), Evolution strategies based particle filters for state and parameter estimation on non-linear models, in *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, pp. 884–890, IEEE, doi:10.1109/CEC.2004.1330954.
- Van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: An extremely efficient particle filter, *Q. J. R. Meteorolog. Soc.*, *136*(653), 1991–1999, doi:10.1002/qj.699.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., 627 pp., Academic Press, Oxford, U. K.
- Wilks, D. S. (2011), On the reliability of the rank histogram, *Mon. Weather Rev.*, *139*(1), 311–316, doi:10.1175/2010MWR3446.1.
- Zhan, W., Y. Chen, J. Zhou, J. Wang, W. Liu, J. Voogt, X. Zhu, J. Quan, and J. Li (2013), Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats, *Remote Sens. Environ.*, *131*(19), 119–139, doi:10.1016/j.rse.2012.12.014.

## 5.4 Résultats et conclusions

La précédente étude a permis d'évaluer les performances du lisseur particulaire pour deux expériences dont la première traite un cas simple de désagrégation sur des données synthétiques et la deuxième traite un cas réaliste sur des données réelles de télédétection acquise sur la région de Crau-Camargue dans le Sud-Est de la France. Dans chacune de ces expériences, les performances de notre approche ont été évaluées par rapport aux performances du modèle sans assimilation, sur une fenêtre d'assimilation journalière et pour une période totale d'étude de 21 jours.

La première étude synthétique traite le cas d'un pixel composé équitablement de quatre classes agricoles caractéristiques de la région d'étude : le sol nu, le blé, le riz et la prairie irriguée. Pour ces quatre classes, deux paramètres sensibles ont été choisis pour définir la particule (les deux paramètres les plus sensibles parmi l'ensemble des paramètres sélectionnés à l'issue de l'analyse de sensibilité).

Plusieurs cas ont été traités afin d'évaluer l'impact de l'échantillonnage des observations sur la fenêtre d'assimilation journalière, la période d'observation, et l'écart-type de l'erreur d'observation sur les performances du lisseur particulaire quant à la réduction des erreurs d'estimation. Cet apport a été quantifié à travers le calcul de l'indice d'efficacité  $\mathcal{I}$  qui est d'autant plus important que la réduction de l'erreur de prédiction est grande (rapport entre l'erreur *a posteriori* et l'erreur *a priori*).

Pour tous les cas traités dans le cadre de l'expérience synthétique, les indices d'efficacité étaient supérieurs à 18%. Ce résultat démontre l'apport de contraindre les simulations du modèle avec des observations HRT/BRS. Les meilleurs indices ont été obtenus pour l'estimation des températures des classes les plus sèches quant aux conditions hydriques. Ceci s'explique par le fait que les classes qui présentent un état hydrique proche de la saturation (comme le riz ou la prairie irriguée) montrent des sensibilités plus faibles de la température par rapport aux paramètres réduisant ainsi la variance de l'erreur *a priori* (cette erreur étant calculée par rapport aux températures de ces classes pour une combinaison de référence des paramètres sensibles). Dans ces conditions, l'assimilation d'observations bruitées peut même dégrader les performances du lisseur particulaire.

L'étude de sensibilité des performances du lisseur à l'amplitude de l'erreur d'observation a montré que ces dernières se dégradent en fonction de la valeur de  $\sigma_o$ . Ce résultat était attendu, car en augmentant la valeur de  $\sigma_o$  on agit sur la sélectivité du lisseur. En effet, le choix des particules est guidé par le calcul de la distance entre les observations et les particules et ceci en fonction de l'erreur d'observation. En effet dans notre cas la matrice  $R$  de l'équation n'est autre que  $\sigma_o$ . D'après l'équation, la distance entre les observations et les simulations ( $y - Hx$ ) est divisée par  $\sigma_o$ . Etant donné que les poids sont calculés en fonction de ce rapport et sont d'autant plus importants que ce rapport est faible, plus la valeur de  $\sigma_o$  est importante, plus le poids est important. Ceci augmente donc la chance qu'une particule soit retenue réduisant par conséquent l'apport des observations pour estimer les "meilleures" températures sous-pixelliques. Si de plus la variance de l'erreur d'observation devient plus importante que celle de l'erreur *a priori*, le lisseur ne fait presque plus de sélection et l'erreur modèle n'est donc plus réduite. Ce cas est rencontré pour les classes dont les erreurs modèles sont les plus petites (exemple :  $\mathcal{I}_{prairie} = [33, 30, 26] \%$  pour  $\sigma_o = [0.5, 2.0, 4.0]$  ).

Afin d'évaluer l'impact de la période d'observation sur la fenêtre d'assimilation journalière ainsi que l'impact du nombre d'observations, quatre scénarios ont été étudiés. Le premier représente le cas standard où l'on assimile toutes les observations disponibles sur la fenêtre d'assimilation (soit 72 observations pour un échantillonnage temporel égal à 20min). Les deuxième et troisième scénarios, traitent respectivement les cas où l'on assimile les observations comprises entre 10h et 18h (25 observations) et entre 10h et 14h (13 observations). Le dernier scénario traite le cas où une seule observation est assimilée sur la journée.

Trois principaux résultats ont été retenus à l'issue de cette expérience :

- le premier résultat est que plus le nombre d'observations est important plus les indices d'efficacité le sont aussi. Ce résultat était attendu car plus on a d'observations, plus on contraint le modèle et la sélection des températures sous-pixelliques est par conséquent meilleure (une réduction de 2% à 6% entre les indices obtenus pour 72 observations et ceux obtenus pour 13 observations).
- le deuxième résultat est que les observations les plus pertinentes quant à la contribution à la réduction de l'erreur *a priori* sont celles acquises durant la journée. En effet la comparaison de la réduction de l'indice d'efficacité obtenue entre les scénarios 1, 2 et 3 est relativement faible ( $\leq 6\%$ ), ce qui montre que l'assimilation des observations nocturnes n'est pas d'un grand apport sur la réduction des erreurs *a priori*. Ceci s'explique en grande partie par le fait que les températures simulées par le modèle le soir ont des valeurs très proches de la température de l'air, réduisant ainsi la variance de l'erreur modèle pour les heures en dehors de la journée. Dans ce cas, l'assimilation des observations nocturnes n'améliore pas l'estimation des températures sous-pixelliques.
- Le dernier scénario montre que même avec une seule observation à 12h les performances du lisseur surpassent celles du modèle et les indices d'efficacité sont  $\geq 20\%$  pour toutes les classes. Une exploration plus fine de ce dernier scénario a été faite en échantillonnant l'intervalle d'heures d'observation [6h, 18h] avec un pas de 2h. Cette expérience a montré que les meilleurs indices sont obtenus pour les heures d'observation entre 10h et 14h.

Suite à cette étude synthétique, on a appliqué notre méthode de désagrégation sur des données réelles en utilisant une série temporelle Météosat de résolution spatiale (3km  $\times$  5km) et de résolution temporelle 15min sur la région de la Crau-Camargue. La carte d'occupation des sols ainsi qu'une image ASTER pour la validation étaient disponibles sur la région d'étude avec une résolution spatiale de 90m. L'image ASTER a été acquise le 26/07/2006 à 10h47 d'où la validation était limitée à cette date et a pu être comparée avec la carte de températures HRS estimée à 10h45 par le modèle. Pour cette expérience, la particule a été définie en choisissant 3 paramètres par classe et la fenêtre d'observation a été définie entre 6h et 12h du matin, compte-tenu des résultats de l'étude synthétique. Un inter-étalonnage par la méthode RANSAC, entre les données d'observations et les données de validation a permis de réduire le biais à  $-1.1K$  (initialement égal à  $3K$  tel que les températures ASTER sont plus chaudes que Météosat à la résolution spatiale de Météosat). L'amplitude de l'erreur d'observation a été estimée en calculant l'erreur quadratique moyenne entre les données d'observation corrigées par RANSAC et les données ASTER agrégées à la résolution spatiale de Météosat. Celle-ci a été estimée à  $1.56K$ . La carte d'occupation des sols a été modifiée pour définir 7 classes de type agricole comme décrit dans la section 4.2 du chapitre précédent. Pour ces sept classes, on a appliqué le lisseur particulière pour une seule fenêtre d'assimilation de 24h soit le 26/07/2006. Les résultats obtenus ont montré que l'assimilation des observations a amélioré l'estimation des températures sous-pixelliques pour la plupart des classes et avec une incertitude sur toute l'image inférieure à  $2.4K$ .



Ces résultats ( $RMSE \leq 2.4K$ ) sont comparables aux méthodes classiques (ex : [In-  
amdar and French \(2009\)](#) ( $\approx 2K$ ), [Merlin et al. \(2010\)](#) ( $\approx 3K$ )) ou aux inversions  
statistiques réalisées par ([Kallel et al. \(2013\)](#) ( $\approx 2K$ )). Les incertitudes obtenues sur  
l'estimation des températures des classes varient entre  $[1.3, 3.1]$  K.

Finalement, les résultats obtenus suite à l'application de notre approche de désa-  
grégation sur des données réelles ont montré des résultats très encourageants quant  
à la réduction de l'erreur modèle qui, pour certaines classes, étaient  $\geq 6K$ . Cette  
étude préliminaire a été poursuivie par la suite, en travaillant sur la base de données  
2009 présentée dans le chapitre 4 qui a permis de tester la méthode sur des périodes  
d'assimilation plus longues, des fenêtres d'assimilation plus larges et en considérant  
une région plus grande avec d'autres types de couvert qui ne sont pas nécessairement  
agricoles (de la forêt par exemple).

---

# Désagrégation de séries temporelles Météosat à plus grande échelle

## 6.1 Introduction

Ce chapitre, présenté sous forme d'un article, décrit l'application du lisseur particulier pour la désagrégation des séries temporelles Météosat acquises sur une période plus longue que celle étudiée dans le chapitre précédent et sur une région d'étude plus large.

Cet article vise à évaluer les limites de l'approche de désagrégation des températures Météosat. Pour cela, une nouvelle carte d'occupation des sols à 30m de résolution (produit Landsat7) nous a été fournie sur la région d'étude pour l'année 2009. La zone décrite par cette carte couvre une surface d'environ 4539km<sup>2</sup> et contient à la fois des couverts agricoles (du blé, du sol nu, de la prairie irriguée, des rizières et des cultures variées) et des couverts de type naturel en particulier de la forêt (voir Figure.4.5 du chapitre 4). Il faut noter que concernant l'occupation des sols, la zone décrite est couverte à plus de 50% par de la forêt. Les données de forçage ainsi que les caractéristiques de la végétation ont été traitées et sont disponibles pour les sept classes décrites par la nouvelle carte d'occupation des sols et dont les détails sont présentés dans le chapitre 4. La validation de l'approche a été réalisée de façon qualitative avec des données Landsat7 à 30m et de façon quantitative avec des données MODIS à 1K de résolution spatiale. Plus de détails sur l'inter-étalonnage des données Météosat avec les données de validation sont présentés dans la section 4.6.2.2 du chapitre 4.

La validation à l'échelle de la classe de végétation a été faite pour les sept classes majoritaires et pour les pixels MODIS les plus homogènes en ce qui concerne l'occupation des sols (pixels contenant plus que 55% d'une classe donnée). La fenêtre d'observation a été choisie de façon à ne considérer que les observations durant la journée compte-tenu des biais observés entre le modèle et les observations durant la nuit d'une part et de la faible sensibilité des paramètres du sol et de la végétation à la température pendant la nuit d'autre part (cette dernière étant fortement contrainte par la température de l'air). La définition de la particule a été faite en utilisant 5 paramètres sensibles pour chacune des classes décrites par la carte d'occupation des sols. La disponibilité des données MODIS et Météosat pour l'année 2009 a permis de tester une période d'assimilation plus longue que celle utilisée lors de l'étude réalisée avec les données de 2006 (sept mois de données Météosat et 29 images MODIS pour la validation). Ceci a permis de tester les performances du filtre non seulement pour

une fenêtre d'assimilation journalière mais aussi pour des fenêtres d'assimilation plus longues (trois jours et sept jours). L'impact des incertitudes sur les données de forçage a aussi été évalué en rajoutant des incertitudes (1K) sur la variable de forçage atmosphérique "température de l'air" dans le cas où l'on assimile les observations sur une fenêtre journalière. Pour toute les expériences menées, les performances du lisseur particulaire ont été comparées aux performances du modèle sans assimilation et la validation a été faite à l'échelle de MODIS.

L'article se divise en quatre parties. La première partie présente l'approche de désagrégation par lissage particulaire proposée dans cette thèse. La deuxième partie présente les données d'observation et de validation à haute et moyenne résolution spatiale ainsi que les différents processus d'inter-étalonnage mis en place. La troisième partie concerne la validation des résultats de désagrégation des séries temporelles Météosat à la résolution des données MODIS. Elle s'articule autour de deux expériences. La première expérience traite le cas d'une fenêtre d'assimilation de données journalière pour laquelle les performances du lisseur particulaire ont été évaluée face à des incertitudes sur les données de forçages. La deuxième expérience traite l'impact de la variation de la taille de la fenêtre d'assimilation de données sur les performances de la méthode de désagrégation. La quatrième et dernière partie est consacrée à la validation qualitative des résultats de désagrégation à la résolution de la carte d'occupation des sols par rapport aux données Landsat7.

## 6.2 Article 2 (version *draft*) : Downscaling of Meteosat Land Surface Temperature over an heterogeneous landscape using a Data Assimilation approach

### Abstract

Land surface temperature (LST) is a key variable for a wide range of environmental applications including urban surface temperature mapping, surface heat island monitoring, vegetation state monitoring, surface energy budgets derivation or evapotranspiration mapping. Nevertheless, reliable monitoring of such environmental applications require LST measurements at frequent intervals and fine spatial resolutions, a difficult task because of tradeoffs among these resolutions. To overcome the technical shortcoming, well established LST downscaling methods have been developed to derive higher resolution LST from available data. Different downscaling algorithms have been so far proposed ranging from linear regression schemes to data assimilation (DA) based ones. In this research a DA based downscaling approach, the Genetic Particle Smoother (GPS) downscaling approach, has been applied to disaggregate Meteosat8 LST ( $3km \times 5km$ ) times series over the Crau-Camargue region in South-East France for 7 months in 2009 based on a land cover map (30m) scale derived from Landsat 7 data. The validation of the downscaled LSTs is based on the moderate resolution LST product from MODIS Aqua/Terra platforms using a set of coincident MODIS clear sky LST images. The performances of the GPS downscaling approach have been assessed in terms of reduction of the root mean square error (RMSE), the mean absolute error (MAE) and the bias at the sub-pixel scale compared to prior (model simulated LSTs without DA), against MODIS data. The impact of the assimilation window size was analyzed. Results showed that GPS is a suitable tool for downscaling Meteosat LST products from  $3 \times 5km$  to 1km scale.

**Keywords :** Land Surface Temperature downscaling, data assimilation, Genetic Particle Smoother, Meteosat.

### 6.2.1 Introduction

Land surface temperature (LST) is a key variable for a wide range of environmental applications including urban surface temperature mapping (Dousset and Gourmelon (2003); Nichol (2009); Bechtel et al. (2012)), surface heat island monitoring (Voogt and Oke (2003); Weng et al. (2004); Stathopoulou and Cartalis (2009); Essa et al. (2012); Zakšek and Oštir (2012)), wildfire detection (Dennison et al. (2006); Dennison and Matheson (2011)) or volcanic eruptive activity monitoring (Pieri and Abrams (2005)) among others. More common applications of LST are monitoring water and energy budgets such as deriving surface energy budgets (Diak et al. (2004); Coudert et al. (2008)), assessing drought severity (Quattrochi and Luvall (2004)), monitoring vegetation physiology (Karnieli et al. (2006)), downscaling evapotranspiration monitoring (Norman et al. (2003); Agam et al. (2008); Kalma et al. (2008); Bindhu et al. (2013)) and model parameter calibration (Coudert (2006); Coudert and Ottlé (2007)). Nevertheless, reliable monitoring of such environmental applications requires measurements of LST at frequent intervals and fine spatial resolutions to capture LST high spatio-temporal variability. Despite the advances of remote sensing, measuring LSTs at high spatial and temporal resolution is not yet possible because of tradeoffs among these resolutions. While high spatial resolution thermal data have limited temporal availability (eg : temporal resolution of 16 days in clear sky conditions for Landsat 7 insufficient for agronomic applications), coarser satellite thermal data are regularly

available (eg : 15 minutes temporal resolution for Meteosat 8). Furthermore, in thermal bands, coarse resolution pixels are often subject to mixture effects which could be presented as a blending of multiple temperature components within the large thermal pixel (Zhan et al. (2013)). This mixture effect is relatively more prevalent in spatially heterogeneous environments such as agricultural (Agam et al. (2007a); Kallel et al. (2013)) and urban areas (Zakšek and Oštir (2012)). The main challenge is then to develop methods to assess LST at higher spatial and temporal resolutions.

A well-established method to derive higher resolution LST from other available data is the disaggregation/downscaling of LST (DLST) (Zhan et al. (2013)). The basic tenet of DLST is to produce finer resolution sub-pixel temperatures from the coarser resolution thermal data through sharpening or unmixing algorithms. Different approaches have been so far developed to downscale LST, ranging from linear regression schemes (Kustas et al. (2003); Agam et al. (2007b,a); Liu and Pu (2008)), to data assimilation (DA) based ones (Ottlé et al. (2008); Zhan et al. (2011); Montzka et al. (2012); Kallel et al. (2013); Mechri et al. (2014)). Linear regression based approaches are built on the inverse relationship between LST and the Normalized Difference Vegetation Index (NDVI) or the fractional vegetation cover ( $f_{cover}$ ) both increasing with the amount of biomass (Agam et al. (2007b,a); Jeganathan et al. (2011)). It should be highlighted that LST variation is more complicated than a linear regression interpolation and can also be subject to several other factors such as photosynthetic activity of vegetation (Merlin et al. (2010)), albedo, soil moisture and emissivity conditions (Ottlé and Stoll (1993); Inamdar and French (2009); Yuan and Bauer (2007); Merlin et al. (2012); Deng and Wu (2013)). The main drawback of these methods is their dependency to high resolution products availability for downscaling to perform the downscaling regression. For instance, to perform downscaling of LST over one whole day, high spatial resolution (HSR) NDVI maps must be provided for a sufficient time steps to describe the daily LST time variability. DA based techniques, give more flexibility in this point since the only high resolution needed product is the land cover map of the study area (Ottlé et al. (2008); Kallel et al. (2013); Mechri et al. (2014)). The main idea of DA approaches is to constrain a physical modeling of the sub-pixel temperatures with large scale observations based on a prior knowledge of vegetation class fractions inside the thermal pixel. In fact DA ensure the estimation of the “best” sub-pixel temperatures through the minimization of a cost function related to the discrepancy between coarse resolution observation and the aggregated modeled sub-pixel LSTs.

In this work, a DA based downscaling approach has been used to enhance the spatial resolution of Meteosat 8 LST time series over the Crau-Camargue region in South Eastern France. The DA approach is based on a genetic particle smoother (GPS) which has been implemented to SETHYS (Coudert (2006); Coudert and Ottlé (2007); Saux-Picart et al. (2009b,a)) land surface model (LSM). This work is a step further the previous one presented in Mechri et al. (2014), as it concerns a longer time period and a more expanded study area.

The paper is organized as follows : a brief description of the downscaling calibration method is provided in Section 2. Section 3 details the actual data settings and the performances of the methodology are discussed in Section 4 followed by the some conclusions and future perspectives of this study.

### 6.2.2 Data and methods

This study focuses on the application of the GPS downscaling approach over the Crau-Camargue region in the South Eastern part of France. The period under analysis spans from March 2009 to September 2009 and the observations assimilated are LST provided by the Satellite Application Facility for Land Surface Analysis (LSA-SAF),

<http://landsaf.meteo.pt/> (Trigo et al. (2011)), derived from Meteosat thermal infrared observations. The evaluation of the downscaling is performed quantitatively at 1km resolution using the Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra LST products and more qualitatively at 30m (spatial resolution of the LC map) using Landsat 7 thermal infrared data.

### 6.2.2.1 GPS downscaling approach

The GPS downscaling approach, is based on the assimilation of low spatial resolution (LSR) and high temporal resolution (HTR) temperature observations in a land surface model able to simulate the sub-pixel temperature variability. It is assumed that the spatial heterogeneity of LST is mainly related to land cover variations. The assimilation is then performed at the LSR pixel scale, each pixel assumed independent (i.e., no spatial nor temporal correlations are considered). Then, the downscaling problem comes down to the estimation of the respective temperatures of the main land cover classes present in the considered pixel, assuming land cover fractions stationary. The estimation of a state variable in DA requires the definition of two models : the dynamic model which describes the dynamics of the state variable in space/time and the observation model which relates the observations to the state variables. In our case, the dynamic model is SETHYS LSM implemented at the scale of the LSR observations (kilometric scale), forced with meteorological data, simulating the energy and water budgets (and therefore, the surface fluxes, temperatures and soil moisture) for the different land cover classes. The observation model is the aggregation model which allows to estimate the surface temperature at the LSR pixel scale. As we are dealing with an equifinality problem (different values of end-member temperatures can lead to the same LSR temperature after aggregation), ensemble DA presents a good framework since the sub-pixel LSTs are estimated from an ensemble of candidate solutions (Pan et al. (2009); Quaife et al. (2008)). For our purpose, we work with the particle filter (Arulampalam et al. (2002); Van Leeuwen (2010); Snyder (2011)), ensemble method successfully applied to state and parameter estimation and downscaling issues (Hill et al. (2011); Mechri et al. (2014)). The method consists in calibrating a set of model parameters to minimize the misfit with the observations. In our case, since the time constant of LST is short and the observations can be irregular (due to cloudiness), the misfit is calculated over a time window. Therefore, a smoothing implementation of the genetic particle filter has been proposed for that purpose (Mechri et al. (2014)). In general, particle filtering includes two main steps : the analysis step and the selection/re-sampling step. The first step consists in weighting each particle (ensemble member, realization of a state variable) according to its “closeness” to the available observations (the closer the particle, the higher the weight). We should note that under the hypothesis of independent observations and error statistics, if the misfit is calculated over a period of time (the assimilation window), the weight can be calculated as the product of all separate “sub-weights” relative to each observation (a normalization is performed so that the final weight remains in the  $[0, 1]$  interval). In this case, we talk about smoothing rather than filtering. The second step, which represents the heart of particle filtering, consists in rejecting low weighted particles and replacing them with copies of the kept ones (re-sampling). In our case, a particle represents a set of SETHYS parameters relative to all the actual end-members inside the LSR pixel (the most sensitive parameters toward LST output). In other words, if we have two land cover classes (end-members) inside a LSR pixel and if for each class, we have selected the most sensitive parameters toward the LST, the particle can be written as follows :

$$x = (P_i^k)_{i \in [1,2], k \in [1,3]} = (P_1^1, P_1^2, \dots, P_1^{r_1}; P_2^1, P_2^2, \dots, P_2^{r_2})$$

where  $x$  is the particle,  $P$  is a model parameter,  $k$  superscript is the selected parameter identifier,  $\{r_1, r_2\}$  are the number of sensitive parameters per class and  $i$  index is the class identifier. Once the particle is defined, the prior ensemble of particles can be generated after defining a probability distribution and a range of variation for each parameter. In our case, a uniform distribution and parameter intervals of variation are prescribed empirically according to previous works. Then, for each parameter set corresponding to the particles ensemble, SETHYS simulates prior estimations of the classes temperatures, given the atmospheric forcing data and vegetation characterization. The aggregation step (the observation model), is performed finally to obtain an ensemble of simulated LSTs at the observations spatial resolution. The role of the particle smoother is to select the simulations which best fit the observations at LSR, corresponding to the “best” estimations of sub-pixel temperatures and parameters set values. In this way, a twofold solution is reached : the calibration of SETHYS parameters for each land cover class, according to LSR available observations over the assimilation window, and the downscaling of the LSR LST time series ensured with the estimation of the different sub-pixel-temperatures. We should note that the aggregation step implies the availability of a HSR Land Cover (LC) map which provides the different land cover fractions inside the LSR thermal pixels. A thorough description of the GPS downscaling-calibration approach is presented in [Mechri et al. \(2014\)](#). As we can note, the particle definition, in other words the parameter selection, is very crucial since the downscaling solution relies on the sensitivity of these parameters to LST model output. To identify the most sensitive SETHYS parameters, a variance based sensitivity analysis (Sobol sensitivity analysis [Chan et al. \(1997\)](#); [Saltelli \(2002\)](#); [Sobol et al. \(2007\)](#); [Saltelli et al. \(2010\)](#)) has been performed for each land cover type of the study area. The results exhibited a maximum of 5 sensitive parameters for each of the considered land cover types as detailed in [Table 6.1](#). These parameters are used to define the particles in the downscaling procedure.

### 6.2.2.2 Study area and forcing data

The study area is located in the “Crau-Camargue” region in South-Eastern France ( $43.53^\circ N$ ;  $4.66^\circ E$ , see [Figure 6.1](#)). The climate is typically Mediterranean, with strong winds, and irregular rainfall (long dry periods in spring and summer). This region is characterized by a wide range of land cover types and important irrigation practices. Micro-meteorological forcing data, vegetation measurements (Leaf Area Index (LAI) and vegetation height) and soil moisture measurements necessary to initialize and force the SETHYS model, were provided from the “Crau-Camargue” data base over the year 2009 for several agricultural fields (the micro-meteorological stations are presented with red location markers in [Figure 6.1](#)). An image classification of Landsat 7 images was performed to map land cover. It allowed to separate seven land cover classes : bare soil, water, prairie, cultures, orchards/woods, wheat and rice. In all the following the orchards/wood class will be designed as “Tree cover” class. We should note that the classification suffers from anomalies especially for the tree cover class mixed with marches and rice lands. Unfortunately such uncertainties are not taken into account at this stage in our downscaling approach because we assume that classification is certain and stable all over the assimilation period. LAI and vegetation height were estimated for each of these classes from NDVI data derived from the Landsat 7 reflectances data (30m spatial resolution). An empirical exponential relationship proposed by [Baret and Guyot \(1991\)](#); [Wilson and Meyers \(2007\)](#) was calibrated by [Bsaibes et al. \(2009\)](#) to inverse LAI from NDVI measurements. Daily values were finally interpolated after a careful merge with the available irregular field observations. Seasonal variations of the vegetation height were also estimated from experimental data and phased with the

mean average seasonal variations of LAI per class.

### 6.2.2.3 Satellite data

**Meteosat8 data** The LSA-SAF provided the LST data to downscale. The LST product Trigo et al. (2011) was obtained with a generalized split-window algorithm (Freitas et al. (2010)) from top-of-atmosphere brightness temperatures measured by Meteosat8 in the thermal infrared, namely in Meteosat8 channels  $IR10.8 \mu m$  and  $IR12.0 \mu m$ . The LSA-SAF LST is produced at full Meteosat8 spatial and temporal resolutions, with a 15 minutes sampling interval and a spatial resolution of about  $3km \times 5km$  ( $3km \times 3km$  at the sub-satellite point). The product is provided over land within the Meteosat8 disk under clear sky conditions along with an error estimation. An automatic Quality Control (QC) is performed on LST data, and the quality information is provided on a pixel-by-pixel basis. The LST confidence level was defined based on the following parameters : viewing angle ; atmospheric characteristics (i.e. surface temperature and column water vapour) ; land surface emissivity (EM) confidence level. The three considered levels of confidence (above nominal, nominal and below nominal) correspond to estimated uncertainties of LST values (respectively less than 1K, between 1 and 2K and above 2K) and with a threshold of 4K. More details are provided in the product user manual available on <http://landsaf.meteo.pt/algorithms.jsp?seltab=0&starttab=0>. This error data-set was used to generate the observation error statistics required by the GPS. The LST time series spans from March to October 2009. For simplicity we will adopt Meteosat as the name for Meteosat8 data in all the following.

**Landsat 7 LST data** The Enhanced Thematic Mapper Plus (ETM+) on board Landsat 7 satellite acquires images of the Earth since July 1999 with a revisit frequency of 16 days. Landsat ETM+ has 8 spectral bands (6 bands in the visible, one thermal band and one panchromatic band). An ETM+ scene has an Instantaneous Field Of View (IFOV) of 30 meters  $\times$  30 meters in bands 1-5 and 7 while band 6 has an IFOV of 60 meters  $\times$  60 meters on the ground re-sampled to 30-meter pixels. The equatorial crossing time is nominally 10am  $\pm$  15min. In the framework of the project Environnement Méditerranéen et Modélisation des Agro-Hydrosystèmes (EMMAH <http://www.umr-emma.fr>), 8 images have been extracted over 2009 from the USGS website (<http://edcsns17.cr.usgs.gov/NewEarthExplorer/>) over the study area and georeferenced according to the Lambert 3 projection. Atmospheric corrections have been performed to estimate LST from top of the atmosphere radiances using MODTRAN5 radiative transfer model and comparison with ground measurements showed good agreement (see Mira et al. (2014)). More details on the Landsat 7 data processing are available in <http://www.cesbio.ups-tlse.fr/multitemp/?p=3825>. The acquisition time over the study area is 10 : 20 am. As the observation data spans from March to September, only 7 images will be considered for the following dates : 03/21/09, 04/06/09, 04/22/09, 06/25/09, 07/27/09, 08/28/09 and 09/29/09.

**Inter-calibration of Landsat 7 and Meteosat LST** Before performing the validation of the downscaling method against HSR Landsat 7 LST it is important to verify whether LST observations and Landsat 7 ones are in good agreement. For that purpose, available Landsat 7 LST maps have been aggregated to Meteosat spatial resolution to be compared to Meteosat LST maps at acquisition time (10 :15 AM for Meteosat and 10 :20 AM for Landsat 7). The few number of available validation data compared to observation ones does not allow us to correct Meteosat time series according to Landsat 7 data. In these circumstances, we decided to calibrate rather validation data according to observation ones. As LST is highly variable in space, it is important, before going



into inter-calibration, to verify whether the correction of HSR Landsat 7 data is possible or not using LSR Meteosat LST. To do so, the Cumulative Distribution Functions (CDF) of the available Landsat 7 LST maps at both HSR and LSR (for aggregated Landsat 7 LST maps) have been calculated and compared to the corresponding Meteosat CDFs as presented in Figure 6.3. The comparison revealed two points. The first point is that for all the validation dates, at Meteosat spatial resolution, Landsat 7 and Meteosat LSTs present important discrepancies and biases that differ from a validation date to the other. The second point is that variance is strongly scale dependent. In fact, the aggregation has drastically smoothed the Landsat 7 LSTs and decreased the variances. Accounting for that, even if a bias correction of LSR Landsat 7 according to Meteosat LSTs is possible at coarse spatial resolution (through image processing methods such as CDF-matching, RANSAC, etc.), it can not be applied to the 30m-scale of the original Landsat 7 images. Therefore, the Landsat 7 derived LSTs will only be used in the following for a qualitative evaluation of the downscaling results at the spatial resolution of the LC map (30m).

**MODIS LST data** The MODIS sensor resides aboard the Terra and Aqua platforms, offering a view of the Earth's surface every 1-2 days. The MODIS sensor collects data within 36 spectral bands, ranging in wavelengths from  $0.4\mu\text{m}$  to  $14.4\mu\text{m}$ . MODIS (Terra/Aqua) Land Surface Temperature and Emissivity (LST/E) products provide per-pixel temperature and emissivity values in a sequence of swath-based to grid-based global products. The MODIS (Terra /Aqua) LST/E Daily L3 Global 1km Grid product (MOD11A1/MYD11A1), is tile-based and gridded in the Sinusoidal projection, and produced daily at 1 km spatial resolution. More details are provided in <http://modis-land.gsfc.nasa.gov/temp.html>. The equatorial passing times for Terra and Aqua are respectively 10h30 am/pm and 1h30 am/pm leading to a maximum of 4 MODIS LST maps a day (Wan and Li (2008)). In this work MODIS level 3 daily LST ( MYD11A1 and MOD11A1 Collection 5) were acquired over the study area for the same observation period (from 01/03/09 to 30/09/09) to evaluate the downscaling results at 1km spatial resolution.

**Inter-calibration of MODIS and Meteosat LST** As MODIS LST product can merge different orbits the acquisition time and the viewing angles could be different. Thus a pre-processing step of co-registration with Meteosat has been applied in order to compare observations acquired at similar time and to mask remaining cloudy pixels. For that purpose, only MODIS LST data with acquisition time-lag less than 3 minutes towards Meteosat data were kept. As the validation is performed for daytime period, only 29 MODIS LST dates have been finally kept for the validation process. As previously mentioned, accounting for LST high temporal variability and the few number of retained MODIS validation LST data compared to the available Meteosat LST time series, the inter-calibration between observation and validation was performed according to observation data. To verify whether MODIS LST and Meteosat ones are in a good agreement MODIS LST maps have been up-scaled to Meteosat spatial resolution and the CDFs at Meteosat-scale and MODIS-scale data have been calculated for the Landsat 7 validation dates (when both MODIS and Landsat 7 data are available). Figure 6.4 present MODIS 1km CDF, upscaled MODIS CDF and Meteosat CDF for the dates where both MODIS and Landsat 7 data are available. The comparison of the LST distribution at MODIS scale and Meteosat ones for the selected dates revealed quite good agreement with the smallest biases obtained for wintertime dates and higher ones for summertime dates (Meteosat LSTs being warmer than MODIS ones).

To investigate the origin of these biases, MODIS aggregated LST have been compared

to Meteosat ones over the most homogeneous pixels (fraction of the majority class larger than 55%) at Meteosat scale. The number of retained pixels for each class is as follows : 9 “Bare soil” pixels, 9 “Water” pixels, 68 “Tree cover” pixels and 1 “Rice” pixel. Then, Meteosat and MODIS temperatures were compared at coarse pixel scale for the different validation dates. Figure 6.5 shows the histograms of temperature differences observed for these most homogeneous pixels. At first sight we can note that Meteosat LSTs are generally warmer than MODIS ones with a mean bias over the different validation dates equal to 0.76K. The features look different from one class to another with larger biases observed for the water and Tree cover pixels. To thoroughly explore the origins of these discrepancies, we presented them, for each class, as a function on the acquisition date and time (see Figure 6.6). The plot shows that larger values are observed at noon and early afternoon which can probably be explained by directional viewing effects. Indeed, Meteosat and MODIS are on different orbits : Meteosat on a geostationary orbit is viewing our study area from the South whereas MODIS on polar orbits is viewing it from the West when descending or from the East in ascending orbits. In those different configurations, the sunlit and shadowed fractions of soil and vegetation which present different temperatures, vary with time leading to different composite temperatures, as demonstrated by various authors (Guillevic et al. (2013); Ermida et al. (2014)). These shadowing effects are often observed when the fraction of trees is significant as shown by Ermida et al. (2014). Over water pixels, the observed bias (about 3.5K) could be explained by differences in the emissivities prescribed in the two products. In the Meteosat product and at these coarse scales, these pixels are considered as land with a prescribed emissivity probably lower than the one prescribed in the MODIS product in which they are considered as pure water.

Such an underestimation of the emissivity could explain the positive biases observed for all the water pixels. Considering all this and the necessity to inter-calibrate our data, we propose to keep Meteosat data unchanged and to correct the MODIS data used for the validation by adding a time dependent weighted bias taking into account the land cover fractions as follows :

$$\hat{T}_{MODIS_{i,j}}^d = T_{MODIS_{i,j}}^d + \sum_{k=1}^4 \alpha_{i,j}^k Bias_k^d, \quad (6.1)$$

where  $d$  corresponds to the validation date,  $i$  and  $j$  are respectively the (row, column) coordinates of the MODIS pixel,  $Bias_k^d$  is the mean bias evaluated for class  $k$  and date  $d$  using the homogeneous pixels as described above and  $\alpha_{i,j}^k$  represents the fraction of class  $k$  inside pixel  $(i, j)$ . It is worth noting that only the four classes defined for Meteosat scale pixels have been included in the correction process and that the contribution of the remaining classes biases has not been included.

### 6.2.3 Results and discussion

This section presents both SETHYS prior and GPS downscaling results. The performances of the GPS downscaling approach have been tested quantitatively in terms of reduction of the root mean square error  $RMSE$ , the mean absolute error  $MAE$  and the bias at MODIS spatial resolution over the study area and for the 29 validation dates selected in the pre-processing step as described in the subsection 6.2.2.3. Landsat 7 data will serve only for a qualitative evaluation of the downscaled LST due to the presence of inter-calibration problems previously presented (see subsection 6.2.2.3). In the following, the prior solution is defined as the mean simulation of  $N = 600$  particles set. GPS downscaling performances will be evaluated against prior solutions for different DA scenarios at MODIS pixel scale over all the validation dates and for the whole study region. The temperatures of each land cover class are validated over the homogeneous

MODIS pixels defined as the pixels where the fraction of the majority class is larger than 55%. Table 6.2 describes the experimental settings of the different DA scenarios. In the first scenario, corresponding to a daily assimilation window, a first case was studied where uncertainties are considered on forcing data to evaluate their impact on GPS-downscaling approach performances. The noise was generated using a first order auto-regressive filter and was added to the air temperature forcing variable (the most sensitive variable) with an error variance equal to 1K. For all the other scenarios, no uncertainties are considered on the forcing data. The two other scenarios : 3-day assimilation window and weekly assimilation window, aims to highlight the impact of the assimilation window size on GPS performances. In all the scenarios, only daytime Meteosat observations, between 7 : 45 am and 4 : 00 pm, have been considered. It is also important to highlight that for all the experiments if at the end of the analysis step more than 90% of the particles are rejected, all the particles are re-sampled using a uniform distribution over the parameter space to avoid the collapse of the PS.

### 6.2.3.1 GPS downscaling approach validation at MODIS spatial resolution

Table 6.3 shows both prior and GPS downscaling results over the study area and for the 29 validation dates. Overall, GPS downscaled LSTs show better agreement with MODIS LSTs presenting lower RMSEs, MAEs and biases for all the scenarios compared to prior estimations. On average, the RMSEs are on the order of 2.5K with a significant reduction of the bias of about 1K as compared to prior error statistics. From a sub-pixel point of view, GPS outperforms the model with a mean error in sub-pixel temperature estimation less than 2K for all the classes and for the best case scenario considered (scenario.1-Case.2). A more thorough analysis of the results is presented in the following subsections.

**Impact of uncertainties in forcing data on GPS downscaling performances** To perform the downscaling, meteorological and vegetation forcing data are required at the coarse pixel scale. Such data are generally subject to errors for example due to lack of representativity when point measurements are used or due to over-smoothing when atmospheric reanalysis are used over the assimilation period. Such errors could have a significant impact on the downscaling performances since the model simulated temperatures are controlled by the available forcing data. To investigate this point, time-correlated a noise has been added to the air temperature forcing variable over the year 2009. It was generated by a first order auto-regressive filter with respect to the temporal distribution of air temperature (through the time correlation). The variance of this noise was equal to 1K. The second and third columns of Table 6.3 show the GPS downscaling results respectively when uncertainties in forcing data are considered and not. Surprisingly, the comparison of GPS downscaling results for the first scenario shows that, despite adding uncertainties, the performances of the GPS downscaling are not affected (RMSEs, MAEs and biases are not significantly different). This statement is also confirmed when comparing the downscaling results for the most homogeneous MODIS pixels as presented in the first two columns of Table 6.2.4. For all the land cover classes, only a slight increase in error statistics has been noticed when comparing the first and the second cases of the first scenario. Such result is important since it proves that even if forcing data are perturbed with an important noise, the assimilation of LSR observations through the GPS ensure a stable solution of the downscaling problem.

**Impact of the assimilation window length** The two last columns of Table 6.3 present GPS downscaling results with larger assimilation window sizes (respectively 3 days and 7 days). Compared to the first scenario error statistics, a slight increase in the RMSE,

MAE is noticed (about 0.13K for Scenario 2 and 0.2K for Scenario 3) with mainly the same bias (about 0.9K). The small degradation caused by the increase of the assimilation window size was somehow expected. In fact when the assimilation window is larger, more observations are assimilated with the GPS. However the availability of many observations is not always beneficial especially if they are locally perturbed with noise or cloud. Because, as we are performing a smoothing of particles, the weight of a particle is the product of the different “sub-weight” evaluated for every observation available over the assimilation window. Thus, if the number of observations is large, the “sub-weights” are smaller and the final weights are nearly equal. In this case, the rejection of the small weighted particles is not evident since all the particles have nearly equal weights. Consequently the variance of the particle ensemble is not optimally reduced leading to an increase in the error variance (larger RMSE and MAE).

A more thorough analysis of the GPS downscaling results has been performed by comparing the estimated sub-pixel temperatures at MODIS scale. Tables 6.4 and 6.2.4 present respectively prior and GPS downscaling results for the different classes present in the LC map. RMSEs, MAEs and biases have been evaluated over the 29 validation dates and the most homogeneous MODIS pixels. It appears from Table 6.4 that for all the scenarios, the estimation of the land cover class temperatures using GPS are better than the prior ones except for the prairie class. In fact, as the prairie class does not represent any homogeneous pixel at Meteosat spatial resolution, the pixels presenting a large fraction of this class have not been well corrected in the inter-calibration process at MODIS pixel scale. Moreover, because of the lower contribution at the coarse resolution pixel, the model parameters are less constrained for prairie class (as well as the other minority classes) class compared to the other majority classes. Their respective temperatures could therefore be subject to larger estimation errors. Apart from prairie class, we can say that the best results are GPS ones and more specifically for the daily assimilation window case. We have also noted that the most important reduction in error statistics was obtained for the most homogeneous Meteosat classes compared to the best prior solution. Figure 6.7 presents an example to this statement for the bare soil class. In this figure, we have superposed the fraction of cover rates and the error statistics maps for the bare soil class. We can see that the most homogeneous bare soil MODIS pixels present in the most homogeneous bare soil Meteosat pixel have lower error statistics than those present in pixels where the bare soil class is not dominant. Such result was somehow expected because the mixing effect increases with the pixel heterogeneity (assessed by the number of classes at the coarse pixel scale) (Agam et al. (2007a)) joining in this way the conclusion obtained in the previous work of Mechri et al. (2014).

### 6.2.3.2 GPS downscaling approach validation at Landsat 7 spatial resolution

As the GPS downscaling solutions are estimated at the LC map spatial resolution it is interesting to verify if the HSR estimated LST map has a similar spatial distribution than the HSR Landsat 7 LST map even if the validation against this data is not possible quantitatively. For that purpose we have zoomed on a region of the LC map where we have both highly heterogeneous structures and homogeneous ones. As the validation against HSR LST data is only qualitative, we have chosen the most contrasted Landsat 7 LST map from the 7 available ones, i.e. the 28<sup>th</sup> April 2009. Figure 6.8 shows respectively LSR Meteosat LST, HSR Landsat 7 LST and HSR GPS estimated LST for the second case of scenario one over a subregion of our study zone. This figure shows that the GPS estimated temperatures have quite similar spatial distribution for the most continuous homogeneous surfaces such as the barren planes (red contours) and the tree cover zones (green contours). A lower agreement has been encountered for highly heteroge-

neous zones and sparse vegetation ones (blue contours). Such result is important even if no quantitative statistics are calculated because it shows that the GPS downscaling of LST more than  $100\times$  coarser than the validation LST has captured more of the spatial variability of the extracted validation zone (except for highly heterogeneous and sparse regions).

#### 6.2.4 Conclusions

The methodology presented here allows to downscale Meteosat high frequency LST time series at higher spatial resolution. The approach is based on the assimilation of the observations in a land surface model able to simulate the temperature of the various land cover types (components) inside the Meteosat pixel and the composite thermal emission. The method assumes that LST is stationary for each component inside the  $3\text{km} \times 5\text{km}$  pixel. It requires the availability of a high resolution land cover map providing the fractions of each vegetation class as well as atmospheric and surface forcing data to simulate prior estimations of the components temperature.

The method was applied over a  $4539\text{km}^2$  region in the South-East part of France and a 7-months period of the year 2009, to downscale Meteosat temperatures to 1km scale, using MODIS LST for the validation. The performances of the method were assessed using a set of 29 non cloudy MODIS images acquired over the study period and the results show that the temperatures at 1km scale can be estimated with an overall accuracy of 2.5K. The precision finally obtained on the class temperatures was estimated over the most homogeneous MODIS pixels and the results show that they can be retrieved with an accuracy better than 2K which is quite satisfactory compared to previous works.

These performances were obtained using a 1-day assimilation window and homogeneous forcing data over the region. The introduction of uncertainties in the atmospheric forcing did not change the performances, demonstrating that the algorithm is rather robust and not too much sensitive to atmospheric forcing uncertainties. The widening of the assimilation window slightly degrades the performances but the results remain still acceptable using a time length of 7 days. As expected, the performances are all the more better that the pixel is less heterogeneous, and the minority land cover classes can be subject to larger errors and more work is needed to adjust prior errors according to class contribution in the observations.

This work faced two main difficulties. The first one consisted in the lack of ground-truth data necessary to better represent the spatial variability of the vegetation characteristics and better define the homogeneous elements to downscale (irrigation practices, LAI, vegetation height, etc.). The second one was the lack of reliable validation data to validate the class temperatures, due partly to inter-calibration issues not resolved at this stage. However, despite these difficulties, the results obtained are reasonably acceptable compared to other approaches based on statistical downscaling approaches or regression relationships against vegetation indices (ex : [Inamdar and French \(2009\)](#) ( $\approx 2\text{K}$ ), [Merlin et al. \(2010\)](#) ( $\approx 3\text{K}$ ), [Kallel et al. \(2013\)](#) ( $\approx 2\text{K}$ )). Furthermore, this method provides downscaled LST times series at the model time step (in our case 30 minutes) as well as all the model output variables like soil moisture or evapotranspiration, even if they are not observed.

The limits of the downscaling have not yet been explored but they are constrained by the resolution of the land cover map, the validity of the assumptions made (stationarity of the class temperature at the coarse pixel scale, stationarity of the land cover map, etc.), the model errors and the uncertainties on the forcing data. All these issues need to be examined and quantified. The inter-calibration issues require much attention because they prevent further developments on the assimilation of multi-instruments mul-

6.2. Article 2 (version *draft*) : Downscaling of Meteosat Land Surface Temperature over an heterogeneous landscape using a Data Assimilation approach

tiscala products. The contribution of other sources of data should also be considered in future studies.



FIGURE 6.1 – Study area : Crau-Camargue Region

TABLE 6.1 – Description of the selected parameters for the considered land cover types

Class	Selected parameters	Definition	Variation ranges
Bare soil	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[0.7, 2.2]
	$Alb_{sec}$	Bare soil albedo	[0.25, 0.85]
Water	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[1.7, 3.2]
	$Alb_{sec}$	Bare soil albedo	[0.25, 0.85]
	$e_s$ $w_{sat}$	Bare soil emissivity Saturation volumetric humidity	[0.93, 0.97] [0.4, 0.5] $m^3.m^{-3}$
Prairie	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[0.6, 2.1]
	$Alb_{sec}$	Bare soil albedo	[0.2, 0.8]
	$V_{max}$	Vegetation photosynthetic capacity	[20.0, 80.0]
	$Alb_g$ $e_g$	Vegetation albedo Vegetation emissivity	[0.1, 0.26] [0.92, 1.0]
Cultures	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[0.75, 2.25]
	$Alb_{sec}$	Bare soil albedo	[0.2, 0.8]
Tree cover	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[1.53, 3.03]
	$Alb_t$	Tree albedo	[0.15, 0.33]
	$lf_t$	Tree leaf width	[0.005, 0.015] $m$
	$e_g$ $e_t$	Vegetation emissivity Tree emissivity	[0.92, 1.0] [0.92, 1.0]
Wheat	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[0.75, 2.25]
	$Alb_{sec}$	Bare soil albedo	[0.2, 0.8]
	$V_{max}$ $lf_g$	Vegetation photosynthetic capacity Vegetation leaf width	[20.0, 80.0] [0.0090.1] $m$
	$e_g$	Vegetation emissivity	[0.92, 1.0]
Rice	$Fact_{therm}$	Factor controlling the volumetric heat soil capacity	[1.75, 23.25]
	$Alb_{sec}$	Bare soil albedo	[0.2, 0.8]
	$w_{sat}$ $V_{max}$	Saturation volumetric humidity Vegetation photosynthetic capacity	[0.4, 0.5] $m^3.m^{-3}$ [20.0, 80.0]
	$e_g$	Vegetation emissivity	[0.92, 1.0]

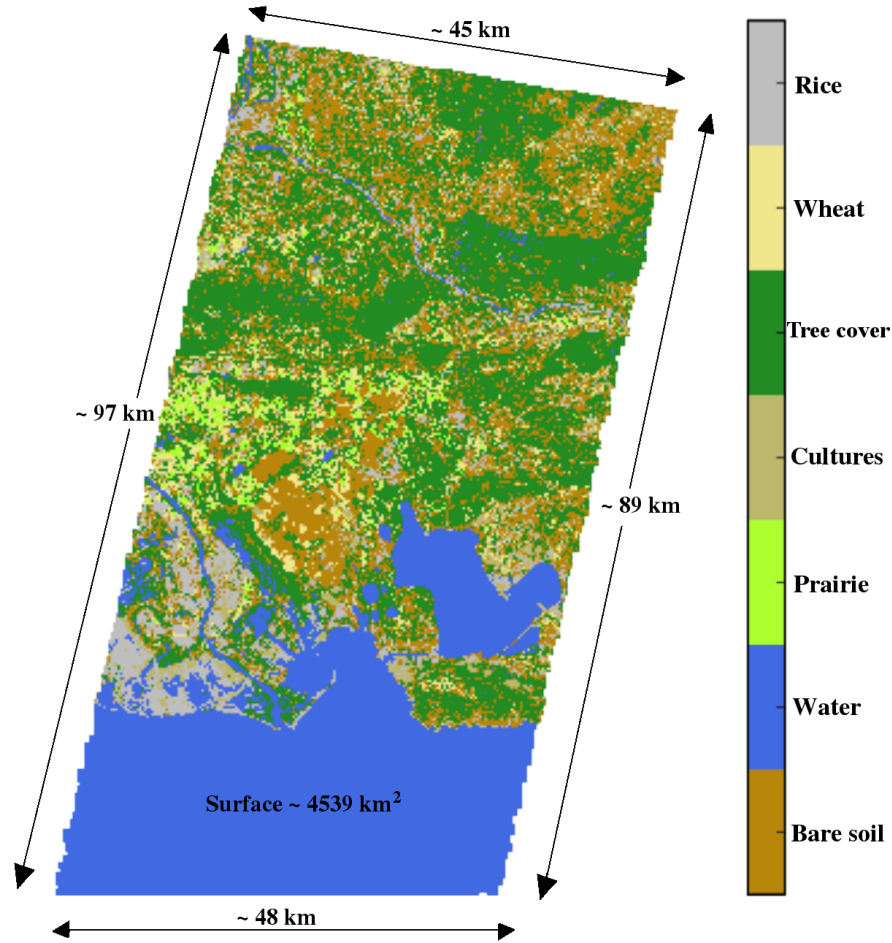


FIGURE 6.2 – Landsat 7 7-class land cover map (30m × 30m spatial resolution).

TABLE 6.2 – Experimental settings for the different experiments

Common settings			
Number of classes	$N_{class} = 7$		
Number of calibrated parameters/class	$N_{param} = 5$		
Assimilation starting day	DOY 60 (03/21/2009)		
Particle ensemble size	$N = 100$		
Observation window	7 : 45 A.M. → 6 : 00 P.M.		
Specific settings : assimilation window size			
	1 day	3 days	7 days
Case 1	Case 2		
Uncertainty on forcing data	No uncertainty on forcing data	No uncertainty on forcing data	No uncertainty on forcing data

TABLE 6.3 – Prior and GPS downscaling results averaged over the whole region and all the validation dates.

	Prior	Scenario 1		Scenario2	Scenario 3
		Case 1	Case 2		
<i>RMSE</i>	2.89	2.50	2.49	2.62	2.70
<i>MAE</i>	2.48	2.10	2.11	2.20	2.27
<i>Bias</i>	-1.72	-0.87	-0.92	-0.89	-0.89



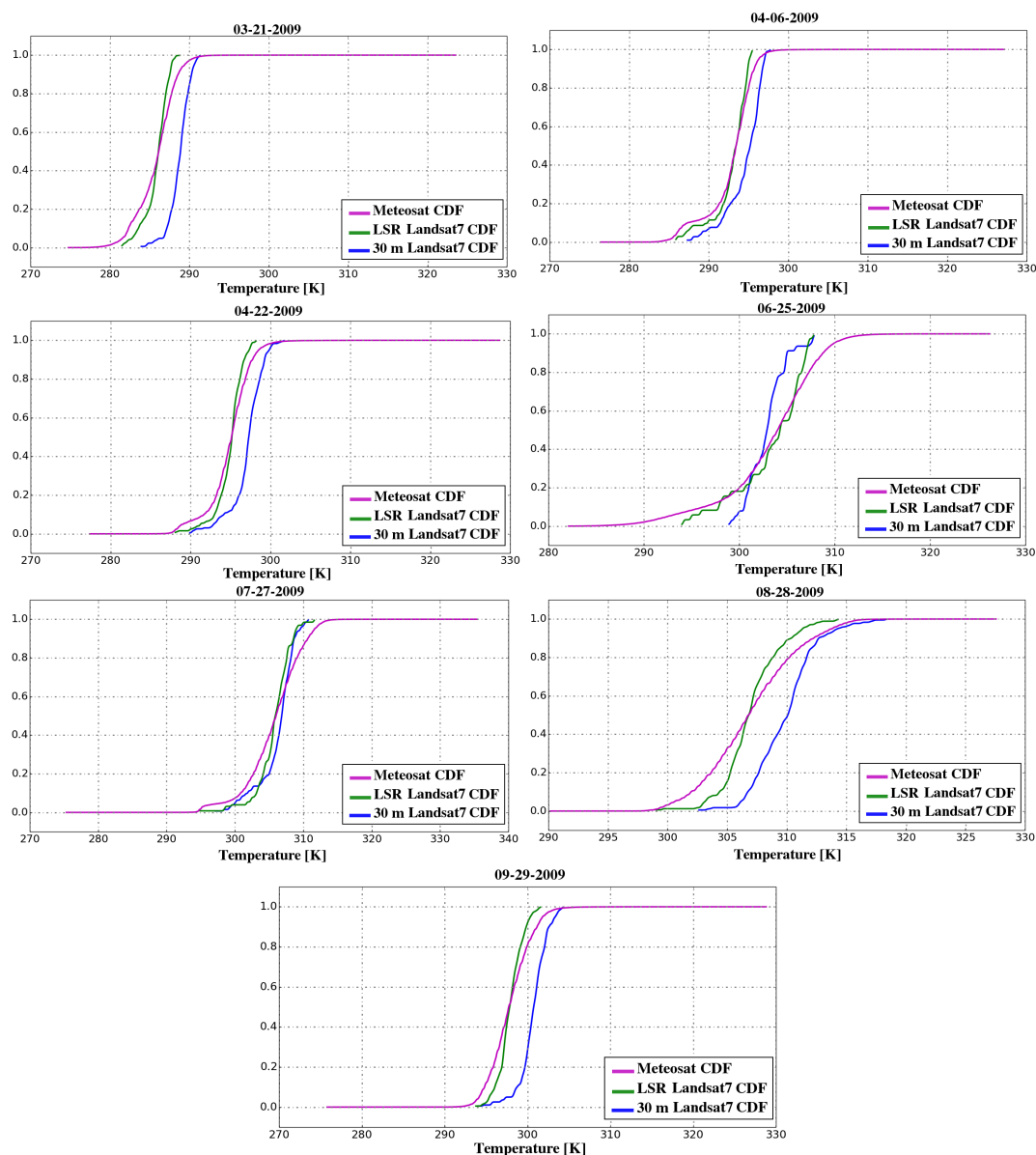


FIGURE 6.3 – Cumulative Distribution Functions (CDF) estimated for Meteosat and Landsat 7 LSTs for the seven validation days over the year 2009 : magenta and green curves represent respectively 30m and LSR Landsat 7 LST CDFs however blue curves represent Meteosat LST CDFs.

TABLE 6.4 – Prior results by class (averaged over MODIS homogeneous pixels and all the validation dates).

	RMSE	MAE	Bias
Bare soil	2.16	1.87	-1.31
Water	2.48	2.16	-0.71
Prairie	0.96	0.80	-0.02
Cultures	1.92	1.66	0.41
Tree cover	2.68	2.38	-2.25
Wheat	1.97	1.60	-0.97
Rice	3.06	2.57	-2.27

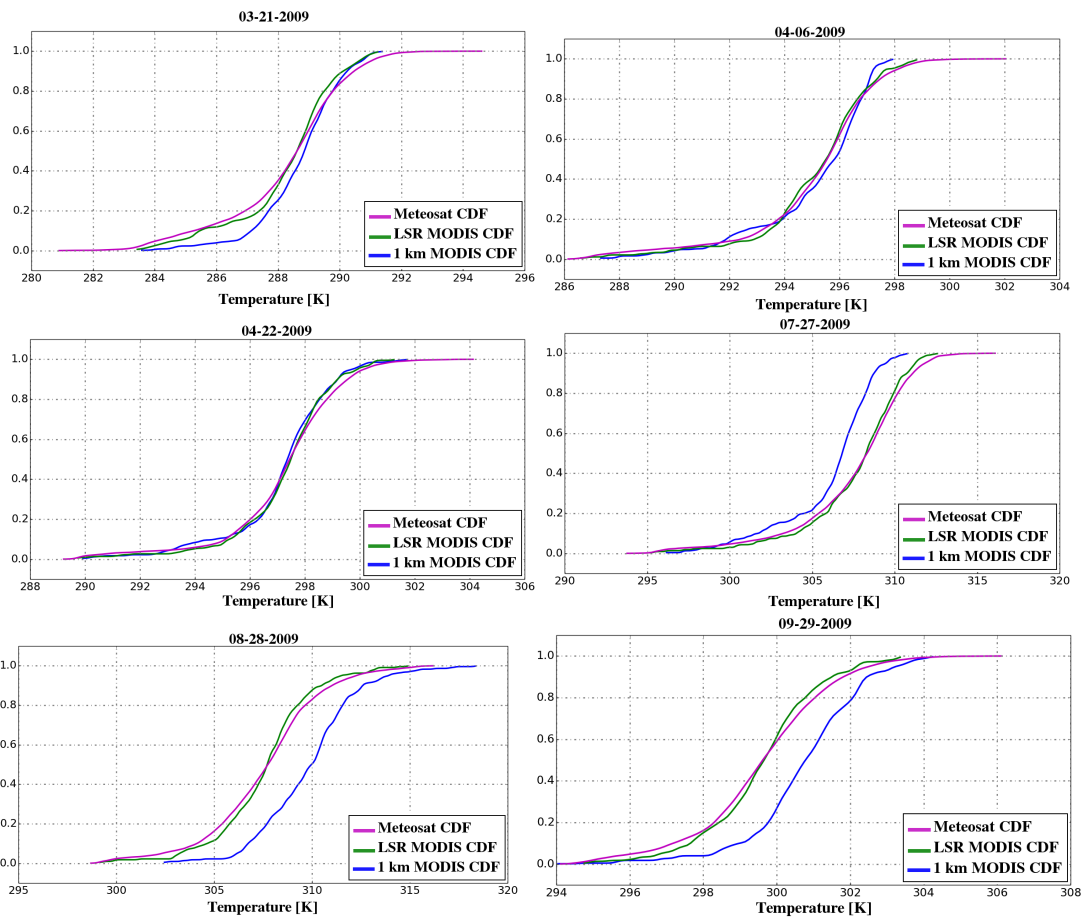


FIGURE 6.4 – Cumulative Distribution Functions (CDF) estimated for Meteosat and MODIS LSTs for the six of Landsat 7 validation dates over the year 2009 : magenta and green curves represent respectively 1km and LSR MODIS LST CDFs however blue curves represent Meteosat LST CDFs.

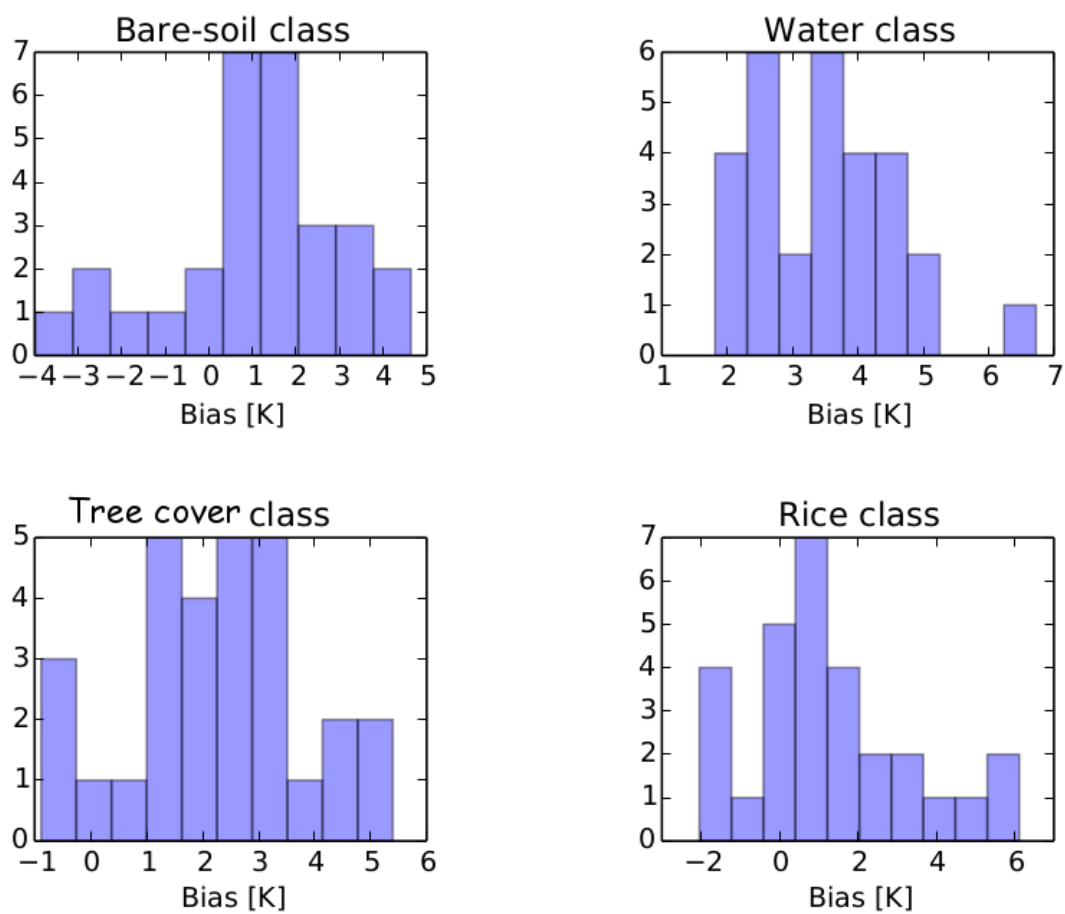


FIGURE 6.5 – Histograms of bias evaluated for the classes of the most homogeneous Meteosat pixels.

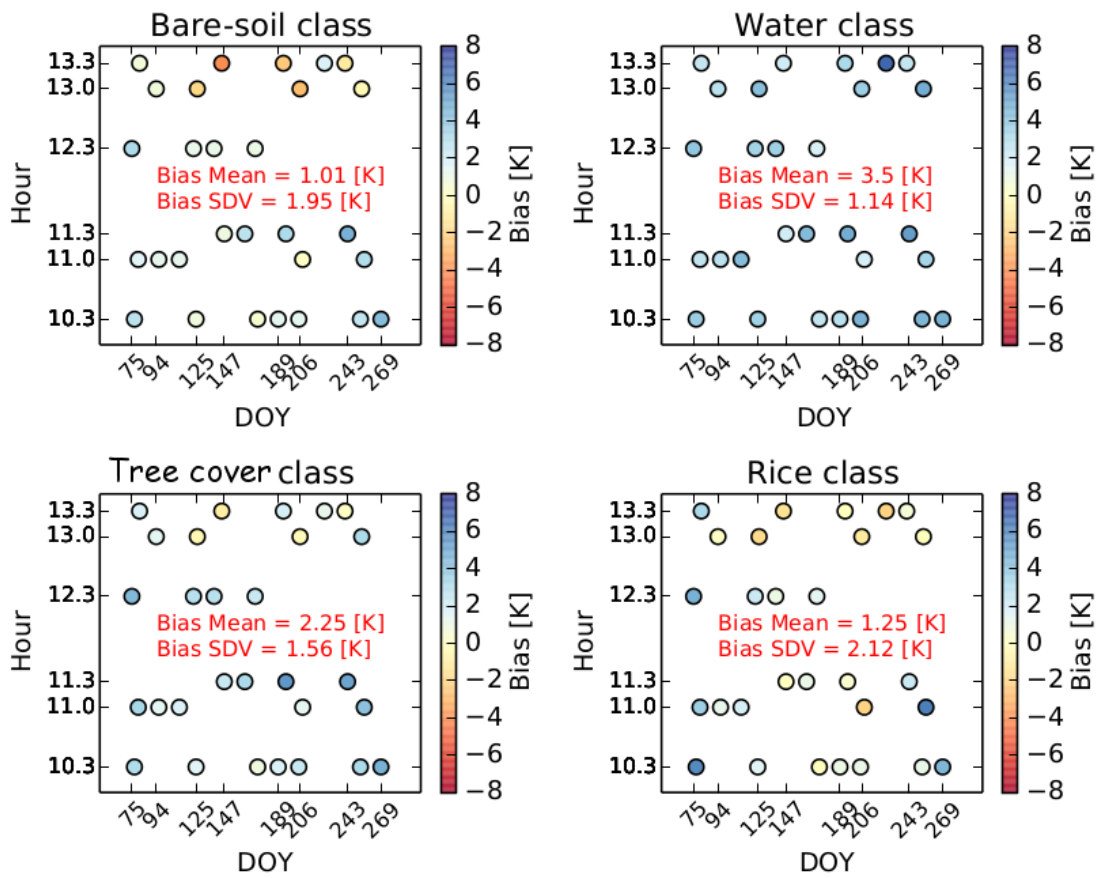


FIGURE 6.6 – Bias evaluated for the classes corresponding to the most homogeneous Meteosat pixels.

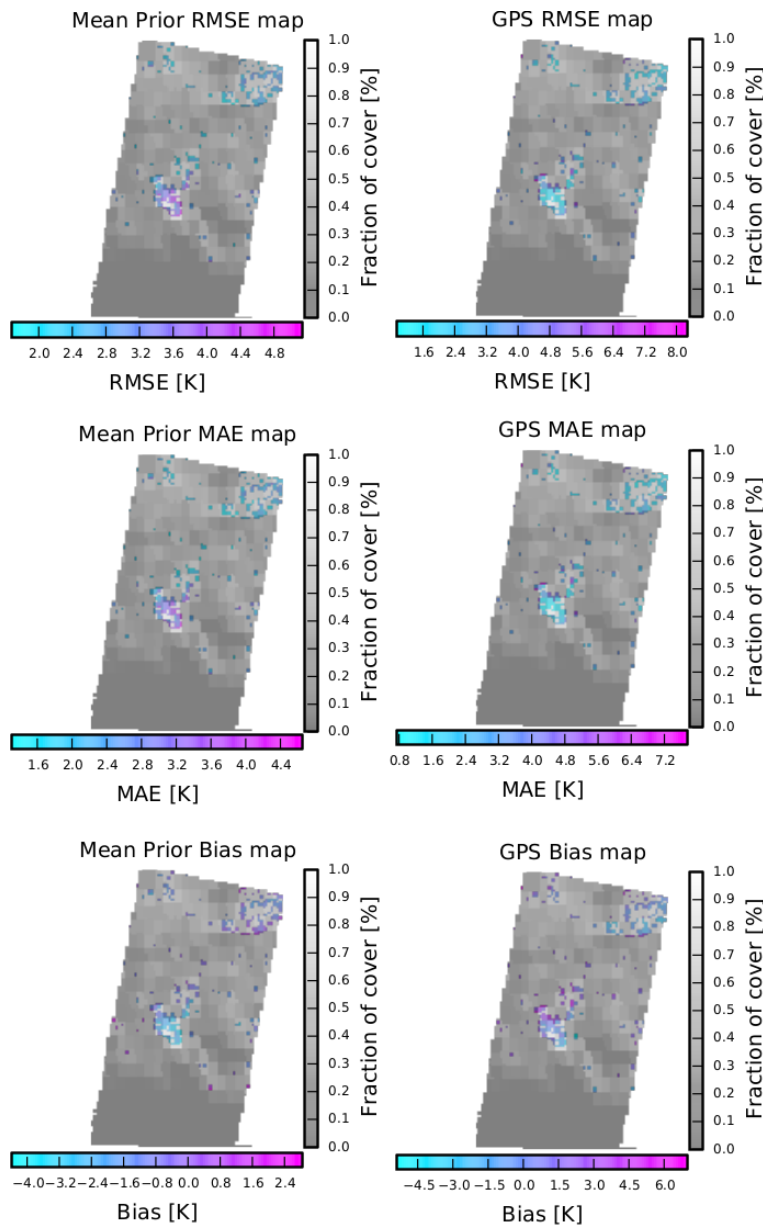


FIGURE 6.7 – Error statistics map versus Meteosat spatial resolution cover fraction rates for bare soil class : the colored scale corresponds to the error statistic value in [K] and the gray scale corresponds to the cover fraction of the bare soil class in each Meteosat pixel.

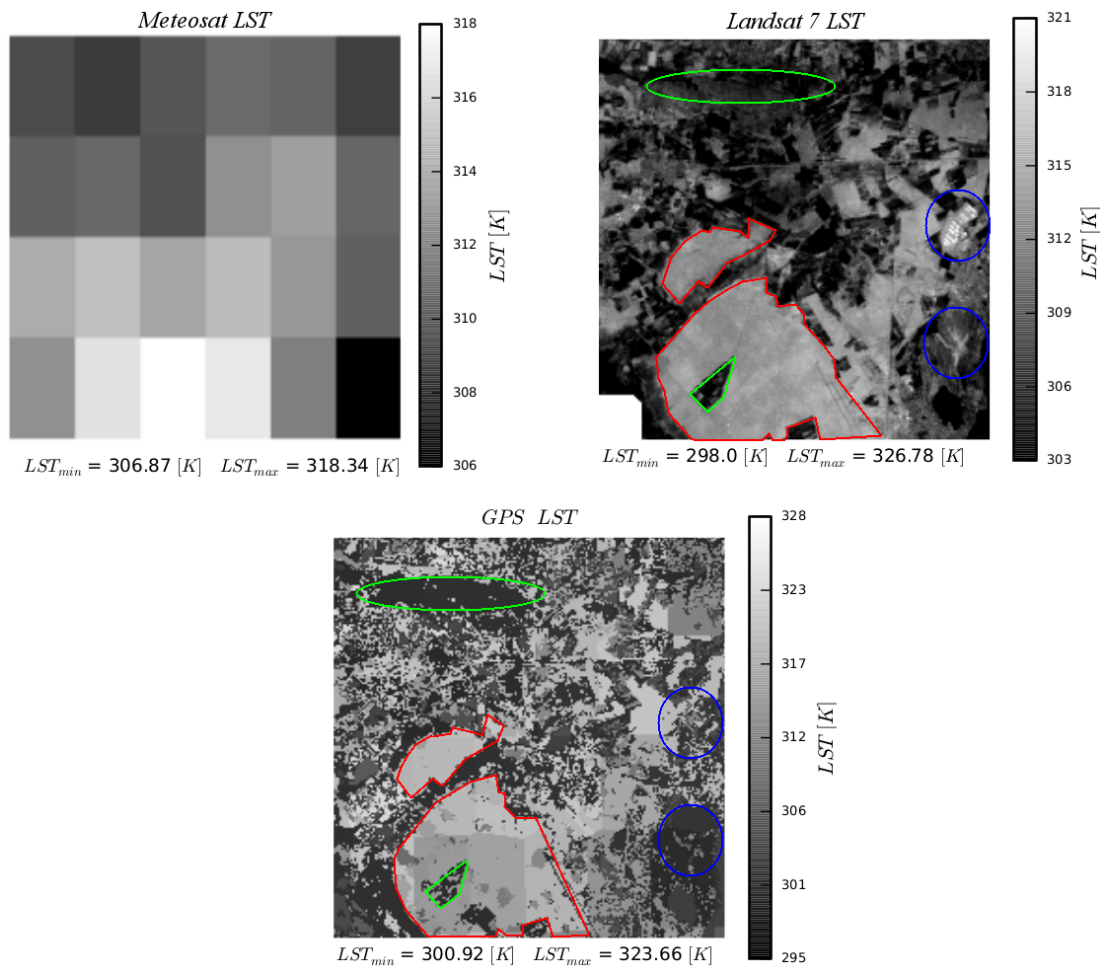


FIGURE 6.8 – HSR GPS downscaling results versus HSR Landsat 7 data and LSR Meteosat data over the Landsat 7 extraction zone.

TABLE 6.5 – GPS downscaling results by class (averaged over MODIS homogeneous pixels and all the validation dates).

	Npixels	Scenario 1			Scenario 2			Scenario 3					
		Case 1 RMSE	Case 1 MAE	Case 1 Bias	Case 2 RMSE	Case 2 MAE	Case 2 Bias	Scenario 2 RMSE	Scenario 2 MAE	Scenario 2 Bias	Scenario 3 RMSE	Scenario 3 MAE	Scenario 3 Bias
Tree cover	1315	2.04	1.68	-1.67	2.01	1.78	-1.73	2.06	1.75	-1.69	2.09	1.79	-1.67
Bare soil	309	1.19	1.00	0.89	1.07	0.94	0.73	1.37	1.14	0.80	1.49	1.20	0.81
Water	255	1.58	1.29	-1.12	1.53	1.24	-1.06	2.00	1.64	-1.19	2.28	1.82	-1.30
Rice	71	2.00	1.74	-1.70	1.99	1.70	-1.67	2.20	1.87	-1.80	2.29	1.97	-1.85
Cultures	17	1.64	1.18	-0.64	1.59	1.20	-0.74	1.71	1.39	-0.99	1.84	1.47	-1.20
Wheat	10	2.00	1.60	-0.84	1.53	1.20	-0.67	1.57	1.30	-0.91	1.70	1.40	-0.99
Prairie	5	1.20	0.89	0.04	1.05	0.86	0.16	1.04	0.83	0.16	1.02	0.81	0.15

## 6.3 Résultats et conclusions

La mise en place de notre approche de désagrégation sur la base de données 2009 a été confrontée à plusieurs difficultés liées soit à la non disponibilité des données de forçage de la végétation (indispensables pour une bonne estimation de la variabilité spatiale et temporelle de la température des végétations) soit à la non disponibilité des données de validation à la résolution de la classe de végétation à cause des problèmes d'inter-étalonnage non résolus. En dépit de toutes les difficultés, les résultats obtenus par notre méthode sont satisfaisants et surpassent les performances du modèle sans assimilation dans tous les cas.

L'assimilation des observations Météosat par le lisseur particulière a permis d'estimer la température des classes majoritaires (par rapport aux pixels MODIS les plus homogènes) avec une incertitude inférieure à 2.5K qui est un résultat très acceptable par rapport aux méthodes de désagrégation classiques basées sur les régressions ou l'inversion statistique ( ex : [Inamdar and French \(2009\)](#) ( $\approx 2K$ ), [Merlin et al. \(2010\)](#) ( $\approx 3K$ )). Les meilleurs résultats ont été obtenus pour la fenêtre d'assimilation de 1 jour avec des incertitudes sur l'estimation des températures inférieure à 2K. Les résultats obtenus en considérant des incertitudes sur la température de l'air, variable de forçage atmosphérique très sensible à la température de surface, étaient très proches de ceux obtenus dans le cas de référence ( $RMSE \in [1.2, 2.0] K$ ). Ce résultat prouve la robustesse de la méthode face à des erreurs liées à des incertitudes sur les données de forçage atmosphérique.

La validation qualitative sur une partie de la région d'étude a montré un bon accord en ce qui concerne la distribution spatiale de la température à la résolution de Landsat7 à l'exception des zones caractérisées par de fortes hétérogénéités spatiales ou dont les caractéristiques de végétation n'étaient pas bien estimées (comme les données de LAI interpolées à partir des cartes de NDVI pour la classe culture). Plusieurs perspectives peuvent être envisagées à la suite de ce travail et feront l'objet du dernier chapitre de la thèse.





---

## Conclusions et perspectives

LE PRÉSENT TRAVAIL DE THÈSE a permis le développement d'une nouvelle approche de désagrégation de la température des surfaces continentales basée sur l'assimilation des données IRT Météosat à BRS et à HRT dans un modèle de surface continentale, le modèle SETHYS.

L'approche de désagrégation proposée vise à contraindre la dynamique des températures simulées par le modèle SETHYS à HRS et à HRT pour différents types d'occupation des sols en minimisant l'écart entre les températures agrégées à l'échelle des observations et les températures de télédétection IRT à BRS et à HRT. Pour simuler les températures des types de couverts présents dans la région d'étude, certains forçages atmosphériques et certaines caractéristiques des végétations doivent être fournis au modèle SETHYS.

Le choix de la méthode d'assimilation de données est guidé par la nature du problème auquel on fait face. Dans notre cas, on cherche à estimer les températures d'un certain nombre de classes de végétation présentes dans un pixel BRS d'une image IRT connaissant les fractions des différentes classes et la température du pixel BRS. Il s'agit donc d'un problème d'équifinalité<sup>1</sup> dont les solutions peuvent être multiples. Les méthodes d'assimilation de données d'ensemble présentent un cadre mathématique adéquat pour résoudre ce genre de problèmes en proposant un ensemble de solutions candidates.

Les méthodes d'assimilation de données d'ensemble les plus connues sont le filtre de Kalman d'ensemble (EnKF) et le filtre particulaire (PF). Nous nous sommes dirigés vers le lissage particulaire en suivant les travaux de Rémy et al. (2012). En effet, différentes études menées en amont de cette thèse ont montré la variation de la sensibilité des paramètres de SETHYS en fonction de la variabilité journalière de la température de surface d'où l'intérêt de faire l'assimilation de données sur des fenêtres d'assimilation qui prennent en compte la variabilité journalière de la LST (Coudert et al. (2006); Coudert and Otlé (2007); Coudert et al. (2008)). Ainsi, associé à cette

---

1. Le principe d'équifinalité indique qu'un même résultat peut être obtenu par des voies et conditions initiales différentes. Dans notre cas étant données que la température du pixel IRT est déduite de la somme pondérée des températures des types d'occupation des sols qui l'occupent la même valeur de cette température peut être retrouvée à partir de différentes valeurs des températures des autres classes d'où la notion d'équifinalité.

sensibilité journalière, le lissage particulière nous a semblé être la solution la plus pertinente pour contrôler l'information sous-pixellique dans le temps, là où auparavant on avait recours à des hypothèses de stationnarité locale. Le rôle du PF est donc de filtrer les températures simulées par le modèle pour les différents couverts présents dans un pixel IRT : il s'agit de sélectionner celles qui, agrégées à la résolution spatiale du pixel IRT, donnent les températures les plus proches des observations.

Pour générer les ensembles de températures de chaque type d'occupation des sols, nous avons fait varier certains paramètres du sol et des végétations du modèle SETHYS. Le choix de ces paramètres a fait l'objet d'une analyse de sensibilité sur la sortie température de SETHYS. Tout au long de cette thèse, notre but était de désagréger les températures Météosat à des résolutions spatiales plus fines. Pour cela, plusieurs expériences ont été menées soit dans un cadre synthétique pour évaluer la convergence de la méthode dans un cas simplifié, soit dans un cadre réaliste pour étudier les performances et les limites de notre méthode pour la désagrégation des températures Météosat à des moyennes (MODIS) et fines (ASTER, Landsat) résolutions spatiales. L'évaluation des performances de notre méthode vise à comparer la réduction des statistiques d'erreurs obtenues en utilisant le PS pour contraindre les simulations de SETHYS par les séries temporelles Météosat et celles obtenues en utilisant le modèle sans assimilation des observations (erreurs *a priori* ou "*prior errors*").

## 7.1 Principaux résultats

Les premiers résultats ont été obtenus pour la désagrégation de pseudo-observations, dont la résolution temporelle (10min) était comparable à des données Météosat (15min), sur un pixel synthétique contenant quatre types de couverts agricoles caractéristiques de la région d'étude : région de Crau-Camargue au Sud-Est de la France. Le pixel se compose de 25% de sol nu (régions steppiques), 25% de blé (blé de la Crau), 25% de prairie (prairies de la Crau) et 25% de riz (rizières de la Camargue). Une simulation de référence pour les températures des différentes classes a été utilisée pour la validation et pour générer les pseudo-observations (par ajout de bruit à la série temporelle obtenue après agrégation des simulations de référence).

L'évaluation des performances de l'approche a été démontrée par l'importante réduction de l'erreur moyenne quadratique  $RMSE$  comparée au *prior* (modèle sans assimilation des observations) avec une précision sur l'estimation des températures des quatre classes de 0.3K sur une période d'assimilation de 23 jours (du 28 Avril 2006 au 20 Mai 2006) et pour une fenêtre d'assimilation journalière avec 72 observations par jour ( $RMSE_{prior} = 1.2K$ ). L'examen de la variance des solutions obtenues pour chaque classe a montré que l'estimation est meilleure pour les classes les plus sèches quant aux conditions hydriques (le sol nu et le blé). En effet dans le cas où les conditions d'humidité du sol sont à saturation (cas des rizières, ou lors de l'irrigation des prairies) les températures simulées pour une classe donnée sont très proches de la température de l'air (forte évapotranspiration) et la variance de l'ensemble des températures simulées par le modèle est par conséquent très faible. Dans ce cas, l'assimilation des observations n'a pas d'apport sur la réduction de la variance de l'ensemble des températures et peut même dégrader les performances du lisseur si les observations sont très bruitées.

L'apport de l'assimilation des observations BRS par le lisseur particulière a été estimé par le calcul de l'indice d'efficacité (déduit du rapport entre l'erreur moyenne quadratique après assimilation des observations et celle du *prior*) pour plusieurs scé-

narios en faisant varier l'amplitude de l'erreur d'observation et la fenêtre d'observation (heure de début et de fin d'acquisition des observations sur une journée). Les meilleures performances ont été obtenues pour les erreurs d'observations les plus faibles et pour les heures d'observation autour de midi (période d'acquisition des observations entre 10h00 et 14h00). Toutefois, pour tous les scénarios traités, les indices d'efficacité étaient positifs et supérieurs à 20%. De même, les meilleurs indices ont été obtenus pour les classes les plus sèches avec des valeurs de 56% pour le sol nu et 59% pour le blé lorsque l'on assimile toutes les observations disponibles sur la fenêtre d'assimilation journalière. Il est aussi important de signaler que même avec une seule observation, le lisseur particulière arrive à réduire l'erreur *a priori* si cette observation est acquise dans la journée (entre 6h00 et 18h00) avec un indice d'efficacité supérieur à 18%.

Cette étude synthétique a prouvé l'intérêt de l'assimilation des observations pour contraindre les températures simulées par le modèle et les résultats satisfaisants obtenus ont été confirmés sur des données réelles de télédétection. La désagrégation des séries temporelles Météosat, dont la résolution spatiale est de  $5\text{km} \times 3\text{km}$ , a été évaluée pour différentes résolutions spatiales des données de validation disponibles sur la région d'étude. Pour les différentes résolutions spatiales étudiées, l'estimation des températures par le lisseur particulière est obtenue avec une incertitude inférieure à 2.5K, résultat meilleur ou comparable aux méthodes classiques basées sur l'interpolation (ex : Inamdar and French (2009) ( $\approx 2\text{K}$ ), Merlin et al. (2010) ( $\approx 3\text{K}$ )) ou sur l'inversion statistique (Kallel et al. (2013) ( $\approx 2\text{K}$ )).

La désagrégation des données Météosat à la résolution spatiale des données MODIS a été faite sur une zone de la région d'étude dont la surface est égale à  $4539\text{km}^2$ . La validation à l'échelle de la classe de végétation a été faite pour sept classes décrites par une carte d'occupation des sols à 30m de résolution et ceci pour les pixels MODIS les plus homogènes en ce qui concerne l'occupation des sols (pixels qui contiennent plus que 55% d'une classe donnée). La fenêtre d'observation a été choisie de façon à ne considérer que les observations durant la journée compte-tenu des résultats obtenus dans l'étude synthétique.

La disponibilité des données MODIS et Météosat pour l'année 2009 a permis de choisir une période d'assimilation plus longue que celle utilisée lors de l'expérience synthétique (sept mois de données Météosat et 29 images MODIS pour la validation). Ceci a permis de tester les performances du filtre non seulement pour une fenêtre d'assimilation journalière mais aussi pour des fenêtres d'assimilation plus longue (trois jours et sept jours). Les meilleurs résultats ont été obtenus pour la fenêtre d'assimilation de données journalières avec des incertitudes sur l'estimation des températures des classes à l'échelle de MODIS qui varient entre 1K et 2K.

L'impact des incertitudes sur les données de forçage a aussi été évalué en rajoutant des incertitudes sur la variable de forçage atmosphérique "température de l'air" dans le cas d'une fenêtre d'assimilation journalière. Les résultats obtenus en considérant des incertitudes sur la température de l'ordre de 1K étaient très proches de ceux obtenus précédemment ( $RMSE \in [1.2, 2.0] \text{K}$ ). Ce résultat prouve la robustesse de la méthode face à des erreurs liées à des incertitudes sur les données de forçage atmosphérique. La désagrégation des séries temporelles Météosat a été validée à de plus fines résolutions à l'aide de données ASTER récupérées sur la région d'étude en 2006 (une zone de surface  $1620\text{km}^2$ ). Malheureusement, les données HRS pour cette année étaient très limitées (une seule image ASTER disponible le 26/07/2006 à 10h47min). Pour cette journée, les séries temporelles Météosat ont été désagrégées à 90m de résolution

et les températures de sept classes décrites par une carte d'occupation de sol à la même résolution spatiale ont été estimées. Comme pour les données MODIS, seules les observations durant la journée ont été prises en compte lors de la désagrégation. Le lisseur particulière a permis l'estimation des températures des différentes classes avec une incertitude qui varie entre [1.3, 3.1] K surpassant ainsi les performances du *prior* pour la plupart des classes.

Il faut noter que des données de validation Landsat7 à plus fine résolution spatiale (30m) étaient aussi disponibles en 2009 sur la région d'étude. Cependant la validation des résultats de la désagrégation des séries temporelles Météosat n'a pas été possible à cause de problèmes d'inter-étalonnage des capteurs Landsat7 et Météosat. Toutefois, une validation qualitative sur une zone de la carte d'occupation des sols fournie en 2009 a montré un bon accord en ce qui concerne la distribution spatiale de la température à la résolution de Landsat7 à l'exception de quelques parties de la zone caractérisées par de fortes hétérogénéités spatiales ou mal caractérisées dans les données de forçage (ces données ont été interpolées grâce à des cartes de NDVI disponibles sur la région d'étude pour l'année 2009).

### 7.2 Limites et difficultés

Globalement, les résultats obtenus pour la désagrégation des séries temporelles Météosat par lissage particulière à des résolutions plus fines (MODIS  $\approx \times 15$ , ASTER  $\approx \times 1800$ ) ont montré des performances satisfaisantes comparées à l'état de l'art, mais de nombreuses contraintes et difficultés ont été rencontrées dans chacune des expériences menées.

La première difficulté était l'absence de données de validation à l'échelle de la classe de végétation bien inter-étalonnées avec les données Météosat. Cette difficulté a pu être surmontée dans le cas où les distributions des températures à HRS et BRS des données de validation n'étaient pas très différentes (cas des données MODIS par opposition aux données Landsat7).

La deuxième difficulté était l'absence de mesures de données de forçage liées aux caractéristiques des végétations pour certaines classes de végétations présentes dans les cartes d'occupation des sols considérées dans les expériences menées sur la région d'étude. En effet, pour simuler la température d'une classe donnée avec SETHYS, il est indispensable d'avoir le forçage des caractéristiques des végétations (LAI et hauteur de végétation) de cette classe. Ces forçages permettent en effet de décrire la variabilité spatiale et temporelle du type de végétation en question. Pour surmonter cette difficulté, deux solutions ont été proposées :

1. La première solution était de regrouper certaines classes dont les forçages ne sont pas disponibles avec d'autres classes dont on dispose des forçages et dont les conditions hydriques et les distributions spatiales des températures à l'échelle des données de validation sont proches. Cependant en modifiant la classification définie par la carte d'occupation des sols, on rajoute implicitement des erreurs sur l'estimation des températures non prises en compte dans notre approche de désagrégation.
2. La deuxième solution, était d'interpoler les données de LAI à partir des données de NDVI à la résolution spatiale de la carte d'occupation des sols. Cette relation étant sensible aux types de couverts considérés (Bsaibes et al. (2009)), les

couverts très faiblement représentés sur la carte d'occupation des sols et dont les relations NDVI-LAI n'étaient pas bien calibrés ont donné des LAI très faibles, probablement sous-estimés. Ce problème a été surtout rencontré pour la classe "cultures" en 2009 avec des LAI interpolés très faibles durant toute l'année. Une autre solution aurait été de regrouper cette classe avec une autre classe dont les caractéristiques hydriques et des végétations étaient proches. Cependant, ne connaissant pas exactement la nature des couverts présents dans cette classe, il nous a été difficile de les regrouper avec une autre classe.

Une troisième difficulté était la faible sensibilité, sur la température, des paramètres pour les classes de végétation qui présentent des conditions d'humidité de sol très proches de la saturation : la forêt, le riz et la prairie. Pour ces classes, la variation des paramètres sensibles donne des variances très faibles des ensembles de températures simulées. Cette faible variance fait que la correction apportée par le lisseur en assimilant les observations BRS est presque nulle et peut même dégrader les performances de la méthode si l'erreur d'observation est plus importante que la variance du *prior*.

Par ailleurs, il faut signaler que l'estimation des températures à l'échelle de la classe de végétation par notre approche a trois grandes limitations :

- la forte hétérogénéité spatiale du pixel IRT à l'échelle de la carte d'occupation des sols (les meilleurs résultats étant obtenus pour les pixels comportant une classe majoritaire),
- la non résolution des corrélations spatiales à l'échelle des données Météosat qui se manifeste par des artéfacts sur les cartes de températures HRS estimées par notre approche,
- la nécessité de refaire une analyse de sensibilité pour identifier les paramètres sensibles aux différents couverts chaque fois que l'on change de région d'étude.

### 7.3 Perspectives

Plusieurs perspectives peuvent être envisagées à la suite de ce travail pour améliorer les performances de la méthode et explorer de nouvelles pistes. La première perspective est l'amélioration de la carte d'occupation des sols supposée stationnaire durant toute la période d'assimilation. Ceci pourrait être envisagé en utilisant des données auxiliaires à HRS (des cartes de NDVI par exemple) qui permettraient d'identifier les erreurs de classification et éventuellement améliorer la connaissance des fractions de couvert à la résolution spatiale des observations et leur variation temporelle. Une deuxième alternative serait de considérer des incertitudes sur les fractions de couverts fournies par la carte d'occupation des sols (supposée parfaitement connue jusqu'à présent).

Une deuxième perspective serait de lever partiellement l'hypothèse de connectivité en attribuant une distribution *a priori* aux températures HRS. En effet, l'idée est d'explorer les distributions estimées, par le lisseur particulière, pour les températures des différentes classes après assimilation des observations Météosat. Jusqu'à présent, la solution de désagrégation a été définie comme étant la moyenne des températures sélectionnées par le lisseur pour une classe donnée. De cette façon on n'utilise qu'une partie de la solution donnée par le lisseur. Il serait donc intéressant de voir si en prescrivant la distribution estimée par le lisseur aux températures des classes présentes dans les pixels BRS, on arrive à améliorer l'estimation des températures à HRS et à reproduire la variation de la distribution spatiale de la température à l'échelle de la classe de végétation.

Une autre perspective à ce travail serait d'explorer l'avantage d'avoir un modèle de surface continentale qui simule non seulement la température mais aussi d'autres variables du sol et des végétations au pas de temps auquel on souhaite échantillonner ses sorties (10 min, 30 min, 1 heure, etc). L'idée est de voir si en assimilant les observations BRS, on arrive à estimer d'autres variables simulées par le modèle à la résolution spatiale de la classe de végétation, comme par exemple l'humidité du sol ou l'évapotranspiration, sans pour autant avoir des observations de ces variables (en assimilant uniquement les températures Météosat).

Une dernière perspective est d'assimiler des données multi-échelles ou multi-variables ou les deux à la fois pour mieux contraindre le modèle avec des observations multiples. Ceci pourrait être envisagé en combinant par exemple des observations MODIS et Météosat, ou en assimilant des données micro-ondes d'humidité du sol et des données Météosat IRT, sous réserve de la possibilité d'inter-étalonner les données des différentes variables/résolutions assimilées.

Finalement, il faut noter que la méthode qu'on propose pour désagréger les données IRT Météosat peut être appliquée à tout type de couvert et pas seulement à des couverts de type agricole, et à tout type de variable si l'on dispose d'un modèle qui permet de simuler ces dernières à la résolution souhaitée après désagrégation.



---

## Bibliographie

- Ades, M. and P. J. van Leeuwen. An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, 139(672) :820–840, 2013. ISSN 00359009. doi :10.1002/qj.1995.
- Agam, N., W. P. Kustas, M. C. Anderson, F. Li, and P. D. Colaizzi. Utility of thermal sharpening over Texas high plains irrigated agricultural fields. *Journal of Geophysical Research D : Atmospheres*, 112, 2007a. ISSN 01480227. doi :10.1029/2007JD008407.
- Agam, N., W. P. Kustas, M. C. Anderson, F. Li, and P. D. Colaizzi. Utility of thermal image sharpening for monitoring field-scale evapotranspiration over rainfed and irrigated agricultural regions. *Geophysical Research Letters*, 35, 2008. ISSN 00948276. doi : 10.1029/2007GL032195.
- Agam, N., W. P. Kustas, M. C. Anderson, F. Li, and C. M. U. Neale. A vegetation index based technique for spatial sharpening of thermal imagery. *Remote Sensing of Environment*, 107 :545–558, 2007b. ISSN 00344257. doi :10.1016/j.rse.2006.10.006.
- Aires, F., C. Prigent, W. B. Rossow, and M. Rothstein. A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations. *J. Geophys. Res.*, 106(D14) :14887, 2001. ISSN 0148-0227. doi : 10.1029/2001JD900085.
- Anderson, M. C., J. Norman, W. Kustas, . Houborg, P. Starks, and N. Agam. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. *Remote Sens. Environ.*, 112(12) :4227–4241, 2008. ISSN 00344257. doi :10.1016/j.rse.2008.07.009.
- Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2) :174–188, 2002. ISSN 1053587X. doi :10.1109/78.978374.
- Asrar, G., M. Fuchs, E. T. Kanemasu, and J. L. Hatfield. Estimating Absorbed Photosynthetic Radiation and Leaf Area Index from Spectral Reflectance in Wheat1. *Agron. J.*, 76(2) :300, 1984. ISSN 0002-1962. doi :10.2134/agronj1984.000219622007600020029x.



- Baehr, C. and O. Pannekoucke. Some issues and results on the EnKF and particle filters for meteorological models. In C. Skiadas and I. Dimotikalis, editors, *Chaotic Systems : Theory and Applications*. World Scientific, 2010. ISBN 9789814299725.
- Ball, J. *An analysis of stomatal conductance*. Ph.D. thesis, Stanford University, 1988.
- Baret, F. and G. Guyot. Potentials and limits of vegetation indices for LAI and APAR assessment. 1991. doi :10.1016/0034-4257(91)90009-U.
- Bartoli, N. and P. Del Moral. *Simulation et Algorithmes Stochastiques*. Cepadues, 2011.
- Bechtel, B., K. Zakšek, and G. Hoshyaripour. Downscaling land surface temperature in an urban area : A case study for Hamburg, Germany. *Remote Sensing*, 4 :3184–3200, 2012. ISSN 2072-4292. doi :10.3390/rs4103184.
- Becker, F. *Observation de la Terre par Télédétection : Fondements Physiques, Méthodologiques et Technologiques*. Number vol. 1 in *Observation de la Terre par télédétection : fondements physiques, méthodologiques et technologiques*. Editions Universitaires Européennes, EUE, 2011. ISBN 9786131565090.
- Becker, F. and Z. L. Li. Surface temperature and emissivity at various scales : Definition, measurement and related problems. *Remote Sens. Rev.*, 12(3-4) :225–253, 1995. ISSN 0275-7257. doi :10.1080/02757259509532286.
- Becker, F., Z. L. Li, F. Beeker, and A. Li. Toward a local split window method over land surface. *Int. J. Remote Sens*, 11(3) :369–393, 1990. ISSN 0143-1161. doi : 10.1080/01431169008955028.
- Ben Mehrez, M., O. Taconet, D. Vidal-Madjar, and C. Valencogne. Estimation of stomatal resistance and canopy evaporation during the HAPEX-MOBILHY experiment. 1992. doi :10.1016/0168-1923(92)90066-D.
- Bernard, R., J. V. Soares, and D. Vidal-Madjar. Differential Bare Field Drainage Properties From Airborne Microwave Observations. 1986. doi :10.1029/WR022i006p00869.
- Bindhu, V. M., B. Narasimhan, and K. P. Sudheer. Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. *Remote Sensing of Environment*, 135 :118–129, 2013. ISSN 00344257. doi : 10.1016/j.rse.2013.03.023.
- Bourgois, L., G. Roussel, and M. Benjelloun. Kalman D'ensemble état-paramètres appliqué au modèle de Lorenz. In *JIME'2011*. France, 2011.
- Brisson, N., F. Bussièrre, H. Ozier-Lafontaine, R. Tournebize, and H. Sinoquet. Adaptation of the crop model STICS to intercropping. Theoretical basis and parameterisation. *Agronomie*, 24(6-7) :409–421, 2004. ISSN 0249-5627. doi :10.1051/agro:2004031.
- Brisson, N., C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussièrre, Y. M. Cabidoche, P. Cellier, P. Debaeke, J. P. Gaudillère, C. Hénault, F. Maraux, B. Seguin, and H. Sinoquet. An overview of the crop model STICS. In *Eur. J. Agron.*, volume 18, pages 309–332. 2003. ISBN 1161-0301. ISSN 11610301. doi :10.1016/S1161-0301(02)00110-7.

- Brisson, N., B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Durr, G. Richard, N. Beaudoin, S. Recous, X. Tayot, D. Plenet, P. Cellier, J.-M. Machet, J. M. Meynard, and R. Delécolle. STICS : a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. 1998. doi : 10.1051/agro:19980501.
- Bsaibes, A., D. Courault, F. Baret, M. Weiss, A. Olioso, F. Jacob, O. Hagolle, O. Marloie, N. Bertrand, V. Desfond, and F. Kzemipour. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. *Remote Sensing of Environment*, 113(4) :716–729, 2009. ISSN 00344257. doi :10.1016/j.rse.2008.11.014.
- Burgers, G., P. Van Leeuwen, and G. Evensen. Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126 :1719–1724, 1998.
- Chan, K., A. Saltelli, and S. Tarantola. Sensitivity analysis of model output : variance-based methods make the difference. In *Proceedings of the 1997 Winter Simulation Conference*, pages 261–268. 1997. doi :10.1109/WSC.1997.640407.
- Chen, X., Y. Yamaguchi, J. Chen, and Y. Shi. Scale effect of vegetation index based thermal sharpening : A simulation study based on aster data. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, volume 9, pages 1235–1238. Graduate School of Environmental Studies, Nagoya University, 464-8601, Japan, IEEE, 2011. ISBN 978-1-4577-1003-2. ISSN 21536996. doi :10.1109/IGARSS.2011.6049422.
- Choudhury, B., N. Ahmed, S. Idso, R. Reginato, and C. Daughtry. Relations between evaporation coefficients and vegetation indices studied by model simulations. *Remote Sens. Environ.*, 50 :1–17, 1994. ISSN 00344257. doi :10.1016/0034-4257(94)90090-6.
- Christopher Frey, H. and S. R. Patil. Identification and Review of Sensitivity Analysis Methods. *Risk Analysis*, 22(3) :553–578, 2002. ISSN 1539-6924. doi :10.1111/0272-4332.00039.
- Coudert, B. *Apport des mesures de température de surface par télédétection infrarouge thermisue pour la modélisation des échanges d'énergie et d'eau à l'interface sol végétation atmosphère*. Ph.D. thesis, Université de Versailles Saint Quentin, 2006.
- Coudert, B. and C. Ottlé. An improved SVAT model calibration strategy based on the optimisation of surface temperature temporal dynamics. *Geophysical Research Letters*, 34, 2007. ISSN 00948276. doi :10.1029/2006GL028778.
- Coudert, B., C. Ottlé, B. Boudevillain, J. Demarty, and P. Guillevic. Contribution of Thermal Infrared Remote Sensing Data in Multiobjective Calibration of a Dual-Source SVAT Model. *Journal of Hydrometeorology*, 7(3) :404–420, 2006. ISSN 1525-755X. doi :10.1175/JHM503.1.
- Coudert, B., C. Ottlé, and X. Briottet. Monitoring land surface processes with thermal infrared data : Calibration of SVAT parameters based on the optimisation of diurnal surface temperature cycling features. *Remote Sensing of Environment*, 112(3) :872–887, 2008. ISSN 00344257. doi :10.1016/j.rse.2007.06.024.
- Courault, D., A. Bsaibes, E. Kpemlie, R. Hadria, O. Hagolle, O. Marloie, J.-F. Hanocq, A. Olioso, N. Bertrand, and V. Desfonds. Assessing the Potentialities of FORMOSAT-2 Data for Water and Crop Monitoring at Small Regional Scale in South-Eastern France. *Sensors*, 8(5) :3460–3481, 2008. ISSN 1424-8220. doi :10.3390/s8053460.

- Courault, D., R. Hadria, F. Ruget, A. Oliosio, B. Duchemin, O. Hagolle, and G. Dedieu. Combined use of FORMOSAT-2 images with a crop model for biomass and water monitoring of permanent grassland in Mediterranean region. *Hydrol. Earth Syst. Sci.*, 14(9) :1731–1744, 2010. ISSN 1607-7938. doi :10.5194/hess-14-1731-2010.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal Royal Meteorological Society*, 120(519) :1367–1387, 1994. doi :10.1002/qj.49712051912.
- Crow, W. T. and E. F. Wood. The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering : A case study based on ESTAR measurements during SGP97. *Adv. Water Resour.*, 26 :137–149, 2003. ISSN 03091708. doi :10.1016/S0309-1708(02)00088-X.
- Dahia, K. *Nouvelles méthodes en filtrage particulière Application au recalage de navigation inertielle par mesures altimétriques*. Ph.D. thesis, Université Joseph Fourier, 2005.
- de Rocquigny, E., N. Devictor, and S. Tarantola. *Uncertainty in Industrial Practice : A Guide to Quantitative Uncertainty Management*. Wiley, 2008. ISBN 9780470770740.
- Deardorff, J. W. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. 1978. doi :10.1029/JC083iC04p01889.
- Decharme, B. and H. Douville. Introduction of a sub-grid hydrology in the ISBA land surface model. *Climate Dynamics*, 26 :65–78, 2006. ISSN 09307575. doi :10.1007/s00382-005-0059-7.
- Del Moral, P. *Feynman-Kac Formulae : Genealogical and Interacting Particle Systems with Applications*. Probability and Its Applications. Springer, 2004. ISBN 9780387202686.
- Demarty, J., C. Ottlé, I. Braud, A. Oliosio, J. P. Frangi, H. V. Gupta, and L. A. Bastidas. Constraining a physically based Soil-Vegetation-Atmosphere Transfer model with surface water content and thermal infrared brightness temperature measurements using a multiobjective approach. *Water Resour. Res.*, 41 :1–15, 2005. ISSN 00431397. doi : 10.1029/2004WR003695.
- Deng, C. and C. Wu. Examining the impacts of urban biophysical compositions on surface urban heat island : A spectral unmixing and thermal mixing approach. *Remote Sensing of Environment*, 131 :262–274, 2013. ISSN 00344257. doi :10.1016/j.rse.2012.12.020.
- Dennison, P. E., K. Charoensiri, D. Roberts, S. Peterson, and R. Green. Wildfire temperature and land cover modeling using hyperspectral data. 2006. doi : 10.1016/j.rse.2005.10.007.
- Dennison, P. E. and D. S. Matheson. Comparison of fire temperature and fractional area modeled from SWIR, MIR, and TIR multispectral and SWIR hyperspectral airborne data. *Remote Sensing of Environment*, 115 :876–886, 2011. ISSN 00344257. doi : 10.1016/j.rse.2010.11.015.
- Diak, G. R., J. R. Mecikalski, M. C. Anderson, J. M. Norman, W. P. Kustas, R. D. Torn, and R. L. DeWolf. Estimating Land Surface Energy Budgets From Space : Review and Current Efforts at the University of Wisconsin—Madison and USDA-ARS. 2004. doi :10.1175/BAMS-85-1-65.

- Dominguez, A., J. Kleissl, J. C. Luvall, and D. L. Rickman. High-resolution urban thermal sharpener (HUTS). *Remote Sensing of Environment*, 115(7) :1772–1780, 2011. ISSN 00344257. doi :10.1016/j.rse.2011.03.008.
- Douc, R., A. Garivier, E. Moulines, and J. Olsson. On the Forward Filtering Backward Smoothing particle approximations of the smoothing distribution in general state spaces models. *Annals of Statistics*, 2009.
- Doucet, A., S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3) :197–208, 2000. doi :10.1023/A:1008935410038.
- Dousset, B. and F. Gourmelon. Satellite multi-sensor data analysis of urban surface temperatures and landcover. In *ISPRS Journal of Photogrammetry and Remote Sensing*, volume 58, pages 43–54. 2003. ISBN 0924-2716. ISSN 09242716. doi : 10.1016/S0924-2716(03)00016-9.
- Ermida, S. L., I. F. Trigo, C. C. DaCamara, F. M. Göttsche, F. S. Olesen, and G. Hulley. Validation of remotely sensed surface temperature over an oak woodland landscape - The problem of viewing and illumination geometries. *Remote Sensing of Environment*, 148 :16–27, 2014. ISSN 00344257. doi :10.1016/j.rse.2014.03.016.
- Essa, W., B. Verbeiren, J. van der Kwast, T. Van de Voorde, and O. Batelaan. Evaluation of the DisTrad thermal sharpening methodology for urban areas. *International Journal of Applied Earth Observation and Geoinformation*, 19 :163–172, 2012. ISSN 15698432. doi :10.1016/j.jag.2012.05.010.
- Evensen, G. Inverse methods and data assimilation in nonlinear ocean models. *Physica D : Nonlinear Phenomena*, 77(1–3) :108 – 129, 1994a. ISSN 0167-2789. doi :http://dx.doi.org/10.1016/0167-2789(94)90130-9.
- Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research : Oceans*, 99(C5) :10143–10162, 1994b. ISSN 2156-2202. doi : 10.1029/94JC00572.
- Evensen, G. The Ensemble Kalman Filter : theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4) :343–367, 2003. ISSN 1616-7341. doi : 10.1007/s10236-003-0036-9.
- Farquhar, G. D., S. von Caemmerer, and J. A. Berry. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta*, 149(1) :78–90, 1980. ISSN 0032-0935. doi :10.1007/BF00386231.
- Faure, H. and S. Tezuka. Another random scrambling of digital (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, chapter Proceeding, pages 242–256. Springer Berlin Heidelberg, 2002. ISBN 978-3-540-42718-6. doi :10.1007/978-3-642-56046-0\_16.
- Faysash, D. A. and E. A. E. Smith. Simultaneous land surface temperature-emissivity retrieval in the infrared split window. *J. Atmos. Ocean. ...*, 16(11) :1673–1689, 1999. ISSN 0739-0572. doi :10.1175/1520-0426(1999)016<1673:SLSTER>2.0.CO;2.
- Fischler, M. A. and R. C. Bolles. Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6) :381–395, 1981. ISSN 0001-0782. doi :10.1145/358669.358692.

- François, C. *Détermination de la température de surface à partir de mesures dans l'infrarouge thermique : Corrections atmosphériques et étude des propriétés directionnelles de l'émissivité et de la température radiative*. Ph.D. thesis, Université de Paris VII, 1995.
- Freitas, S. C., I. F. Trigo, J. M. Bioucas-Dias, and F. M. Göttsche. Quantifying the uncertainty of land surface temperature retrievals from SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 48 :523–534, 2010. ISSN 01962892. doi :10.1109/TGRS.2009.2027697.
- Guillevic, P. C., A. Bork-Unkelbach, F. M. Göttsche, G. Hulley, J.-P. Gastellu-Etchegorry, F. S. Olesen, and J. L. Privette. Directional Viewing Effects on Satellite Land Surface Temperature Products Over Sparse Vegetation Canopies—A Multisensor Analysis. *IEEE Geosci. Remote Sens. Lett.*, 10(6) :1464–1468, 2013. ISSN 1545-598X. doi :10.1109/LGRS.2013.2260319.
- Haykin, S., editor. *Kalman Filtering and Neural Networks*. John Wiley & Sons, Inc., New York, USA, 2001. ISBN 9780471221548. doi :10.1002/0471221546.
- Hill, T. C., T. Quaife, and M. Williams. A data assimilation method for using low-resolution Earth observation data in heterogeneous ecosystems. *Journal of Geophysical Research*, 116(D8) :D08117, 2011. ISSN 0148-0227. doi :10.1029/2010JD015268.
- Hlinka, O., O. Sluciak, F. Hlawatsch, P. M. Djuric, and M. Rupp. Likelihood consensus : Principles and application to distributed particle filtering. *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, pages 349–353, 2010. doi :10.1109/ACSSC.2010.5757533.
- Hoeffding, W. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3) :293–325, 1948. ISSN 0003-4851. doi :10.1214/aoms/1177730196.
- Hol, J. D. Resampling in particle filters. *Institutionen för systemteknik, Linköping University, Department of Electrical Engineering*, (0283) :44, 2004.
- Hol, J. D., T. B. Schon, and F. Gustafsson. On Resampling Algorithms for Particle Filters. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82. IEEE, 2006. ISBN 978-1-4244-0579-4. doi :10.1109/NSSPW.2006.4378824.
- Homma, T. and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52 :1–17, 1996.
- Hoteit, I., X. Luo, and D.-T. Pham. Particle Kalman Filtering : A Nonlinear Bayesian Framework for Ensemble Kalman Filters\*. *Monthly Weather Review*, 140(2) :528–542, 2012. ISSN 0027-0644. doi :10.1175/2011MWR3640.1.
- Houtekamer, P. L. and H. Mitchell. Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review*, 126 :796–811, 1998.
- Idso, S. B. A set of equations for full spectrum and 8- to 14- $\mu\text{m}$  and 10.5- to 12.5 - $\mu\text{m}$  thermal radiation from cloudless skies . *Water Resources Research*, 17(2) :295–304, 1981.
- Inamdar, A. K. and A. French. Disaggregation of GOES land surface temperatures using surface emissivity. *Geophysical Research Letters*, 36(2) :n/a–n/a, 2009. ISSN 00948276. doi :10.1029/2008GL036544.

- Inamdar, A. K., A. French, S. Hook, G. Vaughan, and W. Luekkett. Land surface temperature retrieval at high spatial and temporal resolutions over the southwestern United States. *Journal of Geophysical Research*, 113(D7) :D07107, 2008. ISSN 0148-0227. doi :10.1029/2007JD009048.
- Iooss, B. Revue sur l'analyse de sensibilité globale de modèles numériques. *Journal de la Société Française de Statistique*, 152(1) :3–25, 2011.
- Jazwinski, A. *Stochastic Processes and Filtering Theory*. Mathematics in Science and Engineering. Elsevier Science, 1970. ISBN 9780080960906.
- Jeganathan, C., N. Hamm, S. Mukherjee, P. Atkinson, P. Raju, and V. Dadhwal. Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *International Journal of Applied Earth Observation and Geoinformation*, 13(2) :178–191, 2011. ISSN 03032434. doi :10.1016/j.jag.2010.11.001.
- Jones, H. *Plants and Microclimate : A Quantitative Approach to Environmental Plant Physiology*. Cambridge University Press, 1983. ISBN 9781107511637.
- Kallel, A., C. Ottlé, S. Le Hegarat-Masclé, F. Maignan, and D. Courault. Surface Temperature Downscaling From Multiresolution Instruments Based on Markov Models. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3) :1588–1612, 2013. ISSN 0196-2892. doi :10.1109/TGRS.2012.2207461.
- Kalma, J., T. McVicar, and M. McCabe. Estimating Land Surface Evaporation : A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surveys in Geophysics*, 29 :421–469, 2008. doi :10.1007/s10712-008-9037-z.
- Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1) :35, 1960. ISSN 00219223. doi :10.1115/1.3662552.
- Karnieli, A., M. Bayasgalan, Y. Bayarjargal, N. Agam, S. Khudulmur, and C. J. Tucker. Comments on the use of the Vegetation Health Index over Mongolia. 2006. doi : 10.1080/01431160500121727.
- Kustas, W. P., J. M. Norman, M. C. Anderson, and A. N. French. Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sensing of Environment*, 85 :429–440, 2003. ISSN 00344257. doi :10.1016/S0034-4257(03)00036-1.
- Le Dimet, F.-X. and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A*, 38(2) :97–110, 1986.
- Liu, D. and R. Pu. Downscaling Thermal Infrared Radiance for Subpixel Land Surface Temperature Retrieval. *Sensors*, 8(4) :2695–2706, 2008. ISSN 1424-8220. doi : 10.3390/s8042695.
- Liu, Y., T. Hiyama, and Y. Yamaguchi. Scaling of land surface temperature using satellite data : A case examination on ASTER and MODIS products over a heterogeneous terrain area. *Remote Sensing of Environment*, 105(2) :115–128, 2006. ISSN 00344257. doi :10.1016/j.rse.2006.06.012.
- Mechri, R., C. Ottlé, O. Pannekoucke, and A. Kallel. Genetic particle filter application to land surface temperature downscaling. *Journal of Geophysical Research : Atmospheres*, 119(5) :2131–2146, 2014. ISSN 2169897X. doi :10.1002/2013JD020354.

- Merlin, O., B. Duchemin, O. Hagolle, F. Jacob, B. Coudert, G. Chehbouni, G. Dedieu, J. Garatuza, and Y. Kerr. Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sensing of Environment*, 114(11) :2500–2512, 2010. ISSN 00344257. doi :10.1016/j.rse.2010.05.025.
- Merlin, O., F. Jacob, J.-P. Wigneron, J. Walker, and G. Chehbouni. Multidimensional Disaggregation of Land Surface Temperature Using High-Resolution Red, Near-Infrared, Shortwave-Infrared, and Microwave-L Bands. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5) :1864–1880, 2012. ISSN 0196-2892. doi : 10.1109/TGRS.2011.2169802.
- Mira, M., A. Olioso, V. Rivalland, D. Courault, O. Marolie, and P. Guillevic. Quantifying uncertainties in land surface temperature due to atmospheric correction : Application to Landsat-7 data over a Mediterranean agricultural region. In *IGARSS 2014*, pages 2375–2378. 2014.
- Mitchell, P. L., P. Houtekamer, and L. Herschel. Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 126(1969) :796–811, 1998.
- Monin, A. and A. Obukhov. Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, 24 :163–187, 1954.
- Monteith, J. L. and G. Szicz. The flux of CO<sub>2</sub> over a field of sugar beet. *Q. J. R. Meteorol. Soc.*, 87(371) :112–113, 1961. ISSN 00359009. doi :10.1002/qj.49708737115.
- Montzka, C., V. R. N. Pauwels, H.-J. H. Franssen, X. Han, and H. Vereecken. Multivariate and multiscale data assimilation in terrestrial systems : a review. *Sensors*, 12(12) :16291–333, 2012. doi :10.3390/s121216291.
- Mood, A. M., F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*, volume 3. Mc Graw Hill, 1974. ISBN 0070854653.
- Nichol, J. An Emissivity Modulation Method for Spatial Enhancement of Thermal Satellite Images in Urban Heat Island Analysis. *Photogrammetric Engineering and Remote Sensing*, 75 :547–556, 2009. ISSN 00991112 (ISSN).
- Norman, J. M., M. C. Anderson, W. P. Kustas, a. N. French, J. Mecikalski, R. Torn, G. R. Diak, T. J. Schmugge, and B. C. W. Tanner. Remote sensing of surface energy fluxes at 10 m pixel resolutions. *Water Resources Research*, 39 :n/a–n/a, 2003. ISSN 00431397. doi :10.1029/2002WR001775.
- Olioso, A., T. N. Carlson, and N. Brisson. Simulation of diurnal transpiration and photosynthesis of a water stressed soybean crop. *Agric. For. Meteorol.*, 81(1-2) :41–59, 1996. ISSN 01681923. doi :10.1016/0168-1923(95)02297-X.
- Ottlé, C., A. Kallel, G. Monteil, S. LeHegarat, and B. Coudert. Subpixel Temperature Estimation from Low Resolution Thermal Infrared Remote Sensing. *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, 3(1) :III – 403–III – 406, 2008. doi :10.1109/IGARSS.2008.4779369.
- Ottlé, C. and M. Stoll. Effect of atmospheric absorption and surface emissivity on the determination of land surface temperature from infrared satellite data. 1993. doi : 10.1080/01431169308954018.
- Ottlé, C. and D. Vidal-Madjar. Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region. 1994. doi : 10.1016/0022-1694(94)90056-6.

- Owen, A. B. Scrambling Sobol' and Niederreiter–Xing Points. *Journal of Complexity*, 14(4) :466–489, 1998. ISSN 0885064X. doi :10.1006/jcom.1998.0487.
- Pan, M., E. F. Wood, D. B. McLaughlin, D. Entekhabi, and L. Luo. A Multiscale Ensemble Filtering System for Hydrologic Data Assimilation. Part I : Implementation and Synthetic Experiment. *Journal of Hydrometeorology*, 10(3) :794–806, 2009. ISSN 1525-755X. doi :10.1175/2009JHM1088.1.
- Perrier, A. Étude physique de l'évapotranspiration dans les conditions naturelles. *Annales agronomiques*, 26(3), 1975.
- Pieri, D. and M. Abrams. ASTER observations of thermal anomalies preceding the April 2003 eruption of Chikurachki volcano, Kurile Islands, Russia. *Remote Sensing of Environment*, 99 :84–94, 2005. ISSN 00344257. doi :10.1016/j.rse.2005.06.012.
- Quaife, T., P. Lewis, M. De Kauwe, M. Williams, B. Law, M. Disney, and P. Bowyer. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112 :1347–1364, 2008.
- Quattrochi, D. and J. Luvall. *Thermal Remote Sensing in Land Surface Processes*. Taylor & Francis, 2004. ISBN 9780203502174.
- Rémy, S., O. Pannekoucke, T. Bergot, and C. Baehr. Adaptation of a particle filtering method for data assimilation in a 1D numerical model used for fog forecasting. *Quarterly Journal of the Royal Meteorological Society*, 138(663) :536–551, 2012. ISSN 00359009. doi :10.1002/qj.915.
- Rosenthal, W. D., E. T. Kanemasu, R. J. Raney, and L. R. Stone. Evaluation of an Evapotranspiration Model for Corn1. *Agron. J.*, 69(3) :461, 1977. ISSN 0002-1962. doi :10.2134/agronj1977.00021962006900030032x.
- Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2) :280–297, 2002. ISSN 00104655. doi : 10.1016/S0010-4655(02)00280-1.
- Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2) :259–270, 2010. ISSN 00104655. doi :10.1016/j.cpc.2009.09.018.
- Sandholt, I., K. Rasmussen, and J. Andersen. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.*, 79(2-3) :213–224, 2002. ISSN 00344257. doi : 10.1016/S0034-4257(01)00274-7.
- Saux Picart, S. *Etude couplée modélisation/téledétection pour le suivi des états et bilans hydriques de surface en zone sahélienne*. Ph.D. thesis, Université Pierre et Marie Curie, 2007.
- Saux-Picart, S., C. Otlé, B. Decharme, C. André, M. Zribi, A. Perrier, B. Coudert, N. Boulain, B. Cappelaere, L. Descroix, and D. Ramier. Water and energy budgets simulation over the AMMA-Niger super-site spatially constrained with remote sensing data. *Journal of Hydrology*, 375 :287–295, 2009a. ISSN 00221694. doi : 10.1016/j.jhydrol.2008.12.023.



- Saux-Picart, S., C. Ottlé, A. Perrier, B. Decharme, B. Coudert, M. Zribi, N. Boulain, B. Cappelaere, and D. Ramier. SETHyS\_Savannah : A multiple source land surface model applied to Sahelian landscapes. *Agricultural and Forest Meteorology*, 149 :1421–1432, 2009b. ISSN 01681923. doi :10.1016/j.agrformet.2009.03.013.
- Seguin, B., F. Becker, T. Phulpin, X. F. Gu, G. Guyot, Y. Kerr, C. King, J. P. Lagouarde, C. Ottlé, M. P. Stoll, A. Tabbagh, and A. Vidal. IRSUTE : A minisatellite project for land surface heat flux estimation from field to regional scale. *Remote Sens. Environ.*, 68 :357–369, 1999. ISSN 00344257. doi :10.1016/S0034-4257(98)00122-9.
- Sheng, X. and Y. Hu. Distributed particle filters for wireless sensor network target tracking. *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 4 :845–848, 2005. doi :10.1109/ICASSP.2005.1416141.
- Snyder, C. Particle filters, the “optimal” proposal and high-dimensional systems. In *ECMWF Seminar on Data assimilation for atmosphere and ocean*, pages 6–9. 2011.
- Sobol, I. M. Sensitivity analysis for non-linear mathematical models. *Math. Model. Comput.*, 1(4) :407–414, 1993.
- Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3) :271–280, 2001. ISSN 03784754. doi :10.1016/S0378-4754(00)00270-6.
- Sobol, I. M., S. Tarantola, D. Gatelli, S. Kucherenko, and W. Mauntz. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety*, 92(7) :957–960, 2007. ISSN 09518320. doi : 10.1016/j.res.2006.07.001.
- Sorooshian S. and Gupta V.K. Model Calibration. In S. V.P., editor, *Comput. Model. watershed Hydrol.*, pages 23–67. Water Resources Pub., Highlands Ranch, 1995. ISBN 9780918334916.
- Stathopoulou, M. and C. Cartalis. Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sensing of Environment*, 113 :2592–2605, 2009. ISSN 00344257. doi :10.1016/j.rse.2009.07.017.
- Sun, D. Estimation of land surface temperature from a Geostationary Operational Environmental Satellite (GOES-8). *J. Geophys. Res.*, 108(D11) :4326, 2003. ISSN 0148-0227. doi :10.1029/2002JD002422.
- Sun, D. and M. Kafatos. Note on the NDVI-LST relationship and the use of temperature-related drought indices over North America. *Geophys. Res. Lett.*, 34, 2007. ISSN 00948276. doi :10.1029/2007GL031485.
- Sznitman, A.-S. Topics in propagation of chaos. In P.-L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251. Springer Berlin Heidelberg, 1991. ISBN 978-3-540-53841-7. doi : 10.1007/BFb0085169.
- Taconet, O., R. Bernard, and D. Vidal-Madjar. Evapotranspiration over an Agricultural Region Using a Surface Flux/Temperature Model Based on NOAA-AVHRR Data. *Journal of Climate and Applied Meteorology*, 25(3) :284–307, 1986. ISSN 0733-3021. doi :10.1175/1520-0450(1986)025<0284:EOAARU>2.0.CO;2.

- Thom, A. S. Momentum, mass and heat exchange of vegetation. *Quarterly Journal of the Royal Meteorological Society*, 98 :124–134, 1972. ISSN 00359009. doi :10.1002/qj.49709841510.
- Thompson, D. A dynamical method of analyzing meteorological data. *Tellus A*, 13 :334–349, 1961.
- Trigo, I. F., C. C. Dacamara, P. Viterbo, J.-L. Roujean, F. Olesen, C. Barroso, F. Camacho-de Coca, D. Carrer, S. C. Freitas, J. García-Haro, B. Geiger, F. Gellens-Meulenberghs, N. Ghilain, J. Meliá, L. Pessanha, N. Siljamo, and A. Arboleda. The Satellite Application Facility for Land Surface Analysis. 2011. doi :10.1080/01431161003743199.
- Van Leeuwen, P. J. Nonlinear data assimilation in geosciences : an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653) :1991–1999, 2010. ISSN 00359009. doi :10.1002/qj.699.
- Van Leeuwen, P. J. and G. Evensen. Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation. 1996. doi :10.1175/1520-0493(1996)124<2898:DAAIMI>2.0.CO;2.
- Voogt, J. A. and T. R. Oke. Thermal remote sensing of urban climates. 2003. doi : 10.1016/S0034-4257(03)00079-8.
- Wan, Z. and Z. Li. Radiance-based validation of the V5 MODIS land-surface temperature product. 2008. doi :10.1080/01431160802036565.
- Weng, Q., D. Lu, and J. Schubring. Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, 89 :467–483, 2004. ISSN 00344257. doi :10.1016/j.rse.2003.11.005.
- Wilson, T. B. and T. P. Meyers. Determining vegetation indices from solar and photosynthetically active radiation fluxes. *Agricultural and Forest Meteorology*, 144 :160–179, 2007. ISSN 01681923. doi :10.1016/j.agrformet.2007.04.001.
- Yuan, F. and M. E. Bauer. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote Sensing of Environment*, 106 :375–386, 2007. ISSN 00344257. doi : 10.1016/j.rse.2006.09.003.
- Zakšek, K. and K. Oštir. Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sensing of Environment*, 117 :114–124, 2012. ISSN 00344257 (ISSN). doi :10.1016/j.rse.2011.05.027.
- Zhan, W., Y. Chen, J. Zhou, J. Li, and W. Liu. Sharpening thermal imageries : A generalized theoretical framework from an assimilation perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 49 :773–789, 2011. ISSN 01962892. doi : 10.1109/TGRS.2010.2060342.
- Zhan, W., Y. Chen, J. Zhou, J. Wang, W. Liu, J. Voogt, X. Zhu, J. Quan, and J. Li. Disaggregation of remotely sensed land surface temperature : Literature survey, taxonomy, issues, and caveats. 2013. doi :10.1016/j.rse.2012.12.014.