



HAL
open science

Avancées méthodologiques dans l'estimation de la température de surface continentale par des observations microondes passives

Samuel Favrichon

► **To cite this version:**

Samuel Favrichon. Avancées méthodologiques dans l'estimation de la température de surface continentale par des observations microondes passives. Géophysique [physics.geo-ph]. Sorbonne Université, 2021. Français. NNT : 2021SORUS339 . tel-03591719

HAL Id: tel-03591719

<https://theses.hal.science/tel-03591719v1>

Submitted on 28 Feb 2022

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.

Avancées méthodologiques dans l'estimation de la température de surface continentale par des observations microondes passives

Une thèse présentée pour l'obtention du grade de
Docteur
de SORBONNE UNIVERSITÉ
ÉCOLE DOCTORALE DES SCIENCES DE L'ENVIRONNEMENT D'ÎLE-DE-FRANCE (N°129)

par Samuel FAVRICHON

Sous la direction de :

Directrice de thèse Catherine PRIGENT LERMA, Observatoire de Paris
Co-directeur de thèse Carlos JIMENEZ Estellus

Composition du Jury :

Richard DE JEU	VanderSat B.V.	Examineur
Patricia DE ROSNAY	European Center for Medium Range Weather Forecast	Rapportrice
Carlos JIMENEZ	Estellus	Directeur
Yann KERR	Centre d'Etudes Spatiales de la Biosphère	Examineur
Zhao-Liang LI	Université de Strasbourg	Rapporteur
Catherine PRIGENT	LERMA, Observatoire de Paris	Directrice
Isabel TRIGO	Instituto Português do Mar e da Atmosfera	Examinatrice



Déclaration

Travail effectué sous la direction de Catherine Prigent et Carlos Jimenez entre 2018 et 2021. Le laboratoire d'accueil a été le Laboratoire d'Études du Rayonnement et de la Matière en Astrophysique et Atmosphères (LERMA) sur le site de l'Observatoire de Paris.

L'encadrement de la formation a été fait par l'école doctorale des sciences de l'environnement d'Île-de-France n°129 à Sorbonne Université. Le financement de la thèse est un contrat CIFRE N°2018-0637 en co-tutelle avec l'entreprise Estellus.

Les financements suivants ont aussi été utilisés dans le cadre de cette thèse :

- European Space Agency (ESA) project LST-CCI (contrat no. 4000123553/18/I-NB).
- Centre National d'Études Spatiales (CNES) projet GPM-R.

« J'ai toujours aimé le désert. On s'assoit sur une dune de sable. On ne voit rien. On n'entend rien. Et cependant quelque chose rayonne en silence... »

ANTOINE DE SAINT-EXUPÉRY

Résumé

La température de la surface des continents est une variable essentielle dans la modélisation du bilan énergétique global. La température de la couche supérieure du sol, ou « température de peau », joue un rôle important dans l'interaction entre la Terre et l'atmosphère. Des mesures précises de la température de surface sont précieuses pour de nombreuses applications, comme les prévisions météorologiques, l'analyse de l'impact du changement climatique sur les écosystèmes ou la planification agricole. La température de la surface est généralement estimée à partir de mesures effectuées par des instruments observant dans le spectre infrarouge en orbite autour de la Terre. Cependant, une limitation majeure de ces mesures est que les radiations infrarouges ne peuvent pas pénétrer les nuages. Le rayonnement microonde, en revanche, peut pénétrer la plupart des nuages. Les imageurs microondes passifs en orbite offrent donc un ensemble de mesures complémentaires aux observations infrarouges.

Néanmoins, plusieurs difficultés surviennent lors de l'utilisation des observations microondes passives pour estimer les températures des surfaces continentales. De nouvelles méthodes ont été développées dans cette thèse pour surmonter certaines des difficultés rencontrées par la communauté scientifique. Les résultats obtenus prennent en compte l'interaction des microondes avec l'environnement physique, par exemple avec les nuages ou la sub-surface dans les zones arides. Les méthodes peuvent être intégrées dans les processus existants d'estimation de la température de la surface, afin de produire des estimations plus précises et pour compléter les séries temporelles.

Pour la climatologie le développement des séries temporelles repose sur des instruments multiples qui ont volés sur les dernières décennies. L'inter-étalonnage des différents instruments est essentiel pour fournir de longues séries de données. Le point de départ est d'obtenir des températures de brillance cohérentes sans artefacts instrumentaux évidents. Les instruments plus récents sont plus précis et peuvent être utilisés pour corriger l'étalonnage des mesures plus anciennes afin de les rendre utilisables. Le premier chapitre propose une correction pour inter-étalonner SMMR (1978-1987) avec GMI (2014-présent) sur les surfaces continentales, facilitant l'inclusion potentielle de ces premières observations dans les séries de données microondes à long terme.

Bien que les microondes dans les canaux fenêtres puissent observer la surface à travers la plupart des nuages, certaines mesures sont affectées par la présence de nuages. La détection des observations qui sont contaminées est donc nécessaire. Le signal des canaux contaminés varie en fonction de la fréquence d'observation et du type de nuage. Le deuxième chapitre décrit une méthode polyvalente pour estimer la probabilité de contamination par les nuages pour une observation multifréquence. Un index de contamination est prédit à partir d'un modèle statistique flexible, basé sur des réseaux de neurones. Cette méthode peut être appliquée à des fréquences comprises entre 10 et 183 GHz et être facilement intégrée dans les chaînes de traitement, pour une détection rapide des observations contaminées par des nuages.

Par rapport à l'infrarouge les observations microondes ont des longueurs d'onde plus grandes, et dans les régions arides, incluent une contribution de la sub-surface dans le signal mesuré. Dans certaines situations, la pénétration des microondes dans la sub-surface ne

peut pas être négligée. Par exemple, dans les zones arides, les microondes à basse fréquence (inférieure à 20 GHz) peuvent pénétrer dans le sol à une profondeur de plus de 15 cm. Comme la température varie dans la colonne de sol, la température de surface mesurée par les microondes n'est pas strictement la température de « peau ». Dans le troisième chapitre, cette question est étudiée et des cartes des zones affectées par l'émission de la sub-surface sont créées. Ces cartes peuvent être utilisées pour identifier les endroits où les températures dérivées des microondes doivent être corrigées afin de fournir une estimation plus cohérente avec celles basées sur l'infrarouge.

De plus, en raison de leur longueur d'onde plus grande, les microondes fournissent des mesures à une résolution spatiale moins bonne que l'infrarouge. Dans le quatrième chapitre, une nouvelle méthode pour augmenter la résolution spatiale des températures de surface à partir des observations microondes est décrite. Cette méthode utilise un terme de correction basé sur une relation statistique entre quatre variables explicatives : l'amplitude et la valeur minimale d'un cycle médian de température diurne, un indice de fraction de végétation et un angle représentant la position du soleil. Ce terme de correction permet de réduire l'échelle des observations SSMIS de leur résolution originale de 25 km à 5 km. Cette méthode peut permettre à la communauté scientifique de produire des températures de surface terrestre dérivées des microondes dans des conditions claires et nuageuses à des résolutions plus comparables aux estimations infrarouges.

Pour conclure, le dernier chapitre montre des séries temporelles de températures de surface terrestre par microondes actuellement disponibles. Une discussion suit sur la façon dont les nouvelles méthodes décrites dans cette thèse ont été et continueront à être incluses dans la production de ces séries de données de température de surface.

Summary

Land surface temperature, the radiative temperature of the skin surface, is an essential variable in modelling the global energy budget, playing an important role in the interaction between the land and the atmosphere. Accurate measurements of land surface temperature are critical for many applications, such as weather forecasting, climate change studies, or agricultural planning. Land surface temperature is commonly estimated by Earth-orbiting instruments observing in the infrared spectrum, but a major limitation is that infrared radiations cannot penetrate clouds. Microwave radiation, however, can penetrate most clouds, so passive microwave instruments offer a complementary set of measurements to infrared observations.

Nevertheless, several difficulties are encountered when estimating land surface temperature from passive microwave observations. These challenges are related to how microwaves interact with the physical environment, for example with clouds, or with the sub-surface in arid regions. To overcome them, new methods that can be used to improve and surface temperature retrievals are proposed and developed. The aim being to produce more accurate and complete microwave surface records for both the present day and going back several decades.

Longer time series rely on multiple instruments sources for the underlying observations. Inter-calibrating different instruments is critical to providing seamless data records. The starting point is to obtain consistent brightness temperatures free from obvious instrumental artifacts. Newer instruments are more accurate and can be used to correct the calibration of older measurements to make them more usable. The first chapter proposes a correction to inter-calibrate SMMR (1978-1987) with GMI (2014-present) over continental surfaces, facilitating the potential inclusion of these early SMMR observations into long-term microwave data records.

Although microwave observations in window channels see through most clouds, some measurements can be affected by clouds. The detection of cloud-contaminated observations is therefore needed. Depending on the observation frequency, and the cloud type, the impact on the observed brightness temperatures varies. The second chapter describes a versatile method developed to estimate the probability of cloud contamination for a given multi-frequency observation. A flexible statistical model based on neural networks provides the prediction. This method can be applied to frequencies between 10 and 183 GHz, and has been incorporated into processing chains for fast detection of cloud-contaminated observations.

Compared with infrared, the longer wavelengths of microwave observations in arid areas include a sub-surface contribution to the measured signal. This is mostly noticeable in arid areas, where low-frequency microwaves (below 20 GHz) can penetrate the soil to a depth of over 15 cm for some soil compositions. Because the temperature varies throughout the soil column, the surface temperature measured by the microwave is not strictly a skin temperature. In the third chapter, this issue is studied and maps of the areas affected by sub-surface emission are created. These maps can be used to identify where microwave-derived temperatures should be corrected to provide a more skin-like estimation consistent with infrared-based estimates.

Also, because of their longer wavelength, microwaves provide measurements at a lower spatial resolution than infrared. In the fourth chapter, a new method to increase the spatial resolution of land surface temperature maps derived from microwave observations is presented. This method uses a correction term based on a statistical relationship between four explanatory variables : the amplitude and minimal value of a median diurnal temperature cycle, a vegetation fraction index and an angle representing the sun position. This correction term allows to downscale SSMIS observations from their original 25 km resolution to 5 km, allowing to produce microwave derived land surface temperatures, in clear and cloudy conditions, at resolutions more comparable to the infrared estimates.

To conclude, the final chapter displays the currently available microwave land surface temperature time series. A discussion follows on how the new methods described in this dissertation have been and will continue to be included in the production of these valuable land surface temperature datasets.

Liste des publications et conférences

ARTICLES : *Les articles publiés sont disponibles en annexe (p. 181).*

1. Favrichon S., Prigent C., Jimenez C., (2022). Global estimation of microwave penetration depths between 10 and 90 GHz over arid areas. *En préparation*
2. Favrichon, S., Prigent, C., and Jiménez, C. (2021). A method to downscale satellite microwave land-surface temperature. *Remote Sensing*, 13(7) :1–20
3. Favrichon, S., Jimenez, C., and Prigent, C. (2020). Inter-calibrating SMMR brightness temperatures over continental surfaces. *Atmospheric Measurement Techniques*, 13(10) :5481–5490
4. Favrichon, S., Prigent, C., Jimenez, C., and Aires, F. (2019). Detecting cloud contamination in passive microwave satellite measurements over land. *Atmospheric Measurement Techniques*, 12(3) :1531–1543

CONFÉRENCES :

Juillet 2021 - International Geoscience and Remote Sensing Symposium (IGARSS) : « Mapping Microwave Penetration Depths Over Arid Areas »

Mai 2021 - CIMR Sci2Ops : « Land Surface Temperature retrievals with CIMR : benefits and opportunities »

Novembre 2020 - MicroRad : « All-weather land surface temperature retrieval from microwave observations with an improved spatial resolution »

Juin 2019 - CMSAF User Workshop : « On the use of CM SAF FCDRs to retrieve LST and other ECVs »

Remerciements

Je souhaiterais remercier tous les lecteurs qui prendront le temps de lire cette thèse. Tout d'abord les rapporteurs de ce travail Mme Patricia De Rosnay et M. Zhao-Liang Li, ainsi que les membres du jury, M. Richard De Jeu, M. Yann Kerr et Mme Isabel Trigo qui trouveront ici les détails de l'avancée actuelle de mes recherches.

Cette thèse n'aurait pas été possible sans l'exigence et l'accompagnement de mes directeurs de thèse, Carlos Jimenez et Catherine Prigent. Je leur suis grandement reconnaissant pour leur patience tout au long du travail. Grâce à leurs explications et à nos discussions, j'ai approfondi de nombreux sujets dont j'ignorais l'existence avant de les rejoindre à l'Observatoire.

Pour les différents chapitres des remerciements spécifiques sont dus à certaines personnes qui m'ont donné accès à leurs jeux de données ou aidé à comprendre les différentes problématiques rencontrées :

- Pour le second chapitre, à Fennig Karsten et Marc Schroder du Satellite Application Facility on Climate Monitoring. Les discussions concernant les versions disponibles des jeux de données inter-étalonnées et les développements en cours ont permis de former ce chapitre.
- Le troisième chapitre doit beaucoup à Filipe Aires, pour les conseils et les discussions sur les problèmes soulevés par l'utilisation de réseaux de neurones sur des observations satellites. Ainsi qu'à Hervé Le Gléau et Gaelle Kerdraon (Centre de Météorologie Spatiale de Météo-France) pour l'aide apportée dans l'utilisation des classifications de nuages de SEVIRI.
- Pour le quatrième chapitre, les discussions avec Cecile Ferrari pour comprendre les modèles de diffusion thermique dans la surface, et les échanges avec Patricia de Rosnay pour l'utilisation des données ERA5-Land ont été très utiles. Un remerciement va aussi à Roland Vogt pour la mise à disposition des données de la station Goabeb.
- Le travail de Sofia Ermida pour la création de la base de données de comparaison des températures en ciel clair et sous nuages a été très utile pour mettre en place la correction des paramètres détaillée dans cette section.

Bien sûr cette thèse est aussi redevable des travaux des nombreux chercheurs qui m'ont précédé et qui œuvrent pour améliorer notre compréhension et la description du monde qui nous entoure. A travers les cours reçus au Centre européen de prévision à météo à moyen terme, à Sorbonne Université et les présentations à différentes conférences j'ai pu avoir des points de vue différents sur le travail effectué. Je n'aurais bien sûr pas produit ce travail sans avoir pu bénéficier de nombreux enseignants passionnés depuis mes premières années de lycées jusqu'à mes études supérieures.

La réussite de cette thèse doit beaucoup à l'entourage à l'Observatoire de Paris. L'ambiance y était assurée par des collègues mémorables, qu'ils soient occupants du bureau 801bis ou plus largement des derniers étages du bâtiment A de l'Observatoire. Venir au

travail a toujours été un plaisir (entre les différents confinements). Je remercie ici aussi les différents compagnons de télétravail que j'ai pu côtoyer. Et enfin tous mes proches, en particulier mes parents, pour leur soutien au cours des dernières années.

Qu'ils soient tous ici remerciés.

Table des matières

I	La température de surface des continents et son estimation par la télé-détection microonde passive	17
1	Température de surface des continents	19
1.1	Définition de la température de surface	19
1.2	Une variable centrale pour étudier les échanges thermiques à la sur- face de la Terre	20
1.3	Caractéristiques spatiales et temporelles des températures de surface continentales	22
2	Température radiométrique et principes de télédétection	24
2.1	Rayonnement électromagnétique d'un corps noir et approximation de Rayleigh-Jeans pour les microondes	24
2.2	Interaction du rayonnement électromagnétique avec la matière et l'équation du transfert radiatif	26
3	Sources de données de température de surface	32
3.1	Stations in situ	32
3.2	Observations infrarouges	33
3.3	Observations microondes	36
3.4	Modèles de surface	37
4	Une méthode d'estimation de la température de surface par des observations microondes passives	38
4.1	Méthode utilisée pour l'estimation de la température de surface dans le projet <i>Land Surface Temperature - Climate Change Initiative</i> . . .	39
4.2	Limites des inversions de température de surface en microondes . . .	42
5	Objectifs et contenu de la thèse	43
II	Imageurs microondes passifs et leur inter-étalonnage au-dessus des sur- faces continentales pour la création de séries temporelles	45
1	Données d'imageurs microondes disponibles	46
1.1	Principe de fonctionnement d'un radiomètre microonde passif	46
1.2	Description des imageurs microondes utilisés	48
1.3	Niveau de produit distribué et étalonnage	52
2	Création de longues séries temporelles de température de brillance : le né- cessaire inter-étalonnage des instruments	53
2.1	Méthodes d'inter-étalonnage d'imageurs microondes	54
2.2	Fundamental Climate Data Records disponibles pour les imageurs microondes	56
3	Inter-étalonnage d'instrument : le cas des températures de brillance de SMR au-dessus des continents	57
3.1	Données utilisées et méthode appliquée	58
3.2	Résultats	61
3.3	Évaluation de l'inter-étalonnage proposé	65

4	Conclusion	67
III La contamination nuageuse des observations microondes utilisées pour la caractérisation des surfaces 69		
1	Typologie des nuages et leur impact sur les observations microondes	69
1.1	Distribution spatiale des types de nuages	72
1.2	Impact des nuages sur les observations microondes	72
1.3	Méthodes de détection ou de mesure de la présence nuageuse en microondes	78
2	Méthode d'apprentissage automatique	79
2.1	Modèles statistiques	80
2.2	Données d'entraînement, de test et de validation	83
3	Estimer une probabilité de contamination nuageuse sur les observations en microondes passives, au-dessus des océans et des continents	85
3.1	Sources de données	85
3.2	Choix de la méthodologie et détails d'implémentation	88
4	Résultats obtenus pour différentes surfaces	90
4.1	Détection des nuages au-dessus des océans	90
4.2	Détection des nuages au-dessus des continents	92
4.3	Amélioration de la définition d'une contamination nuageuse au-dessus des continents	92
4.4	Évaluation du modèle de détection de contamination nuageuses	97
5	Améliorer les modèles : utilisation des réseaux de convolutions	101
5.1	Les réseaux de convolutions	102
5.2	Application	103
6	Conclusion	106
IV La pénétration des microondes dans la sub-surface des zones arides 109		
1	Interaction du rayonnement microonde avec la sub-surface en milieu aride	111
1.1	Permittivité diélectrique	112
1.2	Profondeur de pénétration et température effective	112
1.3	Modèles de permittivité diélectrique pour les sols	114
2	Estimation du profil de température dans le sol et de son évolution temporelle	116
2.1	Résolution de l'équation de Fourier	116
2.2	Comparaison de la solution aux données in situ et de réanalyses	121
3	Estimation de la profondeur de pénétration des microondes	125
3.1	Domaine d'étude et données utilisées	126
3.2	Application	127
4	Analyse des résultats	129
4.1	Cartes d'émissivités estimées	129
4.2	Cartes des profondeurs de pénétration	130
4.3	Comparaison des structures géologiques connues	131
4.4	Analyse de la dépendance en fréquence des profondeurs de pénétration	133
5	Conclusion	134
V Amélioration de la résolution spatiale des estimations de température de surface 139		
1	Résolutions spatiales et temporelles des produits de température de surface issus d'observations satellites	140
2	Méthodes pour la désagrégation de la température de surface	141
2.1	Pour les températures issues de l'infrarouge	143
2.2	Pour les variables issues d'observations microonde passives	143

2.3	Données auxiliaires à haute résolution	145
3	Description de la méthode de désagrégation des températures de surface issues d'observations microondes	146
3.1	Recherche d'une relation statistique entre les données aux différentes échelles	147
3.2	Sélection des données auxiliaires	148
3.3	Implémentation du modèle d'estimation de la structure spatiale . . .	150
3.4	Application à des données synthétiques à basse résolution	152
4	Prise en compte de l'effet des nuages dans le modèle de désagrégation . . .	156
4.1	Analyse de l'effet des nuages sur les températures de surface	157
4.2	Paramétrisation de l'impact des nuages dans la désagrégation	159
5	Méthode globale « tout temps » de désagrégation	162
6	Conclusion	167
VI Conclusions et perspectives		171
1	Séries de données existantes du projet LST CCI	171
2	Résumé des résultats obtenus : application au projet LST CCI et aux ob- servations microondes	173
3	Développements et perspectives	175
A Appendix		179
1	Notations	179
2	Acronymes	180
B Articles publiés		181

Chapitre I

La température de surface des continents et son estimation par la télédétection microonde passive

Dans son mémoire de 1827 Joseph Fourier tente d'apporter une réponse à « *La question des températures terrestres* ». Ces travaux sont fondateurs par leur analyse des causes qui affectent la température de la Terre. En utilisant de nouvelles méthodes de calcul (l'équation de Fourier, le développement en série, etc.) et quelques observations, il comprend que le rayonnement solaire est le facteur le plus important expliquant la variation de la température de la Terre. Le bilan énergétique est effectué entre la chaleur interne de la Terre et le rayonnement solaire, malgré des erreurs réfutées depuis sur la température de l'espace environnant. Les conclusions tirées pour les masses continentales, océaniques ou les pôles font de ce travail une des prémices de la climatologie.

La température du sol à la surface de la Terre, à bien différencier de la température de l'air près de la surface, varie très fortement suivant les environnements. On peut ainsi estimer des températures entre -100°C à la surface de l'Antarctique jusqu'à près de 80°C dans certains déserts chauds. En plus de ces hétérogénéités spatiales, une grande variabilité temporelle est observée au cours des cycles diurnes et saisonniers de température.

La température de surface de la Terre est un paramètre essentiel dans les bilans d'échanges d'énergie à la surface de notre planète. La connaissance de la température de surface est donc primordiale pour des applications climatiques ou météorologiques. Les résultats les plus récents des analyses climatiques, comme celles produites par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) établissent la nécessité de mesurer avec précision les paramètres géophysiques sur des périodes temporelles les plus longues possibles. Ces efforts sont fournis afin de saisir l'ampleur des changements en cours dans le système climatique et d'améliorer la prédiction des scénarios d'évolution possibles pour les prochaines décennies. Ces changements s'opèrent à l'échelle d'écosystèmes (e.g. aridification du bassin méditerranéen), ou à des échelles plus locales (e.g. phénomènes d'ilots de chaleurs).

De même, la prévision numérique du temps avec une précision la plus importante possible est cruciale pour nos sociétés. Cependant, en dépit du développement de l'humanité à la surface de la Terre, une part importante de celle-ci n'est pas occupée. Dans les régions reculées et peu ou pas peuplées il est donc extrêmement compliqué de mesurer les proprié-

tés physiques avec une grande précision et avec une fréquence importante. L'avènement de la télédétection par satellite, au cours de la fin du XX^{ème} siècle, a offert de nouvelles opportunités de mesurer différents paramètres dont la température de surface à une échelle globale et temporelle satisfaisante.

Dès la fin des années 1960, les premiers résultats d'estimation de paramètres géophysiques, comme la température des océans, ou les concentrations en vapeur d'eau de l'atmosphère ont été obtenus par des mesures satellites. Les techniques d'observation passive du rayonnement comme la mesure dans l'infrarouge ont été développées par exemple pour mesurer des profils de température et d'humidité dans l'atmosphère (Aires et al., 2011b). Des méthodes actives comme le LiDAR (*Light Detection and Ranging*), ou le Radar (*Radio Detection and Ranging*) sont utilisées pour mesurer des variables comme la couverture végétale ou les phénomènes atmosphériques (Briggs, 1980; Ma et al., 2000; Rapinel et al., 2015). Les microondes passives sont des observations clés dans cette couverture du spectre de fréquence électromagnétique et sont utilisées pour mesurer différents paramètres de surface au-dessus des océans (vitesse du vent, salinité, température), ou dans l'atmosphère (précipitations, vapeur d'eau). Ces observations ont aussi des applications au-dessus des continents, parmi lesquelles on peut citer l'estimation de l'étendue des glaces de mer (Comiso et al., 1997), l'étendue des eaux de surface (Prigent et al., 2007), la quantité de neige (Pulliainen, 2006), la température de surface (McFarland et al., 1990), et l'humidité des sols (Kerr et al., 2001).

Le principal avantage des microondes passives pour la mesure de la température de surface, par rapport aux observations dans l'infrarouge, est la possibilité d'être partiellement insensible à la couverture nuageuse. Les nuages couvrent 60 % du globe en moyenne, et ce en étant inégalement répartis. Sur certaines régions, les conditions de ciel clair, nécessaires aux observations infrarouges de la surface, sont rares. C'est pourquoi le développement des mesures en microondes est nécessaire, pour pallier ces observations manquantes, mais aussi fournir des données complémentaires en synergie avec les autres sources d'observation. L'utilisation de toutes ces sources de données et de nouvelles méthodologies (modèles statistiques, augmentation de la puissance de calcul disponible) s'est traduite par exemple par l'amélioration des prédictions météorologiques par des modèles numériques (English et al., 2013).

Parmi les méthodes de télédétection, l'observation passive mesure le rayonnement naturel provenant d'un corps à une longueur d'onde donnée. Quelle que soit la fréquence à laquelle on effectue la mesure, le rayonnement émis dépend de la température du corps observé. Cette propriété peut être utilisée pour estimer la température des corps observés, que ce soit la Terre, mais aussi les planètes. De nombreux travaux ont été effectués pour comprendre et utiliser les observations satellites microondes passives pour estimer la température de la surface des continents avec des résultats extrêmement prometteurs. Les modèles inverses qui permettent d'estimer une quantité géophysique en se basant sur des observations sont maintenant opérationnels. Ils utilisent des combinaisons de données a priori et de modèles physiques pour contraindre les estimations. Pour atteindre une plus grande maturité, les méthodes de télédétection requièrent dorénavant une grande fiabilité des mesures sur le long terme, ainsi que la compréhension des possibles sources de contamination. De plus, certaines améliorations pour compenser des points faibles des observations satellites microondes comme la résolution spatiale de ces données sont possibles.

Ce chapitre présente les bases théoriques qui cadrent le travail de cette thèse. Tout d'abord, la variable de température de surface est discutée avec ses propriétés. Puis dans un second temps, le principe de la télédétection, en particulier en microondes, est introduit.

Les différentes sources permettant d'estimer la température de surface sont ensuite comparées. Plus de détails sont fournis sur une méthode permettant d'estimer la température en microonde qui est utilisée pour produire les données utilisées dans cette thèse. Finalement les problématiques liées à l'estimation de la température de surface en microondes qui seront abordées dans cette thèse sont posées.

1 Température de surface des continents

Il convient tout d'abord de définir précisément ce que l'on souhaite étudier et le rôle de la température de surface dans la description physique de l'environnement. Il est aussi important d'examiner les spécificités de la température de surface continentale qui rendent son étude délicate.

1.1 Définition de la température de surface

La **température thermodynamique** d'un corps est la quantité qui correspond à l'énergie liée à l'agitation des particules qui le constitue. Elle peut être mesurée en utilisant un thermomètre à l'équilibre thermodynamique avec l'objet ciblé. On a alors une relation entre une grandeur observable (volume d'une quantité fixe de mercure à pression constante par exemple) et la température du corps.

Dans le cas de la Terre, on peut définir la **température de surface** de manière simplifiée comme la température qui serait ressentie en touchant la surface de la Terre. Elle varie suivant le type de surface, l'heure de la journée, la saison ou les conditions météorologiques. Les surfaces continentales peuvent être composées de glace ou de neige, de terre nue ou couverte de végétation, ou encore de constructions humaines. Pour permettre l'étude de la température, on considère qu'elle a une valeur moyenne sur une certaine aire aux propriétés homogènes. Dans les modélisations de la surface, elle est parfois appelée *skin temperature* (T_{skin}), elle correspond alors à la température qui est requise pour satisfaire le bilan énergétique de la surface (décrit par l'équation I.1).

Plusieurs sources de mesures ou d'estimation permettent son étude, les principales seront listées dans la section 3. Elle peut être modélisée ou estimée indirectement à partir de l'observation de variables dont la valeur dépend de la température de la surface. De manière générale, les variations de la température de surface sont étudiées sur une échelle de la dizaine de mètres à la dizaine de kilomètres. Les variations temporelles sont quant à elles échantillonnées, suivant les sources entre la minute et une dizaine d'heures.

Il faut bien distinguer la température de la surface dont il est question dans cette thèse de la température de l'air à 2 m (T_{2m}). Cette variable est mesurée en routine par les stations météo à la surface de la Terre et sert pour dans les prévisions météorologiques mais aussi pour le suivi climatique (Hansen et al., 2010). Les deux températures présentent des évolutions temporelles différentes, comme on peut le voir sur la figure I.1. Le pic de la température de surface est plus matinal que celui de la température de l'air en raison de différences des propriétés thermiques de l'air et du sol. La différence entre la T_{2m} et la température du sol régit le flux de chaleur sensible, mais intervient aussi dans d'autres échanges du bilan radiatif.

On ne parlera pas dans cette thèse de la température de la surface des grandes étendues

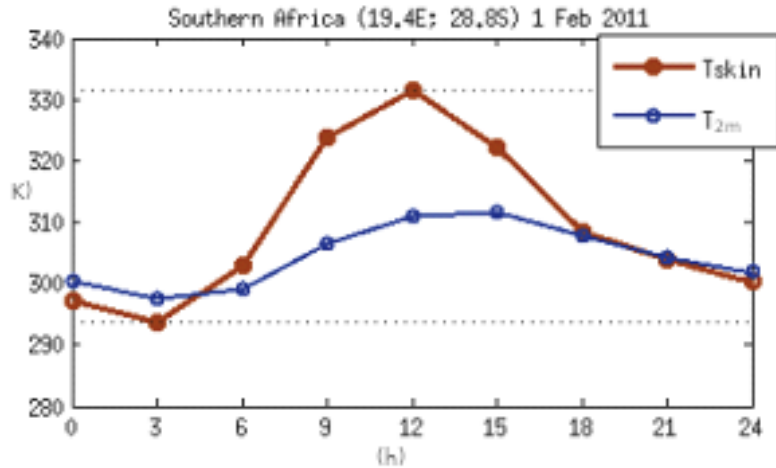


FIGURE I.1 – Comparaison de la température de surface (en rouge) et de la température de l’air à 2 m (T_{2m} en bleu) en kelvin (K), au-dessus d’un point en Afrique du Sud pour une journée du 01 février 2011 (extraite de EUMeTrain Land Surface Temperature tutorial).

d’eau (océans et lacs) qui ont un comportement physique différent de celui des surfaces continentales. Les étendues d’eau sont plus homogènes spatialement et temporellement, la température des océans varie dans un plage beaucoup plus faible (0 à 40° C) qui s’explique par la disponibilité constante d’eau pour l’évaporation et une inertie thermique plus importante. Tous ces éléments en font un problème bien distinct.

1.2 Une variable centrale pour étudier les échanges thermiques à la surface de la Terre

La température de surface doit être mesurée, car elle intervient dans de nombreux échanges d’énergie à la surface de la Terre. Les cycles de l’eau et de l’énergie sont interdépendants sur Terre. Connaître la température de surface permet d’évaluer les modélisations des échanges à la surface et de contraindre les bilans énergétiques au niveau du sol. Son étude peut aussi servir de manière indépendante comme indicateur des changements météorologiques ou climatiques d’une région. Comprendre dans quelle mesure les prévisions météorologiques, les écosystèmes ou le cycle de l’eau seraient modifiés par des perturbations météorologiques ou climatiques nécessite un suivi fin de la température de la surface.

Bilan thermique

Le bilan thermique à la surface relie l’ensemble des paramètres qui influent sur changement de la température de surface. Pour une surface supposée homogène et isotherme, d’épaisseur négligeable, le flux de température dans ce volume unitaire s’écrit (Polcher et al., 1998) :

$$\rho C \frac{\partial T_{skin}}{\partial t} = L_n + S_n + SH + LE + G. \quad (I.1)$$

Dans cette équation on note

- T_{skin} la température de la surface moyenne sur la couche infinitésimale considérée,

avec ρ la densité du sol et C la capacité thermique massique des propriétés physiques qui dépendent de la composition du sol.

- L_n est le flux net de rayonnement de grande longueur d'onde.
- S_n est le flux net de rayonnement à ondes courtes.
- SH est le flux de chaleur sensible entre le sol et l'atmosphère qui dépend de la différence de température entre le sol et la couche inférieure de l'atmosphère (le plus souvent assimilée à T_{2m}). Des effets liés au vent ou à la rugosité de surface modifie la valeur de cet échange.
- LE est le flux de chaleur latente lié au changement de phase de l'eau au niveau du sol. L'humidité spécifique de la couche inférieure de l'atmosphère contraint cet échange.
- G est le flux de chaleur échangé par conduction avec les couches inférieures du sol.

La quantité d'eau dans le sol (W) joue aussi un rôle important à la fois dans le flux de chaleur latente, mais aussi dans les paramètres thermiques de la surface (C). La conservation de l'eau à la surface permet aussi d'écrire que

$$\frac{\partial W}{\partial t} = P - R - E,$$

avec P les précipitations, R l'écoulement vers les cellules voisines, et E l'évaporation. Ce terme d'évaporation lie donc le bilan énergétique et le bilan hydrique en surface.

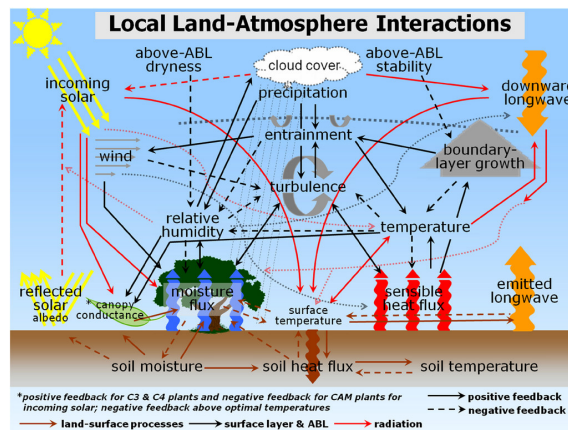


FIGURE I.2 – Schéma représentant différents processus à l'interaction entre l'atmosphère et la surface. Les flèches marrons représentent les processus de surface, avec les interactions positives en flèches continues et négatives en trait discontinu (figure extraite de Dirmeyer et al., 2019).

La figure I.2 représente par des flèches les différentes interactions entre des composants à la surface de la Terre. Les rétroactions positives dans les flèches continues et négatives dans les flèches discontinues. Les flux d'énergie ou d'eau sont aussi représentés par les flèches ondulées, et les précipitations en hachures grises. La variable de température de surface intervient dans un grand nombre des phénomènes représentés ici. Lorsque l'on s'intéresse aux variations de température de surface à une échelle plus grande que le kilomètre, on peut négliger les échanges de chaleur latéraux avec des régions voisines. On considère alors le problème en une dimension sur une colonne de sol et d'atmosphère. D'autres phénomènes non représentés ici comme l'effet de la neige (e.g. fonte, accumulation) influencent aussi la température de surface.

Les modèles climatiques ou de surface subdivisent souvent les cellules de grille¹ suivant la couverture du sol (neige, végétation haute/basse, sol nu...). Des modélisations différentes peuvent être appliquées à chaque type de couverture du sol ce qui améliore la représentation. Parmi les modèles de surface, on peut citer le *Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL)* (Balsamo et al., 2011). Ce modèle fait partie du système utilisé pour produire les données ERA5 (détails dans la section 3).

Utilisations de la température de surface continentale

La mesure de la température de surface est donc centrale pour estimer les échanges décrits. Apporter une mesure précise de la température en surface permet de mieux comprendre les phénomènes en jeu, contraindre les différents flux et améliore l'estimation des différentes variables interconnectées. De plus la température de surface est directement utilisée dans certaines applications. On peut lister un certain nombre de domaines pour lesquels la température de surface est une donnée nécessaire :

- Pour l'étude à long terme du changement climatique une variable telle que la température de surface permet de mesurer l'évolution et la sensibilité des différentes parties du globe aux changements en cours. De telles études requièrent des mesures correctement étalonnées et disponibles sur la plus grande durée possible pour analyser avec confiance les tendances détectées. C'est l'objectif de l'initiative de l'Agence Spatiale Européenne (ESA) *Climate Change Initiative (CCI)*, qui cible un certain nombre de variables climatiques essentielles (*Essential Climate Variable - ECV*) dont fait partie la température de surface (Bojinski et al., 2014).
- La température de surface intervient dans les modèles de surface, qu'ils soient utilisés de manière indépendante ou inclus dans des modèles globaux, par exemple pour la prédiction numérique du temps. Estimer la température indépendamment et la comparer aux sorties des modèles permet de guider les développements requis pour améliorer la fiabilité des modélisations.
- Le troisième domaine d'application regroupe les utilisations directes de la température de surface. En agriculture, la mesure de la température de surface permet d'optimiser les rendements, en détectant les conditions inappropriées à la croissance des plantes. La mesure de l'humidité des sols, critique en période de sécheresse, utilise aussi des estimations de température de surface (Kustas et al., 2003). Enfin la détection des zones d'extrême concentration de chaleur (comme les zones urbaines) repose aussi sur la mesure de la température de surface.

1.3 Caractéristiques spatiales et temporelles des températures de surface continentales

Les variations de la température de surface peuvent être importantes à l'échelle du globe, les records de température du sol mesuré par des satellites sont ainsi de 80°C et -110°C sur Terre (Zhao et al., 2021). Comme l'indique l'équation I.1, plusieurs termes peuvent conduire à des changements de température du sol. La source la plus importante de variation est le rayonnement solaire arrivant à la surface. La variation quotidienne du rayonnement due à l'alternance des jours et des nuits crée un cycle diurne dont on voit un exemple sur la figure I.1. Les variations annuelles liées aux saisons apparaissent aussi

1. Une cellule de grille est l'unité qui définit l'aire de la colonne sur laquelle est appliquée le modèle

dans les données mesurées. A ces variations causées par le rayonnement solaire incident, une variabilité additionnelle peut être ajoutée par les conditions météorologiques, comme des variations de l'humidité du sol, de la température ou de l'humidité de la couche d'atmosphère la plus proche du sol ou encore de la couverture nuageuse.

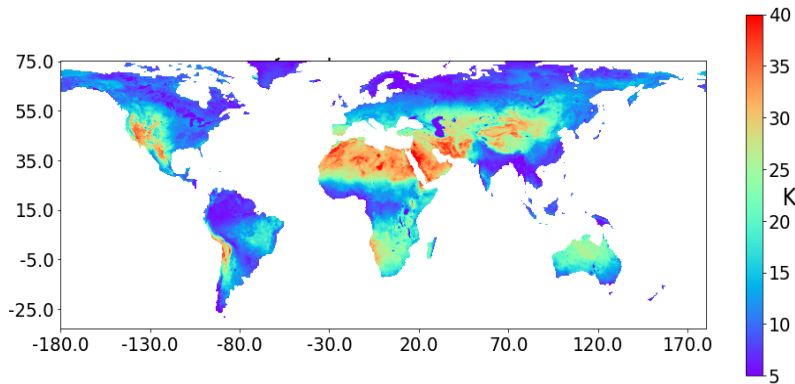


FIGURE I.3 – Carte des amplitudes des cycles diurnes de température de surface moyens (K) pour le mois de juin 2016 dans les données de réanalyse ERA5.

On retrouve sur la figure I.3 l'amplitude du cycle diurne moyen calculé pour le mois de juin 2016 dans les données de la réanalyse ERA5. Les amplitudes des cycles diurnes varient de près de 40 K sur les déserts à moins de 5 K au-dessus des forêts denses. Le contraste entre les zones arides, du Sahara, d'Asie centrale, de l'ouest américain avec de fortes amplitudes et les forêts dans la zone tropicale ou aux hautes latitudes est marqué.

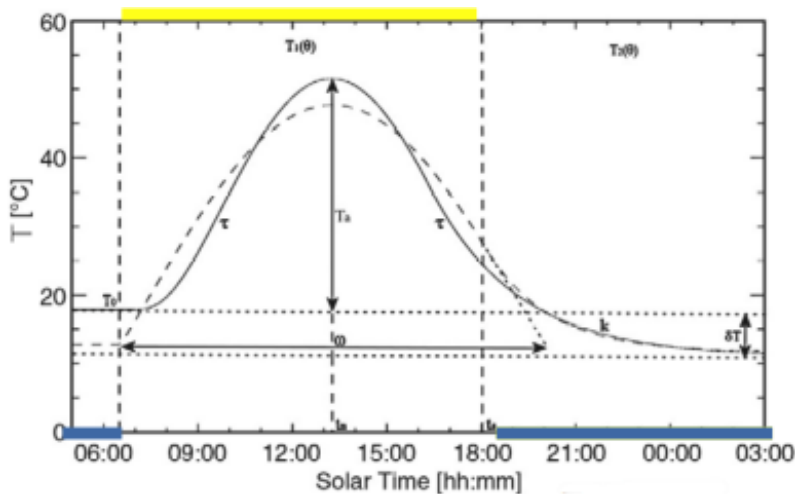


FIGURE I.4 – Modélisation simple de l'évolution de la température de surface ($^{\circ}\text{C}$) sur une journée par Gottsche and Olesen (2009). En jaune la durée du jour et en bleu de la nuit.

Une façon simple de résumer les variations quotidiennes de température en fonction du rayonnement est proposée par Gottsche and Olesen (2009). On retrouve sur la figure I.4 l'évolution idéalisée de la température d'une surface soumise à un rayonnement solaire au cours du jour. La modélisation est basée sur une décroissance exponentielle pour simuler le refroidissement de la température la nuit et une sinusoïde partielle pour la température de jour. Cette approximation du comportement des températures de surface peut être ajustée empiriquement pour de nombreuses situations, mais ne prend pas en compte des variations ponctuelles, par exemple causées par la présence de nuages.

L'équation I.1 met en évidence des causes possibles de la différence de température entre des surfaces pourtant soumises à la même insolation et aux mêmes conditions atmosphériques. Le changement de température est en effet contraint par les propriétés physiques du sol, en particulier la densité ρ et la capacité thermique massique C . C représente la quantité d'énergie nécessaire pour changer la température d'une unité de masse d'un degré. Elle dépend de la composition du sol, en particulier de son humidité. La valeur de C de l'eau liquide est environ 5 fois plus élevée que celle de la roche. C'est ce qui explique en partie la différence importante du comportement de la température au-dessus des surfaces océaniques et terrestres. De même, au-dessus des continents, où les zones humides continentales (e.g., marais, zones d'inondations) ont des variations quotidiennes plus faibles de la température que les zones arides. D'autres effets peuvent modifier la température de la surface, comme la présence d'ombres portées sur la surface, qui limitent le rayonnement incident (Ermida et al., 2014) dues à l'orographie qui limite l'exposition de certaines pentes ou à la présence de nuages.

Les figures I.1 et I.3 mettent en évidence la forte dynamique du cycle diurne, à la fois spatialement et temporellement. Les nombreux phénomènes qui affectent la température à la fois liés aux conditions météorologiques, aux flux radiatifs, au cycle de l'eau et au type de sol rendent le suivi de la température à une échelle globale une tâche délicate. C'est ce qui rend nécessaire l'étude et l'estimation de la température de surface par différentes méthodes dont les observations satellites (Jin and Dickinson, 2010).

2 Température radiométrique et principes de télédétection

Pour mesurer une température de surface, on a mentionné la mesure d'une grandeur physique en équilibre thermodynamique avec l'objet que l'on veut mesurer. Cette méthode n'est pas applicable de manière satisfaisante pour la mesure des températures de surface à une échelle globale. On utilise donc la mesure indirecte de la température que l'on peut observer à distance. On définit la **température radiométrique** qui correspond à la température à l'origine du rayonnement émis par toute la surface dans le champ de vision d'un détecteur en direction de celui-ci (Becker and Li, 1995). Pour comprendre le principe de cette mesure radiométrique par des instruments en orbite, en particulier dans le domaine microonde, il faut détailler quelques principes de télédétection.

Cette section expose les principes qui régissent le fonctionnement de la mesure à distance (télédétection) principalement dans le domaine des microondes. Une première section couvre le rayonnement électromagnétique, puis les interactions du rayonnement avec les milieux traversés. Pour finir l'équation du transfert radiatif est donnée.

2.1 Rayonnement électromagnétique d'un corps noir et approximation de Rayleigh-Jeans pour les microondes

Le spectre des fréquences électromagnétiques est donné à la figure I.5 et les rayonnements dans le domaine visible puis infrarouge² et enfin microondes apparaissent pour des fréquences décroissantes. Plus la fréquence augmente plus le rayonnement est énergétique. Le domaine des microondes³ commence à partir d'une fréquence de 1 GHz et la plage

2. L'abréviation IR pour le rayonnement InfraRouge est couramment utilisée.

3. L'abréviation MW pour *MicroWave* est utilisée en anglais.

utilisée dans cette thèse se limite aux fréquences inférieures à 200 GHz.

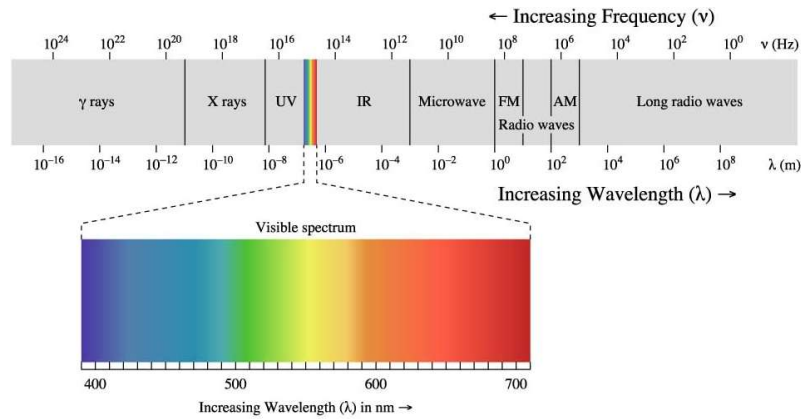


FIGURE I.5 – Spectre électromagnétique. En haut l'échelle des fréquences (Hz), et en bas celle des longueurs d'ondes (m). Le détail du spectre visible est ajouté en couleur (figure extraite de Wikimedia Commons).

Une des propriétés fondamentales de l'électromagnétisme est que tout corps à une température non nulle (-273.15°C) émet un rayonnement électromagnétique. La loi de Planck décrit la distribution de la luminance aux différentes fréquences en fonction de la température pour un corps noir en équilibre thermodynamique. La luminance énergétique spectrale (en $\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{Hz}^{-1}$) est la densité du flux énergétique émis par une surface pour un angle solide donné à une fréquence ν et pour une température thermodynamique T de la surface du corps noir.

$$L(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_{\text{B}}T}} - 1}. \quad (\text{I.2})$$

Dans l'équation I.2, h est la constante de Planck, c la vitesse de la lumière dans le vide et k_{B} la constante de Boltzmann. T la température en Kelvin et ν la fréquence en Hertz sont les deux variables. C'est ce phénomène qui explique l'émission de lumière visible par des corps chauffés à plusieurs centaines de degrés. Le soleil a une température proche de 6000 K, la quantité maximale de rayonnement fournie est proche des longueurs d'onde du visible. Pour des températures similaires à celles rencontrées sur Terre ($\sim 290 \text{ K}$) un corps noir émet principalement dans le domaine infrarouge. Mais une part du rayonnement est émise dans la bande des microondes. L'espace lointain a lui une température faible et donc un rayonnement principalement dans le domaine des microondes. C'est ce qui est résumé sur la figure I.6.

Le corps noir est un modèle théorique parfait pour lequel tout le rayonnement provenant d'une source est dû à l'émission thermique, c'est-à-dire sans réflexion. Pour des objets réels, on parle de corps gris pour lesquels l'émission n'est pas parfaite et on a donc un rapport entre la luminance émise par un corps réel ($L_{\text{corps gris}}$) et un corps idéal ($L_{\text{corps noir}}$) que l'on appelle émissivité. Cette émissivité dépend des propriétés du matériau, de sa température et de la fréquence d'observation. L'émissivité e_{ν} à une fréquence donnée varie donc entre 0 et 1 qui est celle d'un corps noir.

$$e_{\nu} = \frac{L_{\text{corps gris}}}{L_{\text{corps noir}}}. \quad (\text{I.3})$$

Les mesures effectuées par un radiomètre sont données sous la forme de températures de brillance (T_b) qui correspondent à la température d'un corps noir qui fournirait le même

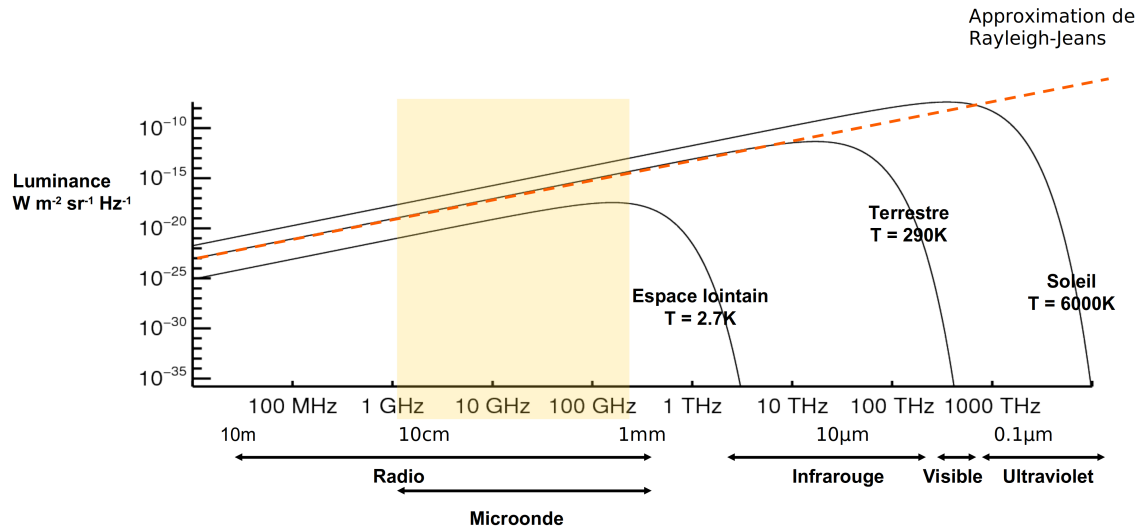


FIGURE I.6 – Distribution de luminance énergétique spectrale ($W \cdot m^{-2} \cdot sr^{-1} \cdot Hz^{-1}$) par la fonction de Planck pour des corps à différentes températures. En jaune la bande des microondes et en pointillé orange l’approximation de Rayleigh-Jeans.

rayonnement que celui observé. La mesure d’une température de brillance en infrarouge ou en microonde permet donc en connaissant l’émissivité de déterminer la température radiométrique d’un corps.

Pour les fréquences dans le domaine des microondes, on a $\frac{h\nu}{kT} \ll 1$. Alors par un développement limité, on obtient l’approximation de Rayleigh-Jeans de la loi de Planck. À une fréquence ν , avec c la vitesse de la lumière dans le vide et k_B la constante de Boltzmann elle s’écrit :

$$L(\nu, T) = \frac{2\nu^2 k_B T}{c^2}. \tag{I.4}$$

La figure I.6 affiche en orange la droite qui correspond à l’approximation de Rayleigh-Jeans pour un corps à une température proche de celles rencontrées sur Terre (~ 290 K). On voit que cette approximation n’est valable que pour des conditions de températures et de fréquences spécifiques. Dans le domaine microonde, c’est sur la base de cette équation que la température équivalente d’un corps noir mesurée par le satellite peut être convertie en une température thermodynamique (Carver et al., 1985).

2.2 Interaction du rayonnement électromagnétique avec la matière et l’équation du transfert radiatif

La section précédente présente l’émission d’un rayonnement par un corps quelconque. Lors de la propagation d’une onde électromagnétique, des interactions avec la matière se produisent. Les principaux effets sont introduits ici mais seront détaillés à nouveau dans les chapitres successifs.

Les milieux rencontrés par les ondes électromagnétiques sont dans le cas de cette thèse des diélectriques, c’est-à-dire des matériaux non conducteurs pour lesquels on considèrera aussi qu’ils ne sont pas magnétiques. Une onde électromagnétique peut s’écrire sous la forme d’une fonction d’onde sinusoïdale. Les composantes électriques et magnétiques sont orthogonales entre elles et par rapport à la direction de propagation de l’onde. Dans

un espace à 3 dimensions (x,y,z), pour une onde se propageant suivant direction de z, les composantes des champs se répartissent sur les axes x et y. La variation de cette répartition entre ces deux directions au cours du temps est décrite par la polarisation de l'onde. On peut décrire l'onde comme la superposition de deux champs orthogonaux entre eux suivant deux polarisations (le plus souvent nommées horizontale et verticale).

Dans les matériaux non conducteurs, la permittivité diélectrique ϵ_ν , pour une fréquence donnée caractérise la propagation d'une onde à cette fréquence dans ce milieu. La notation complexe utilisée $\epsilon = \epsilon' - j\epsilon''$ sépare la partie réelle qui décrit la vitesse de propagation de l'onde dans le milieu de la partie imaginaire qui décrit les pertes par absorption.

Par conservation de l'énergie lorsqu'un rayonnement rencontre un objet, la totalité du rayonnement incident est soit transmis, soit absorbé, soit diffusé par le milieu. Un milieu est dit transparent à une longueur d'onde si la totalité du rayonnement incident est transmis (le cas du vide pour toute les fréquences). Si le rayonnement subit une extinction, son opacité augmente. Cette extinction permet de définir la **profondeur de pénétration**. Cette distance correspond à la profondeur pour laquelle l'intensité du rayonnement est réduite d'un facteur e . En notant le coefficient d'atténuation du rayonnement causé par un phénomène de diffusion κ_s et celui de l'atténuation par absorption κ_a . La profondeur de pénétration s'écrit (Ulaby et al., 2014) :

$$\delta_p = \frac{1}{\kappa_a + \kappa_s}.$$

Par absorption

A l'inverse de l'émission, l'absorption décrit la transformation de l'énergie incidente apportée par une onde électromagnétique en énergie interne sous forme de chaleur. Une atténuation exponentielle du champ électrique se produit en pénétrant dans un milieu, on peut écrire le coefficient d'absorption volumique qui dépend des propriétés diélectriques, avec λ_0 la longueur d'onde dans le vide :

$$\kappa_a = -2\frac{2\pi}{\lambda_0}\Im(\sqrt{\epsilon}).$$

En supposant nulle la diffusion dans le milieu : $\delta_p = \frac{1}{\kappa_a}$. Cette notion caractérise l'épaisseur de la couche de matière qui contribue à un rayonnement. Aux longueurs d'ondes proches de l'infrarouge, la couche de matière contribuant au rayonnement mesuré provenant de cette surface est fine ($\delta_p \approx 50\mu m.$). On peut donc supposer cette couche en équilibre thermodynamique. On peut alors l'identifier à T_{skin} la température introduite pour le bilan radiatif. En microondes, la présence d'eau augmente fortement l'atténuation et donc l'épaisseur de la couche contribuant au rayonnement est aussi faible. En l'absence d'eau les valeurs de ϵ'' pour les fréquences microondes peuvent devenir très faibles et la profondeur de pénétration devient alors importante. Ce cas sera examiné au chapitre IV.

Par diffusion et réflexion

Dans un corps absorbant un rayonnement s'il est à l'équilibre thermodynamique (pas de changement de chaleur) il y a ré-émission d'un rayonnement correspondant à celui absorbé. Cette ré-émission se fait dans de multiples directions et plusieurs configurations peuvent

se présenter suivant le type de milieu et la taille des particules rencontrées par rapport à la longueur d'onde. La section efficace de rétrodiffusion σ_b est une aire qui correspond au rapport entre le rayonnement retourné dans la direction d'origine et le rayonnement incident.

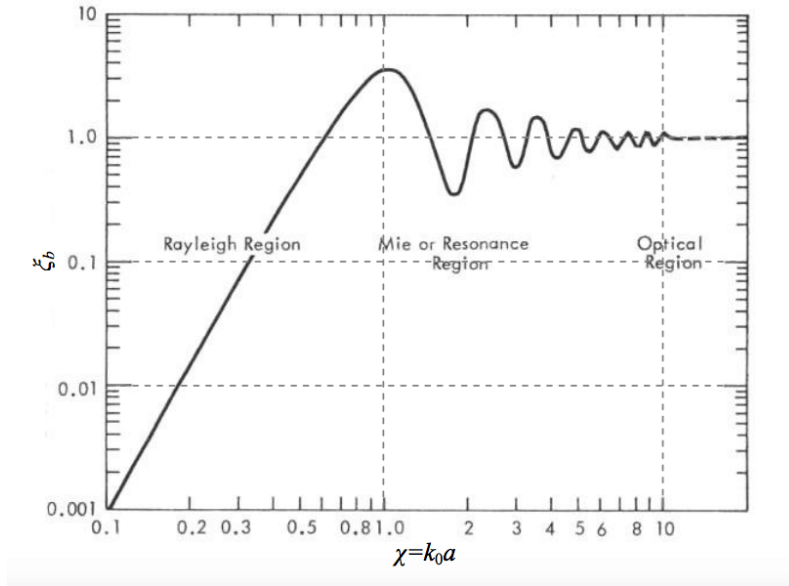


FIGURE I.7 – Efficacité de la rétrodiffusion en fonction du ratio entre le diamètre de la particule (pour des sphères métalliques) et le nombre d'onde ($k_0 = \frac{2\pi}{\lambda}$). Sont spécifiées les zones pour lesquelles les approximations de Rayleigh et Mie sont valables (figure extraite de Ulaby et al., 2014).

On peut mesurer l'efficacité de la rétrodiffusion qui dépend du rapport entre la section efficace de rétrodiffusion et la taille des particules ou des structures internes et la longueur d'onde. Elle se note alors $\xi_b = \frac{\sigma_b}{\pi a^2}$ avec a le diamètre des particules rencontrées. La figure I.7 décrit l'évolution de l'efficacité de la rétrodiffusion par rapport au ratio entre le diamètre de la particule et le nombre d'onde ($k_0 = \frac{2\pi}{\lambda}$). Quand $a \leq \frac{\lambda}{2\pi}$, c'est-à-dire des particules petites devant la longueur d'onde, la diffusion diminue fortement avec la diminution de $\chi = \frac{2\pi a}{\lambda}$, c'est la diffusion de Rayleigh. Lorsque l'ordre de grandeur entre la longueur d'onde et les particules observées est le même, le régime de Mie s'applique.

On voit dans cette section que les tailles des particules rencontrées ont une influence importante sur le rayonnement transmis et mesuré. Dans le domaine des infrarouges, les longueurs d'onde sont de l'ordre de grandeur du micromètre. Cette longueur est proche de la taille des gouttes d'eau en suspension dans les nuages. Dans ce cas, le régime de Mie s'applique. Cela rend les observations très sensibles à la présence de nuages. En microonde la longueur d'onde est plus importante ce qui permet de ne pas être sensible à la plupart des hydrométéores⁴. Plus de détails sur ces effets sont examinés au chapitre III.

Dans le cas de la réflexion, pour des surfaces spéculaires, les lois de l'optique géométrique s'appliquent. Le calcul de l'angle du rayon réfléchi et l'intensité de rayonnement associé dépendent uniquement de l'angle du rayon incident ainsi que des propriétés diélectriques des deux milieux en contact. L'angle pour lequel la réflexion de la polarisation normale au plan d'incidence est nulle est appelé angle de Brewster ($\sim 53^\circ$). Les surfaces réelles sont rarement planes et des effets liés à la rugosité de la surface complexifient

4. Les hydrométéores correspondent à l'ensemble des formes prises par l'eau en suspension dans un nuage : gouttelettes, cristaux de neige, grêle, etc.

souvent le calcul de la réflexion.

Équation du transfert radiatif en microondes

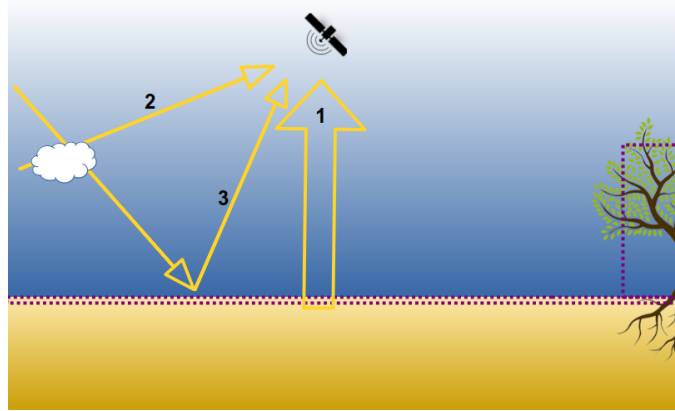


FIGURE I.8 – Schéma représentant les différentes contributions au rayonnement mesuré au niveau du satellite. En pointillé violet, la zone correspondant à la température de surface telle qu’observée par un satellite en microonde. En jaune, les composantes de la température de brillance mesurée : (1) le signal émis par la surface transmis à travers l’atmosphère, (2) l’émission des composants atmosphériques (gaz/nuages), (3) le rayonnement émis par l’atmosphère réfléchi à la surface et transmis à travers l’atmosphère.

Les températures de brillance mesurées par un satellite en orbite observant la Terre en microonde résultent donc de plusieurs phénomènes. Comme présenté dans la figure I.8, la surface de la Terre et son atmosphère ont leur part dans le rayonnement mesuré. La section précédente met en évidence la variété des processus qui affectent le rayonnement à une longueur d’onde, qui dépend à la fois de la température et des paramètres physiques des milieux. En microonde, cette mesure se compose des termes suivants, qui répartissent les contributions entre différents milieux, pour une fréquence ν une polarisation P et un angle d’observation θ :

$$T_{b(\nu,P,\theta)} = T_{b\uparrow(\nu,P,\theta)} + \tau_{\nu,\theta}[e_{\nu,P,\theta}T + (1 - e_{\nu,P,\theta})T_{b\downarrow(\nu,\theta)}]. \quad (\text{I.5})$$

Avec $T_{b,\nu,P,\theta}$ la température de brillance mesurée par le satellite, $T_{\uparrow,\nu,\theta}$ et $T_{\downarrow,\nu,\theta}$ les températures de brillance montante et descendante de l’atmosphère. La contribution due aux sources extraterrestres comme celle du fond cosmique (à ~ 3 K) est comprise dans $T_{\downarrow,\nu,\theta}$. On suppose ici que les contributions atmosphériques ne sont pas polarisées. La température thermodynamique de la surface est T . $\tau_{\nu,\theta}$ correspond à l’extinction causée par l’atmosphère qui dépend des conditions atmosphériques. $e_{\nu,P,\theta}$ est l’émissivité de la surface, qui varie avec la fréquence la polarisation et l’angle d’observation. Pour la plupart des mesures par un imageur microondes passives ayant un angle de visée fixe, le paramètre θ est constant ($\sim 53^\circ$) et donc sous-entendu.

C’est cette équation qui sert pour l’estimation de nombreux paramètres géophysiques à partir d’observations satellites en microonde. Résoudre le problème inverse consiste à déterminer les paramètres géophysiques qui correspondent à l’observation $T_{b,\nu,P}$ effectuée. Pour mieux contraindre ce problème, de multiples observations à des fréquences, polarisations et/ou angles d’incidences différents sont utilisées simultanément, et des données auxiliaires peuvent être ajoutées. Dans le cas de l’estimation de la température de surface,

la contribution de l'atmosphère n'est pas la variable d'intérêt. Cependant, il est nécessaire de connaître la contribution de l'atmosphère et les valeurs de la transmission atmosphérique pour chacune des fréquences utilisées. La variation de la transmittivité atmosphérique en microonde explique en partie le choix des canaux utilisés.

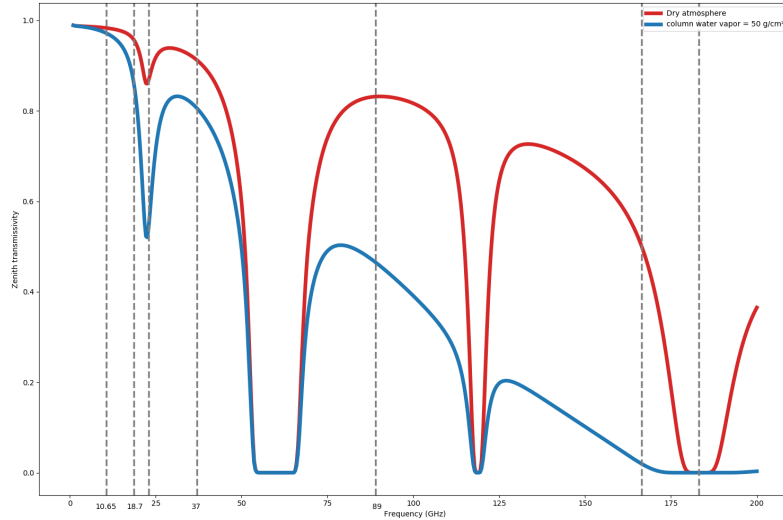


FIGURE I.9 – Transmission atmosphérique estimée pour une atmosphère pour les fréquences entre 1 et 200 GHz en moyenne pour le mois de janvier (obtenue des données de réanalyse ERA5) par le modèle de Rosenkranz (Rosenkranz, 1998) au-dessus du Sahara. En rouge une atmosphère sèche (*Column Water Vapour Content, CWVC* = $0 \text{ kg} \cdot \text{m}^{-2}$) et en bleu avec un contenu en vapeur d'eau, (*CWVC* = $50 \text{ kg} \cdot \text{m}^{-2}$). Les lignes pointillées grises indiquent la position de certaines des fréquences utilisées à bord de l'instrument microonde GMI à 10, 18.7, 23, 36, 89, 166 et 183 GHz.

La figure I.9 affiche la transmittivité atmosphérique pour des profils atmosphériques moyens au-dessus du Sahara en janvier. Les raies liées à l'absorption de différents composants gazeux sont notables, ainsi que le continuum de vapeur d'eau dont la contribution augmente avec la fréquence. On note les raies du dioxygène à ~ 60 et ~ 120 GHz, et les raies de vapeur d'eau à ~ 22 et ~ 183 GHz. Ces phénomènes d'absorption à des fréquences spécifiques permettent, par exemple, le sondage atmosphérique pour l'établissement de profils de température et de vapeur d'eau. Pour l'observation de la surface, la transmission maximale du signal est souhaitée, ce qui conduit à utiliser les canaux en dehors de ces raies d'absorption. On voit aussi que les canaux aux plus basses fréquences seront moins sensibles à la présence de vapeur d'eau. L'impact de la variation du contenu en vapeur d'eau est illustré par la différence entre la courbe rouge et la courbe bleue. En augmentant le contenu en vapeur d'eau (courbe bleu,) la transmittivité décroît très fortement, et ce d'autant plus que la fréquence est élevée.

La figure I.10 montre un exemple de carte de températures de brillance observées par un radiomètre observant la Terre à 19 GHz en projection polaire au mois de décembre 2014. On peut noter sur cette figure différentes structures spatiales qui correspondent à des phénomènes différents. Les différences entre les T_b observées pour différentes surfaces proviennent à la fois des différences de température, mais aussi des différences d'émissivité. L'émissivité de l'océan en microonde est plus faible que celle des continents, de la neige ou de la glace, ce qui conduit les T_b observées à être très différentes. Les caractéristiques de la surface de la cryosphère ne sont pas identiques à celles des surfaces continentales non glacées. La présence de neige peut expliquer les températures de brillance froides au-dessus de l'Amérique du Nord et de la Sibérie. La banquise peut être vue à la transition entre les

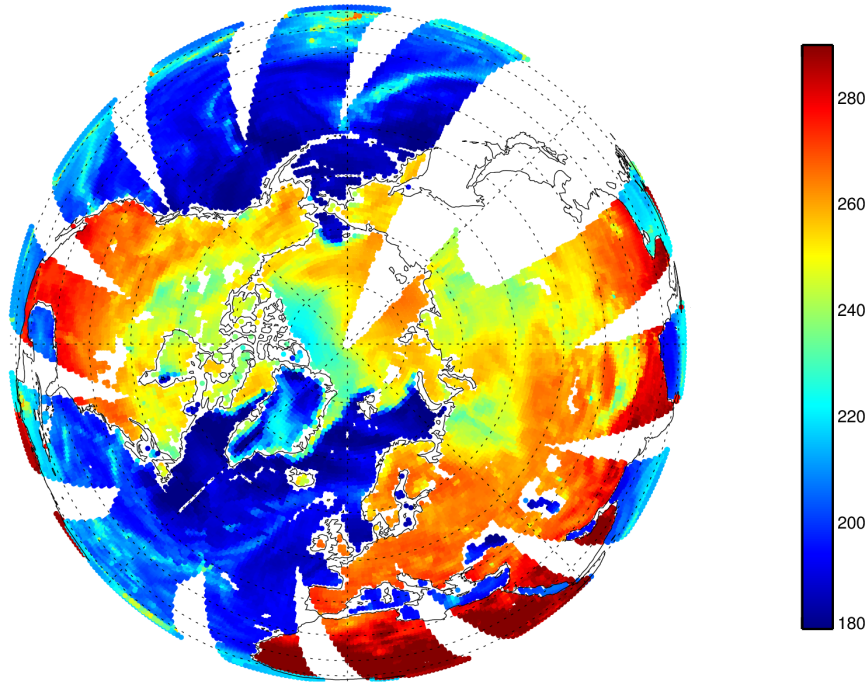


FIGURE I.10 – Carte des températures de brillance (K) en projection polaire observée par l'instrument SSMIS à 19 GHz en polarisation verticale le 3 décembre 2014. (A. Geer ECMWF)

températures de brillance de l'océan autour de 180 K à celles de la glace autour de 240 K. Des structures liées à des perturbations atmosphériques peuvent aussi être remarquées au-dessus des océans comme au sud de l'Islande ou à l'est de la côte américaine.

Température effective

On a défini la température radiométrique et la température de la surface de la Terre à la section précédente. En supposant que les profondeurs de pénétration sont faibles et que la couche contribuant au signal est à l'équilibre thermodynamique, les deux variables peuvent être identifiées. En supposant la surface homogène dans le champ de vision d'une observation (température et émissivité constante) on a vu la relation entre la température de brillance observée et la température de la surface. Dans le cas d'une observation satellite, l'hypothèse d'une surface aux propriétés homogènes ne peut pas être vérifiée à l'échelle du champ de vision de l'instrument en orbite (de l'ordre de la dizaine de kilomètres).

On considère donc que ce qui peut être estimé par le satellite est la **température moyenne effective** d'une portion de la Terre, qui contient des contributions de tous les types de surfaces présents dans le champ de vision. Ces types de surface peuvent être de la végétation, ou un sol nu et correspondent aussi à une épaisseur définie par la profondeur de pénétration à la fréquence donnée. On supposera cette région homogène à l'échelle de l'observation satellite.

Le terme de température de surface, noté T , employé par la suite de cette thèse, se réfère donc à la température effective de l'ensemble hétérogène qui forme la couverture du sol⁵ sur

5. Elle est parfois aussi appelée *Land Surface Temperature* (LST) ou T_s .

une certaine épaisseur généralement très fine. Les problématiques liées à cette définition, à savoir la profondeur de pénétration dans les sols des zones arides et l'hétérogénéité de la couverture sont abordés dans les chapitres IV et V.

3 Sources de données de température de surface

La définition et le principe de l'estimation de la température de surface par des observations satellites en microondes a été présenté. Cette section introduit les principales sources de mesures existantes de la température de surface qui sont issues soit de mesures in situ, soit d'observations satellites, soit de modélisations de la surface.

Les avantages et inconvénients des différentes méthodes sont comparés. Ces sources seront utilisées au cours du travail de thèse comme points de comparaison ou pour contraindre les problèmes rencontrés. Les stations in situ sont locales, tandis que les estimations fournies par les modèles de surface et les observations microondes ou infrarouges sont globales ou quasi-globales. La figure I.11 permet de comparer schématiquement les résolutions spatiales et temporelles des différentes sources de données pour un point à la surface de la Terre.

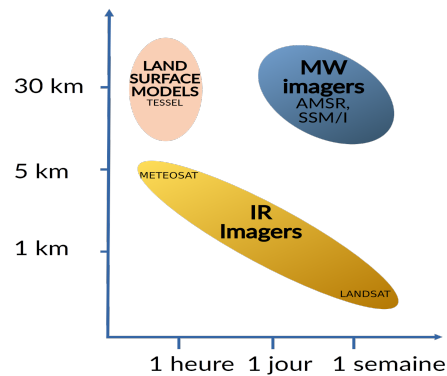


FIGURE I.11 – Différences de résolution spatiale et temporelle entre les instruments microondes (*MW imagers*), infrarouges (*IR Imagers*) et les données issues de modèles (Land Surface Models).

3.1 Stations in situ

La température de la surface peut être obtenue localement par des stations équipées de capteurs infrarouges radiométriques. Les stations disposant de tels instruments de mesures sont rares, car ce n'est pas un paramètre mesuré en routine par les stations météorologiques comme peut l'être la température de l'air à 2 m.

En connaissant l'émissivité infrarouge de la surface sous le capteur infrarouge placé sur une tour (partie gauche de la figure I.12) il est possible d'effectuer des mesures de température avec une résolution temporelle très importante et ce même en présence de nuages. En sélectionnant des sites avec une surface homogène sur de larges régions, il est alors possible de comparer les mesures effectuées localement à celles observées à la résolution des satellites (Guillevic et al., 2018).

La figure I.12 à droite affiche la localisation de quelques stations des réseaux de validation de température de surface situés en Amérique du Nord, en Europe et en Afrique. D'autres stations sont disponibles dans d'autres réseaux (par exemple, *Ground based observation for validation, Baseline Surface Radiation Network*) mais le nombre total reste faible (≤ 100). Des radiomètres mobiles sont parfois utilisés pour conduire des expérimentations sur certaines surfaces. La précision de la mesure de la température atteinte pour les stations in situ est importante (≤ 1 K) mais leur couverture spatiale est extrêmement

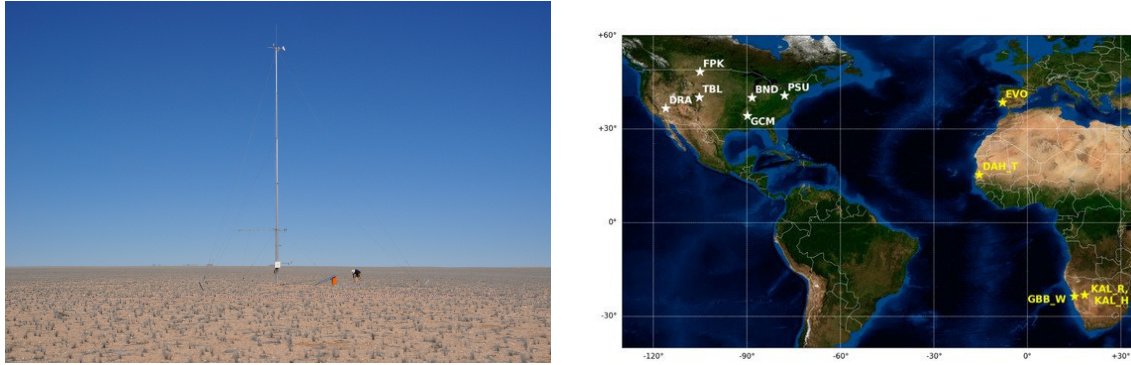


FIGURE I.12 – A gauche : photo du site de mesure in situ de Goabeb Plains station, Namibie (extraite de Basel University). À droite : Carte de la localisation des stations in situ disponibles au sein du réseau *Surface Radiation Network* (SURFRAD) en blanc et *Karlsruhe institute of technology* (KIT) en jaune (figure extraite de Martin et al., 2019).

limitée. Elles sont principalement utilisées pour le développement de modèles ou pour la validation des produits de température de surface (Martin et al., 2019).

Quelques rares stations effectuent des mesures à différentes profondeurs qui permettent de valider les flux thermiques modélisés (G dans l'équation I.1). Des données des profils de température du sol sont aussi disponibles comme sous-produit des stations de mesures de l'humidité des sols (*International Soil Moisture Network*). Malheureusement ces stations ne mesurent pas la température radiométrique en surface, mais uniquement la température à différentes profondeurs.

3.2 Observations infrarouges

L'observation radiométrique en infrarouge est la méthode de mesure pour les températures de surfaces dans les stations in situ. Elle peut aussi être employée avec des instruments à bord de satellites en orbite autour de la Terre. Le principe de la télédétection en infrarouge est similaire à celui présenté pour les microondes, c'est-à-dire par la mesure d'une température de brillance. En connaissant l'émissivité de la surface dans les conditions de l'observation, on en déduit la température de la surface. Ces mesures sont effectuées en utilisant des canaux proches du maximum de la fonction de Planck pour les températures issues de la surface terrestre et situées dans les bandes d'absorption minimale de l'atmosphère ($\lambda \approx 11 \mu m$). Quelques différences par rapport aux microondes méritent d'être soulignées :

En premier lieu, l'approximation de Rayleigh-Jeans n'est pas applicable aux observations infrarouges, mais les fréquences utilisées sont proches du maximum d'émission d'un corps à la température de la Terre. Dans ce cas la loi de Stefan-Boltzmann, avec σ la constante de Stefan-Boltzmann, T la température de surface et e l'émissivité en infrarouge, établit la relation entre le flux énergétique émis (F) et la température du corps. Elle est de la forme : $F \propto e_{IR}\sigma T^4$.

En second lieu, le fait que les émissivités de la Terre dans l'infrarouge sont très proches de 1. On constate aussi sur la figure I.13 que celles-ci sont très homogènes sur la plupart des surfaces terrestres, excepté les zones de désert. De plus, elles sont stables face à des variations d'humidité des sols.

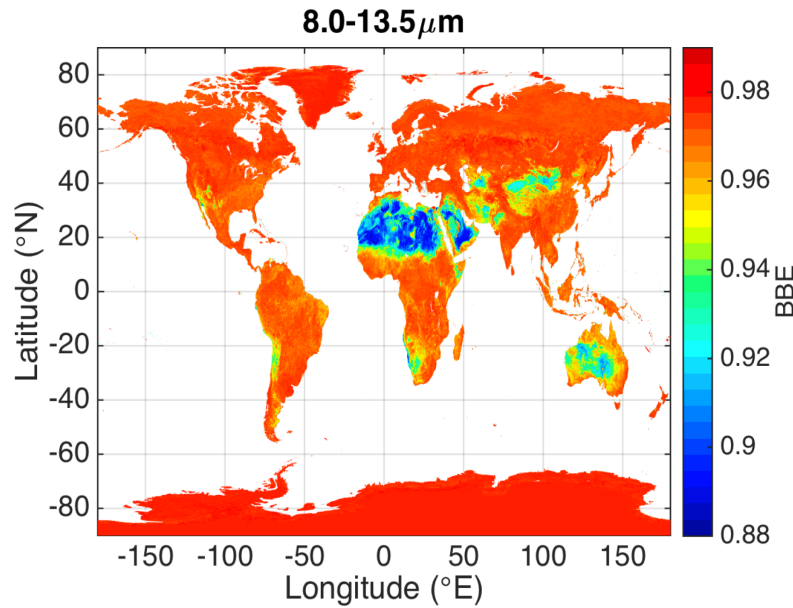


FIGURE I.13 – Emissivité des surfaces en infrarouge. (figure extraite de Feltz et al., 2018)

Ces effets combinés font que le signal mesuré est très sensible aux variations de la température de la surface. La température mesurée par l'infrarouge est donc directement liée à la température de surface utilisée dans le bilan radiatif.

Les observations peuvent être effectuées à la fois par des satellites en orbite polaire à une altitude de ~ 800 km (Prata et al., 1995) avec une résolution spatiale allant jusqu'à la centaine de mètres et des temps de revisite variables suivant la fauchée de l'instrument de 2 fois par jours à quelques jours d'écarts. Les satellites en orbite géostationnaire (36 000 km) fournissent des observations avec une résolution spatiale allant jusqu'à 2 km et une résolution temporelle proche de 15 minutes (Schmit et al., 2017). Cette haute résolution temporelle permet de bien représenter la dynamique du cycle diurne ainsi que de pouvoir bénéficier au mieux des fenêtres sans nuages. Cependant, les observations des satellites géostationnaires sont limitées par une couverture réduite en latitude ($\pm 60^{\circ}\text{N}$), et la nécessité de combiner plusieurs instruments géostationnaires pour observer toutes les longitudes (Freitas et al., 2013).

Parmi les instruments utilisés pour la mesure de températures de brillance infrarouges, on peut citer les familles d'instruments suivantes qui se différencient par les fréquences des canaux embarqués, et la résolution spatiale et temporelle des produits fournis.

- Les imageurs à bord de satellites en orbite polaire héliosynchrone comme le *Moderate Resolution Imaging Spectroradiometer* (MODIS) ou l'*Advanced Very-High-Resolution Radiometer* (AVHRR).
- Les satellites géostationnaires par exemple *Geostationary Operational Environmental Satellite* (GOES) muni du *Advanced Baseline Imager*, ou le *Spinning Enhanced Visible and Infrared Imager* (SEVIRI) à bord de Météosat. Ce dernier est utilisé dans la création des données de températures issues de l'infrarouge du *Land Surface Analysis Satellite Application Facility* (LSA SAF).
- Les imageurs hyperspectraux comme *Infrared atmospheric sounding interferometer* (IASI) peuvent aussi être utilisés pour estimer la température de surface (Paul et al., 2012).

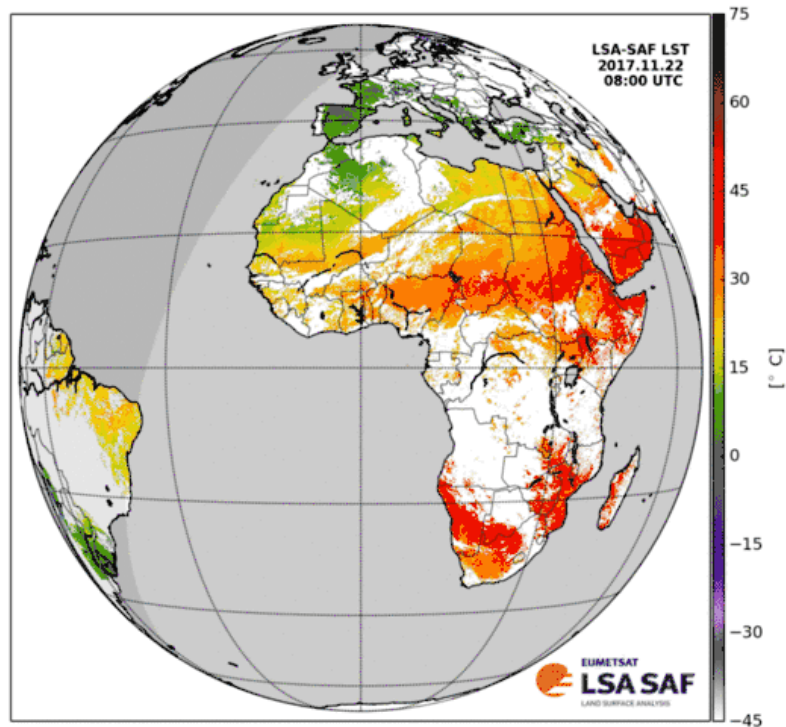


FIGURE I.14 – Exemple d’estimation de température de surface des continents ($^{\circ}\text{C}$) estimées à partir des mesures par un instrument infrarouge à bord d’un satellite géostationnaire le 22 novembre 2017 à 08:00 (figure extraite du site du LSA SAF).

Pour estimer la température de surface, il faut connaître l’effet de l’atmosphère sur l’observation (absorption et diffusion) et l’émissivité de la surface observée. Plusieurs méthodes ont été développées pour estimer la température à partir d’observations en infrarouge, avec l’estimation ou non des émissivités en parallèle.

En supposant une connaissance de l’émissivité et du profil atmosphérique, des tables de correspondance (*Lookup tables*) ont été construites permettant d’estimer la température. Quand plusieurs canaux sont disponibles, une combinaison linéaire ou polynomiale des températures de brillance mesurées pour chacun des canaux permet d’éliminer la nécessité de connaître l’effet de l’atmosphère (Wan and Li, 1997). Les émissivités peuvent quant à elles être calculées en effectuant par exemple des combinaisons pondérées des émissivités des types de surface et de la végétation, par leur fraction respective à la surface d’une observation (Trigo et al., 2008). Une revue de différentes méthodes d’estimation de la température de surface a été faite par Li et al. (2013).

L’inconvénient majeur de ces observations provient de la sensibilité très importante aux nuages qui sont opaques aux fréquences utilisées. Les observations de la surface ne sont donc possibles que par temps clair. On peut le constater sur la figure I.14 où pour une observation à une heure donnée de nombreuses températures ne peuvent pas être estimées. L’utilisation de produits composites, moyennes quotidiennes, ou par 10 jours permet d’obtenir une couverture quasi-complète.

Les sources principales d’erreur de ces produits proviennent :

- Des contaminations nuageuses malgré les efforts effectués pour filtrer ces points, la présence de nuages ayant des tailles inférieures à la résolution spatiale de l’instrument

pouvant poser des problèmes.

- La quantification de la vapeur d'eau atmosphérique.
- Des estimations des émissivités (en particulier sur les déserts).
- Des étalonnages et inter-étalonnages entre les multiples instruments.
- Des incertitudes causées par l'angle d'observation de la surface qui est variable pour les satellites géostationnaires ont aussi été observés (Ermida et al., 2014).

3.3 Observations microondes

De multiples satellites en orbite polaire fournissent des observations à différentes fréquences en microonde passive. En utilisant l'approximation de Rayleigh-Jeans (équation I.4), l'équation de transfert radiatif et des données auxiliaires, il est possible d'estimer la température de la surface. Les fréquences les plus utilisées sont celles situées dans les fenêtres de transmission atmosphérique, pour obtenir la contribution maximale de la surface à la température de brillance mesurée.

Avec l'approximation de Rayleigh-Jeans, le rayonnement de la surface est proportionnel au produit entre l'émissivité et la température de la surface. Par rapport aux observations dans l'infrarouge, le signal mesuré est moins sensible aux variations de température et plus sensible aux variations d'émissivité. Comme dans le cas des observations infrarouges, l'estimation de la température de surface est un problème sous-contraint c'est-à-dire que le nombre d'observations est inférieur au nombre de variables. En microonde, si on utilise une fréquence mesurée dans les deux polarisations, (V et H), il faut déterminer 2 émissivités ($e_{\nu,H}, e_{\nu,V}$) et une température.

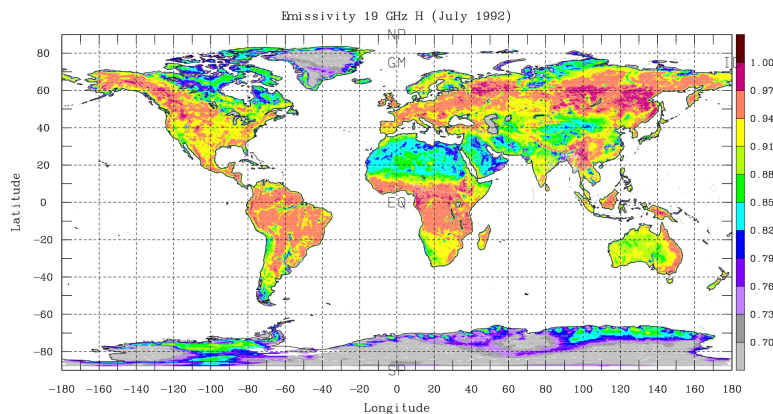


FIGURE I.15 – Carte des émissivités des surfaces continentales à 19 GHz en polarisation horizontale (figure extraite de Prigent et al., 1998a).

Comme on peut le voir sur la figure I.15 les émissivités varient fortement avec la surface en microondes (de ~ 0.6 à ~ 1 en polarisation horizontale). On distingue des structures spatiales associées aux zones arides, aux forêts, et à la présence d'eau en surface. Les émissivités sont aussi en moyenne plus faibles et plus variables temporellement en microonde qu'en infrarouge (figure I.13).

Plusieurs méthodes ont été développées dès la fin des années 1990 pour estimer la température de surface à partir d'observations microondes, une revue méthodique en est effectuée par Duan et al. (2020). Pour toutes les méthodes, le choix des canaux utilisés

résulte d'un compromis entre la résolution spatiale de l'observation (qui diminue avec les fréquences) et l'augmentation de la contribution atmosphérique avec la fréquence. De plus, dans les zones arides la profondeur de pénétration des microondes diminue avec la fréquence et l'hypothèse d'une équivalence entre la T_{eff} et la température de surface n'est pas vérifiée aux basses fréquences.

Une approche simple utilise uniquement le canal à 37 GHz (Holmes et al., 2009). La polarisation verticale est utilisée car l'émissivité est plus élevée et qu'elle varie moins avec l'humidité des sols qu'en polarisation horizontale. De plus au-dessus des continents la contribution de l'atmosphère reste limitée à cette fréquence. Les paramètres de la relation linéaire entre $T_{b,37V}$ et la température de surface sont estimés pour des colocalisations d'observations de l'instrument AMSR-E et de stations in situ de mesure de température. Des méthodes plus complexes estiment la température comme une combinaison linéaire d'un plus grand nombre de canaux (Pulliainen et al., 1997) ou sur des zones géographiques limitées (Zhao et al., 2011). Elles font aussi des hypothèses fortes sur les émissivités de canaux proches pour réduire le nombre de variables de l'inversion (Weng and Grody, 1998).

D'autres méthodes font une inversion variationnelle de multiples canaux, pour estimer simultanément la température à la surface, l'émissivité et les paramètres atmosphériques (Prigent et al., 1997; Prigent and Rossow, 1999). Cette méthode utilise une première estimation des émissivités microondes (Prigent et al., 1998b). En reprenant les résultats de l'inversion variationnelle, une autre approche a été introduite par Aires et al. (2001) pour inverser les observations microondes passives et estimer les températures de surface. Elle repose sur un modèle empirique basé sur de l'apprentissage automatique d'un réseau de neurones qui permet d'obtenir plus rapidement des résultats similaires à l'approche variationnelle. La régression non linéaire de plusieurs canaux s'appuie aussi sur des a priori d'émissivité et permet à la fois de prendre en compte l'effet de l'atmosphère, mais aussi des changements d'émissivité de surface. Plus de détails sur cette méthodologie qui est employée pour effectuer les estimations de température de surface au sein du projet LST CCI sont donnés dans la section 4.

3.4 Modèles de surface

La température de surface peut aussi provenir de modèles de surface en réalisant le bilan de l'équation I.1 (Boone et al., 1999). Ces modèles sont utilisés dans les centres de prévision numérique du temps (Nogueira et al., 2020) mais aussi dans les analyses de sensibilité climatiques (Van Den Hurk et al., 2016; Wyser et al., 2020).

Dans les centres de prévision météo, des analyses sont effectuées où les paramètres provenant d'une modélisation globale du système Terre (incluant un modèle de surface terrestre) sont propagées par des codes de transfert radiatif pour simuler les températures de brillance observées par divers instruments. L'écart entre ces observations modélisées et les observations réelles des instruments est corrigé pour contraindre la prévision du modèle physique en tenant compte des erreurs du modèle et des erreurs d'observations. Cette assimilation de données entre l'état du système Terre fourni par un modèle physique avec des observations provenant de multiples sources sert de base à la prévision numérique du temps. Elle peut aussi être employée avec des données historiques et permet la création de réanalyses. Ces résultats de réanalyses sont utilisés pour mieux comprendre les évolutions de la météo et du climat passés. Dans les réanalyses, l'utilisation de nouvelles versions du modèle physique ou l'ajout de données utilisables peuvent permettre d'améliorer les résultats par rapport aux analyses météorologiques en temps réel. La résolution spatiale

et temporelle est fixe et dépend principalement des capacités de calcul et des modèles utilisés.

Produit	Résolution	Durée	Couverture
ERA5	31 km - 1 h	1950-2021	Globale
ERA5-Land	9 km - 1 h	1981-2021	Globale
ERA-Interim	80 km - 6 h	1979 -2019	Globale
NCEP NCAR	250 km - 6 h	1948 - 2021	Globale
JRA55	150 km - 3 h	1958 - 2021	Globale

TABLE I.1 – Caractéristiques des principales réanalyses disponibles. (*reanalyses.org*)

Parmi les réanalyses listées dans la table I.1, on peut citer celles fournies par le Centre européen pour les prévisions météorologiques à moyen terme ou *European Centre for Medium-Range Weather Forecasts* (ECMWF) dont la dernière version ERA5 (Hersbach et al., 2020) a une résolution spatiale de 31 km et temporelle d’une heure de 1950 à 2021. ERA5-Land (Muñoz-Sabater et al., 2021) fournit aussi les données pour les couches dans la sub-surface (température à différentes profondeurs), mais sans effectuer d’assimilation d’observations. Pour ERA5-Land, le bilan radiatif d’ERA5 est utilisé et interpolé à la résolution spatiale de 9 km entre 1981 et 2021. C’est le modèle *Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology* (HTESSEL) qui est employé pour représenter les interactions dans le sol dans ces analyses.

La température de surface est centrale dans la modélisation des bilans thermiques. Elle est donc restituée à chaque cycle d’assimilation du modèle. Cependant, les observations ne sont pas directement assimilées pour cette variable, mais son évolution est contrainte dans le modèle physique de surface par les échanges avec les couches les plus basses de l’atmosphère (T_{2m}). Un second inconvénient lié à l’utilisation de températures de surface issues d’une réanalyse est que la résolution spatiale est moins bonne que celle provenant de satellites infrarouges par exemple.

Des modèles sont aussi employés pour estimer l’évolution de la température sous les nuages dans les observations infrarouges. Les données manquantes sont estimées en contraignant un modèle de surface avec les observations disponibles (Martins et al., 2019).

La table I.2 résume les caractéristiques des sources de données présentées dans cette section ainsi que leurs principaux avantages et inconvénients. On constate bien la complémentarité entre les différentes sources, qu’elle soit sur la couverture spatiale, sur la résolution spatiale et temporelle des observations ou sur la disponibilité « tout temps » des estimations. Une revue de différentes méthodes d’estimation de la température de surface a été faite par Li et al. (2013) qui identifie différentes pistes pour l’amélioration des produits de température de surface issus des observations satellites.

4 Une méthode d’estimation de la température de surface par des observations microondes passives

Cette section détaille le fonctionnement de l’algorithme utilisé pour produire la température de surface à partir d’observations microondes qui sont du projet *Land Surface Temperature - Climate Change Initiative* (LST CCI). La seconde partie recense les princi-

Source	Résolution spatiale	Résolution temporelle	Couverture spatiale	Inconvénients	Avantages
Infrarouge géostationnaire	2 - 5 km	15 min	$\pm 60^\circ\text{N}$	Opacité des nuages, couverture spatiale limitée	Précision des estimations
Infrarouge polaire	100 m - 1 km	1 - 7 jours	Globale	Opacité des nuages	Haute résolution spatiale, précision des estimations
Modèles de surface	9 - 25 km	1 h	Globale	Résolution spatiale, dépendance aux autres paramètres	« Tout temps »
In situ	10 m	1 min	Locale	Couverture très limitée	Précision des estimations, stabilité temporelle, « Tout temps »
Microonde	15 - 25 km	1 jour	Globale	Résolution spatiale, précision des mesures, pénétration de la surface	« Tout temps »

TABLE I.2 – Table résumant les résolutions spatiales et temporelles, la couverture géographique, les inconvénients et les avantages des principales sources de données de température de surface.

pales limites rencontrées par les méthodes utilisant des observations microondes passives pour estimer la température de surface.

4.1 Méthode utilisée pour l'estimation de la température de surface dans le projet *Land Surface Temperature - Climate Change Initiative*

La méthode présentée dans cette section a fait l'objet de nombreux développements. On pourra citer les travaux de Prigent et al. (1998a); Aires et al. (2001), le document de référence du LST CCI (2019), et les évaluations conduites par Jiménez et al. (2017); Ermida et al. (2017). Cette méthode utilise des canaux entre 19 et 89 GHz. À partir de l'équation du transfert radiatif, en supposant que la même température est mesurée par toutes les fréquences, la température de surface T , pour une fréquence ν s'écrit :

$$T = \frac{T_{b,\nu,P} - T_{\uparrow,\nu} - \tau_\nu(1 - e_{\nu,P})T_{\downarrow,\nu}}{\tau_\nu e_{\nu,P}}. \quad (\text{I.6})$$

En dépit de la relative insensibilité de ces fréquences au contenu de la colonne d'atmosphère, sa contribution doit être prise en compte pour déterminer les composantes de l'équation du transfert radiatif (eq. I.6). Pour estimer la température, l'émissivité de la surface doit aussi être calculée. Résoudre ce système d'équation pour toutes les fréquences est donc un problème sous contraint.

La figure I.16 détaille les différentes étapes de la création d'un produit de température de surface estimés à partir d'observations microondes. Toutes les données sont utilisées à une échelle globale.

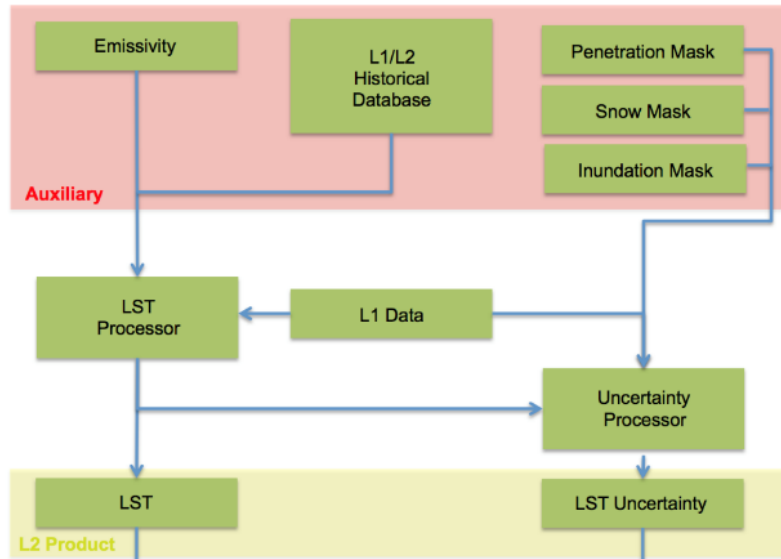


FIGURE I.16 – Diagramme représentant les différentes étapes de la création du jeu de données de température de surface, dérivées des observations microondes pour le projet LST CCI (figure extraite de l’*Algorithm Theoretical Basis Documents* - Document de Description Théorique des Algorithmes du LST CCI).

Tout d’abord des données auxiliaires sont regroupées :

1. Les émissivités moyennes mensuelles pré-calculées par l’outil *Tool to Estimate Land Surface Emissivity in the Microwave* (TELSEM) (Aires et al., 2011c).
2. Les climatologies de données historiques de températures de brillance aux fréquences considérées qui sont utilisées pour l’entraînement
3. Les masques contenant des indicateurs de zones présentant des difficultés pour l’estimation comme les zones avec une forte profondeur de pénétration, de la neige, ou des inondations

Le cœur du traitement (*LST Processor*) utilise les températures de brillance observées par un instrument (*L1 Data*), et les données d’émissivités et les données historiques en entrée d’un réseau de neurones pour effectuer l’inversion. Aucune information additionnelle provenant d’analyses n’est utilisée, l’information sur l’état de l’atmosphère étant présente à travers les canaux à ~ 23 GHz (proche d’une ligne d’absorption de vapeur d’eau) et à 89 GHz.

Le réseau de neurones est un modèle statistique (plus de détails sur le fonctionnement de ce type de modèles sont donnés au chapitre III) qui est au préalable entraîné. Des observations coïncidentes de températures de brillance, et de données auxiliaires sont fournies en entrée. En sortie, les paramètres de surface observés provenant du projet *International Satellite Cloud Climatology Project* (ISCCP) sont utilisés. Les paramètres de cette régression non linéaire entre les T_b et les émissivités en entrée et les températures de surface sont donc appris pour ces données d’entraînement en minimisant l’erreur quadratique entre les sorties prédites par le modèle et les températures de surface de la base d’apprentissage. L’erreur finale du modèle sur des données de validation est inférieure à 2 K pour 70 % des données.

Une estimation de l’incertitude est effectuée en parallèle, en détaillant plusieurs types d’incertitudes :

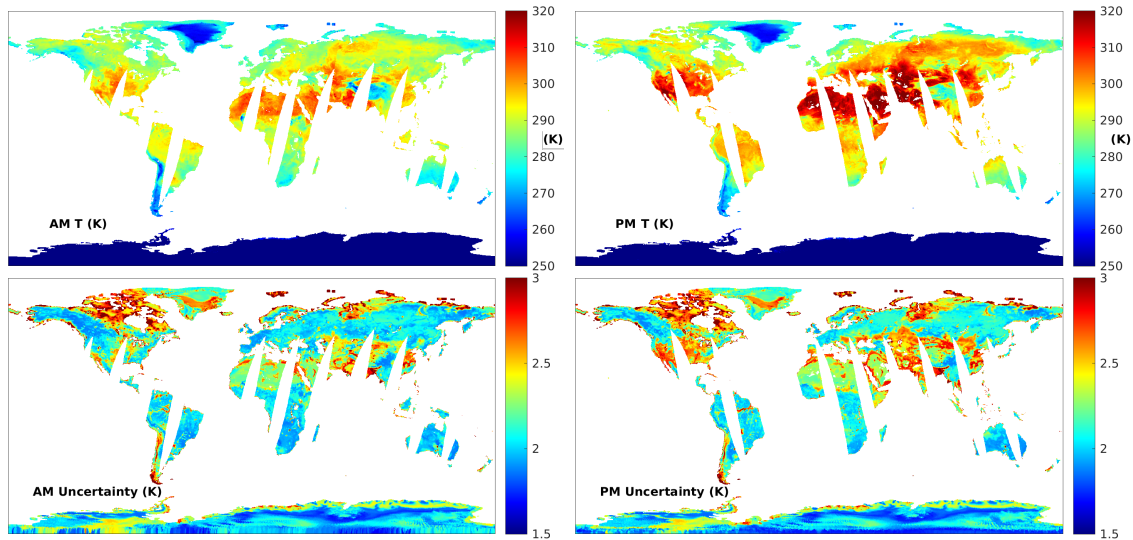


FIGURE I.17 – Estimations des températures de surface (K) le 2 juillet 2012 pour l’instrument SSMIS par la méthode décrite dans cette section. À gauche pour une observation à 06:00 heure locale à l’équateur et à droite pour une observation à 18:00 heure locale à l’équateur. En haut la température et en bas l’incertitude sur l’estimation de température de surface (figure extraite des rapports du LST CCI).

- La variabilité liée aux paramètres du réseau de neurones : elle est estimée en entraînant un ensemble de 100 réseaux de neurones et calcule la moyenne et la déviation standard de la sortie. Cela permet de cibler les zones pour lesquelles la convergence du modèle n’est pas bonne due par exemple à un manque d’observation ou une variabilité naturelle des émissivités importante (e.g., zones de neige).
- L’incertitude liée aux entrées du modèle : les entrées contiennent un bruit lié à la mesure de température de brillance ou des erreurs dans les a priori des émissivités. L’effet de ces incertitudes dans la prédiction du modèle est calculé en fonction des paramètres d’entrée.

Ces sources d’incertitudes ne prennent pas en compte l’effet des erreurs potentielles dans la température de surface utilisée pour l’entraînement. Celle-ci est en effet elle-même issue d’une estimation et n’est donc pas dénuée d’erreur.

Toutes ces étapes permettent de calculer la température de surface avec son incertitude à partir des observations microondes d’un instrument. Ce procédé a été appliqué aux données de plusieurs instruments dont SSMIS F13 (figure I.17). La résolution du canal à 19 GHz de cet instrument est de 60 km, mais des tests ont montré que le canal à 37 GHz est celui avec le plus de poids dans l’estimation de la température. On considère donc que les températures de surface sont restituées avec une résolution de ~ 30 km.

La figure I.17 présente les cartes des températures de surface et de l’incertitude associée estimées pour l’instrument SSMIS le 02 juillet 2012. Les différentes orbites de SSMIS sont agrégées pour fournir la température de surface à une échelle globale pour cette journée. On constate bien l’écart de température induit par le cycle diurne entre le passage du matin (06:00) et du soir (18:00). Les incertitudes les plus élevées apparaissent pour certaines zones arides, les zones avec de la neige (e.g. Nord de la Sibérie) ou les zones de cotes.

4.2 Limites des inversions de température de surface en microondes

La section précédente présente une méthodologie développée pour résoudre le problème inverse de la télédétection de la température de surface à partir d'observations microondes. Cette méthode, comme les autres méthodes développées pour estimer les températures de surface à partir d'observations microondes, a certaines limites. Duan et al. (2020) liste un certain nombre des problèmes qui affectent les estimations de température de surface à partir d'observations microondes de manière générale :

- Le processus d'estimation de l'émissivité et de la température de surface basé sur l'inversion est un problème sous-contraint. Le nombre d'observations disponibles est toujours inférieur au nombre de degrés de liberté du problème. L'incertitude sur l'émissivité par exemple cause directement une incertitude sur la température de surface.
- En première approximation, de nombreuses méthodes ignorent l'impact de l'atmosphère sur les observations. Cependant, cela conduit à des erreurs systématiques quand l'atmosphère n'est pas compatible avec les hypothèses effectuées. Même quand l'atmosphère est prise en compte, des erreurs dans le profil de vapeur d'eau ou d'eau liquide atmosphérique transparaissent immédiatement dans le résultat de l'estimation de la température de surface. Dans la méthodologie utilisée ici, l'atmosphère est en grande partie prise en compte. Mais en cas de nuages la méthode échoue.
- La profondeur de pénétration de l'infrarouge et des microondes est différente, surtout dans les régions arides. La contribution d'une épaisseur plus importante du sol au rayonnement microonde dans certaines zones arides est visible et conduit à des biais systématiques entre ces deux sources d'observation.
- Finalement, les résolutions spatiales des produits de température de surface microondes sont difficilement comparables aux mesures in situ en raison de la très large empreinte au sol comparée à la représentativité de ces stations. L'hétérogénéité spatio-temporelle de la température limite l'utilisation de la température estimée par microonde pour un certain nombre d'applications.

Les méthodes existantes traitent certains des points soulevés, par exemple en utilisant les informations des canaux à 23 et 89 GHz plus sensibles au contenu en eau de l'atmosphère.

A ces défis posés à la télédétection microonde, s'ajoutent les contraintes inhérentes à la construction de séries temporelles. De longues séries chronologiques d'estimations de la température de la surface terrestre à l'aide d'observations microondes passives sont en cours de construction au sein du projet LST CCI.

- Des instruments différents sont nécessaires pour couvrir de longues séries temporelles, car la durée de vie des instruments ne dépasse pas 10 années. Les différents instruments peuvent avoir des caractéristiques différentes (e.g., fréquences, heures de passage, angle d'incidence) et n'utilisent pas tous les mêmes méthodes d'étalonnage. Des corrections peuvent être nécessaires.
- La stabilité des mesures effectuées par un instrument au cours du temps doit être garantie. Pour pouvoir estimer des tendances environnementales de long terme, il est nécessaire de supprimer les évolutions artificielles liées à l'instrument.
- Enfin, les mesures doivent être cohérentes entre elles malgré des observations qui diffèrent par des aspects physiques, comme l'angle de visée ou l'heure de passage.

5 Objectifs et contenu de la thèse

Ce chapitre introductif présente la définition d'une température de surface telle qu'observée par un satellite. Le principe de l'estimation de la température de surface à partir de mesures de température de brillance en microonde est détaillé. L'utilité des mesures microondes pour compléter les autres sources de température de surface en particulier en présence de nuage apparaît clairement. Mais quelques limitations de ces méthodes et des travaux nécessaires à la création de séries temporelles ont été identifiés (section 4.2). C'est donc sur ces axes que le travail de cette thèse s'est orienté. Pour répondre à un certain nombre des défis soulevés, cette thèse développe plusieurs méthodologies, qui feront l'objet des prochains chapitres.

Tout d'abord, pour créer de longues séries temporelles des instruments différents doivent être utilisés. Les différences entre des observations de provenances hétérogènes doivent être analysées. Les données fournies par les imageurs microondes disponibles sont présentées. L'analyse des observations fournies par l'instrument le plus ancien, le *Scanning Microwave Multichannel Radiometer* (SMMR), nécessaire pour étendre la durée des séries temporelles de 10 années supplémentaires, met en évidence des erreurs au-dessus des surfaces continentales. Une correction est donc développée en utilisant un instrument de référence, le *Global Precipitation Mission Microwave Imager* (GMI), plus récent et mieux étalonné. Les méthodes d'inter-étalonnage et les résultats obtenus sont présentés dans le chapitre II (*Imageurs microondes passifs et leur inter-étalonnage au-dessus des surfaces continentales pour la création de séries temporelles*).

Les estimations de températures reposent principalement sur les canaux fenêtres dans lesquels l'effet de l'atmosphère est faible. On a aussi vu que les microondes en raison de leur longueur d'onde étaient peu sensibles à la présence des nuages. Cependant, on note la présence de contaminations dues aux nuages épais dans les estimations des températures de surface. Pour détecter ces contaminations et limiter leur impact sur les estimations, un modèle polyvalent de détection des nuages en microondes est développé basé sur un réseau de neurones. Il permet d'estimer la probabilité qu'une observation soit contaminée, en fonction des fréquences disponibles dans l'observation. Les résultats présentés dans le chapitre III (*La contamination nuageuse des observations microondes utilisées pour la caractérisation des surfaces*) mettent en évidence une contamination variable suivant le type de nuages ce qui permet d'améliorer le filtrage de ces phénomènes.

Dans les zones arides, la faible humidité des sols peut engendrer des valeurs importantes de profondeur de pénétration en microondes. De plus la température dans le sol n'est pas homogène et la température en surface peut être très différente de celle dans la sub-surface ce qui pose des problèmes lors de l'inversion des températures. Pour quantifier cette pénétration des ondes, la température dans le sol est modélisée ainsi que les contributions atmosphériques aux différentes longueurs d'ondes. Grâce à l'instrument GMI, des cartes de profondeur de pénétration sont estimées, entre 10 GHz et 90 GHz, sur l'Afrique du Nord et la péninsule arabique (chapitre IV *La pénétration des microondes dans la sub-surface des zones arides*).

Pour favoriser l'utilisation des températures de surface estimées à partir d'observations microondes pour le plus grand nombre d'applications possibles, une méthode est développée visant à augmenter la résolution spatiale de ces données. Les températures de surface sont disponibles en infrarouge à une plus haute résolution. En combinant celles-ci avec des variables auxiliaires qui décrivent la dynamique temporelle du cycle diurne de tempéra-

ture et les structures spatiales fines, il est possible de construire une méthode permettant de désagréger les températures de surface issues des microondes à une résolution 5 fois plus fine. Les étapes pour appliquer cette procédure, « tout temps », est détaillée dans le chapitre V (*Amélioration de la résolution spatiale des estimations de température de surface*).

En conclusion, quelques résultats présentant l'état actuel des séries temporelles créées par le projet LST CCI sont montrées. L'application des résultats de cette thèse pour ce projet et pour l'utilisation des microondes passives est détaillé. Pour finir les axes de recherche prometteurs identifiés dans cette thèse sont listés.

Publications

- Le second chapitre traite des imageurs passifs et de l'inter-étalonnage de l'instrument SMMR avec GMI. Une publication en est issue : Favrichon, S., Jimenez, C., and Prigent, C. (2020). Inter-calibrating SMMR brightness temperatures over continental surfaces. *Atmospheric Measurement Techniques*, 13(10) :5481–5490.
- Le troisième chapitre aborde le problème de la détection des contaminations nuageuses dans les observations microondes. Les résultats ont été publiés : Favrichon, S., Prigent, C., Jimenez, C., and Aires, F. (2019). Detecting cloud contamination in passive microwave satellite measurements over land. *Atmospheric Measurement Techniques*, 12(3) :1531–1543.
- Le quatrième chapitre étudie la profondeur de pénétration des microondes dans les zones arides, en vue de proposer une correction pour rendre comparable les estimations issues des microondes à celles issues de l'infrarouge. Un article est en cours de préparation sur cette thématique.
- Le cinquième chapitre propose une méthode de désagrégation applicable aux observations de température de surface issues des microondes passives, quelle que soit l'heure de l'observation et la couverture nuageuse. Une publication en est tirée : Favrichon, S., Prigent, C., and Jiménez, C. (2021). A method to downscale satellite microwave land-surface temperature. *Remote Sensing*, 13(7) :1–20.

Chapitre II

Imageurs microondes passifs et leur inter-étalonnage au-dessus des surfaces continentales pour la création de séries temporelles

Ce chapitre explore les données microondes disponibles pour la restitution de la température de surface. Les imageurs passifs microondes existants en détaillant ceux utilisés dans cette thèse sont présentés. Les problèmes posés par l'utilisation de plusieurs instruments indépendants pour créer de longues séries temporelles de mesures cohérentes sont décrits, en particulier pour le cas de l'inter-étalonnage de l'instrument Scanning Microwave Multichannel Radiometer (SMMR). Pour celui-ci, une correction des données observées est proposée pour permettre une utilisation des mesures homogène avec les instruments plus récents. Les problèmes du choix de l'instrument de référence, des techniques d'étalonnage ainsi que des plages de températures étalonnées au-dessus des continents sont ainsi abordés.

Pour créer une série temporelle de température de surface issue d'observations microondes couvrant les 40 dernières années, l'utilisation d'observations provenant de plusieurs instruments est nécessaire. Cette section s'intéresse donc spécifiquement aux sources de données satellites disponibles et à la validation de l'étalonnage de ces instruments.

Dans le domaine des microondes, les observations disponibles sont effectuées grâce à des instruments sur des plateformes satellites comme le *Scanning Microwave Multichannel Radiometer* (SMMR) à partir de 1978, les *Special Sensor Microwave / Imager* (SSM/I) et le *Special Sensor Microwave Imager / Sounder* (SSMIS) de 1987 jusqu'à nos jours, les instruments *Advanced Microwave Sounding Radiometer* (AMS-R) depuis 2002 ou encore le récent *Global Precipitation Mission Microwave Imager* (GMI) lancé en 2015. De futures missions sont prévues pour poursuivre l'observation de la Terre dans ce domaine de fréquence, permettant ainsi de pérenniser les séries temporelles.

Pour assurer la cohérence des estimations dérivées des observations microondes sur de

longues séries temporelles, les mesures effectuées par des instruments différents requièrent un traitement préliminaire. Une fois les données homogènes, les mêmes méthodes d'estimation de la température de surface décrites au chapitre précédent peuvent être appliquées. C'est dans cette optique que sont créés les *Fundamental Climate Data Records* (FCDR) où les données brutes (niveau 1 voir section 1.3) de plusieurs instruments sont rendues homogènes et cohérentes. Plusieurs FCDR ont été produits à partir des mesures effectuées par les imageurs microondes passifs, depuis les instruments pionniers comme SMMR jusqu'à GMI. Jusqu'à présent, les efforts se sont essentiellement concentrés sur les océans où les simulations de transfert radiatif alimentées par la réanalyse peuvent servir de référence pour inter-étalonner des instruments.

Au-dessus des surfaces continentales, l'inter-étalonnage des imageurs microondes passifs est un défi, en particulier pour les satellites polaires héliosynchrones qui ont des heures de passage différentes à l'équateur. C'est pourquoi une méthode d'inter-étalonnage est décrite, permettant de mettre en lumière et de corriger des différences majeures entre SMMR et les autres instruments microondes, pour des fréquences à 18 GHz et 37 GHz, au-dessus des surfaces continentales.

1 Données d'imageurs microondes disponibles

Plusieurs instruments fournissent des mesures de température de brillance dans les microondes, à des fréquences utiles pour l'estimation de la température de surface (chapitre I, section 3.3). Certains sont actuellement en orbite autour de la terre, plusieurs générations les ont précédés, et de nouveaux instruments sont prévus pour les prochaines décennies. Cette section donne un aperçu des instruments existants, des jeux de données fournis et détaille les caractéristiques de ceux qui seront utilisés dans cette thèse. Les instruments décrits ci-dessous sont tous des radiomètres multi-fréquences, utilisant des canaux entre 6 et 183 GHz.

Tout d'abord le principe de fonctionnement des radiomètres microondes passifs est introduit, puis les caractéristiques de plusieurs instruments sont détaillées et enfin les niveaux de produits satellite explicités. Les instruments décrits sont : l'instrument pionnier des observations multi-canaux SMMR, puis SSM/I et SSMIS qui fournissent la plus longue couverture temporelle par des instruments similaires, les deux imageurs AMSR et GMI le plus récent. Les principales caractéristiques de ces instruments ainsi que quelques autres imageurs qui ne sont pas utilisés dans cette thèse comme le *Tropical Rainfall Measuring Mission's* (TRMM) *Microwave Imager* (TMI) ou le *Microwave Analysis and Detection of Rain and Atmospheric Structures* (MADRAS) sont données dans le tableau II.1.

1.1 Principe de fonctionnement d'un radiomètre microonde passif

La réception du signal à une fréquence donnée est effectué grâce à un récepteur microonde combiné à des systèmes d'amplification (Dicke, 1946). Le signal provenant de la Terre est capturé par une antenne parabolique réceptrice, qui concentre le rayonnement vers un certain nombre de cornets de réception correspondants aux différents canaux et polarisations mesurés derrière lesquels sont installés les chaînes de détection. À chaque fréquence, les différents éléments de cette chaîne de réception/détection participent au bruit instrumental. Une bande de fréquence d'une largeur sélectionnée autour de la fréquence d'intérêt est mesurée et le temps d'intégration de la mesure est ajusté pour obtenir une

Imageur	Canaux (GHz) (Polarisation)						Résolution spatiale (à 37 GHz)	Angle d'incidence	Années d'opération
	~ 18 GHz	~ 23 GHz	~ 37 GHz	~ 89 GHz	165.5 GHz	183.3 GHz			
SMMR	18.0 (V,H)	21.0 (V,H)	37.0 (V,H)	–	–	–	29km * 17km	50.2°	1978–1987
SSM/I	19.4 (V,H)	22.2 (V)	37.0 (V,H)	85.5 (V,H)	–	–	36km * 24km	53.1°	1987–2006
TMI	19.35 (V,H)	21.3 (V)	37 (V,H)	85.5 (V,H)	–	–	16km * 9km	53.0°	1998–2015
AMSR-E	18.7 (V,H)	23.8 (V,H)	36.5 (V,H)	89.0 (V,H)	–	–	14km * 9km	55.0°	2002–2011
MADRAS	18.7 (V,H)	23.8 (V)	36.5 (V,H)	89.0 (V,H)	157 (V,H)	–	40km * 16km	56.0°	2013
SSMIS	19.4 (V,H)	22.2 (V)	37.0 (V,H)	91.6 (V,H)	150 (H)	H	44km * 28km	53.1°	2003–
AMSR2	18.7 (V,H)	23.8 (V,H)	36.5 (V,H)	89.0 (V,H)	–	–	12km * 7km	55.0°	2012–
GMI	18.7 (V,H)	23.8 (V)	36.64 (V,H)	89.0 (V,H)	V166.5 (V,H)	V	15km * 9km	52.8°	2014–

TABLE II.1 – Caractéristiques de quelques imageurs microondes. Les notations V, ou H précisent les polarisations disponibles quand plusieurs canaux observent à la même fréquence dans des polarisations différentes. (Les canaux de sondage ne sont pas précisés dans cette table).

donnée avec une précision acceptable. La qualité doit être mesurée et surveillée au cours de la durée de vie de chaque instrument Atkinson (2015). Sa mesure dépend du bruit intrinsèque de l'instrument (nommée température de bruit T_{sys}), de la bande passante (B) et du temps d'intégration (τ). Le niveau de signal détectable en température de brillance s'écrit (Ulaby et al., 2014) :

$$Ne\Delta T = \frac{T_{sys}}{\sqrt{B\tau}}.$$

Pour les récepteurs passifs, la résolution spatiale de l'observation dépend de la fréquence et du diamètre de l'antenne (D) à bord du satellite, une antenne plus grande permet d'atteindre des résolutions plus élevées. La même antenne étant généralement utilisée pour tous les canaux d'un même instrument, il en résulte que les observations aux fréquences les plus hautes (associées à de faibles longueurs d'ondes λ) ont des résolutions plus importantes qu'à basses fréquences : la résolution de l'observation est proportionnelle à $\frac{\lambda}{D}$.

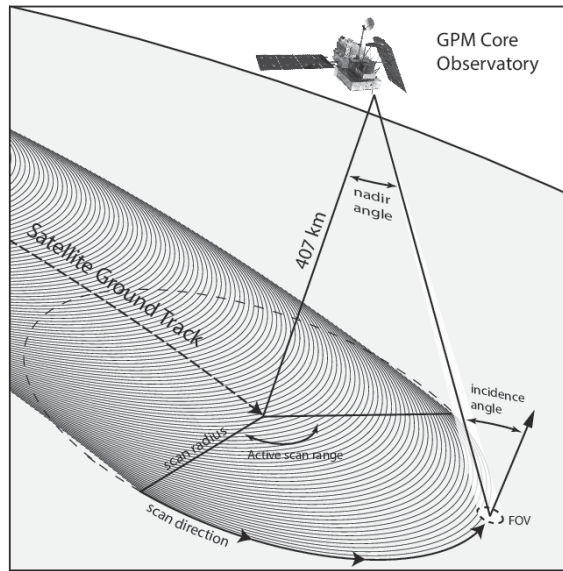


FIGURE II.1 – Représentation schématique du fonctionnement de l'instrument GMI à bord du *GPM Core observatory* (figure extraite du site de présentation de la mission GPM GMI - NASA).

Comme illustré sur la figure II.1 la plupart des radiomètres sont des imageurs coniques, l'instrument est disposé sur la plateforme et tourne autour d'un axe ce qui permet de balayer une section conique à la surface de la Terre avec un angle constant, l'angle d'incidence (*Earth incidence angle* - EIA). Celui-ci est le plus souvent proche de 53° . Cet angle est proche de l'angle de Brewster qui permet d'obtenir une différence importante entre le rayonnement réfléchi dans la polarisation horizontale et celui dans la polarisation verticale. Au cours de la rotation de l'instrument, des sources de rayonnement à des températures prédéfinies sont observées pour permettre l'étalonnage de l'instrument. L'angle de l'orbite de la plateforme ou inclinaison de l'orbite, définit l'angle que forme l'orbite du satellite par rapport à une orbite équatoriale.

1.2 Description des imageurs microondes utilisés

Comme présenté ci-dessus les paramètres des imageurs microondes peuvent différer sur de nombreux points. La plupart des imageurs présentés ont des orbites héliosynchrones,

c'est-à-dire que nonobstant une dérive non planifiée de la plateforme, les passages au-dessus de l'équateur se font à heure fixe, deux fois par jour. Ce n'est pas le cas de GMI dont l'orbite est dérivante. Ce paramètre fait partie du cahier des charges des instruments, tout comme les résolutions des instruments et les canaux utilisés. Ce cahier des charges répond aux besoins spécifiques d'application ciblées telles que la mesure des précipitations ou la prévision numérique du temps. L'utilisation des données pour la climatologie, peut présenter des contraintes supplémentaires, liées à la stabilité des instruments sur le long terme.

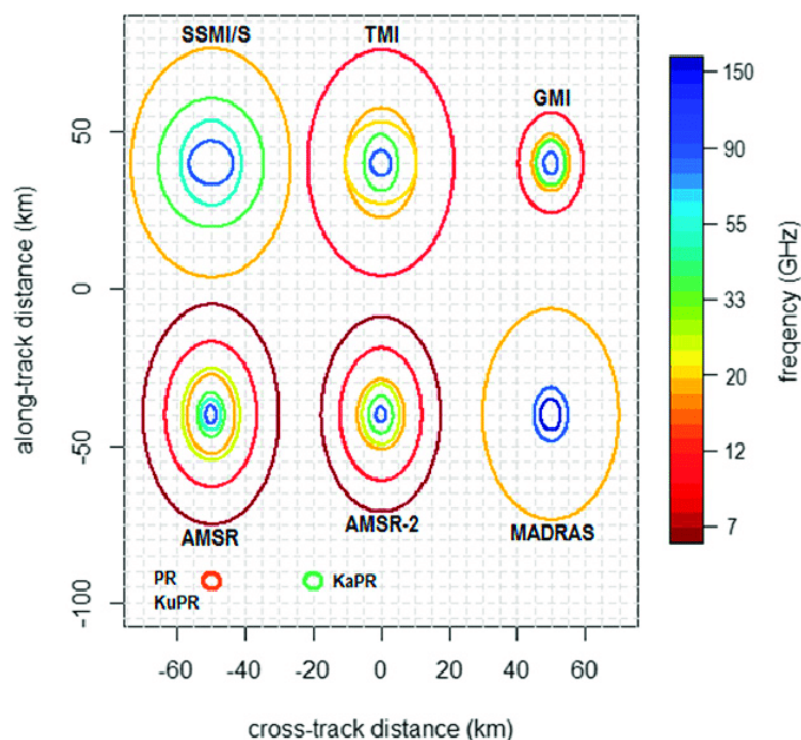


FIGURE II.2 – Comparaison des empreintes au sol de différents imageurs microondes (km). Les fréquences sont représentées par des ellipses de la même couleur (figure extraite de Guilloteau et al., 2017).

La figure II.2 présente les tailles des empreintes au sol pour un certain nombre d'imageurs microondes. Avec une augmentation de la taille des antennes, la résolution spatiale des instruments est améliorée.

Scanning Multichannel Microwave Radiometer

Le Scanning Multichannel Microwave Radiometer (SMMR) est le premier radiomètre microonde multi-canaux. En 1978, deux versions de cet instrument sont envoyées sur les satellites Seasat et Nimbus 7. Seule la mission à bord de Nimbus 7 a été utilisable plus de 6 mois. Sur les 10 années suivantes, SMMR a été le seul radiomètre multi-fréquence, avec des canaux à 6.63, 10.69, 18.0, 21.0 et 37.0 GHz en polarisations verticale et horizontale. Il a été conçu principalement pour des applications océaniques afin d'estimer la vitesse du vent de surface et la température de la surface de la mer (Gloersen and Barath, 1977). Il a aussi permis de nombreuses études exploratoires des propriétés de la surface terrestre observées par les microondes (van de Griend and Owe, 1994; Windsor and Charlton, 1991). En raison de la limitation de la puissance à bord, les mesures n'ont été effectuées qu'un jour sur deux. L'angle d'incidence (EIA) diminue de 50.2° à 49.3° à partir de 1986, diminution

due à une dérive non corrigée de la plateforme. Compte tenu de l'orbite polaire et de la fauchée d'environ 780 km, la couverture complète de la Terre est effectuée en environ 6 jours.

La résolution de l'instrument à 37 GHz est d'environ 25 km et la précision de l'instrument est limitée à 2 K (Gloersen and Barath, 1977). L'étalonnage utilise la température du fond cosmologique comme référence froide (2.7 K), une charge d'étalonnage chaude embarquée autour de 300 K ainsi que des études climatologiques pour estimer les biais dans les étalonnages. Le canal à 21 GHz est défectueux à partir de 1985 et une dérive sensible dans les mesures effectuées est apparue en 1986.

Special Sensor Microwave / Imager (SSM/I) et Special Sensor Microwave Imager Sounder (SSMIS)

Sous ces deux noms sont regroupés une série de capteurs aux caractéristiques similaires pour les canaux dans les fenêtres de transmission atmosphérique. Le premier, SSM/I, a été lancé en 1987 à bord du satellite F08 du *Defense Meteorological Satellites Program* (DMSP) (Hollinger et al., 1990). Le premier instrument SSMIS a été lancé en 2003 et actuellement les instruments à bord des plateformes F16, F17 et F18 fournissent des observations. L'ensemble du programme a été développé initialement pour les besoins de l'armée des Etats-Unis d'Amérique. Les canaux à 19.3, 23.3, 37 et 85.5 GHz, communs à SSM/I et SSMIS, mesurent les températures de brillance en polarisations horizontale et verticale (uniquement verticale pour le canal à 23 GHz), et la résolution est d'environ 30 km pour le canal à 37 GHz. En plus des canaux fenêtres, SSMIS possède des canaux dans les raies d'absorption de la vapeur d'eau et de l'oxygène autour de 60 et 183 GHz pour le sondage atmosphérique. L'inclinaison de l'orbite de 98.8° et la fauchée de 1400 km assure une revisite de la surface terrestre tous les 3 jours, jusqu'à $\pm 87^\circ$ de latitude (la table II.1, reprend ces informations).

Les données évaluées par Hollinger et al. (1990) sont d'une qualité suffisante pour un grand nombre d'applications dans des domaines variés (McFarland et al., 1990; Prigent et al., 1999; Andersson et al., 2010). Ce sont des instruments clés pour créer des FCDR à partir d'imageurs microondes, compte tenu de la longue durée d'observation disponible. Toutefois, les orbiteurs polaires DMSP ne sont pas équipés de moteur permettant de compenser la dérive de l'orbite due aux frottements. Par conséquent, les instruments sont sujets à des dérives aux heures de passage à l'équateur comme affiché sur la figure II.3, ce qui rend l'inter-étalonnage des instruments plus difficile.

La figure II.3 illustre l'évolution des heures de passage de chaque plateforme au-dessus de l'équateur. On voit que pour certains instruments l'écart de l'heure locale de passage au-dessus de l'atmosphère entre le début et la fin de vie de l'instrument peut être très élevé jusqu'à ~ 6 h.

Global Precipitation Measurement (GPM) Microwave Imager (GMI)

Comme illustré sur la Figure II.1, l'instrument GMI (Hou et al., 2014) est à bord de la plateforme GPM lancée par la *Japan Aerospace Exploration Agency* (JAXA) en 2015, principalement pour l'étude des précipitations. Les 13 canaux mesurent les températures de brillance entre 10 et 190 GHz avec une résolution à 37 GHz d'environ 15 km (voir la table II.1). Le satellite a une inclinaison de 65° , et permet des observations non

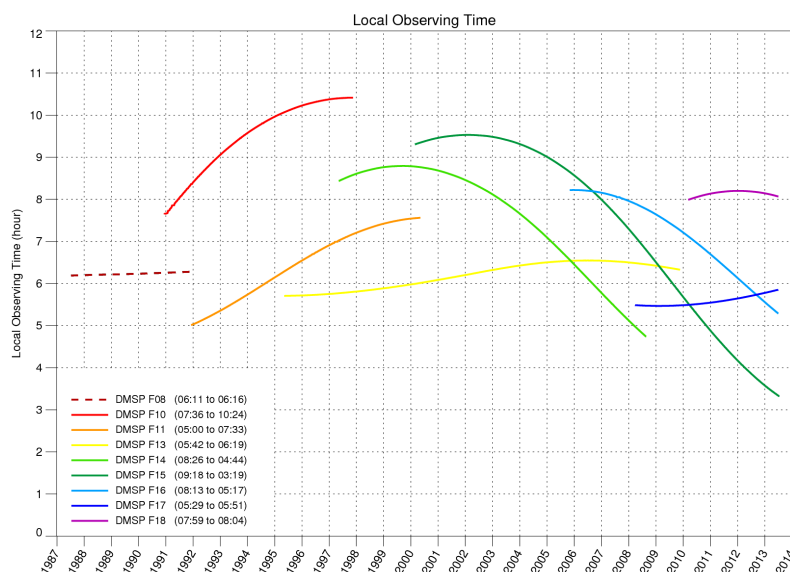


FIGURE II.3 – Comparaison des heures locales de passage à l'équateur des plateformes portant les instruments de la série SSM/I et SSMIS (figure extraite de Sapiano et al., 2013).

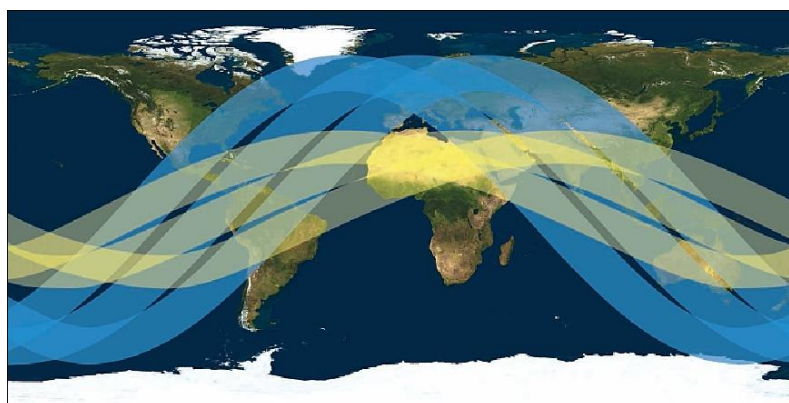


FIGURE II.4 – Illustration de la fauchée à la surface de la terre de GMI avec une orbite inclinée à 65° (en bleu) comparée à celle de l'instrument TMI (en jaune) avec une orbite à 35° . (NASA)

héliosynchrones de la Terre, des tropiques aux hautes latitudes. L'angle d'incidence des observations est de 52.8° et la fauchée de ~ 900 km. Cette largeur de fauchée associée à l'inclinaison signifie que les régions polaires ne sont pas entièrement couvertes.

Contrairement aux instruments précédents, l'inclinaison de l'orbite est plus faible, ce qui limite la couverture polaire de cet instrument. La figure II.4 illustre en bleu la couverture du globe de cet instrument.

L'étalonnage du GMI est décrit par Wentz and Draper (2016) : en plus de la charge chaude habituelle, GMI utilise des diodes à bruit qui améliorent la précision de l'étalonnage. Le satellite peut également effectuer des manœuvres de vol pour corriger la dérive. Lean (2017) a constaté de faibles biais pour tous les canaux par rapport aux simulations de l'ECMWF (inférieurs à 0.8 K). Toutes ces spécificités techniques font de GMI une excellente référence à des fins d'inter-étalonnage.

Les fréquences les plus élevées (≥ 150 GHz) et les plus basses (≤ 100 GHz) ne pointent pas sur la même aire au sol au même moment et ont une résolution différente. Le produit distribué nommé 1C-R effectue la collocalisation des observations indépendamment de leur résolution spatiale. La procédure de collocalisation associe la mesure haute fréquence la plus proche pour chaque cellule à partir des fréquences les plus basses. Elle est décrite dans le document de la NASA (2015).

Advanced Microwave Scanning Radiometer (AMSR)

Cette série de 3 instruments embarqués sur différentes plateformes avec des fréquences entre 6 et 90 GHz. L'angle d'incidence est d'environ 55° avec des heures de passage à l'équateur autour de 01:30 et 13:30. Parmi ceux-ci seuls deux ont fourni des observations sur des périodes de temps conséquentes (≥ 1 an) : AMSR-E entre 2002 et 2011 et AMSR-2 depuis 2012 (Kawanishi et al., 2003). L'utilisation ciblée est l'estimation de paramètres de surface des océans tels que la vitesse du vent, la concentration ou le déplacement des glaces. Ces instruments peuvent corriger leur orbite et ont donc des heures de passage à l'équateur extrêmement stables. L'altitude de 700 km avec une fauchée de 1450 km pour AMSR-E qui permettent une couverture globale bi-quotidienne.

Autres instruments existants

D'autres radiomètres ont été où sont encore opérationnels, mais ne sont utilisés que de manière circonstancielle dans ce travail. Certains sont des instruments de recherche et développés pour valider des concepts instrumentaux, d'autres sont utilisés de manière opérationnelle. On pourra ainsi lister les instruments suivants :

- L'imageur de la *Tropical Rainfall Measurement Mission* (TRMM) Kummerow et al. (1998) (TMI), prédécesseur de la mission GMI avec des fréquences entre 10 et 90 GHz, et une orbite très peu inclinée, comme illustré sur la figure II.4.
- L'instrument *WindSat* (Gaiser et al., 2004), couvrant une plage de fréquences de 6 à 37 GHz et mesurant les 4 paramètres de Stokes pour les canaux à 10, 18 et 37 GHz, pour l'observation de la surface des océans.

1.3 Niveau de produit distribué et étalonnage

Pour chacun des instruments, une série de traitements est appliquée aux données brutes mesurées par les capteurs de l'instrument jusqu'à obtenir des variables géophysiques, comme la température de surface, à une résolution temporelle et spatiale fixe. Les produits intermédiaires de ces étapes de traitement sont décrits par une succession de niveau de produit distribués. Dans la suite du texte pour chaque jeu de données utilisé, le niveau du produit correspondant est précisé. La table II.2 résume ces niveaux de données.

Le processus peut se résumer ainsi :

1. La donnée brute est téléchargée depuis le satellite (niveau 0) en incluant tous les paramètres de vol de la plateforme et ceux associés à la mesure (par exemple la température des références d'étalonnage).

Niveau	Variable	Résolution	Description
0	Donnée satellite brute	native non géoréférencée	Donnée brute téléchargée de la plateforme
1	T_b (K)	native par orbite	Température de brillance étalonnée et géoréférencée
2	Variable géophysique (température, précipitation, ...)	native par orbite	Unité de la variable géophysique (K pour la température de surface), par orbite
3 et +	Variable géophysique ou T_b	résolution temporelle et spatiale fixe	Produits agrégés pour obtenir des résolutions spatiales et temporelles prédéfinies

TABLE II.2 – Résumé des caractéristiques des niveaux de produits issus d’instruments satellites.

2. Un premier traitement est effectué en utilisant les caractéristiques de la plateforme comme l’altitude, les heures de passage, l’angle d’observation qui permettent de géo-référencer l’observation et convertir les mesures en températures de brillance. Ce niveau de produit peut être décliné en plusieurs versions (1B, 1C..) suivant les traitements appliqués : étalonnage, co-localisation et intégration d’observations. Au niveau 1, les données sont fournies par orbite et à leur résolution originale.
3. Le niveau 2 correspond à des variables géophysiques dérivées, comme la température de surface. Chaque méthodologie d’inversion se base sur un produit de niveau 1 et estime une valeur physique en gardant la résolution du produit initial.
4. Finalement, les niveaux de produits suivants (3 et +) combinent des produits satellites de niveau 1 ou 2 pour fournir des données à des résolutions spatiales et temporelles choisies, plus adaptée à l’utilisateur final.

Concernant l’étalonnage des instruments, une partie est effectuée au sol pour estimer le bruit de l’instrument, mais chaque instrument est étalonné en utilisant des points de référence tels que la température du fond diffus cosmologique (2.7 K) ou des sources d’étalonnage à bord de l’instrument à des températures prédéfinies (Wentz, 2013). Ces différentes références, et les différences techniques entre les plateformes, comme l’isolation des effets du soleil, peuvent expliquer une partie des différences observées d’un instrument à l’autre et la nécessité de créer des FCDR.

Parmi les étapes du traitement, un certain nombre concernent l’élimination de mesures qui peuvent être des sources potentielles d’erreur dans les analyses. Les sources d’erreurs pouvant poser des problèmes dans les inversions sont multiples et différentes solutions peuvent être mises en œuvre pour limiter leur impact. La section suivante détaille certaines de techniques en place dans les observations satellites pour garantir la qualité des produits dérivés.

2 Création de longues séries temporelles de température de brillance : le nécessaire inter-étalonnage des instruments

Pour créer une longue série temporelle de données de températures de surface à partir d’observations microondes, il faut utiliser des mesures provenant d’instruments différents.

En effet, la durée de vie d'un satellite étant d'une dizaine d'années au mieux, une couverture de données sur des décennies requiert l'utilisation de plusieurs instruments. Obtenir des mesures cohérentes sur une longue période de temps permet de suivre l'évolution des variables géophysiques sur de longues durées ce qui est nécessaire pour les études de tendances climatiques. Pour utiliser différents instruments de manière homogène et produire des séries temporelles les plus cohérentes possibles, un travail préliminaire sur les données est nécessaire.

L'étalonnage entre instruments ou inter-étalonnage regroupe un ensemble de méthodes qui visent à rendre comparables les mesures obtenues par des instruments différents. L'inter-étalonnage permet de corriger un biais entre les mesures réalisées par deux instruments d'une scène à la même température. Ce biais qui peut varier suivant la température ou le type de surface peut ensuite être pris en compte dans la production de produits dérivés et être attribué à des différences fondamentales dans la source de mesure, comme l'étalonnage de l'appareil.

Plusieurs équipes de recherche travaillent sur l'inter-étalonnage des températures de brillance des radiomètres microondes pour fournir des jeux de données homogènes au cours du temps (Berg et al., 2013; Wentz, 2013; Fennig et al., 2019). Ces inter-étalonnages sont principalement effectués au-dessus des zones océaniques qui sont plus homogènes dans l'espace et dans le temps. Ces données inter-étalonnées sont la base des FCDR. Les FCDR sont des données inter-étalonnées provenant de multiples capteurs qui ont été rendues homogènes et dont la qualité a été contrôlée pour que les données soient précises et stables sur de longues périodes. Ce sont les enregistrements de données bruts à partir desquelles les produits géophysiques peuvent être calculés et agrégés pour former des séries temporelles sur le long terme des variables climatiques essentielles (*Essential Climate Variables - ECV*) (Bojinski et al., 2014). Il est aussi possible d'inter-étalonner les mesures au niveau des variables géophysiques dérivées (Cavalieri et al., 2012), mais ce n'est pas le sujet du travail présenté.

Plusieurs méthodes existent pour inter-étalonner les instruments, un bref résumé de celles applicables aux microondes est proposé ci-dessous (pour plus de détails voir Chander et al. (2013)). Pour la plupart des instruments microondes, les corrections proposées par les inter-étalonnages sont inférieures à 1 K.

2.1 Méthodes d'inter-étalonnage d'imageurs microondes

Les méthodes d'étalonnage s'appliquent donc à trouver des points de comparaisons entre différents instruments, pour ensuite corriger de manière systématique une différence ou en tracer l'origine à un problème d'étalonnage initial. Avant de procéder à un inter-étalonnage, un choix de la référence doit être effectué. Cet instrument est souvent choisi pour la qualité de son étalonnage et la stabilité des mesures fournies tout au long de sa vie.

Les **comparaisons directes** (*Simultaneous Nadir Overpass*) sont la méthode la plus évidente d'inter-étalonnage. Elle est envisageable quand la même scène est observée par deux instruments sous des conditions d'observation proches : mêmes heures d'observation, et mêmes caractéristiques instrumentales. Il est alors directement possible de comparer les mesures effectuées par chacun des deux instruments de la scène. Cette opération répétée pour des scènes ayant des propriétés spectrales différentes permet d'estimer l'écart entre l'instrument de référence et celui à étalonner. Cette méthode est utilisée par

exemple avec des observations polaires (Cavalieri et al., 1999; Sapiano et al., 2013) dont de multiples observations sont effectuées par les satellites en orbite héliosynchrone. Les températures de brillance mesurées au-dessus des pôles, présentent l'inconvénient d'être uniquement des températures basses par rapport aux autres températures terrestres observées.

L'étalonnage "délégué" (*vicarious calibration*) (Brown and Ruf, 2005) utilise des données au-dessus de scènes dont les propriétés sont connues. Par exemple des zones de l'océan pour des mesures de températures de brillances froides, et au-dessus de la forêt Amazonienne pour les points les plus chauds (Ruf, 2000). Pour ces scènes, il est alors possible de comparer la mesure effectuée par l'instrument de référence et par l'instrument à inter-étalonner en pouvant supposer la stabilité temporelle de la scène ciblée. Des évolutions de ces méthodes ont été proposées en prenant en compte la variabilité saisonnière des surfaces observées, permettant ainsi d'inclure un plus grand nombre de sites de référence (Yang et al., 2016).

La méthode dite de **double différence** (Kroodsma et al., 2012) est une solution pour comparer des instruments ayant des caractéristiques différentes. Elle est nécessaire quand deux instruments présentent des différences dans la fréquence des canaux utilisés, l'angle d'incidence ou l'heure de passage à l'équateur. Le principe consiste à introduire un troisième point de mesure, appelé pivot, par rapport auquel on peut comparer l'instrument de référence et l'instrument à étalonner. En infra-rouge, on utilise parfois des instruments géostationnaires comme pivots, mais pour les microondes le plus souvent un modèle et des données auxiliaires servent comme pivot (Meirink et al., 2013). La différence de chacun des instruments utilisés avec le pivot est calculée et c'est celle-ci qui est étalonnée. Cette méthode présente l'avantage d'être utilisable même en l'absence de colocation temporelle proche. On peut décrire cette procédure d'utilisation d'un modèle comme pivot, pour un instrument A , avec un instrument de référence REF , ayant des caractéristiques différentes. On étalonne les températures à une fréquence donnée selon la formule :

$$T_A - T_{modele,A} = T_{REF} - T_{modele,REF}.$$

Pour garantir la stabilité de l'inter-étalonnage, une combinaison entre les paramètres fournis par une réanalyse d'une part et un modèle de transfert radiatif d'autre part peut être utilisée comme pivot. Le modèle de transfert radiatif est alimenté par des informations atmosphériques et de surface et tient compte des différences d'angle d'incidence, des caractéristiques des canaux et même des différences dans les heures de passage au-dessus de l'équateur. La double différence est le plus souvent appliquée au-dessus des océans étant donnée la plus grande difficulté à effectuer un transfert radiatif fiable au-dessus des continents. En effet, les modèles de transfert radiatif microonde sur les continents, ainsi que tous leurs paramètres d'entrée nécessaires (par exemple, l'humidité du sol, la densité de la végétation, l'équivalent en eau de la neige), ne sont pas disponibles avec la qualité requise pour de nombreuses surfaces. De plus, la possible forte variabilité diurne des températures de la surface terrestre n'est pas décrite avec suffisamment de précision et de résolution temporelle pour tenir compte des différences des heures locales de passage des satellites.

Pour ce second problème, différentes méthodologies ont été employées : Dai and Che (2009) ont testé une modélisation de la variation diurne de la température de surface pour inter-étalonner des instruments ayant des heures de passage différentes sur les continents. Pour ceux-ci l'approche principale consiste à utiliser des correspondances avec un instrument de référence ayant une orbite différente, ce qui permet de fournir des comparaisons quasi-directes sur une large plage de latitudes, même pour des satellites ayant des heures

différente de passage à l'équateur. Grâce à leurs orbites non héliosynchrones, TMI et GMI peuvent être utilisés comme références intermédiaires pour inter-étalonner les instruments héliosynchrones dont les heures de survol sont radicalement différentes, comme la série SSM/I (heure de survol de l'équateur au crépuscule et à l'aube) et la série AMSR (heures de survol de l'équateur à 01:30 et à 13:30). Berg et al. (2018) ont appliqué ces méthodes pour inter-étalonner tous ces instruments.

Pour finir, des comparaisons statistiques des distributions des températures obtenues par les instruments permettent de limiter les erreurs aléatoires causées par le choix des points de comparaison. La plupart des équipes travaillant sur ces inter-étalonnages utilisent des combinaisons des différentes méthodes pour améliorer la qualité des résultats. Par exemple, Grody et al. (2004) proposent de compléter et améliorer l'inter-étalonnage empirique des *Microwave Sounding Unit* par une modélisation des références de l'étalonnage entre les instruments. Cette correction permet d'expliquer une partie des différences d'étalonnage entre les instruments et de les corriger. Les distributeurs de données fournissent des produits inter-étalonnés mais mettent aussi à disposition les coefficients à appliquer pour corriger chaque instrument par rapport à la référence choisie.

2.2 Fundamental Climate Data Records disponibles pour les imageurs microondes

Instrument	Reference	SMR	SSM/I	SSMIS	TMI	AMSR-E	AMSR2	GMI
CDR CSU V01	Berg et al. (2013)	-	✓	✓	*	-	-	-
CSU V03	Berg et al. (2018)	-	✓	✓	✓	✓	✓	*
CDR RSS	Wentz (2013)	-	✓	✓	-	-	-	-
CM SAF	Fennig et al. (2019)	✓	*	✓	-	-	-	-
CM SAF	(en développement)	✓	*	✓	-	✓	✓	✓

TABLE II.3 – Caractéristiques des Fundamental Climate Data Records existant pour les imageurs microondes passifs. Les ✓ indiquent un instrument inter-étalonné et les * indiquent l'instrument de référence de chaque FCDR.

Le tableau II.3 liste les principaux FCDR pour les instruments microondes passifs disponibles pour la communauté. Pour chacun sont détaillés les instruments inclus dans l'inter-étalonnage (coches) ainsi que l'instrument utilisé comme référence (étoiles).

Les FCDR fournis par EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF), Colorado State University (CSU) et Remote Sensing Systems (RSS) utilisent tous SSM/I et SSMIS dans leur inter-étalonnage. Quelques différences importantes existent cependant entre ces différents jeux de données :

- GMI n'est utilisé comme instrument de référence que dans la version la plus récente de l'inter-étalonnage par Berg et al. (2018), les versions précédentes étant basées sur l'utilisation de TMI. Plusieurs méthodes d'inter-étalonnage sont employées simultanément pour réduire l'incertitude du jeu de données. Les différentes méthodes se sont bien accordées entre elles, offrant une confiance accrue dans l'inter-étalonnage proposé. La version initiale (V01) est améliorée par l'ajout de nouveaux satellites dans le FCDR et l'utilisation de GMI comme référence. Cet ajout permet une comparaison plus juste sur l'ensemble du cycle diurne des observations au-dessus des

continents.

- L’inter-étalonnage proposé par *Remote Sensing System* (RSS) repose quasi exclusivement sur l’utilisation de modèles de transfert radiatifs au-dessus des océans (Wentz, 2013), et n’inclut pas les mesures de SMMR.
- SSM/I F11 est utilisé comme référence par le CM SAF. A ce jour, seul le CM SAF inclut les données de SMMR dans leur FCDR (Fennig et al., 2019). Un étalonnage linéaire a été effectuée pour réduire le biais de l’instrument dans le domaine ciblé, c’est-à-dire pour les estimations de paramètres de la surface océanique (Andersson et al., 2010). L’inter-étalonnage effectué par le CM SAF est basé sur un chaînage d’inter-étalonnages des instruments SSM/I F11 et F08 puis la méthode de double différence est appliquée entre SSM/I F08 et SMMR, en utilisant des simulations de transfert radiatif issues de réanalyses pour tenir compte des changements de fréquences, de bandes passantes et d’EIA. Cette correction est calculée uniquement sur les océans sans nuages.

L’évaluation des corrections proposées par les différentes solutions d’inter-étalonnages pour les instruments de la série SSM/I fournissent des résultats similaires. Cependant, l’inter-étalonnage proposé par le CM SAF est le seul à inclure SMMR. Mais il a été développé principalement pour les surfaces océaniques. La correction appliquée n’a donc que peu d’effet sur les températures de brillance élevées (principalement au-dessus des continents). Il est donc nécessaire d’effectuer cet inter-étalonnage au-dessus des surfaces continentales comme présenté dans la section 3.

3 Inter-étalonnage d’instrument : le cas des températures de brillance de SMMR au-dessus des continents

La section suivante est composée des résultats présentés dans l’article : Favrichon, S., Jimenez, C., and Prigent, C. : Inter-calibrating SMMR brightness temperatures over continental surfaces, Atmos. Meas. Tech., 13, 5481–5490, <https://doi.org/10.5194/amt-13-5481-2020>, 2020.

Comme décrit ci-dessus, des observations microondes multifréquences globales sont disponibles depuis 1978 grâce à l’instrument SMMR et à tous ses successeurs. Malgré des caractéristiques communes aux différents instruments, les sauts technologiques entre les différentes générations d’imageurs microondes et les différences d’étalonnage entre des instruments identiques à bord de plateformes différentes sont la source de biais systématiques dans les données fournies. Ces biais ne sont pas constants ni nécessairement linéaires, mais affectent de manière systématique les mesures effectuées. Certaines causes peuvent être attribuées à des problèmes techniques (étalonnage, etc.) mais d’autres ne trouvent pas d’explication simple étant donné la difficulté à diagnostiquer les instruments en orbite. Pour créer des jeux de donnée homogènes couvrant plusieurs décades, un inter-étalonnage entre les instruments doit être effectué.

Cette section s’intéresse spécifiquement à l’inter-étalonnage de l’instrument SMMR décrit dans la section 1.2 par rapport à d’autres imageurs microondes au-dessus des surfaces continentales.

Les données inter-étalonnées fournies par CM SAF dans leur FCDR pour l’instrument SMMR sont décrites ci-dessus. Cependant, une évaluation de cet inter-étalonnage

montre que la correction pour les surfaces continentales (spécifiquement les températures de brillance élevées) n'est pas suffisante pour les canaux communs à tous les imageurs et surtout ceux les plus utiles pour l'estimation de la température de surface (à 18 et 36 GHz). Il est donc nécessaire d'effectuer un étalonnage spécifique pour ceux-ci.

La première étape est le choix d'un instrument de référence. L'instrument SMMR ayant une fin de vie en août 1987 et les premières observations par SSM/I (F08) datant de juillet 1987, la superposition des mesures n'est que de quelques semaines seulement. En raison du faible chevauchement des observations et d'une différence de ~ 6 h dans les heures de passage à l'équateur entre SMMR et SSM/I (F08), ce choix de référence n'apparaît pas être compatible avec un inter-étalonnage robuste. Pour ces raisons, GMI a été choisi comme référence malgré l'absence d'une période d'observation commune. Dans cette section, les cycles diurnes de trois ans de températures de brillance GMI sont d'abord calculés, puis utilisés pour évaluer les différences de T_b mesurées par rapport à SMMR. Sur la base d'une analyse statistique des différences, une correction linéaire est mise en œuvre pour étalonner SMMR par rapport à GMI. Il est démontré que cette correction réduit également les biais entre SMMR et SSM/I. Elle peut donc être appliquée aux observations SMMR pour les rendre plus cohérentes avec les données existantes des températures de brillance microondes sur les surfaces continentales.

Les paragraphes suivants décrivent les données utilisées, ainsi que le pré-traitement avant l'analyse. Les résultats, ainsi que les coefficients de l'inter-étalonnage proposé, sont finalement décrits.

3.1 Données utilisées et méthode appliquée

Les données satellites

Pour effectuer ce travail les données provenant de trois instruments sont collectées. La table II.1 précise les fréquences et caractéristiques exactes des différents instruments et la description des missions est disponible dans la section 1.

Les données originales de **SMMR** disponibles sont des données de niveau 1B, et non des comptages bruts. Elles comprennent des corrections liées au diagramme de rayonnement l'antenne et à la dérive du capteur (Njoku et al., 1980). Les coefficients des corrections apportées ne sont plus disponibles on utilise donc des données dont l'étalonnage a déjà fait l'objet de corrections. Dans cette section, seules les données SMMR présentant la meilleure qualité sont utilisées. Les données présentant une intrusion solaire, une intrusion dans le champ de vision ou une erreur de balayage sont supprimées.

Les données de l'instrument **GMI** sont utilisées comme référence. L'utilisation de cet instrument de référence repose sur les hypothèses suivantes :

- ✓ les conditions environnementales moyennes n'ont pas connu de modifications extrêmes entre la période couverte par SMMR et celle couverte par GMI.
- ✓ La différence entre l'instrument à étalonner et la référence suit une relation linéaire.

Il est alors possible de comparer des séries de mesures obtenues par les deux instruments et d'en déduire une correction à appliquer. Cette approche ne permet pas d'effectuer un

inter-étalonnage extrêmement précis, mais elle doit permettre de corriger les larges biais qui rendent hasardeuse l'utilisation des observations de SMMR pour créer des séries de données au-dessus des continents. Dans cette étude, les T_b GMI étalonnées du niveau 1C (Hou et al., 2014) sont utilisées. GMI, avec son orbite non héliosynchrone, observe le cycle diurne complet incluant donc les heures de passage à l'équateur de SMMR et SSM/I, ce qui permet d'évaluer les deux instruments.

Finalement les données fournies par l'instrument **SSM/I** à bord de F08 sont également utilisées dans cette section, car il s'agit du seul instrument dont la période d'observation chevauche celle de SMMR (bien que seulement sur 28 jours). Les données SSM/I fournies par le CM SAF, sans la couche d'inter-étalonnage, sont utilisées pour évaluer la cohérence de l'inter-étalonnage de SMMR et la validité des hypothèses.

Méthode

La méthode développée repose sur la comparaison des températures de brillance observées pour deux canaux (~ 18 et ~ 37 GHz) aux polarisations verticale et horizontale. L'heure de passage au-dessus de l'équateur de SMMR étant fixe sur la majeure partie de sa durée de vie, les mesures au-dessus des continents ne peuvent être comparées qu'à des observations faites à des heures proches. GMI permet d'obtenir des observations à la même heure que SMMR en utilisant le fait qu'un point donné de la terre est vu à des heures différentes à chaque passage du satellite. Cela permet de reconstruire un cycle diurne complet de températures de brillance et donc d'effectuer la comparaison à l'heure de passage de SMMR.

Compte tenu de leurs années d'opération, SMMR et GMI n'ont pas de période d'observation commune. Il faut donc passer par une analyse statistique des températures de brillance mesurées par chacun des instruments sur différentes années. Cette comparaison permet d'estimer la correction nécessaire pour ajuster les données issues de SMMR à la référence choisie au-dessus des surfaces continentales.

Les deux différences principales entre les instruments (décrits dans la section II.1) sont une petite variation de l'angle d'incidence (50.2° pour SMMR et 52.8° pour GMI) et des différences dans la largeur de bande et la fréquence centrale des canaux utilisés. Mais celles-ci ne peuvent pas expliquer totalement le biais observé.

La différence d'angle d'incidence est $\sim 3^\circ$. Au-dessus des océans, cette différence peut très fortement modifier le signal mesuré au niveau du satellite, car l'émissivité dépend fortement de l'angle d'incidence. De plus, la contribution atmosphérique est relativement plus importante dans la température de brillance mesurée (compte tenu de la faible émissivité de la surface) et une petite variation de l'angle peut donc avoir un effet important sur la température de brillance mesurée.

Pour les surfaces continentales, l'émissivité est forte, avec des valeurs supérieures à 0.9 pour la plupart des surfaces (Prigent et al., 2006) et les changements dus à l'angle d'incidence sont limités (Prigent et al., 2000). Avec une émissivité élevée, la contribution de l'atmosphère au signal mesuré est relativement faible et les changements d'angle d'incidence n'influeront pas sensiblement le résultat. De la même façon, les émissivités de surface étant élevées et variant peu suivant la fréquence (Prigent et al., 2000), la légère différence de fréquences entre SMMR et GMI ne devrait pas affecter les signaux de manière significative. La figure II.5 modélise avec un simple code de transfert radiatif l'impact

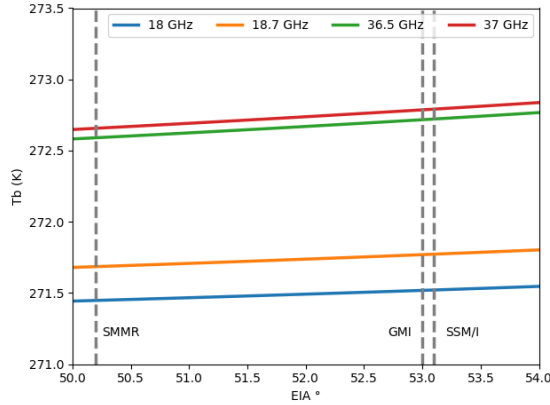


FIGURE II.5 – Changement de la température de brillance simulée (K) à l’aide du modèle de transfert radiatif de Rosenkranz en faisant varier les paramètres d’EIA et pour différentes fréquences avec une émissivité et une température de surface fixes à 300 K. Les lignes interrompues grises indiquent les angles d’incidence de SMMR, GMI et SSM/I.

des changements d’EIA et de fréquence sur les températures de brillance mesurées toutes choses égales par ailleurs. L’effet simulé apparaît être inférieur à 0.1 K pour les deux instruments, et donc négligeable par rapport aux erreurs de mesures et aux différences observées.

Pour effectuer la comparaison, chacun des jeux de données satellites est projeté sur une grille *Equal Area EASE* globale à une résolution de 25 km entre 60° S et 70° N, ainsi que sur la grille *EASE southern hemisphere azimuthal* pour les données au-dessus de l’Antarctique (Brodzik et al., 2012). Un masque des pixels contenant de l’eau est appliqué sur les données pour les exclure du traitement pour limiter les contaminations. Une sélection de données représentatives est faite, couvrant des situations saisonnières différentes. Les observations pour deux paires de mois, juillet/août et janvier/février sont choisies pour chacun des instruments.

Les données GMI sont collectées pour les années de 2015 à 2017. Les quatre mois choisis incluent des températures typiques des saisons d’été et d’hiver et, avec des données sélectionnées sur toute la surface de la Terre, la majorité des types de surfaces possibles est représentée. La résolution des observations GMI étant plus fine que celle de la grille de projection, toutes les observations tombant dans un rayon de ~ 12 km sont moyennées, pour obtenir une unique valeur pour chaque pixel. Pour la dimension temporelle, les valeurs sont moyennées sur une fenêtre de 15 minutes, pour chaque cellule. Si aucune observation n’est disponible dans un intervalle de 15 minutes donné, une interpolation linéaire avec les points existants les plus proches est effectuée. Enfin, pour limiter le bruit dans le cycle diurne reconstitué, en adéquation avec la réalité physique du phénomène avec des variations lisses de la température, une moyenne glissante sur 75 minutes est appliquée.

Pour SMMR, les T_b sont collectées pour les mêmes mois que GMI, pour les années 1981, 1982 et 1987. Les observations sont disponibles deux fois par jour, pour l’orbite montante et descendante, et tous les deux jours comme imposé par le mode de fonctionnement de l’instrument (plus de détails dans la section 1.2). La résolution des observations de SMMR est proche de celle de la grille de projection, donc la valeur de l’observation la plus proche de chaque pixel est utilisée. L’utilisation de plusieurs années d’observation limite le risque d’observer un biais causé par la variabilité inter-annuelle (dues à des phénomènes cycliques comme El Niño). Le jeu de données final utilisé contient environ 2 millions de

points représentant diverses surfaces continentales et deux saisons contrastées.

3.2 Résultats

Comparaisons entre SMMR et GMI à l'échelle régionale

Pour apprécier les différences existantes entre les deux sources d'observation, une sélection de zones présentant des conditions environnementales variées est effectuée. Elle couvre des surfaces arides froides et chaudes, des forêts tropicales et des zones couvertes d'une végétation clairsemée.

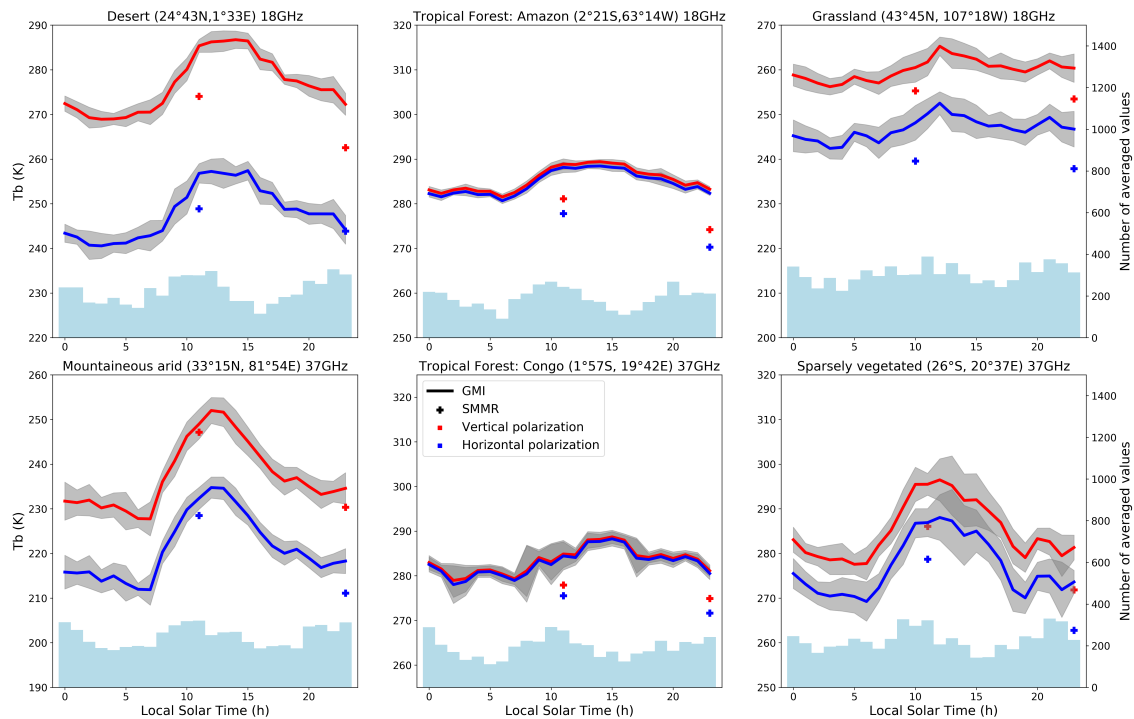


FIGURE II.6 – Cycles diurnes de température de brillance (K) de GMI (2015-2017) comparés aux valeurs mesurées par SMMR (1981,1982,1987), à différentes fréquences et pour différents lieux pour les mois de janvier/février (la polarisation horizontale en bleu et la polarisation verticale en rouge). Les mesures sont moyennées sur 25 pixels et par heure. La déviation standard de chacune des températures moyennes est affichée. L'histogramme affiche pour chaque heure le nombre de points de mesure associés.

Le cycle diurne moyen de T_b dérivé de GMI est calculé pour des zones échantillons couvrant 5×5 pixels, sur deux mois pour trois années de donnée (2015-2017), indépendamment pour chaque saison. Sur ces zones couvrant environ $1^\circ \times 1^\circ$ en surface, la moyenne des T_b SMMR pour les orbites ascendantes et descendantes est également calculée, sur les mêmes mois, pour 1981, 1982 et 1987.

La figure II.6 présente les résultats pour les régions sélectionnées, pendant l'hiver septentrional (janvier et février) à 18 et 37 GHz, pour les deux polarisations orthogonales. Les résultats sont présentés pour des régions montagneuses froides, pour des déserts arides avec des T_b élevées pendant la journée et une grande amplitude du cycle diurne, ainsi que pour des forêts tropicales chaudes avec une variation limitée de T_b . Les régions forestières présentent une faible amplitude du cycle diurne des T_b (≤ 10 K), avec presque aucune

différence entre les polarisations V et H en raison de l'effet d'émission et de diffusion de la végétation dense. Les régions arides (froides ou chaudes) présentent de grandes variations diurnes de T_b , directement liées au cycle diurne de la température de la surface. La différence de polarisation est significative en raison d'un comportement presque spéculaire de ces surfaces plutôt planes. Avec l'augmentation de la couverture végétale (végétation clairsemée ou prairie), associée à la présence d'humidité dans le sol, la différence entre les polarisations augmente par rapport à celles avec un couvert forestier total et l'amplitude des cycles diurnes est réduite par rapport aux zones arides.

La variabilité associée à chaque température moyenne pour GMI est calculée et également affichée (la bande grise autour du cycle diurne indique un écart-type). Le nombre de pixels individuels utilisés dans le calcul du cycle diurne est également indiqué (en bleu). La faible variabilité des signaux sur l'Amazonie confirme la grande stabilité des T_b dans cette région qui est régulièrement utilisée comme référence radiométrique chaude (Brown and Ruf, 2005).

Les valeurs de SMMR ont été calculées pour les données initiales du niveau 1B ainsi que pour les données inter-étalonnés par le CM SAF mais seules les données du niveau 1B sont affichées sur la figure II.6. Les résultats obtenus à partir des données inter-étalonnées fournies par le CM SAF montrent des valeurs très similaires à celles obtenues avec les données de niveau 1B originales (différence ≤ 1 K).

Les températures diurnes et nocturnes de SMMR suivent la même évolution que pour GMI, mais avec une sous-estimation systématique de la T_b , quelle que soit la fréquence et la polarisation. Ces différences sont généralement beaucoup plus importantes que ce que l'on attend de la variabilité du signal GMI, indiquée par l'écart-type autour de la moyenne. Cette différence entre SMMR et GMI est très proche pour les deux polarisations. La différence semble diminuer pour les endroits les plus froids. La variabilité est plus élevée sous les latitudes moyennes où des variations d'une année sur l'autre sont possibles et où des événements météorologiques peuvent avoir un impact sur les mesures.

Ces différences ne peuvent cependant pas être expliquées uniquement par les caractéristiques différentes des instruments (EIA ou fréquences), ou par les changements environnementaux entre les périodes couvertes par les deux instruments (les années 1980 pour le SMMR et les années 2010 pour le GMI).

Différentes causes peuvent être à l'origine de tels écarts, comme une mesure erronée de la température de la charge d'étalonnage chaude ou encore une erreur dans la correction du diagramme d'antenne qui causerait une contamination des observations par un signal froid provenant du fond cosmique. Toutefois, corriger ces potentielles sources d'erreur aurait un impact sur tous les étalonnages SMMR existants. D'autres sources d'erreur sont encore possibles qu'on ne peut pas isoler compte tenu du niveau de données accessibles pour SMMR.

La couche d'inter-étalonnage du CM SAF pour SMMR ne semble pas améliorer les résultats. Sur Terre, la correction moyenne ajoutée par l'étalonnage est inférieure à 0.5 K pour les polarisations verticale et horizontale à 18 GHz et pour le canal horizontal à 37 GHz. La correction pour la polarisation verticale à 37 GHz est en moyenne d'environ 2 K. L'inter-étalonnage a été conçu principalement pour les observations océaniques avec de faibles T_b et non pour les applications terrestres et les corrections sont faibles. Les données en ajoutant la couche d'inter-étalonnage proposée par le CM SAF ne sont pas représentées sur la figure. D'autres lieux et saisons ont été analysés, avec des conclusions similaires (non présentées ici).

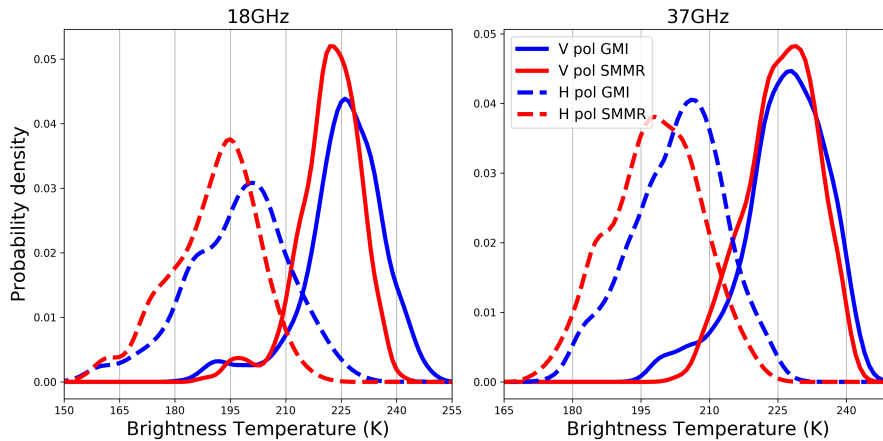


FIGURE II.7 – Densité de probabilité des températures de brillance (K) observées en juillet/août au-dessus de l’Antarctique par GMI en 2015 et SMMR en 1987.

Les observations au-dessus de la calotte glaciaire Antarctique sont aussi analysées pour étendre la plage de température étudiée à des températures plus basses. L’angle de l’orbite de GMI limite les observations à 80° et les pixels en coïncidence sont donc situés à la limite du continent. Pour éviter toute contamination par des observations océaniques ou de glace de mer, les mois d’hiver de l’hémisphère sud (juillet/août) sont utilisés et un filtre des pixels terrestres est appliqué. L’absence de cycle diurne durant la longue nuit australe permet de comparer directement la distribution des T_b pour chacun des deux instruments (figure II.7), sans avoir à reconstruire le cycle diurne. On observe, pour ces températures de brillance, un écart entre les données de SMMR et GMI de l’ordre de 5 à 10 K pour tous les canaux, sauf le canal à 37 GHz en polarisation verticale pour lequel l’écart est plus faible.

Encore une fois, la différence significative observée entre les deux jeux de données ne peut pas être attribuée uniquement à des changements dans les propriétés spectrales des surfaces, ou à des variations environnementales, même avec plus de 30 ans d’écart. Jezek et al. (1991) ont comparé les T_b observées par SSMI F08 et SMMR au-dessus de la calotte glaciaire Antarctique, sur la période de recouvrement. Ces travaux mettent aussi en évidence des différences importantes pour les canaux à 18 et 37 GHz. L’attribution de ces différences à un changement d’angle d’incidence ou de fréquences entre les instruments doit donc être écartée. Seuls des défauts d’étalonnage peuvent raisonnablement expliquer que les températures de brillance de SMMR soient plus froides que celles observées par GMI.

Dérivation d’une correction pour SMMR au-dessus des surfaces continentales

Afin d’évaluer la possibilité d’établir une correction pour la différence d’étalonnage de SMMR par rapport à GMI, les observations des deux instruments sont comparées à une échelle globale. Les orbites ascendantes et descendantes sont comparées aux températures correspondantes de GMI, à la même heure. La comparaison est faite pour janvier/février et juillet/août sur une période de 3 ans pour chacun des instruments (1981-1982-1987 pour SMMR et de 2015 à 2017 pour GMI, comme précédemment).

Pour proposer un inter-étalonnage indépendant des précédents, la couche d’inter-étalonnage proposée par CM SAF n’est pas appliquée aux données SMMR de niveau

1B. Les points pour lesquels la différence entre GMI et SMMR est supérieure à 3 fois la déviation standard par rapport à la différence moyenne sont éliminés. Une inspection visuelle permet de comprendre que la majorité de ces points sont situés dans des régions côtières : étant donné que la résolution spatiale de GMI est bien supérieure à celle de SMMR, les mesures de SMMR le long des côtes ont une contribution plus forte de l’océan.

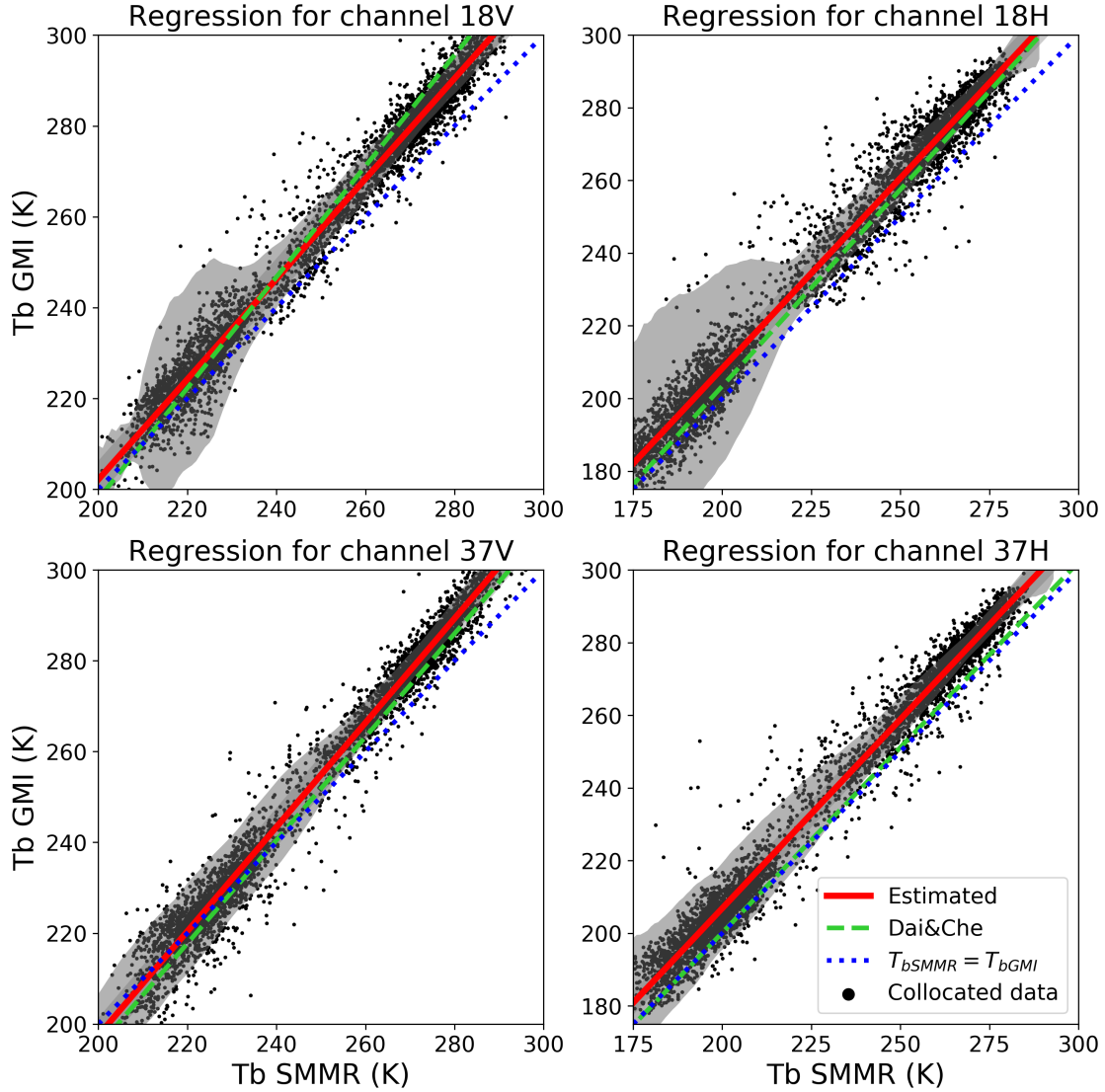


FIGURE II.8 – Nuages de points correspondants aux T_b de SMMR et GMI colocalisées. La régression linéaire entre les valeurs de SMMR et GMI après un filtrage des points est affichée en rouge. La correction estimée par Dai and Che (2009) (et corrigée de l’écart d’étalonnage entre SSM/I F08 et GMI) apparaît en vert. La ligne en pointillés bleus correspond à la droite pour laquelle les valeurs de T_b des deux instruments sont égales : $T_{SMMR} = T_{GMI}$.

La figure II.8 affiche la distribution des T_b de GMI par rapport à celles de SMMR, en polarisations verticale et horizontale pour les canaux à 18 et 37 GHz. La sous-estimation des T_b par SMMR par rapport à celles de GMI augmente clairement pour des valeurs de température plus élevées. Dans cette étude une correction linéaire est proposée pour inter-étalonner les observations de SMMR vers celles de GMI. Une correction plus sophistiquée ne semble pas justifiée. Premièrement, il est évident qu’il s’agit d’un inter-étalonnage de premier ordre, puisque la comparaison statistique porte sur différentes années. Deuxième-

ment, un modèle plus complexe serait notablement plus compliqué à mettre en œuvre dans les chaînes de traitement. De plus, augmenter la complexité de l'inter-étalonnage mis en place pourrait entraîner des problèmes de sur-apprentissage ce qui affecterait une partie du signal observé.

La correction pour obtenir les données inter-étalonnées T^{IC} pour chacun des canaux est proposée telle que

$$T_{SMMR}^{IC} = a \times T_{SMMR} + b.$$

La pente a et l'ordonnée à l'origine b sont estimées par une minimisation de la somme des moindres carrés $\sum_{i=1}^N (T_{GMI} - a \times T_{SMMR} - b)^2$ en supposant que $T_{SMMR}^{IC} = T_{GMI}$, avec N le nombre de points utilisés pour estimer les coefficients de régression. Les données collectées couvrent l'ensemble des T_b rencontrées au-dessus des continents, en incluant des régions polaires.

La distribution des points entre les températures froides (≤ 240 K) et chaudes est inégale, avec moins de points pour les T_b les plus froides. Pour limiter ce biais, un échantillonnage aléatoire des points des extrémités froides et chaudes du jeu de données est effectué, permettant de simuler une distribution également répartie sur l'intervalle de température. L'échantillonnage est répété et testé pour confirmer la stabilité des coefficients de régression estimés. Les droites obtenues avec ces coefficients sont ajoutées sur la Figure II.8. L'erreur quadratique moyenne de la régression linéaire est indiquée par la zone ombrée grise. Il apparaît clairement que l'incertitude augmente pour la gamme de T_b avec le moins de données disponibles (entre les mesures au-dessus de l'Antarctique et le reste du globe ou à la limite de l'intervalle de températures couvert). Les coefficients de la régression linéaire sont fournis dans le tableau II.4. Les intervalles de confiance sur les coefficients sont calculés avec un test t de Student avec un niveau de 99%.

Canal	Pente	Ordonnée	R^2
18V	1.10±0.01	-18.7±2.2	0.976
18H	1.05±0.01	-1.29±1.9	0.971
37V	1.15±0.01	-32.2±2.2	0.976
37H	1.04±0.01	-1.23±1.9	0.976

TABLE II.4 – Coefficients de la correction linéaire estimés pour l'inter-étalonnage des observations de SMMR à 18 et 37 GHz aux polarisations verticale et horizontale, par rapport aux observations de GMI sur les continents. L'intervalle de confiance calculé avec un test t de Student à 99% est indiqué.

3.3 Évaluation de l'inter-étalonnage proposé

La correction suggérée est appliquée aux valeurs mesurées par SMMR. Les résultats peuvent être comparés aux autres imageurs microondes. L'inter-étalonnage étant dérivé de GMI, on s'attend à un bon accord avec ce capteur. La première vérification consiste à comparer les observations SMMR étalonnées aux observations SSM/I, pour leur période de chevauchement. Le cycle diurne de T_b dérivé de GMI est utilisé comme un pont entre les observations SMMR et SSM/I qui ont des heures différentes de passage à l'équateur.

La figure II.9 présente les comparaisons de T_b issues de SMMR et SSM/I F08, pour différentes zones moyennées sur leur période commune d'observation en juillet et août 1987, ainsi que le cycle diurne des T_b dérivé de GMI (estimé sur trois années différentes entre 2015 et 2017). Elle montre que les T_b de SSM/I, utilisées sans aucun inter-étalonnage,

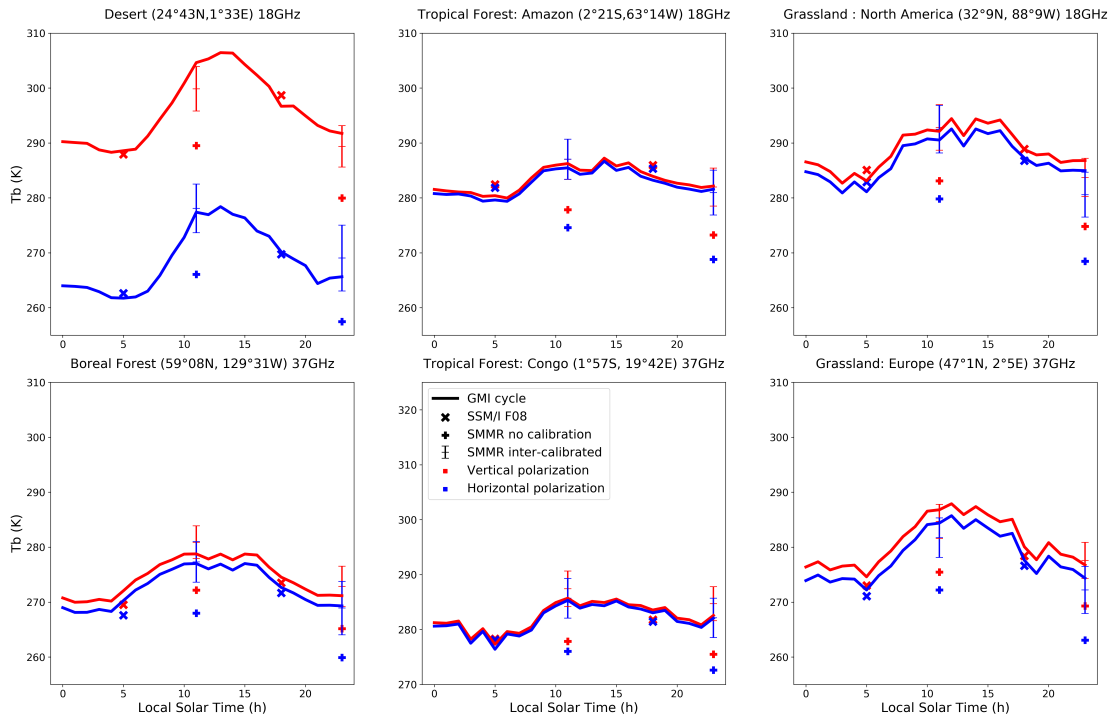


FIGURE II.9 – Cycle diurne de température de brillance (K) de GMI en juillet/août comparé aux valeurs mesurées par SMMR avec et sans la correction proposée, et aux valeurs SSM/I pour les mois de juillet/août 1987 sur différentes zones et fréquences. L’incertitude sur l’inter-étalonnage de SMMR est affichée pour chaque point corrigé.

ne présentent pas un biais extrême par rapport au cycle diurne des T_b estimé à partir de GMI. La différence moyenne entre SSM/I F08 (sans application d’une méthode d’inter-étalonnage) et GMI est d’environ 2.5 K pour les canaux à 18 GHz et d’environ 0.5 K pour les canaux à 37 GHz. Un biais est attendu entre les mesures qui n’ont pas été inter-étalonnées, mais il est plus faible que celui détecté entre SMMR et GMI. Cela confirme les hypothèses concernant l’influence de l’émissivité sur les surfaces continentales ainsi que la faible variation des conditions environnementales dans l’intervalle de 30 ans entre les années 1980 et les années 2010.

La concordance entre les valeurs inter-étalonnées de SMMR et de GMI est clairement améliorée par rapport aux résultats précédents avec les valeurs de SMMR au niveau 1B. Ces résultats montrent que SSM/I F08 et SMMR nouvellement inter-étalonné sont cohérents, du moins dans les limites de l’incertitude résultant de cet inter-étalonnage.

En plus de l’évaluation avec d’autres capteurs, les résultats de notre inter-étalonnage sont comparés avec une méthode alternative développée pour SMMR. Dai and Che (2009) ont construit leur méthode d’inter-étalonnage à partir d’observations dans des zones désertiques et polaires. L’instrument SSM/I F08 est utilisé comme référence et un modèle est développé pour tenir compte des variations diurnes des températures. Les corrections linéaires qui en résultent sont ajoutées sur la figure II.8 (lignes pointillées vertes). Les deux méthodes d’inter-étalonnage sont basées sur des principes très différents. Néanmoins, elles concordent bien pour tous les canaux, avec des pentes similaires et des ordonnées à l’origine légèrement différentes. Cela renforce la confiance dans la méthodologie employée pour dériver les coefficients de l’inter-étalonnage.

4 Conclusion

Ce chapitre s'intéresse aux données d'imageurs microondes utilisables pour l'estimation de la température de surface continentale et leur évaluation pour la réalisation de séries temporelles. En évaluant le FCDR contenant les données de SMMR, il apparaît que les corrections proposées ne permettent pas de corriger les différences observées pour l'instrument SMMR au-dessus des surfaces continentales. Ce chapitre présente une méthode pour étendre les FCDR pour SMMR au-dessus des continents à 18 et 37 GHz, en utilisant l'instrument GMI comme référence, malgré l'absence d'une période d'observation commune. Les observations de GMI sont utilisées pour reconstruire les cycles diurnes de T_b qui contiennent les heures des observations de SMMR, avec l'hypothèse que les conditions environnementales n'ont pas changé radicalement au cours des 30 dernières années et que les différences de fréquences et d'EIA entre les deux instruments peuvent être négligées au-dessus des surfaces continentales. Avec ces hypothèses, l'objectif est de corriger les différences majeures entre les capteurs.

Avant l'inter-étalonnage, la comparaison des observations SMMR et GMI montre une sous-estimation significative des T_b avec SMMR. Cette sous-estimation tend à augmenter avec l'augmentation des T_b . Une régression linéaire est proposée pour les canaux à 18 et 37 GHz dans les polarisations horizontale et verticale afin d'étalonner les observations SMMR sur celles de GMI. La correction SMMR a ensuite été évaluée par rapport aux observations SSM/I F08 sur leur période commune d'observation en orbite, en juillet et août 1987. Le cycle diurne de T_b dérivé de GMI agit comme un intermédiaire entre les deux instruments qui ont des heures de passage différentes à l'équateur. Le résultat donne un bon accord entre toutes les mesures des capteurs.

Cet inter-étalonnage de l'instrument SMMR sur les continents fait partie d'un effort pour étendre les séries de mesure microondes passives sur les continents pour neuf années supplémentaires, de 1978 à 1987. Le travail présenté dans ce chapitre est en cours d'évaluation par le CM SAF pour être intégré dans le prochain cycle de développement du FCDR. Cependant, compte tenu de l'utilisation de l'instrument GMI actuellement en activité comme référence d'étalonnage, de l'incertitude sur les sources d'erreur de l'étalonnage ainsi que de la validité de l'hypothèse de stabilité environnementale, toute comparaison entre les différentes époques doit être menée avec une prudence. L'utilisation de cette correction pour dériver des séries climatiques doit être soigneusement évaluée avant sa mise en œuvre.

L'inter-étalonnage proposé ici est spécifique à un type de surface, mais une correction des observations pourrait être contrainte sur toute la plage de température entre les températures de brillance froides et chaudes. Unifier la correction au-dessus des continents et des océans serait aussi rendue plus facile si les données des coefficients utilisés pour créer les données de niveau 1B étaient disponibles.

Toutefois, les observations microondes au-dessus des surfaces continentales ne sont pas utilisées uniquement pour l'estimation des températures de surface et l'inter-étalonnage proposé ici au niveau des températures de brillance peut trouver de nombreuses autres applications. En effet, le signal observé dans les données de niveau 1 pour certaines applications est plus marqué comme pour l'étendue des glaces (Comiso and Nishio, 2008) ou d'inondation (Sippe et al., 1998). Malgré leur manque de précision pour détecter des signaux fins, les coefficients fournis dans cette étude pourraient être utilisés dans des études sur de longues séries temporelles pour ces variables géophysiques.

En résumé

- Les instruments d'observation dans les microondes se succèdent avec des caractéristiques différentes et des améliorations. Toutes ces mesures doivent être utilisées de manière cohérente pour créer des longues séries temporelles homogènes.
- Les FCDR fournissent des jeux de données de température de brillance inter-étalonnées permettant des estimations de paramètres comme la température de surface continentale sur de longues séries temporelles.
- Le plus ancien imageur microonde multifréquence est SMMR. Son inter-étalonnage global au-dessus des continents par rapport à l'instrument GMI n'avait pas été effectué. En utilisant GMI comme référence, une correction pour les canaux à 18 et 37 GHz de SMMR est proposée.
- Néanmoins, l'incertitude sur les températures de brillance inter-étalonnée reste importante, limitant leur pertinence pour l'étude des tendances climatiques de température de surface. Cependant, l'utilisation de cet inter-étalonnage pour des applications comme l'estimation de l'étendue de la banquise semble pertinente.

Chapitre III

La contamination nuageuse des observations microondes utilisées pour la caractérisation des surfaces

Comme présenté dans l'introduction, l'avantage majeur de la télédétection de la température de surface en microondes par rapport à celle dans le domaine de l'infrarouge est la capacité à observer en dépit d'une couverture nuageuse de la Terre. Une description de la présence de nuages à la surface de la Terre, de leur typologie est présentée. Puis ce chapitre aborde plus en détail l'interaction des rayonnements microondes avec les nuages, parfois accompagnés de précipitations, dans l'atmosphère. Enfin pour mieux filtrer les observations possiblement contaminées par un nuage dans les séries de données, il faut connaître les limites de cette insensibilité aux nuages. Un modèle est développé pour quantifier la probabilité qu'une mesure effectuée par des observations microondes soit contaminée par un nuage, pour différentes fréquences et différentes surfaces. Ce modèle permet d'obtenir un filtre des observations dans le domaine des microondes qui soit adaptable aux différentes plages de fréquences disponibles sur les imageurs microondes.

1 Typologie des nuages et leur impact sur les observations microondes

Les nuages se forment lors de la condensation de la vapeur d'eau contenue dans des masses d'air, en présence d'aérosols. Les mouvements atmosphériques, causés par de la convection, l'orographie, ou par d'autres déplacements de masses d'air provoquent sous certaines conditions l'apparition de nuages. Quand les températures atteignent le niveau de congélation, les gouttelettes peuvent se transformer en cristaux de glace. Ces agrégats de gouttelettes d'eau, d'aérosols, et parfois de glace en suspension dans l'atmosphère sont des phénomènes très facilement observés à l'œil nu. Leur observation et leur classification font partie des premières techniques développées pour la météorologie dès le début du XIX^{ème} siècle (Howard, 1803). Leur étude est aussi critique pour l'analyse des cycles de l'eau et de l'énergie sur Terre. Dans le cycle de l'eau, les nuages sont à l'origine des précipitations, principale source d'eau. Ils influencent aussi le bilan radiatif en réfléchissant une partie du rayonnement solaire incident, mais aussi du rayonnement émis par la Terre.

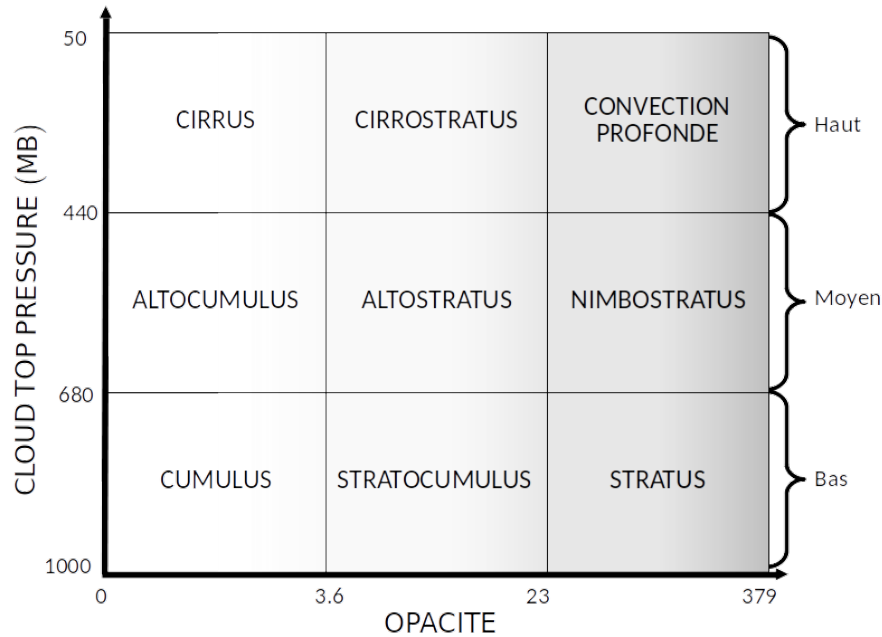


FIGURE III.1 – Table présentant une typologie simplifiée des nuages distribuée suivant la hauteur du sommet du nuage et l’opacité (figure extraite du site de présentation de l’ISCCP).

Cette classification basée sur des observations a été complétée au cours des siècles et la version actuelle présentée dans l’Atlas International des Nuages sous l’égide de l’organisation météorologique mondiale comprend neuf classes majeures de nuages et plusieurs dizaines de variétés et sous genres (Cohn, 2017). On retrouve cette présence concentrée sur les latitudes moyennes, autour des tropiques et au-dessus des océans. Pour l’analyse, on peut utiliser des typologies simplifiées comme celle présentée dans la figure III.1. Basée sur les travaux de l’*International Satellite Cloud Climatology Project* (ISCCP), elle discrimine les types de nuages en fonction de l’altitude du sommet (représentée par la *Cloud Top Pressure*) et de l’épaisseur optique du nuage (*Cloud Optical Thickness*) en se basant sur des observations dans les domaines visible et infrarouge.

Les causes de la formation de ces différentes structures nuageuses sont potentiellement différentes et la variabilité intra-classe est toujours importante. Il ne faut pas oublier que la description s’applique à un continuum d’états nuageux : la différence entre des observations appartenant à deux classes peut être faible. Les particules associées à la présence de nuages ont des tailles variables, de la gouttelette à l’agglomérat de glace comme dans les grêlons. Les plus grosses particules se retrouvent dans les formations nuageuses intenses de convection profonde.

D’autres classifications sont disponibles, basées sur différentes observations satellites. Pour celles-ci, la classification s’effectue au niveau du pixel dont la taille dépend du type de satellite. Par exemple, Meteosat est un satellite géostationnaire européen positionné au-dessus de l’équateur couvrant principalement l’Afrique, une partie de l’Amérique du Sud, l’Europe et le Moyen-Orient, entre $\pm 60^\circ$ de latitude et $\pm 60^\circ$ de longitude. Les canaux du *Spinning Enhanced Visible and InfraRed Imager* (SEVIRI) englobent les domaines visible et infrarouge (Schmid, 2000), avec une taille de pixel d’environ $5 \text{ km} \times 5 \text{ km}$. La classification des nuages de SEVIRI est utilisée dans cette thèse comme référence en raison de sa couverture spatiale et de la disponibilité des données.

Des algorithmes ont été développés pour fournir des informations sur les nuages, telles que la température et la hauteur du sommet des nuages, la teneur en eau, et également le type de nuage, toutes les 15 minutes sur l'ensemble du champ de vision (Derrien and Le Gléau, 2005). Ces algorithmes se basent sur des arbres de décision appliqués aux différentes fréquences pour déterminer les types de nuages. Les classes de nuages sont décrites dans le tableau III.1. Les descriptions des nuages sont différentes de la description données par l'ISCCP mais les approches présentent des similarités. Par exemple les nuages hauts semi-transparentes sont principalement des cirrus d'épaisseur variable, éventuellement au-dessus de nuages plus bas. La classe de nuages fractionnés correspond à des pixels qui ne sont que partiellement nuageux ou à une couverture nuageuse hétérogène.

Type de nuage	Description (<i>English name</i>)
1	Surface terrestre sans nuages (<i>Cloud-free land</i>)
1	Surface océanique sans nuages (<i>Cloud-free sea</i>)
2	Nuages très bas (<i>Very low clouds</i>)
3	Nuages bas (<i>Low clouds</i>)
4	Nuages moyens (<i>Medium clouds</i>)
5	Nuages hauts et opaques (<i>High opaque clouds</i>)
6	Nuages très hauts et opaques (<i>Very high opaque clouds</i>)
7	Nuages hauts, semi-transparentes, fin (<i>High semi-transparent thin clouds</i>)
8	Nuages hauts, semi-transparentes, moyennement épais (<i>High semi-transparent meanly thick clouds</i>)
9	Nuages hauts, semi-transparentes, épais (<i>High semi-transparent thick clouds</i>)
10	Nuages hauts semi-transparentes, au-dessus de nuages (<i>High semi-transparent above lower clouds</i>)
11	Nuages fractionnés (<i>Fractional clouds</i>)

TABLE III.1 – Classification simplifiée des types de nuages pour l'instrument SEVIRI. (Derrien and Le Gléau, 2005)

Les observations provenant de satellites et les classifications effectuées depuis le sol en utilisant des observations dans le visible ou l'infrarouge diffèrent principalement à cause de l'opacité des nuages dans les données visible et infrarouge. Seule l'extrémité de la colonne nuageuse la plus proche de l'observation est visible e.g., le sommet du nuage pour un satellite et la base de celui-ci pour des observateurs sur la Terre. Les cas de colonne atmosphérique avec des couches de nuages à plusieurs altitudes sont donc plus délicats à quantifier avec des observations dans le visible et l'infrarouge.

Une autre observation, qui est consubstantielle à la présence de nuages est celle de précipitations. Certains des nuages sont susceptibles de précipiter, c'est-à-dire que les gouttelettes d'eau ou les particules de glace ne restent plus en suspension dans le nuage et chutent vers la surface de la Terre. Des cas de précipitations sont donc possiblement associés aux classes présentées ci-dessus. Les précipitations ont un effet sur l'observation de la température de surface comme décrit dans la section 1.2 mais peuvent aussi engendrer des changements à la surface. L'émissivité et la température de la surface changent par exemple lorsque l'humidité du sol est modifiée. Ces effets se produisant simultanément, il peut être délicat d'isoler la cause des changements dans les observations satellites microondes.

1.1 Distribution spatiale des types de nuages

Un second point important à prendre en compte pour l'étude de l'effet des nuages sur les observations microondes est leur répartition à la surface de la Terre. On l'a vu, différents phénomènes peuvent être à l'origine des nuages, et ceux-ci se produisent sur différentes parties du globe de manière non homogène. Il y a donc des zones où l'absence de nuages est notable (déserts) et des zones où la présence de nuage est très régulière au cours de l'année (reliefs, zone de convergence inter-tropicale). En moyenne, les nuages couvrent $\sim 60\%$ du globe (Rossow and Schiffer, 1999), et cette distribution inégale renforce encore la nécessité d'obtenir des informations de surface pour les zones où la couverture nuageuse est quasi-constante, pour lesquelles peu d'observations en infrarouge sont exploitables.

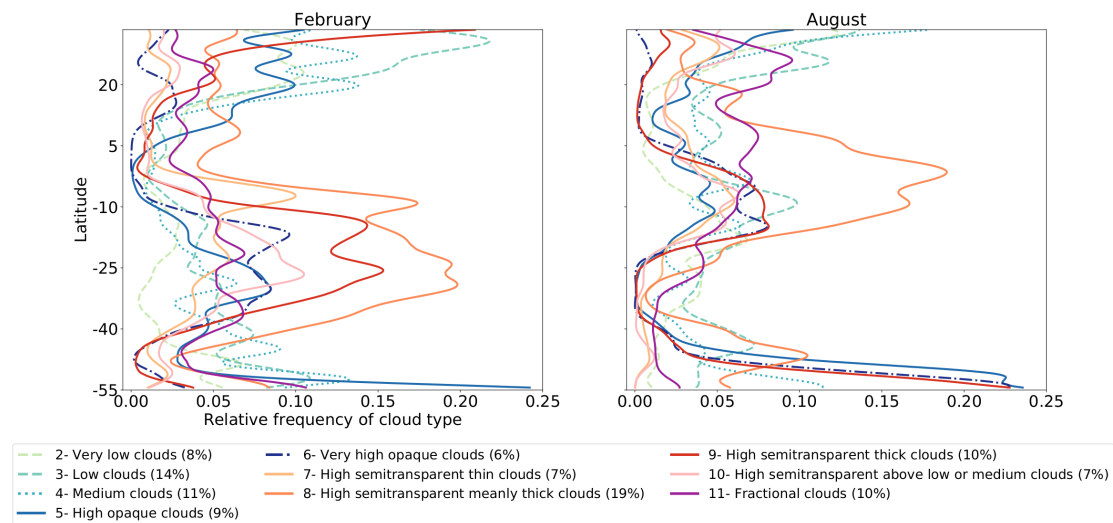


FIGURE III.2 – Fréquence relative des types de nuages en fonction de la latitude, pour février (à gauche) et août (à droite) en 2015, au-dessus des continents situés dans le disque SEVIRI. La fréquence relative moyenne de chaque type de nuage sur ces 2 mois est indiquée dans la légende.

La figure III.2 montre la variation latitudinale des types de nuages présentés dans la section précédente au-dessus des continents dans le disque SEVIRI, pour février et août 2015. Comme attendu la zone de convergence inter-tropicale se déplace entre les deux saisons. Pour les latitudes moyennes de l'hémisphère nord la fréquence des nuages est plus élevée en février qu'en août. La fréquence relative moyenne de chaque type de nuage est affichée, montrant que tous les types de nuages sont bien représentés dans la classification effectuée avec les observations dans le visible et l'infrarouge.

Un tour d'horizon des types de nuages rencontrés sur Terre et de leur répartition ayant été effectué, il faut maintenant comprendre comment ceux-ci interagissent avec les observations microondes passives qui sont utilisées pour mesurer les températures de surfaces continentales.

1.2 Impact des nuages sur les observations microondes

Comme décrit précédemment, les observations microondes utilisées pour mesurer les températures de surface sont effectuées par des instruments placés sur des orbites à quelques centaines de kilomètres d'altitude. Le signal mesuré correspond donc à celui

de la surface observée par l'antenne (cf. chapitre II) et à l'ensemble des autres éléments présents dans le champ de vision, dont l'atmosphère. L'équation de transfert radiatif décrit les sources qui produisent le rayonnement mesuré au niveau du satellite (cf. chapitre I). Dans le cadre de cette thèse la résolution du problème inverse pour estimer la température de surface repose sur des méthodes statistiques, comme mentionné au chapitre I, pour prendre en compte ces différentes sources de variabilité.

Différents phénomènes déjà introduits dans le chapitre I affectent les rayonnements électromagnétiques à différentes fréquences. Cette interaction peut se produire par absorption, émission ou diffusion avec les constituants suivants :

1. Pour les gaz dans l'atmosphère, comme le dioxygène ou la vapeur d'eau, des raies d'absorptions existent affectant les fréquences. On en voit l'effet sur le spectre I.9. Les fréquences pour les mesures de la surface sont choisies en fonction de ces fenêtres de transmission.
2. Comme détaillé dans le chapitre I, tout corps émet un rayonnement à des fréquences dépendant de sa température. Les constituants de l'atmosphère n'échappent pas à la règle et sont donc à l'origine d'une partie du signal mesuré au niveau du satellite. On retrouve bien cette contribution dans l'équation I.5 avec les termes $T_{\uparrow,\nu}$ et $T_{\downarrow,\nu}$.
3. Par absorption et diffusion les gouttes d'eau et cristaux de glace interagissent aussi avec le signal. Ces effets dépendent de la relation entre la taille de la particule et la longueur d'onde. Plus les particules sont grosses plus elles sont diffusantes. En microonde passive en dessous de 40 GHz, les effets de la diffusion sur les nuages sont limités. Ces phénomènes sont décrits par la modélisation de Mie et de Rayleigh.

En utilisant des combinaisons de mesures à des fréquences avec des sensibilités différentes à la surface et à l'atmosphère on peut déduire une estimation de la température de surface, sans avoir besoin d'explicitement spécifier le profil atmosphérique. La plupart des signaux atmosphériques sont pris en compte lors de l'inversion statistique. Les modèles développés reposent sur la synergie entre les différentes longueurs d'ondes pour maîtriser l'impact des variables autres que la température de surface sur les mesures en microondes.

Les interactions possibles des nuages et des précipitations avec les microondes se font soit par des processus d'absorption, d'émission ou encore par diffusion. On peut détailler les processus qui entrent en jeu dans ces phénomènes et les effets attendus pour les T_b mesurées au niveau d'un satellite.

Les cartes de la figure III.3 affichent les températures de brillance composée de plusieurs orbites mesurées par l'instrument AMSR-E à 89 GHz en polarisation horizontale en haut, et à 36.5 GHz en polarisation horizontale en bas. Plusieurs observations sur cette figure sont importantes pour comprendre l'interaction du contenu atmosphérique avec les mesures microondes. Tout d'abord le contraste entre les observations au-dessus des océans et des terres est très clairement visible (du aux différences des émissivités des océans ~ 0.5 et des continents ~ 0.9). À 89 GHz, les structures liées à la présence de vapeur d'eau sont très visibles principalement autour des tropiques. Ces mêmes phénomènes sont par contre indétectables au-dessus des continents. A 36.5 GHz, le contraste entre les terres émergées et l'océan est encore renforcé. Mais les structures liées aux composantes atmosphériques sont beaucoup moins visibles. Il est cependant possible de remarquer des points radiométriquement chaud au centre de structures atmosphériques, par exemple dans l'océan indien. Celles-ci sont attribuées à des structures nuageuses importantes associées à des précipitations. Au-dessus des continents, aucune structure ne semble pouvoir être attribuées à la présence de contamination atmosphérique sur ces cartes, mais la présence de contamination nuageuse reste possible.

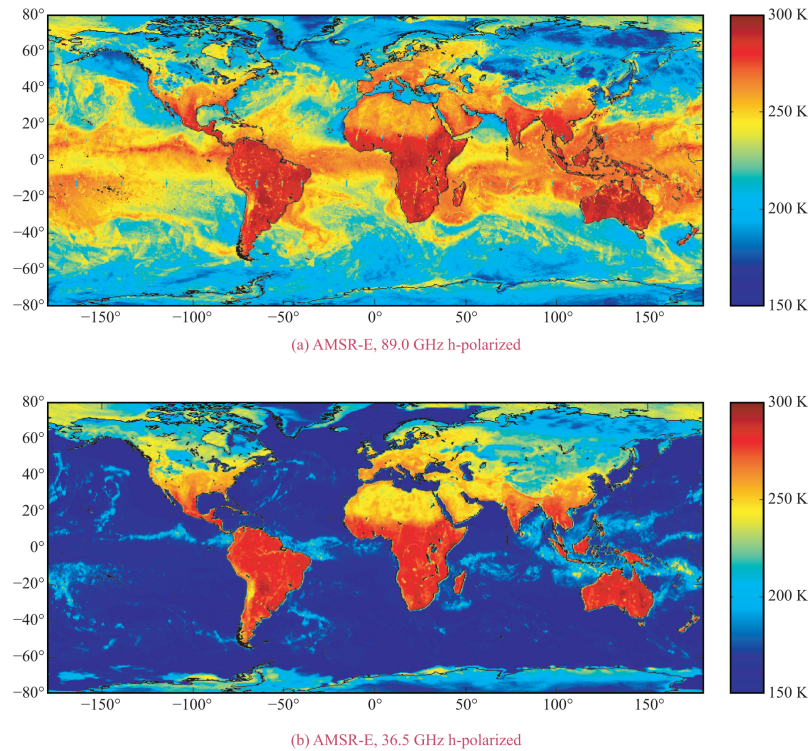


FIGURE III.3 – Cartes de température de brillance (K) obtenues par le radiomètre microonde passif AMSR-E à 89 GHz en polarisation horizontale (a) et 36.5 GHz en polarisation horizontale (b) (figure extraite de Ulaby et al., 2014).

Les mêmes conclusions sont tirées de l'analyse de l'impact de nuages sur les températures de brillance représentés dans la figure III.4. Les courbes correspondent à la différence entre les températures de brillance simulées avec de faibles (en bleu) et de fortes (en rouge) quantités d'eau liquide dans l'atmosphère par rapport à une atmosphère sans nuages. L'effet est très facilement détectable sur la figure de gauche, au-dessus des océans, où la température observée est beaucoup plus chaude en présence de nuages. À droite, au-dessus des continents l'effet de la quantité d'eau liquide sur les températures est négligeable, à cause de la forte émissivité des surfaces continentales qui limite le contraste entre les contributions de la surface et de l'atmosphère.

Emission

La première contribution d'un nuage ou de la pluie consiste en l'émission d'un rayonnement dans le domaine microonde lié à la température physique du nuage ou de la pluie. Cette contribution se superpose aux autres dans la mesure effectuée au niveau du satellite. Donc ce rayonnement n'est détectable à une fréquence donnée que par le contraste que ce rayonnement peut avoir par rapport à des observations de la surface effectuées en l'absence de nuages.

L'effet des nuages ou de la pluie qui peut être détecté dépend fortement du type de surface. La contribution de la surface aux observations microondes passives est proportionnelle à l'émissivité de la surface, qui varie de ~ 0.5 au-dessus des océans à ~ 0.9 au-dessus des sols secs ou des forêts denses. La pluie et les nuages émettent à des températures proches de celles de la surface. Ainsi le contraste entre les gouttelettes composant le nuage ou la pluie et la surface sera généralement plus important sur les océans que sur les

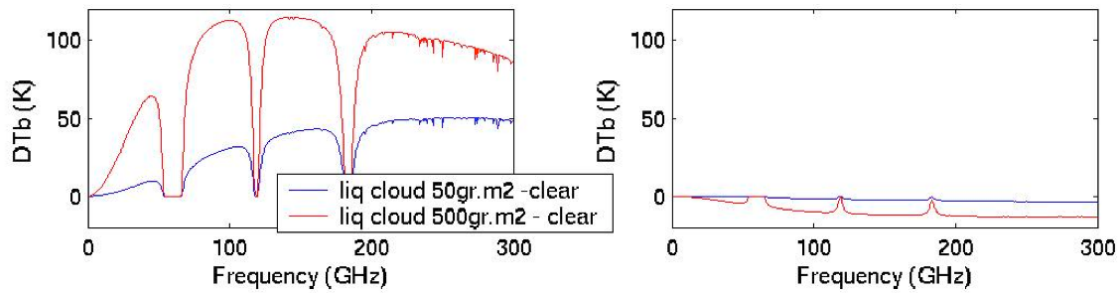


FIGURE III.4 – Différence entre les températures de brillance (K) mesurées pour une atmosphère avec un contenu en eau liquide nuageuse faible ($50 \text{ g} \cdot \text{m}^{-2}$ en bleu) et fort ($500 \text{ g} \cdot \text{m}^{-2}$ en rouge) par rapport à une atmosphère sans nuage. La figure de gauche est donnée au-dessus de l’océan et à droite au-dessus des continents (figure communiquée par C. Prigent).

continents : l’émission de l’eau liquide des nuages et de la pluie augmente la température de brillance sur les océans radiométriquement froids, mais ne présentera pas beaucoup de contraste sur les terres déjà radiométriquement chaudes.

Il en résulte qu’au-dessus des océans, les microondes passives peuvent être utilisées pour quantifier l’eau liquide des nuages et les quantités de pluie (e.g. Greenwald et al. (1993); Kummerow et al. (1998)). Pour les applications à la surface des océans, l’estimation des paramètres de surface est possible en prenant en compte la contribution de la quantité d’eau liquide dans le profil atmosphérique lorsque la transmission atmosphérique est encore suffisamment élevée pour avoir dans le signal observé une contribution significative de la surface.

Extinction

L’extinction du rayonnement traversant le nuage est due pour partie à une absorption et pour partie à la diffusion. Des variables intégrées sur la totalité de l’atmosphère comme le total d’eau liquide dans la colonne atmosphérique fournissent des informations suffisantes pour calculer l’extinction du rayonnement due à l’absorption.

Pour les particules, il est aussi difficile de déterminer précisément la distribution des tailles de particule dans un nuage. Mais on peut appliquer les lois de Mie et Rayleigh, pour les différentes fréquences et les différentes tailles de particules présentes dans un nuage. Les gouttelettes d’eau et de pluie ont un effet diffusif qui augmente avec la fréquence. Pour rappel, les solutions pour la diffusion de Mie et de Rayleigh s’appliquent sous certaines conditions à l’interaction entre une onde et une particule. Basés sur ces résultats, des outils numériques permettent de calculer les effets de la diffusion pour différentes combinaisons d’ondes et de particules (Bauer et al., 2006).

- L’approximation de **Rayleigh** s’applique quand la particule est beaucoup plus petite que la longueur d’onde, alors la diminution de l’intensité du signal causée par la diffusion est faible, et ce, d’autant plus que la taille de la particule est faible devant la longueur d’onde.
- La formulation de **Mie** peut s’utiliser quant à elle lorsque la taille de la particule est de l’ordre de grandeur de la longueur d’onde, on a alors une diffusion variable suivant l’angle d’incidence du rayonnement par rapport à la particule.

- Lorsque la taille de la particule rencontrée par une onde dépasse la taille de celle-ci les lois de l’optique géométrique s’appliquent, ce qui n’arrive pas avec les particules rencontrées dans des nuages pour des fréquences dans le domaine des microondes.

Ces calculs peuvent être effectués pour des cas idéaux (sphères avec une distribution des tailles connue, réparties de manière homogène dans une couche atmosphérique) mais ne sont que des approximations de la réalité qui permettent bien cependant d’expliquer la semi-transparence des nuages aux observations dans le domaine microonde.

Particule	Taille	Type de diffusion à 37 GHz	Type de diffusion à 183 GHz
Gouttelettes d’eau	5-50 μm	Négligeable - Rayleigh	Rayleigh
Gouttes d’eau	0.1-3 mm	Rayleigh - Mie	Mie
Cristaux de glace	10-100 μm	Rayleigh	Rayleigh - Mie
Neige	1-10 mm	Mie	Mie
Grêle	~ 10 mm	Mie	Mie

TABLE III.2 – Description des particules associées à la présence de nuages, leur taille moyenne et le type de diffusion associé à 37 et 183 GHz. (A. Geer ECMWF)

La table III.2 rappelle les ordres de grandeurs de ceux-ci. On voit que pour les basses fréquences, seules les plus grosses particules peuvent avoir un effet diffusif (effet négligeable pour les gouttelettes d’eau de petite tailles) et que la taille minimale des particules ayant un impact augmente avec la fréquence. On constate aussi grâce à cette table que les plus grosses particules sont celles composées de glace (comme la grêle) qui se rencontrent uniquement dans les épisodes convectifs les plus forts (section 1).

On retrouve ces résultats sur la figure III.5, où le comportement de la diffusion en fonction de la longueur d’onde et de la taille des particules est résumé. Les frontières ne sont pas strictes, comme on a pu le voir au chapitre I. En présence de glace, la température de brillance diminue. Ces basses températures induites par les nuages sont plus faciles à détecter au-dessus des continents qu’au-dessus des océans, car ceux-ci sont radiométriquement plus froids que les continents.

La figure III.6 résume les différents effets dus à la présence de nuages en microonde à 37 et 85 GHz. À une fréquence peu élevée (37 GHz), le phénomène dominant dans la signature radiométrique du nuage sera l’émission due principalement à l’eau liquide du nuage ou des précipitations. La diffusion y est faible mais avec une augmentation de la fréquence la diffusion prend plus d’importance et fait donc diminuer la T_b mesurée. Cet effet est visible au nord ouest de l’Australie dans la figure III.3 où le cœur d’un système nuageux est plus froid que les nuages environnants.

En combinant les effets de diffusion et d’émission pour les nuages, on peut utiliser les mesures microondes pour caractériser les distributions de particules ou le contenu en glace dans les nuages (Turk et al., 1994; Petty, 2001) au-dessus de l’océan. Suivant les fréquences, la sensibilité des canaux aux constituants atmosphériques est variable, qu’ils soient gazeux, liquides ou solides. La combinaison des informations provenant de plusieurs canaux permet une estimation robuste de la température de la surface continentale dans la plupart des cas. Mais au-dessus des continents, la détection des nuages et de la pluie à l’aide de microondes passives est beaucoup plus compliquée comme illustré dans les travaux de Spencer et al. (1989) ou Aires et al. (2001). La présence d’un nuage ou de pluie

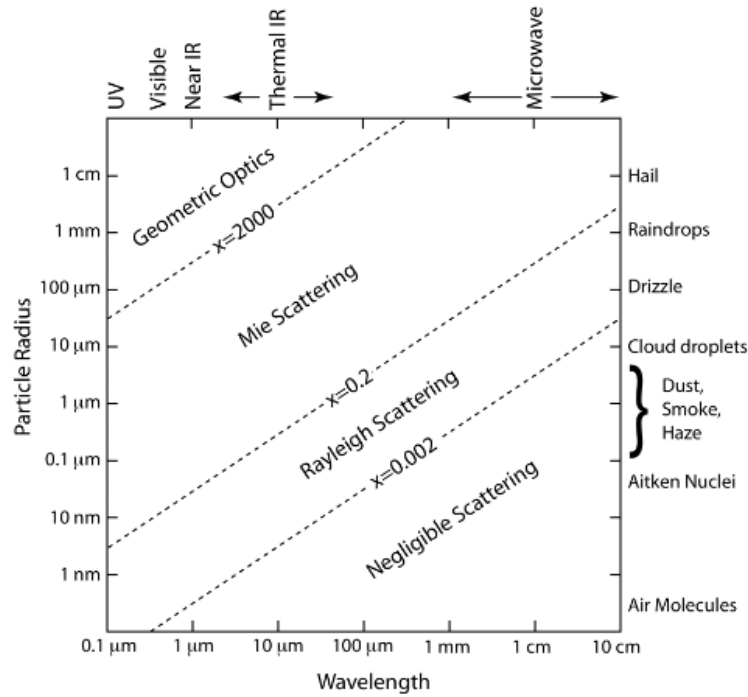


FIGURE III.5 – Graphique représentant les limites des domaines de diffusion de Rayleigh, Mie et de la diffusion géométrique, en fonction de la longueur d'onde et des tailles de particules (figure extraite de Petty, 2006).

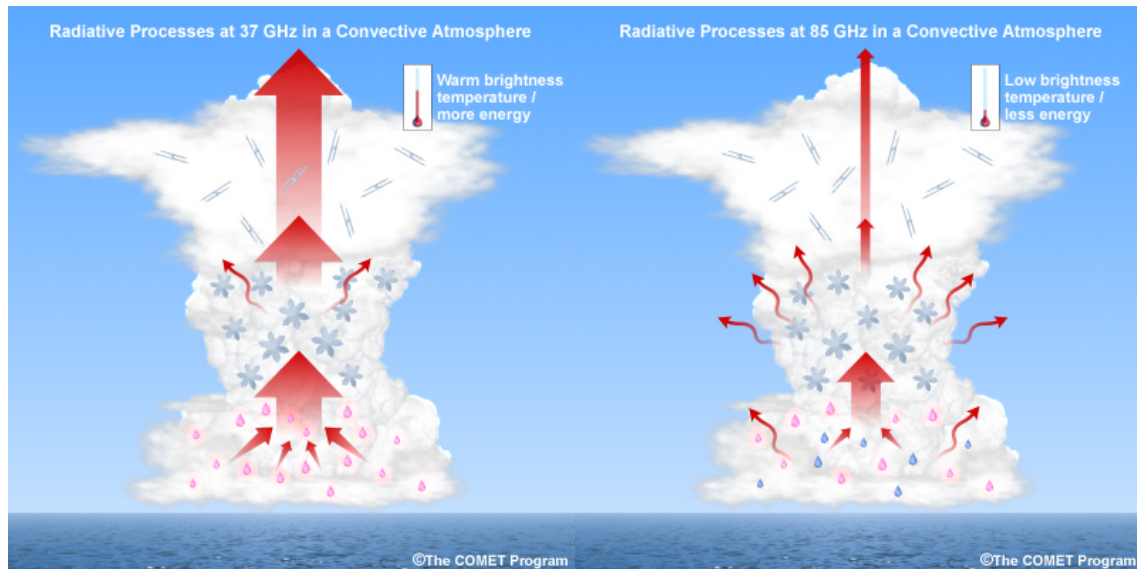


FIGURE III.6 – Contribution d'un nuage au rayonnement observé au-dessus de l'océan pour une fréquence de 37 GHz (gauche) et 85 GHz (droite) (figure extraite de *University Corporation for Atmospheric Research* COMET Program).

peut alors compromettre l'estimation des paramètres de la surface terrestre.

1.3 Méthodes de détection ou de mesure de la présence nuageuse en microondes

La différence entre la signature des nuages en microondes passifs au-dessus des océans et des surfaces continentales est importante. La quantification des précipitations à base d'observations multi-fréquence (et parfois multi-instruments) fait partie des objectifs des missions comme la *Global Precipitation Mission* (GPM) (Hou et al., 2014), mais les nuages et le contenu en eau liquide sont plus délicats à mesurer. Il faut donc bien distinguer les méthodes au-dessus des surfaces continentales de celles au-dessus des océans mais aussi les mesures de précipitations de celles de détection de nuages.

Les méthodes estimant la présence de précipitations ou de nuages au-dessus des océans sont nombreuses, on pourra citer :

- Le contenu en eau liquide des nuages est estimé par différentes paramétrisations et les meilleurs résultats sont obtenus en combinant tous les canaux entre 19 et 85 GHz (Karstens et al., 1994). Ces études évaluent aussi la possibilité de comparer les résultats obtenus par les observations microondes par rapport à celles dans le visible/infrarouge.
- Le taux de pluie peut être estimé avec des combinaisons de canaux aux deux polarisations à 19, 37 et 85 GHz pour créer des indices de diffusion et d'absorption (Petty, 1994).
- L'analyse de l'impact de la présence de nuages sur les températures de brillance au-dessus des océans est poursuivie par Wentz (1997) pour limiter les erreurs dans les estimations des paramètres géophysiques à la surface de l'océan pour l'instrument SSM/I (section 1.2).
- Enfin en regroupant des observations de plusieurs instruments sur 18 années, une climatologie du contenu en eau liquide des nuages a pu être créée (O'Dell et al., 2008). Les résultats mettent bien en évidence la distinction délicate à faire entre le contenu précipitant et le contenu liquide des nuages.

Les avancées méthodologiques ne changent pas le fait que les fréquences les plus basses sont moins sensibles aux nuages et à la présence de précipitations.

Au-dessus des continents, la quantification est plus délicate et les algorithmes de filtrage de données bruitées ou uniquement de détection sont principalement développés :

- Des méthodes de détection de la pluie ont été développées pour le capteur SSM/I en s'appuyant le signal de diffusion à 85 GHz. Elles utilisent des arbres de décision avec uniquement les canaux à 19 GHz (V et H) et à 23 et 85 GHz en polarisation verticale (Grody, 1991). Des améliorations à cette méthode de détection ont été apportées par Ferraro (2004) qui utilise la combinaison des polarisations verticales à 19, 22 et 85 GHz pour déterminer un indice de diffusion permettant de détecter les précipitations.
- Des méthodes de filtrage des nuages ont également été dérivées, comme par Long et al. (1999) qui ont analysé des séries chronologiques de la température de brillance à 85 GHz avec différentes méthodes pour supprimer la perturbation des nuages sur les images SSM/I pour des applications à la surface terrestre.
- Pour l'estimation de l'humidité de la haute troposphère avec des mesures satellites autour de la raie de vapeur d'eau à 183.31 GHz, Buehler et al. (2007) ont déve-

loppé des filtres se basant sur des différences entre des canaux pour éviter les pixels contaminés par les nuages.

- Aires et al. (2011a) ont utilisé une méthode de réseau de neurones entraîné sur les produits de nuages SEVIRI (*Spinning Enhanced Visible and Infrared Imager*) de deuxième génération de Meteosat pour créer un masque de nuages et une classification à partir des mesures en microonde *Advanced Microwave Sounding Units A* et *B* (AMSU-A / AMSU-B) avec des canaux de 23 GHz à 183 GHz. Des modèles statistiques ont été construits séparément au-dessus des continents et des océans pour détecter les nuages et les classer en ciel clair, nuages bas, moyens ou élevés.
- Un classificateur Bayésien et un réseau de neurones ont été développés en utilisant les mesures de sondage de l'instrument SSMIS avec des fréquences jusqu'à 183 GHz pour détecter les observations pour lesquelles le contenu en eau liquide disponible dans les réanalyses de l'ECMWF est non nul (*Liquid Water Path* ≥ 0). La précision atteignant 90% au-dessus de l'océan et 80% pour les continents (Islam et al., 2015).

On constate, à travers ce rapide aperçu de la littérature sur ce sujet, que les méthodes sont spécifiquement développées pour des canaux disponibles sur certains instruments et reposent soit sur des modèles d'apprentissage automatique, soit sur une simple détection d'anomalie dans une série de données.

Le filtrage, en analysant les anomalies, permet de détecter les valeurs qui s'éloignent trop de la distribution normale des températures, et qui donnent lieu le plus souvent à des estimations incohérentes. La méthode la plus simple est de supposer que les températures de brillance pour un pixel donné ont une distribution normale et que les écarts supérieurs ou inférieurs à plusieurs déviations standards sont trop extrêmes pour pouvoir être prises en compte. Ces méthodes reposent sur des connaissances assez précises soit de l'évolution temporelle des T_b en un point donné, soit de l'historique des températures mesurées pour un pixel.

Les méthodes reposant sur un apprentissage automatique sont quant à elles plus générales et peuvent être appliquées à des mesures indépendamment de l'historique sur ce point. En général elles se basent sur la complémentarité des multiples canaux dans la mesure des surfaces par rapport aux nuages. Certaines études ciblent explicitement la prévision numérique du temps pour essayer de limiter l'impact dans l'assimilation de données des mesures bruitées par un nuage. Les travaux de Islam et al. (2015) mettent en évidence la difficulté de détecter les nuages fins associés aux valeurs faibles de contenu en eau liquide. Cependant, le cas des contaminations dans les observations à basse fréquence ≤ 40 GHz (utilisées pour l'estimation de la température de surface) ne font pas l'objet d'une étude spécifique.

Une méthode générale pouvant s'accommoder des changements de canaux disponibles à bord des différents imageurs microondes serait donc bénéfique pour améliorer la détection de ces contaminations pour les produits de surface en microonde. C'est ce qui est développé dans les sections suivantes.

2 Méthode d'apprentissage automatique

Pour construire des relations empiriques entre des observations et un phénomène qualitatif comme la présence de nuage ou quantifié comme une estimation de la précipitation,

les méthodes d'apprentissage automatiques sont performantes. Cette section décrit le cadre général des méthodes empiriques utilisées dans ce chapitre, mais aussi pour d'autres applications au cours de la thèse (chapitre V). Quelques principes méthodologiques sur l'apprentissage automatique en particulier en utilisant des réseaux de neurones sont détaillés.

2.1 Modèles statistiques

Les modèles statistiques sont une description approximative de données par une combinaison de distributions et d'opérateurs mathématiques. Ils sont construits empiriquement en se basant sur la relation entre des données observées et une ou des variables ciblées. Ils ne sont pas basés sur des relations décrites par des lois physiques, mais utilisent l'information contenue dans les données pour simuler celles-ci.

Pour un vecteur de donnée d'entrée \mathbf{x} et des valeurs observées \mathbf{y} , on cherche les paramètres d'un modèle f tel que $f(\mathbf{x}) = \mathbf{y}$. La variable ciblée peut être une valeur quantitative ou catégorique, et dans ce cas elle peut être binaire.

Parmi les modèles statistiques on peut citer les régressions linéaires, logarithmiques (Dreiseitl and Ohno-Machado, 2002) ou polynomiales, qui comportent peu de paramètres libres ; les arbres de décisions ou encore les réseaux de neurones, où le nombre de paramètres peut être très élevé.

Les réseaux de neurones (Rumelhart et al., 1986) sont des systèmes de traitement d'information reposant sur une combinaison d'opérations élémentaires successives appliquées de manière méthodique à des données fournies en entrée. L'inspiration provient de la description biologique du fonctionnement des systèmes nerveux (Lettvin et al., 1959) dans laquelle on aurait remplacé les processus biologiques par des opérations mathématiques. Ils trouvent de nombreuses applications pour résoudre des problèmes de traitement d'image (Bottou et al., 2002) ou d'intelligence artificielle pour des jeux (Tesauro, 1994) mais aussi simuler les phénomènes physiques (Krasnopolsky, 2020). L'unité de base d'un réseau de neurones est le neurone qui effectue deux opérations :

1. La somme pondérée des entrées,
2. Le calcul d'une fonction d'activation sur la somme pondérée.

Ces neurones sont ensuite combinés suivant une structure qui définit la succession des opérations à appliquer pour estimer la valeur de sortie étant donné une entrée. L'apprentissage des poids du réseau peut se faire de manière supervisée (minimisation d'une fonction de perte) ou non supervisée (c'est-à-dire en extrayant des structures des données sans connaître la structure a priori). Le cas des réseaux non supervisés ne sera pas traité ici. La structure la plus simple des réseaux supervisés est le réseau de neurones multicouches (*Multilayer Perceptron* - MLP) complètement connecté (Hornik et al., 1989) (aussi appelé dense), dans lequel chaque neurone est connecté à tous les neurones de la couche précédente, comme décrit sur le schéma de la figure III.7.

Pour un réseau tel que celui représenté sur la figure III.7 avec 5 neurones dans la couche cachée ($N_{neuron} = 5$), avec \hat{y} la prédiction binaire qui représente la présence ou l'absence d'un nuage et \mathbf{T} le vecteur des températures de brillance en entrée ($N_{input} = 4$). On note Φ et Φ_{sortie} la fonction d'activation utilisée pour les neurones cachés et en sortie :

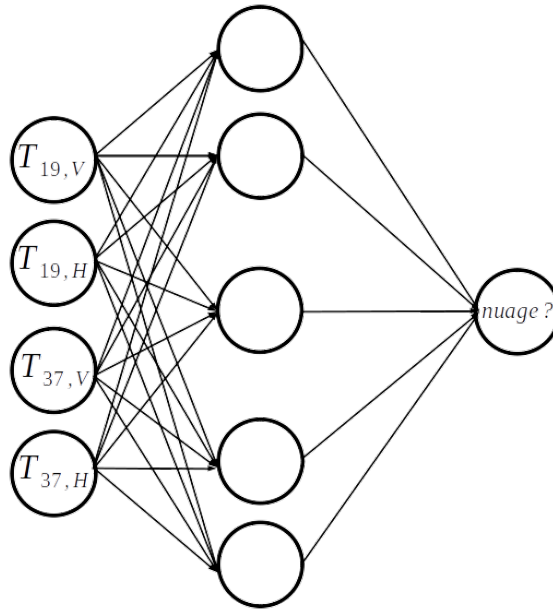


FIGURE III.7 – Schéma représentant un réseau de neurones multicouche avec 4 entrées (par exemple des températures de brillance), 5 neurones dans une couche cachée et 1 neurone en sortie.

$$\hat{y} = \Phi_{sortie} \left(\sum_{j=1}^{N_{neuron}=5} w_j \times \Phi \left(\sum_{i=1}^{N_{input}=4} w_{ji} \mathbf{T}_i + b_j \right) + b_{sortie} \right). \quad (\text{III.1})$$

Les paramètres entraînés sont w_{ji} et w_i les poids dans les connexions du réseau et les termes b_j et b_{sortie} du biais de chaque neurone.

Les MLPs sont des approximateurs universels qui peuvent, avec suffisamment de paramètres, approximer n'importe quelle fonction (Hornik, 1991). Ils peuvent être utilisés pour des problèmes de régression (déterminer une valeur en fonction des entrées) ou de classification (attribuer une catégorie à une observation en entrée). Leur usage pour résoudre des problèmes inverses en particulier en présence de phénomènes non-linéaires est très développé (Krasnopolsky and Schiller, 2003).

On distingue deux types de paramètres dans un réseau de neurones :

- Les **hyperparamètres** donné pour la phase d'entraînement pour le problème donné. Ils peuvent être adaptés en fonction des performances du modèle sur un jeu de données de validation. On comprend dans ceux-ci le nombre et l'organisation des neurones, le nombre d'itérations sur les données d'apprentissage, les fonctions d'activation utilisées, la fonction de perte, l'algorithme de mise à jour des poids, mais aussi la méthode d'initialisation de ceux-ci. Un changement d'un de ces paramètres peut avoir un impact non-négligeable sur les résultats obtenus.
- Les **poids** du réseau permettent de déterminer les relations entre les entrées et sorties du réseau. Ils correspondent aux coefficients de pondération et de biais w, b dans l'équation III.1. Ils peuvent être initialisés de manière aléatoire et sont ensuite corrigés au fur et à mesure de l'apprentissage par l'algorithme de mise à jour des poids suivant la fonction de perte.

Parmi ces hyperparamètres il est important de préciser trois éléments :

La **fonction de perte (ou de coût)** définit l'objectif du modèle. Cet objectif est approché en minimisant l'écart entre les sorties prédites par le modèle et les valeurs observées. La descente du gradient de la fonction de coût permet de s'approcher d'une valeur minimale, plusieurs approches existent dont la descente de gradient stochastique (Amari, 1993). On détermine ainsi les paramètres du réseau en appliquant l'algorithme de mise à jour des poids de manière itérative. Suivant l'objectif à atteindre différentes fonctions de perte peuvent être utilisées, les plus simples étant l'erreur quadratique ou l'erreur absolue.

Les **fonctions d'activation** sont les fonctions qui sont appliquées en sortie de chaque neurone à la somme des entrées pondérées. Le plus souvent des fonctions non linéaires sont utilisées comme des tangentes hyperboliques, la fonction *softmax*, ou la fonction unité linéaire rectifiée. La figure III.8 présente quelques exemples de fonctions qui peuvent

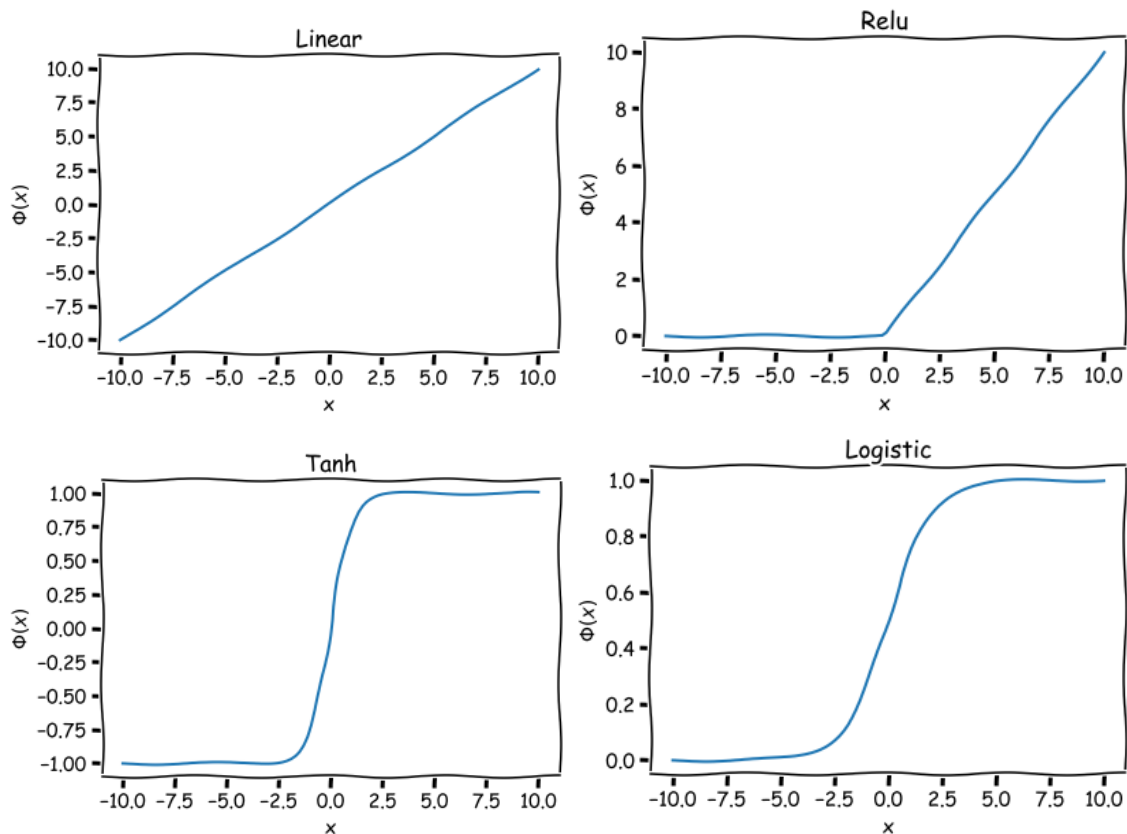


FIGURE III.8 – Exemples de fonction d'activation Φ utilisées pour les neurones. Dans l'ordre de haut à gauche en bas à droite : fonction linéaire, unité linéaire rectifiée, tangente hyperbolique et logistique (sigmoïde).

être utilisées pour l'activation des neurones. La sortie linéaire ne présente pas de non-linéarité et un réseau de neurones construit uniquement avec des fonctions d'activation linéaires est équivalent à une régression multilinéaire. Les autres fonctions présentent des comportements comme la continuité de la dérivée qui peuvent être souhaitables.

L'**algorithme de mise à jour des poids** décrit comment les poids utilisés à chaque connections sont mis à jour en fonction de l'erreur commise sur les exemples. L'algorithme de rétropropagation du gradient (Rumelhart et al., 1986) calcule pour chacun des poids du réseau la correction à apporter en fonction de l'erreur telle que définie par la fonction de perte, calculée en sortie du réseau. De nombreuses améliorations tentent d'accélérer la convergence vers la combinaison de poids optimale en limitant les fluctuations aléatoires

du gradient (Kingma and Ba, 2015; Ruder, 2016).

Tous les modèles de cette thèse sont implémentés à l'aide de la librairie *Keras* disponible en python (Chollet et al., 2015). L'entraînement est arrêté lorsque la perte ne diminue pas pendant 5 itérations consécutives, ce qui arrive après quelques centaines d'itérations sur la base d'apprentissage selon le réseau et la taille des entrées. Le matériel utilisé pour cette étape est un ordinateur portable de bureau standard, avec 4 cœurs et 16 Go de mémoire vive.

Les réseaux ne garantissent pas l'obtention de la combinaison optimale des poids pour minimiser la fonction de perte choisie et, compte tenu de la haute dimensionnalité des problèmes étudiés, des minima locaux peuvent être trouvés. Cependant, dans la pratique, les capacités de ces méthodes à fournir de « bons » résultats ont été démontrées pour une grande variété de problèmes et la solution trouvée sur la base d'apprentissage est toujours testée sur des exemples différents (comme décrit dans la section 2.2) pour valider les capacités de généralisation du modèle ainsi construit.

2.2 Données d'entraînement, de test et de validation

Dans les problèmes d'apprentissage automatique, où les paramètres d'un modèle sont adaptés pour réduire la valeur d'une fonction d'erreur sur un problème donné, les données utilisées pour obtenir ces paramètres sont cruciales. On définit le modèle comme la paire composée d'un ensemble des opérations mathématiques appliquées aux entrées et d'un jeu de données qui représente un sous-ensemble de l'espace du problème. Comme vu dans la section précédente, les hyper-paramètres du modèle qui définissent sa structure ou le type de relation utilisé entre les entrées et les sorties ne sont pas changés entre les itérations d'apprentissage, mais sont adaptés pour définir le modèle le plus adapté au problème. L'entraînement consiste à minimiser une fonction de perte de telle sorte que la sortie du modèle se rapproche au plus près de la classe ou de la valeur réelle qui est observée pour une entrée donnée. Comme pour tout modèle, c'est la capacité à généraliser les résultats pour des entrées inconnues qui détermine sa performance.

Dans la pratique, on utilise au moins des jeux de données définis comme suit :

- Le **jeu d'entraînement**, correspond aux données sur lesquelles les paramètres du modèle sont appris. Plusieurs itérations successives sur l'ensemble des données sont souvent nécessaires. On utilise le plus souvent de 50 à 70 % du total de donnée disponible.
- Le **jeu de validation** sert à valider les résultats du modèle pour des données non utilisées lors de l'apprentissage. Ces données permettent de valider et choisir le modèle et les hyper-paramètres choisis et aussi de stopper les itérations sur le jeu d'entraînement pour éviter les problèmes de sur-apprentissage. Ce jeu de donnée représente de 15 à 30 % des données totales disponibles.
- Le **jeu de test** est finalement utilisé, il n'entre pas en jeu lors de la définition du modèle ni pendant les différentes étapes d'entraînement. On utilise des données qui n'ont pas été présentées au modèle lors des phases précédentes. Celles-ci peuvent provenir de données complémentaires disponibles en dehors de la base, et sinon environ 15 % de la taille de la base totale est utilisé. Cela permet de déterminer effectivement les performances du modèle et ses capacités de généralisation.

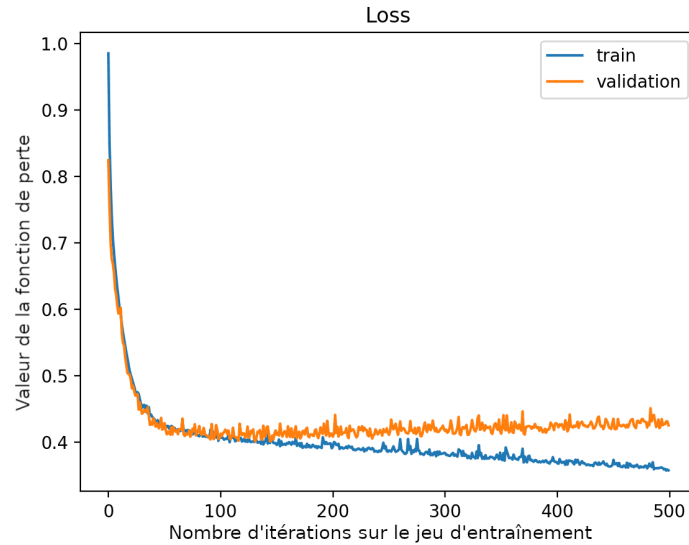


FIGURE III.9 – Exemple de courbes de la perte au cours de l’entraînement d’un réseau de neurones. En bleu la courbe pour l’ensemble d’entraînement, en orange pour l’ensemble de validation. (figure extraite du site de J. Brownlee)

Si la taille et la représentativité du jeu de données pour le problème à modéliser est suffisante alors l’erreur de test et de validation sont sensiblement identiques. Lors de la phase d’entraînement, l’erreur diminue pour le jeu de donnée de validation et pour le jeu de donnée d’entraînement. Si en poursuivant l’entraînement l’erreur sur le jeu de donnée de test augmente à nouveau alors il y a un phénomène de sur-apprentissage qui diminue la qualité du modèle entraîné. On peut observer un exemple de ce phénomène sur la figure III.9 avec la courbe bleu de l’erreur sur le jeu d’entraînement qui décroît malgré une augmentation de l’erreur sur le jeu de données de validation.

Pour que l’entraînement du modèle fonctionne même lorsque les entrées ou les sorties représentent des quantités différentes, il peut être nécessaire d’appliquer un pré-traitement aux données. En plus d’un nettoyage des données pour éviter les valeurs aberrantes, une mise à l’échelle est souvent appliquée. Celle-ci permet de rendre plus homogènes des valeurs aux distributions très différentes. Une différence importante dans la magnitude des sorties peut poser des problèmes lors du calcul de la fonction de perte. En entrée, le pré-traitement améliore souvent la convergence en permettant aux sommes pondérées des variables d’être dans les parties non saturées des fonctions d’activation. Dans les parties linéaires des fonctions d’activations le un gradient est non nul et permet l’apprentissage.

La **normalisation**, recentre les valeurs dans un intervalle pré-défini, par exemple $[-1; 1]$. En appelant X la variable d’entrée à normaliser alors :

$$X_{normalise} = \frac{X - X_{min}}{X_{max} - X_{min}}.$$

La **standardisation** centre la moyenne de l’échantillon en 0 et limite la déviation standard à 1. Pour une variable X , avec une moyenne μ et un écart type σ :

$$X_{standardise} = \frac{X - \mu}{\sigma}.$$

3 Estimer une probabilité de contamination nuageuse sur les observations en microondes passives, au-dessus des océans et des continents

La section suivante est partiellement issue de l'article : S. Favrichon, C. Prigent, C. Jimenez, and F. Aires, "Detecting cloud contamination in passive microwave satellite measurements over land," Atmos. Meas. Tech., vol. 12, no. 3, pp. 1531–1543, 2019.

Comme abordé dans la section 1.3, les méthodes de détection de nuages et de précipitations applicables aux instruments microondes sont spécifiques à un instrument et à la gamme de fréquence associée.

L'objectif de cette section est donc de développer une méthode qui permette d'indiquer la contamination nuageuse/pluvieuse d'une observation microonde passive au-dessus des continents, pour différentes gammes de fréquences disponibles à bord des générations successives d'instruments. Pour ce faire, une approche similaire à celle de Aires et al. (2011a) est utilisée pour développer un indicateur de nuage/pluie tel que détecté par différentes gammes de fréquences microondes entre 10 et 190 GHz.

En comparant les observations de l'imager GMI à une classification de nuages fondée sur des observations dans le visible/infrarouge, un modèle statistique est entraîné permettant de détecter la présence d'un nuage dans les températures de brillance mesurées à différentes fréquences. Dans les sections suivantes les données utilisées pour l'étude sont décrites, puis la méthodologie mise en place. Les résultats sont présentés pour les océans, puis au-dessus des continents.

3.1 Sources de données

Les différentes sources de données sont décrites ici, à savoir la classification des nuages de SEVIRI et les températures de brillance de GMI. Comme mentionné dans les sections et chapitres précédents, SEVIRI est une source de mesure fiable, disponible sur une longue période de temps et surtout avec un échantillonnage temporel de 15 minutes. Ce qui rend la combinaison avec GMI, dont les heures de passage à l'équateur sont variables, très pratique pour observer à toutes les heures du jour.

Information sur la présence de nuages

La classification de nuages développée pour l'instrument SEVIRI par Derrien and le Gléau (2010) est fournie par le *Climate Satellite Application Facility* (CM SAF) de manière quotidienne depuis 2004. La version 2013 de l'algorithme de classification des nuages SEVIRI est utilisée, présentée dans la section 1. La pluie n'est pas détectée séparément du nuage : certains nuages sont susceptibles de précipiter et les pixels nuageux incluent évidemment des cas de précipitation. Six jours complets de chaque mois de 2015 forment un ensemble de 72 journées différentes qui représentent une grande variation des types de nuages et des conditions de surface possibles, couvrant l'ensemble des cycles diurnes et annuels.

Observations microondes

Cette classification nuageuse doit être combinée avec des observations microondes pour pouvoir comprendre l'interaction des microondes avec les nuages et détecter les situations nuageuses qui ont un impact sur les microondes. Comme décrit dans le chapitre précédent, les nombreux imageurs observant dans le domaine microonde sont différents. Des fréquences proches sont utilisées sur les instruments successifs et ils ont des caractéristiques suffisamment proches (voir Tableau II.1) qui peuvent permettre un traitement similaire des données à partir de 1978.

On peut simplement regrouper les imageurs en fonction des canaux disponibles à bord des appareils pour étudier les différentes sensibilités de ces groupes de canaux aux nuages :

- En dessous de 40 GHz : Cet ensemble particulier de canaux est disponible entre autre à bord de SMMR. Les fréquences disponibles pour l'instrument GMI sont : 18.7 GHz (V, H), 23.8 GHz (V), et 36.5 GHz (V, H).
- Avec des fréquences allant jusqu'à 90 GHz : par exemple sur les instruments de la série SSM/I (cf. section 1.2). À bord de GMI, la fréquence de 89 GHz (V, H) est aussi présente.
- Avec des fréquences allant jusqu'à 190 GHz : C'est par exemple le cas de SSMIS ainsi que GMI avec des canaux à 165.5 GHz (V, H) et 183.3 GHz (V).

Les observations microondes passives proviennent de la mission GMI car elle inclut toutes les fréquences disponibles à bord des autres imageurs et ses heures de passage à l'équateur ne sont pas fixes (comme décrit dans la section 1.2). Dans cette étude les T_b calibrées disponibles dans le produit de niveau 1C-R sont utilisées. Les données GMI de 2015 ont été téléchargées, pour les 72 jours correspondant à la sélection SEVIRI décrite ci-dessus.

Préparation du jeu de données et analyse préliminaire

Les données SEVIRI et GMI ont des résolutions spatiales et temporelles très différentes. Il faut donc trouver les observations les plus proches et les colocaliser sur une grille commune pour un traitement ultérieur. Les pixels avec un indicateur annonçant une qualité dégradée sont évités, tant pour GMI que pour SEVIRI. Pour chaque observation par GMI, la mesure SEVIRI la plus proche dans le temps est sélectionnée. Avec des observations SEVIRI toutes les 15 minutes, il y a une différence maximale de 7.5 minutes entre les mesures GMI et la classification SEVIRI correspondante. Compte tenu de la résolution spatiale, plusieurs pixels SEVIRI correspondent à un seul pixel de GMI. Donc dans les données d'entraînement, seules les observations GMI associées à une unique classe de SEVIRI sont conservées. Une variabilité sera toujours présente entre la radiance observée et le type de nuage de SEVIRI, en raison de nuages inhomogènes, à une échelle inférieure à la taille de l'empreinte au sol de la mesure en microondes, en particulier pour les canaux aux plus basses fréquences.

Les pixels situés au-dessus de 55° Nord et au-dessous de 50° Sud sont écartés : ils sont de plus grande taille dans les données SEVIRI et sont sujets à plus de contamination par la neige et la glace. Le masque terrestre de GMI est appliqué pour séparer les surfaces terrestres et les surfaces d'eau. Pour une première analyse de la sensibilité des microondes aux nuages, les distributions des températures de brillance des microondes (T_b) sont tra-

cées dans la Fig. III.10, pour les différents types de nuages et pour des fréquences GMI sélectionnées, sur les océans (colonne de gauche) et les terres (colonne de droite).

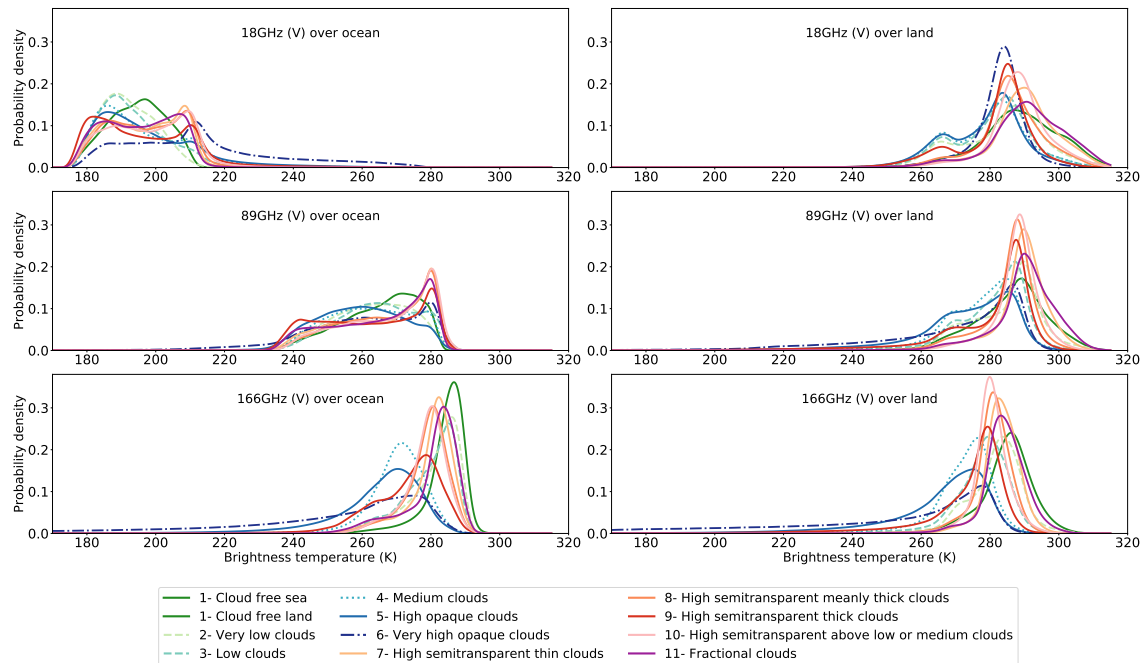


FIGURE III.10 – Distributions de probabilité des T_b observés par GMI pour différents types de nuages à 18 GHz (haut), 89 GHz (milieu), et 166 GHz (bas) pour la polarisation verticale, sur l’océan (gauche) et la terre (droite) à partir de l’ensemble de données.

Avec l’augmentation de la fréquence, l’atténuation atmosphérique augmente et la contribution de la surface au signal diminue : la différence de la valeur moyenne des T_b entre les situations océaniques et terrestres est donc réduite et les distributions dans la dernière ligne sont beaucoup moins dispersées que celles de la première ligne sur la figure III.10.

Un changement dans le signal reçu par l’instrument peut être dû à l’effet des nuages, mais peut aussi être lié à des changements dans les propriétés de la surface (température de surface de l’océan ou de la terre, vitesse du vent à la surface de l’océan, humidité du sol ou densité de végétation sur la terre). Les types de nuages peuvent être associés de préférence à certains environnements, et la modification de l’émissivité de la surface en fonction des conditions de surface rend difficile l’établissement de relations simples entre un signal et la présence de nuages. En outre, la vapeur d’eau module le signal microonde, et cet effet augmente avec les fréquences dans les canaux fenêtre jusqu’à parfois occulter tout signal provenant de couches plus basses qu’elles soient atmosphériques ou de surface.

Sur l’océan jusqu’à 100 GHz, les nuages apparaissent détectables et, dans une certaine mesure, leurs types peuvent être distingués : il y a suffisamment de contraste entre le fond océanique radiométriquement froid et le rayonnement des nuages. Au-dessus de 100 GHz, la contribution de la surface diminue radicalement. Les nuages très opaques peuvent présenter de faibles T_b (la longue queue gauche de la distribution) qui sont liés à la diffusion par la phase gelée du nuage (sur le graphique en bas à gauche).

Au-dessus de la terre, à 18 GHz, les pics les plus bas dans les histogrammes pour la plupart des types de nuages (autour de 265 K) sont probablement liés à la présence d’eau à la surface. Sinon, à 18 GHz, les histogrammes sont très similaires pour toutes les situations au-dessus des continents, ce qui signifie que cette fréquence a une sensibilité

très limitée à la présence et au type de nuage. Cela peut être considéré comme un atout pour la caractérisation de la surface terrestre avec ces fréquences, car le signal ne sera pas affecté par la présence de nuages. À haute fréquence, les nuages très opaques présentent de faibles T_b (queues de gauche des distributions), en raison de la diffusion de la glace dans les nuages (comme à 166 GHz au-dessus de l'océan). Ces nuages opaques seront probablement détectés au-dessus des terres avec ces hautes fréquences.

Les données d'entraînement, de validation et de test sont construites en utilisant les observations GMI colocalisées avec informations sur les nuages de SEVIRI. Pour couvrir toute la diversité des situations nuageuses, 72 jours sur une année complète sont sélectionnés (section 3.1). Le disque d'acquisition de SEVIRI exclut les régions de haute latitude (avec le plus d'observations de pixels couverts de neige ou de glace) et ne couvre pas non plus la totalité des continents. Cependant, il a été montré dans Aires et al. (2011a) que l'utilisation d'une classification de nuages sur le disque SEVIRI avec des observations microondes peut être extrapolée aux autres continents. Le modèle produit sera donc applicable en dehors des zones couvertes par l'instrument SEVIRI.

Dans la base de données construite, un nombre de cas égal pour chacune des classes de nuages est sélectionné. Ce procédé garantit que la classification obtenue ne sera pas biaisée en faveur des situations nuageuses les plus fréquentes négligeant celles qui le sont moins. Le même nombre de situations claires et nuageuses est donc échantillonné et chaque type de nuage est également représenté pour les cas nuageux. Chaque échantillon correspond à un ensemble d'observations. On obtient ainsi 1 million d'échantillons pour chacun des 10 types de nuages, et 10 millions d'échantillons sans nuage. Pour un modèle de classification des nuages, avec 11 classes de sortie possibles, la base de données est construite avec une répartition similaire des classes, ce qui donne environ 11 millions d'observations. Les bases de données résultantes sont ensuite divisées de manière aléatoire entre les ensembles de données d'apprentissage, de validation et de test.

Des tests supplémentaires sont effectués sur des scènes n'ayant pas servi à la création de ces jeux de données respectant ainsi le protocole nécessaire pour développer des modèles d'apprentissage automatiques robustes sur un jeu de donnée.

3.2 Choix de la méthodologie et détails d'implémentation

L'objectif est de détecter la contamination par les nuages des observations microondes au-dessus des continents, pour pouvoir en limiter l'impact sur les estimations des paramètres de surface. Comme décrit dans la première partie de ce chapitre l'effet d'un nuage sur les observations dépend des éléments qui le composent. Certains éléments prédominent dans certains types de nuages. Il n'est pas question à ce stade de déterminer le type du nuage associé à chaque observation. Il est plutôt de déterminer les classes de nuages qui affectent les observations pour les différentes plages de fréquences microondes. Les classifications de nuages obtenues dans le domaine du visible et de l'infrarouge contiennent des nuages qui ne sont pas détectés en microondes.

Comparaison de deux modèles statistiques : arbres de décision et réseaux de neurones

Le choix de la méthode statistique utilisée se base sur les expériences précédentes dans ce domaine, mais aussi sur des comparaisons de plusieurs méthodes. Pour illustrer le léger

avantage présenté par les réseaux de neurones, une comparaison entre une méthode basée sur des arbres de décisions et un réseau de neurones est présentée dans la table III.3.

Les résultats sont obtenus pour une classification binaire, à partir de tous les canaux disponibles. Les résultats sont présentés sous la forme d’une matrice de confusion, elle se lit de la façon suivante : la première colonne correspond aux classes observées dans le jeu de test (ciel nuageux ou clair) ; puis pour chacun des modèles testés, les deux colonnes suivantes présentent le pourcentage de pixels pour lesquels la prédiction a été nuageuse ou claire suivant la classe de l’observation. Les vrais positifs sont en haut à gauche (classe observée nuageuse - classe prédite nuageuse) et les vrais négatifs sont en bas à droite (observé clair - prédit clair).

observation \ prédiction	Réseau de neurones		XGBoost	
	nuageux	clair	nuageux	clair
nuageux	75.1	24.9	78.3	21.7
clair	14.5	85.5	22.5	77.5

TABLE III.3 – Comparaison de résultats de classification de nuages obtenus par un réseau de neurones puis un modèle d’arbres de décisions (XGBoost) au-dessus des continents en utilisant toutes les fréquences possibles.

Les arbres de décisions « boostés » sont une combinaison d’arbres de décision où différents critères de décision sont successivement appliqués. Le *boosting* consiste à combiner plusieurs arbres de décision pour améliorer la prédiction finale. XGBoost est une implémentation reconnue et standard de cette méthodologie utilisée dans de nombreuses applications (Chen and Guestrin, 2016). Le réseau de neurones quant à lui utilise 13 neurones en entrée, puis 34 dans une couche cachée et enfin 1 seul neurone de sortie.

Les deux méthodes comparées dans la table III.3 donnent des résultats similaires avec un léger avantage pour les réseaux de neurones. Cette table permet aussi de présenter le compromis entre la précision¹ et le rappel², c’est-à-dire la capacité du modèle à ne pas se tromper sur les cas nuageux identifié, mais aussi à identifier le plus de cas possibles.

Les sections suivantes utilisent uniquement des réseaux de neurones. Cependant, la simple comparaison effectuée ici ne permet pas de conclure sur le fait qu’une des méthodes offre systématiquement de meilleurs résultats que l’autre. Les résultats présentés confirment encore que la performance d’un modèle dépend autant de sa structure que des données qui lui sont fournies.

Il peut y avoir d’autres avantages à utiliser certaines méthodes statistiques, et les arbres de décisions permettent par exemple d’explorer les différents arbres créés et d’en déduire les canaux qui permettent le plus de discrimination entre les cas clairs et nuageux parmi les entrées. Sans surprise, les résultats obtenus empiriquement sont en accord avec la sensibilité physique décrite dans la section 1 en mettant en avant les canaux à 183 GHz en polarisation verticale, puis ceux à 89 GHz en polarisation verticale puis horizontale.

1. La précision correspond au ratio entre le nombre de vrais positifs et le total de prédictions positives

2. Le rappel est le ratio entre le nombre de vrais positifs et le total de positifs dans les données.

Détails d'implémentation du modèle choisi

La structure du réseau utilisé pour le modèle de détection des nuages a un seul neurone de sortie qui fournit un résultat binaire (0 pour la présence de nuages et 1 pour le ciel clair) en utilisant une fonction sigmoïde. Les entrées du réseau de neurones sont les canaux microondes comme illustré sur la figure III.7 de la section 2.1, leur nombre dépendant des gammes de fréquences (5, 7, ou 11). Cinq neurones (resp. 7 et 9) dans la couche cachée sont utilisés. Un nombre plus important de neurones et un réseau plus grand ont été testés, mais ils n'ont pas offert d'améliorations significatives dans la précision résultante. Pour une classification par type de nuages, 11 neurones de sortie pour chacune des classes sont utilisés. Dans ce cas la fonction d'activation de la couche de sortie est une fonction *softmax*, qui permet d'avoir une somme des prédictions entre toutes les classes égale à 1.

Les paramètres du classificateur MLP sont trouvés pendant la phase d'apprentissage où une fonction de perte binaire d'entropie croisée est minimisée avec l'algorithme de rétropropagation. Cette fonction de perte mesure pour un exemple donné l'écart entre la sortie du réseau et la présence ou absence de nuage. L'utilisation de cette fonction de perte permet d'interpréter la sortie continue du réseau de neurones comme une probabilité de classification (Bridle, 2011) dans chacune des classes et qui équivaut à la recherche du maximum de vraisemblance. Le résultat de la sortie continue du réseau peut ensuite être converti en une décision binaire à l'aide d'un seuil à définir. Dans les graphiques et résultats suivants, sauf indication contraire, un seuil de décision de 0.5 est appliqué pour obtenir la classification binaire. Pour les sorties multi-classes, la valeur la plus élevée parmi les neurones de sortie est sélectionnée comme classe prédite. Les résultats affichés montrent le pourcentage de vrais positifs (pixels nuageux correctement détectés), et de vrais négatifs (pixels clairs correctement prédits) de tous les échantillons d'un ensemble de test.

4 Résultats obtenus pour différentes surfaces

Les résultats sont présentés en deux parties, une première présente les résultats au-dessus de l'océan, où les nuages sont censés être plus faciles à détecter et à quantifier, comme vu dans les distributions de la figure III.10. Cela constitue un terrain d'essai pour la méthode, avant de l'étendre au cas terrestre, plus difficile.

4.1 Détection des nuages au-dessus des océans

Comme décrit dans la section 3.1, la base de données est créée avec une distribution égale des cas nuageux et clairs et une répartition équilibrée entre les différents types de nuages. La détection des nuages est évaluée pour les trois gammes de fréquences microondes (tous les canaux, seulement ceux en dessous de 100 GHz, seulement ceux en dessous de 40 GHz), et les résultats sont présentés dans le tableau III.4 pour l'ensemble de données de test. La détection des nuages donne de bons résultats sur l'océan, atteignant une précision d'au moins 80 %, même avec un nombre réduit de canaux. La faible émissivité de l'océan (~ 0.5) et sa relative homogénéité permettent de détecter correctement la présence de nuages, même uniquement avec de basses fréquences microondes.

Ces résultats de détection de nuages sont très encourageants. On peut donc envisager d'étudier une classification des nuages sur l'océan, avec les mêmes plages de fréquences

	Tous canaux (%)	En dessous de 100 GHz (%)	En dessous de 40 GHz (%)
Pixels clairs correctement prédits	91	89	89
Pixels nuageux correctement prédits	81	74	72

TABLE III.4 – Résultats d’une classification binaire sur l’océan pour différentes gammes de fréquences microondes.

microondes. L’ensemble de données utilisé pour l’entraînement contient toutes les classes disponibles échantillonnées de manière égale. Des réseaux de neurones similaires sont appliqués, avec 11 neurones de sortie possibles représentant les 10 classes de nuages et le cas clair, pour les trois plages de fréquences. Les matrices de confusion (Fig. III.11) affichent les résultats de la classification, montrant pour chaque classe (axe des ordonnées) le pourcentage d’échantillons prédits comme appartenant à l’une des 11 classes SEVIRI possibles (axe des abscisses). La diagonale montre le pourcentage de classification correcte pour chaque type de nuage.

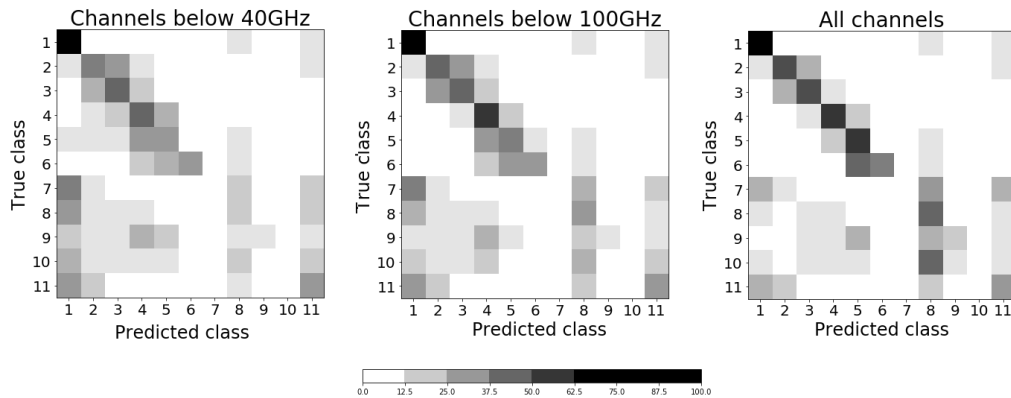


FIGURE III.11 – Matrice de confusion sur l’océan pour la classification des types de nuages pour les trois gammes de fréquences microondes : en dessous de 40 GHz (à gauche), en dessous de 100 GHz (au milieu), et tous les canaux (à droite). Les numéros des types de nuages sont détaillés dans le tableau III.1. *Pour rappel la classe 1 correspond au ciel clair, les classes 2 à 6 aux nuages opaques à des hauteurs croissantes, les classes 7 à 10 aux nuages semi-transparentes et la classe 11 aux nuages fractionnés.*

La plus grande précision est atteinte pour la détection des cas de ciel clair (classe 1) au-dessus de l’océan, pour les trois gammes de fréquences microondes. Il est parfois confondu avec les nuages semi-transparentes moyennement épais (classe 8) ou les nuages fractionnés (classe 11), car ils n’affectent pas de manière significative les T_b mesurées. Pour les nuages opaques (classes 2 à 6), les pourcentages les plus élevés se situent près de la diagonale : ces types de nuages sont correctement classés ou confondus avec des nuages à une altitude proche. On constate une augmentation de la détection des nuages très opaques (classes 4 et 5) lorsque le canal à 89 GHz est disponible. Cela s’explique par la détection accrue du contenu en glace qu’offre ce canal par rapport aux fréquences plus basses. Lorsque tous les canaux sont disponibles, la discrimination entre les couches de nuages est encore plus facile, ce qui permet une meilleure classification. Les nuages élevés semi-transparentes (classes 7, 8, 9, 10) sont parfois classés de manière incorrecte en tant que ciel clair, en particulier avec les fréquences les plus basses (en raison de la moindre sensibilité de ces canaux à ces faibles perturbations). Ils sont aussi confondus avec les nuages élevés semi-transparentes épais (classe 8) lorsque les fréquences plus élevées sont utilisées ce qui est

attendu étant donné qu'ils partagent des propriétés similaires (telle que la hauteur des nuages). Les nuages fractionnés (classe 11) ne sont pas bien classés, la classe prédite étant soit sans nuage, soit des nuages élevés semi-transparents (classe 8).

4.2 Détection des nuages au-dessus des continents

Une méthode similaire de détection des nuages est appliquée sur les continents. La classification par un réseau de neurone est construite en utilisant les trois différentes plages de fréquences microondes comme entrées et avec une sortie indiquant la probabilité d'un ciel clair ou nuageux.

Le modèle statistique et la construction de la base de données sont les mêmes que pour l'océan, et sont décrit dans les sections 2.1 et 2.2. Tout comme la table III.4 sur l'océan, le tableau III.5 (partie supérieure) présente les précisions atteintes au-dessus des continents par les trois gammes de fréquences. Comme anticipé, les performances de classification se détériorent par rapport au cas de l'océan. Néanmoins, même dans le pire des cas (avec seulement 3 canaux basses fréquences disponibles), les détections positives et négatives réelles sont proches de 70 %.

Le résultat de la détection a été analysé plus en détail, en fonction du type de nuage (partie inférieure du tableau III.5). Ces résultats ne sont que le détail des résultats précédents (partie supérieure du tableau III.5) séparés par type de nuage original. De grandes différences sont observées entre les types de nuages. Pour les nuages non semi-transparents (classes 2 à 6), plus le nuage est élevé, meilleur est le taux de détection : ceci est directement lié à la présence de glace dans les nuages élevés qui peuvent diffuser les microondes. Plus la fréquence est élevée, meilleure est la détection de la phase de glace. De même, les nuages semi-transparents hauts ne peuvent être détectés que lorsqu'ils sont suffisamment épais (classes 9 et 10).

4.3 Amélioration de la définition d'une contamination nuageuse au-dessus des continents

Les résultats précédents ont montré que les microondes ne peuvent pas détecter tous les nuages vus par les observations dans le visible/infrarouge, en particulier lorsque seules les fréquences les plus basses sont disponibles. Ce comportement rend les microondes très utiles pour les applications « tout temps » au-dessus de la surface terrestre. Cependant, certains nuages semblent être détectés correctement avec les microondes, même uniquement en utilisant des basses fréquences. Pour une caractérisation précise de la surface terrestre avec les microondes, seules les situations nuageuses qui contaminent réellement les microondes doivent être détectées. À cette fin, les résultats du modèle précédent permettent de définir de manière plus appropriée les sources de la contamination par les nuages dans les microondes : pour toutes les plages de fréquences, les nuages minces semi-transparents élevés, les nuages moyennement épais semi-transparents élevés et les nuages fractionnés (c'est-à-dire les classes 7, 8 et 11). La précision de la classification est proche de 50 %, ce qui correspond à une affectation aléatoire des classes.

Pour mieux formaliser la faible performance du classificateur sur ces classes on peut calculer le score de performance d'Heidke (Hyvärinen, 2014). Ce score quantifie l'amélioration apportée par le modèle par rapport à une assignation aléatoire pour les prédictions

Type de nuage	Tous canaux (%)	En dessous de 100 GHz (%)	En dessous de 40 GHz (%)
Pixels clairs correctement prédit	83	73	69
Pixels nuageux correctement prédit	77	73	73
2 - Nuages très bas	63	70	71
3 - Nuages bas	77	78	77
4 - Nuages moyens	92	85	83
5 - Nuages hauts et opaques	97	85	83
6 - Nuages très hauts et opaques	98	92	90
7 - Nuages hauts, semi-transparentes, fins	59	56	54
8 - Nuages hauts, semi-transparentes, moyennement épais	66	61	64
9 - Nuages hauts, semi-transparentes, épais	89	80	80
10 - Nuages hauts, semi-transparentes, au-dessus de nuages	84	74	71
11 - Nuages fractionnés	53	48	46

TABLE III.5 – Partie supérieure : pourcentage de détection correcte des nuages à partir de l’ensemble de test au-dessus des continents. Partie inférieure : détail du pourcentage de chaque type de nuage prédit comme nuageux. Les résultats sont présentés pour les trois gammes de fréquences microondes. Un seuil de 0.5 est appliqué en sortie du réseau pour calculer ces prédictions.

Type de nuage	Tous canaux	En dessous de 100 GHz	En dessous de 40 GHz	Moyenne
2	0.28	0.19	0.16	0.21
3	0.36	0.22	0.18	0.25
4	0.43	0.25	0.19	0.30
5	0.45	0.25	0.20	0.30
6	0.46	0.27	0.22	0.32
7	0.26	0.13	0.09	0.16
8	0.30	0.15	0.12	0.19
9	0.41	0.23	0.19	0.28
10	0.39	0.20	0.16	0.25
11	0.23	0.09	0.06	0.13

TABLE III.6 – Valeur du *Heidke skill score* pour les différentes classes de nuages. Les résultats sont présentés pour les trois gammes de fréquences microondes et en moyenne.

d’une classe. La table III.6 regroupe les scores obtenus pour les différents types de nuages aux différentes fréquences. On observe effectivement que les résultats pour les classes 7 et 11 sont très proches de 0 en particulier quand seule les fréquences ≤ 40 GHz sont disponibles. La classe 8 est incluse car elle est la continuité de la classe 7 et on peut donc supposer qu’une partie des cas de la classe 7 sont très similaires à ceux de la classe 8. Ces résultats signifient que les fréquences ne sont pas suffisamment affectées par ces types de nuages pour pouvoir les détecter.

Redéfinition du jeu d’entraînement

Pour étudier plus précisément les nuages qui ont un impact sur les microondes, un

ensemble de données d'entraînement plus approprié pour détecter les observations de microondes contaminées par les nuages est construit en supprimant les trois classes ambiguës mentionnées précédemment (à savoir les classes 7, 8 et 11). L'idée derrière cette nouvelle base de données d'entraînement est que la suppression des ambiguïtés au stade de l'apprentissage améliorera la classification. En effet, la suppression des types de nuages SEVIRI ambigus de la base de données d'entraînement permet au modèle d'ignorer ces phénomènes principalement détectables dans le visible/infrarouge. La sensibilité plus faible aux nuages dans les microondes est donc prise en compte dans le nouvel ensemble de données d'apprentissage. Avec un faible taux de détection, les classes ambiguës introduisent également plus de confusion lors de la classification des pixels de ciel clair. Pour éviter cette confusion et augmenter la détection correcte des pixels de ciel clair, le nouveau jeu de données d'entraînement a les mêmes propriétés que celui décrit dans la section 4.2, mais en ayant retiré les 3 types de nuages mentionnés précédemment.

Le modèle est capable de généraliser la classification aux types de nuages non contaminants ambigus (classes 7-8-11), même s'ils ne sont pas présents dans l'ensemble de données d'entraînement. Les types de nuages formant un continuum, un nuage appartenant à une classe peut être aussi proche d'un pixel nuageux dans une autre classe que d'un pixel appartenant à la même classe. Le modèle pourra donc les détecter de la même façon. L'objectif de la nouvelle classification n'est pas de détecter les pixels clairs et nuageux tels que définis par l'imagerie visible et infrarouge, mais de séparer les pixels microondes qui sont contaminés par les nuages ou les précipitations de ceux qui ne le sont pas.

La nouvelle définition des pixels détectés est donc : "probablement contaminées par un nuage ou des précipitations en microondes". Pour éviter toute confusion entre les résultats obtenus avec le nouveau jeu de données d'entraînement, les types de nuages issus de SEVIRI originaux sont appelés "Clair" ou "Nuageux" et l'indice construit spécifiquement pour les microondes donne des pixels "Non-contaminés" ou "Contaminés par des nuages ou des précipitations".

Les résultats de cette nouvelle classification sont fournis dans le tableau III.7, séparément pour les pixels clairs (classe 1), pour les pixels nuageux avec des nuages qui contaminent les microondes (les pixels microondes contaminés par les nuages, c'est-à-dire les classes 2,3,4,5,6,9,10), et pour les pixels nuageux correspondant aux 3 types de nuages difficiles à détecter avec les microondes (les pixels ambigus ignorés dans l'ensemble de données d'entraînement, c'est-à-dire les classes 7, 8, 11).

Situation prédite pour le pixel	Tous canaux (%)	En dessous de 100 GHz (%)	En dessous de 40 GHz (%)
Ciel clair	88	77	71
Contaminant-en-microondes	84	76	78
Contaminant-en-microondes pour un pixel ambigu	49	43	52

TABLE III.7 – Résultats donné en pourcentage de classification correcte pour les différentes populations claires et nuageuses, pour les trois gammes de fréquences microondes. Voir le texte pour plus de détails.

Les résultats montrent que la précision de la détection du ciel clair augmente ainsi que la précision de la détection des pixels contaminés par les nuages en microondes. Pour cette nouvelle classe on obtient 84 % de détection avec toutes les fréquences par rapport à une détection des pixels nuageux dans le tableau III.5 de 77 % (avec toutes les fréquences). Ce résultat est attendu, car les cas ambigus ont été retirés des statistiques ; il est également

cohérent avec le nombre de pixels ambigus (ignorés dans les données d'entraînement) qui sont prédits comme étant contaminés par des nuages en microondes par la nouvelle classification (près de 50% quelle que soit la gamme de fréquences). La détection des pixels nuageux tels que définis par la classification SEVIRI est réduite par rapport aux modèles précédents (voir tableau III.5) mais cela est attendu étant donné que tous les types de nuages sont utilisés dans le test mais seulement certaines classes de nuages (nuages contaminants) sont utilisées pour l'entraînement. Ce nouveau modèle a de meilleures capacités de discrimination sur la classe contaminée nouvellement définie. Certains types de nuages, qui ne contaminent pas le signal microonde, ne sont pas détectés, ce qui entraîne une diminution du nombre de pixels nuageux détectés. Ce résultat correspond à ce qui était recherché lors de la suppression de certains types de nuages SEVIRI de l'ensemble de données des pixels contaminés.

Étude du seuil de détection

La sortie originale de la classification n'est pas binaire, mais un nombre entre 0 et 1 (voir la section 2.1). Dans les résultats présentés jusqu'à présent, un seuil de décision de 0.5 a été adopté pour séparer les deux classes. Cependant, comme mentionné dans le commentaire de la table III.3 le taux de rappel et la précision sont un compromis qu'il s'agit de faire. Peut-on alors ajuster ce seuil pour une meilleure détection des observations contaminées par les nuages ?

La figure III.12 présente les sorties du réseau de neurones, pour les trois populations précédemment définies dans le tableau III.7 et pour chaque plage de fréquences microondes. Les panneaux du haut et du milieu confirment que les pixels clairs et les pixels contaminés par les nuages en microondes sont classés avec confiance, comme le montre les distributions des valeurs de la sortie du réseau très distinctes pour ces deux populations, 0 indiquant une confiance élevée dans la classe contaminée par les nuages en microondes et 1 une confiance élevée dans un pixel classé comme clair. Néanmoins, lorsque des canaux supérieurs à 100 GHz ne sont pas disponibles, une fraction non négligeable de la population de pixels clairs est classée entre 0.1 et 0.4, ce qui signifie que la confiance dans la prédiction est plus faible. Pour les types de nuages ambigus qui n'ont pas été pris en compte lors de l'apprentissage (panneau inférieur), la distribution des sorties couvre une large gamme de valeurs, ce qui traduit l'incertitude de la prédiction. Cependant, avec la gamme de fréquence complète, il y a un certain nombre d'observations étiquetées comme contaminées avec confiance (pic dans les valeurs faibles de la sortie du réseau de neurone), ceci peut être attendu en raison de la meilleure sensibilité des canaux hautes fréquences aux nuages fins. La figure III.12 montre clairement qu'en fonction du seuil de décision sélectionné pour les valeurs de sortie du réseau, il est possible de filtrer des pixels plus ou moins ambigus. Ce seuil a été fixé jusqu'à présent à 0.5, mais il pourrait être modifié. La sélection de ce seuil doit dépendre de la gamme de fréquences et de l'application finale de la détection des cas nuageux.

Par exemple, pour les estimations de la température de la surface terrestre, l'idée est d'éviter les nuages qui affectent réellement les T_b dans les basses fréquences microondes (en dessous de 40 GHz) qui sont utilisées pour l'estimation de ce paramètre (comme introduit dans la section 3.3). Ces méthodes n'excluent toutefois pas l'utilisation des fréquences plus élevées pour la détection de la contamination par les nuages, si ces fréquences sont également disponibles. De plus, l'intérêt des microondes pour l'estimation de la température de surface terrestre est de compléter les estimations infrarouges qui ne sont pas dispo-

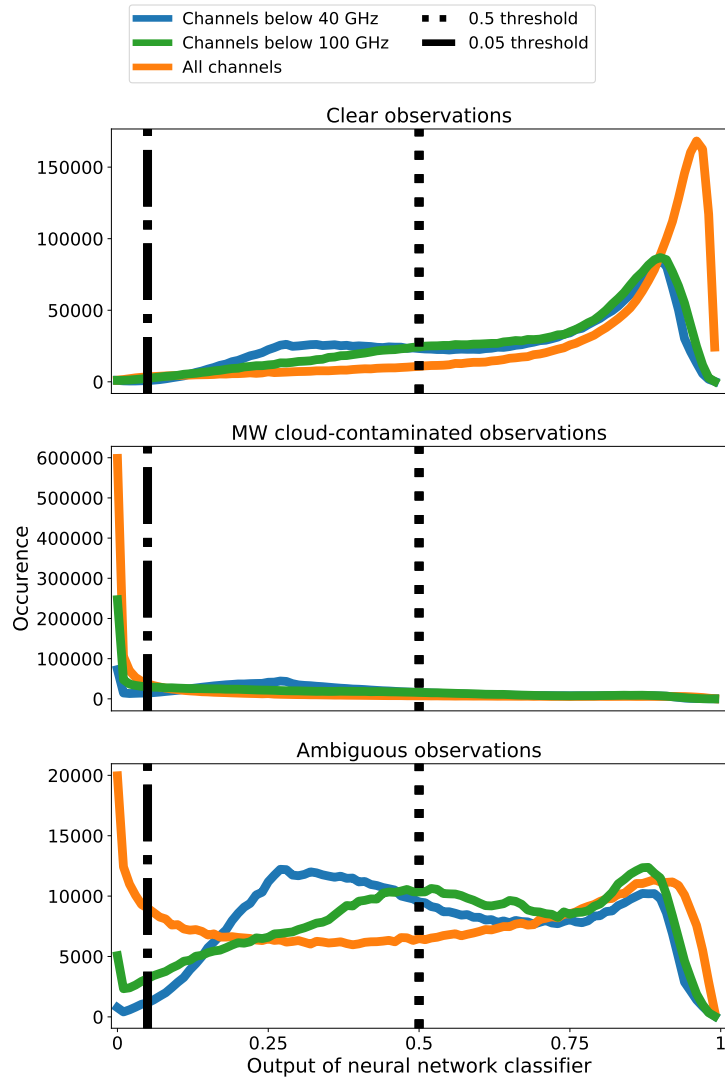


FIGURE III.12 – Distributions de probabilité des sorties du modèle pour les pixels de ciel clair (en haut), les pixels contaminés par les nuages en microondes (au milieu) et pour les pixels ambigus (en bas), pour les trois plages de fréquences de microondes. Voir le texte pour plus de détails sur les trois populations.

nibles dans des conditions nuageuses : en conséquence, seules les observations microondes sérieusement contaminées par les nuages devraient être détectées, afin de maintenir une couverture quasi « tout temps » par les estimations en microondes tout en limitant les estimations erronées dans des situations très nuageuses / pluvieuses. Dans ce cadre, le rôle de la classification des nuages est de s'assurer que les observations contaminées par les nuages sont correctement détectées. La détection correcte des vrais cas clairs est de moindre importance. La première étape consiste à déterminer quel seuil doit être appliqué pour filtrer uniquement les observations nuageuses avec certitude.

La figure III.13 présente le pourcentage d'observations microondes prédites comme étant contaminées par les nuages, en fonction du seuil sur la sortie du réseau de neurones, à la fois pour les cas microondes contaminés par les nuages (les vrais positifs, ligne pleine)

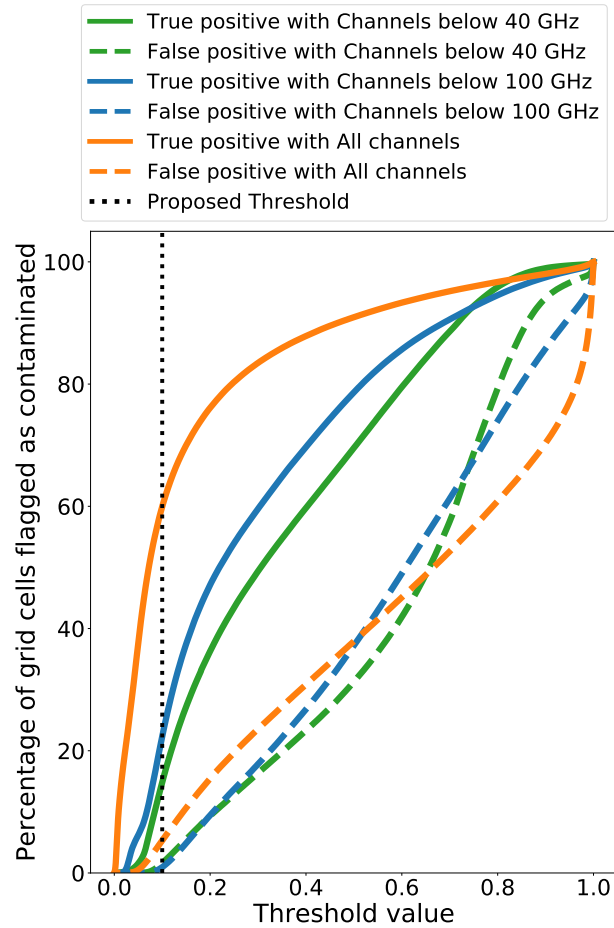


FIGURE III.13 – Évolution du pourcentage d’observations microondes correctement classées comme étant contaminées par les nuages (vrais positifs, lignes pleines), et des pixels de ciel clair incorrectement classés comme étant contaminés (faux positifs, lignes pointillées), en fonction du seuil de sortie du modèle, pour les trois gammes de fréquences microondes. Pour cet ensemble de données, la moitié des observations sont nuageuses selon SEVIRI.

et les cas de ciel clair (les faux positifs, ligne pointillée). Il montre qu’un seuil inférieur à 0.1 maintient le pourcentage de cas de ciel clair mal classés à un niveau bas (faible pourcentage de faux positifs). Combiné avec les résultats de la Figure III.12 (panneau central), un seuil à 0.05 ou 0.01 pourrait également être testé, afin de classer uniquement les observations contaminées par les nuages avec un degré de confiance élevé.

4.4 Évaluation du modèle de détection de contamination nuageuses

Pour l’instrument GMI

Une journée d’observation de GMI, le 15 juin 2015, est sélectionné pour illustrer le potentiel de la classification de la contamination des nuages en microondes. Cette journée n’a pas été utilisée dans les jeux de données d’entraînement et de test utilisés précédemment. Pour les trois gammes de fréquences microondes, la classification est appliquée avec les seuils sélectionnés (0.1, 0.05, 0.01). Le tableau III.8 fournit le pourcentage d’observations classées comme contaminées par les nuages pour chaque configuration, ainsi que les résultats des algorithmes de détection des précipitations de Ferraro (2004) basés sur un

arbre de décision et des seuils sur les canaux.

Seuil utilisé	Tous canaux (%)	En dessous de 100 GHz (%)	En dessous de 40 GHz (%)
0.1	9.7	5.1	1.5
0.05	7.9	3.6	0.8
0.01	5.2	2.0	0.3
Ferraro (2004)	–	1.4	0.4

TABLE III.8 – Pourcentage d’observations microondes classées comme contaminées par les nuages, pour les trois gammes de fréquences microondes, avec différents seuils sur la sortie du réseau de neurones. Les résultats sont présentés pour le 15 juin 2015, sur les surfaces continentales à l’intérieur du disque SEVIRI. La dernière ligne du tableau présente le pourcentage d’observations détectées comme précipitant avec la méthode de Ferraro (2004) en utilisant des canaux jusqu’à 100 GHz ou seulement en dessous de 40 GHz.

Comme prévu, lorsque les canaux à haute fréquence sont inclus, la sensibilité de notre méthodologie à la contamination par les nuages augmente, tout comme le pourcentage d’observations contaminées par les nuages. En utilisant un seuil à 0.1 et la sortie du modèle appliqué à tous les canaux environ 10 % des observations sont contaminés. Sur cette journée, environ 30 % des observations SEVIRI sont affectées par des nuages.

En utilisant uniquement les fréquences inférieures à 40 GHz, le pourcentage d’observations contaminées par les nuages diminue. Cela illustre l’intérêt d’utiliser des canaux de basses fréquences microondes pour la caractérisation de la surface terrestre « tout temps ». Le nombre d’observations contaminées en utilisant les fréquences jusqu’à 89 GHz est 4 fois plus fort que celui en utilisant uniquement des canaux ≤ 40 GHz (avec un seuil de 0.05). Pour toutes ces combinaisons de seuils/modèles, le nombre d’observations de ciel clair (selon SEVIRI) incorrectement classée comme contaminée reste inférieur à 0.5 %.

À des fins de comparaison, les algorithmes de détection de pluie de Ferraro (2004) sont également exécutés et comparés à la fois en utilisant les canaux jusqu’à 85 GHz et en se limitant aux fréquences inférieures à 40 GHz. Les résultats de la dernière ligne du tableau montrent le nombre d’observations qui sont marquées comme précipitantes. Comme attendu, le nombre de situations précipitantes est inférieur au nombre d’observations de microondes contaminées par des nuages. Pour les modèles avec des canaux au-dessus de 40 GHz, plus de 90 % des observations précipitantes sont détectées par la méthode développée. Le modèle utilisant uniquement des canaux à des fréquences inférieures à 40 GHz détecte toujours plus de 50 % des observations de précipitation lorsque le seuil de 0.1 est utilisé.

Le premier résultat apparent est que le nombre d’observations marquées comme étant contaminées est plus élevé avec tous les canaux disponibles (plus de 10 fois plus de pixels). Ceci n’est pas une surprise étant donné les résultats obtenus dans la section 4.3. Ceci est également lié au nombre attendu de pixels contaminés qui devrait diminuer lorsque les fréquences les plus sensibles ne sont pas présentes.

Avec l’estimation du nombre de pixels qui sont marqués comme étant contaminés par chaque modèle avec différents seuils, il est possible d’afficher une carte globale de l’emplacement de ces pixels contaminés. Cela permet à la fois de vérifier la cohérence spatiale de la détection effectuée et de détecter des cas systématiquement mal pris en compte par le modèle. Dans la figure III.14, les modèles sont appliqués aux données terrestres, pour créer 3 cartes différentes. Pour chaque carte, un seuil différent est appliqué, 0.1 avec les canaux

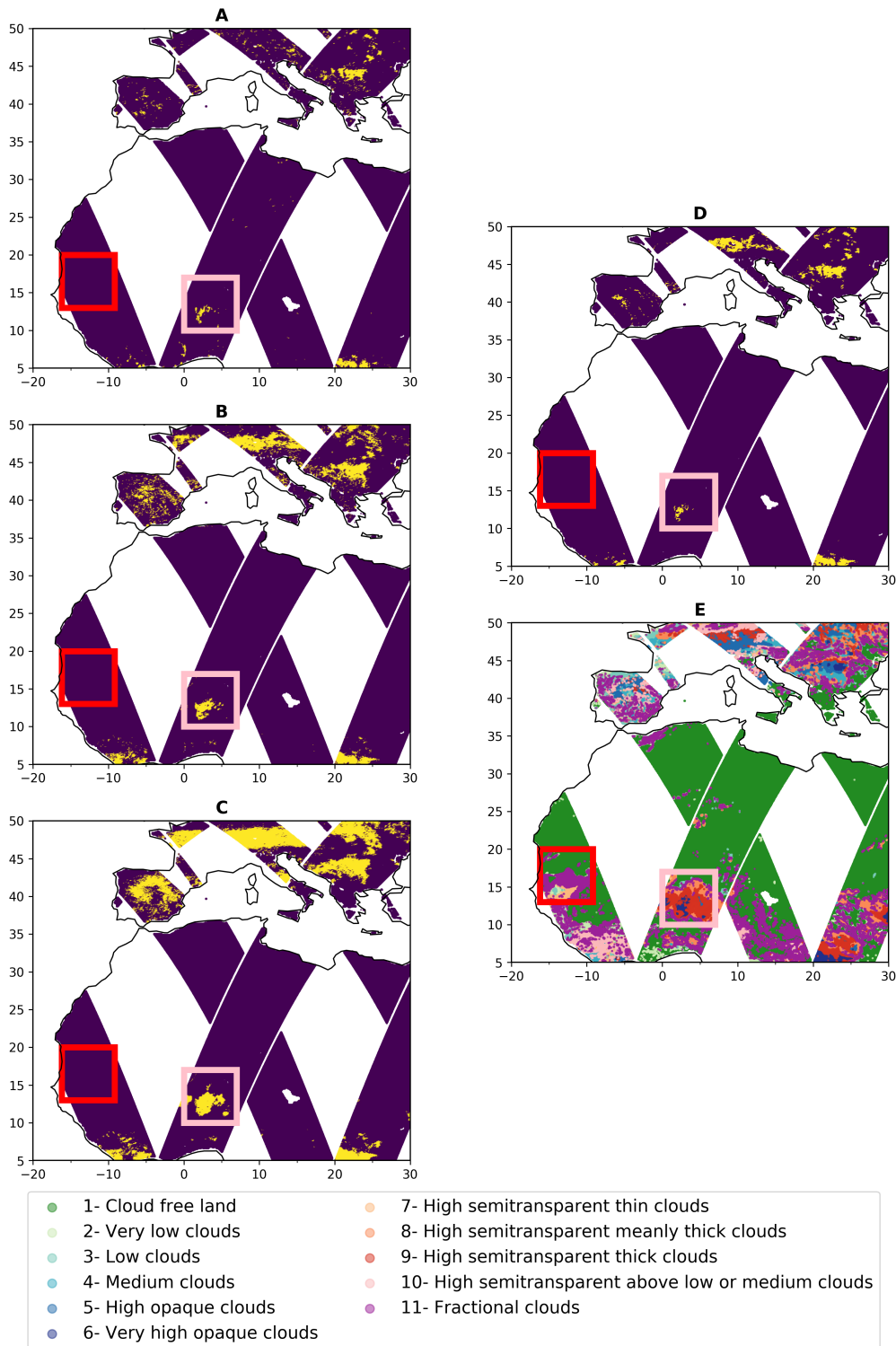


FIGURE III.14 – Cartes montrant pour le 15 juin 2015 : les pixels contaminés prédits par le modèle en utilisant des canaux inférieurs à 40 GHz avec un seuil de 0.1 (A) ; par le modèle en utilisant des canaux inférieurs à 100 GHz avec un seuil de 0.05 (B) ; par le modèle en utilisant des canaux inférieurs à 190 GHz avec un seuil de 0.01 (C) ; les pixels précipitants détectés selon Ferraro (2004) (D) ; les classes de nuages de SEVIRI (E). Les cases carrées rouges et roses mettent en évidence deux régions plus petites discutées dans la section 4.3

les plus bas (A), 0.05 avec les canaux jusqu'à 100 GHz (B) et 0.01 avec tous les canaux dis-

ponibles (C). La quatrième sous-figure (D) représente les observations précipitantes selon l'algorithme utilisant le canal à 89 GHz de Ferraro (2004). La cinquième sous-figure (E) montre le type de nuage dans la classification de SEVIRI. Plusieurs observations peuvent être faites en analysant les cartes ainsi produites :

- L'accord entre les modèles et l'augmentation du nombre de pixels signalés comme contaminés quand un nombre plus important de canaux sont utilisés est clairement visible (A B C).
- Dans certaines zones, les pixels nuageux ne semblent pas être détectés (par exemple la zone carrée rouge). En regardant le détail des types de nuages SEVIRI (sous-figure E) dans cette zone, on peut constater qu'il s'agit principalement de nuages fractionnés, semi-transparents ou de nuages bas, ce qui explique le faible taux de contamination, selon la définition établie à la section 4.3.
- Dans la zone rose, un nombre plus important de pixels sont détectés comme contaminés. En effet, les types de nuages les plus représentés sont : les nuages hauts, semi-transparents, épais (23%), les nuages hauts, semi-transparents au-dessus de nuages (20%) et les nuages très hauts et opaques (17%). Tous ces types de nuages sont ceux qui peuvent le plus affecter la mesure.
- On peut aussi constater que les observations de pixels précipitant se trouvent correctement dans les pixels nuageux détectés avec toujours un plus grand nombre de pixels nuageux que de pixels précipitant.

Cette application globale des modèles montre leur utilisation possible suivant les gammes de fréquences disponibles pour détecter les observations contaminées. Bien que l'ajout d'informations supplémentaires en utilisant les canaux plus sensibles au contenu en glace conduise à une meilleure détection de la contamination par les nuages, les résultats montrent qu'il est possible de filtrer les mesures contaminées par les nuages même au-dessus des terres avec un nombre restreint de canaux. Les seuils utilisés ici sont adaptés pour l'application spécifique de cette thèse, avec un faible nombre de pixels en ciel clair mal classés, et aussi avec l'occurrence réelle de phénomènes convectifs profonds qui contaminent le plus les observations. En effet, les données du projet international de climatologie des nuages par satellite (ISCCP) montrent une occurrence moyenne de 2.6 % pour la convection profonde (Rossow and Schiffer, 1999) qui est du même ordre de grandeur que la fraction de nuages contaminants détectés en utilisant uniquement les microondes basses fréquences (≤ 40 GHz).

Pour l'instrument AMSR-E Les différentes versions du modèle présentés ont été développées en utilisant une définition d'un nuage dérivé d'une typologie développée pour l'instrument SEVIRI mais la visée de ce travail est de fournir un outil utilisable pour les différents imageurs microondes, avec des fréquences entre 10 et 200 GHz.

En utilisant les données du satellite AMSR-E (décrit dans la section 1.2) qui dispose de canaux à des fréquences inférieures à 90 GHz on peut appliquer le modèle développé pour en évaluer la robustesse et l'utilité pour compléter les méthodes existantes de détection d'observation contaminées par des nuages. L'indicateur d'une observation contaminée est ajouté dans le traitement des données par Jiménez et al. (2017) et correspond à une valeur d'observation supérieure ou inférieure à 3 déviations standard par rapport à la valeur de référence pour ce pixel. Ce filtre est appliqué avec succès pour éliminer les cas les plus visibles de contamination dans les observations microondes. On applique le modèle utilisant les fréquences jusqu'à 100 GHz.

indicateurs existants	Sortie du modèle	En dessous de 0.5 (%)	Au-dessus de 0.5 (%)
	Nuageux		88
Ciel clair		58	42

TABLE III.9 – Comparaison de résultats de la classification par pixel pour 3 journées de données à une échelle globale entre les cas nuageux et clair obtenus par un modèle MLP pour des valeurs d’indicateurs existants pour l’instrument AMSR-E. Le pourcentage de pixels classés avec un seuil à 0.5 dans chacune des catégories est donné.

On obtient les résultats présentés dans la table III.9 qui donnent le pourcentage de pixels détectés comme étant contaminés avec un seuil à 0.5 par le modèle MLP utilisant les canaux jusqu’à 100 GHz. Le nombre de pixels utilisés est d’environ 4 millions, répartis équitablement entre les observations de nuit et de jour, pour les 14, 15 et 16 juillet 2015. La comparaison de la sortie du modèle avec le filtre existant signalant les cas nuageux suspicieux donne des résultats comparables. Près de 90 % des points déjà signalés sont aussi marqués par le modèle et de nombreuses autres situations sont détectées.

Le modèle peut donc aussi être appliqué à des observations provenant d’autres instruments microondes que GMI. Cela est permis en partie par l’utilisation de GMI lors de l’apprentissage qui contient des observations avec des heures locales de passage variables et par les capacités de généralisation du modèle entraîné.

5 Améliorer les modèles : utilisation des réseaux de convolutions

Étant donné que les événements nuageux ne sont que très rarement circonscrits à un seul pixel, mais s’étendent parfois sur une grande région, il peut être intéressant d’ajouter les informations des pixels voisins aux informations disponibles en entrée du modèle. Plusieurs approches peuvent être envisagées pour introduire cette information spatiale :

- Il a déjà été mentionné la possibilité d’utiliser une détection de valeur aberrante par rapport aux plus proches voisins par exemple dans Sun and Chawla (2004), mais qui se heurte sur les continents à l’hétérogénéité spatiale des surfaces qui peut être très forte.
- La possibilité d’inclure des informations décrivant les pixels voisins comme des informations additionnelles en entrée des modèles de classification a aussi été testée par exemple pour des méthodes non supervisées de classification de nuages dans le domaine visible (Beaudoin et al., 2001).
- Plus récemment, le développement des méthodes de traitement d’images basées sur des réseaux de neurones a permis à de nouvelles techniques d’émerger pour la détection des nuages (Ye et al., 2017; Segal-Rozenhaimer et al., 2020). Ces outils ayant démontré leur intérêt pour l’amélioration des performances des algorithmes de traitement d’image, ils sont de plus en plus appliqués dans le domaine de la télédétection y compris pour la détection de nuages (Zhu et al., 2017; Kemker et al., 2017).

On peut donc envisager de traiter le problème de détection de contaminations nuageuses comme un problème de segmentation d’image, c’est-à-dire d’attribuer à chaque pixel de l’image une classe nuageuse (ou contaminée) ou ciel clair.

5.1 Les réseaux de convolutions

Les développements des méthodes de calculs et la disponibilité de données massives ont permis d'améliorer les méthodes basées sur les réseaux de neurones. Les principes sont les mêmes que ceux décrits dans la section 2.1, où des données d'apprentissage sont utilisées de manière répétée pour optimiser les poids d'un réseau, mais le nombre de paramètres est augmenté et l'organisation des couches modifiées (LeCun et al., 1998; Krizhevsky et al., 2012). A la différence des MLP, dans les réseaux de convolution (CNN), tous les neurones ne sont pas connectés entre des couches cachées successives pour un pixel donné : seuls ses plus proches voisins sont utilisés dans la somme pondérée du neurone de la couche suivante. En convoluant (un produit matriciel) des filtres de poids et une partie de l'entrée, seuls les pixels proches spatialement en entrée sont connectés. Les mêmes poids sont donc réutilisés pour différentes parties de l'image en entrée qui permet une économie du nombre de paramètres à entraîner.

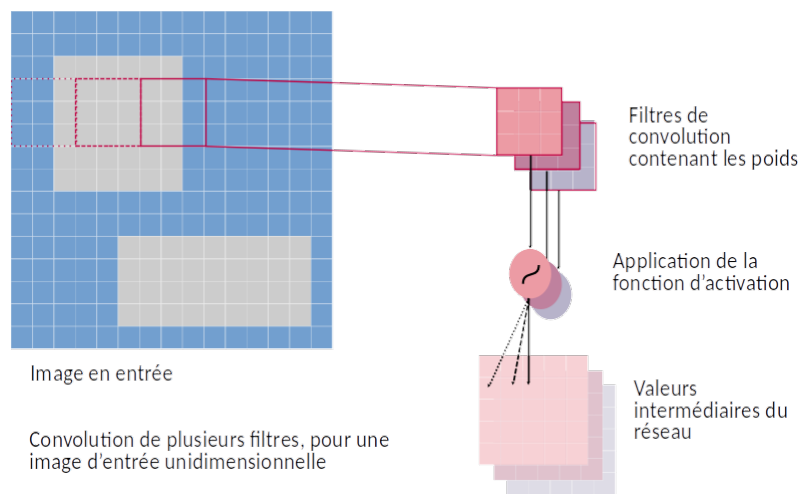


FIGURE III.15 – Schéma explicatif du fonctionnement simplifié de la convolution d'un filtre de poids pour une image en entrée.

Le schéma sur la figure III.15 présente le fonctionnement de la convolution de trois filtres de poids sur une image en entrée. La fonction d'activation est appliquée à la somme pondérée. Les filtres sont utilisés avec un pas de trois entre chaque répétition ce qui réduit la taille entre l'image d'origine et la carte des caractéristiques intermédiaires d'un facteur 3. La combinaison de filtres de convolution et de l'application des fonctions d'activation non-linéaires forme le bloc de base d'un CNN. On voit bien que cette architecture accentue les effets de proximité entre pixels, tout en gardant la possibilité d'utiliser des relations à plus grande distance en multipliant les couches. Avec l'ajout de fonction de multiplication, ces combinaisons de filtres permettent de détecter des caractéristiques de l'image d'entrée quelle que soit la position de cette caractéristique. De la même façon que pour les MLP, des hyperparamètres doivent être définis pour un réseau : le nombre de couches, la taille et le nombre de filtres, les fonctions d'activation, le pas d'application des filtres...

Le modèle U-Net (Ronneberger et al., 2015), originalement développé pour la segmentation d'images biomédicale, propose une architecture où des couches de convolution puis de dé-convolution sont appliquées successivement pour créer une prédiction en sortie de la même dimension que le problème en entrée. Cette architecture a été appliquée avec succès à des problèmes de télédétection (Yao et al., 2018). La dé-convolution est l'opération inverse de la convolution : elle ajoute des pixels en interpolant les valeurs de pixels voisins en fonction de filtres de poids. Ces opérations sont répétées jusqu'à l'obtention d'une carte

de prédiction de la taille de l'image donnée en entrée. Des connections entre des couches non-consécutives sont ajoutées qui permettent la création de la carte de segmentation en sortie. L'entraînement de ce type de réseau se fait d'une façon tout à fait similaire à celle décrite plus haut avec la rétropropagation du gradient calculé pour une fonction de coût, et avec en entrée des images échantillonnées dans l'ensemble des données d'apprentissage et en sortie des cartes des pixels nuageux. La taille des données manipulées peut imposer d'utiliser des capacités de calcul plus importantes.

5.2 Application

De la même façon que dans la section précédente les données sont sélectionnées à partir des mesures GMI reprojctées sur la grille SEVIRI au-dessus des continents en utilisant des images d'une taille de (128×128) pixels (avec au moins 80 % des pixels ayant des valeurs définies). La taille de la fenêtre est choisie pour être inférieure à la largeur d'une fauchée de GMI (~ 850 km) pour éviter que de trop nombreuses parties de l'image soient non définies, mais tout en ayant une couverture spatiale la plus large possible. Les tailles de fenêtres choisies sont souvent des puissances de deux car cela facilite le calcul des tailles des couches successives après chaque convolution. Les données sont réparties entre des bases d'entraînement, de validation et de test et sont utilisées pour entraîner un modèle de réseau de convolution U-Net avec trois couches de convolution suivies de trois couches de dé-convolution. Les données incluent les pixels ambigus pour lesquels la détection d'une contamination nuageuse pose problèmes.

On obtient les résultats suivants (table III.10) :

observation \ prédiction	MLP		U-Net	
	nuageux	clair	nuageux	clair
nuageux	77	23	82	18
clair	17	83	12	88

TABLE III.10 – Comparaison de résultats de classification de nuages obtenus par un modèle MLP puis un modèle U-Net au-dessus des continents. Le pourcentage de pixels classés dans chaque classe est donné.

Les résultats présentés dans la table III.10 sont du même type que dans la table III.5 pour le MLP. On note l'amélioration apparente des performances du modèle plus complexe basé sur un réseau de convolution. Cette amélioration est très nette surtout au vu du peu d'optimisation effectuée sur les hyperparamètres contrairement aux modèles de la section précédente. La séparation entre les cas nuageux et clairs est bien meilleure, mais il est intéressant d'étudier la répartition de cette amélioration pour les différents types de nuages au-dessus des surfaces continentales.

La figure III.16 représente les résultats de la classification par différents modèles pour plusieurs orbites de GMI sur la journée du 16 juin 2019 qui n'a pas été utilisée lors de l'entraînement. Les prédictions sont faites pixel par pixel pour les modèles MLP, et sont effectuées sur des fenêtres de (128×128) pixels pour le modèle U-Net. Pour couvrir l'ensemble de la surface, la fenêtre de 128×128 est appliquée de manière glissante sur les orbites et les prédictions sont moyennées.

Il est aussi intéressant de constater que le modèle U-Net donne des résultats qui ap-

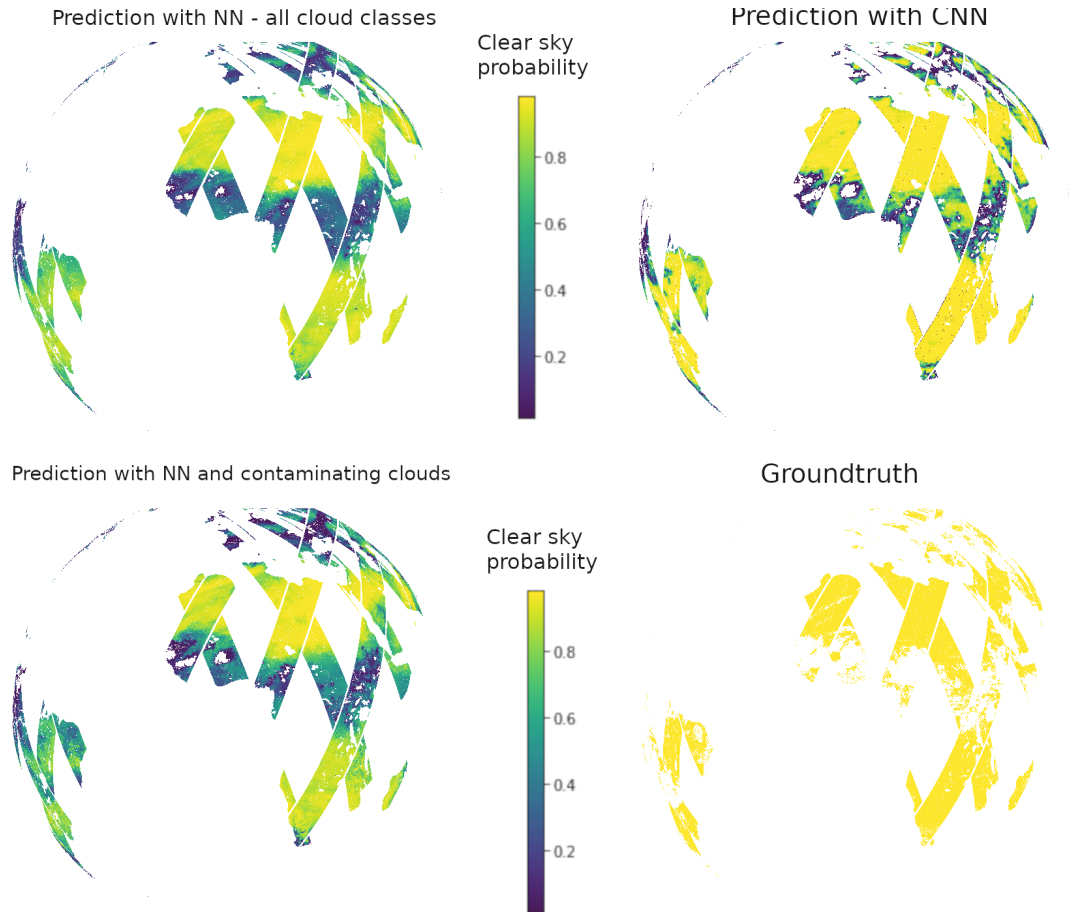


FIGURE III.16 – Comparaison de la prédiction effectuée par différents modèles utilisant les fréquences entre 18 et 183 GHz de GMI pour les orbites du 16 juin 2015 colocalisées avec le type de nuages de SEVIRI. Les cartes affichent une probabilité (entre 0 et 1) d’avoir un ciel clair pour un pixel donné. Dans les cartes de gauche, en haut la prédiction avec un MLP entraîné sur toutes les classes, et en bas en utilisant uniquement les classes contaminantes (détails dans la section 4.2). En haut à droite la carte des prédictions par le réseau de convolution, avec les prédictions moyennées pour des fenêtres de 128×128 moyennées. Enfin en bas à droite, le masque de ciel clair fourni par les données de SEVIRI.

paraissent bien meilleurs sur des zones tropicales et aux moyennes latitudes (au-dessus de l’Europe par exemple). Plusieurs hypothèses pourraient expliquer cette amélioration, en particulier le fait que les T_b autour d’un pixel fournissent une information de contraste qui permet de mieux distinguer les cas nuageux, ou la présence plus importante de formations nuageuses continues (bandes).

On peut aussi comparer la performance des différents modèles par bande de latitude comme présenté sur la figure III.17. On constate effectivement une amélioration des résultats sur les zones tropicales et tempérées où la présence de nuages est importante. Les erreurs sur les latitudes autour de 40°S s’expliquent par le faible nombre de points sur cette zone qui limite la possibilité d’appliquer les réseaux de convolution. Des tests plus systématiques n’ont pas été conduits en raison du temps de calcul plus élevé associé à l’emploi de ce modèle.

Cette première application d’un réseau de convolution à la détection de contamination nuageuse dans les observations microondes montre que cette méthode peut fournir une amélioration par rapport aux modèles statistiques traditionnellement employés. Cette brève

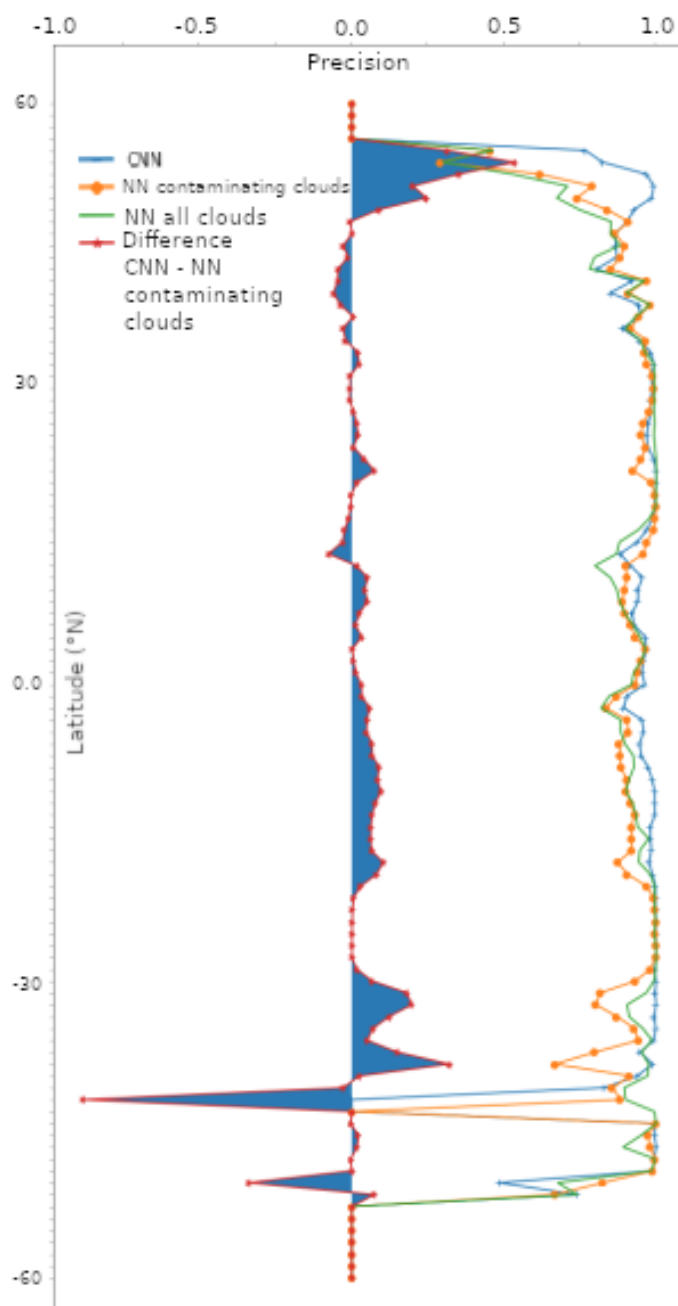


FIGURE III.17 – Précision des différents modèle appliqués aux observations avec tous les canaux pour la classification moyennée sur la latitude pour une journée de juillet 2015. En jaune le MLP entraîné uniquement sur les observations contaminantes, en vert le MLP entraîné sur toutes les classes de nuages et en bleu les résultats du CNN. La différence entre la précision du CNN et du modèle MLP entraîné sur les classes contaminantes est présentée en rouge avec une aire bleue.

étude permet aussi de mettre en évidence un certain nombre de points qui nécessitent de l'attention avant une utilisation plus générale :

- L'utilisation est plus complexe, car nécessite des données sur une fenêtre de taille fixe. Cette contrainte peut poser des difficultés d'intégration avec les chaînes de traitement existantes qui fonctionnent pour la plupart pixel par pixel. Ce type de méthode s'appliquerait peut-être plus facilement à des valeurs fournies sur des grilles

régulières, ou au niveau d'une orbite complète.

- Une seconde difficulté qui n'a pas été développée ici est le traitement de données manquantes dans les entrées. Ici la valeur moyenne des pixels de l'image est affectée à l'ensemble des pixels manquants en entrée. Ces pixels où la mesure de température de brillance ou le type de nuage ne sont pas disponibles sont masqués dans la carte de contamination créée par un modèle, mais l'introduction de ces valeurs artificielles peut avoir un impact sur le résultat de la classification qu'il faudrait mesurer.

On peut voir dans ces modèles de nouvelles pistes prometteuses pour mieux utiliser l'ensemble des données obtenues par des observations satellites. Des améliorations des hyper-paramètres sont encore à étudier et l'utilisation de données temporelles d'une façon similaire aux données spatiales pourrait être envisagée, mais dans une optique pratique d'intégration aux chaînes de traitement actuellement en service pour la création de longue série de données, il est préférable d'utiliser des modèles pixel par pixel tels que les MLP présentés précédemment.

6 Conclusion

Une méthodologie a été développée pour détecter la contamination nuageuse des observations microondes passives sur les continents (à l'exception des zones couvertes de neige et de glace). Elle est basée sur une classification par un réseau de neurones, entraînée sur des types de nuages SEVIRI colocalisés. Différentes méthodologies ont été comparées qui ont conduit à ce choix de modèle. La sortie du réseau de neurones indique la probabilité d'une contamination par un nuage du signal microondes, pour une gamme de fréquences donnée. L'indice de contamination par des nuages ou des précipitations est fourni avec des valeurs comprises entre 0 et 1 : le seuil appliqué à cet indice peut être personnalisé pour s'adapter à l'application requise et signaler les observations contaminées. Bien que l'objectif soit ici la détection des nuages au-dessus des surfaces continentales, le modèle a également été testé dans le cas plus simple de la détection au-dessus des océans. Comme attendu, la fiabilité de l'indice augmente avec le nombre de canaux disponibles et les performances sont meilleures sur l'océan. Dans tous les cas, même avec un nombre réduit d'informations sur terre, la détection des observations contaminées est effectuée dans plus de 70 % des cas.

Certaines difficultés subsistent, liées par exemple à la résolution spatiale des observations qui ne peuvent pas détecter un phénomène de petite taille. C'est le cas par exemple de certaines convections très localisées qui n'impacteront pas l'ensemble d'un pixel et ne seront donc pas correctement prises en compte. Ces cas de figures ne sont pas détectables par les instruments actuels, mais peuvent tout de même influencer l'estimation de la température moyennée à l'échelle du pixel.

Un second point mérite d'être soulevé : celui de la représentativité des situations observées, fondamentale lors de l'utilisation de méthodes statistiques. La sélection de données sur les hautes latitudes fait défaut lors de l'entraînement en raison de la couverture de l'instrument SEVIRI limitée à $\pm 60^\circ$ en latitude. Ce biais peut être compensé en utilisant d'autres sources d'observation dans le visible provenant par exemple d'instruments en orbite polaire. Le type de surface pourrait aussi être une information supplémentaire à fournir en entrée du modèle permettant d'améliorer les cas de détection limite, par exemple au-dessus des zones couvertes de glace ou de neige.

L'utilisation de données auxiliaires (comme des climatologies d'émissivité de la surface terrestre) a été envisagée et pourrait aider à caractériser les conditions de la surface. Cela faciliterait la détection des nuages, mais au prix d'une complexité accrue de l'application du modèle développé qui n'est pas souhaitable.

Le modèle de détection des contaminations nuageuses a été testé pour signaler la contamination des nuages pour des observations microondes de l'instrument GMI mais peut aussi être appliqué à différents imageurs y compris ceux ne disposant que de fréquences inférieures à 40 GHz tels SMMR (cf. section 1.2) ou CIMR (chapitre VI). Jusqu'à présent, pour la production de températures de surface à partir d'observations microondes, les indices de détection de nuages et/ou de pluie sont basés sur des seuils liés à des canaux autour de 85 GHz (Jiménez et al., 2017).

Les résultats obtenus dans cette section peuvent être appliqués à l'échelle mondiale dans les zones sans glace ni neige et sont potentiellement utiles pour de nombreuses applications où il est intéressant d'identifier les éventuelles contaminations nuageuses dans les radiances microondes observées. La versatilité de cet indice fait qu'il peut être utile pour sélectionner les scènes claires pour une estimation précise de l'émissivité des microondes (Moncet et al., 2011) mais aussi pour détecter uniquement des scènes très nuageuses pour l'analyse des convections profondes (Prigent et al., 2011). Des exemples d'applications à différents instruments sont présentés ci-dessous.

Dans la pratique selon les applications et l'instrument utilisé, si la gamme de fréquences complète jusqu'à 100 GHz est disponible, il peut être pertinent d'utiliser toutes les fréquences jusqu'à 100 GHz pour filtrer les pixels contaminés par les nuages, même si seules les fréquences inférieures à 40 GHz sont utilisées dans l'estimation. Par exemple, si la température de la surface doit être estimée avec une très faible incertitude à partir des observations SSM/I (un instrument qui possède des canaux allant jusqu'à 85 GHz), il peut être judicieux d'utiliser toute la gamme de fréquences pour détecter la contamination par les nuages, même si seules les fréquences inférieures à 40 GHz sont utilisées pour l'estimation de la température de surface.

Les différentes versions du modèle sont disponibles sur demande, elles ont déjà pu être appliquées dans le cadre de plusieurs études :

- Dans la chaîne de traitement des températures de surface dérivées de l'instrument SSM/I dans le cadre du projet *Land Surface Temperature Climate Change Initiative Essential Climate Variable* (LST CCI ECV). Un indicateur avec la probabilité de contamination est fourni en parallèle des températures estimées pour permettre aux utilisateurs finaux de sélectionner les points en fonction de la sensibilité à des contaminations qui est acceptable pour leur étude.
- Le modèle permettant la détection de nuages sur l'océan a été utilisé dans des études préliminaires sur l'instrument CIMR (Kilic et al., 2018) en particulier le modèle permettant de filtrer avec uniquement les fréquences en dessous de 40 GHz disponibles.
- Dans le travail présenté dans le chapitre IV, pour filtrer les observations nuageuses lors de la reconstruction de cycles diurnes de températures de brillances de GMI.

En résumé

- L'impact des nuages et des précipitations sur des observations microondes passives à différentes fréquences est évalué. Même si les observations qui sont effectuées sont beaucoup moins sensibles à la présence de nuages que celles dans le visible ou l'infrarouge, de possibles effets sont détectables même avec des fréquences inférieures à 40 GHz.
- Pour une estimation précise des paramètres de la surface terrestre, les observations microondes contaminées par les nuages doivent être détectées pour éviter d'interpréter la présence de nuages comme un changement des caractéristiques de la surface.
- L'effet du nuage sur la mesure de la température de brillance dépend à la fois du type de nuage et de la fréquence à laquelle l'observation est faite. Les nuages épais et haut ont une signature microonde plus importante que les nuages fins, fractionnés, ou bas.
- La comparaison de la sensibilité aux nuages pour les différentes fréquences met en évidence l'utilité des canaux aux fréquences inférieures à 40 GHz pour fournir des estimations quasi « tout temps ». Ceci justifie d'autant plus la méthode d'inversion de la température de surface choisie qui repose en priorité sur ces canaux.
- Un indice de contamination des observations par des nuages a été développé en utilisant un réseau de neurones pour pouvoir être intégré dans les chaînes de traitement de températures. Le modèle est disponible pour différentes gammes de fréquences et permet d'obtenir une détection des cas contaminés dans plus de 70 % même en utilisant uniquement des fréquences ≤ 40 GHz.

Chapitre IV

La pénétration des microondes dans la sub-surface des zones arides

Une des limites rencontrées par la télédétection microonde passive de la température de surface se situe dans les zones arides. Les déserts couvrent $\sim 12\%$ de la surface des terres émergées non gelées et occupent donc une place importante dans le bilan énergétique global. Pour ce type de surface, la couche du sol contribuant au signal mesuré peut dépasser la dizaine de centimètres, introduisant des biais dans l'estimation de la température par rapport aux autres sources de mesure (infrarouge et modèles). Cette pénétration plus importante dans les zones arides s'explique par les propriétés diélectriques des sols. Les résultats offrent aussi la possibilité de développer une méthode pour rendre cohérentes les estimations de température de surface microondes avec celles issues des instruments infrarouges et des modèles de surface.

Les chapitres précédents décrivent les méthodes développées pour éviter les écueils dus aux nuages dans l'estimation de la température de la surface avec des observations microondes passives. On pourrait penser que les zones arides¹ sont mesurées avec une grande précision en raison de la faible présence de nuages et de précipitations et de la relative stabilité de ces surfaces dans le temps. Cependant, comme noté par Prigent et al. (2016), on remarque pour ces régions une baisse de la fiabilité de l'estimation de température de surface. Ce chapitre détaille pourquoi les mesures de température de surface en microondes au-dessus des zones arides peuvent présenter un écart par rapport aux autres sources et comment cet écart peut être corrigé.

Dès 1970 et les premières observations par un radiomètre observant à 37 GHz, l'instrument *Electrically Scanning Microwave Radiometer* (ESMR), Allison (1977) remarque de larges structures dans les températures de brillance mesurée au-dessus de différentes zones du Sahara. Ces structures spatiales sont liées aux différences de composition entre les zones de sable, de roches métamorphiques et de roches carbonatées. Plusieurs hypothèses expliquent ces différences de T_b , comme des températures annuelles différentes, une porosité différente, des changements d'émissivité de surface ou encore une humidité dans les sols due à la présence d'aquifères.

1. Malgré les très faibles précipitations en Antarctique qui en font une région aride, on considère uniquement les zones arides non couvertes de glace.

CHAPITRE IV. LA PÉNÉTRATION DES MICROONDES DANS LA SUB-SURFACE DES ZONES ARIDES

Des études ont par la suite été menées utilisant les nouveaux instruments disponibles avec des capteurs multi-fréquences. Choudhury et al. (1981), puis Prigent et al. (1999) ont expliqué cette différence constatée par la contribution aux températures de brillance de couches à des profondeurs différentes et qui ont donc des températures thermodynamiques différentes.

Cet effet a des conséquences directes sur les températures de surface estimées par les observations microondes. Il a été régulièrement constaté que pour une scène donnée dans une zone aride, les températures mesurées par les microondes présentaient un biais par rapport aux températures provenant d'observations infrarouges (Jiménez et al., 2017; Martins et al., 2019). Zhou et al. (2018) constatent par exemple un biais chaud entre les températures estimées par l'instrument AMSR-E pour le passage à l'équateur à 01:30 et les températures mesurées par l'instrument MODIS au-dessus des zones désertiques de Chine. Ce biais est de l'ordre de 5 °C et peut s'expliquer par la contribution à la température de brillance mesurée par l'instrument microonde de couches en profondeur plus chaudes que la surface la nuit. Cependant, des erreurs possibles dans les estimations infrarouges ne sont pas exclues.

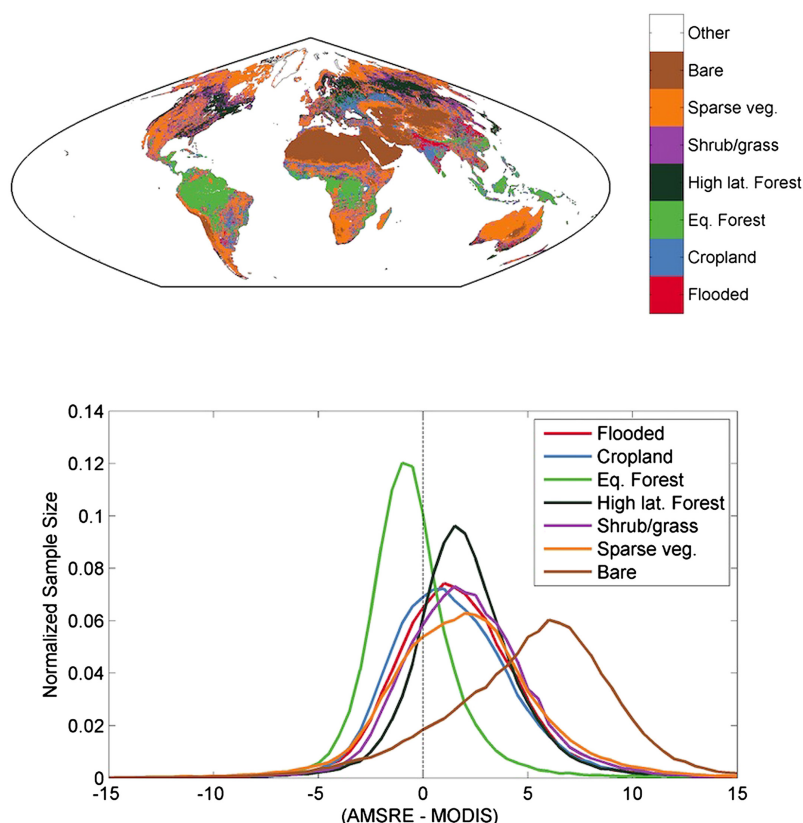


FIGURE IV.1 – Types de surface tels que définis par ESA Globcover (en haut) et la distribution des biais entre les températures provenant de AMSR-E et MODIS à 01:30 pour chacun des types de surface (en bas). Les aires sont normalisées par le nombre total d'observations dans chaque classe. (Ermida et al., 2017)

On constate bien ce biais systématique dans la figure IV.1, qui présente pour différents types de surfaces la distribution des différences de températures mesurées par l'instrument AMSR-E et l'instrument MODIS observant dans l'infrarouge pour une observation de nuit (01:30). Cette figure issue des travaux de Ermida et al. (2017) montre un biais entre les estimations de température infrarouges et microondes au-dessus des surfaces arides de l'ordre de 5 à 10 °C. Ce biais est positif, car les températures de surface estimées de nuit

par AMSR-E proviennent d'une profondeur plus importante que les températures estimées en infrarouge. A cette profondeur plus importante, la nuit, la température de la sub-surface est plus élevée que celle de la surface ce qui explique le biais observé.

L'explication repose sur deux phénomènes qui conjointement affectent les observations :

- L'extinction d'un rayonnement microonde, en particulier aux fréquences les plus basses, est plus faible dans un sol sec qu'en présence d'humidité. La couche de sol contribuant au rayonnement est donc plus large.
- Comme décrit dans le chapitre I, dans les déserts, la température dans le sol varie fortement au cours du cycle diurne. L'amplitude de cette variation change en fonction de la profondeur dans le sol.

Ces deux effets combinés font que la température mesurée par le satellite peut ne pas correspondre à la température de surface du sol telle que définie par les modèles de surface et estimée par les observations infrarouges. Pour produire des jeux de données de température à partir d'observations microondes comparables à ceux obtenus par d'autres sources d'observations, il faut donc quantifier la pénétration des microondes dans le sol pour pouvoir proposer une correction de ces erreurs constatées.

Une revue de différentes études effectuées qui ont mis en évidence ce phénomène où la contribution d'un signal ne provenant pas d'une couche micrométrique à la surface du sol est observé par les microondes est d'abord effectuée. Puis ce chapitre détaille l'interaction des microondes avec la matière qui explique la pénétration des microondes dans la sub-surface. Cela conduit à préciser la définition de la température de surface mesurée définie dans l'introduction. Ensuite une description de la variation de la température dans le sol est étudiée, et la modélisation est comparée à des données in situ et de réanalyses. Une méthode est ensuite développée pour estimer la profondeur réelle des observations effectuées, en utilisant des observations microondes et la modélisation de la température du sol. Ces résultats peuvent ensuite être comparés à des propriétés géologiques existantes et utilisés pour modéliser les changements des propriétés diélectriques.

1 Interaction du rayonnement microonde avec la sub-surface en milieu aride

On développe ici les interactions du rayonnement électromagnétique avec un milieu solide, d'abord une explication théorique puis les modèles proposés.

Pour expliquer la pénétration des microondes dans le sol, il convient de décrire la propagation d'une onde électromagnétique dans un milieu (ici le sol) que l'on suppose homogène, isotrope et non-conducteur. Ce milieu est composé de matériaux diélectriques qui présentent des propriétés électro-magnétiques qui conditionnent la propagation d'un rayonnement dans le milieu considéré. On ne considère pas ici les effets du champ magnétique en supposant que tous les matériaux rencontrés ne sont pas magnétiques.

1.1 Permittivité diélectrique

La **permittivité (constante) diélectrique** d'un milieu caractérise les propriétés du matériau dans un champ électromagnétique. Celle-ci exprimée en Farad par mètre ($F \cdot m^{-1}$) se note avec la valeur complexe :

$$\epsilon = \epsilon' - j\epsilon''.$$

Elle dépend de la composition, mais aussi de la température et de la structure (porosité) de la matière. Le paramètre ϵ' contrôle la vitesse de l'onde électrique dans le milieu, et par là l'indice de réfraction à l'interface entre deux milieux. La partie imaginaire ϵ'' de la permittivité décrit quant à elle les pertes par absorption dans le milieu qui dissipe l'énergie du rayonnement sous forme de chaleur. On définit le facteur de perte comme le rapport entre la partie imaginaire et la partie réelle de la permittivité :

$$\tan\delta = \frac{\epsilon''}{\epsilon'}. \quad (\text{IV.1})$$

Matériau	ϵ'	ϵ''	$\tan\delta$	Référence
Eau liquide (20 GHz)	35	35	1	Ulaby et al. (2014)
Sable (20 GHz)	2.6	0.03	0.011	Njoku and Kong (1977)
Sable (10 GHz)	2.6	0.01	0.004	Mätzler (1998)
Sable (10 GHz)	4	0.08	0.02	Grody and Weng (2008)
Rhyolite (35 GHz)	3.41	0.024	0.007	Campbell and Ulrichs (1969)
Basaltes (35 GHz)	7.36	0.41	0.06	Campbell and Ulrichs (1969)
Carbonates (35 GHz)	7.76	0.24	0.03	Thomas (2004)
Vide / Air	1	0	0	-

TABLE IV.1 – Quelques valeurs de permittivités diélectriques données pour des matériaux différents : de l'eau liquide, du sable, de l'air et quelques minéraux fréquemment rencontrés dans les zones arides. Les valeurs sont données pour des températures standard à la surface de la Terre (~ 20 °C).

La table IV.1 donne quelques valeurs fournies dans la littérature pour les propriétés diélectriques de matériaux. Une valeur de perte nulle indique un milieu transparent au rayonnement, tandis qu'une valeur importante signifiera qu'une plus grande partie du rayonnement est dissipée sous forme de chaleur dans le matériau considéré. On constate la valeur importante de la perte pour l'eau, qui explique l'utilisation des microondes pour le chauffage d'aliments, et aussi le fait que la présence d'eau inhibe toute pénétration dans les sols. Des dispositifs existent pour mesurer ces valeurs pour des échantillons types, mais les études menées soulignent la complexité d'utiliser ces résultats obtenus en condition de laboratoire sur des échantillons sélectionnés à des échelles plus larges (Zheng et al., 2005). Des études sur des échantillons de sable ou de roches rencontrés dans les zones arides ont été menées (Thomas, 2004) dont les valeurs sont reportées dans la table IV.1.

1.2 Profondeur de pénétration et température effective

Le chapitre I a déjà présenté la notion de profondeur de pénétration. Cette section présente plus en détail les notations et les relations entre les différentes variables qui interviennent dans la profondeur de pénétration. Pour déterminer la profondeur de laquelle un signal microonde peut provenir à travers le sol, il faut tout d'abord mesurer l'atténuation de cette onde.

Pour la propagation d'une onde électromagnétique que l'on suppose uniforme et plane, dans un milieu diélectrique avec perte on peut écrire l'amplitude du champ électrique dans un plan (noté x) variant suivant la profondeur z de la façon suivante (Ulaby et al. (2014) équation 2.63) :

$$\|E_x(z)\| = \|E_{x0}\|e^{-\mu z},$$

avec $\|E_{x0}\|$ l'amplitude à l'origine du champ et μ la constante d'atténuation de la décroissance exponentielle dans le milieu propageant cette onde. La constante d'atténuation μ s'écrit :

$$\mu = \frac{2\pi}{\lambda_0} \sqrt{\frac{\epsilon'}{2} \left[\sqrt{1 + \left(\frac{\epsilon''}{\epsilon'}\right)^2} - 1 \right]} \quad (\text{IV.2})$$

Alors à une distance $\delta_s = \frac{1}{\mu}$ l'amplitude est égale à $E_{x0}e^{-1}$ (diminution de 63%), c'est ce qu'on appelle la **profondeur de peau** (δ_s). La densité d'énergie électrostatique est égale au carré de l'amplitude de l'onde électrique. On a donc dans un milieu sans pertes dues à la diffusion, une décroissance de l'intensité qui est le double de celle de l'amplitude. On définit alors la **profondeur de pénétration** (δ_p) comme la profondeur pour laquelle l'intensité est diminuée d'un facteur e^{-1} , qui vaut donc :

$$\delta_p \approx \frac{\delta_s}{2} = \frac{1}{\kappa_a}. \quad (\text{IV.3})$$

On ignore dans ce cas l'effet de la diffusion volumique des particules dans le sol (κ_s) que l'on considère négligeable devant l'atténuation causée par le matériau.

Pour un facteur de perte faible (e.g. $\epsilon'' \ll \epsilon'$) en utilisant $\lambda_0 = \frac{c}{\nu}$ la longueur d'onde dans le vide, la profondeur de pénétration s'écrit par un développement limité d'ordre 1 :

$$\delta_p = \frac{\lambda_0 \sqrt{\epsilon'}}{2\pi \epsilon''}. \quad (\text{IV.4})$$

On suppose ici que le sol est un milieu homogène stratifié, le rayonnement émis par chaque couche est en partie absorbé par les couches supérieures. Chacune des couches peut avoir une température différente. On peut alors définir la température effective (Choudhury et al., 1981) :

$$T_{eff} = \int_0^\infty T(x) \mu(x) \exp \left[- \int_0^x \mu(x') dx' \right] dx, \quad (\text{IV.5})$$

avec T la température thermodynamique et x la profondeur d'une couche ($x=0$ à la surface). μ est le coefficient d'atténuation du rayonnement. Cette valeur de T_{eff} est aussi la température moyenne des couches contribuant au rayonnement ou la valeur d'une couche dont la température et l'émissivité produisent la même température de brillance que la superposition des contributions de toutes les couches du sol.

La valeur de la profondeur de pénétration, parfois appelée profondeur de sensibilité thermique (Wilheit, 1978), dépend donc des propriétés diélectriques et de la fréquence. D'un matériau à l'autre des différences importantes de permittivité diélectrique, qui permet d'estimer la profondeur de pénétration des microondes, peuvent être mesurées. Cela explique que la pénétration dans les sols humides est beaucoup plus faible que dans les sols secs, et que la pénétration décroît avec les fréquences croissantes.

Il est difficile d'effectuer des mesures systématiques des compositions des sols des zones arides pour en déterminer les propriétés diélectriques à toutes les fréquences. La section suivante présente deux modèles existants pour estimer la permittivité diélectrique des sols.

1.3 Modèles de permittivité diélectrique pour les sols

On a vu dans la section précédente quelques valeurs de la permittivité diélectrique mesurées pour des matériaux connus. On a aussi défini la profondeur correspondant à la profondeur pour laquelle l'intensité du signal mesuré est diminuée de 63%.

Modèle pour les sols humides

Les sols sont le plus souvent caractérisés par plusieurs paramètres : l'humidité, la densité du sol et la distribution des tailles de particules. Un modèle de la permittivité des sols a été proposé par Dobson et al. (1985) qui prend en compte la granularité du sol, décrite comme la part de sable, de limon et d'argile (tailles de grains différents), la densité, la température ainsi que la quantité d'eau dans le sol. Les particules de sables ont le diamètre le plus large ($\geq 50 \mu m$) et celles d'argile le plus petit. La modélisation de Dobson et al. (1985) prend en compte à la fois l'eau liée aux grains et l'eau libre entre les grains. On peut donc utiliser cette paramétrisation pour simuler simplement les ordres de grandeurs de profondeur de pénétration qui peuvent être attendus.

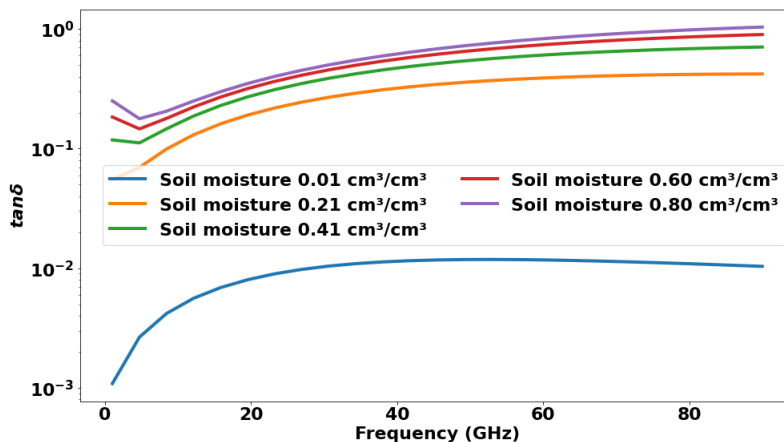


FIGURE IV.2 – Valeurs de la perte à différentes fréquences pour un sol contenant 50% de sable et d'argile à une température de 30°C avec une humidité variable des sols.

La figure IV.2 donne une idée du facteur de perte (eq. IV.1) à différentes fréquences en microondes pour plusieurs valeurs d'humidité du sol tel que modélisée par Dobson et al. (1985). On suppose que le sol est composé à 50% d'argile et de sable, à une température de 30°C . On retrouve bien l'effet important de la présence d'eau qui rend le facteur de perte proche de 1, ainsi que son augmentation pour des fréquences croissantes. Ceci explique la sensibilité variable des fréquences microondes à la sub-surface en fonction de l'humidité du sol et de la fréquence.

Modèle pour les sols secs

Pour les sols secs, la modélisation précédente n'est pas adaptée, car elle modélise surtout l'effet de la présence d'eau dans un milieu avec une granularité variable. Des études sur des échantillons de sols et de roches ont été menées pour tenter de modéliser la variation de la permittivité diélectrique en fonction de la fréquence (Blinn et al., 1972; Ulaby et al., 1990; Mätzler, 1998).

La partie réelle de la constante diélectrique est considérée constante pour les plages de fréquences microondes. En notant ρ la densité du sol, la paramétrisation d'Ulaby et al. (1990) pour les roches s'écrit :

$$\epsilon' = 2\rho. \quad (\text{IV.6})$$

Celle de Dobson et al. (1985) pour les échantillons de sols secs s'écrit :

$$\epsilon' = (1 + 0.44\rho)^2. \quad (\text{IV.7})$$

Ces deux équations donnent des résultats proches (- de 5% d'écart) sur la plage des densités de sol courantes.

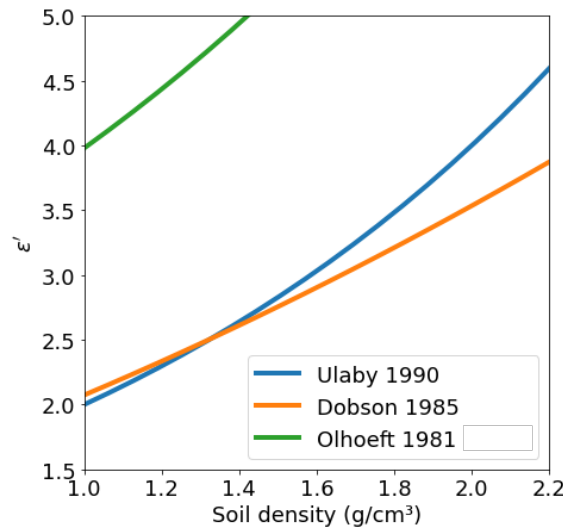


FIGURE IV.3 – Comparaison de modèles de la partie réelle de la constante diélectrique reportés dans la littérature en fonction de la densité de l'échantillon de roche ou de sol.

La figure IV.3 compare différents modèles ou valeurs utilisées. Jinkai (1990) reporte la relation déterminée par Olhoeft pour des basaltes (en vert), et les valeurs obtenues sont très différentes des autres paramétrisations. La valeur obtenue par l'équation IV.7 (en orange) et en bleu celle par l'équation IV.6 sur un plus grand nombre d'échantillons de roche. Pour la partie imaginaire de la permittivité, aucune relation avec la densité n'est obtenue pour les échantillons étudiés. Mais la dépendance en fréquence pour les sols secs dans le domaine microonde prend la forme :

$$\epsilon''(\nu) = a + \frac{b}{\nu}, \quad (\text{IV.8})$$

avec a et b des paramètres à déterminer en fonction du type de roche et ν la fréquence du rayonnement. En utilisant une valeur fixe de $\epsilon' = 2.6$, Mätzler (1998) détermine ϵ'' pour des échantillons de sable collecté dans un erg du Sahara. La valeur est reportée dans la table IV.1. Il note de manière similaire une décroissance de la partie imaginaire de la permittivité pour des valeurs de fréquences croissantes.

Les résultats précédents mettent en évidence la différence importante de comportement diélectrique des sols, en fonction de la présence ou absence d'eau. Pour les sols arides, la

composition minéralogique du sol peut influencer fortement les valeurs de permittivité diélectrique.

2 Estimation du profil de température dans le sol et de son évolution temporelle

Connaître la température effective définie dans la section précédente implique de connaître le profil de la température dans la colonne de sol. On peut mesurer cette température en effectuant des mesures à plusieurs profondeurs. On observe alors des profils de température très variables, suivant les heures de la journée et les saisons comme illustré sur la figure IV.4. Les amplitudes diurnes des changements de température à la surface peuvent atteindre plusieurs dizaines de degrés pour les sols non gelés alors que la température reste constante à partir d'une certaine profondeur (de l'ordre de quelques mètres).

Le problème de la détermination de la température d'un corps soumis à des changements de température est étudié depuis plusieurs siècles. Joseph Fourier en 1820 a proposé une équation différentielle permettant de décrire l'évolution au cours du temps de la température dans un solide. C'est à partir d'une résolution de cette équation présentée dans la section suivante que la température dans la colonne de sol pourra être déterminée.

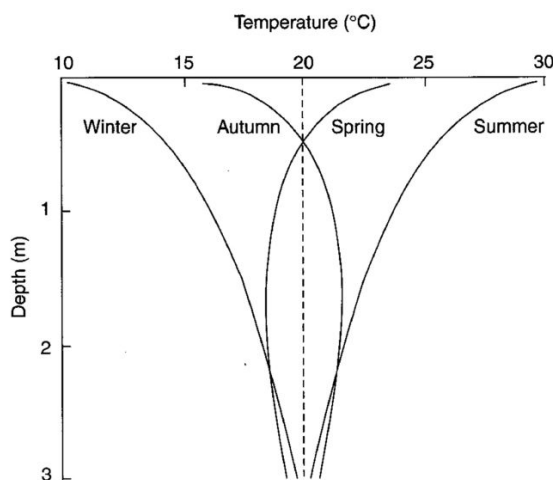


FIGURE IV.4 – Profil de température dans le sol pour différentes saisons dans une région sans gel (figure extraite de Hillel, 1998).

2.1 Résolution de l'équation de Fourier

Il est possible d'utiliser l'équation de Fourier pour déterminer l'évolution de la température dans le sol en fonction de la profondeur et du temps. Dans le cadre de cette thèse, un certain nombre d'hypothèses permettent de simplifier le problème pour faciliter le calcul d'une solution.

- ✓ De la même façon que pour l'étude des propriétés diélectriques d'un milieu, on suppose que les différentes couches du sol sont homogènes et avec une composition stable au cours du temps. On a alors un seul matériau avec une densité et des propriétés thermiques constantes à la résolution spatiale considérée.

- ✓ Dans une zone aride, les quantités d'eau dans le sol sont faibles. On peut alors ignorer le transport de chaleur occasionné par les déplacements d'eau. Les changements d'états de l'eau (gel ou évaporation) sont négligés.
- ✓ On considère aussi les échanges de chaleurs latéraux négligeables face à ceux dans la direction verticale.
- ✓ Il n'y a pas de production de chaleur interne par un phénomène exothermique dans le sol.

L'équation doit donc uniquement être résolue en une dimension, c'est-à-dire que l'on ramène le problème de l'estimation de la température dans le sol à l'étude de la température dans une colonne de sol unidimensionnelle avec des propriétés thermiques connues.

La colonne de sol a une extrémité soumise à une température variable (la surface) et l'autre extrémité à une profondeur de l'ordre de quelques mètres à une température constante. Dans ce cas, le flux de chaleur dans le milieu considéré et le gradient de la température dans ce milieu sont proportionnels. C'est l'équation de Fourier :

$$k \frac{\partial^2 T}{\partial z^2} = \rho C \frac{\partial T}{\partial t}. \quad (\text{IV.9})$$

Dans cette équation, z est la profondeur, t le temps et T la température. On note k la conductivité thermique ($W \cdot m^{-1} \cdot K^{-1}$), ρ la densité ($kg \cdot m^{-3}$) et C la capacité thermique massique ($J \cdot kg^{-1}$).

Différentes méthodes existent pour résoudre cette équation différentielle qu'elles soient numériques ou analytiques. Les paramètres thermiques du sol et les méthodes de résolution font l'objet des sections suivantes.

Paramètres thermiques des sols

Dans l'équation de Fourier trois paramètres définissent les propriétés thermiques des sols. La capacité thermique massique (C) multipliée par la densité du sol (ρ) mesure la quantité de chaleur stockée par unité de volume dans le sol : on la nomme alors capacité thermique volumique. La conductivité thermique (k) représente la vitesse de diffusion de la chaleur dans le milieu. Pour obtenir une valeur de ces paramètres à l'échelle globale, trois contraintes existent :

- La première est de connaître précisément les types de sol à l'échelle de la Terre, alors que de larges régions du monde en particulier dans les zones arides n'ont que partiellement été sondées pour qualifier le type de sol.
- Les surfaces observées à la résolution des observations microondes sont très hétérogènes, et donc constituées de plusieurs types de sols.
- Enfin, une difficulté réside dans le choix des modèles utilisés pour décrire les relations entre les constituants du sol et les paramètres thermiques.

Pour la **conductivité thermique** (k), les principaux facteurs sont la distribution des tailles de grains et surtout la quantité d'eau dans le sol. Différents modèles existent pour estimer k en fonction des caractéristiques physiques du sol. Une comparaison de ces modèles pour l'obtention des cartes globales de conductivité est effectuée par Dai et al. (2019). L'ECMWF implémente le modèle de Viterbo et al. (1999) dans le modèle physique du sol (ECMWF, 2017) (IFS) en calculant une somme pondérée entre une valeur sèche et une valeur saturée de conductivité thermique.

La **densité** (ρ) est évaluée en se basant sur des échantillonnages du sol qui sont ensuite utilisées avec des fonctions de transferts pédométriques pour obtenir des cartes globales de densité. Pour certaines parties du globe, en particulier les déserts les seules données globales sont celles issues des travaux de la *Food and Agricultural Organisation* (FAO) dans les années 1970. Cependant, certaines zones ont bénéficié de mises à jour plus récentes.

De la même façon, la **capacité thermique massique** (C) d'un milieu composé de plusieurs matériaux peut être estimée comme une somme pondérée des capacités thermiques des différents constituants. La valeur donnée repose donc sur des mesures en laboratoire de propriétés de matériaux et sur des sondages des sols.

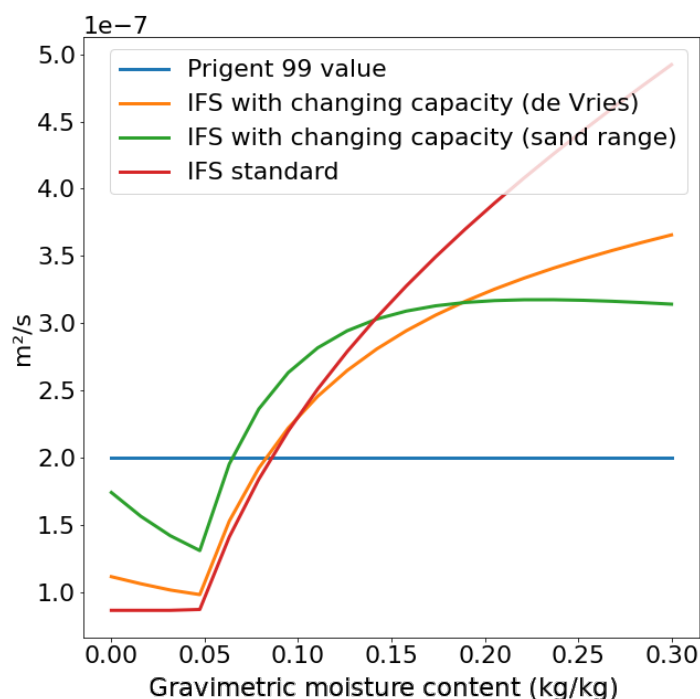


FIGURE IV.5 – Comparaison de la diffusivité thermique (α) en fonction du contenu en eau du sol pour différents modèles de l'IFS évalués par l'ECMWF. Une valeur fixe utilisée par défaut pour les zones arides dans les travaux de Prigent et al. (1999) est aussi indiquée (en bleu).

Pour simplifier la notation, un unique paramètre appelé diffusivité thermique $\alpha = \frac{k}{\rho C}$ est utilisé en lieu et place de ces 3 coefficients. La figure IV.5 illustre la variation de la diffusivité thermique (α) en fonction de la saturation en eau dans le sol. Les 3 courbes diffèrent par la valeur de la capacité thermique massique utilisée, et par les modèles de diffusivité décrits dans Peters-Lidard et al. (1998). Les courbes sont comparées à une valeur utilisée par défaut pour le désert dans l'étude de Prigent et al. (1999). Le comportement des modèles diffère très fortement pour les valeurs élevées de contenu en eau dans le sol. Dans les zones arides, les différentes implémentations dépendent fortement des valeurs sèches de capacité thermique et de conductivité. Il peut y avoir un rapport 2 entre les diffusivités estimées en l'absence d'eau. Pour résoudre au mieux l'équation de Fourier, il est important d'utiliser des valeurs les plus précises possibles de ces paramètres qui varient suivant les zones géographiques.

Les différents jeux de données existants utilisent les données de la FAO (1957) comme

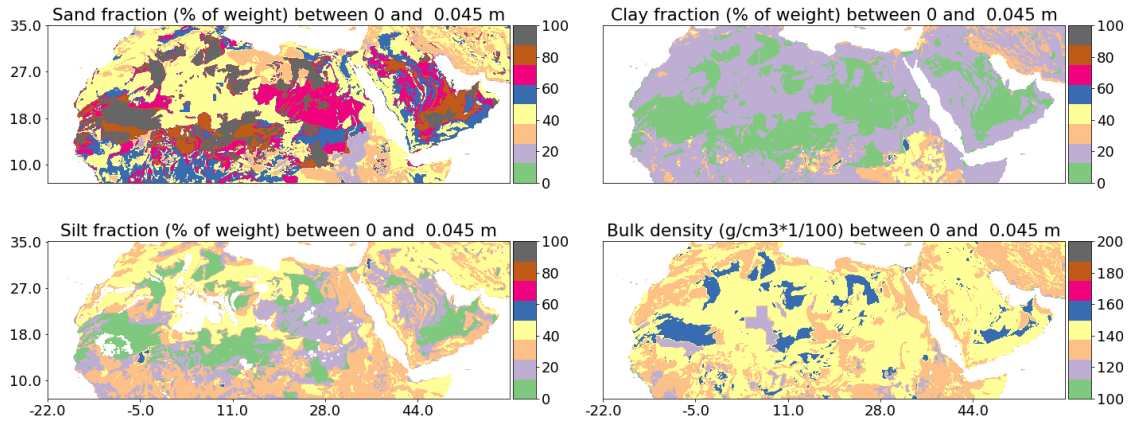


FIGURE IV.6 – Paramètres de composition du sol en Afrique du nord et dans la péninsule arabique fournies par Shangguan et al. (2014) pour la couche de surface : pourcentage de sable (haut gauche), argile (haut droite), limon (bas gauche) et la densité (bas droite).

source principale dans les zones arides et peu d'actualisations ont eu lieu depuis. Certaines cartes disponibles présentent des artefacts flagrants comme les discontinuités à la frontière Egypto-libyenne dans les données de Nachtergaele et al. (2009). On notera que dans les données utilisées à l'ECMWF, ces discontinuités sont présentes. Les cartes fournies par Shangguan et al. (2014) résolvent certains de ces problèmes et ont donc été choisies comme source de données (dont certaines composantes sont présentées dans la figure IV.6). Les données sont disponibles sur plusieurs couches dans le sol. La variabilité des paramètres en fonction de la profondeur a été évaluée pour les zones arides et celle-ci est faible pour le jeu de données fourni par Shangguan et al. (2014). Ceci légitime notre hypothèse de l'homogénéité du milieu dans la colonne de sol.

Méthodes de résolution de l'équation unidimensionnelle de la chaleur

Une fois les paramètres de l'équation correctement définis, on peut s'intéresser à sa résolution. Pour cela deux approches sont possibles. La première, numérique, effectue des successions d'opérations pour approcher le résultat à une échelle prédéfinie. La seconde suppose des hypothèses fortes sur les conditions aux limites, mais permet d'obtenir une expression analytique de la température en tout point de la colonne de sol. Le bilan radiatif en surface n'est pas calculé dans cette résolution. Seule la température en surface et l'hypothèse d'échanges effectuées uniquement par conduction sont nécessaires. L'effet des flux de chaleur sensible ou latente sur la température sont les entrées qui permettent de calculer la variation du cycle diurne en surface.

Pour les deux approches, la température en surface doit donc être fournie. Ici les données de la réanalyse ERA5-Land (Muñoz-Sabater et al. (2021), détails dans la section 3.4 du chapitre I) sont utilisées. Elles sont disponibles quotidiennement sur les continents à une échelle de $0.1^\circ \times 0.1^\circ$ et fournissent les températures de surface *skin temperature* (skt). Elles sont sélectionnées sur la zone d'analyse qui couvre l'Afrique du Nord et la péninsule arabique. Ces données fournissent une estimation de la variation de température à la surface, contrainte par les flux radiatifs du modèle physique de l'ECMWF, et aussi des températures intégrées sur plusieurs couches (0-7, 7-28, 28-100, 100-289 cm). La température dans le sol au-delà de 289 cm est considérée comme constante. Le modèle physique de l'ECMWF approche la solution de l'équation de Fourier de manière discrète sur ces 4 couches, en utilisant les paramètres détaillés ci-dessus et des flux issus de la réanalyse

ERA5 et son modèle physique global en surface. Des composantes supplémentaires existent pour prendre en compte les effets causés par les changements d'état de l'eau (ECMWF, 2017). Peu d'évaluations du profil de température fourni par ERA5-Land ont été effectuées (Cao et al., 2020) et elles ne concernent pas les zones arides. Dans la suite de ce chapitre ERA5-Land est utilisée pour l'évaluation du profil thermique du sol. Par la suite, lorsque les données atmosphériques sont requises les données ERA5 sont utilisées.

Les **méthodes numériques** se basent sur la recherche de solutions sur des grilles discrètes de l'équation IV.9. Ces méthodes par différences finies font l'objet de très nombreuses études et la méthode de Crank-Nicolson (Smith et al., 1985) est une des méthodes employées pour trouver une solution de l'équation de la chaleur. On peut en définissant les pas appropriés aux échelles de temps et d'espace considérées (cycles diurnes de 24 h et profondeur maximale des variations ≤ 3 m) et les conditions aux limites (variation de la température en surface et température fixe en profondeur) calculer la température de la colonne avec une résolution verticale suffisante pour notre application de l'ordre du centimètre. Ces méthodes numériques s'adaptent très bien à tout type de variation de température en surface, mais elles présentent l'inconvénient de requérir un temps de calcul important.

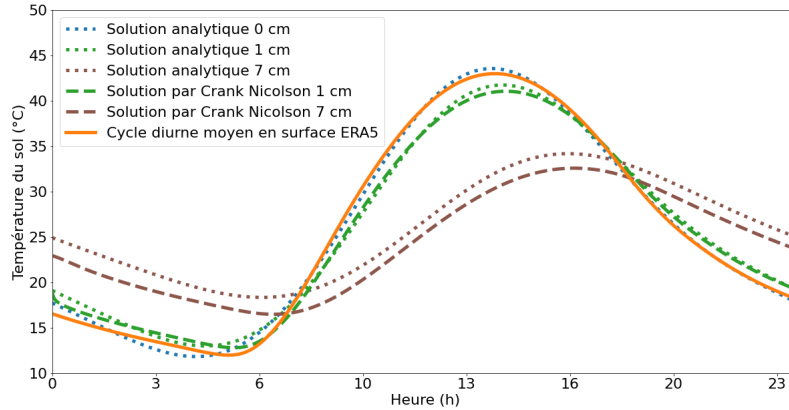


FIGURE IV.7 – Comparaison de solutions de l'équation de Fourier par deux différentes méthodes : une solution analytique (ligne pointillée), une solution numérique (trait discontinu). Les solutions sont trouvées pour une journée moyenne d'avril 2016, à une position de N 31°N 7.5°E en surface et à des profondeurs de 1 et 7 cm. Le cycle diurne moyen d'ERA5 en avril 2016 est en orange.

Les **solutions analytiques** donnent des solutions exactes pour des conditions aux limites ayant des formes connues, par exemple pour le cas d'une variation périodique de la température en surface. Cette solution pour la température du sol, proposée par Fourier est détaillée dans les travaux de Vries (1963).

En effet, pour une variation de température périodique à une extrémité de la colonne de sol (ici à la surface) de la forme :

$$T(z = 0, t) = \bar{T} + \sum_n A_n \cos(n\omega t + \phi_n). \quad (\text{IV.10})$$

En considérant une température fixe au delà d'une certaine profondeur on peut déterminer une de l'équation différentielle IV.9. Le cycle diurne de température peut être correctement approximé avec une somme de deux sinusoides. Dans ce cas, \bar{T} est la valeur moyenne sur le cycle diurne considéré, $n = 2$, $\omega = \frac{2\pi}{86400}$ (la pulsation en rad.s^{-1}) et A_n, ϕ_n

décrivent l'amplitude et le déphasage des composantes sinusoïdales. Ils sont obtenus par la méthode des moindres carrés par rapport à un cycle diurne de température connu.

A une profondeur donnée, on a donc :

$$T(z, t) = \bar{T} + \sum_n \exp(-z\sqrt{\frac{n\omega}{2\alpha}}) A_n \cos(n\omega t + \phi_n - z\sqrt{\frac{n\omega}{2\alpha}}). \quad (\text{IV.11})$$

On retrouve les composantes du cycle de surface (l'amplitude et la valeur moyenne de la température sur un cycle), avec un déphasage des cycles et une atténuation exponentielle des amplitudes des cycles de température avec la profondeur. La figure IV.7 présente un exemple de cycles modélisés à différentes profondeurs par les deux méthodes. La courbe orange correspond aux données de réanalyses d'ERA5-Land moyennes sur le mois d'avril 2016 autour d'un point dans le Sahara (31°N, 7.5°E). Ces données sont utilisées pour la résolution numérique et pour estimer les composantes de la solution analytique. Une valeur de $\alpha = 3.5 \times 10^{-7} \text{m}^2 \cdot \text{s}^{-1}$ est utilisée. On constate que les deux composantes utilisées pour la solution analytique suffisent à approximer correctement le cycle diurne (pointillés bleus). Les deux méthodes de résolution donnent des résultats proches aux différentes profondeurs et diffèrent dans les couches les plus profondes en raison des valeurs différentes utilisées pour la température à l'extrémité de la colonne de sol.

Compte tenu de ces résultats et de la plus grande facilité de la solution analytique pour obtenir la température à une profondeur quelconque cette solution est préférée par la suite.

2.2 Comparaison de la solution aux données in situ et de réanalyses

Pour évaluer la solution de l'équation de diffusion thermique proposée, il est possible de comparer les températures modélisées à la fois à des données issues de réanalyses, ou à des données mesurées par des stations in situ. Pour la plupart des comparaisons, les amplitudes des cycles deviennent très faibles au-delà de 10 cm, c'est pourquoi les comparaisons sont principalement effectuées sur les couches supérieures du sol.

Comme décrit ci-dessus la réanalyse ERA5-Land fournit une température intégrée sur une certaine épaisseur (e.g. 7 cm pour la première couche) qui est différente de la valeur ponctuelle obtenues par notre solution, pour effectuer une comparaison sensée avec la solution analytique, il faut intégrer le résultat sur la même épaisseur que les couches de la réanalyse.

Les mesures effectuées in situ sont le plus souvent faites par des capteurs de température enterrés à des profondeurs prédéterminées. Comme détaillé dans le chapitre I, peu de sites de mesures de la température de surface par des capteurs infrarouges existent, plus nombreux sont les sites étudiant l'humidité des sols pour lesquels les températures à différentes profondeurs sont aussi mesurées. Chaque capteur mesure la variation au cours du temps de la température en un point de la colonne de sol, ces séries de mesures peuvent donc être directement comparées aux résultats issus de la solution analytique. On peut effectuer une interpolation entre les températures de différentes couches pour estimer une température moyenne sur une certaine épaisseur mais ce procédé reste approximatif. Sur les figures suivantes, des comparaisons de valeurs intégrées ou ponctuelles sont faites.

Station Goabeb

La station Goabeb (21.1°S, 14.6°E) est équipée d'un système de mesure du profil de la température dans le sol, opéré par l'Université de Bâle qui nous a communiqué les données de l'année 2017. Les données de la même année pour le pixel le plus proche de la station sont aussi extraites des données ERA5-Land. Le cycle diurne de température provenant de la réanalyse sont utilisés pour déterminer les paramètres de la solution analytique. Le paramètre de diffusivité thermique α est choisi en fonction des cartes décrites ci-dessus (section 2.1).

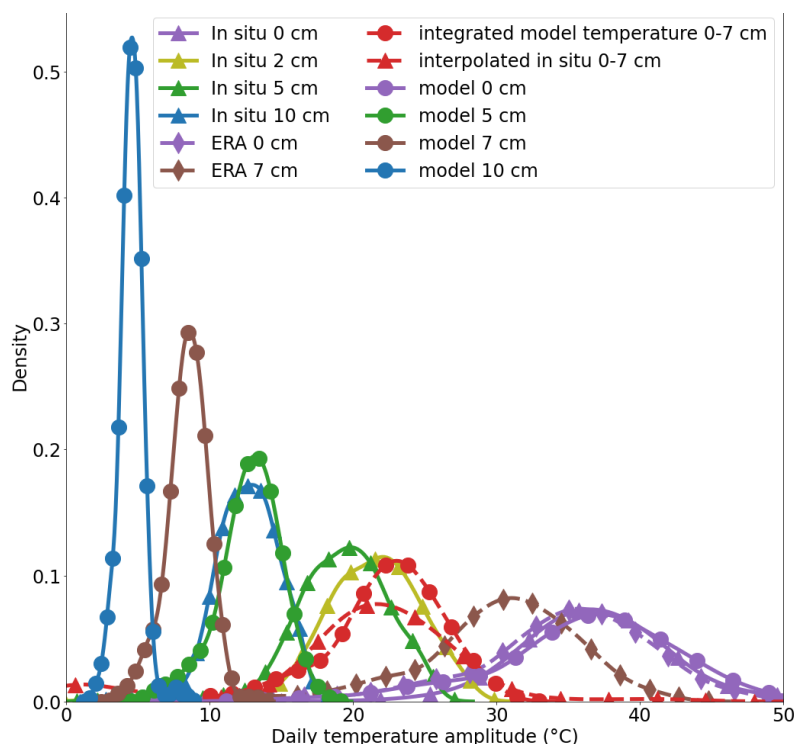


FIGURE IV.8 – Comparaison des amplitudes des cycles de température à différentes profondeurs pour la station Goabeb en 2017. Les données in situ (triangle), les données de réanalyse (diamant), les amplitudes obtenues par le modèle analytique à différentes profondeurs (rond). En rouge la valeur intégrée pour la solution analytique (trait discontinu - marqueur rond) entre 0 et 7 cm et la valeur interpolée (triangle) pour les données in situ est affichée.

Les mesures in situ sont effectuées à 0, 2, 5 et 10 cm et la distribution des amplitudes aux différentes profondeurs correspondent aux lignes continues sur la figure IV.8. On a bien des amplitudes décroissantes en fonction de la profondeur, comme attendu selon l'équation IV.9. Les distributions des amplitudes des températures pour le cycle de surface et la couche entre 0 et 7 cm de la réanalyse sont affichées en trait pointillé. La température en surface affiche une amplitude très similaire (courbes violettes) pour les différentes sources, mais différent dans les couches plus profondes. Les amplitudes des cycles semblent décroître plus vite dans les données in situ (triangle) que dans la modélisation (rond).

Plusieurs facteurs peuvent expliquer les écarts d'amplitudes entre les différents cycles diurnes de température :

- La diffusivité thermique a un effet important sur le modèle analytique. On a vu que α dépendait fortement des hypothèses d'humidité des sols. En supposant l'absence

d'humidité du sol, une sous-estimation de α est possible.

- Pour les mesures in situ, de petits changements de la profondeur réelle peuvent avoir un impact important sur la température mesurée. On peut en particulier tester la décroissance exponentielle attendue des amplitudes aux différentes profondeurs. On observe que les amplitudes à 2 cm pourraient être sous-estimées.

Les amplitudes du cycle diurne calculées dans la réanalyse sont intégrées sur des couches et donc difficilement comparables aux valeurs in situ ou issues du modèle de propagation thermique décrit ci-dessus. Pour comparer, une interpolation de la température entre 0 et 7 cm est effectuée pour la température in situ (courbe rouge discontinue marqueur triangulaire) et la solution analytique est intégrée entre les mêmes profondeurs (courbe rouge marqueur rond). Ces courbes peuvent donc être comparées à la courbe des amplitudes de cycle de la couche entre 0 et 7 cm de la réanalyse (diamant). Les amplitudes sont beaucoup plus vite atténuées dans les données de réanalyse, alors qu'elles semblent mieux concorder entre les données in situ intégrées et la solution analytique intégrée. De la même façon que pour les différences entre le modèle analytique et in situ, la variation de l'humidité du sol peut avoir une influence importante sur la diminution de l'amplitude des cycles diurnes de température et pourrait expliquer ces différences.

Il est aussi intéressant de comparer les profils de températures pour les extrema du cycle diurne, car ils permettent de vérifier que la représentation des températures par le modèle analytique est proche de celles mesurées in situ à des moments clés de la journée.

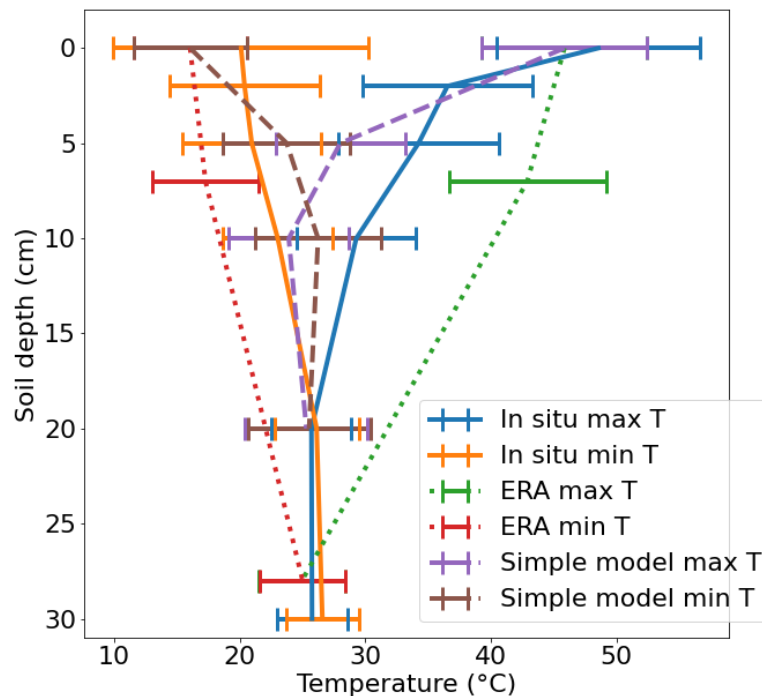


FIGURE IV.9 – Profils de température moyens du sol obtenus par différentes méthodes à l'heure des extrema du cycle de température pour la station Goabeb en 2017. En trait continu pour les capteurs in situ, en pointillé pour les données de réanalyses et en trait discontinu pour la solution analytique de l'équation de Fourier. Pour chaque valeur, une déviation standard autour de la moyenne est indiquée par le segment horizontal.

La figure IV.9 affiche le profil de température moyen dans le sol pour la station Goabeb en 2017, en trait continu mesuré par les capteurs in situ, en pointillés dans les données de

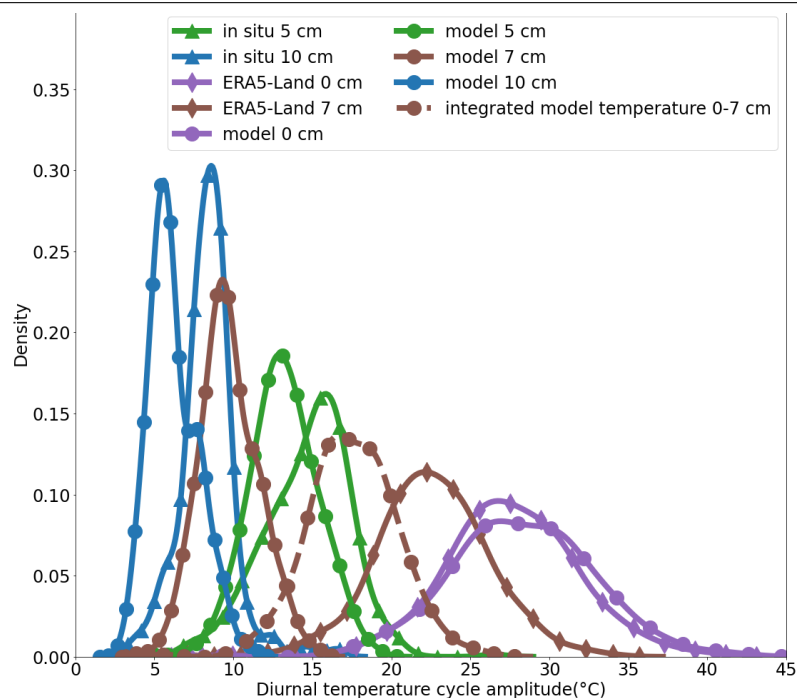


FIGURE IV.10 – Distribution des amplitudes des cycles de températures pour l'année 2009 à la station d'Ekia, avec les données in situ (triangle), les données de réanalyse (diamant), les amplitudes obtenues par le modèle analytique à différentes profondeurs (rond) et le modèle analytique intégré sur des couches (rond et ligne discontinue).

réanalyses et en trait discontinu par la solution analytique de l'équation de Fourier. Ce profil est mesuré aux heures des extrema du cycle en surface, c'est-à-dire proche de 13:00 pour le maximum et en fin de nuit ($\sim 06 : 00$) pour le minimum. Les profils concordent avec ce qui est attendu, avec une température tendant vers une valeur constante dans les couches les plus basses. Les profils calculés près de l'heure du maximum du cycle diurne (bleu/violet/vert) montrent bien une décroissance de la température avec la profondeur et inversement pour les profils proches du minimum de température (orange/rouge/marron). On peut aussi constater que les températures près du minimum sont plus froides en surface qu'en profondeur. L'écart des températures obtenues par la solution analytique par rapport à la température réelle, peut s'expliquer par une différence de diffusion thermique qui cause le déphasage. L'écart des températures fournies par la réanalyse par rapport aux autres données s'explique par le fait que ce sont des températures intégrées sur une couche et donc les variations sont atténuées.

Station Ekia

La station d'Ekia ($16^{\circ}\text{N } 1.5^{\circ}\text{O}$) fait partie du réseau de station in situ du projet *African Monsoon Multidisciplinary Analysis* (AMMA) utilisé pour la validation des méthodes de mesure de l'humidité des sols (de Rosnay et al., 2009). La température de surface radiométrique n'est pas relevée de manière systématique. Seules des mesures à des profondeurs de 5, 10, et 30 cm sont effectuées.

La comparaison est effectuée de la même façon que pour la station de Goabeb en sélectionnant le pixel le plus proche dans les données de la réanalyse ERA5-Land. La figure IV.10 donne pour l'année 2009 les distributions des amplitudes des cycles diurnes mesurées à différentes profondeurs.

Les amplitudes des cycles diurnes des données in situ (marqueur triangulaire) et obtenus par la solution analytique (marqueur rond) à 5 (courbes verte) et 10 cm (courbes bleus) ont des magnitudes similaires. La température in situ semble avoir des cycles plus faibles, mais l'absence de mesure de température de surface in situ ne permet pas d'attribuer cela à une différence de diffusivité thermique ou à un cycle diurne in situ avec une amplitude plus faible.

De la même façon que pour la station Goabeb, les températures modélisées par la solution analytique sont intégrées entre 0 et 7 cm (marqueur rond et ligne discontinue). L'écart moyen des amplitudes avec l'amplitude fournie par la réanalyse (marqueur diamant) est de ~ 5 K. Cette forte valeur met en évidence une incertitude importante dans les paramètres utilisés dans l'équation de Fourier.

En utilisant les développements effectués dans cette section, on peut donc utiliser la solution analytique de l'équation de Fourier pour estimer le profil de température dans un sol aride. La température de surface fournie par les réanalyses de manière globale peut être utilisée pour contraindre les composantes sinusoïdales de la solution analytique. Mais les températures intégrées par couche disponibles dans la réanalyse ne permettent pas d'estimer facilement la température à une profondeur quelconque dans le sol. Ce qui rend l'utilisation de la solution analytique nécessaire.

3 Estimation de la profondeur de pénétration des microondes

Pour les microondes dans les zones arides une quantification de la pénétration des ondes dans le sol en utilisant un modèle des propriétés diélectriques basé sur des cartes de caractéristiques géologiques est délicat. Pour rappel, l'équation du transfert radiatif (IV.12) établit la relation entre une température de brillance mesurée au niveau du satellite et les différentes contributions dans le champ de vision. Celles-ci proviennent de la couche effective du sol qui contribue à la température, avec l'émissivité effective associée, mais aussi de l'atmosphère. Pour une fréquence donnée ν , à une polarisation P , on peut écrire :

$$T_{B,\nu,P} = T_{\uparrow,\nu} + \tau_{\nu}[e_{eff,\nu,P}T_{eff,\nu} + (1 - e_{eff,\nu,P})T_{\downarrow,\nu}]. \quad (IV.12)$$

Avec $T_{\uparrow,\nu}$ et $T_{\downarrow,\nu}$ la contribution ascendante et descendante de l'atmosphère, τ_{ν} la transmittance atmosphérique, l'émissivité effective $e_{eff,\nu,P}$ et la température effective du sol $T_{eff,\nu}$. Plus de détails sur les termes de l'équation du transfert radiatif sont donnés dans le chapitre I.

Estimer la température effective d'une mesure impose donc d'estimer la profondeur de pénétration des microondes dans le sol à une fréquence donnée. Différentes approches pour tenir compte de cet effet existent.

Tout d'abord, pour les observations à basses fréquences (≤ 5 GHz), utilisées pour la mesure de l'humidité des sols par exemple, des paramétrisations de T_{eff} dépendant uniquement de la température en surface et à grande profondeur constante (≥ 1 m) ont été développées (Choudhury et al., 1981). Ces expressions sont de la forme :

$$T_{eff} = T_{x=0} + B \times (T_{x=0} - T_{x=1m}).$$

Des améliorations de l'estimation du paramètre B ont été proposées pour prendre en compte l'effet de variations de l'humidité des sols ou de la structure du sol (Holmes et al., 2006; Wigneron et al., 2008; Lv et al., 2018).

Pour les zones semi-arides, au-dessus de 10 GHz une paramétrisation de la relation entre la $T_{B,37GHz[V]}$ et la température à 5 cm a pu être utilisée (De Jeu, 2003).

A des fréquences plus élevées et pour les zones arides, différentes approches ont été employées pour estimer la profondeur de pénétration des microondes, et donc la T_{eff} d'une mesure. Les méthodes emploient, soit des observations de températures de brillance à plusieurs fréquences couplées à une modélisation du profil de température pour estimer la profondeur de pénétration, soit des mesures in situ de pénétration en microonde qui permettent d'estimer les propriétés diélectriques du sol à ces fréquences.

- Le travail de Prigent et al. (1999) emploie une valeur fixe de conductivité thermique pour les zones désertiques (figure IV.5) et les températures de surface proviennent des observations infrarouges thermiques de l'*International Satellite Cloud Climatology Project* (ISCCP) (Rossow et al., 1991). Des observations de l'imageur SSM/I, en orbite polaire, mais à des heures de passage sensiblement différentes (quelques heures seulement) sont utilisées pour déterminer simultanément les émissivités et la profondeur de pénétration. Pour le calcul du profil atmosphérique les données du *National Center for Environmental Prediction* (NCEP) sont utilisées. Des pénétrations dépassant les 10 cm sont estimées pour des fréquences à 19 GHz et des cartes de la zone couvrant la péninsule arabe et le Sahara sont fournies.
- Grody and Weng (2008) développent une procédure similaire et estiment des profondeurs de pénétration jusqu'à 15 cm à 19 GHz. Le coefficient de diffusivité thermique utilisé est de $0.002 \text{ cm}^2/\text{s}$ et les profils atmosphériques de température et de vapeur d'eau proviennent de modèles de prévision numérique du temps. La température de surface provient de mesures par un satellite infrarouge géostationnaire.
- Des expérimentations utilisant un radiomètre monté sur un camion ciblant une plaque d'aluminium enterrée dans le sol ont été menées par Zhang et al. (2012). Ils déterminent empiriquement la profondeur de pénétration en fonction de la composition du sol à différentes fréquences jusqu'à 36 GHz.

Ici une approche similaire à celle présentée dans Prigent et al. (1999) est utilisée pour estimer les profondeurs de pénétration à plusieurs fréquences microondes. En supposant que la pénétration est la même aux deux polarisations, et en connaissant les contributions atmosphériques, le profil de température du sol, et les températures de brillance, on peut estimer les émissivités et la profondeur de pénétration à chaque fréquence. Les températures de brillance de l'instrument GMI sont utilisées. Le profil thermique du sol est déterminé en utilisant la méthode présentée plus haut, contrainte sur les données fournies par la réanalyse ERA5. Les paragraphes suivants détaillent les données utilisées et les résultats obtenus.

3.1 Domaine d'étude et données utilisées

La zone d'étude se limite au Sahara et à la péninsule arabe entre 10 et 35°N et 22°W et 50°E. Cette zone présente l'avantage d'être le plus grand désert chaud du monde, avec différents types de surface rocheuses ou de sables. Cette zone est celle pour laquelle les problématiques de pénétration des microondes ont été remarquées pour la première fois. Cependant, rien n'empêche une extension de la méthodologie à une échelle globale.

Les **températures de brillance** utilisées sont celles de l'instrument GMI (décrit dans la section 1.2) mesurées à 10.65 (V,H), 18.7 (V,H), 36.5 (V,H) et 89.0 (V,H) GHz. Cet

instrument présente l'avantage d'avoir une orbite non-héliosynchrone. À chaque orbite de l'instrument, l'heure locale de l'observation change ce qui peut être exploité pour effectuer des mesures tout au long du cycle diurne. Moyenner les observations sur plusieurs mois et plusieurs années permet de calculer un cycle diurne robuste par rapport à la variabilité des mesures. La fenêtre de 30 minutes est un compromis entre un filtrage fort tout en gardant la dynamique temporelle du signal. En utilisant les données entre 2015 et 2020, on peut reconstruire les cycles diurnes de température avec une résolution temporelle de 30 minutes, sur une grille de $0.25^\circ \times 0.25^\circ$ utilisée pour les données ERA5. Les données contaminées par des nuages (rares sur le désert) sont filtrées en utilisant la méthode décrite dans le chapitre III (Favrichon et al., 2019).

Pour estimer le **profil de température du sol** par la solution analytique présentée dans la section précédente, deux sources de données sont nécessaires :

- Les paramètres thermiques introduits dans la section 2.1 fournis par Shangguan et al. (2014) sont projetés sur la même grille à 0.25° , en moyennant la valeur des couches ≤ 1 m.
- La température de surface est celle de la variable *skt* en moyenne mensuelle dans la réanalyse ERA5. Elle est disponible toutes les heures et une interpolation linéaire est effectuée pour obtenir des points toutes les demi-heures. Les températures dans les couches inférieures du sol n'étant pas utiles, les données ERA5-Land à une résolution différentes ne sont pas utilisées.

Enfin pour la **contribution atmosphérique** les paramètres du profil atmosphérique de pression, température et vapeur d'eau des données mensuelles moyennes par heure de la réanalyse ERA5 sont utilisées.

Toutes les données sont colocalisées sur la grille $0.25^\circ \times 0.25^\circ$ et agrégées par trimestre en couvrant les années de 2015 à 2020. Pour respecter les contraintes d'humidité des sols faibles un filtre permet de ne sélectionner que les points pour lesquels la valeur moyenne d'humidité du sol sur le trimestre ne dépasse pas $0.1 \text{ m}^3 \text{ m}^{-3}$. L'utilisation de données en moyenne trimestrielle sur 5 années est acceptable compte tenu de la stabilité des conditions sur les déserts. On obtient une base de données de températures de surface, de profils atmosphériques, et de températures de brillance moyenne, toutes les demi-heures pour les 4 trimestres.

3.2 Application

En utilisant les données présentées ci-dessus, les étapes suivantes permettent d'obtenir le profil de température dans le sol et la contribution atmosphérique attendue pour chacune des situations dans la base de données :

1. La transmittivité, $T_{\uparrow,\nu}$ et $T_{\downarrow,\nu}$ sont estimés en utilisant le profil atmosphérique moyen par heure de l'ECMWF en température, humidité, pression par le modèle de Rosenkranz (1998). Le calcul est fait pour chacune des fréquences, en moyenne trimestrielle sur 5 ans.
2. La température de surface d'ERA5 est utilisée pour estimer les paramètres \bar{T} , A_1 , ϕ_1 , A_2 et ϕ_2 des composantes de la modélisation de la température.
3. En combinant les paramètres thermiques avec ces coefficients, on peut modéliser la température à toute profondeur de la colonne de sol.

4. On collecte aussi pour une fréquence aux deux polarisations les températures de brillance moyennes trimestrielle observées par GMI entre 2015 et 2020.

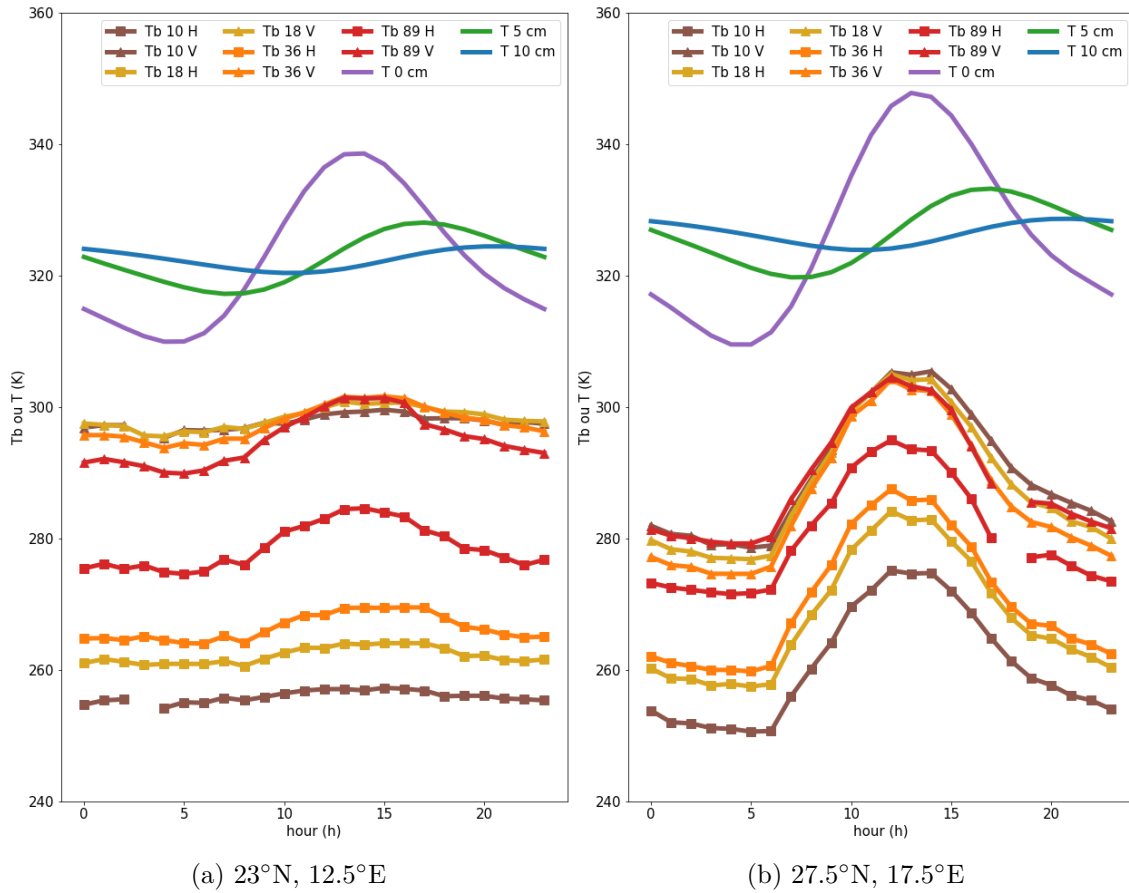


FIGURE IV.11 – Colocalisation des cycles de température modélisés ainsi que les températures de brillance de GMI moyennes (2015-2020) durant le trimestre de juillet-août-septembre pour deux pixels : à gauche (23°N , 12.5°E) et à droite (27.5°N , 17.5°E). En trait continu sans marqueur, la température obtenue par la solution analytique de l'équation de Fourier avec les paramètres obtenus par les données de réanalyse. Les températures de brillance de GMI aux polarisations verticales (triangles) et horizontales (carrés) à plusieurs fréquences (10, 19, 37 et 89 GHz) sont affichées.

La figure IV.11 présente deux exemples de la colocalisation des données moyennées pour le trimestre de juillet à septembre. Sur la même figure, les températures de brillance de l'instrument GMI et les températures à différentes profondeurs estimées par le modèle analytique sont représentées. On voit bien l'émissivité plus faible pour les T_b mesurées aux polarisations horizontales (marqueurs carrés) qui conduit à des températures de brillance plus faibles qu'à la polarisation verticale (marqueurs triangulaires). La situation des deux scènes présentées est très contrastée : à gauche dans un désert de sable (erg de Bilma) le cycle diurne de température de brillance est très atténué pour les fréquences basses, ce qui est lié à la profondeur de pénétration. A droite le cas d'un désert rocheux au-dessus du champ volcanique d'Haruj est présenté. Les différentes températures de brillance ont des cycles très similaires, avec des maxima proches, ce qui correspond à une profondeur de pénétration plus faible.

On peut alors utiliser la méthode des moindres carrés sur les 48 valeurs moyennes (toutes les 30 minutes) du cycle diurne de T_b pour chaque polarisation à une fréquence pour estimer la valeur des émissivités à chaque polarisation $e_{eff,\nu,H}$, $e_{eff,\nu,V}$ et la profondeur de

pénétration δ_p telle que $T_{eff,\nu} = T(z = \delta_p, t)$. Cette opération est répétée pour chacun des pixels pour les 4 trimestres sur la zone d'étude. L'ajustement de ces paramètres permet d'obtenir les cartes d'émissivités pour chaque polarisation à chaque fréquence, et les cartes de profondeur de pénétration pour chacune des fréquences.

4 Analyse des résultats

Les différentes cartes produites sont analysées dans la section suivante. Plusieurs axes d'analyse sont intéressants, comme la comparaison des émissivités estimées par rapport à des cartes existantes, la stabilité temporelle des profondeurs de pénétration, ou encore la comparaison des structures spatiales avec des caractéristiques géologiques connues.

4.1 Cartes d'émissivités estimées

La figure IV.12 présente les émissivités calculées. Elle sont comparées aux émissivités de TELSEM (*Tool to Estimate Land-Surface Emissivities at Microwave frequencies*) à deux fréquences : 18 GHz en polarisation verticale, et 36 GHz en polarisation horizontale. TELSEM fournit des estimations de l'émissivité pour différentes fréquences (entre 19 et 100 GHz) et angles d'incidences en polarisation verticale et horizontale. Ces estimations sont basées sur une combinaison des climatologies d'émissivités moyennes pour les surfaces continentales dans le domaine microonde à partir d'observations de l'instrument SSM/I et des méthodes d'interpolation (Aires et al., 2011c). Les données TELSEM disponibles par mois sont colocalisées et moyennées pour obtenir une valeur moyenne trimestrielle de l'émissivité sur la zone d'étude. On peut alors comparer les émissivités estimées à partir des observations GMI à celles disponibles dans la base TELSEM.

Les structures spatiales d'émissivité entre le résultat de la procédure décrite ci-dessus et les données de TELSEM sont très semblables. La différence entre les deux ne montre pas de différence systématique localisées sur certaines zones, et l'erreur moyenne est faible ($\sim 2\%$). L'émissivité fournie par TELSEM présente des structures moins abruptes que celles obtenues à partir des observations GMI. Cette différence peut s'expliquer par l'interpolation appliquée aux données TELSEM pour fournir des émissivités cohérentes entre elles, alors que les estimations basées sur les observations de GMI sont indépendantes et donc moins lissées.

Le coefficient de corrélation linéaire entre les émissivités calculées et celles de TELSEM est estimé pour tous les trimestres. Pour les fréquences à 18 et 36 GHz des valeurs de 0.89 et 0.86 sont obtenues pour les émissivités en polarisation verticale et 0.76 et 0.80 en polarisation horizontale.

Ces résultats confortent la méthodologie employée pour obtenir les résultats aux fréquences à 18 et 36 GHz. Pour les émissivités à toutes les fréquences, l'erreur de l'ajustement par la méthode des moindres carrés des émissivités sur les données est inférieur à 0.002 pour les fréquences jusqu'à 37 GHz et inférieur à 0.003 aux deux polarisations à 89 GHz.

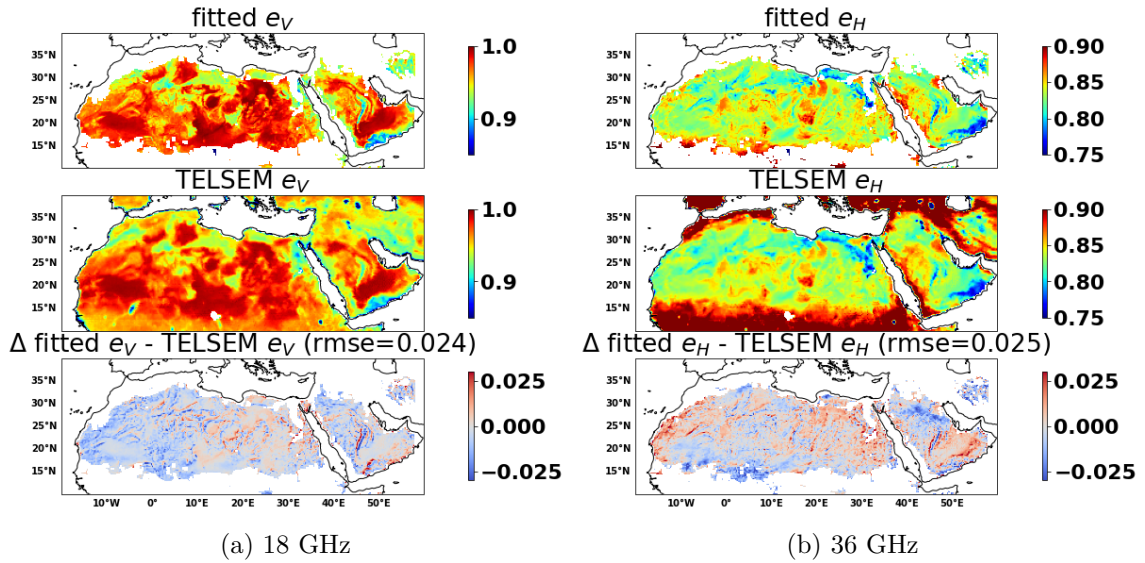


FIGURE IV.12 – Carte des émissivités estimées à 18 GHz au deuxième trimestre (avril-mai-juin) en polarisation verticale (à gauche) et 36 GHz au quatrième trimestre (octobre-novembre-décembre) en polarisation horizontale (à droite). Pour chaque fréquence, la figure du haut correspond aux émissivités calculées pour l’instrument GMI par la méthode décrite, en dessous aux données obtenues de TELSEM. La troisième ligne présente la différence entre les deux premières cartes. L’erreur moyenne (RMSE) sur les données est dans le sous-titre.

4.2 Cartes des profondeurs de pénétration

Pour la quantification de la profondeur de pénétration, on peut tout d’abord étudier les cartes qui résultent de l’application de la méthode décrite ci-dessus (section 3.2). Les résidus de l’ajustement varient entre 0.22 cm à 10 GHz et 0.16 cm à 89 GHz. Une profondeur de pénétration maximale d’environ 35 cm à 10 GHz est obtenue pour certains points dans la zone d’étude.

La figure IV.13 présente les cartes de la profondeur de pénétration estimée sur la zone d’étude. On remarque des structures spatiales très marquées et similaires aux différentes fréquences. Cette correspondance entre les estimations effectuées de manière indépendante aux différentes fréquences conforte la robustesse de la méthodologie. L’évolution des profondeurs est aussi conforme à ce qui est attendu, avec les valeurs les plus importantes atteintes aux plus basses fréquences.

La déviation standard des estimations entre les différents trimestres est aussi calculée et affichée sur la colonne de droite de la figure IV.13. Ces valeurs indiquent que les zones avec une profondeur de pénétration importante sont aussi celles avec la plus grande variabilité. En comparaison des travaux précédents ayant étudiés la profondeur de pénétration des microondes, les mêmes structures spatiales que dans Prigent et al. (1999) peuvent être observées. Mais le plus grand nombre d’observations disponibles sur le cycle diurne des températures améliorent la précision des résultats. En supplément la fréquence à 10 GHz fournit des nouvelles mesures de la profondeur de pénétration.

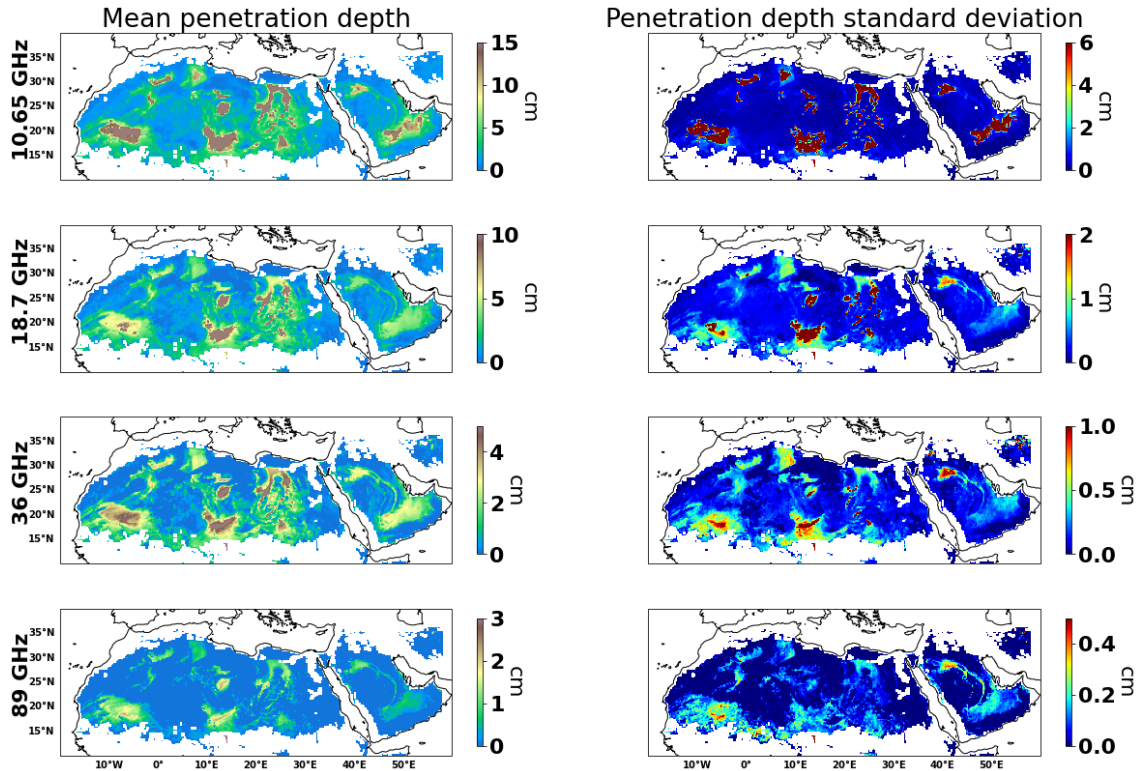


FIGURE IV.13 – Cartes des moyennes annuelles des profondeurs de pénétration de haut en bas à 10.65, 18.7, 36, 89 GHz (à gauche). Dans la colonne de droite la déviation standard calculée pour les 4 trimestres de la profondeur de pénétration est affichée.

4.3 Comparaison des structures géologiques connues

Les structures spatiales peuvent être comparées avec des structures géologiques connues. Les données de l'*United States Department of Agriculture (USDA) Global Soil Regions* fournissent une classification des types de surface à l'échelle globale. Le jeu de données à une résolution de 3 km est reprojété sur la grille utilisée pour l'analyse. Cette carte des types de sol est construite en ciblant prioritairement l'analyse des surfaces agricoles et ne représente pas forcément au mieux la réalité des zones arides.

La distribution des profondeurs de pénétration est représentée sur la figure IV.14 pour différents types de sols. Les fortes profondeurs de pénétration semblent correspondre partiellement à la présence de deux types de sols : le sable et les *Entisol Orthents* qui est un type de sol marqué par l'érosion. On remarque que la distribution présente deux modes, qui semblent avoir des comportements en microondes très différents. On peut aussi représenter la répartition géographique de ces types de sols.

La figure IV.15 représente en haut la carte des zones répertoriées comme étant couvertes de sable (*Shifting sands*) mais où la profondeur de pénétration est faible (≤ 15 cm à 10.65 GHz). En bas, les types de sols correspondants à de fortes profondeurs de pénétration sont représentés. On retrouve bien une certaine cohérence spatiale pour les zones avec des profondeurs de pénétration des microondes importantes. Elles se situent au cœur des zones d'erg, ou de zones à l'érosion éolienne très forte. Ces résultats mettent en évidence que la pénétration des microondes se trouve de manière préférentielle associée aux zones de sables. Cependant des profondeurs de pénétration importante peuvent aussi être mesurées au dessus d'autres types de surfaces. La cartographie précise des zones présentant

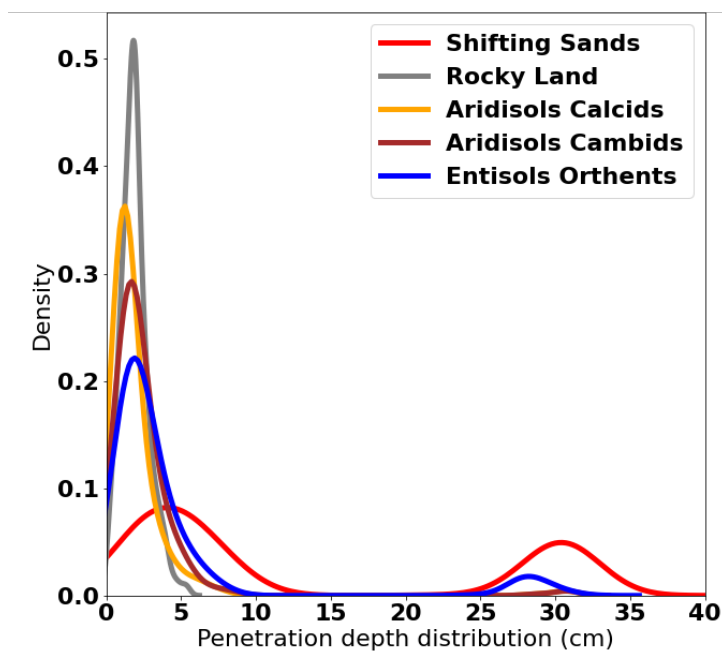


FIGURE IV.14 – Distributions des profondeurs de pénétration à 10 GHz pour différents types de sol dans les zones arides obtenus dans les données *Global Soil Regions*.

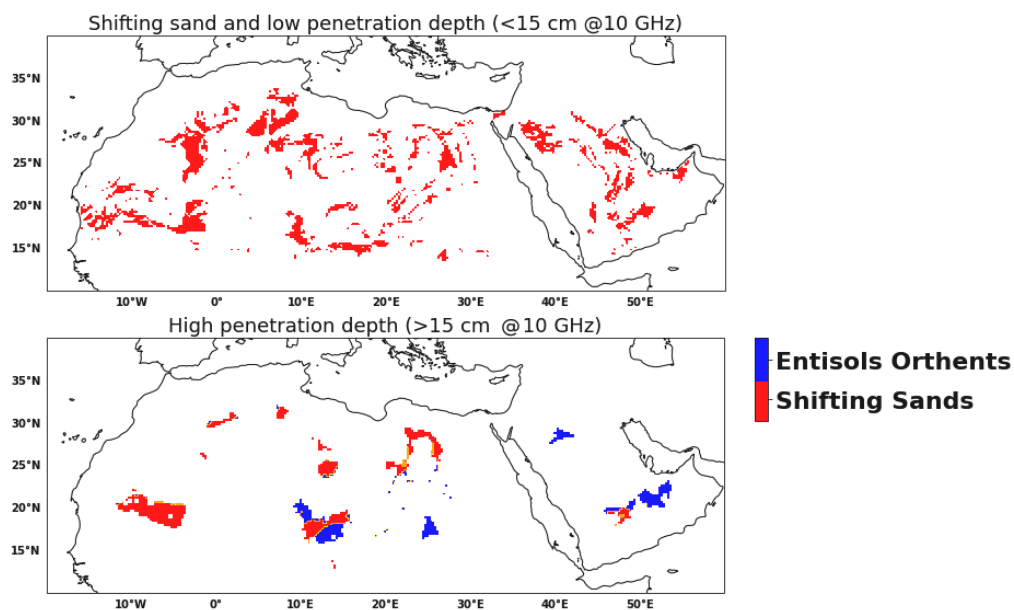


FIGURE IV.15 – Cartes comparant la présence de sables à l'obtention de faibles profondeurs de pénétration (≤ 15 cm à 10 GHz) en haut et types de sol présents pour les profondeurs de pénétration importante en bas (≥ 15 cm à 10 GHz).

une profondeur de pénétration importante comme calculé ici est donc nécessaire.

4.4 Analyse de la dépendance en fréquence des profondeurs de pénétration

Les données obtenues par l'ajustement de la profondeur de pénétration aux différentes fréquences permettent aussi de comparer les résultats obtenus aux modélisations basées sur des échantillons (section 1.3) ou aux résultats d'études précédentes (section 3). Ces

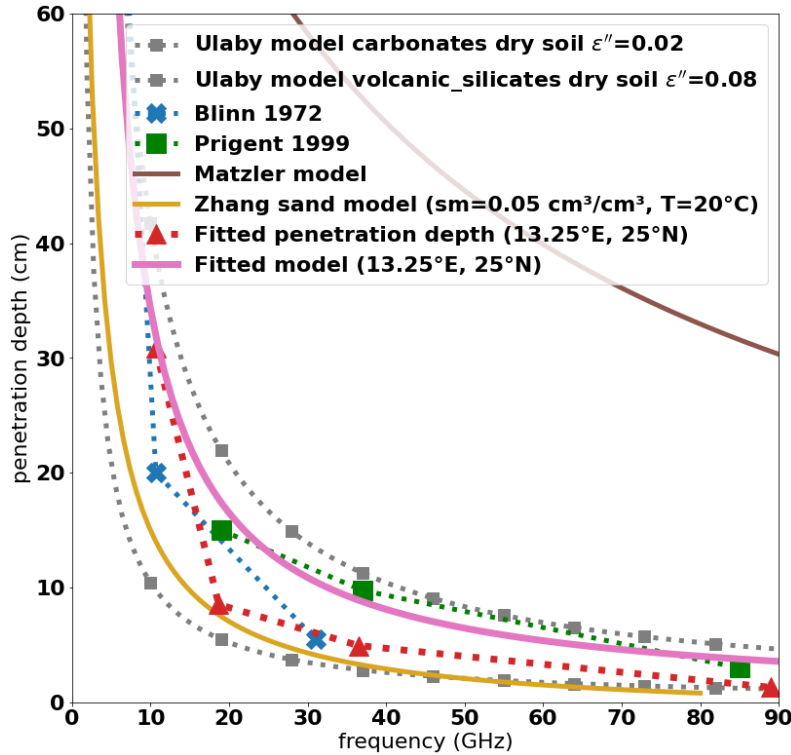


FIGURE IV.16 – Comparaisons des profondeurs de pénétration observées (en pointillé) ou modélisées (trait continu) par différentes études : les marqueurs triangulaires rouges correspondent aux valeurs obtenues par la méthode décrite ci-dessus autour de 13.25°E/25°N et modélisées par la courbe rose (plus de détails dans le texte). Les carrés gris correspondent aux valeurs obtenues par Ulaby et al. (1990), les croix bleues par Blinn et al. (1972), et les carrés verts par Prigent et al. (1999). La courbe marron correspond au modèle de Mätzler (1998) et en jaune de Zhang et al. (2012) pour les sols composés de sable.

résultats ont montré que la profondeur de pénétration décroissait avec des fréquences plus élevées et différentes modélisations ont été proposées pour ce phénomène. On retrouve cette décroissance pour tous les modèles et observations de la figure IV.16. Les profondeurs de pénétrations calculées par différentes études pour les échantillons de sable ou les zones avec une forte profondeur de pénétration donnent des résultats similaires.

Les valeurs maximales de pénétration observées par Prigent et al. (1999) sur la même zone d'étude (carrés verts) correspondent bien aux profondeurs de pénétration estimées avec l'instrument GMI dans ce travail. La profondeur maximale observée à 19 GHz est de 15 cm environ similaire entre les deux études réalisées avec des sources de données différentes. Les structures spatiales présentées sont similaires mais très marquées. Et comme on l'a vu ci-dessus la variation spatiale de la profondeur de pénétration ne peut pas être associée de manière évidente à un type de surface.

En utilisant les équations IV.4, IV.7 et IV.8 on peut déterminer la valeur de ϵ' en fonction de la densité du sol sur un pixel et modéliser la variation de ϵ'' en fonction de la fréquence. En combinant cette modélisation d' ϵ' et ϵ'' on obtient la courbe rose sur la figure IV.16. Cette équation utilise la densité du sol pour déterminer la valeur constante de la partie réelle de la permittivité, diélectrique puis une relation inverse avec la fréquence pour la partie imaginaire de la permittivité diélectrique (paramètres a et b de l'équation IV.8). Des corrections peuvent être nécessaire pour corriger de l'angle de visée du satellite.

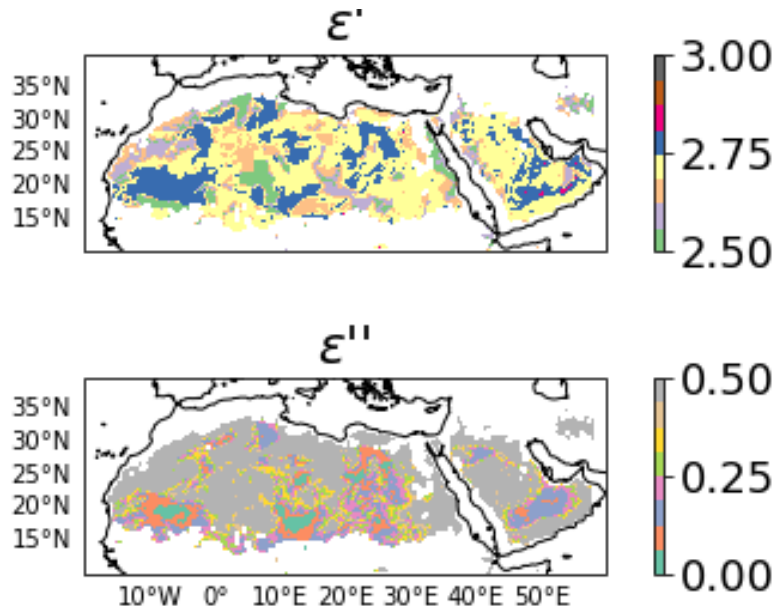


FIGURE IV.17 – Carte représentant les valeurs de ϵ' et ϵ'' à 18 GHz obtenues à partir des profondeurs de pénétration, de la densité du sol et du modèle décrit par l'équation IV.4.

Cette relation permet de déterminer des cartes des paramètres de permittivité diélectrique (ϵ' , ϵ''). La figure IV.17 représente ces valeurs calculées à 18 GHz, en utilisant les profondeurs de pénétration estimées. La partie supérieure présente la carte des valeurs de ϵ' , la densité varie peu donc les valeurs sont assez constantes sur toute la zone d'étude. Des structures spatiales similaires à celles de la profondeur de pénétration sont observées pour ϵ'' . La variabilité de la profondeur de pénétration provient donc surtout de la partie imaginaire de la permittivité diélectrique.

5 Conclusion

L'étude met en évidence la variabilité importante qui existe dans les profondeurs de pénétrations des microondes dans les sols arides. Cette profondeur de pénétration peut poser des problèmes dans l'estimation de la température de surface. L'épaisseur de la couche de sol qui contribue effectivement au signal mesuré a une température qui peut être très différente de la température en surface. On peut estimer la température en tout point de la colonne de sol en utilisant l'équation de Fourier. Cela requiert de connaître la diffusivité thermique du sol et le cycle diurne des températures en surface que l'on obtient des données ERA5. Ce profil de température dans le sol peut être utilisé conjointement avec les contributions atmosphériques à différentes fréquences et les mesures du cycle diurne moyen des températures de brillance observées par GMI pour résoudre l'équation

du transfert radiatif. Pour chaque fréquence, on estime alors les émissivités aux deux polarisations et la profondeur de pénétration. Une carte des émissivités et des profondeurs de pénétration à 10, 18, 36 et 89 GHz pour une zone couvrant le Sahara et la péninsule arabique est disponible.

Les résultats obtenus pour les émissivités à 18 et 39 GHz sont cohérents avec les émissivités TELSEM. Pour les cartes de profondeur de pénétration, les structures spatiales présentent des similarités avec la distribution des zones de sables mais ne se recoupent pas totalement. En supposant que la profondeur de pénétration est causée par une diminution de l'absorption dans le sol, en supposant les effets de diffusion faibles on peut modéliser son évolution. Différents modèles proposés permettent de simuler la profondeur de pénétration, mais dépendent de paramètres qui sont à déterminer pour chaque pixel. Le paramètre le plus influent est la partie imaginaire de la permittivité diélectrique que l'on peut donc modéliser à partir des estimations de profondeurs de pénétration. On obtient ainsi des cartes de permittivité diélectriques entre 10 et 90 GHz. De nombreux développements sont encore nécessaires pour tirer complètement parti de la modélisation de ce phénomène, pour mieux comprendre les propriétés diélectriques des sols et pour corriger les températures de surface estimées par les microondes.

Dans le but d'améliorer les estimations de la température de surface à partir d'observations microondes passives au-dessus des zones arides, il faut prendre en compte le phénomène de la pénétration des microondes dans un milieu sec décrit dans ce chapitre. Une fois la profondeur de pénétration connue à toutes les fréquences la question de la correction à apporter pour déterminer la température de surface se pose. La solution de ce problème correspond à la résolution d'un problème inverse de conduction de la chaleur.

Plusieurs cas sont à distinguer :

- Pour les estimations basées sur une unique fréquence (Holmes et al., 2009) la profondeur de T_{eff} peut être déterminée de manière précise en utilisant les cartes développées ci-dessus. En connaissant les paramètres de la température moyenne sur un cycle diurne et la diffusivité thermique il est possible de déterminer la température à une autre profondeur : l'amplitude de l'écart entre T_{eff} et \bar{T} suit une relation exponentielle qui change avec la profondeur, et l'heure de la mesure est corrigé du déphasage attendu (Holmes et al., 2008).
- Dans le cas où l'estimation de la température ne repose pas sur une seule fréquence, il est possible d'intégrer la profondeur effective des températures contribuant aux mesures de températures de brillances dans l'équation du transfert radiatif. Dans ce cas, au-dessus d'une zone aride, on associe à chaque fréquence une profondeur de pénétration, et donc une température différente dépendant des paramètres thermiques et du développement en série de Fourier. La paramétrisation de Zhang et al. (2012) de la profondeur de pénétration est utilisée par Zhou et al. (2017) pour développer une méthode pour estimer la température de surface en se basant sur des observations microondes de l'instrument AMSR-E.
- Au-dessus des zones arides, il est aussi envisageable de modifier la méthode d'estimation de la température de surface. Pour ces zones les fréquences supérieures à 85 GHz sont utilisables pour estimer la température de surface avec des contaminations atmosphériques qui sont faibles dans les zones arides. A 89 GHz, on l'a vu ci-dessus, la profondeur de pénétration reste $\leq 2cm$, en utilisant des fréquences plus élevées, on peut donc avoir $T_{eff} \approx T_{z=0}$.

Avant une mise en place systématique de telles corrections, un certain nombre d'amé-

liorations des résultats obtenus ici peuvent être proposés.

Ce chapitre aborde la problématique de la pénétration des microondes dans les zones arides, mais peut encore être améliorée sur un certain nombre de points :

1. Le premier axe est l'extension de la zone d'étude à l'ensemble du globe. Toutes les données utilisées ici sont aussi disponibles à cette échelle et dans l'optique d'une utilisation systématique toutes les zones arides présentant des profondeurs de pénétration élevées doivent être déterminées.
2. Un second point est le développement des méthodes de correction de la température de surface. Certaines pistes ont été évoquées dans la section précédente, mais des alternatives plus simples pour des instruments spécifiques observant à des heures locales fixes pourraient être envisagées dans un premier temps.
3. Enfin, l'indicateur des zones avec une profondeur de pénétration importante des observations microondes utilisé dans le cadre du projet LST CCI peut être amélioré en utilisant les résultats obtenus dans cette étude qui précisent les zones affectées.

En plus de ces résultats qui feront l'objet d'une publication en préparation, l'étude des propriétés diélectriques des sols et des structures géologiques sur les déserts offre des perspectives intéressantes. Elles pourraient être utilisées dans les études d'émission d'aérosols (Foucart, 1987; Jounet et al., 2014) qui proviennent majoritairement des zones d'érosion éoliennes (Bodélé) ou couvertes majoritairement de sable. Le peu de données disponibles sur les zones arides rend ces analyses importantes pour mieux connaître ces régions. Rattacher les structures spatiales à des structures géologiques connues pourrait faire l'objet d'études approfondies.

Ici les données entre 10 et 90 GHz ont été traitées. L'extension des résultats à des fréquences plus basses (1.4 GHz) peut être envisagée, ou utilisant des sources d'observations complémentaires (radar). Des études ciblant spécifiquement ces basses fréquences pourraient apporter des informations complémentaires pour adapter les modélisations du comportement de la permittivité diélectrique dans le domaine des microondes à basse fréquence. Ces résultats trouveront des applications à la fois sur Terre, mais aussi dans le domaine de l'exploration spatiale (Bonney, 2020), où de nombreuses planètes présentent des aridités extrêmes.

En résumé

- La présence d'eau influence très fortement les propriétés diélectriques des sols. Les propriétés diélectriques caractérisent l'atténuation d'un rayonnement électromagnétique dans la matière. Dans les zones arides, la faible quantité d'eau fait diminuer cette atténuation et donc augmenter la profondeur de pénétration.
- Avec l'équation de Fourier, il est possible de modéliser la température de la colonne de sol, en connaissant les propriétés thermiques du sol et le cycle diurne de la température en surface.
- En combinant les modélisations de la température dans le sol et des observations des cycles diurnes de températures de brillance à différentes fréquences mesurées par l'instrument GMI, on peut déterminer les émissivités et les profondeurs de pénétration des microondes entre 10 et 90 GHz sur les zones arides.
- Les estimations de la profondeur de pénétration aux différentes fréquences sont fournies sur le Sahara et la péninsule arabique et peuvent être utilisées pour estimer les propriétés diélectriques des sols dans les zones arides.
- La connaissance de la profondeur de pénétration des microondes dans les zones arides doit permettre d'améliorer la restitution des températures de surface issues des microondes dans ces régions. Connaître cette profondeur peut aussi permettre de mieux prendre en compte la contribution radiométrique de la surface, dans l'assimilation des observations microondes pour le sondage atmosphérique des couches les plus basses de l'atmosphère.

Chapitre V

Amélioration de la résolution spatiale des estimations de température de surface

Le principal désavantage rencontré par les utilisateurs des produits dérivés des observations microondes passives est la résolution spatiale bien plus faible que les produits géophysiques équivalents issus des observations infrarouges. Une amélioration souhaitable du produit de température de surface pour en faciliter l'usage serait donc d'en augmenter la résolution spatiale. Les progrès techniques de l'instrumentation permettront d'accomplir ce travail, mais il n'est pas possible de modifier les instruments en orbite ni les observations déjà effectuées. La problématique de l'augmentation de la résolution spatiale du produit de température de surface est donc la thématique développée dans ce chapitre. La méthode s'appuie sur des informations à haute résolution provenant d'observations infrarouges pour déterminer la structure spatiale des températures à l'intérieur d'un pixel microonde. Elles sont ensuite combinées avec les estimations à basse résolution issues des observations en microonde. Pour parvenir à une méthode applicable aux observations microondes quelle que soit la couverture nuageuse, les informations à haute résolution doivent être sélectionnées et potentiellement corrigées pour réduire les biais des variables auxiliaires obtenues en ciel clair uniquement. On peut alors obtenir des valeurs de température de surface avec une résolution spatiale élevée, « tout temps » par désagrégation des estimations issues des microondes.

Les mesures de température de brillance microondes peuvent servir à estimer la température des surfaces continentales dans de nombreuses conditions. Malgré un certain nombre de défis abordés dans les chapitres précédents, il est déjà possible d'obtenir des séries de données sur au moins 30 ans en combinant les observations de capteurs à bord de différentes plateformes. Mais les mesures obtenues par les capteurs passés et actuels (SSM/I, AMSR ou même GMI) ont une résolution qui est au mieux de 15 km, mais plus souvent proche de 25 km (cf. chapitre I). Pour un grand nombre d'applications des résolutions plus précises sont très bénéfiques et parfois nécessaires (Atkinson, 2013). C'est en particulier le cas pour l'aide aux pratiques agricoles de précisions qui nécessitent effectivement des résolutions spatiales bien plus élevées que celles fournies actuellement par les instruments microondes. Des estimations globales de la température de surface à haute résolution (HR) sont disponibles grâce aux observations de capteurs dans le domaine de l'infrarouge, mais sans la capacité d'observer par tous les temps.

Les produits de température de surface diffèrent par leur résolution spatiale (taille au sol du champ de vision de l'instrument) mais aussi par leur résolution temporelle (durée entre deux mesures d'un même point à la surface). La section suivante détaille tout d'abord les résolutions des différents produits existants. L'accent est mis sur la résolution spatiale qui fait l'objet de ce chapitre. Dans les sections suivantes la construction d'une méthode de désagrégation applicable aux estimations microondes est détaillée. Après la collecte des informations auxiliaires disponibles et la création d'un modèle permettant de déterminer la structure spatiale des températures de surface, des températures simulées sont utilisées pour valider la méthodologie. La méthodologie est ensuite adaptée pour prendre en compte l'effet des nuages sur les données auxiliaires. La méthode peut finalement être appliquée à des estimations issues de mesures par l'instrument SSMIS.

1 Résolutions spatiales et temporelles des produits de température de surface issus d'observations satellites

Les avantages et inconvénients des différents produits de température de surface (T) sont résumés dans cette section. Comme cela a été décrit dans le chapitre I, la température de la surface à une échelle globale, peut être estimée depuis des mesures satellites : le plus souvent des observations dans le domaine infrarouge (IR) (Wan and Dozier, 1989), mais aussi des observations microondes.

On a vu que la résolution spatiale des observations microondes était au mieux de 15 km avec les instruments actuels. En infrarouge, quand l'observation est faite par des satellites géostationnaires, on obtient des cartes de T avec un temps de revisite de 15 min à une résolution spatiale de ~ 5 km (Trigo et al., 2008). Des résolutions spatiales plus élevées sont atteintes avec les satellites en orbite polaire (jusqu'à quelques centaines de mètres), mais au détriment du temps de revisite (Wan and Li, 1997).

Comme on le voit sur la figure I.11 du chapitre I, et dans le chapitre II, la résolution spatiale d'une observation dépend de la taille de l'antenne et de l'altitude du satellite. Les instruments placés à des orbites plus hautes, doivent donc déployer des capteurs de plus grande taille pour obtenir la même résolution spatiale que ceux à des orbites plus basses. La résolution temporelle résulte, elle, du type d'orbite et du fonctionnement de l'instrument. Les satellites géostationnaires (avec une orbite haute ~ 36000 km) par définition restent à une position fixe au-dessus du même point à la surface de la Terre, ils peuvent donc effectuer des observations avec un échantillonnage temporel important de la zone couverte par le capteur. En revanche pour les orbites polaires (≤ 1000 km) deux observations par jour d'un point sont effectuées et la revisite dépend de la largeur de la fauchée de l'instrument.

La figure V.1 illustre l'effet des différences de résolution de plusieurs produits au-dessus de l'Europe. Les détails spatiaux visibles à très haute résolution (figure de gauche) sont perdus dans les observations infrarouges géostationnaires (figure du milieu). Pour les observations infrarouges, les vignettes ont été choisies pour éviter la couverture totale de la zone par des nuages, cependant, les problèmes causés par la présence nuageuse sont très visible. L'estimation de T à partir d'observations IR n'est effectuée qu'en ciel clair, les distributions des températures mesurées sont donc biaisées, car ne représentent que des cas de ciel clair (Ermida et al., 2019). Enfin, la température de surface obtenue par les observations microondes montre bien des variations spatiales de température. L'information le long des côtes est en par contre dégradée. Cela est dû au fait que l'intrusion de

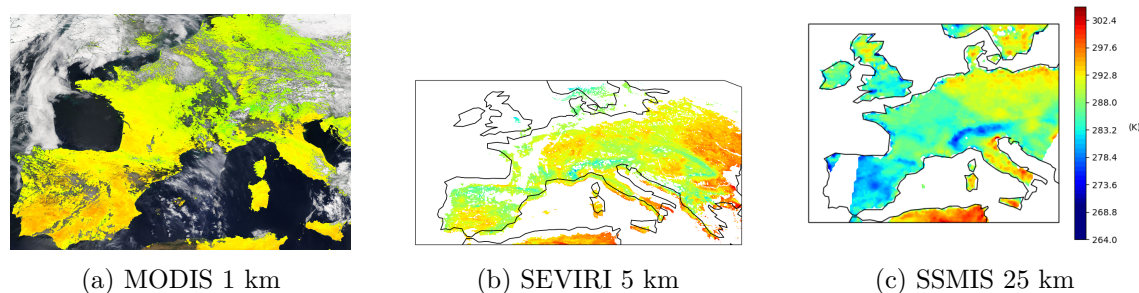


FIGURE V.1 – Cartes sélectionnées au-dessus de l’Europe pour différents produits de température de surface (K) issus des données MODIS avec une résolution de 1 km (10 mai 2015), SEVIRI avec une résolution de 5 km (27 aout 2016), et SSMIS à ~ 25 km (10 mai 2016).

portions de surface couverte d’eau dans une observation contamine très fortement les températures de brillance : les émissivités sont plus faibles sur les océans que sur les continents et l’estimation de température sur les côtes en est affecté.

Instrument	Résolution spatiale	Résolution temporelle	Couverture spatiale
Landsat 7	60 m	16 jours	\sim Globale
MODIS	1 km	2 passages \times 2 jours	\sim Globale
MSG SEVIRI	5 km	toutes les 15 minutes	Géostationnaire (~ 20 % du globe)
SSMIS	25 km	2 passages \times jour	\sim Globale

TABLE V.1 – Table résumant les résolutions spatiales et temporelles et la couverture spatiale de certains instruments utilisés pour fournir des produits de température de surface.

La table V.1 met bien en évidence la grande différence de résolution spatiale des mesures issues des microondes comparées à l’infrarouge. Cet écart important justifie la nécessité de développer des méthodes de désagrégation pour compléter les produits infrarouges pour les cas nuageux et rendre disponibles des observations à haute définition pour toutes les conditions météorologiques.

Pour surmonter cette limitation des produits infrarouge, certains produits utilisent les T dérivées d’un modèle de bilan énergétique de la surface terrestre pour combler les lacunes dans les estimations T dérivées de l’infrarouge (Martins et al., 2019). Une revue des méthodes de complétion des données manquantes causées par les nuages est donnée dans Mo et al. (2021). D’autres approches permettant de fournir une température de surface terrestre « tout temps » reposent sur la méthode de fusion des informations en utilisant une combinaison de mesures IR et de distribution attendue des températures (Wu et al., 2019, 2021; Xu and Cheng, 2021). Enfin une alternative consiste à utiliser les estimations de température issues des microondes comme point de départ et leur appliquer une méthodologie de désagrégation qui permette d’atteindre des résolutions similaires à celles des produits infrarouges. C’est cette piste qui est développée dans ce chapitre.

2 Méthodes pour la désagrégation de la température de surface

Pour effectuer cette augmentation de la résolution spatiale, on applique une désagrégation aussi appelée réduction d’échelle. D’une observation à basse résolution (échelle large),

on veut obtenir un nouveau produit avec une résolution plus élevée (échelle fine). On appelle désagrégation l'utilisation d'une méthode permettant d'augmenter la résolution spatiale d'une valeur estimée.

Le principe de la désagrégation est d'utiliser des modèles ou des algorithmes pour déterminer la température des pixels à haute résolution inclus dans un pixel à plus basse résolution. Un certain nombre d'hypothèses sont associées à l'augmentation de la résolution des cartes de température :

- ✓ L'hétérogénéité spatiale de la température est plus fine que l'observation à basse résolution et l'on dispose d'informations sur la distribution spatiale des températures à haute résolution. Cela justifie la nécessité de développer des méthodes de désagrégation pour estimer une température à haute résolution à partir de la température à basse résolution.
- ✓ Une relation existe entre des propriétés physiques mesurables des pixels à haute résolution et la variable de température ciblée. Cette relation doit être suffisamment générale pour être applicable à de nombreux pixels. Ceci justifie le recours à des données auxiliaires permettant de guider la structure spatiale à haute résolution.
- ✓ Aussi on suppose que tous les pixels à haute résolution inclus dans une observation contribuent à la température à basse résolution. En première approximation cette contribution est supposée égale pour tous les sous-pixels.

La traduction de ces hypothèses en équation est détaillée dans la section 3.1 lors de l'implémentation de la méthode de désagrégation.

Cette désagrégation s'applique à des données de niveau 2 ou 3 (variable géophysique de température). Des méthodes qui ne sont pas liées aux spécificités des microondes peuvent donc être étudiées.

A la différence des surfaces océaniques, les continents ont une capacité thermique volumique plus faible. La température de surface, exposée au rayonnement solaire subit donc des variations quotidiennes et saisonnières plus importantes. On s'attend aussi à ce que la végétation et l'élévation affectent la distribution spatiale de T , la végétation augmentant la capacité thermique de la surface et l'élévation diminuant les T locales (Anderson et al., 2003). De plus les caractéristiques de la surface, comme le taux d'humidité du sol, conduisent à une variabilité spatiale qui peut aussi être importante (Bechtel, 2015). De plus la température de la surface est influencée par le rayonnement solaire qui est modifié par la saison, l'heure locale et aussi en présence de nuages. L'impact de ces effets doit donc être également pris en compte.

On veut déterminer une relation entre la température et des paramètres liés à ces différentes caractéristiques pour fournir des informations utiles sur le motif à haute résolution des températures de surface. Cette section décrit les méthodes existantes pour améliorer la résolution des températures et les informations qui peuvent être utilisées.

De manière générale la désagrégation pour les températures de surface s'appuie sur un ou plusieurs des axes suivants (Atkinson, 2013) :

- Des hypothèses ou des connaissances a priori sur les caractéristiques de la température à haute résolution, comme des climatologies de distribution de température à haute résolution.
- Une interpolation entre des valeurs connues. Si on suppose que la variation spatiale des températures entre deux observations est linéaire, alors on peut déterminer la température à des intervalles discrets entre les deux observations.

- L'utilisation de relations entre des observations simultanées à une meilleure résolution. Par exemple en microonde, les observations à des fréquences plus élevées ont une meilleure résolution spatiale que celles à des fréquences plus basses. Si on peut déterminer la relation entre les T_b mesurées par les différents canaux, on peut améliorer la résolution spatiale de l'observation à basse fréquence vers celle de la haute fréquence.

2.1 Pour les températures issues de l'infrarouge

La littérature est particulièrement abondante pour le cas des températures issues de l'infrarouge. On a vu dans la section précédente que selon l'orbite et l'instrument, les données infrarouges pouvaient avoir un rapport 10 entre leurs résolutions, mais présentent pour autant des caractéristiques physiques similaires. De nombreuses études ont donc été conduites pour augmenter artificiellement la résolution des estimations de températures issues de l'infrarouges. Zhan et al. (2013) résument différentes approches que l'on peut regrouper de la façon suivante :

L'approche basée sur des **méthodes physiques** où une description physique de l'évolution de la température couplée à des informations existantes au sein d'un système d'assimilation de données permet de contraindre la température des sous-pixels compte tenu des informations fournies (Mechri et al., 2016).

Une seconde approche vise à déterminer une **relation empirique** entre des informations qui sont disponibles à une meilleure résolution que la variable à désagréger. Cette approche peut être utilisée avec des canaux à des fréquences différentes qui présentent des résolutions supérieures, mais peut surtout utiliser des informations provenant d'autres sources (Kustas et al., 2003). Les régressions statistiques bénéficient des informations fournies par les données auxiliaires comme l'indice de végétation *Normalized Difference Vegetation Index* (NDVI) ou les Modèles Numériques de Terrain (MNT) mais peuvent aussi être enrichies par des mesures d'humidité des sols, ou tout autre paramètre dont la valeur peut être liée à celle de la température.

Dans chacun de ces cas, un modèle physique ou empirique de la distribution de température est utilisé. Dans les travaux de Zhan et al. (2012) et Peng et al. (2017) sur les méthodologies de désagrégation, il est souligné l'importance de la sélection des informations auxiliaires à haute résolution spatiale qui sont utilisées pour fournir les motifs spatiaux à haute résolution. Les données disponibles seront détaillées dans la section 2.3.

2.2 Pour les variables issues d'observations microonde passives

Dans le domaine des microondes la désagrégation a, jusqu'à présent, été plus largement appliquée à l'estimation d'autres paramètres, tels que l'humidité du sol (Merlin et al., 2008; Peng et al., 2017). Dans ce domaine, l'utilisation d'observations provenant de mesures de radar, de température de surface dans le domaine infrarouge, ou encore reposant sur des sur-échantillonnages d'observations à basse résolution (De Jeu et al., 2017) ont été employées. Quelques études ont été effectuées pour la désagrégation des températures de surface. On retrouve toutefois l'utilisation des MNT (Abrams et al., 2020) en raison de la relation simple entre l'altitude et la température de surface, de l'ordre de 6.5 K/km (Minder et al., 2010). Parmi les travaux pertinents on peut lister les suivants :

- Duan et al. (2017) ont combiné les estimations de T issues des observations AMSR-E à ~ 25 km et de T infrarouge du MODIS à 1 km. Sous un ciel nuageux, l'information à haute résolution spatiale est fournie par un MNT et une correction linéaire est appliquée pour corriger l'effet de l'altitude sur les T microondes à basse résolution spatiale. L'estimation est appliquée sur la Chine, aux heures de passage du satellite Aqua (01:30 et 13:30).
- Avec des données similaires, Sun et al. (2019) ont développé une estimation de la T à haute résolution à partir d'observations AMSR-E à basse résolution combinées avec des données infrarouges MODIS et *Geostationary Operational Environmental Satellite* (GOES). Les pixels infrarouges contaminés par les nuages sont comblés par l'information microonde, avec l'aide de données NDVI et MNT à haute résolution pour augmenter la résolution spatiale de la T microonde au-dessus de l'Amérique du Nord.
- Une alternative est proposée par Zhang et al. (2019) qui constate que les méthodes adoptées pour la descente d'échelle de T microondes ne parviennent pas à quantifier l'effet des variations temporelles liées au cycle diurne de T , limitant leur utilisation à une seule heure locale de passage du satellite. Un indice de végétation à haute résolution, des informations sur l'altitude à une résolution de 1 km et les sorties d'une réanalyse sont utilisés pour réduire l'échelle des T observées par AMSR-E. Une décomposition de la variation du cycle diurne en différentes composantes (variation annuelle, variation météorologique et cycle) est adoptée en chaque point pour corriger les estimations de T dérivées des microondes. La méthode est appliquée sur le nord-est de la Chine. La complexité des données d'entrée requises entrave son applicabilité à une plus large échelle.
- Zhang et al. (2020) utilisent une succession de régressions linéaires pour augmenter étape par étape, la résolution des données AMSR-E vers la résolution des données MODIS. La résolution passe ainsi de 0.25° à 0.08° , puis 0.03° et enfin 0.01° . Les données auxiliaires utilisées sont le NDVI et les mesures d'élévation et de pente de la mission topographique de la navette spatiale (SMRT). L'avantage des régressions linéaires successives, démontré sur des scènes en Chine continentale, est de limiter l'impact d'un changement de relation entre les variables explicatives et la variable cible à des échelles trop différentes.

Les deux premières méthodes listées ont une approche similaire, elles n'utilisent les observations microondes que pour les pixels IR nuageux. Cela repose sur une hypothèse très forte : l'homogénéité des températures estimées par infrarouge en ciel clair et les microondes peut être garantie. Dans le cas contraire cela risque d'introduire un biais entre les zones nuageuses et en ciel clair.

Toutes ces études utilisent des paramètres auxiliaires similaires, représentant la topographie ou la végétation et seule l'étude de Zhang et al. (2019) prend en compte l'effet d'un changement d'heure sur les relations entre les températures à haute et basse résolution. Les autres méthodes sont spécifiques à un temps de passage à l'équateur donné, pour un satellite. Au sein d'une empreinte au sol de T issues des microondes à basse résolution, les surfaces hétérogènes peuvent avoir des capacités thermiques différentes et des cycles diurnes de température différents : ceci doit être pris en compte. L'utilisation de sources de données à une plus haute résolution que les observations microondes disponibles pour toute observation et tout au long du cycle diurne est donc nécessaire.

Les études portant sur la désagrégation de variables géophysiques ont montré qu'appliquées individuellement à chaque pixel à basse résolution, ce type de méthodes a tendance

à produire des discontinuités parasites aux bords de chaque pixel de basse résolution (Liu, 2000; Aires et al., 2017). Dans ces études, différentes méthodes sont appliquées utilisant des filtres passe-bas ou des contraintes de régularisation pour lisser les solutions.

2.3 Données auxiliaires à haute résolution

Estimer les structures spatiales de température de surface à haute résolution et à toute heure du jour nécessite de s'appuyer sur des données auxiliaires. Dans la section précédente on a vu que les données utilisées étaient le plus souvent issues de deux sources : un index de végétation, comme le NDVI et les MNT qui décrivent la topographie de la surface. D'autres études proposent d'utiliser des informations sur le cycle diurne comme la température maximale ou la température minimale (McVicar and Jupp, 2002; Kalma et al., 2008). Des informations sur le cycle diurne de T lui-même peuvent être disponible à partir d'un modèle ou d'une observation. La modélisation simple de Göttsche and Olesen (2009) décrite dans l'introduction résume le cycle diurne avec peu de paramètres.

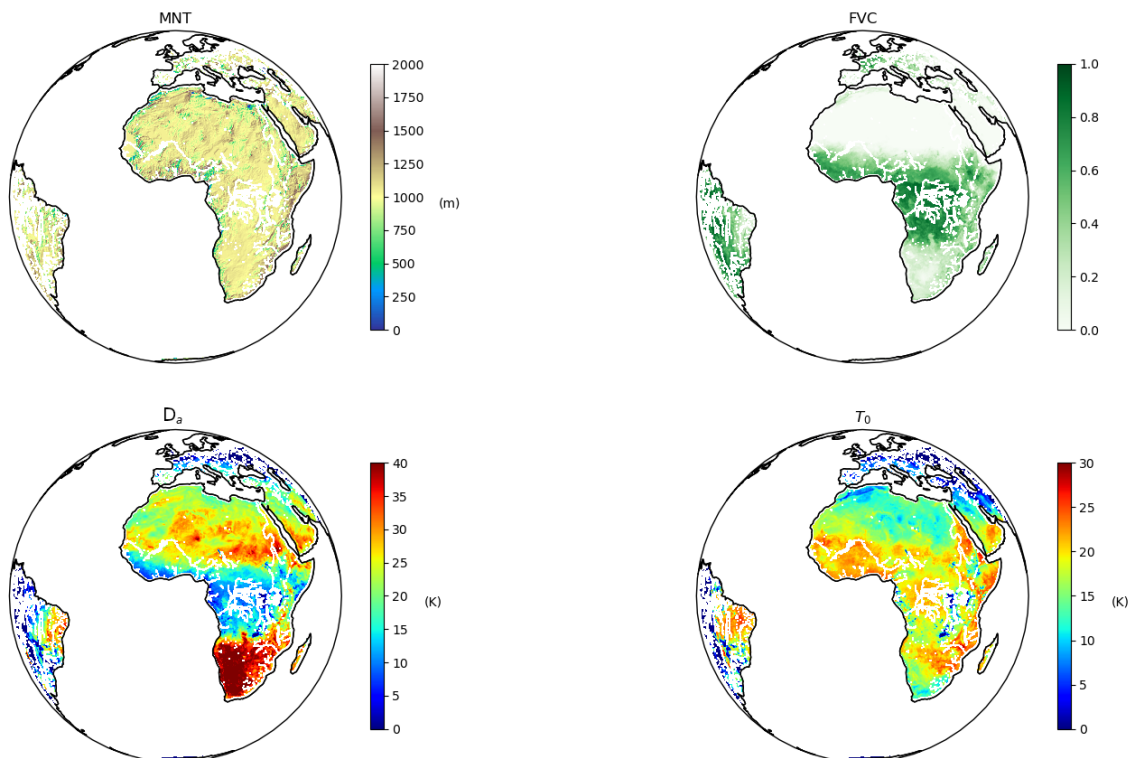


FIGURE V.2 – Modèle numérique de terrain à une résolution de 10 km obtenue par l'instrument ASTER (Abrams et al., 2020). Fraction de Végétation par pixel (*Fractional Vegetation Cover*- FVC), amplitude du cycle diurne de températures (*Diurnal cycle amplitude* - D_a) et température minimale (T_0) du cycle diurne modélisé dans les données LSA SAF novembre 2016.

La figure V.2 présente quelques variables disponibles à haute résolution. Les pixels contenant de l'eau sont masqués. En haut à gauche, un modèle numérique de terrain estimé à partir des données ASTER est présenté. Ces données peuvent être disponibles à une très haute résolution mais sont statiques pour un pixel donné. Les 3 sous-figures suivantes proviennent des données du LSA SAF. Les observations de l'instrument SEVIRI à bord d'un satellite géostationnaire couvrent la Terre entre $\pm 60^\circ$ N. La résolution est de ~ 5 km et la disponibilité supérieure à 10 ans. Les données SEVIRI permettent de dériver

des produits de température de surface toute les 15 min ainsi que des données agrégées sur 10 jours, tels que des indices de végétation. La fraction de végétation par pixel est un de ces indices, présenté sur la figure V.2. La valeur minimale T_0 et l'amplitude D_a du cycle diurne médian sont calculés sur 10 jours avec le modèle de Göttsche and Olesen (2009).

De très claires structures spatiales apparaissent sur les différentes variables (figure V.2) qui peuvent être liées au couvert de surface comme la végétation, à la présence d'eau ou encore à l'orographie. L'échantillonnage temporel des variables D_a et T_0 est suffisant (10 jours) pour permettre de prendre en compte la dynamique saisonnière des températures de surface. La contrainte imposée par l'utilisation d'observations visibles et infrarouges est que les variables agrégées sur 10 jours ne sont basées que sur des observations par temps clair. Cette fenêtre de 10 jours permet dans la majeure partie des cas d'avoir suffisamment de points de mesure pour calculer les paramètres du cycle moyen.

Il est cependant nécessaire de développer une méthode de désagrégation qui puisse fonctionner efficacement à la fois en ciel clair et couvert. Dans les exemples tirés de la littérature, les méthodes de désagrégation ont été appliquées jusqu'à présent aux estimations de T microondes, sans tenir compte de la couverture nuageuse. Cependant, on s'attend à ce que les nuages affectent la distribution spatiale des T locales. Par exemple, l'étude de Ermida et al. (2019) illustre la différence significative entre les T diurnes plus froides sous les nuages que par temps clair. Cette différence devrait dépendre des types de surface, les zones à grande capacité thermique (e.g. les surfaces végétalisées et humides) étant moins sujettes aux variations de T que les autres surfaces (telles que les zones arides). Les caractéristiques des nuages (e.g. leur épaisseur optique), ainsi que l'heure de la journée ont aussi un impact sur les variations de températures causées par les nuages.

En s'appuyant sur les produits issus de l'instrument SEVIRI décrits ci-dessus, une méthodologie adaptée aux spécificités des observations microondes est développée dans la section suivante pour effectuer une désagrégation des T . Cette méthode doit être applicable globalement, quelle que soit l'heure locale de l'observation ou la couverture nuageuse de la scène.

3 Description de la méthode de désagrégation des températures de surface issues d'observations microondes

La section suivante est partiellement issue d'une publication : Favrichon, S. ; Prigent, C. ; Jiménez, C. 2021. "A Method to Downscale Satellite Microwave Land-Surface Temperature" Remote Sensing 13, no. 7 : 1325. <https://doi.org/10.3390/rs13071325>

Pour obtenir un produit de T microonde à haute résolution, la procédure décrite ici repose sur une relation statistique entre des informations auxiliaires à haute résolution et les estimations de T . Cette fonction statistique est d'abord dérivée pour des conditions de ciel clair en utilisant la T IR et un certain nombre de données auxiliaires obtenues par des observations infrarouges de SEVIRI. Les estimations originales de T à 5 km par SEVIRI sont utilisées comme résolution fine, la résolution grossière requise pour tester la méthode étant produite synthétiquement à partir de l'agrégation des T_{SEVIRI} originales à une résolution de 25 km.

Pour les scènes nuageuses, la même fonction statistique sera utilisée, mais avec une correction préalable des données auxiliaires pour tenir compte des conditions nuageuses. Cette correction (décrite dans la section 4) résulte de l'étude des relations entre les données auxiliaires utilisées pour la désagrégation et les variables de contenu en eau liquide dans la colonne d'atmosphère (*total column liquid water*) liées à la présence de nuages dans la réanalyse ERA5. Enfin, la désagrégation avec la correction des nuages est illustrée en appliquant la méthode à des températures de surface produite par SSMIS (à une résolution de 25 km) pour estimer les températures de surface à une résolution de 5 km.

3.1 Recherche d'une relation statistique entre les données aux différentes échelles

La désagrégation statistique repose sur des informations à la haute résolution spatiale visée. Comme mentionné dans la section précédente, l'objectif est de développer une méthode qui soit suffisamment flexible pour être applicable aux différentes heures locales de passage d'un satellite. Il est donc nécessaire de disposer d'informations de surface échantillonnées à différents moments de la journée, ce qui restreint le choix de la plateforme aux satellites géostationnaires. SEVIRI est choisi ici pour fournir des mesures toutes les 15 minutes, entre $\pm 60^\circ\text{N}$ et $\pm 60^\circ\text{E}$. Lors du développement de la fonction de désagrégation, ces estimations constitueront la résolution spatiale fine ciblée (*High Resolution - T_{HR}*), tandis que l'agrégation par une simple moyenne des estimations T à ~ 5 km sera utilisée comme résolution synthétique grossière (*Low Resolution - T_{LR}*) pour tester la désagrégation.

La couverture spatiale de SEVIRI n'englobe pas les latitudes extrêmes, mais la diversité des situations est jugée suffisante pour développer et tester la méthodologie. Un mois par saison (février/mai/août/novembre) est téléchargé de l'archive du LSA SAF pour l'année 2016 afin de représenter la variabilité saisonnière. Le choix des données auxiliaires utilisées est discuté dans la section 3.2, après l'exposé du cadre théorique de la méthode.

On peut expliciter un certain nombre d'hypothèses sur lesquelles reposent cette méthodologie :

- ✓ L'évolution des températures est similaire à toutes les échelles, c'est-à-dire que les relations estimées à l'échelle globale sont valables à l'échelle locale d'un pixel à basse résolution.
- ✓ On suppose aussi que la température estimée par les observations microondes est la même que celle estimée dans le domaine de l'infrarouge. Même si dans certaines situations, cette hypothèse est erronée (chapitre IV).

Cette section détaille les notations employées pour la désagrégation en utilisant un modèle empirique de description des structures de température à haute résolution. Cela consiste à estimer les T à haute définition $T_{downscaled}$, à partir des T_{LR} mesurées à basse résolution et d'informations auxiliaires $A_{HR}^{(i)}$ à haute résolution. On a dans ce cas T_{LR} qui s'écrit comme la moyenne des T_{HR} :

$$T_{LR} = \frac{1}{N} \sum_{i=1}^N T_{HR}^{(i)}. \quad (\text{V.1})$$

Dans la figure V.3, une observation à basse résolution T_{LR} est découpée en 25 pixels à haute résolution $T_{HR}^{(i)}$ ($N=25$), par homogénéité avec les produits de température issus

de SSMIS et de SEVIRI à respectivement ~ 25 et ~ 5 km.

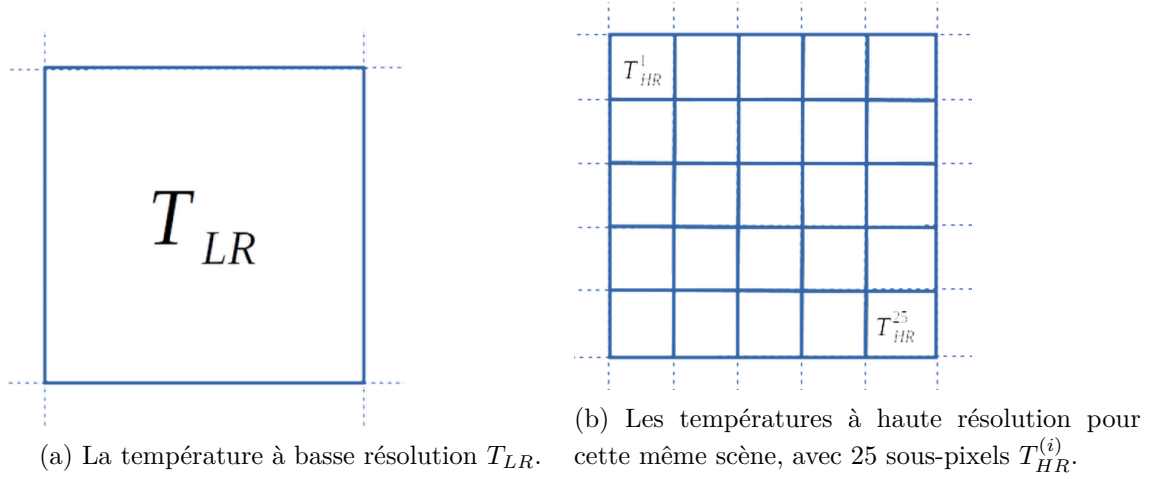


FIGURE V.3 – Schéma représentant les températures de surface à différentes résolutions.

La désagrégation repose sur l'hypothèse qu'une relation statistique générale F peut relier les informations auxiliaires $A_{HR}^{(i)}$ et les $T_{HR}^{(i)}$. Cette relation permet de paramétrer un modèle spatial à haute résolution, c'est-à-dire :

$$\tilde{T}_{HR}^{(i)} = F(A_{HR}^{(i)}). \quad (\text{V.2})$$

Pour chaque pixel haute résolution, une correction $\Delta\tilde{T}_{HR}^{(i)}$ sera appliquée à T_{LR} pour décrire la structure spatiale fine :

$$\tilde{T}_{HR}^{(i)} = T_{LR} + \Delta\tilde{T}_{HR}^{(i)}. \quad (\text{V.3})$$

On peut montrer que cette correction est une fonction de $A_{HR}^{(i)}$:

$$\Delta\tilde{T}_{HR}^{(i)} = \tilde{T}_{HR}^{(i)} - T_{LR} = \tilde{T}_{HR}^{(i)} - \frac{1}{N} \sum_{j=1}^N \tilde{T}_{HR}^{(j)} = F(A_{HR}^{(i)}) - \frac{1}{N} \sum_{j=1}^N F(A_{HR}^{(j)}). \quad (\text{V.4})$$

Cette équation décrit l'augmentation ou la diminution de T pour un pixel à haute résolution i donné. Cet écart calculé correspond au motif à haute résolution : c'est la différence entre la valeur de la fonction F calculée pour le pixel i et la valeur moyenne des températures calculées par F pour tous les sous-pixels j dans le pixel à basse résolution.

On peut alors appliquer globalement la correction obtenue par l'équation V.4, pour obtenir une température à haute résolution sur un pixel $T_{downscaled}$:

$$T_{downscaled} = \Delta\tilde{T}_{HR}^{(i)} + T_{LR} + \text{lissage} \quad (\text{V.5})$$

L'application de cette équation permet de déterminer la température à haute résolution produite par la méthode d'augmentation de la résolution, le terme de lissage est précisé dans la section 3.3.

3.2 Sélection des données auxiliaires

On veut donc sélectionner les sources d'information les plus pertinentes pour contraindre la fonction F et donner une estimation la plus précise possible de la structure spatiale fine des températures. En se basant sur les résultats présentés dans les études précédentes

(section 2.2), sur la compréhension des phénomènes qui contraignent l'évolution de la température de surface et sur les performances du modèle, on peut sélectionner un certain nombre de données auxiliaires pour obtenir la « meilleure » performance du modèle F . Ces variables ont été introduites dans la section précédente, celles utilisées dans le modèle sont listées ci-dessous.

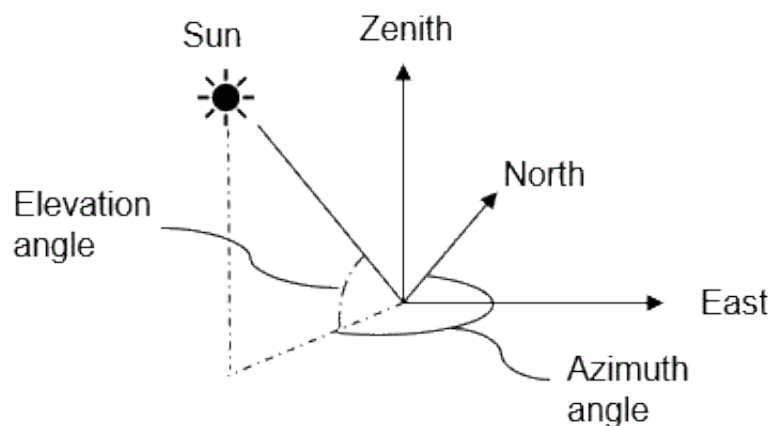


FIGURE V.4 – Schéma représentant l'angle d'élevation solaire pour un point à la surface de la Terre dans l'hémisphère nord.

- Pour caractériser la couverture végétale de surface, la moyenne sur 10 jours de la FVC de SEVIRI est sélectionnée. Elle représente la fraction de végétation verte par unité de surface horizontale du sol (García-Haro et al., 2013).
- Le LSA SAF fournit également un certain nombre de paramètres dont D_a et T_0 , qui décrivent le cycle diurne médian des températures (T) sur la même période de 10 jours. Ce cycle médian est estimé à partir des températures de SEVIRI, en supposant que le cycle diurne suit le modèle décrit dans Göttsche and Olesen (2009), introduit dans la section précédente.
- La configuration spatiale des températures T change avec l'heure locale. Par exemple, les rivières et les zones arides peuvent présenter des T similaires la nuit, mais des T très différentes à midi. En raison de sa capacité thermique, T pour de l'eau est plutôt stable pendant la journée, mais T sur la zone aride varie fortement. Afin de prendre en compte les changements de la distribution spatiale des T en fonction de l'heure locale, une variable liée au rayonnement solaire est ajoutée comme quatrième entrée à la fonction F : l'angle d'élevation du soleil au-dessus de l'horizon est calculé, comme représenté sur la figure V.4. Son sinus ($\sin(\alpha_S)$) est utilisé comme prédicteur continu du modèle spatial, pour tenir compte de l'heure locale et de la saison. Il présente l'avantage de donner une indication de l'ensoleillement reçu par unité de surface, et donc représente partiellement le cycle diurne. Il est aussi symétrique par rapport à l'équateur et varie en fonction des saisons. Cet angle est toujours défini par rapport à l'heure locale.

Les variables FVC , T_0 , et D_a distribuées par le LSA SAF à une résolution spatiale de ~ 5 km ainsi que le sinus de l'angle d'élevation du soleil calculé pour chacun des pixels sont utilisées pour établir la relation F .

La figure V.5 montre la distribution conjointe de ces variables pour illustrer leurs relations avec T . Chaque variable est groupée par intervalles pour représenter la distribution d'une autre variable sous différentes conditions. Les quartiles supérieurs et inférieurs (lignes pointillées) ainsi que les extrema de chaque distribution sont indiqués. Une corrélation positive entre T et T_0 est évidente (en haut à gauche). Une tendance similaire n'est pas visible entre T et D_a (en haut à droite) : une faible T_0 est associée à une faible température pendant la journée, mais il n'y a pas de corrélation claire entre D_a et T . Ceci est attendu si l'on considère que les températures élevées sont rencontrées dans des situations contrastées telles que la forêt équatoriale (faible D_a) et les régions désertiques (forte D_a). Ceci est corroboré par la corrélation négative entre D_a et FVC (en bas à droite) liée à la plus faible amplitude du cycle diurne sur les zones végétalisées et par l'absence de corrélation entre T et FVC (en bas à gauche). Ce dernier point montre que la végétation ne peut pas être la seule source d'information pour distinguer les températures chaudes des températures froides, confirmant la nécessité pour dériver la correction $\Delta\tilde{T}_{HR}^{(i)}$ d'avoir plus de prédicteurs qu'un produit lié seulement à la présence de végétation.

3.3 Implémentation du modèle d'estimation de la structure spatiale

La fonction statistique F a pour but de modéliser la relation entre les 4 données auxiliaires et T_{HR} . Cette estimation par F est utilisée pour calculer les écarts entre les pixels à haute résolution au sein d'une observation à basse résolution. Ces écarts forment le motif des températures à haute résolution. Ce motif est ensuite ajouté à T_{LR} à basse résolution. La sortie de la fonction F est une température qui est utilisée dans l'équation V.4 pour fournir la structure spatiale (correction) des températures haute résolution pour un pixel donné.

Ici, le modèle implémenté est un réseau de neurones multicouches au fonctionnement similaire à celui introduit au chapitre III. La structure contient 4 neurones d'entrée, 1 couche cachée avec 10 neurones (avec une fonction sigmoïde pour l'activation), et 1 sortie. En entrée sont les variables FVC , T_0 , D_a , et $\sin(\alpha_S)$, et en sortie les T_{HR} . Une base de données de $\sim 50M$ points des 4 mois de 2016 extraits de l'archive SEVIRI sont utilisés pour entraîner, valider et tester le modèle.

Le disque SEVIRI est centré sur l'équateur en Afrique, et avant l'entraînement la distribution originale de la base de données est ré-échantillonnée pour éviter une sur-représentation des situations désertiques (avec une végétation très faible) et un meilleur échantillonnage des angles d'élévation solaire pour lesquels les valeurs basses sont plus fréquentes. Cette sur-représentation des faibles angles d'élévation solaire est due au faible nombre de points pour lequel le soleil est directement au zénith et où l'angle vaut 90° (uniquement entre les tropiques, pour certains jours de l'année).

Les résidus, c'est-à-dire l'écart quadratique moyen (*Root Mean Square Error* - RMSE) entre la cible T_{HR} et la sortie du réseau de neurones après l'entraînement, sont affichés pour la plage de T couverte par la base de données d'entraînement sur la figure V.6.

Pour illustrer la pertinence du choix des entrées, des réseaux de neurones similaires par leurs hyper-paramètres (taille de réseau, nombre d'itérations pour l'entraînement) sont créés en utilisant des combinaisons de T_0 , D_a , FVC et $\sin(\alpha_S)$ en entrée et leurs résidus sont affichés. Sur l'ensemble de la plage de T , le réseau de neurones utilisant les quatre prédicteurs affiche une performance plus régulière, en particulier pour les valeurs très basses et élevées de T . Il est intéressant de noter que l'indice de végétation ne semble

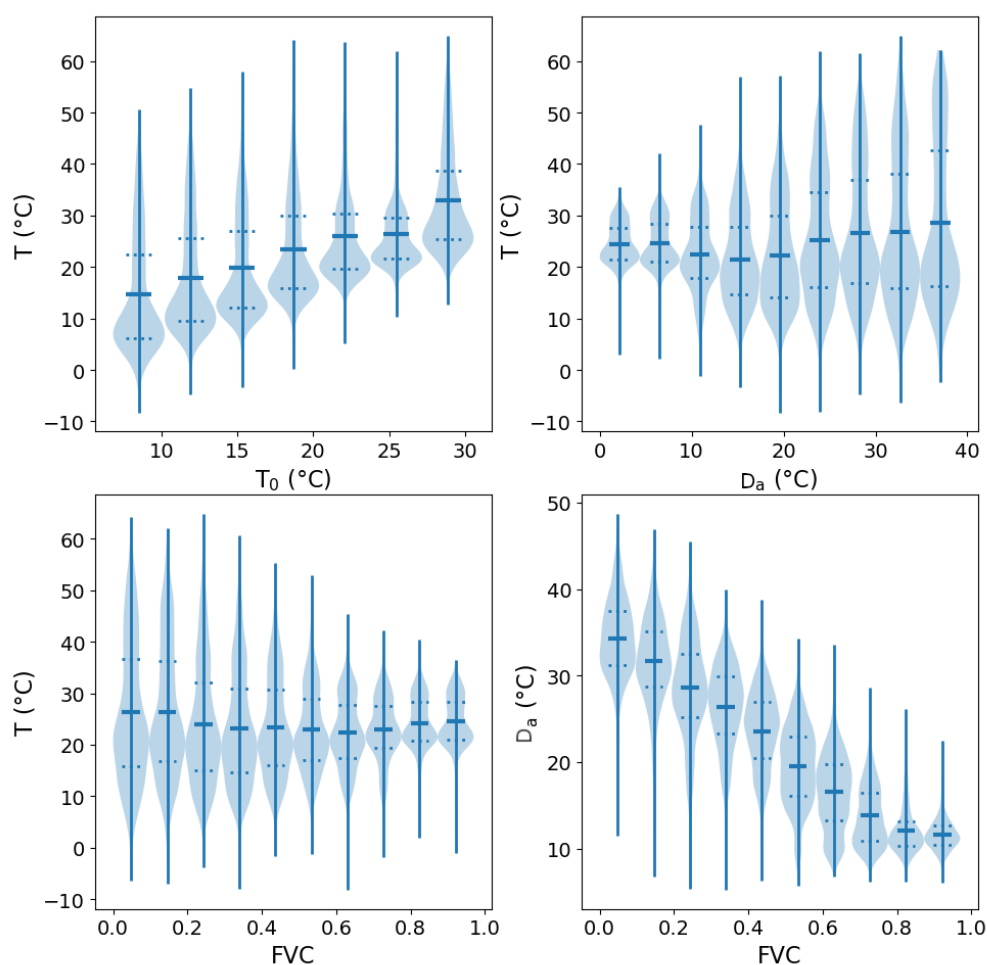


FIGURE V.5 – Température de surface à haute résolution T en fonction des variables auxiliaires à haute résolution proposées comme prédicteurs pour la désagrégation. T en fonction de la température minimale (T_0) (en haut à gauche), en fonction de l’amplitude du cycle diurne (D_a) (en haut à droite), en fonction de la couverture végétale fractionnelle (FVC) (en bas à gauche). L’évolution de D_a en fonction de FVC est également représentée (en bas à droite). Pour chaque figure, la distribution de la variable de l’axe des y pour un intervalle de valeur de la variable de l’axe des x est représentée en bleu. De plus, la moyenne (ligne solide), les quartiles supérieurs et inférieurs (lignes pointillées) sont indiqués, ainsi que les extrema de la variable.

pas suffire à lui seul à fournir une distinction claire entre les températures, pour toutes les heures de la journée.

Il est également à noter que, même pour le réseau de neurones avec quatre entrées, les résidus de la fonction « semblent » élevés. Toutefois, cela est prévisible, car l’objectif de cette fonction n’est pas de prédire la température à haute résolution exacte à partir des quatre variables auxiliaires, mais plutôt d’informer sur les différences locales relatives liées au contenu informatif de ces entrées.

Il convient de noter que de manière similaire à l’étude de Aires et al. (2017), une procédure de lissage est également appliquée en considérant que la température d’un pixel LR donné résulte aussi d’une contribution des T des pixels basse résolution voisins. Ainsi, la contribution de chaque pixel LR est moyennée spatialement avec un poids proportionnel

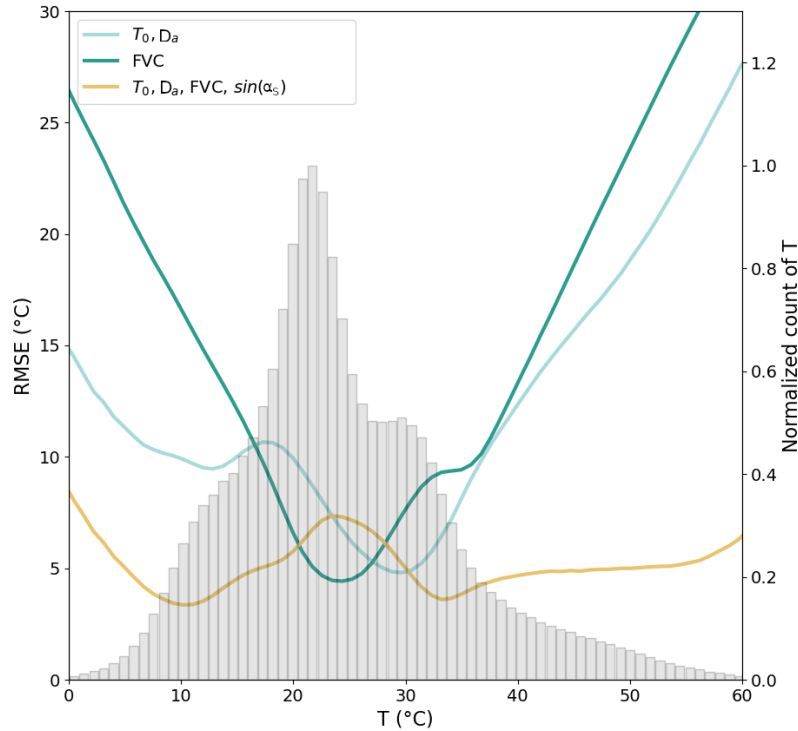


FIGURE V.6 – Racine carrée de l’erreur quadratique moyenne (RMSE) de la T estimée avec la fonction F pour différentes combinaisons de prédicteurs en fonction de T . En bleu F est entraîné avec T_0 et D_a , uniquement en utilisant FVC en vert, et enfin en utilisant T_0 , D_a , FVC , et $\sin(\alpha_S)$ en marron. La distribution de T dans le jeu de test est indiquée par les barres grises (échelle sur l’axe de droite).

à sa distance au pixel haute résolution considéré. La contribution partielle des pixels LR voisins est pertinente physiquement étant donné que les mesures réelles proviennent d’une tache au sol qui peut être plus large qu’un unique pixel.

3.4 Application à des données synthétiques à basse résolution

Avec l’implémentation de la fonction F sur les données disponibles, il est possible de mettre en place des tests de la méthodologie pour des cas en ciel clair. Pour ces tests, les estimations synthétiques de T_{LR} à ~ 25 km sont produites en faisant la moyenne des T infrarouge SEVIRI originales à ~ 5 km sur des fenêtres de 5×5 pixels. Cela permet d’évaluer la méthode de désagrégation indépendamment des différences potentielles entre les températures microondes et IR. Cela permet aussi de tester la méthodologie à différents moments de la journée, car le cycle diurne complet est observé par l’instrument IR géostationnaire (au contraire des satellites polaires héliosynchrones microondes avec seulement deux passages à heure fixe au-dessus d’un point par jour).

Tout d’abord un test sur une scène à deux saisons avec des températures à basse résolution différentes sont utilisées. Une scène observée par SEVIRI est sélectionnée dans le sud de l’Afrique (16.7°S 21.0°E à 19.8°S 24.7°E), dans la région de l’Okavango qui présente de fortes hétérogénéités de surface. Les structures spatiales marquées sont causées par le delta intérieur de l’Okavango, où la présence d’eau et la variabilité de la végétation ont un impact sur les T . Les résultats sont présentés pour deux saisons et deux moments contrastés de la journée, en février (11 février 2016) à 05:00 et en août (13 août 2016) à

12:00.

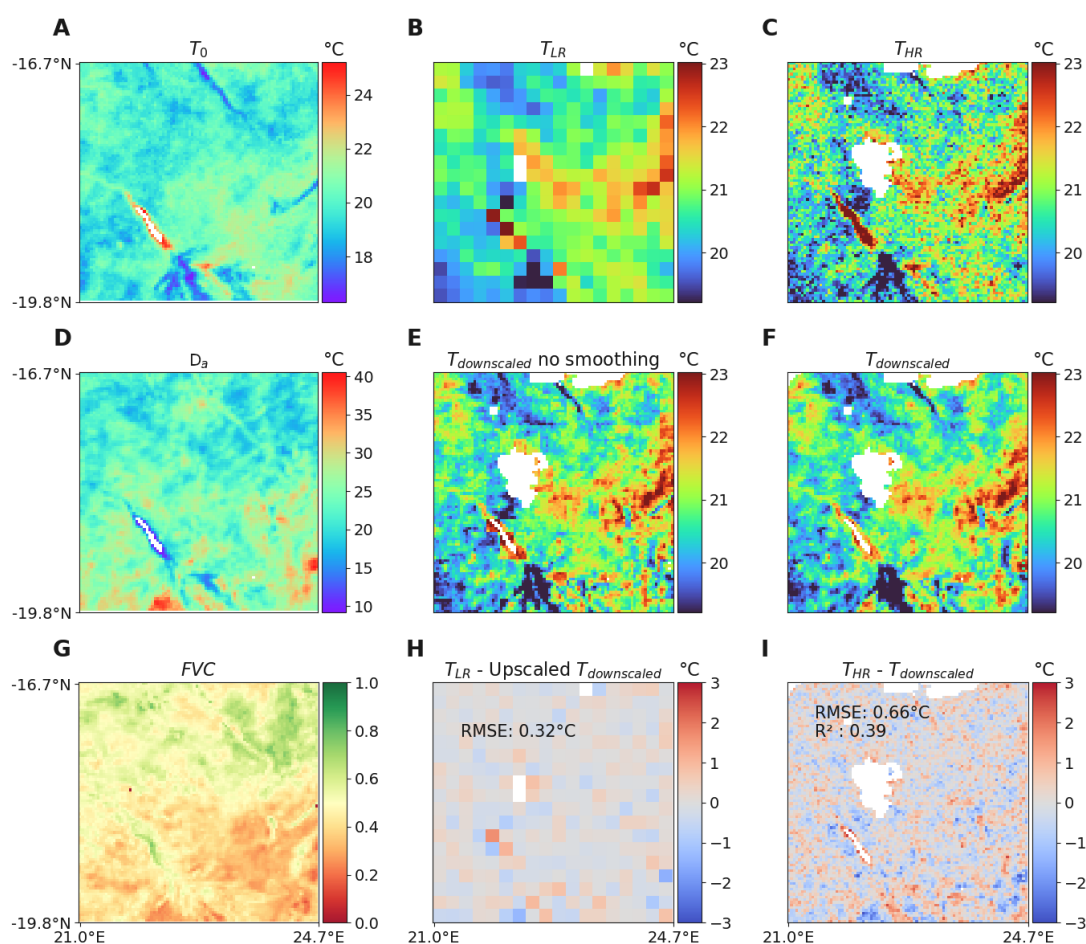


FIGURE V.7 – Différentes étapes de la méthodologie appliquées à une scène du 11 février 2016 à 05:00 au-dessus du delta de l'Okavango (16.7° S 21.0° E à 19.8° S 24.7° E). Les variables auxiliaires d'entrée (T_0 , D_a , FVC) sont présentées en A, D et G. Les T à basse résolution et à haute résolution originales dans les figures B et C. Les T obtenues par la méthode de désagrégation, avant et après lissage, dans la deuxième ligne (E et F). La figure H contient la différence entre T en entrée et $T_{downscaled}$ moyennée (5×5) à basse résolution avec l'erreur RMSE superposée sur la scène. Enfin, la figure en bas à droite (I) montre la différence entre T_{HR} et $T_{downscaled}$ produite par la méthodologie. Les valeurs intégrées de RMSE et de coefficient de détermination (R^2) sont indiquées en noir en H et I.

Pour les deux figures V.7 et V.8, les trois sous-figures de gauche (A, D, G) montrent les cartes des variables auxiliaires T_0 , D_a et FVC . Des différences dans les structures spatiales des trois variables sont visibles, avec un comportement opposé entre D_a et T_0 autour du delta intérieur. La différence de couverture végétale entre les 2 scènes sélectionnées (à des saisons différentes) est également forte. La sous-figure B montre les T_{LR} calculées à partir des températures infrarouges, chaque pixel représentant environ $25 \times 25 \text{ km}^2$, tandis que la sous-figure C affiche la T originale de SEVIRI (T_{HR}), c'est-à-dire la T cible à reproduire par la désagrégation. La gamme de température diffère entre les 2 scènes sélectionnées : la température moyenne est d'environ 20°C à 05:00, et environ 40°C à 12:00. Comme T_{LR} est construit synthétiquement à partir de T_{HR} , certaines valeurs sont manquantes en présence de nuages ou de valeurs invalides (par exemple pas d'estimation de T sur les surfaces d'eau).

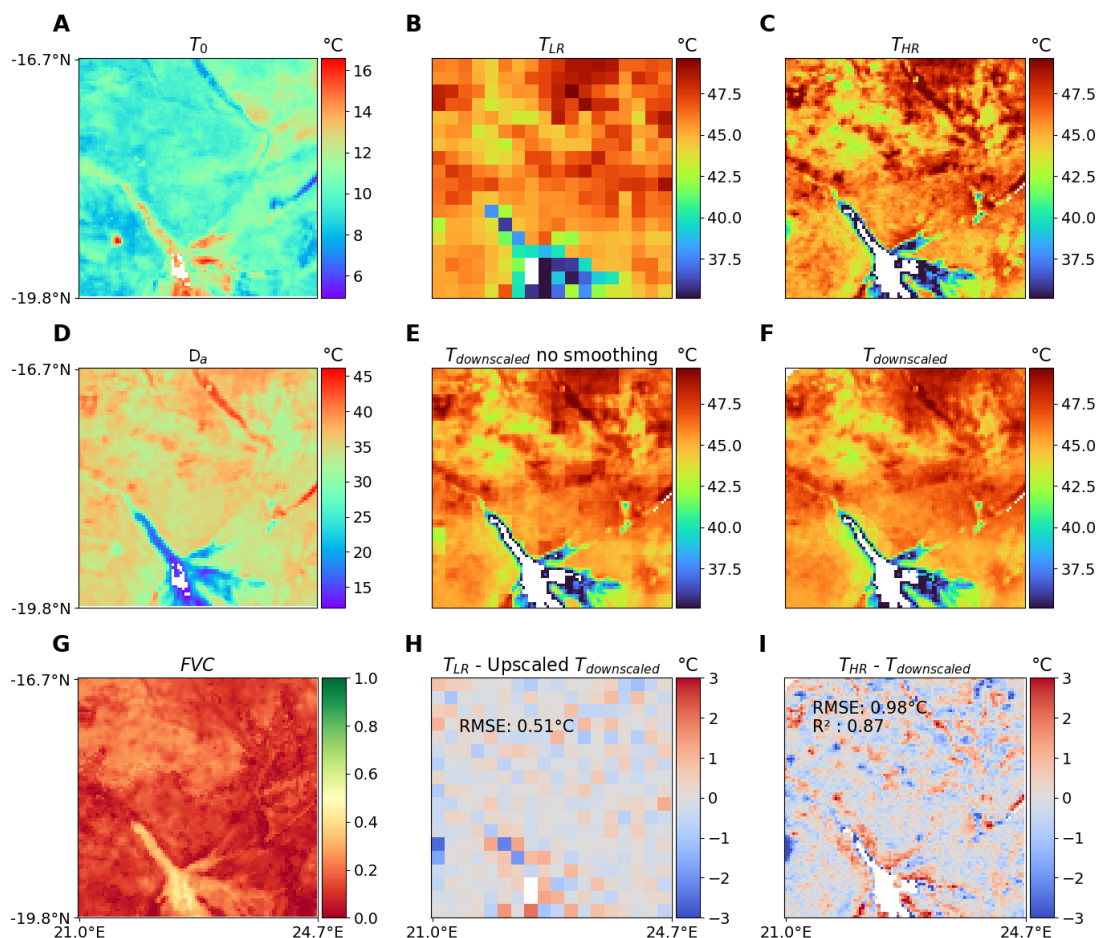


FIGURE V.8 – Même figure que la figure V.7 pour le 13 aout 2016 à 12:00.

La sous-figure E montre la température après la désagrégation mais avant l'étape de lissage ($T_{downscaled\ no\ smoothing}$) et la sous-figure F les T finales ($T_{downscaled}$) après lissage. L'effet de lissage est apparent par la disparition des artefacts en forme de blocs sur le produit final. Par rapport à T_{HR} , $T_{downscaled}$ apparaît plus lisse, avec moins de contraste et quelques structures spatiales manquantes. La procédure de lissage peut introduire des erreurs suivant la distribution des T_{HR} . La sous-figure I affiche la différence entre la T_{HR} originale (figure C) et $T_{downscaled}$ (F) : c'est le résidu de la méthode de désagrégation. Les pixels avec des écarts importants se trouvent principalement autour des valeurs extrêmes locales de T_{HR} . Cette moindre performance sur les valeurs extrêmes est inhérente à l'entraînement du réseau de neurones utilisé pour la fonction F . La minimisation de l'erreur quadratique moyenne conduit à une sous-estimation de la variance en sortie qui peut parfois être traitée avec des méthodes d'inflation de la variance (Von Storch, 1999). Certaines erreurs importantes apparaissent également près des valeurs manquantes, par exemple près de l'étendue d'eau dans le delta.

Dans la sous-figure H, la moyenne spatiale de $T_{downscaled}$ est calculée afin d'obtenir un produit à basse résolution qui peut être comparé au T_{LR} synthétique (figure B) utilisé comme entrée de la méthode de désagrégation. Cette comparaison met en évidence les zones pour lesquelles la désagrégation dégrade les informations de température fournies en entrée. C'est un critère important pour juger de la qualité d'une méthode. Les seules zones présentant des écarts importants sont celles qui se situent autour des données manquantes ou avec un fort contraste avec les pixels voisins.

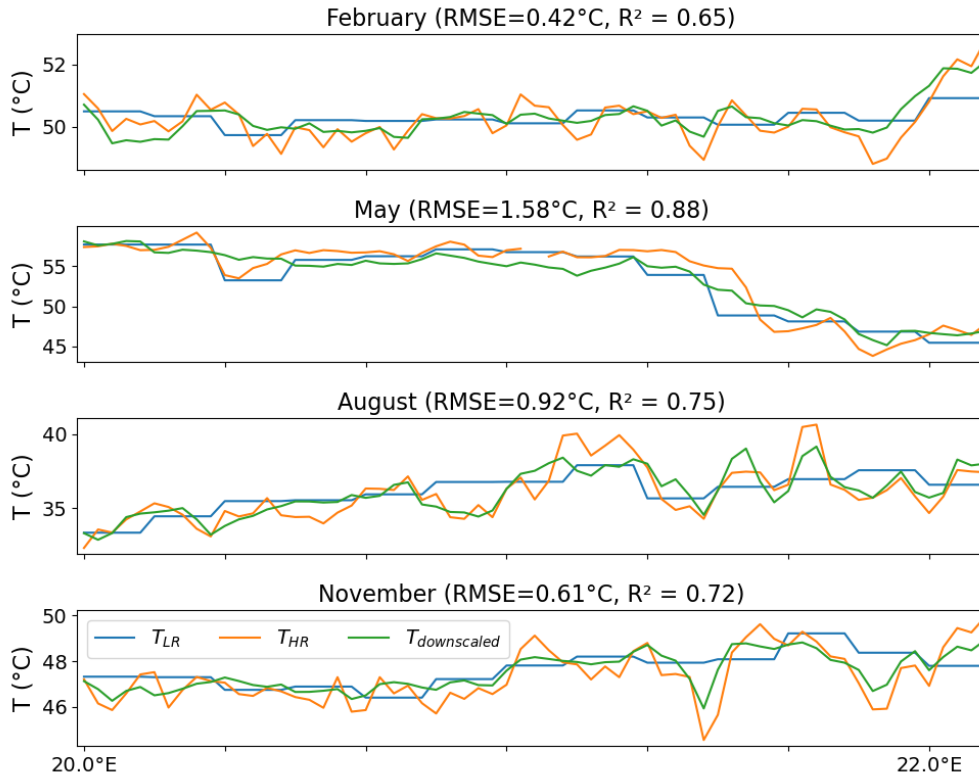


FIGURE V.9 – Coupe latitudinale de la température autour 12.5°N entre 20°E et 22°E à 10:00 le 18 des mois de février, mai, août et novembre (haut en bas). Pour chaque mois, la température SEVIRI à haute résolution T_{HR} est en orange, la moyenne (5×5) T_{LR} est en bleu et $T_{downscaled}$ en vert. Pour chaque graphique, la RMSE et le coefficient de détermination R^2 entre $T_{downscaled}$ et T_{HR} sont affichés dans le titre.

Les résultats de la désagrégation sont aussi mis en évidence sur la figure V.9 où est représentée une coupe le long de la latitude à 12.5°N de $T_{downscaled}$, de T_{HR} et de la moyenne spatiale de l'entrée T_{LR} . La coupe est intégrée spatialement entre 12.4°N et 12.7°N , s'étendant de 20°E à 22°E pour le 18 des mois de février/mai/août/novembre 2016 à 10:00. La variabilité présente dans T_{HR} est amortie dans T_{LR} , tandis qu'elle est bien représentée dans $T_{downscaled}$, avec des valeurs de corrélation relativement importantes (coefficients de détermination R^2 entre 0.65 et 0.88). La méthode de désagrégation fonctionne bien pour une plage de températures entre 30°C et 60°C . La sous-estimation des extrema déjà mentionnée est également visible sur la coupe.

La figure V.10 reprend les mêmes dates que la figure précédente sur une géographique couvrant une zone entre 35°S - 35°N et 22 - 35°E et calcule pour chaque heure :

1. L'erreur moyenne entre $T_{downscaled}$ (résultat de la méthode de désagrégation) et T_{HR} les températures de SEVIRI (en trait pointillé), noté HR dans la légende.
2. L'écart entre T_{LR} la valeur moyennée et T_{HR} (en trait plein), noté LR dans la légende.

Cette figure permet de mettre en évidence que le bénéfice lié à la désagrégation est faible durant la nuit quand les températures sont proches de la température minimale sur le cycle. Par contre, durant le jour les hétérogénéités spatiales de température sont plus fortes et le bénéfice apporté par la désagrégation est très clair pour tous les mois étudiés.

Les exemples précédents montrent que la fonction F reproduit correctement les varia-

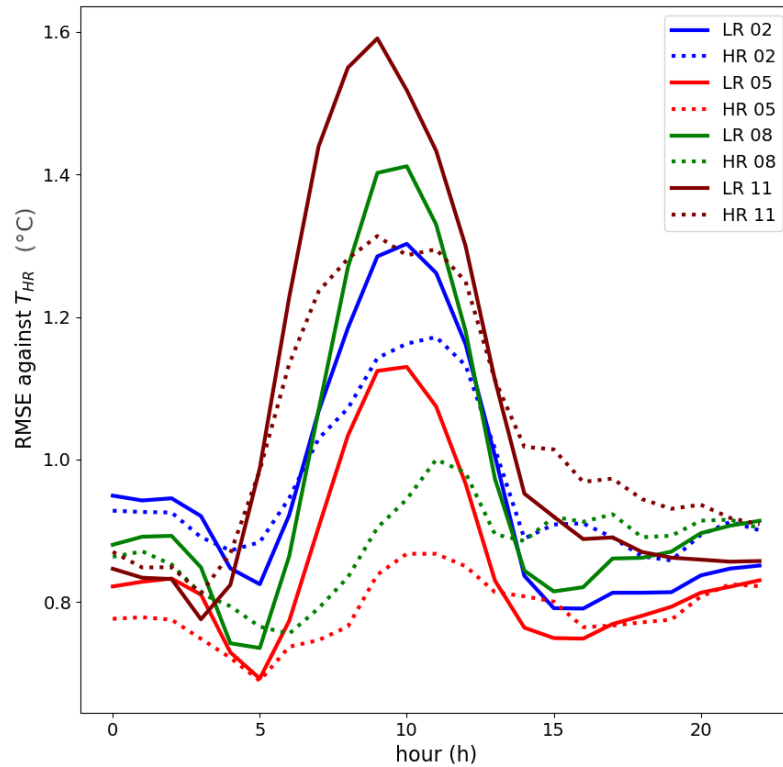


FIGURE V.10 – Figure présentant la racine carrée de l’erreur quadratique moyenne entre $T_{downscaled}$ et la température originale T_{HR} de SEVIRI (trait pointillé - HR) et entre T_{LR} et T_{HR} (trait continu - LR). Le calcul est effectué pour une bande couvrant des latitudes entre 35°S - 35°N et des longitudes entre 22 - 35°E pour le 18 des mois de février (02), mai (05), août (08) et novembre (11), en fonction des heures de la journée.

tions de T à fine échelle, bien qu’elle soit moins performante pour les valeurs de T les plus extrêmes au sein des scènes. Dans l’ensemble, la désagrégation produit une bonne approximation des températures à haute résolution. Elle s’adapte également bien aux différentes saisons, heures de la journée et conditions de surface. L’écart quadratique moyen (RMSE) entre T_{HR} et $T_{downscaled}$ a été calculé pour le 18^{ème} jour des mois de février/mai/août et novembre 2016 à toutes les heures entre 35°S - 35°N et 22 - 35°E . Les valeurs de la RMSE sont comprises entre 0.62°C en mai et 1.07°C en novembre. La RMSE calculée entre T_{HR} et la moyenne spatiale de T_{HR} se situe entre 0.74°C et 1.16°C , confirmant l’efficacité de la méthode de désagrégation.

4 Prise en compte de l’effet des nuages dans le modèle de désagrégation

La section précédente a mis en place le cadre théorique et a exploré les résultats obtenus sur des données synthétiques. Cependant, les tests effectués ne s’appliquent que sur des données en ciel clair. Comme l’utilisation des températures issues des microondes trouve son principal bénéfice dans les situations nuageuses, il faut étudier l’applicabilité de cette méthodologie dans ces situations. La première étape consiste à analyser l’effet de la présence d’un nuage sur les températures de surface, puis dans un second temps de

proposer une méthode permettant d'estimer une correction liée aux nuages pouvant être prise en compte dans la méthodologie de désagrégation.

4.1 Analyse de l'effet des nuages sur les températures de surface

Intuitivement, on s'attend à obtenir des températures de surface de jour plus faible quand le rayonnement provenant du soleil est bloqué par les nuages. De nuit, une plus forte concentration de nuages devrait créer une plus forte réflexion du rayonnement émis par la surface et donc moins de refroidissement.

Détailler les effets de la présence des nuages sur le rayonnement a été abordée par de nombreux modèles (Beersma et al., 2002; Morcrette et al., 2008), mais ils sont complexes à mettre en œuvre. Pour estimer cet impact, on peut utiliser des données de température de surface obtenues dans les microondes avec un masque indiquant les observations nuageuses et celles en ciel clair. Par exemple, le travail de Ermida et al. (2019) fournit des comparaisons de température de surface microonde avec ou sans nuages. De la même façon les données des réanalyses fournissent des températures de surface pour différentes conditions nuageuses.

Dans les données de Ermida et al. (2019), la température de surface est estimée pour les observations microondes AMSR-E en utilisant la méthodologie de Jiménez et al. (2017). La base de données contient des moyennes trimestrielles de température entre 2008 et 2010. La présence de nuages provient du produit de niveau 3 du satellite MODIS sous la forme d'une fraction de couverture nuageuse agrégée à une résolution de 1° . Toutes les données sont donc colocalisées à cette résolution. Enfin, pour discriminer les résultats en fonction du type de couverture du sol on peut utiliser des données Globcover (Arino et al., 2012) qui fournissent une classification statique des sols entre 85°N et 60°S .

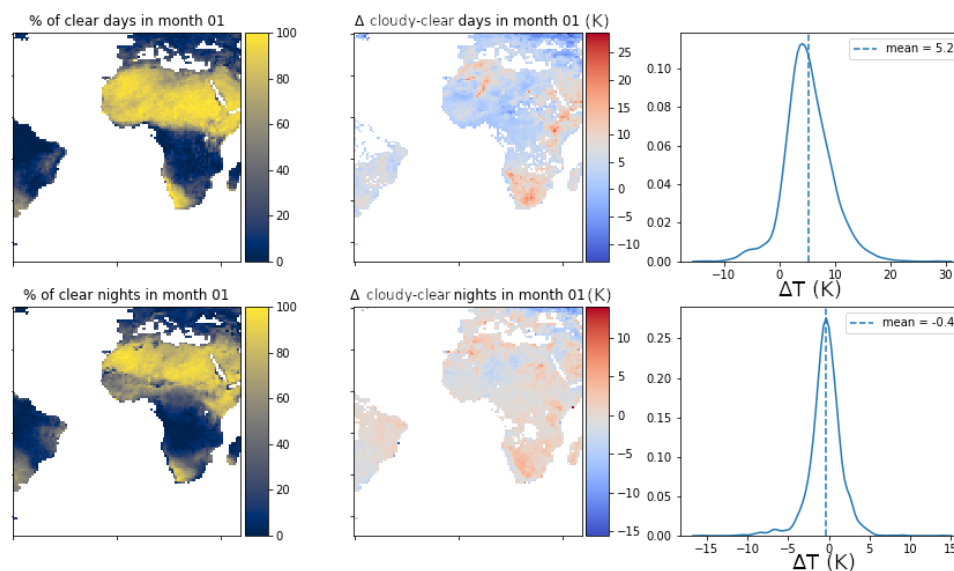


FIGURE V.11 – Comparaison de la répartition spatiale de la fraction de couverture nuageuse moyenne MODIS (première colonne), des écarts de température (K) entre ciel nuageux et ciel clair pour chaque pixel et la distribution des écarts de température entre ciel clair et nuageux (K) pour les données de température de surface issues d'AMRS-E pour le passage de jour (première ligne) et de nuit (seconde ligne) en janvier entre 2008 et 2010 dans les données de Ermida et al. (2019).

CHAPITRE V. AMÉLIORATION DE LA RÉOLUTION SPATIALE DES ESTIMATIONS DE TEMPÉRATURE DE SURFACE

En analysant cette base de données, on peut comparer les cartes de répartition de la présence de nuages mois par mois, et la distribution des températures entre les pixels en ciel clair et ceux nuageux. La figure V.11 affiche, pour le passage de jour (13:30 ligne du haut) et de nuit (01:30 ligne du bas) de AMSR-E, en janvier entre 2008 et 2010 l'écart entre les températures de surface estimées pour un ciel clair et sous un ciel nuageux. Les figures sont de gauche à droite, la fraction moyenne mensuelle de couverture nuageuse dans les données MODIS, puis l'écart de température pixel par pixel entre les jours nuageux et clairs et enfin la distribution de ces écarts pour tous les pixels présentés.

Les résultats montrent, outre la distribution inégale de la présence des nuages dans le champ d'observation, la distribution des écarts de température entre les jours clairs et nuageux. De jour, comme attendu, la température est plus chaude de quelques degrés en l'absence de nuages. De nuit, la différence est bien plus faible et les températures de surface en présence de nuages sont en moyenne très légèrement plus chaudes que celles en ciel clair. Ceci est cohérent avec un rayonnement émis par la surface et réfléchi par le nuage, avec une magnitude qui dépend de la hauteur de la base des nuages et de la quantité de nuages (Chen et al., 2000).

On retrouve de tels résultats pour tous les mois de l'année dans la base de données de mesures des T issues de AMSR-E. Pour confirmer et améliorer ces résultats, il est souhaitable de :

1. Pouvoir effectuer cette analyse à des heures différentes. En effet l'écart attendu entre les températures de surface en ciel clair et en ciel couvert peut varier au cours de la journée même en supposant une couverture nuageuse constante.
2. Comprendre l'impact du type de nuage (ou de son opacité) sur les différences observées.
3. Vérifier si la variation est la même pour des surfaces différentes.

Les données de la réanalyse ERA5 peuvent fournir des informations susceptibles de répondre à ces interrogations.

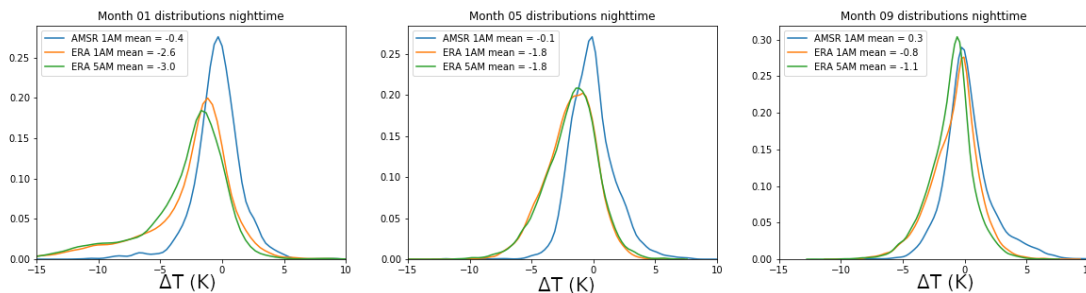


FIGURE V.12 – Comparaison entre les distributions des différences de températures de surface entre ciel clair et ciel nuageux (K) estimées par AMSR-E, avec les nuages provenant de MODIS et ERA5 avec le seuil de $TCLW \geq 0$ pour les mois de janvier, mai et septembre entre 2008 et 2010, pour les températures la nuit (à 01:30 pour AMSR-E et ERA5 et aussi à 05:00 pour ERA5).

La figure V.12 présente les résultats obtenus de l'écart entre les températures de surface des pixels clairs et nuageux pour différentes heures de passage la nuit à différents mois. Les distributions présentent des résultats similaires à ceux obtenus avec la température estimée par Ermida et al. (2019) issue de AMSR-E. Pour les températures mesurées la nuit les données ERA5 à deux heures différentes (01:30 et 05:00) montrent un écart qui semble augmenter plus la mesure est proche de la fin de la nuit, ce qui est conforme à l'effet attendu du nuage.

La réduction de la quantité de rayonnement solaire atteignant la surface de la Terre pendant la journée dépend de l'épaisseur optique du nuage. Au niveau de la surface, les changements de T dépendront essentiellement du type de surface, avec un amortissement attendu de la variabilité spatiale de T sous un ciel couvert par rapport aux conditions de ciel clair. Ceci est illustré dans la figure V.13, où l'impact de la nébulosité (quantité d'eau liquide totale dans la colonne atmosphérique *Total column liquid water TCLW*) provenant des données ERA5 est affichée pour deux surfaces avec une végétation contrastée. La distribution du ΔT , c'est-à-dire la différence de T entre les cas de ciel clair et des nuages fins ($TCLW < 50 \text{ g/m}^2$), ou des nuages épais ($TCLW > 50 \text{ g/m}^2$), est présentée au-dessus de forêts sempervirentes¹ et de savanes issues de la classification Globcover, à différentes heures locales.

Les heures sélectionnées sont proches de l'heure de passage des capteurs SSMIS (06:00/18:00) et AMSR-E (01:30/13:30) et correspondent aussi à des instants de la journée proches du minimum et du maximum du cycle diurne des températures. Les deltas de T de surface entre les cas de ciel clair et de ciel nuageux de nuit (01:00 et 06:00) sont pour le plus souvent négatifs, ce qui confirme que la couverture nuageuse tend à augmenter T , en raison du piégeage du rayonnement thermique émis par la surface. Le ΔT négatif de 06:00 semble être légèrement plus grand que celui de 01:00, ce qui est cohérent avec la baisse des températures de surface jusqu'à ce que le soleil recommence à éclairer la surface. D'autre part, la T diurne tend à diminuer de jour en présence d'un nuage avec un effet plus fort vers le milieu de la journée (13:00), lorsque le cycle diurne de température est proche de son maximum. Des nuages plus épais peuvent également être associés à un effet plus fort, positif ou négatif sur ΔT . Les distributions pour la savane (avec une végétation clairsemée) et la forêt sempervirente (avec une végétation dense) sont présentées sur la figure V.13 et la densité de végétation semble également réduire l'ampleur de l'effet du nuage sur ΔT .

Ces différentes analyses de jeux de données indépendants permettent d'estimer le comportement global des températures sous différentes conditions de couvertures nuageuse. Bien que des effets supplémentaires puissent également affecter les distributions de T , les distributions de la figure V.13 montrent que les variations de la quantité d'eau liquide due au nuages ($TCLW$) et de la végétation peuvent être suffisamment informatives pour une description de premier ordre de l'impact des nuages sur T .

4.2 Paramétrisation de l'impact des nuages dans la désagrégation

Pour pouvoir prendre en compte l'effet des nuages dans le modèle de désagrégation, il faut être capable de modéliser les conséquences de la présence d'un nuage sur les entrées du modèle F . L'étude des changements de T dus aux nuages pour une variété de conditions de surface et de nuages est un défi. Isoler l'effet des nuages nécessiterait une estimation de la T pour la même zone géographique, dans les mêmes conditions de surface et d'insolation, pour différentes heures de la journée et pour une variété de couvertures nuageuses. Ici, l'objectif est uniquement de développer une solution pragmatique qui peut être intégrée systématiquement et facilement dans la méthode de désagrégation. Dans ce but, si les informations disponibles sur les nuages indiquent que les estimations de T_{LR} correspondent à une scène nuageuse, une correction sera appliquée aux données auxiliaires A_{HR} (qui existent pour les scènes en ciel clair) pour tenir compte de l'impact des nuages sur les estimations à haute résolution $\tilde{T}_{HR}^{(i)}$.

1. Forêt dont les plantes gardent leurs feuilles tout au long de l'année - forêts persistantes.

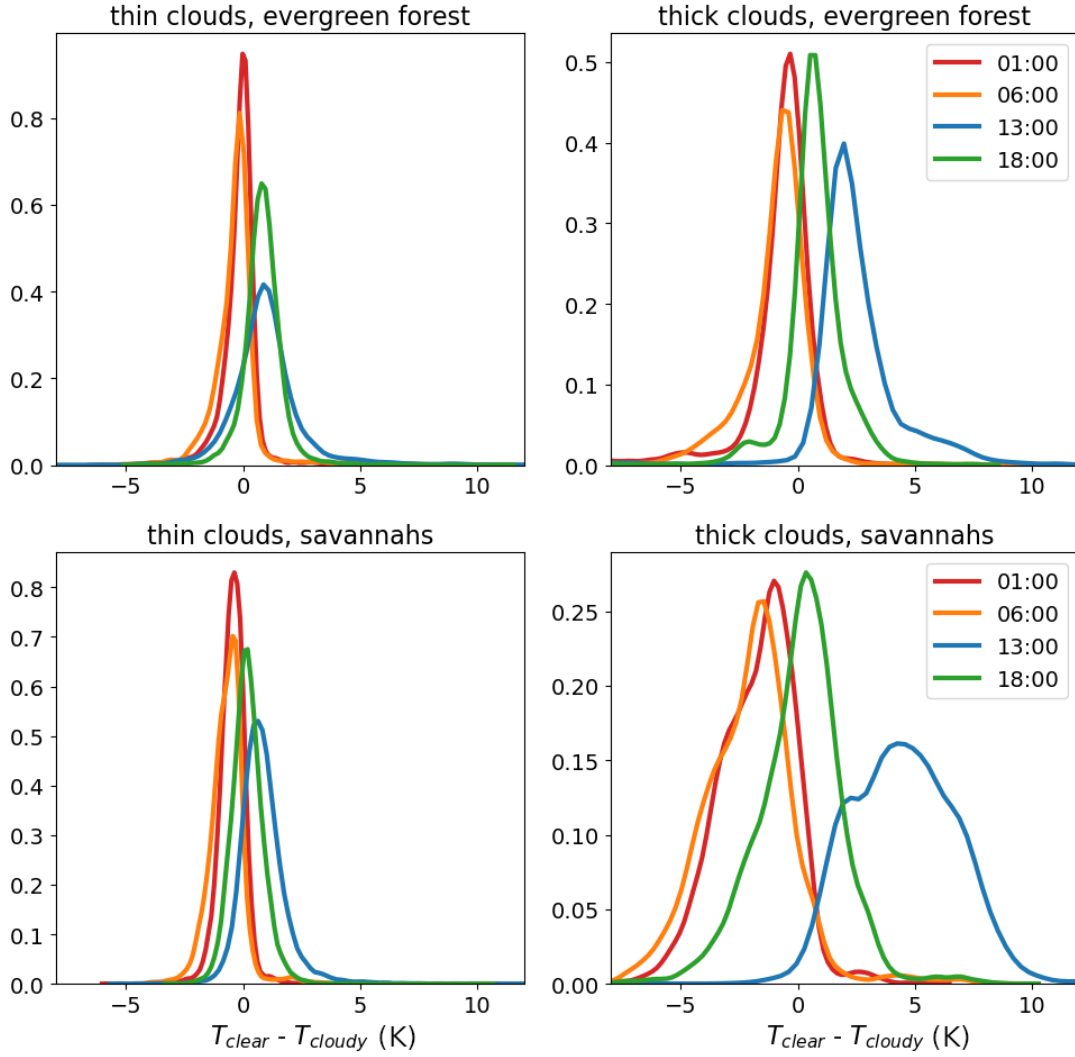


FIGURE V.13 – Distribution de ΔT , la différence entre les températures de surface modélisées dans ERA5 en ciel clair et en présence de nuages en janvier, mai et septembre 2016. La différence de température est calculée entre les cas clairs et ceux avec des nuages fins ($TCLW < 50g/m^2$) pour la première colonne, et entre les cas de ciel clair et ceux avec des nuages épais ($TCLW > 50g/m^2$) dans la seconde colonne. La première ligne montre les distributions au-dessus des forêts sempervirentes et la deuxième ligne pour les observations au-dessus des savanes. Chaque distribution est calculée pour les températures à différents moments de la journée : 01:00, 06:00, 13:00 et 18:00.

La correction $\Delta A_{HR}^{(i)}$ est basée sur une fonction simple C paramétrant l'effet des nuages en fonction des valeurs observées pour une sélection de variables d'entrée $B_{HR}^{(i)}$, que l'on peut exprimer comme suit :

$$\Delta \tilde{A}_{HR}^{(i)} = C(B_{HR}^{(i)}). \quad (V.6)$$

La relation statistique F restant inchangée, la nouvelle correction $\Delta \tilde{T}_{HR}^{(i)}$ pour une scène nuageuse est la suivante :

$$\Delta \tilde{T}_{HR}^{(i)} = F(A_{HR}^{(i)} + C(B_{HR}^{(i)})) - \frac{1}{N} \sum_{j=1}^N F(A_{HR}^{(j)} + C(B_{HR}^{(j)})). \quad (V.7)$$

Afin de disposer d'une variable continue du contenu de la végétation, le *Leaf Area Index* (*LAI*) distribué par ERA5 remplace la classification statique de Globcover utilisé dans la section précédente. Les données ERA5 de température de surface (*skt*), du contenu total en eau liquide dans la colonne atmosphérique (*TCLW*) et de végétation (*LAI*) disponibles sur le globe toutes les heures à une résolution spatiale de ~ 25 km permettent d'étudier ces changements. La résolution est plus faible que celle de SEVIRI, et l'estimation de T par ERA5 ne reproduit pas toujours les T observées (Johannsen et al., 2019). Cependant, les T de la réanalyse et les conditions nuageuses devraient être physiquement cohérentes et suffisamment réalistes pour une première tentative de prise en compte de l'information nuageuse dans la méthodologie proposée.

On a vu dans la figure V.11 que l'effet des nuages sur ΔT (entre les températures en ciel clair ou nuageux) nocturne est négligeable par rapport à celui sur ΔT diurne. L'effet sur T_0 est donc faible. L'effet est le plus fort près du pic du cycle diurne, lorsque la température la plus chaude du cycle est atteinte T_{max} . Par conséquent, une correction sur la température maximale du jour est proposée. Comme $T_{max} = T_0 + D_a$ cela se traduit par une correction de l'entrée D_a de la fonction F dépendante des variations de *TCLW* et de *LAI*.

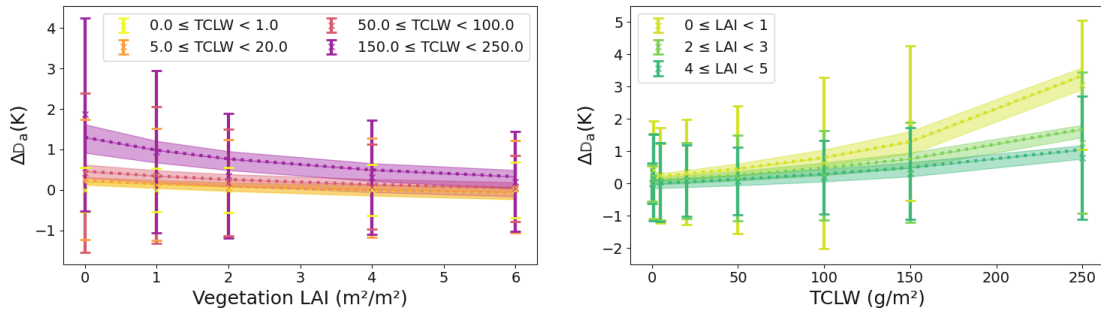


FIGURE V.14 – Différence moyenne d'amplitude des cycles diurnes (K) entre un ciel clair et un ciel nuageux à partir des données ERA5 pour les mois de janvier, mai et septembre 2016 sous forme de lignes solides (avec 1 écart-type) en fonction de l'indice de surface de végétation (*LAI*) (à gauche) et du contenu en eau liquide de la colonne (*TCLW*) (à droite). Sur chaque figure, des couleurs différentes sont utilisées pour les intervalles de *TCLW* (à gauche) et de *LAI* (à droite). Les lignes pointillées montrent les valeurs prédites de ΔD_a par la paramétrisation avec la même entrée et la zone ombrée montre l'incertitude de 1σ sur les paramètres estimés.

Le paramétrage de la fonction C est basé sur les distributions présentées dans la figure V.14. Des mois de l'année 2016 sont téléchargés à partir de l'archive de l'ECMWF pour estimer une correction : *TCLW* et *LAI* sont sélectionnés comme entrées de la fonction C . Pour chaque pixel dans les données, on calcule l'amplitude du cycle D_a qui correspond à la différence entre la valeur maximale et la valeur minimale. La différence observée entre l'amplitude du cycle pour un ciel clair et nuageux ΔD_a pour différentes valeurs de *LAI* (à gauche) et de *TCLW* (à droite) est affichée. Une journée est considérée nuageuse si la valeur de $TCLW \geq 0$ à la fois pour la température maximale et la température minimale. On retient dans ce cas la valeur moyenne du contenu en eau liquide qui correspond à cette amplitude.

Pour chaque intervalle de végétation ou de quantité d'eau dans la colonne atmosphérique, l'écart type autour de la moyenne de ΔD_a est calculé et indiqué par la ligne continue.

De même que pour les observations de la figure V.13, un ΔD_a important est observé plus les nuages sont épais et moins la végétation est dense. Une paramétrisation polynomiale estimée des distributions est également affichée sous forme de lignes pointillées, et elle est utilisée pour implémenter en pratique la fonction C . L'incertitude de 1σ sur les paramètres estimés pour cette fonction est indiquée par une zone colorée de la courbe. L'incertitude augmente pour des $TCLW$ plus grands et des FVC plus petits, en raison du plus petit nombre d'échantillons dans la base de données pour ces extrêmes.

En substance, la paramétrisation agit comme un coefficient d'amortissement réduisant l'amplitude du cycle diurne de température en présence de nuages. Les surfaces nues répondent plus fortement aux variations de rayonnement dues aux nuages, et la quantité d'eau dans la colonne liée à l'opacité du nuage occasionne une diminution de l'amplitude jusqu'à un plateau pour les valeurs très importantes d'eau liquide dans la colonne atmosphérique. La paramétrisation est plutôt simple, mais est néanmoins considérée comme suffisamment efficace pour que des informations sur les nuages soient incluses dans la méthode de désagrégation.

5 Méthode globale « tout temps » de désagrégation

Résumé de la méthode de désagrégation

Le déroulé global de la méthode de désagrégation est illustré par la figure V.15. Dans une première phase d'apprentissage, la fonction statistique F est entraînée sur un ensemble représentatif de données du LSA SAF T_{HR} avec les entrées A_{HR} correspondantes : FVC , T_0 , D_a et $\sin(\alpha_S)$ comme décrit dans la section 3.3. La fonction de correction simple C est paramétrée à partir des données issues d'ERA5 $TCLW$ et LAI pour corriger l'amplitude de la température D_a en conditions nuageuses. Une fois ces deux fonctions paramétrées, elles peuvent être utilisées dans la phase d'application pour augmenter la résolution des mesures effectuées à basse résolution T_{LR} .

Dans la phase d'application, pour les T_{LR} concernées, on sélectionne les entrées A_{HR} correspondantes (FVC , T_0 et \tilde{D}_a , $\sin(\alpha_S)$). La fonction F est utilisée pour estimer les structures à haute résolution $\Delta\tilde{T}_{HR}$, qui sont ensuite ajoutées aux T_{LR} originales comme décrit par l'équation V.3. Si la scène contient des estimations de T_{LR} considérées comme nuageuses dans les données ERA5, les valeurs de $TCLW$ et LAI correspondantes sont extraites. La fonction C détermine la correction à appliquer à l'amplitude D_a avant son utilisation dans la fonction F .

Application aux températures de surface estimées à partir de SSMIS

Une fois la méthodologie de désagrégation développée à l'aide des jeux de données SEVIRI et ERA5, elle est appliquée aux estimations de T microonde de SSMIS, basées sur la méthode décrite au chapitre I, section 4. Pour illustrer l'utilisation de la méthode de désagrégation, trois journées (26 novembre 2016, 26 et 27 mai 2016) d'estimation de T SSMIS pour deux régions d'Afrique sont téléchargées depuis les archives LST CCI.

Tout d'abord, la désagrégation est appliquée aux températures SSMIS sur la même zone du delta de l'Okavango que dans la section 3.4. Les résultats sont présentés dans

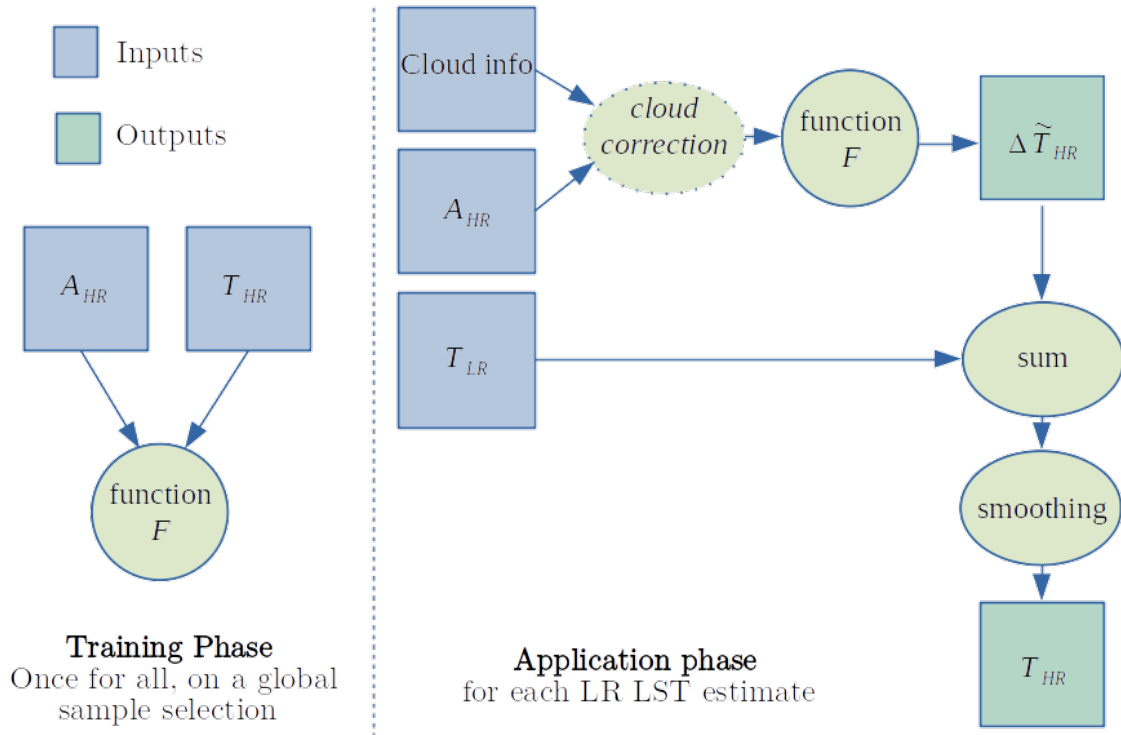


FIGURE V.15 – Schéma représentant la méthodologie de désagrégation, sur la gauche pour la phase d'entraînement, et sur la droite pour la phase d'application.

la figure V.16. Le 26 novembre 2016 à 05:00 est sélectionné, car il présente à la fois des conditions claires (la partie sud de la scène) et nuageuses (pour la partie nord où aucune température infrarouge n'est disponible en raison de la présence de nuages). Les deux sous-figures A et D montrent les variables auxiliaires D_a et FVC . Leurs valeurs sont différentes de celles des figures V.7 et V.8, reflétant les changements saisonniers de ces variables. Il est notable que même si la partie nord de la scène est nuageuse, les estimations de D_a et FVC dérivées de l'IR existent, en raison de leur intégration temporelle sur 10 jours. Par conséquent, la désagrégation peut être appliquée partout sur cette scène. La sous-figure B montre les T_{LR} issues de SSMIS, c'est-à-dire les T basse résolution de température dont il faut augmenter la résolution. Dans la sous-figure C, les étapes de correction des nuages, d'agrégation et de lissage sont appliquées pour produire $T_{downscaled}$. La sous-figure E montre la moyenne (5×5) des températures haute résolution issues de SEVIRI. Enfin, la sous-figure F présente la T_{HR} originale de SEVIRI.

La comparaison des T_{LR} originales de SSMIS, des $T_{downscaled}$ et des T_{HR} de SEVIRI montre que les structures spatiales des T_{HR} de SEVIRI sont bien représentées par les estimations de $T_{downscaled}$. Dans la partie de la scène où le ciel est clair, $T_{downscaled}$ reproduit correctement le gradient de T du nord au sud et les T plus froides dans le lit de la rivière, comme observées par SEVIRI T_{HR} . La valeur moyenne et l'écart-type des pixels clairs de la scène sont calculés pour les T_{LR} de SSMIS et pour les T_{HR} de SEVIRI dans la figure V.16, et montrent un biais chaud de $\sim 1^\circ\text{C}$ pour les mesures de SSMIS. Ce biais est lié à une différence initiale entre les T microondes et infrarouges dans cette zone, et non à un artefact de la désagrégation. Pour la partie de la scène où le ciel est clair, la RMSE calculée entre T_{HR} de SEVIRI et T_{LR} de SSMIS donne une valeur de 1.72°C . Cet RMSE est améliorée à 1.60°C entre T_{HR} et $T_{downscaled}$ après la désagrégation.

La désagrégation pour la partie nuageuse de la scène est raisonnable, sans artefacts

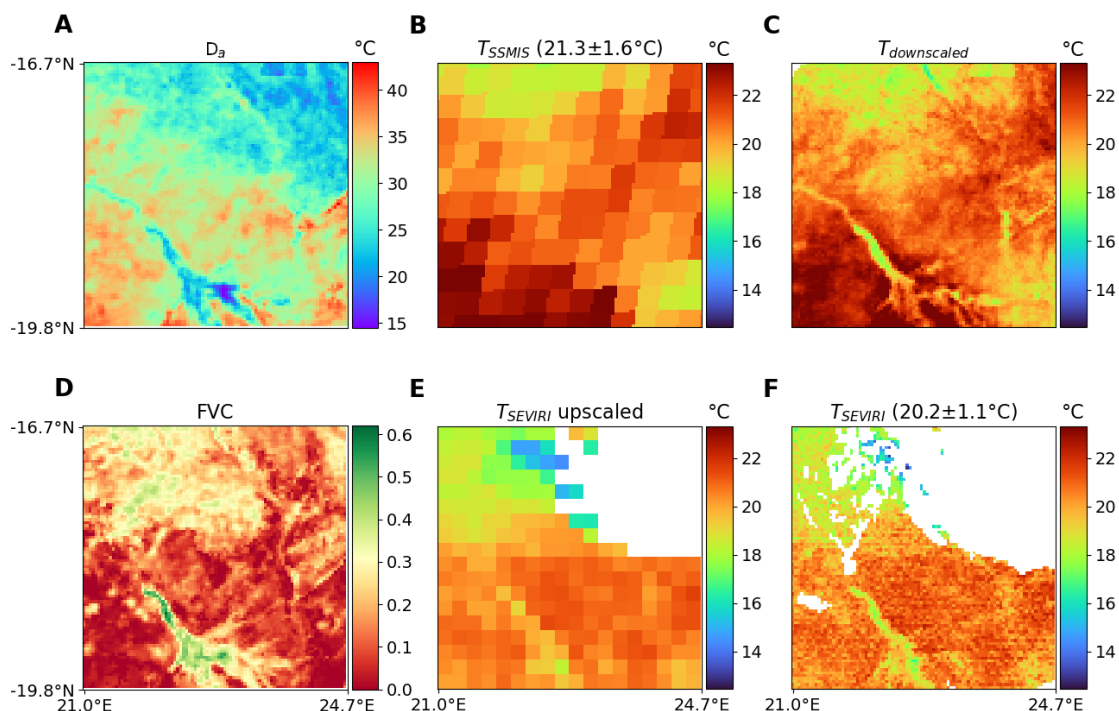


FIGURE V.16 – Pour une scène du 26 novembre 2016 à 05:00 sur le delta de l’Okavango (16.7° S 21.0° E à 19.8° S 24.7° E) : la première colonne montre 2 des variables auxiliaires d’entrée : D_a (A) et FVC (D). La sous-figure B montre la température SSMIS originale (T_{LR}) et la sous-figure C la $T_{downscaled}$, enfin la sous-figure F montre la T_{SEVIRI} à haute résolution et la sous-figure E la T IR moyennée spatialement sur une zone de 5×5 . La température moyenne et l’écart-type pour la partie ciel clair de la scène sont indiqués pour T_{SSMIS} et T_{SEVIRI} .

évidents par rapport à la partie claire. L’impact des nuages sur la variable D_a , calculé par la fonction C , donne des valeurs de l’ordre de 3°C pour cette scène. Mais son impact sur $T_{downscaled}$ n’est pas très important par rapport à une situation en ciel clair, car l’effet de D_a dans la fonction F est plus faible vers les premières heures du matin que vers la mi-journée.

La comparaison de $T_{downscaled}$ avec les T_{HR} de SEVIRI près des bords des nuages révèle quelques incohérences. Certaines estimations des T de SEVIRI semblent beaucoup plus froides ($\sim 10^{\circ}\text{C}$) que les T moyennes environnantes de la scène ($\sim 20^{\circ}\text{C}$) et que les estimations $T_{downscaled}$ correspondantes de SSMIS, ce qui suggère une probable contamination nuageuse dans les estimations T IR.

Dans un second temps, pour mieux mettre en évidence la correction apportée par la paramétrisation de l’effet des nuages, la figure V.17 montre le résultat de l’application de la désagrégation à 2 jours consécutifs sur une scène du nord de la Tanzanie (2.4°S 32°E à 4.9°S 35.4°E). Pour chaque jour, en 2 colonnes, la première ligne (sous-figures A, B) montre la T SSMIS originale et indique la température moyenne et l’écart-type sur les parties de la scène où le ciel est clair. Puis la deuxième ligne (C, D) montre le résultat après l’application de la désagrégation. La troisième ligne (E, F) affiche T_{SEVIRI} avec la température moyenne et l’écart type sur la scène. Enfin, la dernière ligne (G, H) montre la correction sur la variable D_a qui est appliquée sur la base de la paramétrisation des nuages et des données ERA5, pour la structure des nuages du sud-ouest au nord-est de la

scène.

La présence de nuages est plus forte dans le nord-est de la scène pour la deuxième journée. Pour les parties claires, le deuxième jour apparaît plus chaud que le premier, que ce soit pour les observations infrarouges ou microondes. Les observations T_{LR} issues de SSMIS dans la partie claire des scènes sont ~ 3 K plus chaudes que celles obtenues par SEVIRI. Le biais est plus important que celui observé dans la figure V.16 (~ 1 K plus chaud) et peut être attribué à une contamination nuageuse dans les données infrarouges due à la difficulté à filtrer des nuages fragmentés.

Au-dessus d'une large section de la scène pour la deuxième journée, les T_{SSMIS} sont plus froides comparées au premier jour, reflétant possiblement l'impact de la présence de nuages sur la température de surface. L'illustration de la correction appliquée à l'amplitude D_a à la même résolution que les données ERA5 montre que cette correction n'excède pas 2.0 K. Enfin $T_{downscaled}$, à la fois pour la partie claire et sous les nuages, montre des structures spatiales comparables, en accord avec les structures géophysiques de la région. On peut le remarquer sur le delta situé au nord-ouest du lac Victoria dans le nord-ouest ou le plateau haut au nord-est où la température de surface moyenne est plus faible.

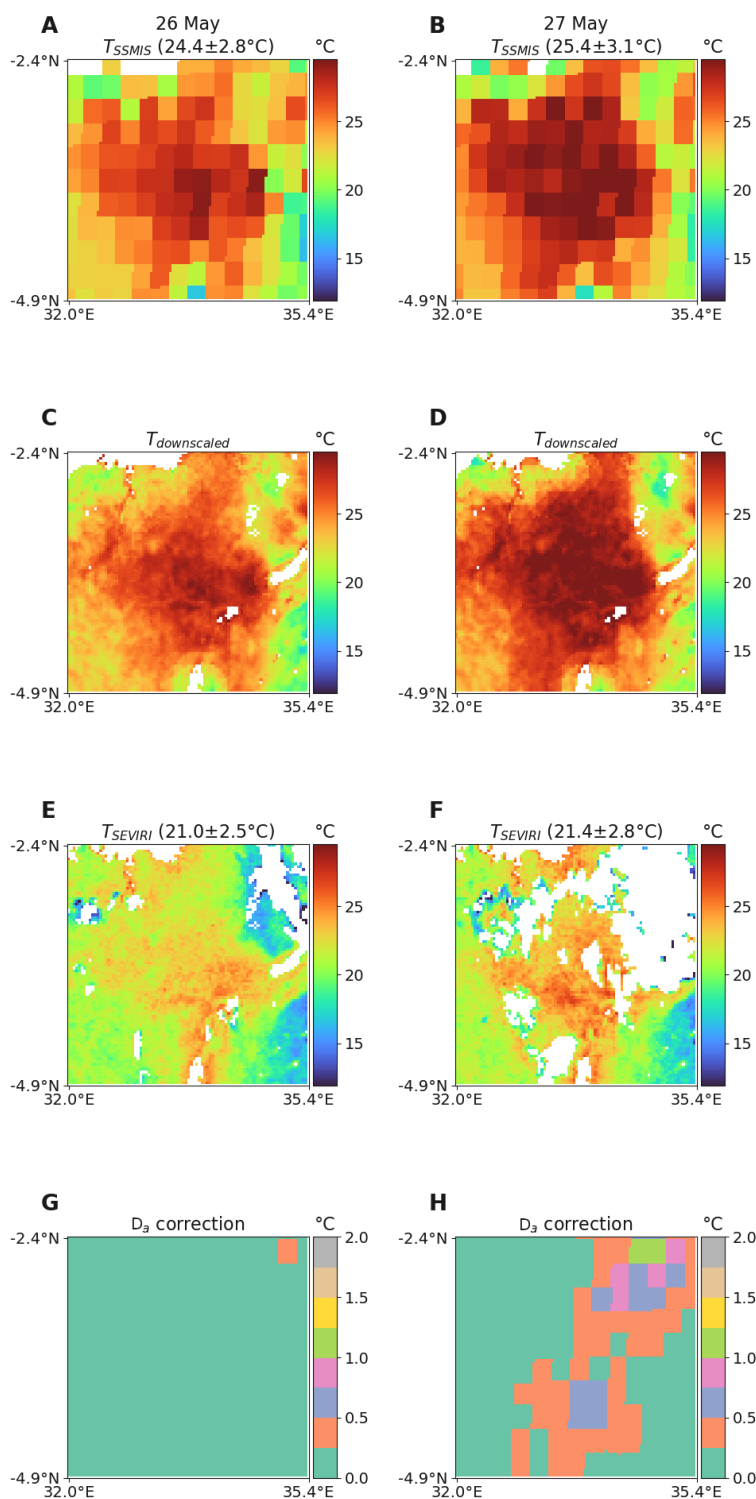


FIGURE V.17 – Scènes du passage de l’après-midi pendant 2 jours consécutifs, les 26 et 27 mai 2016 au nord de la Tanzanie (2.4°S 32°E à 4.9°S 35.4°E). Sur la première ligne (A et B) se trouve l’observation SSMIS à basse résolution, puis C et D montrent le résultat après l’application de la désagrégation, et E et F montrent la T_{SEVIRI} au même moment. Enfin, la dernière ligne (G, H) montre la correction sur le paramètre D_a obtenu par la paramétrisation appliquée aux données de la réanalyse ERA5. La moyenne et l’écart-type de T_{SSMIS} et T_{SEVIRI} pour les parties de la scène en ciel clair sont indiqués dans les sous-titres des figures.

6 Conclusion

Les sections précédentes de ce chapitre illustrent les performances de la méthodologie développée pour la désagrégation sur des données synthétiques, mais aussi l'application de la méthode à la température de surface dérivée des observations SSMIS.

Pour fournir une température de surface terrestre « tout temps » à haute résolution, la méthodologie développée permet d'augmenter la résolution spatiale originale des températures microondes de 25 km à une résolution de 5 km, similaire aux observations géostationnaires en infrarouge. À cette fin, une relation statistique est d'abord établie entre les informations auxiliaires à haute résolution et des mesures de température à haute résolution, afin d'estimer les motifs de la température à une résolution spatiale fine dans des conditions de ciel clair. Le modèle est entraîné à l'aide de T infrarouges et de données auxiliaires provenant d'observations géostationnaires SEVIRI à ~ 5 km. Différentes combinaisons de variables sont testées et l'erreur du modèle est la plus faible sur toute la gamme de température, en utilisant la valeur minimale et l'amplitude du cycle diurne médian de la température en ciel clair (T_0 et D_a), ainsi que la couverture végétale fractionnelle FVC , et l'angle d'élévation solaire ($\sin(\alpha_S)$). Dans une phase d'application, la fonction statistique est ensuite utilisée pour produire un modèle de température haute résolution. Ce modèle permet de calculer l'écart attendu de température entre les pixels à haute résolution contenu dans une observation à basse résolution. Ce motif est ajouté à la basse résolution T_{LR} et fournit des estimations spatialement plus fines $T_{downscaled}$.

Plusieurs points forts de cette méthodologie sont à souligner. La méthodologie développée ici est indépendante de l'heure locale de passage de la plate-forme transportant l'instrument microonde utilisé pour estimer les T . Les différentes heures locales de passage de chaque instrument (06:00 et 18:00 pour SSMIS et 01:30/13:30 pour AMSR par exemple) sont traitées par la prise en compte de l'angle d'élévation solaire dans le modèle. Le même modèle peut donc être appliqué à un produit de température de surface à basse résolution à tout moment de la journée. C'est un avantage pour la création de produits de T cohérent à haute résolution à partir d'observations microondes, étant donné les heures locales de passage variables des satellites selon les missions. La méthode peut déjà être utilisée pour toutes les zones pour lesquelles les données des satellites géostationnaires sont disponibles, y compris les zones tropicales souvent nuageuses. Les variables auxiliaires utilisées dans l'entraînement du modèle F offrent de nouvelles sources d'information pour décrire les modèles T à haute résolution avec une précision plus fine. L'avantage apporté par l'utilisation de D_a et T_0 notamment, est illustré dans la figure V.6, est compréhensible.

Cependant, la performance de cette méthode dépend beaucoup de la performance de l'estimation de température à partir des observations microondes, pour laquelle des évaluations ont été menées (Prigent et al., 2016; Jiménez et al., 2017; Ermida et al., 2017). L'évaluation des températures de surface estimées à partir d'observations microondes ne fait pas partie des objectifs de cette étude, mais des biais systématiques sur certaines surfaces ont été mis en évidence. Cela signifie également que les erreurs dans les données de désagrégation ne sont pas forcément dues à la méthodologie de désagrégation (comme l'illustrent les exemples des figures V.7 et V.8) mais peuvent être inhérents à la méthode d'estimation des T en microonde.

Une poursuite de ces travaux pourrait évaluer la possibilité de généraliser les prédicteurs pour obtenir une désagrégation globale, par exemple en utilisant des capteurs infrarouges haute résolution en orbite polaire. Les T obtenues par l'instrument MODIS et

des indices de végétation qui sont disponibles à l'échelle mondiale pourraient être utilisés. MODIS étant à bord d'un orbiteur polaire avec des heures de passage à l'équateur à 01:30 et 13:30, les informations sur l'ensemble du cycle diurne de T ne sont pas disponibles, mais T à 01:30 est proche du minimum du cycle T et sa différence avec T à 13:30 peut être un indicateur de l'amplitude du cycle diurne D_a . Ces données permettraient de réduire l'échelle des microondes T à une plus haute résolution, jusqu'à ~ 1 km.

Enfin, la correction des valeurs auxiliaires sous les nuages, développée ici, n'est pas seulement une composante majeure de cette méthodologie, mais aussi une contribution précieuse à la compréhension de l'impact attendu des nuages sur T . Comme soulevé dans le travail de Zhang et al. (2019), l'hypothèse selon laquelle le coefficient dérivé du ciel clair peut également être utilisé pour les scènes nuageuses est ici étudié plus en détail. La paramétrisation de l'effet des nuages proposée ici est une première tentative pour mieux comprendre ce qui reste une source élevée d'incertitude dans les mesures globales de T « tout temps ». L'évaluation de cette correction est toujours délicate, car les données pour la validation de la température en ciel nuageux sont peu nombreuses.

D'autres pistes de développement restent aussi possibles en particulier pour l'implémentation de la fonction F établissant la relation entre les données auxiliaires et la structure à haute résolution des températures. La première évolution pourrait être d'utiliser des modèles plus complexes permettant de restituer la variance de la donnée ciblée comme les modèles de *Mixture density* (mélanges de densité) (Christopher M. Bishop, 1994) pour obtenir en même temps que la structure de la température une estimation de sa variabilité en chacun des points. Cette information pourrait fournir une estimation de l'erreur qui servirait à la fois à améliorer la qualité de la désagrégation et à estimer l'erreur introduite par une telle méthode. De même les méthodes développées pour la super-résolution d'images pourraient être adaptées (Hughes and Ramsey, 2010; Yue et al., 2016) aux images provenant d'observations microondes, permettant ainsi de bénéficier des avancées méthodologiques et de l'inclusion d'information spatiale de façon plus fine qu'avec la méthode actuellement développée. Ces résultats font ressortir un certain nombre de points qui devraient être pris en compte lors de la mise en œuvre d'une telle méthode de désagrégation, ou du développement de méthodes alternatives dans le futur.

En résumé

- Les applications utilisant les produits de température de surface requièrent des résolutions spatiales les plus fines possibles. Pour les instruments déjà existants la seule façon de développer des méthodes de désagrégation.
- Une méthode a été développée en s'appuyant sur la relation entre les caractéristiques physiques de la surface (végétation, amplitude et valeur minimale du cycle diurne de température), l'élévation du soleil, et les températures à haute résolution en ciel clair. En utilisant les structures de température à haute résolution découlant de cette relation, on peut désagréger les températures microondes à basse résolution à la résolution des informations auxiliaires utilisées (ici ~ 5 km) en ciel clair quelle que soit l'heure de l'observation microonde à basse résolution effectuée.
- Pour pouvoir appliquer cette méthode sur des données altérées par une couverture nuageuse, une correction de l'amplitude du cycle diurne de la surface en fonction du type de végétation et du contenu en eau liquide atmosphérique est développée. Cette correction peut être appliquée aux données d'entrée de la fonction servant à l'estimation de la structure haute résolution des températures.
- La méthode présentée a été appliquée à des températures issues de l'instrument SSMIS et des températures de surfaces synthétiques pour lesquelles elle réduit l'écart entre les températures à haute et basse résolution.

Chapitre VI

Conclusions et perspectives

L'estimation de la température de surface par des observations satellites microondes passives est un outil précieux pour de nombreuses applications. Elle est nécessaire pour contraindre le bilan énergétique à la surface afin d'estimer diverses caractéristiques des surfaces continentales. De plus, le suivi de l'évolution des conditions climatiques de notre environnement se rappelle fréquemment à tous les humains. À cet effet, il convient de construire des séries temporelles avec la meilleure résolution temporelle et spatiale possible, tout en garantissant la cohérence et la qualité des estimations. Les résultats obtenus dans cette thèse apportent des réponses pour un certain nombre de défis soulevés par ce cahier des charges. Dans ce chapitre, l'état actuel de la création de séries temporelles est présenté et l'accent est mis sur les résultats obtenus déjà appliqués, ou leur mise en œuvre envisagée. Après un résumé des résultats obtenus dans la thèse les perspectives ouvertes par ce travail sont développées.

1 Séries de données existantes du projet LST CCI

À la suite de la Convention-cadre des Nations unies sur les changements climatiques, l'agence spatiale européenne (ESA) a parmi ses missions celle de contribuer aux archives climatiques des variables climatiques essentielles (ECV). Le projet *Climate change Initiative* (CCI) est la traduction de cet effort pour créer de longues séries temporelles de variables géophysiques observées par des satellites. La température de surface est la cible du projet *Land Surface Temperature* LST CCI en combinant les estimations de différentes sources, infrarouge et microonde, en provenance de multiples instruments pour approcher une couverture de 30 années. La méthode d'estimation des températures à partir des observations microondes a été présentée dans l'introduction de cette thèse. Elle a déjà été appliquée aux observations de différents imageurs dont SSM/I et SSMIS.

Les données quotidiennes produites par l'application de la méthode sont affichées dans la figure I.17 du chapitre I. Des moyennes mensuelles de température à une échelle globale peuvent aussi être construites. La figure VI.1 présente la carte des températures moyennes de l'observation de fin de journée en heure locale par l'instrument SSM/I. Des séries temporelles de ces températures de surface issues des microondes ont été créées pour les premières itérations du projet LST CCI. Elles sont disponibles à l'adresse : <https://climate.esa.int/en/projects/land-surface-temperature/>. En se basant sur les températures de brillance inter-étalonnées fournies par le CM SAF pour l'instrument SSM/I à bord de la plateforme F13 (1995-2015) et SSMIS à bord de F17 (2006-2021),

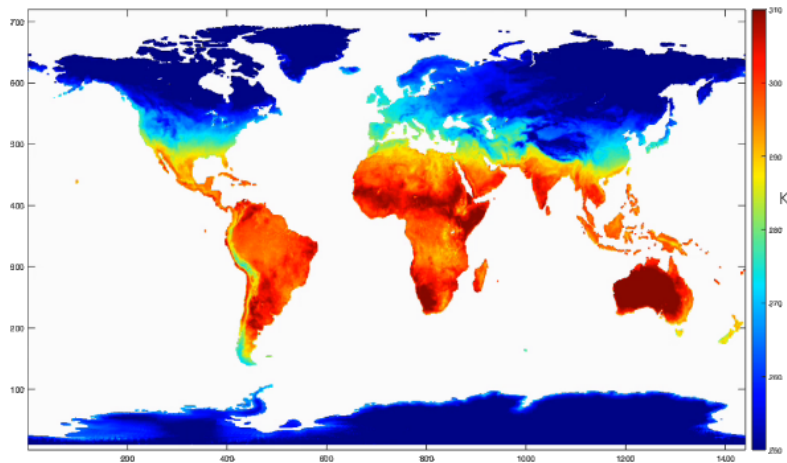


FIGURE VI.1 – Température moyenne (K) du passage en fin de journée en heure locale ($\sim 18:00$) de l'instrument SSM/I F13 pour le mois de janvier 1996 (figure extraite des données LST CCI).

une couverture temporelle de 25 ans est atteinte.

La figure VI.2 présente une comparaison entre les anomalies de séries temporelles obtenues en microondes par le projet LST CCI et les anomalies dans les températures de surface de la réanalyse ERA5. Le changement d'instrument microonde utilisé est signalé par le trait gris au 160^{ème} mois. Les séries temporelles sont données pour des températures de surface chaudes au-dessus de l'Amazonie et froides au-dessus de l'Antarctique. Ces deux surfaces présentent l'avantage d'une relative stabilité des émissivités, mais aussi des cycles diurnes de température quasi-nuls qui limitent les tendances artificielles liées à la dérive des heures de passage à l'équateur des satellites. On constate que les anomalies ont des variations similaires et les estimations microondes représentent remarquablement bien la variabilité interannuelle des températures des surfaces continentales.

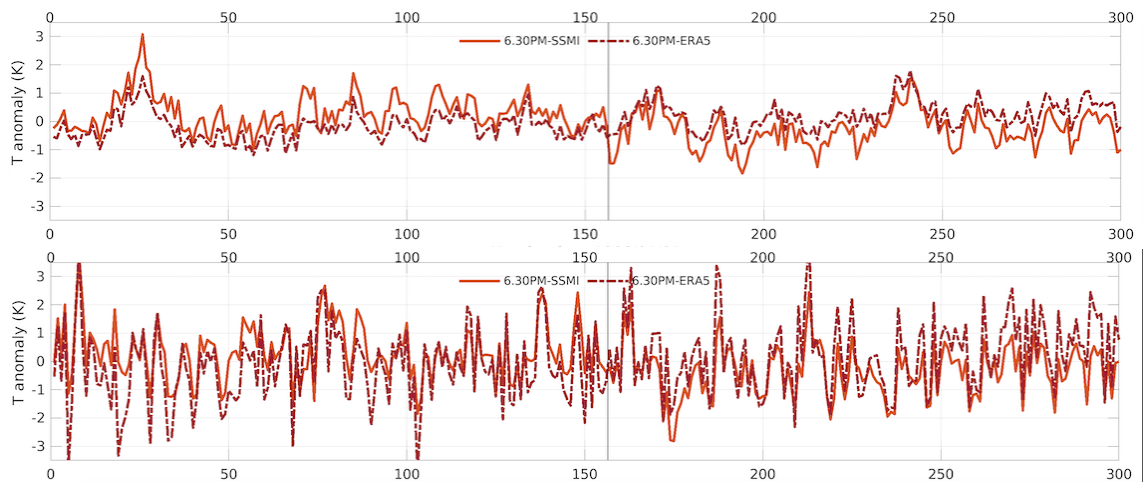


FIGURE VI.2 – Anomalies de températures (K) de surface mensuelles pour le passage au-dessus de l'équateur à 18:00 heure locale de l'instrument SSM/I et SSMIS (trait plein) et des données ERA5 (trait discontinu). En haut au-dessus de l'Amazonie, en bas au-dessus de l'antarctique. (communication de C. Jimenez)

L'accord entre les deux sources de données est globalement très bon et permet se projeter avec confiance dans la suite du projet LST CCI qui a été prolongé récemment.

2 Résumé des résultats obtenus : application au projet LST CCI et aux observations microondes

C'est dans l'optique de fournir des améliorations pour ce projet qu'une partie des développements de cette thèse ont été effectués, mais le champ d'application des résultats obtenus ne se limitent pas à ce cadre.

Pour améliorer les résultats fournis le premier point à traiter a été celui des données utilisées. Soit en augmentant la quantité de données disponibles ou en améliorant la qualité de celles utilisées.

Pour étendre la durée des séries temporelles de température de surface de 30 ans à plus de 40 ans, l'utilisation de l'instrument SMMR est envisagée. C'est en effet le seul imageur microonde multifréquence disponible entre 1979 et 1987. Dans le chapitre II, un inter-étalonnage a été proposé pour les températures de brillance des canaux en polarisations verticale et horizontale à 18 et 36 GHz de SMMR, par rapport à GMI utilisé comme référence. Mais les résultats montrent que l'incertitude associée aux mesures et à l'inter-étalonnage rendent délicate toute utilisation pour la construction de séries climatiques. La prolongation des séries temporelles de température sur une décennie supplémentaire ne semble pas suffisamment robuste. Ces données ne seront donc pas utilisées au sein du projet LST CCI dans l'immédiat. Les résultats ont à nouveau souligné l'importance de la qualité des étalonnages proposés par les services de traitement de données climatiques tels que le CM SAF. Les résultats mettent cependant en évidence la différence de l'étalonnage des mesures de température de brillance en microonde entre les océans et les surfaces continentales. Les coefficients proposés ici, et plus généralement la méthodologie, peuvent servir pour toutes les utilisations de longues séries de T_b microondes comme l'étude de l'étendue des glaces. Le CM SAF envisage d'adopter la correction proposée ici pour les températures de brillance continentale de SMMR dans les prochaines versions.

Dans la production des données de température de surface de niveau 2 détaillée dans le chapitre I, un certain nombre de filtres sont appliqués. Des indicateurs sont ajoutés au cours du traitement pour signaler les pixels pour lesquels le résultat fourni n'est pas de la meilleure qualité possible.

Par exemple ces indicateurs signalent les cas où la taille du pixel englobe des surfaces non terrestres (inondation ou océans) qui faussent la température estimée, ou que le modèle n'est pas dans des conditions d'application optimales, comme au-dessus des zones couvertes de neige ou de glace. Pour améliorer les méthodes dans ces situations, le chapitre III et IV apportent des réponses partielles.

Dans le chapitre III, un modèle a été développé pour améliorer le filtre détectant les contaminations nuageuses. Un modèle statistique pour la détection des contaminations, basé sur un réseau de neurones, a été présenté. Il est construit sur des observations colocalisées des températures de brillance de l'instrument GMI et d'une typologie de nuages créée avec l'instrument SEVIRI. Ce modèle propose une méthode pour déterminer la probabilité d'une contamination nuageuse dans les mesures de température de brillance en microonde. Les résultats, par bande de fréquence correspondant à différents instruments (≤ 40 GHz, ≤ 90 GHz, jusqu'à 183 GHz), fournissent un index adaptable aux applications finales. Les poids des réseaux de neurones entraînés sont disponibles ce qui rend la méthode simple à implémenter. Le filtrage est amélioré par rapport aux précédentes méthodologies, car entraîné spécifiquement sur les types de nuages contaminant pour les

observations microondes. Le modèle a déjà pu être intégré à la chaîne de traitement des observations microondes dans la dernière version du produit de température de surface LST CCI. La méthode de détection des contaminations nuageuses peut aussi être appliquée à toutes les observations en microonde passives, pour améliorer les estimations de variables géophysiques en surface. Une des applications déjà mise en œuvre est de filtrer les contaminations nuageuses au-dessus de l’océan dans une étude d’évaluation de simulation d’émissivités microondes de l’océan par comparaison avec des observations (AMSR, GMI, etc.) (Kilic et al., 2018).

Le chapitre IV étudie la pénétration des microondes dans les zones arides. Dans le cadre de la LST CCI, l’indicateur en service pour les zones avec de fortes profondeurs de pénétration n’est pas spécifique aux observations microondes passives. Pour estimer la profondeur de pénétration des microondes dans les zones arides, plusieurs éléments ont été regroupés. Tout d’abord une connaissance des propriétés physiques des sols et de la température en surface du sol (provenant des données ERA5) sont combinés à une solution de l’équation de Fourier pour fournir le profil de température dans le sol. Les cycles diurnes des températures de brillance (issus des observations de GMI entre 2015 et 2020) peuvent être utilisés pour estimer simultanément la profondeur de pénétration à chaque fréquence et les émissivités pour chaque fréquence et polarisation, dans l’équation du transfert radiatif. Les profondeurs de pénétration sont estimées jusqu’à 35 cm à une fréquence de 10 GHz. Les cartes obtenues permettent de préciser les pixels pour lesquels la méthode d’estimation à partir d’observations microondes passives ne donne pas des résultats comparables aux estimations issues de l’infrarouge. Une correction pour rendre comparable les températures issues des microondes à celles issues des infrarouges au-dessus des zones arides est en développement. Les cartes de pénétration des microondes à différentes fréquences dans la sub-surface offrent des perspectives intéressantes dans des domaines connexes :

1. La modélisation des profondeurs de pénétration jusqu’à des fréquences basses (≤ 10 GHz) est nécessaire pour l’analyse des mesures à basse fréquence (~ 1.4 GHz) issues des instruments *Soil Moisture Active Passive* (SMAP) ou *Soil Moisture and Ocean Salinity* (SMOS). Les travaux effectués ici pourront être extrapolés vers ces basses fréquences.
2. La valeur des paramètres diélectriques inférées des cartes de pénétration peut améliorer la connaissance de la composition des surfaces dans les zones arides, rarement étudiées. Ces résultats sont en effet un outil pour lier les propriétés minéralogiques et de granularité des sols à des structures géologiques observées en microonde.

Pour finir, la problématique de la résolution spatiale des estimations de température est traitée, en développant une méthode de désagrégation des températures issues des microondes (chapitre V). Elle est basée sur des variables auxiliaires à haute résolution représentant la dynamique du cycle diurne (valeur minimale et amplitude des températures pour un cycle diurne ainsi que l’angle d’élévation du soleil) et un indice de végétation. Ces variables servent à estimer la structure spatiale des températures à haute résolution. En ajoutant une correction à ces variables auxiliaires pour prendre en compte l’effet des nuages sur les températures de surface, la méthode peut être appliquée par « tout temps » et à toute heure pour désagréger des observations de température de surface. L’amélioration dans la représentation des températures, en passant d’une résolution de ~ 25 km à une résolution de ~ 5 km, est notable, en particulier quand les structures spatiales sont contrastées (autour de midi heure solaire). Cette méthode est une étape précieuse pour effectuer des comparaisons avec des températures de surface obtenues par des observations

infrarouges.

3 Développements et perspectives

Les travaux présentés ici s'inscrivent dans une continuité des développements des méthodes de mesure de température de surface et aussi dans le champ plus large de la télédétection satellite microonde. Il est possible dès maintenant de se projeter dans les prochaines améliorations attendues dans le domaine et aussi cibler certains domaines d'étude prometteurs. Les avancées pourront être méthodologiques, basées sur l'utilisation de nouveaux instruments à bord de satellites, ou sur de nouvelles campagnes de mesures terrain. Elles offriront l'opportunité de prolonger et améliorer les résultats actuels.

Travaux futurs

Les résultats de cette thèse ont mis en évidence certains besoins, pour les prochaines versions des produits de températures de surface issus des microondes, en particulier ceux produits pour le LST CCI.

- À moyen terme une production unifiée combinant les données issues des microondes à celles issues de l'infrarouge pourrait être envisagée. Une telle unification pourrait être le but d'un produit de niveau 3+ et être proposée dans les prochains plans de travail du projet LST CCI. Cela reposerait à la fois sur la correction des différences de mesure dans les zones arides et sur la réduction d'échelle pour obtenir un produit « tout temps » haute résolution de température de surface.
- Concernant les modèles de restitution de la température, des améliorations peuvent être apportées. Il pourrait être utile de mettre à jour les données utilisées dans l'inversion, par exemple les variables auxiliaires comme les bases de données climatologiques d'émissivités utilisées.
- D'autres limites n'ont pas été abordés dans cette thèse, par exemple l'amélioration de la restitution des températures de surface au-dessus de zones avec des variabilités très fortes de l'émissivité. C'est le cas des surfaces de neiges et de glaces ou encore des zones d'inondations.

Nouvelles données

Les données d'observations ou les données auxiliaires sont critiques pour les résultats obtenus. Les futurs instruments ou l'ajout de données pour compléter les données manquantes pourraient apporter des résultats nouveaux. Les promesses les plus tangibles et les plus évidentes pour la mesure de la température par des observations microondes passives viennent des futurs instruments. Un certain nombre d'instruments sont en cours de développement pour permettre de poursuivre les observations dans des gammes de fréquence similaires. Ces mesures après inter-étalonnage pourront ainsi pérenniser les séries temporelles commencées. Parmi les projets on peut citer :

- La continuation de la série d'instruments AMSR, avec AMSR-3 dont le lancement est prévu en 2023, ajoutera aux fréquences actuelles d'AMSR-E des canaux à 165.5 et 183.3 GHz (Kasahara et al., 2020). Les mesures effectuées par ces instruments sont particulièrement utiles, avec des heures de passage à l'équateur qui restent stables

et proches des maxima et minima des cycles diurnes de température.

- La série d'instruments *Micro Wave Radiation Imager* (MWRI) à bord des plateformes chinoises Feng-Yun en orbite polaire fournissent des données dans des canaux entre 10 et 89 GHz (Yang et al., 2011) et avec des caractéristiques instrumentales proches de AMSR. Le programme est prévu pour durer au moins 10 ans de plus et pourrait donc aussi compléter les données existantes. Un travail d'inter-étalonnage devra cependant être effectué au préalable.
- Enfin, le *Copernicus Imaging Microwave Radiometer* (CIMR) est un instrument très prometteur pour l'observation passive en microonde avec un début des opérations prévu en 2028. Les canaux seront disponibles entre 1.4 GHz et 36 GHz avec une résolution de 5 km à 36 GHz. Pour la première fois, des mesures à toutes ces fréquences seront disponibles par le même instrument. Et si l'objectif principal de cette mission est l'observation des pôles et de la surface de l'océan (Kilic et al., 2018), ces observations permettront aussi d'améliorer notre connaissance des surfaces continentales.

Les mesures de T_b dans les gammes de fréquences des microondes ne seront pas interrompues et les prochains instruments promettent de nouvelles observations à des résolutions actuellement inaccessibles. Ces données pourraient être directement comparées avec des produits infrarouges à des résolutions similaires. On peut aussi anticiper que l'importance des questionnements climatiques rendra plus robuste les étalonnages effectués, pour garantir la stabilité des observations sur le long terme. Étant donné les caractéristiques proches des différents instruments, les modèles actuels d'estimation de la température de surface à partir d'observations microondes devraient pouvoir être employés sans difficultés majeures.

Dans les données qui seront amenées à être exploitées, toutes ne proviennent pas de nouveaux instruments en orbite. Pour utiliser de manière optimale les observations multifréquences aux résolutions variables des imageurs microondes, des évaluations sur des zones in situ doivent être poursuivies. Ces données sont nécessaires que ce soit pour la validation des séries temporelles de température ou des profils de température dans le sol. On peut ajouter à ces campagnes de mesure in situ ou aux stations pérennes, des campagnes de sondage, pour connaître avec précision les caractéristiques des sols sur Terre. Ces données pourraient apporter des bénéfices considérables, en mesurant systématiquement des propriétés thermiques ou diélectriques des sols rencontrés, et ce même en dehors des zones agricoles.

Perspectives

Cette thèse a traité de quelques aspects de la télédétection d'une variable très spécifique : la température de la surface de la terre. Mais on peut faire le constat que dans les estimations par télédétection de nombreuses variables et phénomènes interagissent. Améliorer la connaissance de la mesure de température de surface peut donc avoir des bénéfices pour d'autres variables d'intérêt. Parmi celles-ci, on peut citer la végétation qui est une composante de la couverture de la surface qui a fait l'objet de nombreuses études. Les questions soulevées par le transfert radiatif dans la végétation et la température effective de celle-ci en font un objet d'étude complexe. Mais cette mesure est nécessaire compte tenu de la criticité de l'estimation de la biomasse dans le cycle du carbone global. (Saatchi et al., 2011; Brandt et al., 2018; Prigent and Jimenez, 2021). Les observations microondes sont donc un outil puissant pour de nombreuses applications et les synergies entre les

mesures de différentes sources permettent d'estimer de nombreux paramètres à la surface de la Terre, des océans, ou dans l'atmosphère.

Un dernier point à souligner est que les méthodes statistiques employées sont très générales. Elles peuvent être employées dans d'autres domaines de la télédétection que ceux abordés dans cette thèse. On a pu par exemple commencer à développer des méthodes de filtrage de nuage pour des observations hyperspectrales infrarouges, en utilisant des réseaux de convolution similaires à ceux du chapitre III. Si tout au long de cette thèse l'importance critique des données utilisées pour l'entraînement et la validation a été soulignée, les méthodes statistiques employées ici seront amenées à évoluer. La capacité à entraîner des réseaux plus complexe permettra sans doute d'améliorer les résultats, mais sous réserve de comprendre les contraintes imposées par la physique. Avec ces nouvelles méthodologies et la quantité de données disponible, des modèles plus spécialisés pourront être introduits, en envisageant une localisation des modèles pour chacun des pixels sur Terre. Ces paramétrisations des phénomènes physiques auront sans doute de meilleure performance que celles obtenues dans cette thèse. Ces approches soulèvent aussi la question d'utiliser les techniques de transfert d'apprentissage pour pré-entraîner des modèles qui partagent un contexte commun avant une optimisation finale pour chaque application spécifique.

La température de surface est au cœur de nombreux problèmes de télédétection, les progrès accompli dans ce domaine peuvent donc profiter à des applications très variées. Les prochaines décennies verront à la fois de nouveaux résultats et une pression croissante pour mesurer les changements en cours dans notre environnement.

Annexe A

Appendix

1 Notations

Symbole	Unité	Nom
δ_p	m	Profondeur de pénétration
α	$m^2 \cdot s^{-1}$	Diffusivité thermique
k	$W \cdot m^{-1} \cdot K^{-1}$	Conductivité thermique
C	$JK^{-1}kg^{-1}$	Capacité thermique massique
D_a	K ou $^{\circ}C$	Amplitude d'un cycle diurne de température
L	$W \cdot m^{-2} \cdot sr^{-1} \cdot Hz^{-1}$	Luminance énergétique spectrale
$L_n, S_n, G...$	$J \cdot s^{-1}$	Flux thermique
T (T_s , LST)	K ou $^{\circ}C$	Température de la surface
T_0	K ou $^{\circ}C$	Température minimale d'un cycle diurne de température
T_b	K	Température de brillance
T_{2m}	K ou $^{\circ}C$	Température de l'air à 2 m du sol
α_S	$^{\circ}$	Angle d'élévation du soleil
δ_s	m	Profondeur de peau
ϵ	$F \cdot m^{-1}$	Permittivité diélectrique
κ_a	m^{-1}	Coefficient d'absorption volumique
κ_s	m^{-1}	Coefficient de diffusion volumique
λ	m	Longueur d'onde
μ	$Np \cdot m^{-1}$	Constante d'atténuation
ν	GHz, Hz, s^{-1}	Fréquence
ω	$rad \cdot s^{-1}$	Pulsation
\bar{T}	K ou $^{\circ}C$	Température moyenne
ρ	$kg \cdot m^{-3}$	Densité
c	$m \cdot s^{-1}$	Vitesse de la lumière dans le vide (≈ 299792458)
h	$J \cdot Hz^{-1}$	Constante de Planck ($\approx 6.62607015 \times 10^{-34}$)
k_B	$J \cdot K^{-1}$	Constante de Boltzman ($\approx 1.380649 \times 10^{-23}$)

2 Acronymes

AMSR	Advanced Microwave Sounding Radiometer
AVHRR	Advanced Very High Resolution Radiometer
CM SAF	Climate Satellite Application Facility
CNES	Centre National d'études Spatiale
CWVC	Column Water Vapor Content
EIA	Earth Incidence Angle
ESA	European Space Agency
FCDR	Fundamental Climate Data Records
FVC	Fractional Vegetation Cover
G(PM)MI	Global Precipitation Mission Microwave Imager
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
HTESSEL	Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology
IASI	Infrared Atmospheric Sounding Interferometer
IFS	Integrated Forecasting System
IR	InfraRed
ISCCP	International Satellite Cloud Climatology Project
JAXA	Japan Aerospace Exploration Agency
KIT	Karlsruher Institut für Technologie
LAI	Leaf Area Index
LSA SAF	Land Surface Application Satellite Application Facility
MADRAS	Microwave Analysis and Detection of Rain and Atmospheric Structures
MLP	MultiLayer Perceptron
MNT	Modèle Numérique de Terrain
MODIS	Moderate Resolution Imaging Spectroradiometer
MW	MicroWave
NASA	National Aeronautics and Space Administration
NCEP	National Center for Environmental Prediction
NDVI	Normalized difference vegetation index
RMSE	Root mean square error
SEVIRI	Spinning Enhanced Visible and InfraRed Image
SMMR	Scanning Microwave Multichannel Radiometer
SSM/I	Special Sensor Microwave / Imager
SSMIS	Special Sensor Microwave Imager / Sounder
SURFRAD	Surface Radiation Budget Network
TELSEM	Tool to Estimate Land-Surface Emissivities at Microwave frequencies
TRMM / TMI	Tropical Rainfall Measuring Mission / TRMM Microwave Imager

Annexe B

Articles publiés

Atmos. Meas. Tech., 13, 5481–5490, 2020
<https://doi.org/10.5194/amt-13-5481-2020>
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.



Atmospheric
Measurement
Techniques
Open Access
EGU

Inter-calibrating SMMR brightness temperatures over continental surfaces

Samuel Favrichon^{1,2}, Carlos Jimenez^{2,1}, and Catherine Prigent^{1,2}

¹Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, Paris, France

²Estellus, Paris, France

Correspondence: Samuel Favrichon (samuel.favrichon@obspm.fr)

Received: 20 December 2019 – Discussion started: 27 January 2020

Revised: 27 August 2020 – Accepted: 31 August 2020 – Published: 14 October 2020

Abstract. Microwave remote sensing can be used to monitor the time evolution of some key parameters over land, such as land surface temperature or surface water extent. Observations are made with instruments, such as the Scanning Microwave Multichannel Radiometer (SMMR) before 1987, the Special Sensor Microwave/Imager (SSM/I) and the subsequent Special Sensor Microwave Imager/Sounder (SSMIS) from 1987 and still operating, and the more recent Global Precipitation Measurement Microwave Imager (GMI). As these instruments differ on some of their characteristics and use different calibration schemes, they need to be inter-calibrated before long-time-series products can be derived from the observations. Here an inter-calibration method is designed to remove major inconsistencies between the SMMR and other microwave radiometers for the 18 and 37 GHz channels over continental surfaces. Because of a small overlap in observations and a ~ 6 h difference in overpassing times between SMMR and SSM/I, GMI was chosen as a reference despite the lack of a common observing period. The diurnal cycles from 3 years of GMI brightness temperatures are first calculated and then used to evaluate SMMR differences. Based on a statistical analysis of the differences, a simple linear correction is implemented to calibrate SMMR on GMI. This correction is shown to also reduce the biases between SMMR and SSM/I, and can then be applied to SMMR observations to make them more coherent with existing data records of microwave brightness temperatures over continental surfaces.

1 Introduction

Since 1978, passive microwave satellite imagers have provided Earth observations at multiple frequencies, over ocean and land, for atmospheric or surface applications such as cloud and precipitation monitoring, surface temperature estimation, ocean wind speed measurement, or sea ice concentration retrievals (Ulaby et al., 1986). With now more than 40 years of data records, climate analysis can be performed from these measurements, provided that the observation time series are well calibrated, consistent, and homogeneous.

The successive microwave imagers share common characteristics, but with technological changes from a generation to the next and possible calibration issues between instruments even from the same series. Major microwave imagers include the Seasat Nimbus-G Scanning Multichannel Microwave Radiometer (SMMR) from 1978 to 1987 (Gloersen and Barath, 1977), the Special Sensor Microwave Imager (SSM/I) from 1987 up to now (Hollinger et al., 1990), and the more recent Global Precipitation Measurement Microwave Imager (GMI) since 2014 (Hou et al., 2014). To create homogeneous measurements from different instruments, each instrument has to be carefully calibrated first, and the different instruments have then to be inter-calibrated. The changes between instruments include differences in overpassing times, in Earth incidence angles (EIAs), in channel center frequencies, and in bandwidths. All these aspects need to be accounted for in the inter-calibration process.

Multiple teams have worked to provide corrections for the brightness temperatures of microwave imagers to ensure homogeneous data records over time (e.g., Berg et al., 2013; Wentz, 2013; Fennig et al., 2020) that can be used as fundamental climate data records (FCDRs). FCDRs are calibrated

Published by Copernicus Publications on behalf of the European Geosciences Union.

data from multiple sensors that have been made coherent and quality-controlled to be accurate and stable over long time periods. These are the basic data records from which geophysical products can be computed and aggregated to form long-term records of essential climate variables (Bojinski et al., 2014).

Different methods exist to inter-calibrate sensors. Quasi-direct comparisons of observations can be performed at the poles, where overpassing of polar satellites are frequent and the surface responses rather stable. However, this calibration is limited to rather cold brightness temperatures (Sapiano et al., 2013). The statistics of the coldest or warmest scenes have also been analyzed to inter-calibrate the sensors. This is the so-called vicarious calibration, applied over ocean for the cold end (Ruf, 2000) and to the Amazon forest for the warm end (Brown and Ruf, 2005).

More recently this vicarious method has been extended to other forested sites, taking into account their seasonal variability (Yang et al., 2016). Over ocean, double-difference methods are often adopted, using radiative transfer simulations as the reference, to bridge the gap between instruments with different characteristics (Kroodsmas et al., 2012). The radiative transfer model is usually fed by atmospheric and surface information from reanalyses, and it can theoretically account for changes in Earth incidence angle (EIA), changes in channel characteristics (e.g., frequency and bandwidth), and differences in overpassing times. However, this method is very challenging over continental surfaces. First, microwave radiative transfer models over land, along with all their necessary input parameters (e.g., soil moisture, vegetation density, and snow water equivalent), are not available with the required quality over a large range of surface types. Second, the possibly strong diurnal variability of the land surface temperatures is not described with enough accuracy and temporal resolution to account for differences in the satellite overpassing times. Nevertheless, Dai and Che (2009) tested a modeling of the diurnal variation of the surface temperature to inter-calibrate instruments with different overpassing times over land.

Another inter-calibration method consists in using matchups with a reference instrument that has a different orbit type, making it possible to provide quasi-direct comparisons over a large range of latitudes, even for satellites with different overpassing times. With their low orbits that sample the diurnal cycle, the Tropical Rainfall Measuring Mission Microwave Imager (TMI), and more recently the GMI, can be used as intermediate references to inter-calibrate Sun-synchronous instruments with drastically different overpassing times such as the SSM/I series (Equator overpassing time around dusk and dawn) and the Advanced Microwave Scanning Radiometer (AMSR) series (Equator overpassing times at midnight and midday). Berg et al. (2018) applied the different methods to inter-calibrate the SSM/I series. The different methods agreed well with each other, offering increased confidence in the proposed inter-calibration.

Table 1 lists the major FCDRs from passive microwave imagers available to the community, indicating the inter-calibrated instruments, and the reference instrument. The EUMETSAT Climate Monitoring Satellite Application Facility (CM SAF), Colorado State University (CSU), and Remote Sensing Systems (RSS) all include the SSM/I and Special Sensor Microwave-Imager/Sounder (SSMIS) in their inter-calibrations. CSU uses for the first time GMI in their recent inter-calibration scheme (Berg et al., 2018). They simultaneously employ multiple calibrations methods to reduce the uncertainty in the data record. Their initial work (V01) is improved by adding more satellites and by using GMI as the reference for its description of the diurnal cycle over land. RSS's inter-calibration effort is essentially based on the use of a radiative transfer model over ocean (Wentz, 2013). The CM SAF uses SSM/I F11 as their reference satellite. The inter-calibration is a scene-dependent correction. It is done by correcting the cold end of the observed SMMR-ERA T_b differences to match the SSM/I-ERA T_b differences but keeping the warm calibration end at the observed hot load target temperature. It is therefore not expected to have an impact on the warm T_b range.

So far, only the CM SAF includes SMMR in their fundamental climate data records (FCDRs). The inter-calibration was developed initially for the monitoring of fluxes over the ice-free ocean (Andersson et al., 2010). To extend the climate record of satellite-derived land surface parameters in time, here we propose to analyze the possibility of inter-calibrating the SMMR instrument over land. The SMMR instrument failed in August 1987, and the first SSM/I (F08) was launched in 1987, with an overlapping time of only a few weeks, and with ~ 6 h differences in their overpassing times at the Equator. Here, we suggest to use GMI as a reference instrument, assuming that the environmental conditions have not changed drastically from the SMMR to the GMI era, to allow the comparison of a large set of observations averaged over time. This strategy does not allow a detailed inter-calibration to be performed, but it makes it possible to correct for major biases that so far have hampered the use of SMMR over land for the generation of climate records of geophysical parameters. We will concentrate on the channels that are common to all the microwave imagers used in the FCDR in Table 1, the Ku (around 18 GHz) and Ka (around 36 GHz) channels. These are key observations for the retrieval of several land surface parameters (e.g., surface water extent (Prigent et al., 2007), snow water equivalent (Pulliainen, 2006), or land surface temperature (Jiménez et al., 2017)). We will use the CM SAF FCDR for SMMR data set as the starting point of our developments.

In Sect. 2, the satellite observations used in this study are briefly described, along with their preprocessing for the analysis. The result of the inter-comparison is presented in Sect. 3, along with the proposed inter-calibration procedure and its evaluation. Section 4 concludes this study.

Table 1. The characteristics of the available fundamental climate data records from passive microwave imagers. The ✓ denotes an inter-calibrated instrument, while the * is for the reference instrument.

Instrument	Reference	SMMR	SSM/I	SSMIS	TMI	AMSR-E	AMSR2	GMI
CDR CSU V01	Berg et al. (2013)	–	✓	✓	*	–	–	–
CSU V03	Berg et al. (2018)	–	✓	✓	✓	✓	✓	*
CDR RSS	Wentz (2013)	–	✓	✓	–	–	–	–
CM SAF FCDR	Fennig et al. (2020)	✓	*	✓	–	–	–	–

2 Data and method

2.1 The satellite data

Table 2 summarizes the major characteristics of the SMMR and GMI instruments for the Ku and Ka channels. SSM/I on board F08 is also included, as it is the only instrument with an overlap period with SMMR (albeit of only 28 d), and it will be used for evaluation. GMI, with its non-Sun-synchronous orbit, observes the full diurnal cycle, including the SMMR and SSM/I overpassing times.

SMMR was launched in 1978 on the Nimbus-7 satellite and operated until August 1987. It is the first multichannel microwave imager, designed mainly for oceanic applications to estimate the surface wind speed and sea surface temperature (Gloersen and Barath, 1977). Due to power limitation on board, measurements were performed only every other day. The EIA decreases from 50.2 to 49.3° starting in 1986. The initial SMMR record used by the CM SAF is the Level 1B data, not the raw counts that are not available. The SMMR Level 1B data are described by Njoku (1980) and include the antenna pattern and spillover correction, as well as sensor drift correction. The calibration uses the cosmic background temperature as a cold reference (2.7 K), an on-board hot calibration load around 300 K, and climatological means to estimate biases in the calibrations. A linear calibration was performed to reduce instrument bias in the target domain, i.e., for ocean surface parameter estimations. The inter-calibration performed by the CM SAF is based on the double-difference technique between the SSM/I F08 and SMMR brightness temperatures (T_{bs}), using radiative transfer simulations from reanalysis to account for the changes in frequencies, bandwidths, and EIA. This correction is computed only over cloud-free water surfaces. It is described in detail in Fennig et al. (2020). In this study, we will only use SMMR data showing the best quality (data with sun intrusion, field-of-view, or scan error are removed).

GMI is a recent microwave imager launched in 2015. It observes between 10 and 190 GHz to measure precipitation across the globe. It will be used here as a reference standard for calibration. The instrument characteristics are summarized in Table 2, listing only the channels relevant to this study. The satellite has a 65° inclination, allowing non-Sun-synchronous observations of the Earth, from the tropics to the high latitudes. The antenna has an EIA of 52.8° and a swath

of ~ 900 km. This swath width associated with the inclination means that polar regions are not fully covered. The GMI calibration is described by Wentz and Draper (2016). In addition to the usual hot load, GMI uses noise diodes, improving calibration accuracy. The satellite can also perform flight maneuvers to correct drift and improve calibration. Lean (2017) found low biases for all channels, as compared to ECMWF simulations (lower than 0.8 K). All these technical specificities make GMI an excellent reference for inter-calibration purposes. In this study, the calibrated T_b Level 1C data (Hou et al., 2014) are used.

The first SSM/I was launched in 1987 on board the F08 satellite of the Defense Meteorological Satellites Program (DMSP). In the following years multiple instruments were launched, such as the TMI in 1997 and the SSMIS in 2003. The DMSP polar orbiters cannot correct orbital degradation; as a consequence, the instruments are subject to drifts in the overpassing times, making instrument inter-calibration more challenging. Here, the SSM/I data provided by the CM SAF without the inter-calibration layer will be used to evaluate the consistency of our SMMR land calibration.

2.2 The method

SMMR and GMI do not have any common observing period. Therefore we need to rely on a statistical analysis of SMMR and GMI observations from different years to provide a correction of the SMMR biases over continental surfaces. The fundamental hypothesis here is that the changes in the environmental conditions affecting the microwave signals are limited between the 1980s and 2010s, as compared to the SMMR and the GMI instrument calibration differences.

The SMMR and GMI EIA difference is ~ 3°. Over ocean, this can strongly affect the signal, due to the sensitivity of the ocean emissivity to the incidence angle as well as to the changing atmospheric contribution with angle. Over land, the surface emissivity is usually high for both polarizations, with values of 0.9 and higher for most surface types (Prigent et al., 2006), with very limited changes with EIA (Prigent et al., 2000). With a high land surface emissivity, the atmospheric contribution to the signal is small compared to the surface one and the changes in the atmospheric contribution with angle will not significantly affect the measurement. By the same token, with the land surface emissivities being rather high and showing very smooth variations with

Table 2. The major characteristics of the passive microwave imagers, directly relevant to this study. The instruments include other channels, but they are not used here.

Instrument	SMMR	SSM/I (F08)	GMI
Earth incidence angle (°)	50.2 to 49.3*	53.1	52.8
Channels (GHz)	18.0 (V, H), 21.0 (V, H)*, 37 (V, H)	19.35 (V, H), 22.235 (V), 37 (V, H)	18.7 (V, H), 23.8 (V), 36.5 (V, H)
Instantaneous field of view (km × km at 37 GHz)	17 × 29	24 × 36	8.6 × 14
Ascending Equator overpassing time (h)	00:00	06:00	non-Sun-synchronous
Operating years	October 1978–August 1987	June 1987–2006	September 2014–

* Failure of the 21 GHz channel in 1985 and drift from 1986.

frequency (Prigent et al., 2000), the differences in frequencies between the Ku and Ka channels of SMMR and GMI are not expected to significantly affect the signals.

To facilitate the comparison, each data set is projected on the Equal-Area Scalable Earth (EASE)-Grid 2.0 globally between 60° S and 70° N and on the EASE Southern Hemisphere azimuthal grid over the area of Antarctica (Brodzik et al., 2012). A land mask is used to filter out water pixels at 25 km resolution.

GMI data are collected for the months of January, February, July, and August from 2015 to 2017. These 4 months per year include the boreal summer and winter T_b diurnal cycles, and with data sampled from the whole Earth most of the possible surface conditions are represented in the analyzed data set. As GMI has a higher spatial resolution than the target grid, all observations falling within a grid cell of ~ 12 km radius are averaged to derive one value per grid cell. Regarding the temporal dimension, the values are averaged with a 15 min window for each grid cell to suppress small-scale variability in the signal. If no observation in a given grid cell is available in a particular 15 min interval, a linear interpolation with the closest existing measurements is performed. Finally, to get a smooth diurnal cycle, a moving average is applied over each grid cell with a 75 min window (two data points before and after the target one).

For SMMR, T_{bs} are collected from 1981, 1982, and 1987 for the same months as GMI. The observations are available twice a day (for the ascending and descending overpasses). Given the comparable resolutions of SMMR observations and the target grid, a nearest-neighbor technique is used to project the T_{bs} onto the EASE grid. Using several years of data for each instrument can alleviate possible effects of strong inter-annual variability in the signals (related to El Niño or La Niña events for instance).

The resulting data sets contain almost 2 million data points spanning all the continents and two contrasted seasons (summer and winter).

3 Results

3.1 SMMR and GMI comparisons at the regional scale

At regional scales, a set of different areas have been selected to represent large variability in environmental conditions. The averaged T_{bs} diurnal cycle derived from GMI is calculated for sample areas covering 25 grid cells, over 2 months, over 3 years (2015–2017). Over these areas, covering roughly a $1^\circ \times 1^\circ$ surface, the averaged SMMR T_{bs} for the ascending and descending orbits are also calculated, over the same months, for 1981, 1982 and 1987.

Figure 1 presents the results for areas in the North Hemisphere, during winter (January and February), at 18 and 37 GHz, for both orthogonal polarizations. It includes cold mountainous regions, arid deserts with high T_{bs} during the day and a large amplitude of the diurnal cycle, and warm rainforest with a limited variation in T_{bs} . Forested regions show a low diurnal cycle amplitude of the T_{bs} (< 10 K), with almost no difference between V and H polarizations due to the emission and scattering effect of the dense vegetation. Arid regions (cold or warm) have large diurnal variations in T_{bs} , directly related to the diurnal cycle of the land surface temperature. The polarization difference is significant, due to an almost specular behavior of these rather flat surfaces. With increasing vegetation cover (sparsely vegetated or grassland) an intermediate behavior is observed. The variability associated to each GMI average is computed and also displayed (the grey shades around the diurnal cycle indicate 1 standard deviation). The number of individual pixels used in the diurnal cycle calculation is also indicated (blue shades). The low variability of the signals over the Amazon confirms the high stability of the T_{bs} in this region, which is regularly used as a warm radiometric reference (e.g., Brown and Ruf, 2005). The SMMR values have been calculated for the initial Level 1B data as well as for the CM SAF inter-calibrated results, but only the Level 1B data are displayed in Fig. 1. Note that the CM SAF inter-calibration is neither designed nor validated for observations over land.

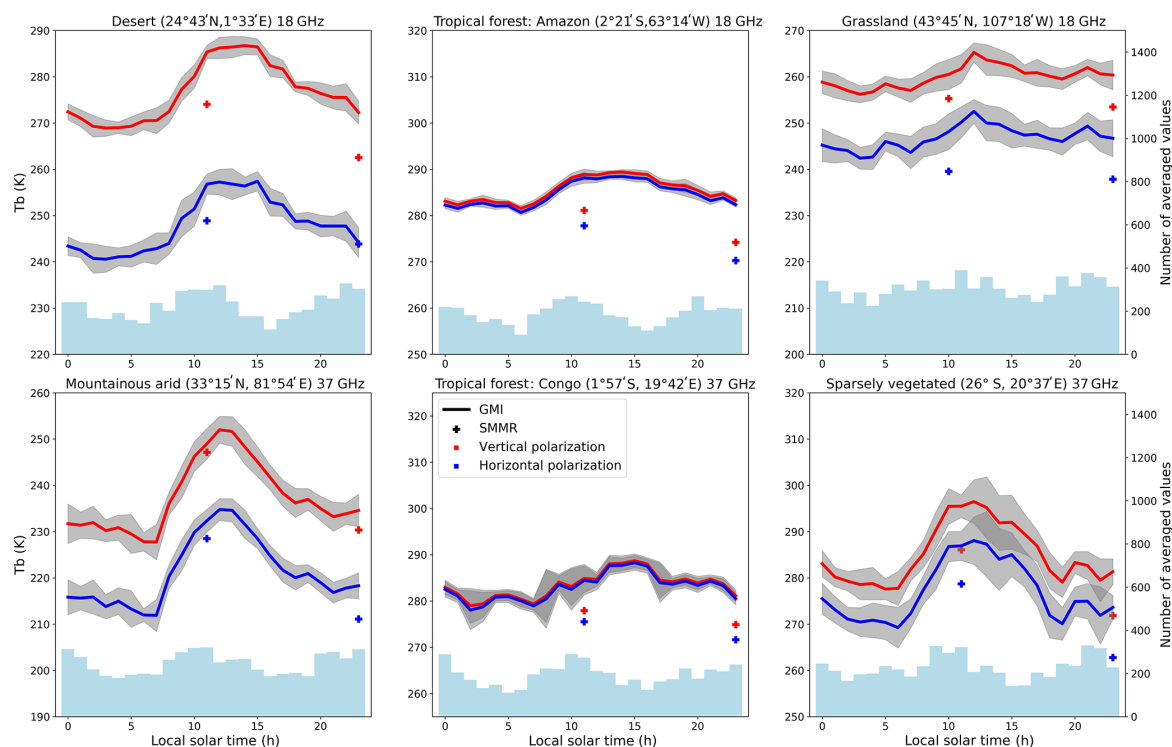


Figure 1. Diurnal cycles from GMI compared to SMMR values at different frequencies and locations averaged for the months of January and February (horizontal polarization in blue and vertical in red). Measurements are averaged over 25 grid cells and by hour, with the standard deviation of each average temperature in the diurnal cycle and the associated number of measurements displayed below.

The SMMR day and night values follow the same diurnal pattern as GMI, but with a systematic underestimation of the T_{bs} , regardless of the frequency and polarization. These differences are usually significantly larger than what is expected from the GMI signal variability, indicated by the standard deviation around the average. The difference between SMMR and GMI appears to be almost the same for both polarizations. The difference seems to decrease for the coldest locations. The variability is higher at midlatitudes, where year-on-year variations are possible and meteorological events can impact the measurements. The differences cannot be explained solely by the instruments' different characteristics (EIA or frequencies) or by environmental changes between the periods covered by the two instruments (the 1980s for SMMR and the 2010s for GMI). Different sources could cause such errors, for instance an erroneous warm calibration load temperature or an error in the correction of antenna pattern that misses the Earth.

The inter-calibration layer from CM SAF for SMMR does not seem to improve the results. Over land the mean correction added by the calibration is below 0.5 K for the 18 GHz vertical and horizontal polarization and 37 GHz horizontal channel. The correction for the 37 GHz vertically polarized

channel is on average around 2 K. The inter-calibration has been designed mainly for ocean observations with low T_{bs} and not for land applications, so it does not adequately correct for calibration issues over land (not shown in the figure). Other locations and seasons were analyzed, with similar conclusions (not shown here).

The observations over the Antarctic ice sheet are also explored, to extend our investigation to lower T_{bs} . Because of the GMI coverage, the comparison is limited to the edge of Antarctica. Care is exercised to avoid contamination by the ocean and sea ice. Southern Hemisphere winter months are selected (July and August). During this long night, there is no diurnal cycle and the T_{bs} distribution for the two satellites are directly compared (Fig. 2). The delta between SMMR Level 1B and GMI T_{bs} distribution is between 5 and 10 K for all channels except the 37 GHz vertical polarization, which shows a lower difference.

Here as well, the significant differences in the SMMR Level 1B and GMI T_{bs} cannot be attributed only to changes in environmental conditions, even over ~ 30 years. Jezek et al. (1993) compared SMMR and SSM/I over the Antarctic ice sheet during their overlap period and also found significant differences in both Ku and Ka bands. They discussed the im-

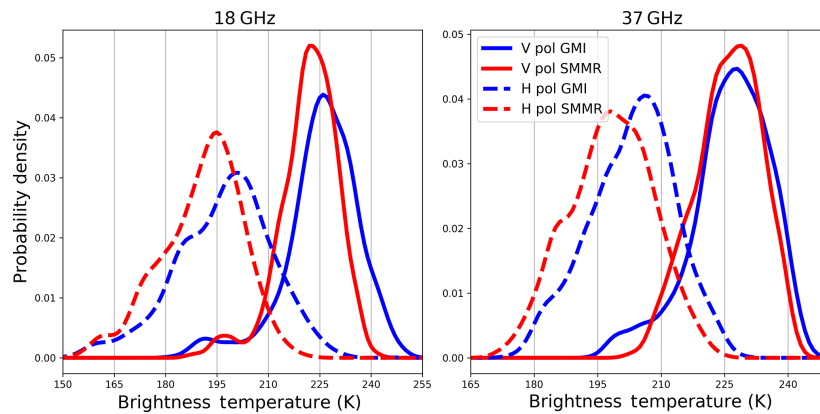


Figure 2. Probability density functions of July/August brightness temperatures over Antarctica for GMI in 2015 and SMMR in 1987.

pect of the change in EIA and frequency between the instruments but concluded that it cannot explain the large observed differences. Only calibration issues can explain the observed differences, with the SMMR T_{bs} colder than the GMI ones.

3.2 Derivation of a SMMR correction over continental surfaces

In order to assess the possibility to correct for the SMMR calibration issue, SMMR and GMI observations are compared at the global scale. The SMMR observations (both ascending and descending orbits) are compared to the corresponding GMI values for the same times in the day, for January and February and for July and August, over 3 years for the two instruments (1981, 1982, and 1987 were used for SMMR and 2015 to 2017 for GMI, as before). Points for which the difference between the GMI and SMMR values is outside 3 standard deviations from the mean difference are suppressed. We checked that these points were essentially located over coastal regions: GMI has a better spatial resolution than SMMR, and over the coasts the SMMR observations will likely include more contributions from the surrounding ocean. Figure 3 displays the distribution of the GMI against the SMMR T_{bs} , at 18 and 37 GHz, for both polarizations. The underestimation of the SMMR T_{bs} compared to the GMI ones clearly increases with T_{bs} . Here we suggest a simple linear correction to inter-calibrate the SMMR observations toward the GMI ones. A more sophisticated correction would not be justified. First, we are aware that this is a first-order inter-calibration, as the comparison involves different years. Second, it would be more complex to implement especially in processing chains. Lastly, it would likely overfit some part of the signal rather than correcting the T_{bs} .

A simple correction is proposed for all channels with the form $\hat{T}_{SMMR} = a \times T_{SMMR} + b$. The coefficients slope a and the intercept b are estimated through the minimization of the sum of squared difference: $\sum_{i=1}^N (T_{GMI} - a \times T_{SMMR} - b)^2$,

Table 3. Estimated linear correction coefficients, for the inter-calibration of the SMMR observations at 18 and 37 GHz, vertical and horizontal polarizations, with respect to GMI observations over the continents. The uncertainties are also added (computed with a t test at the 99 % confidence level).

Channel	Slope	Intercept	R^2
18V	1.10 ± 0.01	-18.7 ± 2.2	0.976
18H	1.05 ± 0.01	-1.29 ± 1.9	0.971
37V	1.15 ± 0.01	-32.2 ± 2.2	0.976
37H	1.04 ± 0.01	-1.23 ± 1.9	0.976

derived from the assumption that $\hat{T}_{SMMR} = T_{GMI}$, with N being the number of data points used for the regression coefficient estimation. The data collection covers the full T_b range observable over continental surfaces, including some polar regions.

The distribution of the data points from the cold and warm ends is uneven, with fewer points for lower T_{bs} . To alleviate this issue, we randomly sample the points from both the cold and warm ends to simulate an even distribution over the full T_b range. Different samplings were tested to confirm the stability of the estimated coefficients. The resulting regression lines are added to Fig. 3. The mean squared errors of the linear regression are indicated with grey shades. It appears clearly that the uncertainty in the areas with fewer data (between the Antarctica temperatures and the land data points) or near the edge of the range of values are the ones with the highest errors. The coefficients of the linear regressions are provided in Table 3. The uncertainties on the regression coefficients are computed with a t test for a 99 % confidence level.

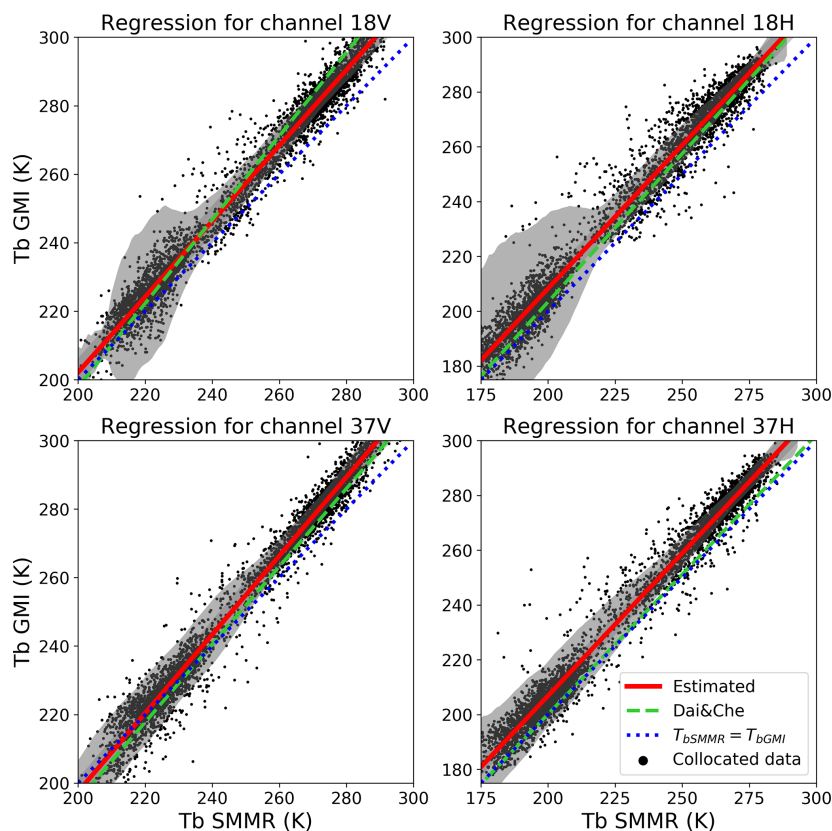


Figure 3. Linear regression on SMMR values against GMI after filtering points. In addition the correction derived by Dai and Che (2009) (and corrected for the SSM/I F08-to-GMI calibration) is displayed.

3.3 Evaluation

The suggested corrections are applied to SMMR values, and the results can be compared to other imager observations. With the inter-calibration being derived from GMI, a good agreement is expected with that sensor. The first check consists in comparing the calibrated SMMR observations to the SSM/I observations, for their overlapping period. The GMI-derived diurnal cycle of T_{bs} is used as a bridge between the SMMR and SSM/I observations that have different overpassing times. Figure 4 presents the comparisons of SMMR and SSM/I F08, for different locations, averaged over their common period in July and August 1987, along with the GMI-derived diurnal cycle of the T_{bs} (estimated over 3 different years). It shows that SSM/I T_{bs} , used without any inter-calibration, are in good agreement with the T_b diurnal cycle estimated from GMI. Contrarily to the SMMR observations before inter-calibration, no obvious large bias is observed between SSM/I and GMI, even for the very warm scenes, at 18 and 37 GHz for both polarizations. The average difference between SSM/I F08 and GMI (without any inter-calibration scheme applied) is around 2.5 K for the 18 GHz channels and

around 0.5 K for the 37 GHz channels. A bias is expected between measurements that have not been inter-calibrated, but it is lower than the one detected for SMMR and confirms our assumptions regarding the emissivity behavior over continental surfaces as well as the small variation in environmental conditions in the 30-year gap between the 1980s and the 2010s.

The agreement between the SMMR inter-calibrated values and GMI is clearly improved, as compared to the previous SMMR Level 1B results. These results show that the SSM/I F08 and the newly inter-calibrated SMMR are consistent, at least within the uncertainty provided with this inter-calibration. More cases have been tested, with similar results (not shown here).

In addition to the evaluation with other sensors, the results of our inter-calibration are compared with an alternative method developed for SMMR. Dai and Che (2009) built their inter-calibration scheme upon observations in desert and polar areas. The SSM/I F08 instrument is adopted as the reference, and a model is developed to account for the diurnal changes in temperatures. The resulting linear regressions are

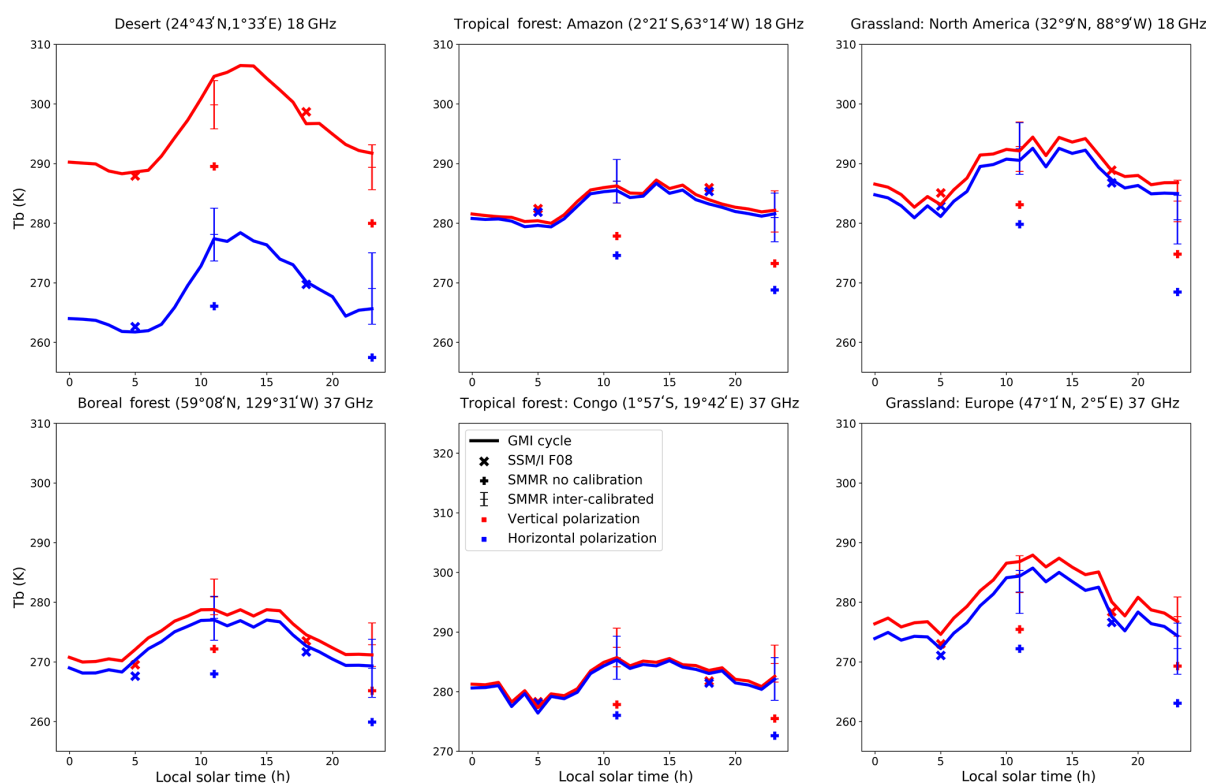


Figure 4. Diurnal cycle of GMI in July/August compared to values measured by SMMR with and without the proposed correction, and SSM/I values for the months of July/August 1987 over various locations and at different frequencies.

added in Fig. 3 (green dotted lines). The two inter-calibration methods are based on very different principles. Nevertheless, they agree well for all channels, with similar slopes and slightly different intercepts. This adds confidence to our methodology.

4 Conclusions

Several FCDRs from passive microwave imagers have been produced, from pioneer instruments such as SMMR to the AMSR series. So far, the efforts have essentially focused on oceanic regions, where radiative transfer simulations fed by reanalysis can serve as a reference to bridge the differences between the instruments, in terms of frequency, EIA, and overpassing time. Over continental surfaces, inter-calibration of passive microwave imagers is challenging, especially for Sun-synchronous polar-orbiting satellites with different overpassing times at the Equator. Here we derived a method to extend the FCDR collection to SMMR over land at 18 and 37 GHz, using the non-Sun-synchronous GMI instrument as a reference, despite the lack of a common observing period. GMI observations are used to reconstruct the diurnal cycles of T_{bs} that should be observed with SMMR, with the as-

sumption that the environmental conditions have not changed drastically over the last 30 years and that the differences in frequencies and EIA between the two instruments can be neglected over land. With these hypotheses, the objective is to correct for the large differences between the sensors. Before inter-calibration, the comparison of the SMMR and GMI observations shows a significant underestimation of the T_{bs} with SMMR, and this underestimation tends to increase with increasing T_{bs} . A linear regression is suggested for the 18 and 37 GHz channels, vertical and horizontal polarizations, to calibrate the SMMR observations toward the GMI estimates. The SMMR correction is evaluated with respect to the SSM/I F08 observations over their overlapping period in orbit, in July and August 1987. The GMI-derived diurnal cycle of T_{bs} acts as a bridge between the two instruments that have different overpassing times. A good agreement is reached between all sensor measurements. This inter-calibration of the SMMR instrument over land will make it possible to extend the passive microwave estimations of land surface variables over 9 more years backward, from 1978 to 1987. However, given the use of the more recent GMI instrument as a calibration reference, and the uncertainty regarding the calibration error sources as well as the validity of the environmental stabil-

ity hypothesis, any comparison between the different epochs should be conducted with extreme care. The use of this correction to derive climate series should be thoroughly assessed before using it. This will be practically tested in the near future for the estimates of, for instance, surface water extent and land surface temperatures from microwave observations, two variables that rely on the 18 and 37 GHz observations for their retrieval, under clear and cloudy-sky conditions.

Data availability. The Satellite Application Facility on Climate Monitoring provides access to the Fundamental Climate Data Record of Microwave Imager Radiances (https://doi.org/10.5676/EUM_SAF_CM/FCDR_MWI/V003; Fennig et al., 2017, 2020). The Global Precipitation Measurement Microwave Imager GMI_R Common Calibrated Brightness Temperatures Collocated L1C 1.5 h 13 km V05 (GPM_1CGPMGMI_R, <https://doi.org/10.5067/GPM/GMI/R/1C/05>; Berg, 2016; Berg et al., 2016) is provided by NASA.

Author contributions. All authors have been involved in interpreting the results, discussing the findings, and editing the paper. SF conducted the main analysis and wrote the draft of the paper. CJ and CP provided guidance on using the data sets and expertise on analyzing the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors are grateful to Karsten Fennig and Marc Schröder for their valuable help on using the CM SAF FCDR data as well as the discussion regarding the SMMR instrument and the work performed in this paper. We would also like to acknowledge Wesley Berg and Christian Kummerow from Colorado State University for the data and information they provided us.

Review statement. This paper was edited by S. Joseph Munchak and reviewed by Linwood Jones and two anonymous referees.

References

- Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., and Schulz, J.: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data – HOAPS-3, *Earth Syst. Sci. Data*, 2, 215–234, <https://doi.org/10.5194/essd-2-215-2010>, 2010.
- Berg, W., Sapiano, M. R., Horsman, J., and Kummerow, C.: Improved geolocation and earth incidence angle information for a fundamental climate data record of the SSM/I sensors, *IEEE T. Geosci. Remote*, 51, 1504–1513, <https://doi.org/10.1109/TGRS.2012.2199761>, 2013.
- Berg, W.: GPM GMI_R Common Calibrated Brightness Temperatures Collocated L1C 1.5 hours 13 km V05, Greenbelt, MD, USA, Goddard Earth Sciences

- Data and Information Services Center (GES DISC), <https://doi.org/10.5067/GPM/GMI/R/1C/05>, 2016.
- Berg, W., Bilanow, S., Chen, R., Datta, S., Draper, D., Ebrahimi, H., Farrar, S., Jones, W. L., Kroodsmma, R., McKague, D., Payne, V., Wang, J., Wilheit, T., and Yang, J. X.: Intercalibration of the GPM microwave radiometer constellation, *J. Atmos. Ocean. Tech.*, 33, 2639–2654, <https://doi.org/10.1175/JTECH-D-16-0100.1>, 2016.
- Berg, W., Kroodsmma, R., Kummerow, C., and McKague, D.: Fundamental Climate Data Records of Microwave Brightness Temperatures, *Remote Sens.*, 10, 1306, <https://doi.org/10.3390/rs10081306>, 2018.
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of essential climate variables in support of climate research, applications, and policy, *B. Am. Meteorol. Soc.*, 95, 1431–1443, <https://doi.org/10.1175/BAMS-D-13-00047.1>, 2014.
- Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H.: EASE-Grid 2.0: Incremental but significant improvements for earth-gridded data sets, *ISPRS Int. Geo-Inf.*, 1, 32–45, <https://doi.org/10.3390/ijgi1010032>, 2012.
- Brown, S. T. and Ruf, C. S.: Determination of an Amazon hot reference target for the on-orbit calibration of microwave radiometers, *J. Atmos. Ocean. Tech.*, 22, 1340–1352, <https://doi.org/10.1175/JTECH1769.1>, 2005.
- Dai, L. and Che, T.: Cross-platform calibration of SMMR, SSM/I and AMSR-E passive microwave brightness temperature, *Data Processing, Proc. SPIE 7841, Sixth International Symposium on Digital Earth: Data Processing and Applications*, 2010, 7841, 784103–784103-10, <https://doi.org/10.1117/12.873150>, 2009.
- Fennig, K., Schröder, M., and Hollmann, R.: Fundamental Climate Data Record of Microwave Imager Radiances, Edition 3, Satellite Application Facility on Climate Monitoring, https://doi.org/10.5676/EUM_SAF_CM/FCDR_MWI/V003, 2017.
- Fennig, K., Schröder, M., Andersson, A., and Hollmann, R.: A Fundamental Climate Data Record of SMMR, SSM/I, and SS-MIS brightness temperatures, *Earth Syst. Sci. Data*, 12, 647–681, <https://doi.org/10.5194/essd-12-647-2020>, 2020.
- Gloersen, P. and Barath, F. T.: A Scanning Multichannel Microwave Radiometer for Nimbus-G and SeaSat-A, *IEEE J. Oceanic Eng.*, 2, 172–178, <https://doi.org/10.1109/JOE.1977.1145331>, 1977.
- Hollinger, J. P., Peirce, J. L., and Poe, G. A.: SSM/I instrument evaluation, *IEEE T. Geosci. Remote*, 28, 781–790, <https://doi.org/10.1109/36.58964>, 1990.
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The global precipitation measurement mission, *B. Am. Meteorol. Soc.*, 95, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>, 2014.
- Jezek, K., Merry, C., and Cavalieri, D.: Comparison of SMMR and SSM/I passive microwave data collected over Antarctica, *Ann. Glaciol.*, 17, 131–136, <https://doi.org/10.3189/S0260305500012726>, 1993.
- Jiménez, C., Prigent, C., Ermida, S. L., and Moncet, J. L.: Inversion of AMSR-E observations for land surface temperature estimation: 1. Methodology and evaluation with station temperature, *J. Geophys. Res.*, 122, 3330–3347, <https://doi.org/10.1002/2016JD026144>, 2017.

- Kroodsmas, R. A., McKague, D. S., and Ruf, C. S.: Inter-calibration of microwave radiometers using the vicarious cold calibration double difference method, *IEEE J. Sel. Top. Appl.*, 5, 1006–1013, <https://doi.org/10.1109/JSTARS.2012.2195773>, 2012.
- Lean: Assimilation of Global Precipitation Mission (GPM) Microwave Imager (GMI) in all-sky conditions, *Boletin Tecnico/Technical Bulletin*, 55, 255–262, available at: <http://www.ecmwf.int/en/research/publications> (last access: November 2019), 2017.
- Njoku, E. G.: Antenna pattern correction procedures for the Scanning Multichannel Microwave Radiometer (SMMR), *Bound.-Lay. Meteorol.*, 18, 79–98, <https://doi.org/10.1007/BF00117912>, 1980.
- Prigent, C., Wigneron, J. P., Rossow, W. B., and Pardo-Carrion, J. R.: Frequency and angular variations of land surface microwave emissivities: Can we estimate SSM/T and AMSU emissivities from SSM/I emissivities?, *IEEE T. Geosci. Remote*, 38, 2373–2386, <https://doi.org/10.1109/36.868893>, 2000.
- Prigent, C., Aires, F., and Rossow, W. B.: Land surface microwave emissivities over the global for a decade, *B. Am. Meteorol. Soc.*, 87, 1573–1584, <https://doi.org/10.1175/BAMS-87-11-1573>, 2006.
- Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E.: Global inundation dynamics inferred from multiple satellite observations, 1993–2000, *J. Geophys. Res.-Atmos.*, 112, 1–13, <https://doi.org/10.1029/2006JD007847>, 2007.
- Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, *Remote Sens. Environ.*, 101, 257–269, <https://doi.org/10.1016/j.rse.2006.01.002>, 2006.
- Ruf, C. S.: Detection of calibration drifts in spaceborne microwave radiometers using a vicarious cold reference, *IEEE T. Geosci. Remote*, 38, 44–52, <https://doi.org/10.1109/36.823900>, 2000.
- Sapiano, M. R., Berg, W. K., McKague, D. S., and Kummerow, C. D.: Toward an intercalibrated fundamental climate data record of the SSM/I sensors, *IEEE T. Geosci. Remote*, 51, 1492–1503, <https://doi.org/10.1109/TGRS.2012.2206601>, 2013.
- Ulaby, F. T., Moore, R. K., and Fung, A. K.: *Microwave Remote Sensing Active and Passive-Volume III: From Theory to Applications*, Artech House Publishers, Boston, USA, 1986.
- Wentz, F. J.: *SSM/I Version-7 Calibration Report*, report number 011012, Remote Sensing Systems, Santa Rosa, CA, USA, 46 pp., 2013.
- Wentz, F. J. and Draper, D.: On-orbit absolute calibration of the global precipitation measurement microwave imager, *J. Atmos. Ocean. Tech.*, 33, 1393–1412, <https://doi.org/10.1175/JTECH-D-15-0212.1>, 2016.
- Yang, J. X., McKague, D. S., and Ruf, C. S.: Boreal, Temperate, and Tropical Forests as Vicarious Calibration Sites for Spaceborne Microwave Radiometry, *IEEE T. Geosci. Remote*, 54, 1035–1051, <https://doi.org/10.1109/TGRS.2015.2472532>, 2016.

Atmos. Meas. Tech., 12, 1531–1543, 2019
https://doi.org/10.5194/amt-12-1531-2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.



Detecting cloud contamination in passive microwave satellite measurements over land

Samuel Favrichon¹, Catherine Prigent¹, Carlos Jimenez², and Filipe Aires¹

¹Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, Paris, France

²Estellus, Paris, France

Correspondence: Samuel Favrichon (samuel.favrichon@obs-orm.fr)

Received: 5 October 2018 – Discussion started: 7 November 2018

Revised: 7 February 2019 – Accepted: 11 February 2019 – Published: 12 March 2019

Abstract. Remotely sensed brightness temperatures from passive observations in the microwave (MW) range are used to retrieve various geophysical parameters, e.g. near-surface temperature. Cloud contamination, although less of an issue at MW than at visible to infrared wavelengths, may adversely affect retrieval quality, particularly in the presence of strong cloud formation (convective towers) or precipitation. To limit errors associated with cloud contamination, we present an index derived from stand-alone MW brightness temperature observations, which measure the probability of residual cloud contamination. The method uses a statistical neural network model trained with the Global Precipitation Microwave Imager (GMI) observations and a cloud classification from Meteosat Second Generation-Spinning Enhanced Visible and Infrared Imager (MSG-SEVIRI). This index is available over land and ocean and is developed for multiple frequency ranges to be applicable to successive generations of MW imagers. The index confidence increases with the number of available frequencies and performs better over the ocean, as expected. In all cases, even for the more challenging radiometric signatures over land, the model reaches an accuracy of $\geq 70\%$ in detecting contaminated observations. Finally an application of this index is shown that eliminates grid cells unsuitable for land surface temperature estimation.

1 Introduction

Visible–infrared (vis–IR) satellite imagers provide excellent information about land surface characterization. Applications include land surface temperature estimations (e.g. Fre-

itas et al., 2013), vegetation characteristics (e.g. Tucker et al., 2005), and surface water extent (e.g. Pekel et al., 2016).

These geophysical parameters can be retrieved accurately and with a good spatial and temporal resolutions from vis–IR observations, but only under clear-sky conditions. With clouds covering $\sim 60\%$ of the globe at any time (Rossow and Schiffer, 1999), there is a need for alternative sources of information. Passive microwave observations from satellites can partly fill this gap: they are much less sensitive to clouds and can provide valuable estimates of surface properties, despite their coarser spatial and temporal resolutions. Today, land surface temperature can be retrieved from IR observations for $\sim 60\%$ of the locations with a spatial resolution of 1 km^2 twice a day from polar orbiters (Prata et al., 1995) and with a spatial resolution of 2 km^2 every 15 min from geostationary satellites (e.g. Schmit et al., 2017). On the other hand, passive microwaves can provide this information with a spatial resolution of $\sim 20\text{ km}^2$ twice a day over $\sim 100\%$ of the continents (Aires et al., 2001). Programmes are underway to merge these different observations for a complete spatial and temporal coverage. For instance, long time series of land surface temperature estimations with passive microwave observations are under construction, using different generations of passive microwave satellite instruments to be used in synergy with IR estimates (e.g. Prigent et al., 2016; Jiménez et al., 2017).

Although microwaves are less sensitive to clouds, the effect of clouds and rain on the microwave radiation increases with frequency. Multiple effects can occur, from liquid water clouds and rain emitting passive microwave radiation at the physical temperature of the cloud or rain to scattering by ice clouds that can lower the measured brightness temperatures,

Published by Copernicus Publications on behalf of the European Geosciences Union.

especially at high frequencies and for large ice contents. The cloud–rain effect that can be detected strongly depends on the surface type. The surface contribution to the passive microwave observations is proportional to the surface emissivity that changes from ~ 0.5 over ocean to ~ 1 over dry soil or dense forests. This means that the contrast between liquid particles in the cloud and rain and the surface will be usually larger over ocean than over land: cloud and rain liquid water emission increases the brightness temperature over the radiometrically cold ocean but will not show much contrast over the already radiometrically warm land. The opposite will prevail for frozen clouds, with the cloud scattering depressing the brightness temperature above the radiometrically warm land surface. Over ocean, passive microwaves have been extensively used to quantify the cloud liquid water and rain amounts (e.g. Greenwald et al., 1993; Kummerow et al., 1998). For ocean surface applications, cloud liquid water amount can usually be accounted for and the surface parameter estimation can compensate for the cloud impact, when atmospheric transmission is still high enough to have a significant contribution from the surface. Over land, cloud and rain detection using passive microwave is much more complicated (e.g. Spencer et al., 1989; Aires et al., 2001). First, surface emissivity is usually close to one, reducing the contrast between cloud and surface, and second, it changes spatially and temporally, e.g. with variations in soil moisture, vegetation density, or snow cover (e.g. Prigent et al., 2006). This can seriously affect the retrieval of land surface parameters when a cloud or rain effect is misinterpreted as a surface change.

The objective of this study is to develop a method that indicates a cloud–rain contamination on the passive microwave (MW) observations over land for different ranges of frequencies available on board the successive generations of passive MW satellite instruments. Rain detection schemes have been developed for the Special Sensor Microwave/Imager (SSM/I) over land: they are based on the scattering signal at 85 GHz and use decision trees (Grody, 1991; Ferraro, 1997). Cloud-filtering methods have also been derived for specific applications or for a given instrument. Long et al. (1999) analysed the brightness temperature time series at 85 GHz with different methods to remove the cloud perturbation on the SSM/I images for land surface applications. For the estimation of upper tropospheric humidity with satellite measurements around the water vapour line at 183.31 GHz, Buehler et al. (2007) developed filters with different channels around the line to avoid cloud-contaminated grid cells. Aires et al. (2011) used a neural network (NN) method trained on Meteosat Second Generation Spinning Enhanced Visible and Infrared Imager (SEVIRI) cloud products to create a cloud mask and a classification from the Advanced Microwave Sounding Units A and B (AMSU-A/AMSU-B) with channels from 23 to 183 GHz: statistical models were built separately over land and ocean to detect clouds or classify them into clear sky or low, medium, or high clouds.

Here, we use a similar approach to Aires et al. (2011) to develop a cloud–rain indicator over land for the passive MW imagers used for the estimation of land surface parameters over the last decades. Starting from the late 1970s with the Scanning Multichannel Microwave Radiometer (SMMR), a number of imagers have been launched over the years, including the Special Sensor Microwave Imagers (SSM/I, SSMIS), the Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI) (Kummerow et al., 1998), the Advanced Microwave Scanning Radiometers (AMSR-E, AMSR2), or the WindSat instrument (Gaiser et al., 2004). The latest instrument is the Global Precipitation Measurement (GPM) Microwave Imager (GMI) launched in 2014. Similar frequencies are used across the successive MW imagers, and they have relatively close characteristics (see Table 1) that could allow for similar processing of the data starting from 1978. We can divide the available instruments into three groups based on the imaging frequencies used on each of them:

- Below 40 GHz, a particular set of channels is available on board SMMR that flew from 1978 to 1987, as well as WindSat since 2003. The frequencies available on board GMI are 18.7 (V, H), 23.8 (V), and 36.5 GHz (V, H).
- Frequencies up to 90 GHz are available on the Special Sensor Microwave/Imager (SSM/I) (1987 to present), the Advanced Microwave Scanning Radiometer (AMSR-E/AMSR2) (2002 to present), and the Tropical Rainfall Measurement Mission (TRMM) Microwave Imager (TMI) between 1998 and 2015. On board GMI, the 89 GHz (V, H) frequency is added.
- Frequencies up to 190 GHz are, for instance, used for the Special Sensor Microwave Imager Sounder (SSMIS) (2003 to present) or GMI (2014 to present) with channels at 165.5 (V, H) and 183.3 GHz (V).

All these instruments observe with a similar incidence angle at the surface (as a consequence the angular dependence is not taken into account as with sounders such as AMSU). The available frequencies are close (e.g. 37 GHz for SSM/I against 36.5 GHz for GMI and AMSR2) and have small differences in the operating bandwidth. Note that frequencies below 18 GHz are available for some of these instruments, but they will not be considered here as their sensitivity to clouds is very limited. In this study, the passive microwave observations will come from GMI as it includes all the possible frequencies that we may want to use. Another benefit is that the GPM mission is not Sun synchronous and, as a result, it covers the full diurnal cycle, whereas the other instruments are Sun synchronous with overpassing times at the Equator in the morning and afternoon (SMMR, SSMI, and SSMIS) or at midday and midnight (AMSR-E and AMSR2). The cloud information comes from SEVIRI on board Meteosat: it provides a cloud mask as well as a cloud classification. Rain is

Table 1. Characteristics of the MW imagers over the years.

Imager	Channels (GHz) (polarization)						Spatial resolution (at 37 GHz)	Viewing angle	Operating years
	~ 18 (<i>V, H</i>)	~ 23	~ 36 (<i>V, H</i>)	~ 89 (<i>V, H</i>)	165.5	183.3			
SMMR	18.0	21.0 (<i>V, H</i>)	37.0	–	–	–	29 km × 17 km	50.2°	1978–1987
SSM/I	19.4	22.2 (<i>V</i>)	37.0	85.5	–	–	36 km × 24 km	53.1°	1987–2006
TMI	19.35	21.3 (<i>V</i>)	37.0	85.5	–	–	16 km × 9 km	53.0°	1998–2015
AMSR-E	18.7	23.8 (<i>V, H</i>)	36.5	89.0	–	–	14 km × 9 km	55.0°	2002–2011
WindSat	18.7	23.8 (<i>V, H</i>)	37.0	–	–	–	8 km × 13 km	50–55°	2003–
SSMIS	19.4	22.2 (<i>V</i>)	37.0	91.6	–	<i>H, H</i>	44 km × 28 km	53.1°	2003–
AMSR2	18.7	23.8 (<i>V, H</i>)	36.5	89.0	–	–	12 km × 7 km	55.0°	2012–
GMI	18.7	23.8 (<i>V</i>)	36.64	89.0	<i>V, H</i>	<i>V, V</i>	15 km × 9 km	52.8°	2014–

not detected separately from the cloud per se: some clouds are likely to precipitate and the detection of these clouds will obviously include the detection of rain.

We first describe the data sets relevant for this study (Sect. 2). In Sect. 3, we will elaborate on the methodology. Results will be presented over land surfaces as well as over ocean (to illustrate the difference in behaviour over these two surface types), focusing on the detection of the cloud contamination on the MW observations over land (Sect. 4). Section 5 concludes this study.

2 Data sources

The different data sources are described here, namely the SEVIRI cloud classification and the GMI brightness temperatures (T_{bs}). The steps to create a consistent data set are described, along with a preliminary analysis of the observations. Using ancillary data to help characterize the atmospheric and surface conditions related to the cloud occurrence (such as land surface emissivity atlases) could help the cloud detection but at the cost of increasing the complexity to apply it. For flexibility and convenience, the detection of the cloud contamination will be exclusively built from passive MW observations.

2.1 Cloud mask and classification from Meteosat SEVIRI

Meteosat is a geostationary satellite positioned over the Equator. It covers mostly Africa, South America, Europe, and the Middle East, from $\pm 60^\circ$ latitude and $\pm 60^\circ$ longitude. The SEVIRI channels on board Meteosat encompass the visible and infrared ranges (Schmid, 2000), with varying pixel sizes around 3 km^2 . Algorithms have been developed to provide cloud information, such as cloud-top height, water content, and also cloud classification, every 15 min over the whole field of view (Derrien and Le Gléau, 2005).

The Satellite Application Facility on Climate Monitoring (CMSAF) at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) has provided daily data since 2004. We used the 2013 version of the SE-

Table 2. Cloud classification from SEVIRI (Derrien and Le Gléau, 2005).

Cloud type number	SEVIRI class description
1	Cloud-free land
1	Cloud-free sea
2	Very low clouds
3	Low clouds
4	Medium clouds
5	High opaque clouds
6	Very high opaque clouds
7	High semi-transparent thin clouds
8	High semi-transparent meanly thick clouds
9	High semi-transparent thick clouds
10	High semi-transparent above lower clouds
11	Fractional clouds

VIRI cloud classification algorithm, which provides a robust overview of the different cloud types that matter for vis–IR observations. Using this classification, the goal is to improve our understanding of the MW interaction with clouds and to detect the cloudy situations that impact the MW. Six full days each month in 2015 provide 72 different daily situations that represent a large variation in the possible cloud types and surface conditions, covering the full diurnal and annual cycles. The cloud classes are described in Table 2. High semi-transparent clouds are mostly cirrus of varying thickness, possibly over lower clouds. The fractional cloud class corresponds to cells that are only partly cloudy and to heterogeneous cloud cover. The other cloud types represent the continuum of possible cloud states, with varying opacity and height. Some of these clouds are likely to precipitate, and rain cases are naturally included in the database.

Figure 1 shows the latitudinal variation in the cloud types over land within the SEVIRI disk for February and August 2015. The intertropical convergence zone (ITCZ) location changes between the two seasons, as expected. Over the mid-latitudes, the cloud frequency in February is higher than in

August. The average relative frequency of each cloud type is displayed, showing that all cloud types are well represented.

2.2 The passive microwave observations from GMI

GPM relies on several instruments to provide a precipitation evaluation around the globe. The GMI is on board the core GPM satellite. The satellite has a 65° inclination that allows a non-Sun-synchronous observation of the Earth. The available frequencies range from 10 to 183 GHz (Hou et al., 2014). In this study, we use the calibrated T_{bs} available in the level 1C-R product, where all the channels are projected to a common scan centre position, consistent with the 89 GHz channel resolution (4 km^2).

GMI covers the full frequency range we want to analyse, with an incidence angle close to 53° . In this study, different subsets of the channels will be tested, corresponding to the different channel ranges available on the instruments since 1978. In addition, it observes at different local times, limiting possible biases related to observations at specific times of the day. The GMI data from 2015 have been downloaded for the 72 days corresponding to the SEVIRI selection.

2.3 Data set preparation and preliminary analysis

The SEVIRI and GMI data have very different spatial and temporal resolutions. We need to find the closest matching observations and relocate them on a common grid for further processing. Grid cells with a low-quality flag are avoided for both GMI and SEVIRI. Each GMI observation has a time stamp that is used to find the closest SEVIRI scan. With SEVIRI data every 15 min, there is a maximum of 7.5 min difference between GMI measurements and the corresponding SEVIRI classification. Given the spatial resolution, several SEVIRI cells will obviously fall in one GMI grid cell. In the training data set, only GMI observations associated with a unique target SEVIRI class are kept. There may be some mismatch between observed radiance and the SEVIRI cloud type due to inhomogeneous clouds at a scale lower than the footprint, especially for the lowest-frequency channels. This does not mean that GMI cells with heterogeneous cloud cover will not be able to be classified: it just limits the effect of ambiguous cases during the training phase.

The grid cells located above 55° N and below 50° S are discarded: they are larger in size in the SEVIRI data and are subject to more contamination by snow and ice. The GMI land mask is adopted to separate land and water bodies.

As a first analysis of the MW sensitivity to clouds, the distributions of the MW brightness temperatures (T_{bs}) are plotted in Fig. 2 for the different cloud types and for selected GMI frequencies over ocean (left) and land (right).

With increasing frequency, the atmospheric attenuation increases and the surface contribution to the signal decreases: the difference in the mean T_{bs} between the ocean and land situations diminishes with higher frequencies. Differences in

the signal received by the instrument when it is not totally absorbed by the atmosphere can be due to the cloud effect but can also be related to changes in the surface properties (surface temperature of the ocean or land, wind speed at the ocean surface, soil moisture or vegetation density over land). Cloud types can be preferably associated with some environments, and the surface emissivity change with the surface conditions makes it difficult to find simple relationships between signals and cloud presence. In addition, water vapour modulates the MW signal, and this effect increases with frequency in the window channels.

Over ocean up to 100 GHz, the clouds are detectable and to some extent, their types can be distinguished: there is enough contrast between the radiometrically cold ocean background and the cloud radiation. Above 100 GHz, the surface contribution decreases drastically. The high opaque clouds can present low T_{bs} (the long left tail of the histogram) that are related to the scattering by the cloud-ice phase.

Over land at 18 GHz, the lowest peaks in the histograms for most cloud types (around 265 K) are likely related to the presence of water at the surface. Otherwise, at 18 GHz, the histograms are very similar for all land situations, meaning that this frequency has very limited sensitivity to the cloud presence and type. This can be seen as an asset for land surface characterization with these frequencies, as the signal will not be affected by the cloud presence. At high frequencies, the high opaque clouds present low T_{bs} (the left tails of the histograms) due to the ice scattering in the clouds (as at 166 GHz over ocean). These opaque clouds will likely be detected over land with these high frequencies.

3 Method

Our goal is to detect cloud contamination in the MW observations over land. It is not, at this stage, to classify cloud types. It will nevertheless be interesting to analyse the effects of each cloud type in the different frequency domains.

We focus here on the cloud detection for which a binary classification is required, but we will also experiment with the cloud-type classification. Several methods are available, some of which are rule-based, mostly by using thresholds for the various cloud types (e.g. the SEVIRI cloud algorithm by Derrien and Le Gléau, 2005, or the cloud filter at 183 GHz from Buehler et al., 2007). In this study, we use a statistical approach, similar to the one presented in Aires et al. (2011).

3.1 The training and testing data sets

The training and testing data sets are constructed using the collocated GMI observations and SEVIRI cloud information. To cover the full diversity of cloud situations, a full year of data have been sampled with 72 days (Sect. 2). The SEVIRI acquisition disk excludes the high-latitude regions and does not cover the full snow- and ice-free continents either. How-

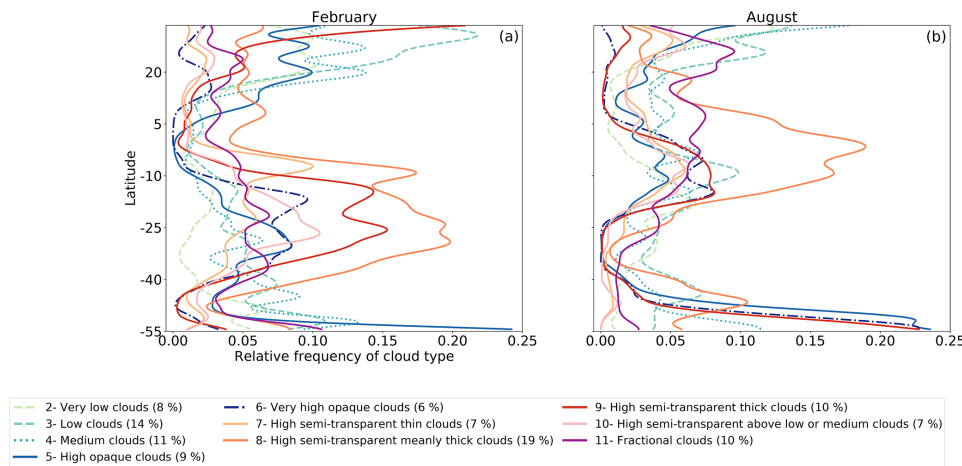


Figure 1. Relative frequency of cloud types as a function of latitude for February (a) and August (b) 2015 over land within the SEVIRI disk. The average frequency of each cloud type over these 2 months is indicated in the legend.

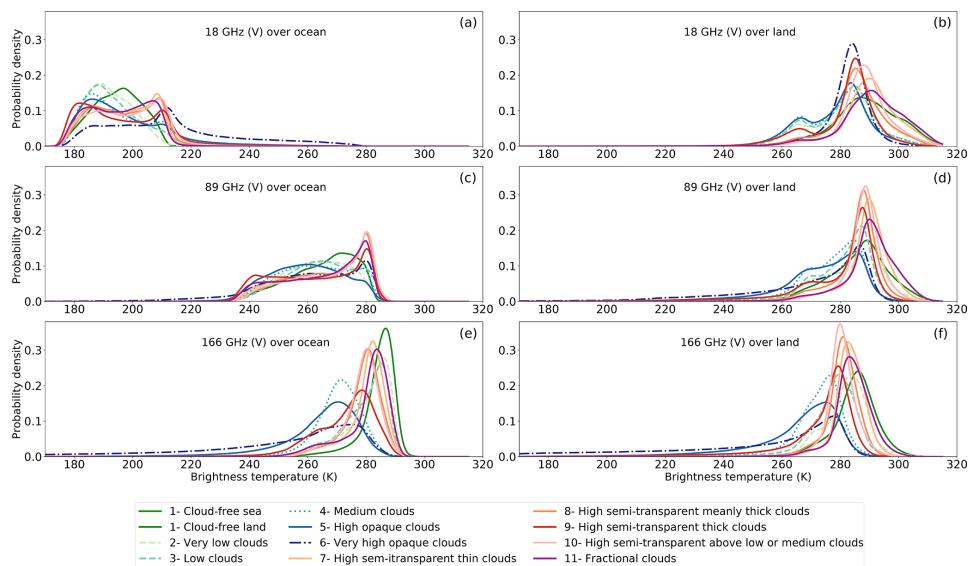


Figure 2. Probability distributions of the GMI-observed T_{b_s} for various cloud types at 18 (a, b), 89 (c, d), and 166 GHz (e, f) for the vertical polarization over ocean (a, c, e) and land (b, d, f) from the filtered data set.

ever, it was shown in Aires et al. (2011) that the calibration of a cloud classification on the SEVIRI disk with MW observations can be extrapolated to the other continents and we are confident that the methodology will be applicable outside the SEVIRI disk, excluding the snow and ice regions.

In the database, we ensure that every cloud type is equally represented. This process ensures that the obtained classification will not be biased towards the most frequent cloud situations, disregarding the less frequent ones. We therefore sample the same number of clear and cloudy situations, with

each cloud type equally represented in the cloudy part. This resulted in 1 million samples for each of the 10 cloud types, and 10 million cloud-free samples. For a cloud classification model, with 11 different possible output classes, the database is built with a similar repartition of classes, giving around 11 million observations. The resulting databases are then randomly divided into the training (80 %) and the testing (20 %) data sets.

3.2 Statistical models

Several statistical models have been tested (e.g. tree-based or logistic regression, results not shown) but we kept an NN classification, based on the MultiLayer Perceptron (MLP) (Rumelhart et al., 1986b). MLPs are universal non-linear approximators, that can, given enough parameters, approximate any function (Hornik, 1991). The NN inputs are the MW channels, their number depending on the frequency ranges (5, 7, or 11). Five neurons (7 and 9) in the hidden layer are used. More neurons and a larger network have been tested, but they did not offer significant improvements in the resulting accuracy (results not shown). The output layer is composed of one binary output (for the cloud detection) or 11 binary outputs (for the cloud classification). The activation in the output layer is a softmax function. The parameters of the MLP classifier are found during the learning stage, in which a binary cross-entropy loss function (Dreiseitl and Ohno-Machado, 2002) is minimized with the back-propagation algorithm (Rumelhart et al., 1986a). Using this loss function allows the continuous output of the NN to be interpreted as a classification probability (Bridle, 1989). The models are implemented using the Keras library (Chollet et al., 2015), and the training is stopped when the loss is not decreasing for five consecutive epochs, which happens after a few hundred epochs depending on the network and input size. The hardware used for this step is a standard office laptop, with four cores and 16 Gb of RAM. After training, the prediction closest to 0 indicates a high probability of having a cloudy grid cell (1 for clear sky). The result of the continuous NN output can then be converted into a binary decision using a threshold to be defined. In the following graphs and results, if not otherwise specified, a decision threshold of 0.5 is applied to derive the binary classification. For multi-class outputs, the highest value among the output neurons is selected as the predicted class. The results are displayed showing the percentage of true positives (cloudy grid cells correctly detected), and true negatives (clear grid cells correctly predicted) from all the samples inside a test set.

4 Results

We first test the methodology over ocean, where clouds are expected to be easier to detect and quantify as we saw from the distributions in Fig. 2. It provides a testing ground for the method, before expanding it to the more difficult land case.

4.1 Detecting clouds over ocean

As described in Sect. 3.1, the database is created with an equal distribution of cloudy and clear conditions and a balanced repartition between the different cloud types. The cloud detection is evaluated for the three MW frequency ranges (all channels, below 100 GHz only, below 40 GHz only), and the results are presented in Table 3 for the test data

set. The cloud detection performs well over ocean, reaching at least 80 % accuracy, even with a reduced number of channels. The low emissivity of the ocean (~ 0.5) and its relative homogeneity makes it possible to correctly detect the cloud presence, even at low MW frequencies.

These cloud detection results are very encouraging and the natural next step is to investigate a cloud classification over ocean, with the same MW frequency ranges. The data set is used with all classes equally sampled, making it suitable for a multiclass classification. Similar NN schemes are implemented, with 11 possible output neurons representing the 10 cloud classes and the clear case for the three frequency ranges. The confusion matrices (Fig. 3) display the results of the classification, showing for each class (y axis) the percentage of samples predicted to belong to 1 of the 11 possible SEVIRI classes (x axis). The diagonal of the confusion matrix shows the correctly classified percentage for each cloud type. The highest accuracy is reached for the cloud-free ocean for the three MW frequency ranges. It is occasionally confused with the high semi-transparent meanly thick clouds (class 8) or the fractional clouds (class 11) as they may not significantly affect the measured T_{bs} . For opaque clouds (classes 2–6), the highest percentages are near the diagonal: these cloud types are correctly classified or classified as a cloud with a similar altitude. We see an increase in the detection of high opaque clouds (classes 4, 5) when the channel at 89 GHz is available. This can be explained by the increased detection of the ice content that this channels offers compared to lower frequencies. When all channels are available the discrimination between cloud layers is even easier, resulting in a better classification. The high semi-transparent clouds (classes 7, 8, 9, and 10) are sometimes incorrectly classified as clear sky, especially with only lower frequencies (due to channels less sensitive to high-altitude phenomena), or high semi-transparent thick clouds (class 8) with higher frequencies, which is expected given that they share similar properties (such as cloud height). Fractional clouds (class 11) are not well classified, the predicted class being either cloud-free or high semi-transparent (class 8).

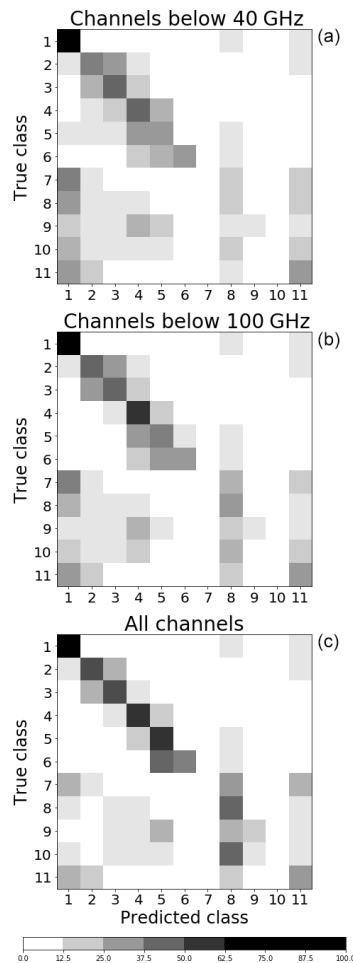
4.2 Detecting clouds over land

A similar cloud detection method is applied over land. The NN classifiers are built using the three different MW frequency ranges as inputs and with one output indicating the clear vs. cloudy probability.

The specifics of the model and database are described in Sect. 3.1 and 3.2. Similar to Table 3 over ocean, Table 4 (top part) presents the accuracies reached over land by the three frequency ranges. The classification performance deteriorates compared to the ocean case, as expected. Nevertheless, even for the worst case (with only five low frequency channels available), true positive and negative detections are close to 70 %.

Table 3. Results of a binary classification over the ocean for different MW frequency ranges.

	All channels (%)	Below 100 GHz (%)	Below 40 GHz (%)
Clear grid cells correctly predicted	91	89	89
Cloudy grid cells correctly predicted	81	74	72

**Figure 3.** Confusion matrix over the ocean for the cloud type classification for the three MW frequency ranges: below 40 GHz (a), below 100 GHz (b), and all channels (c). The cloud type numbers are detailed in Table 2.

The result of the detection has been analysed further, as a function of the cloud type (lower part of Table 4). Note that these are only a detail of the previous results (top part of Table 4) separated by each original cloud type. Large differences are observed between cloud types. For non-semi-transparent clouds, the higher the cloud the better the detection rate: this is directly related to the presence of ice in high

clouds that can scatter the MWs. The higher the frequency, the better the detection of ice phase. Likewise, high semi-transparent clouds can be detected only when they are thick enough.

4.3 Detecting cloud-contaminated microwave observations over land

The previous results showed that MWs cannot detect all clouds seen by vis–IR measurements, especially when only a subset of the frequencies is available. This behaviour is actually very attractive for “all weather” land surface applications with MWs. However, for accurate land surface characterization with MW, we need to identify the cloudy situations that really contaminate the MW. To that end we use the results from the previous model to select an appropriate definition of cloud contamination in the MW. For all frequency ranges, high semi-transparent thin clouds, high semi-transparent meanly thick clouds, and the fractional clouds (i.e. classes 7, 8, and 11), the classification accuracy is close to 50 %, similar to a random class assignment, meaning that these frequency ranges are not affected enough by these cloud types to be able to detect them. To focus on the clouds that do impact the MWs, we rebuild a training data set, suppressing the three ambiguous classes previously mentioned (namely classes 7, 8, and 11). The idea behind this new training database is that removing ambiguities at the learning stage will improve the classification. In other words, removing the ambiguous SEVIRI cloud types from the training database allows the model to ignore these phenomena, which are mostly detected in vis–IR. The lower sensitivity to clouds in MW is thus accounted for in the new training data set.

The results of this new classification are provided in Table 5, separately for the clear grid cells (class 1), for the cloudy grid cells with clouds that do contaminate the MW (the MW cloud-contaminated grid cells, i.e. classes 2, 3, 4, 5, 6, 9, and 10), and for the cloudy grid cells corresponding to the three cloud types that are difficult to detect with MW (the ambiguous grid cells, ignored in the training data set, i.e. classes 7, 8, and 11).

The results show that the clear-sky detection increases and so does the detection of MW cloud-contaminated cells (84 % with all frequencies) compared to the detection of cloudy cells in Table 4 (77 % with all frequencies). This is expected, as the ambiguous cases have been removed from the statistics; it is also consistent with the number of ambiguous cells (ignored in the training data sets) that are predicted as MW

1538 S. Favrichon et al.: Detecting cloud contamination in passive microwave satellite measurements over land

Table 4. Top part shows percentage of correct cloud detection from the test set over land. Lower part shows details of the percentage of each cloud type predicted as cloudy. The results are presented for the three MW frequency ranges.

	All channels (%)	Below 100 GHz (%)	Below 40 GHz (%)
Clear grid cells correctly predicted	83	73	69
Cloudy grid cells correctly predicted	77	73	73
2 – Very low clouds	63	70	71
3 – Low clouds	77	78	77
4 – Medium clouds	92	85	83
5 – High opaque clouds	97	85	83
6 – Very high opaque clouds	98	92	90
7 – High semi-transparent thin clouds	59	56	54
8 – High semi-transparent meanly thick clouds	66	61	64
9 – High semi-transparent thick clouds	89	80	80
10 – High semi-transparent above lower clouds	84	74	71
11 – Fractional clouds	53	48	46

Table 5. Classification results for the different clear and cloudy populations for the three MW frequency ranges. See text for more details.

	All channels (%)	Below 100 GHz (%)	Below 40 GHz (%)
Clear cells correctly predicted	88	77	71
MW cloud-contaminated cells correctly predicted	84	76	78
Ambiguous grid cells predicted as MW cloud contaminated	49	43	52

cloud contaminated by the new classification (close to 50 % regardless of the frequency range).

The original output of the classification is not binary, but a number between 0 and 1 (see Sect. 3.2). In the results shown so far a decision threshold at 0.5 has been adopted to separate the two classes. Would it be possible to adjust this threshold for a better detection of the cloud-contaminated observations? Figure 4 presents the outputs of the NN classifier for the three populations previously defined in Table 5 and for each MW frequency range (Fig. 4).

Figure 4 (top and middle panels) confirms that the clear grid cells and the MW cloud-contaminated grid cells are confidently classified, with very distinct output distributions for these two populations, 0 indicating a high confidence to be in the MW cloud-contaminated class and 1 a high confidence to be in the clear class. Nevertheless, when channels above 100 GHz are not available, a non-negligible fraction of the clear grid-cell population is classified between 0.1 and 0.4, meaning that the confidence in the prediction is lower. For the ambiguous cloud types that were ignored during the training (bottom panel), the distribution of the outputs covers a large range of values, conveying the uncertainty in the prediction. However, with the full frequency range there are a number of observations labelled as confidently contaminated (peak in low NN output values); this can be expected due to the better sensitivity of the high-frequency channels to thin clouds. Figure 4 clearly shows that, depending on the decision threshold selected for the NN output values, it is possible to filter out more or less ambiguous grid cells. So far

it has been at 0.5, but it could be modified. The selection of this threshold should depend on the frequency range and the application.

For instance, for land surface temperature estimates, the idea is to avoid the clouds that really affect the low microwave T_{bs} (below 40 GHz) that are used for the retrieval of this parameter (e.g. Prigent et al., 2016; Jiménez et al., 2017). Note, however, that this does not exclude the use of the higher frequencies for cloud-contamination detection if these frequencies are also available. In addition, the interest of the MW for the land surface temperature estimation is to complement the infrared estimations that are not available under cloudy conditions: as a consequence, only the seriously cloud-contaminated MW observations should be detected to maintain a quasi “all weather” coverage of the MW estimates while limiting erroneous estimates under very cloudy/rainy situations. In that framework, the role of the cloud classification is to make sure the cloud-contaminated observations are correctly detected. The correct detection of the true clear cases is of lesser importance.

Figure 5 presents the percentage of MW observations predicted as cloud contaminated, as a function of the threshold on the NN classifier output for both the MW cloud-contaminated cases (the true positive, solid line) and the clear-sky cases (the false positive, dashed line). It shows that a threshold below 0.1 keeps the percentage of misclassified clear-sky cases low (low percentage of false positives). Combined with the results from Fig. 4 (middle panel), a threshold at 0.05 and 0.01 could also be tested to only classify the

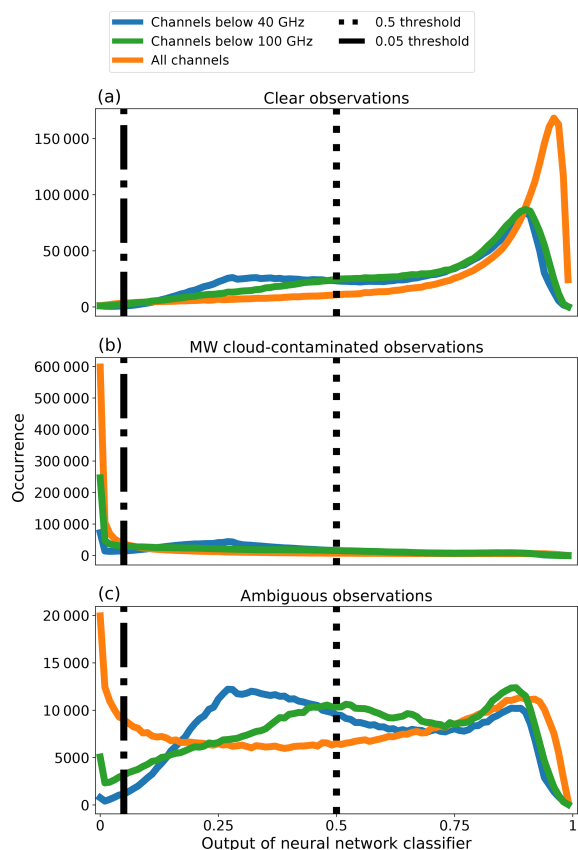


Figure 4. Model output probability distributions for the clear grid cells (a), the MW cloud-contaminated grid cells (b), and for the ambiguous grid cells (c), for the three MW frequency ranges. See text for more detail about the three populations.

cloud-contaminated observations with a high degree of confidence.

A day of GMI observations, 15 June 2015, is selected to illustrate the potential of the classification of the MW cloud contamination. Note that this day is not included in the training or testing data sets previously used. For the three MW frequency ranges, the classification is applied with the selected thresholds (0.1, 0.05, 0.01). Table 6 provides the percentage of observations classified as cloud contaminated for each set-up, along with the results from the Ferraro (1997) precipitation detection algorithms based on a decision tree and thresholds on channels. As expected, when the high-frequency channels are included, the sensitivity of our methodology to the cloud contamination increases, as does the percentage of cloud-contaminated observations, with $\sim 10\%$ cloud-contaminated observations for this frequency range. Note that, for that day, the coincident SEVIRI observations are cloudy at 29%, i.e. 3 times more than the results from the highest detection of the high MW frequency

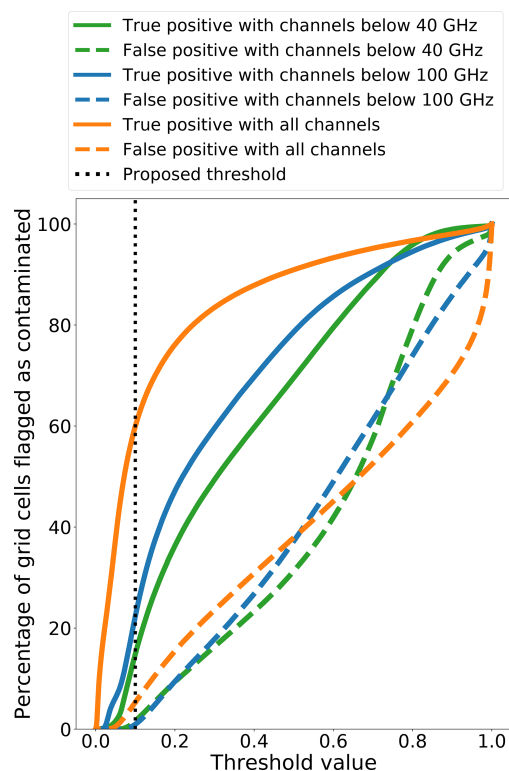


Figure 5. Evolution of the percentage of MW observations correctly classified as cloud contaminated (true positive, solid lines), and clear-sky grid cells incorrectly classified as being contaminated (false positive, dashed line), as a function of the NN output threshold for the three MW frequency ranges. Note that, for this data set, half the observations are cloudy according to SEVIRI.

range. Using only frequencies below 40 GHz, the percentage of cloud-contaminated observations decreases. This illustrates the benefit of using lower MW frequency channels for “all weather” land surface characterization, with a ratio of 4 between the number of contaminated observations when adding the 89 GHz to the frequencies below 40 GHz (using the 0.05 threshold). For all these threshold–model combination, the number of clear-sky observations (according to SEVIRI) incorrectly flagged stays below 0.5% of all observations.

For comparison purposes, the Ferraro (1997) rain detection algorithms are also run and compared to both the algorithm using the 85 GHz channel and the one limited to the frequencies below 40 GHz. The results in the last line of the table show the number of observations that are flagged as precipitating. As expected the number of precipitating situations is lower than the number of cloud-contaminated MW observations. For the models with channels above 40 GHz, more than 90% of the precipitating observations are detected by our method. The model with only channels below 40 GHz

still retrieves more than 50 % of the precipitating observations when the 0.1 threshold is used.

Nevertheless, depending on the applications and the degree of uncertainty required in the land surface product, if the full frequency range up to 100 GHz is available on the instrument, it can be relevant to use all the frequencies up to 100 GHz to filter out the cloud-contaminated grid cells, even if only the frequencies below 40 GHz are used in the retrieval of the land surface parameter. As an example, if the land surface temperature is to be retrieved with very low uncertainty from SSM/I observations (an instrument that has channels up to 90 GHz), it can be wise to use the full frequency range to detect the cloud contamination, even if only the lower frequencies below 40 GHz are used in the retrieval.

Now that we have an estimate of the number of points that are flagged by each model with different thresholds we can plot the global map of the locations of these contaminated cells. Figure 6 shows the results for the three different frequency groups and with three thresholds applied. The thresholds were chosen based on the results in Table 6 to illustrate how different thresholds might be applied to each model while still providing coherent estimates of cloud-contaminated grid cells.

In Fig. 6, models are applied to the data over land to create the three maps. For each map a different threshold is applied: 0.1 with the lowest channels (a), 0.05 with channels up to 100 GHz (b) and 0.01 with all channels available (c). The fourth subplot (d) is the precipitating observations according to Ferraro's 89 GHz algorithm. The fifth subplot (e) shows the SEVIRI cloud type. We can analyse the output of this map:

- The agreement between models and the increased number of flagged points with more channels is clearly visible (a, b, c).
- In some areas, the cloudy grid cells do not appear to be detected (i.e. red area). When looking at the detail of the SEVIRI cloud types (subplot e) in that area we find out they are mostly fractional/semi-transparent or low clouds, which explains the low contamination rate, according to our definition.
- In the pink area, we have a stronger detection of contaminated grid cells. Indeed the most represented cloud types are high semi-transparent thick clouds (23 %), high semi-transparent clouds above low or medium clouds (20 %), and very high opaque clouds (17 %). All these cloud types are the ones that might affect the measurement the most.
- We find that the precipitating observations are correctly found within the detected cloudy cells, but there are more cloud-contaminated observations.

This global application of our models shows the possible use of different frequency ranges to detect contaminated ob-

servations. Although adding more information by using the channels more sensitive to ice content leads to a better detection of cloud contamination, we show here that it is possible to filter out cloud-contaminated measurements even above land with a restricted number of channels. The thresholds used here are coherent for the specific application shown in this study, with a low number of misclassified clear-sky grid cells and also with the real-world occurrence of deep convective phenomena that contaminate the observations the most. Indeed, the International Satellite Cloud Climatology Project (ISCCP) data show that they have an average occurrence of 2.6 % for deep convections that is of the same magnitude as our cloud index associated with the proposed thresholds (Rossow and Schiffer, 1999).

5 Conclusions

Passive microwave observations from satellites are less sensitive to clouds than visible–infrared measurements and can provide an almost “all weather” land surface characterization. However, cloud (and possible rain) can affect the microwave observations, even at frequencies below 40 GHz. For an accurate estimation of land surface parameters, cloud-contaminated MW observations have to be detected to avoid interpreting a cloud presence as a surface change.

A methodology has been developed to detect cloud contamination on passive MW observations over land (except snow- and ice-covered areas). It is based on a NN classification, trained on collocated SEVIRI cloud types. The NN output indicates the probability of cloud contamination in the MW signal for a given MW frequency range. The cloud-contamination index is provided with values in the 0–1 range: the threshold applied to this index can be customized to fit the required application needed to flag out the contaminated observations. Although the target here is cloud detection over land surfaces, the model was also tested over the simpler case of detection over ocean. The index confidence increased with the number of channels available and performed better over the ocean as expected. In all cases, even with a reduced number of information over land, the detection of contaminated observations is performed with more than 70 % accuracy.

An example of a possible application of this cloud-contamination index was shown to eliminate grid cells unsuitable for land surface temperature estimation. The index proved useful to signal cloud contamination for this particular application and will soon be applied to the quality control of a long time record of land surface temperatures (Prigent et al., 2016). The land surface temperature estimate is essentially based on passive microwave frequencies between 18 and 40 GHz, from a succession of satellite imagers since 1978 (SMMR, SSM/I, and SSMIS). The first instrument only measured up to 36 GHz, contrarily to the last instruments. So far, the cloud and/or rain detection indices are based on thresholds related to channels around 85 GHz (Jiménez et al.,

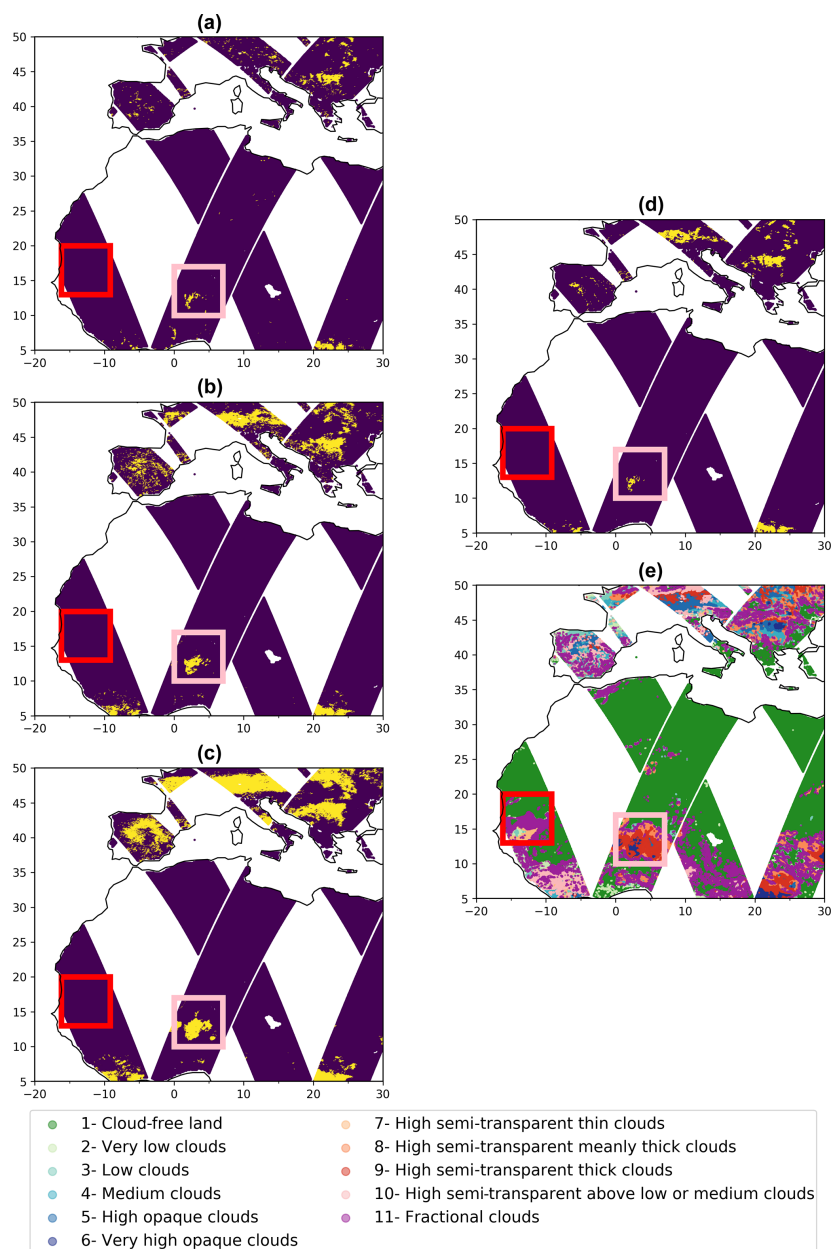


Figure 6. Maps showing for the 15 June 2015: (a) the predicted grid cells flagged by the model using channels below 40 GHz with a 0.1 threshold; (b) by the model using channels below 100 GHz with a 0.05 threshold; (c) by the model using channels below 190 GHz with a 0.01 threshold; (d) the detected precipitating cells according to Ferraro (1997); and (e) the cloudy classes from SEVIRI. The red and the pink square boxes highlight two smaller regions further discussed in Sect. 4.3.

1542 S. Favrichon et al.: Detecting cloud contamination in passive microwave satellite measurements over land

Table 6. Percentage of MW observations classified as cloud contaminated for the three MW frequencies ranges, with different thresholds on the NN classifier output. Results are presented for 15 June 2015 over land surfaces within the SEVIRI disk. The last line of the table presents the percentage of observations detected as precipitating with the Ferraro method using channels up to 100 GHz or only below 40 GHz.

Threshold used	All channels (%)	Below 100 GHz (%)	Below 40 GHz (%)
0.1	9.7	5.1	1.5
0.05	7.9	3.6	0.8
0.01	5.2	2.0	0.3
Ferraro (1997)	–	1.4	0.4

2017). This frequency is not available on board SMMR and the new methodology for the frequency range below 40 GHz will be applied to the full data set, with possible comparisons with the current method up to 100 GHz, when these channels are available. Overall the models developed in this study can be applied globally in ice- and snow-free areas and are potentially useful for numerous applications where it is of interest to identify possible cloud contaminations in observed MW radiances. In addition to the land surface temperature example, this index can be useful for selecting clear scenes for accurate MW emissivity estimation (Moncet et al., 2011) or to detect cloudy scenes for the analysis of deep convections (Prigent et al., 2011).

Data availability. The CLAAS-2 Cloud property data set using SEVIRI – Edition 2 (CLAAS-2, https://doi.org/10.5676/EUM_SAF_CM/CLAAS/V002; Finkensieper et al., 2016) is publicly available from the Satellite Application Facility on Climate Monitoring (CM SAF). The GPM GMI_R Common Calibrated Brightness Temperatures Collocated L1C 1.5 h 13 km V05 (GPM_1CGPMGMI_R, <https://doi.org/10.5067/GPM/GMI/R/1C/05>; Berg, 2016) is provided by NASA.

Author contributions. All authors have been involved in interpreting the results, discussing the findings, and editing the paper. SF conducted the main analysis and wrote the draft of the paper. CJ, FA, and CP provided guidance on using the data sets and expertise on analysing the results.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study was partly funded by the Centre National Centre National d'Études Spatiales (CNES, projects GPM-R and ISMAR) and the European Space Agency (ESA) project LST-CCI (contract no. 4000123553/18/I-NB). We acknowledge Hervé Le Gléau and Gaelle Kerdraon (Centre de Météorologie Spatiale de Météo-France) for providing guidance to interpret the vis-IR cloud classification, Martin Stengel (Deutscher Wetterdienst) for providing expertise on using the SEVIRI data, and Die

Wang (Brookhaven National Laboratory) and Victorial Galligani (Centro de Investigaciones del Mar y la Atmósfera) for providing valuable comments to improve the manuscript. We also extend our thanks to the referees for the constructive insights that were given during the review process. The final revisions to this study were conducted with the help of Marloes Gutenstein-Penning de Vries (Associate Editor for *Atmospheric Measurement Techniques*).

Edited by: Marloes Gutenstein-Penning de Vries

Reviewed by: two anonymous referees


References

- Aires, F., Prigent, C., Rossow, W. B., and Rothstein, M.: 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.-Atmos.*, 106, 14887–14907, <https://doi.org/10.1029/2001JD900085>, 2001.
- Aires, F., Marquisseau, F., Prigent, C., and Sèze, G.: A Land and Ocean Microwave Cloud Classification Algorithm Derived from AMSU-A and -B, Trained Using MSG-SEVIRI Infrared and Visible Observations, *Mon. Weather Rev.*, 139, 2347–2366, <https://doi.org/10.1175/MWR-D-10-05012.1>, 2011.
- Berg, W.: GPM GMI_R Common Calibrated Brightness Temperatures Collocated L1C 1.5 h 13 km V05, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), available at: <https://doi.org/10.5067/GPM/GMI/R/1C/05>, 2016.
- Bridle, J. S.: Probabilistic Interpretation of Feedforward Classification Network Outputs with Relationships to Statistical Pattern Recognition, NATO ASI Series in Systems and Computer Science, 227–236, https://doi.org/10.1007/978-3-642-76153-9_28, 1989.
- Buehler, S. A., Kuvatov, M., Sreerekha, T. R., John, V. O., Rydberg, B., Eriksson, P., and Notholt, J.: A cloud filtering method for microwave upper tropospheric humidity measurements, *Atmos. Chem. Phys.*, 7, 5531–5542, <https://doi.org/10.5194/acp-7-5531-2007>, 2007.
- Chollet, F. et al.: Keras, available at: <https://keras.io> (last access: 1 January 2018), 2015.
- Derrien, M. and Le Gléau, H.: MSG/SEVIRI cloud mask and type from SAFNWC, *Int. J. Remote Sens.*, 26, 4707–4732, <https://doi.org/10.1080/01431160500166128>, 2005.
- Dreiseitl, S. and Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review,

- J. Biomed. Eng., 35, 352–359, [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0), 2002.
- Ferraro, R. R.: Special sensor microwave imager derived global rainfall estimates for climatological applications, *J. Geophys. Res.*, 102735, 715–716, <https://doi.org/10.1029/97JD01210>, 1997.
- Finkensieper, S., Meirink, J.-F., van Zadelhoff, G.-J., Hanschmann, T., Benas, N., Stengel, M., Fuchs, P., Hollmann, R., and Werscheck, M.: CLAAS-2: CM SAF CLOUD property dAtAset using SEVIRI – Edition 2, Satellite Application Facility on Climate Monitoring, https://doi.org/10.5676/EUM_SAF_CM/CLAAS/V002, 2016.
- Freitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., and Perdigão, R.: Land surface temperature from multiple geostationary satellites, *Int. J. Remote Sens.*, 34, 3051–3068, <https://doi.org/10.1080/01431161.2012.716925>, 2013.
- Gaiser, P. W., St. Germain, K. M., Twarog, E. M., Poe, G. A., Purdy, W., Richardson, D., Grossman, W., Jones, W. L., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R. M., and Chang, P. S.: The windSat spaceborne polarimetric microwave radiometer: Sensor description and early orbit performance, *IEEE T. Geosci. Remote.*, 42, 2347–2361, <https://doi.org/10.1109/TGRS.2004.836867>, 2004.
- Greenwald, T. J., Stephens, G. L., Vonder Haar, T. H., and Jackson, D. L.: A physical retrieval of cloud liquid water over the global oceans using special sensor microwave/imager (SSM/I) observations, *J. Geophys. Res.*, 98, 18471, <https://doi.org/10.1029/93JD00339>, 1993.
- Grody, N. C.: Classification of snow cover and precipitation using the special sensor microwave imager, *J. Geophys. Res.*, 96, 7423–7435, <https://doi.org/10.1029/91JD00045>, 1991.
- Hornik, K.: Approximation capabilities of multilayer feedforward networks, *Neural Networks*, 4, 251–257, [https://doi.org/10.1016/0893-6080\(91\)90009-t](https://doi.org/10.1016/0893-6080(91)90009-t), 1991.
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T.: The global precipitation measurement mission, *B. Am. Meteorol. Soc.*, 95, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>, 2014.
- Jiménez, C., Prigent, C., Ermida, S. L., and Moncet, J. L.: Inversion of AMSR-E observations for land surface temperature estimation: 1. Methodology and evaluation with station temperature, *J. Geophys. Res.*, 122, 3330–3347, <https://doi.org/10.1002/2016JD026144>, 2017.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J.: The Tropical Rainfall Measuring Mission (TRMM) sensor package, *J. Atmos. Ocean. Tech.*, 15, 809–817, [https://doi.org/10.1175/1520-0426\(1998\)015<0809:TTRMMT>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<0809:TTRMMT>2.0.CO;2), 1998.
- Long, D. G., Remund, Q. P., and Daum, D. L.: A cloud-removal algorithm for SSM/I data, *IEEE T. Geosci. Remote.*, 37, 54–62, <https://doi.org/10.1109/36.739119>, 1999.
- Moncet, J.-L., Liang, P., Galantowicz, J. F., Lipton, A. E., Uymin, G., Prigent, C., and Grassotti, C.: Land surface microwave emissivities derived from AMSR-E and MODIS measurements with advanced quality control, *J. Geophys. Res.-Atmos.*, 116, D16104, <https://doi.org/10.1029/2010JD015429>, 2011.
- Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, *Nature*, 540, 418–422, <https://doi.org/10.1038/nature20584>, 2016.
- Prata, A. J., Caselles, V., Coll, C., Sobrino, J. A., and Oettle, C.: Thermal remote sensing of land surface temperature from satellites: current status and future prospects, *Remote Sens. Rev.*, 12, 175–224, <https://doi.org/10.1080/02757259509532285>, 1995.
- Prigent, C., Aires, F., and Rossow, W. B.: Land surface microwave emissivities over the global for a decade, *B. Am. Meteorol. Soc.*, 87, 1573–1584, <https://doi.org/10.1175/BAMS-87-11-1573>, 2006.
- Prigent, C., Rochetin, N., Aires, F., Defer, E., Grandpeix, J.-Y., Jimenez, C., and Papa, F.: Impact of the inundation occurrence on the deep convection at continental scale from satellite observations and modeling experiments, *J. Geophys. Res.-Atmos.*, 116, D24118, <https://doi.org/10.1029/2011JD016311>, 2011.
- Prigent, C., Jimenez, C., and Aires, F.: Toward “all weather”, long record, and real-time land surface temperature retrievals from microwave satellite observations, *J. Geophys. Res.*, 121, 5699–5717, <https://doi.org/10.1002/2015JD024402>, 2016.
- Rossow, W. B. and Schiffer, R. A.: Advances in Understanding Clouds from ISCCP, *B. Am. Meteorol. Soc.*, 80, 2261–2287, [https://doi.org/10.1175/1520-0477\(1999\)080<2261:AIUCFI>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2), 1999.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, 323, 533–536, <https://doi.org/10.1038/323533a0>, 1986a.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Learning Internal Representations by Error Propagation, 318–362, MIT Press, Cambridge, MA, USA, 1986b.
- Schmid, J.: The SEVIRI instrument, Proceedings of the 2000 EUMETSAT meteorological satellite data user’s conference, Bologna, Italy, 29 May–2 June 2000, 2000.
- Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., and Lehair, W. J.: A closer look at the ABI on the goes-r series, *B. Am. Meteorol. Soc.*, 98, 681–698, <https://doi.org/10.1175/BAMS-D-15-00230.1>, 2017.
- Spencer, R. W., Goodman, H. M., and Hood, R. E.: Precipitation Retrieval over Land and Ocean with the SSM/I: Identification and Characteristics of the Scattering Signal, *J. Atmos. Ocean. Tech.*, 6, 254–273, [https://doi.org/10.1175/1520-0426\(1989\)006<0254:PROLAO>2.0.CO;2](https://doi.org/10.1175/1520-0426(1989)006<0254:PROLAO>2.0.CO;2), 1989.
- Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *Int. J. Remote Sens.*, 26, 4485–4498, <https://doi.org/10.1080/01431160500168686>, 2005.

Article

A Method to Downscale Satellite Microwave Land-Surface Temperature

 Samuel Favrichon ^{1,2,*} , Catherine Prigent ^{1,2} and Carlos Jiménez ^{2,1}

¹ Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, 75006 Paris, France; catherine.prigent@observatoiredeparis.psl.eu (C.P.); carlos.jimenez@estellus.fr (C.J.)

² Estellus, 75002 Paris, France

* Correspondence: samuel.favrichon@observatoiredeparis.psl.eu

Abstract: High-spatial-resolution land-surface temperature is required for several applications such as hydrological or climate studies. Global estimates of surface temperature are available from sensors observing in the infrared (IR), but without ‘all-weather’ observing capability. Passive microwave (MW) instruments can also be used to provide surface-temperature measurements but suffer from coarser spatial resolutions. To increase their resolution, a downscaling methodology applicable over different land environments and at any time of the day is proposed. The method uses a statistical relationship between clear sky-predicting variables and clear-sky temperatures to estimate temperature patterns that can be used in conjunction with coarse measurements to create high-resolution products. Different predicting variables are tested showing the need to use IR-derived information on vegetation, temperature diurnal evolution, and a temporal information. To build a true ‘all-weather’ methodology, the effect of clouds on surface temperatures is accounted for by correcting the clear-sky diurnal cycle amplitude, using cloud parameters from meteorological reanalysis. Testing the method on a coarse IR synthetic data at ~25 km resolution yields a Root Mean Square Deviations (RMSD) between the ~5 km high-resolution and downscaled temperatures smaller than 1 °C. When applied to observations by the Special Sensor Microwave Imager Sounder (SSMIS) at ~25 km resolution, the downscaling to ~5 km yields a smaller RMSD compared to IR observations. These results demonstrate the relevance of the methodology to downscale MW land-surface temperature and its potential to spatially enhanced the current ‘all-weather’ satellite monitoring of surface temperatures.

Keywords: microwave remote sensing; land-surface temperature; all-weather; spatial resolution



Citation: Favrichon, S.; Prigent, C.; Jiménez, C. A Method to Downscale Satellite Microwave Land-Surface Temperature. *Remote Sens.* **2021**, *13*, 1325. <https://doi.org/10.3390/rs13071325>

Academic Editor: Zhao-Liang Li

Received: 4 March 2021

Accepted: 26 March 2021

Published: 31 March 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land-Surface Temperature (T) is the physical temperature of the uppermost layer of the Earth continental surfaces. It plays a key role in the land–atmosphere exchanges, driving the longwave surface-emitted radiation, and controlling the partitioning between latent and sensible heat fluxes. It is one of the Essential Climate Variable (ECV) of the Global Climate Observing System of the World Meteorological Organization [1]. Development of global, long time series of T data records is currently encouraged, but it faces difficulties. Contrarily to ocean surfaces, land surfaces have a lower heat capacity, and when exposed to solar fluxes, T undergoes strong diurnal and seasonal variations, with high spatial variability related to surface properties such as vegetation density or soil moisture [2].

T at the global scale can be estimated from satellite observations of the thermal emission in the infrared (IR) [3]. From geostationary satellites, T maps with a 15 min sampling time at ~5 km spatial resolution are derived [4]. Higher resolutions are reached with polar orbiters (up to a few hundreds of meters), but at the expense of the revisit time [5]. However, IR cannot penetrate clouds, and consequently, T estimation from IR observations suffers from a clear-sky bias [6]. To overcome this limitation, some products

use T derived from a land-surface energy balance model to fill the cloud gaps in the IR-derived T estimates (e.g., [7,8]). Alternative methods to provide ‘all-weather’ land-surface temperature rely on sensor fusion method [9–11], using a combination of model or IR measurements.

Satellite passive microwave (MW) observations are an attractive alternative for the T estimation under cloudy conditions, as they are much less affected by clouds than the IR. Ref. [12–14], or [15] developed methodologies to retrieve T from MW polar imagers such as the Special Sensor Microwave Imager (SSM/I), the Special Sensor Microwave Imager Sounder (SSMIS), or the Advanced Microwave Scanning Radiometer (AMSR). Some of these instruments have been in operation for a long time now (~40 years), opening up the prospect of creating long term series of consistent ‘all-weather’ T estimates. However, the spatial resolution of the MW T is significantly lower than the IR one (~25 km), and MW T is generally less accurate than the IR one due to the larger dependence on the more varying surface emissivities in the MW compared to the IR, the smaller dependence of the measurements to T , and the impact of subsurface emissions for very dry soils [16,17].

Downscaling of MW T , i.e., the increase of the spatial resolution of these products, would greatly benefit the T user community, and would facilitate the comparison of IR and MW retrievals. Downscaling has already been applied to improve the spatial resolution of the IR T itself [18,19], but in the MW it has so far been more widely applied to other parameter estimates, such as soil moisture [20,21]. Downscaling methods are typically based on establishing either a statistical relationship, or a physically based model, between the initial parameter at a coarse spatial resolution and the ancillary variables that will describe the fine spatial structures. Ref. [19,21] review different downscaling methodologies and insist on the importance of the selection of the ancillary information at high spatial resolution needed to provide the high-resolution spatial patterns. This information is typically extracted from vegetation-related satellite products derived from visible and infrared observations, or from digital elevation models. Both vegetation and elevation are expected to affect the T spatial distribution, with the vegetation increasing the surface heat capacity and the elevation decreasing the local T . Another potential source of information is the T diurnal cycle itself if it can be made available from model or observation, as the T cycle is modulated by the surface conditions [22,23].

A few attempts have already been conducted to downscale the MW T . [24] combined MW T estimates from AMSR-E observations at ~25 km and Moderate-Resolution Imaging Spectroradiometer (MODIS) IR T at 1 km. Under cloudy sky, the high-spatial-resolution information is provided by a digital elevation map (DEM) and a linear correction is applied to correct for the altitude effect on the MW T at low spatial resolution. The retrieval is applied over China, at the overpassing times of the Aqua satellite (01:30 and 13:30). Similarly, [25] calibrated a MW T retrieval from AMSR observations with infrared MODIS and Geostationary Operational Environmental Satellite (GOES) T estimates, to fill the cloudy gaps with the MW and they used the normalized difference vegetation index (NDVI) and elevation information to downscale the MW estimates over North America. In the two methods described only static information (such as DEM or vegetation index) are used under clouds and the T patterns of cloudy scenes are expected to be the same as the ones for clear sky. An alternative is proposed by [26], also noticing that the methods adopted for downscaling MW T fail to quantify the effect of temporal variations related to the diurnal T cycle, limiting their use to a single overpassing time of the satellite. They used high-resolution vegetation, elevation information at 1 km resolution and data assimilation model outputs to downscale the MW T from AMSR-E. A decomposition of the diurnal cycle variation at each location into different components (annual variation, meteorological variation, and a daily cycle) is adopted to correct the MW-derived high-resolution T estimates and it is applied over northeastern China. However, the complex input required for this method hinder its applicability.

In this study, a simpler methodology is proposed to downscale MW T for global application, regardless of the overpassing time of the satellite, for both clear and cloudy scenes.

The main benefits of this method are the possible application to different instruments, with a downscaling method valid regardless of the local time of the satellite observations. So far, only the study in [26] accounted for this issue, the other methods being specific to a given overpassing time. Within a coarse resolution MW T ground footprint, heterogeneous surfaces can have different heat capacities and related T diurnal cycles, and this must be taken into account. In addition, it is also important to develop a downscaling scheme that can operate efficiently under both clear and overcast scenes. To our knowledge, downscaling methods were so far applied to the MW T estimates, without considering the cloud cover. However, it is expected that clouds affect the local T . For instance, the study by [6], showed significantly colder daytime T under clouds than under clear conditions. This difference is expected to depend upon the surface types, with large heat capacity areas (e.g., vegetated and wet surfaces) less prone to T changes than other surfaces (such as dry arid areas), upon the cloud characteristics (e.g., its optical thickness), as well as upon the time of the day.

To obtain a high-resolution MW T product, the procedure described here relies on a statistical relationship between high-resolution ancillary information and the T estimates. This statistical function is first derived for clear-sky conditions using IR T and several ancillary products from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) [27]. The original ~ 5 km T estimates are used as the fine resolution T , with the coarse resolution T required to test the method being synthetically produced from aggregation of the original T to a ~ 25 km resolution. For cloudy scenes, the same statistical function will be used, but with a prior correction of the input to account for the change caused by cloudy conditions. This correction is derived by studying the relationships between the downscaling inputs and cloud-relevant variables from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis [28]. Finally, the downscaling with the cloud correction is illustrated by deriving a real MW T scene at ~ 5 km spatial resolution from an original T scene at ~ 25 km resolution from a Special Sensor Microwave Imager / Sounder (SSMIS) product currently under development by the European Space Agency (ESA) Land-Surface Temperature (LST) Climate Change Initiative (LST-CCI) project [29,30].

The study is organized as follows: the study datasets are described in Section 2, then the downscaling methods and overall processing are given in Section 3, and the methodology is evaluated in Section 4, followed by a discussion of the results obtained in Section 5. Finally, Section 6 presents the study conclusions.

2. Data

2.1. SEVIRI

Statistical downscaling relies on information at the high spatial resolution aimed for. As mentioned in the introduction, the objective is to develop a method that is flexible enough to be applicable to different local times. Therefore, satellite surface information sampled at different times of the day is required, which restricts the choice of platform to geostationary satellites. SEVIRI is selected here to provide measurements every 15 min, between $\pm 60^\circ$ N and $\pm 60^\circ$ E. The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) Land-Surface Analysis of the Satellite Application Facility (LSA SAF) distributes SEVIRI products at ~ 5 km resolution, including instantaneous T under clear-sky conditions [31]. When developing the downscaling function, these estimates will be the targeted fine spatial resolution T , while aggregation by simply averaging the ~ 5 km T estimates will be used as a synthetic coarse resolution T to test the downscaling.

To characterize the surface vegetated cover, the SEVIRI 10-day average of Fraction of Vegetation Cover (FVC) representing the fraction of green vegetation per unit of horizontal soil surface [32] is selected. LSA SAF also provides several parameters that describe the median T diurnal cycle, over the same 10-day period. The SEVIRI T estimates are used to derive these parameters, assuming that the diurnal T cycle follows the model described in [23]. This simple model summarizes the diurnal cycle with a harmonic and an exponential term and is parameterized with a few parameters among which T_0 , the

minimal T of the median diurnal cycle, and well as T_a , the median amplitude of the T diurnal cycle over the same period that are considered in this study.

The SEVIRI spatial coverage does not encompass extreme latitudes but the diversity of situations is deemed enough to develop and test the methodology. Estimates of T , FVC , T_0 , and T_a were extracted from the LSA SAF archive for 2016. One month per season (February/May/August/November) is downloaded to capture the seasonal variability.

2.2. ERA5

To study the changes in the downscaling function inputs under cloudy conditions a dataset providing ‘all-weather’ T is required. ERA5 skin temperature, Total Column Liquid Water ($TCLW$), and Leaf Area Index (LAI) are available over the globe every hour at a spatial resolution of ~ 25 km, and are selected here to study these changes. The resolution is lower than the one from SEVIRI, and ERA5 estimation of T does not always reproduce observed T [33]. However, the reanalysis T and cloud conditions are expected to be physically consistent and realistic enough for a first attempt to take the cloud information into account in the proposed methodology.

Similar months from 2016 are downloaded from the ECMWF archive to derive the cloud correction.

2.3. SSMIS

Once the downscaling methodology is developed using the SEVIRI and ERA5 datasets, it is applied to MW T estimates from SSMIS. The MW T estimates are based on inverting satellite observations from 19 to 90 GHz, for both perpendicular polarizations. The methodology is currently applied to the SSM/I and SSMIS time series (overpassing times around 06:00 and 18:00) as well as to AMSR data (overpassing time at 01:30 and 13:30), providing T at a spatial resolution of ~ 25 km [34]. When MW T is compared with in situ T time series over a large range of environments, good agreement is typically found in vegetated environments, but larger uncertainties in the T retrievals are noted under cold conditions due to the large variability of snow and ice surface emissivities, and over very dry surfaces due to microwave penetration problems in sandy terrain [16].

To illustrate the use of the downscaling method, one day (26 November 2016) of SSMIS T retrievals for a region in Southern Africa are downloaded from the LST-CCI archive.

3. Methods and Processing

3.1. Downscaling

3.1.1. Methodology

The statistical downscaling consists of finding an estimate of T at high-resolution $T_{HR}^{(i)}$, given T at low-resolution T_{LR} and ancillary information $A_{HR}^{(i)}$ provided at high resolution. T_{LR} can be written as the average of T_{HR} :

$$T_{LR} = \frac{1}{N} \sum_{i=1}^N T_{HR}^{(i)}. \quad (1)$$

In Figure 1, T_{LR} is downscaled into 25 high-resolution pixels T_{HR} ($N = 25$), for consistency with SSMIS and SEVIRI T products at respectively ~ 25 and ~ 5 km.

The downscaling relies on the hypothesis that a general statistical relationship F can link the ancillary information $A_{HR}^{(i)}$ and $T_{HR}^{(i)}$, to provide a high-resolution spatial pattern to be applied to low-resolution T_{LR} , i.e.,

$$\tilde{T}_{HR}^{(i)} = F(A_{HR}^{(i)}). \quad (2)$$

For each high-resolution pixel, a correction $\Delta \tilde{T}_{HR}^{(i)}$ will be applied to T_{LR} to describe the fine spatial structure:

$$\tilde{T}_{HR}^{(i)} = T_{LR} + \Delta \tilde{T}_{HR}^{(i)}. \quad (3)$$

It can be shown that this correction is a function of $A_{HR}^{(i)}$ by doing:

$$\begin{aligned}\Delta\tilde{T}_{HR}^{(i)} &= \tilde{T}_{HR}^{(i)} - T_{LR} = \tilde{T}_{HR}^{(i)} - \frac{1}{N} \sum_{j=1}^N \tilde{T}_{HR}^{(j)} \\ &= F(A_{HR}^{(i)}) - \frac{1}{N} \sum_{j=1}^N F(A_{HR}^{(j)}).\end{aligned}\quad (4)$$

This equation describes the increase or decrease induced by the T pattern at high resolution, computed with the average T on the pixel at low resolution as estimated by the F relationship.

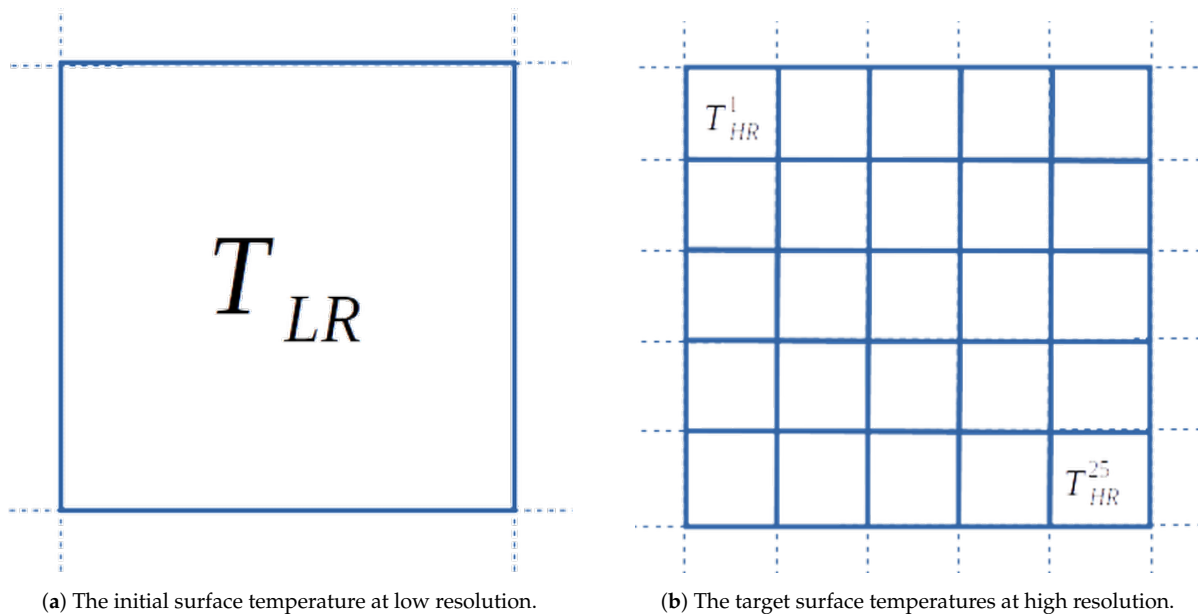


Figure 1. Schematic of the downscaling problem.

3.1.2. Input Selection

The LSA SAF variables FVC , T_0 , and T_a at ~ 5 km spatial resolution are used as the A_{HR} inputs to the F statistical function. Figure 2 shows the joint distribution of these variables to illustrate their relationships with T . Each variable is binned to represent the distribution of another variable under different conditions. The upper and lower quartiles (dotted lines) as well as the extrema for each distribution are shown. A positive correlation between T and T_0 is evident (upper-left panel), but similar trend is not visible between T and T_a (upper-right). Low T_0 is associated with lower T during the day but there is no clear correlation between T_a and T . This is expected considering that high temperatures are encountered in contrasting situations such as equatorial forest (small T_a) and desert regions (large T_a). This is corroborated by the negative correlation between T_a and FVC related to the lower amplitude of the diurnal cycle over vegetated areas, and the lack of correlation between T and FVC (bottom-left). The latter shows that vegetation alone cannot be the only source of information to discriminate warm and cold temperatures, confirming the need to have more predictors than just a vegetation-related product to derive the $\Delta\tilde{T}_{HR}^{(i)}$ correction.

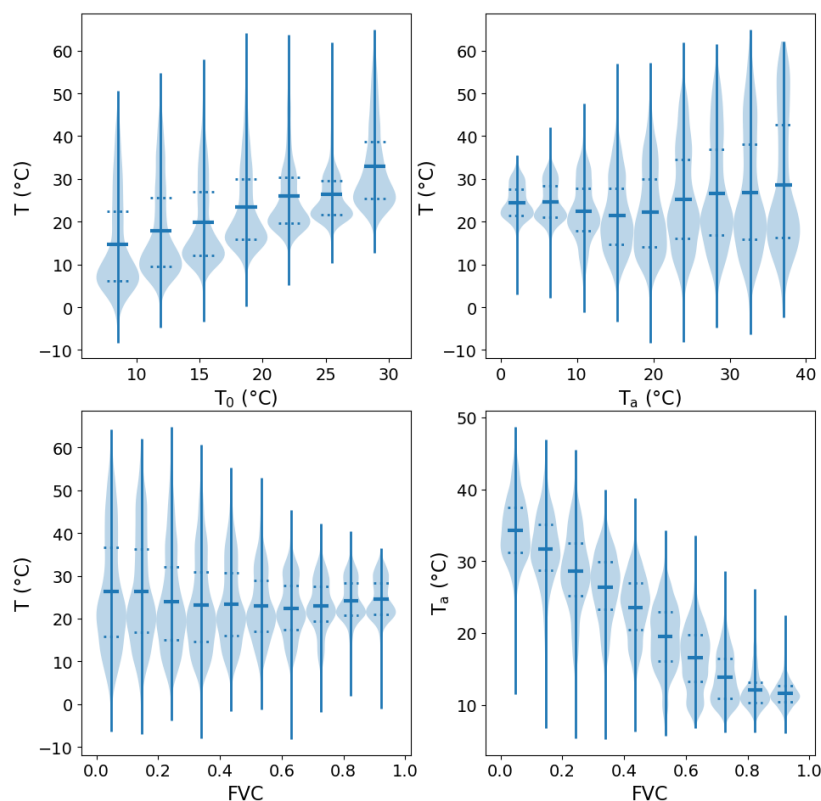


Figure 2. High-resolution surface-temperature T as a function of the high-resolution ancillary variables suggested as predictors for the downscaling. T versus the minimal temperature (T_0) (**top left**), versus the amplitude of the diurnal cycle (T_a) (**top right**), versus the fractional vegetation cover (FVC) (**bottom-left**). T_a versus FVC is also shown (**bottom-right**). For each panel the distribution of the y-axis variable for a bin of value for the x-axis variable is shown in the blue shade. Additionally, the mean (solid line), upper and lower quartiles (dotted lines) are indicated, as well as the extrema of the variable.

The spatial pattern of T is expected to change with local time. For instance, rivers and arid areas can show similar T at night but very different T at midday, with the water T being rather stable during the day, but the arid area undergoing large T changes. To account for changes in the T spatial pattern with local time, a variable related to the solar radiation is added as a fourth input to the F function: the Sun elevation angle over the horizon is calculated and its sinus ($\sin(\alpha_S)$) is used as a continuous predictor of the spatial T pattern, to account for the local time and the season.

3.1.3. Function Implementation

The statistical function F aim is to model the relationship between the 4 inputs and T to provide the high-resolution spatial pattern to be added to the low-resolution T_{LR} . The function provides an absolute estimate of the temperature that is used with the low-resolution T (Equation (4)) to provide the high-resolution temperature pattern for a given pixel. Here it is implemented by a feed-forward multilayer perceptron [35], with 4 input neurons, 1 hidden layer with 10 neurons (with sigmoid activation), and 1 output. A neural network architecture can represent any continuous function given enough internal parameters [36], so it is widely used to approximate this type of mappings [19]. Its inputs are the A_{HR} variables FVC , T_0 , T_a , and $\sin(\alpha_S)$, its output the T_{HR} , and it is trained with a database of ~ 50 M data points from the 4 months in 2016 extracted from the SEVIRI archive.

The SEVIRI disk is centered over the equator in Africa, and before training the original database distribution is resampled to avoid an over representation of desert situations (with very low vegetation) and a better sampling of solar elevation angles for which the low values are more common.

The residuals, i.e., the Root Mean Square Deviation (RMSD) between the targeted T_{HR} and the neural network output after training, are displayed for the range of T covered by the training database in Figure 3. To illustrate the choice of inputs, similar neural networks are trained using T_0 and T_a , or FVC only, as inputs and their residuals are displayed. Over the whole T range, the neural network using the four predictors displays a more even performance, especially for the very low and high T . It is interesting to note that FVC on its own does not seem to be enough to provide a clear distinction between temperatures, for all hours of the day. It is also noticeable that even for the four-input neural network, the function residuals appear to be high. However, this is expected, as this function aim is not to predict the exact high-resolution T from the four ancillary variables, but rather to inform about the relative local differences related to the information content of its inputs. The result of the F model are used in the downscaling procedure (Equation (4)) and yields much lower error as apparent in Section 4.1.

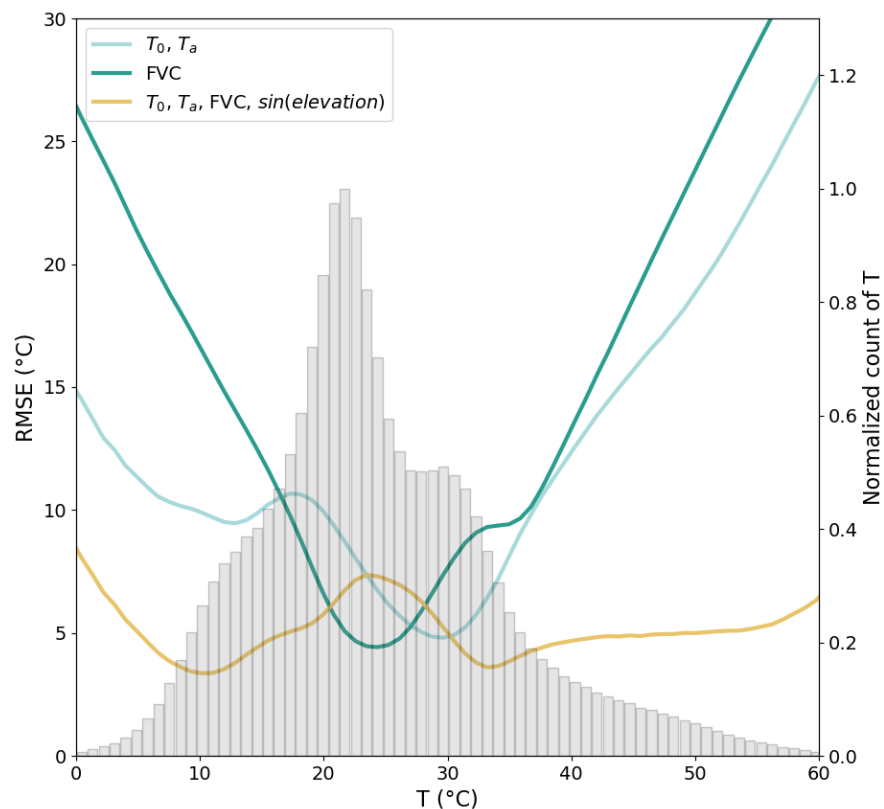


Figure 3. Root Mean Square Deviation (RMSD) of the T estimated with the statistical function F as a function of T , for different combinations of the predicting ancillary variables. In blue F is trained with T_0 and T_a , with only FVC in green, and finally with T_0 , T_a , FVC , and $\sin(\alpha_S)$ in brown. The distribution of T in the testing set is shown in gray bars (labels on the right axis).

3.2. Cloud Correction

3.2.1. Methodology

Exploring T changes due to clouds for a variety of cloud and surface conditions is challenging. Isolating the cloud effect would require an estimate of T for the same area,

under the same surface and insolation conditions, for different times of the day, and for a variety of cloud cover. Here the objective is only to develop a pragmatic solution that can be systematically and easily integrated into the downscaling processing. To achieve this, if the available cloud information indicates that the T_{LR} estimates correspond to a cloudy scene, a correction is applied to the A_{HR} ancillary data (that exists for clear-sky scenes) to account for the impact of the clouds on the $\tilde{T}_{HR}^{(i)}$ high-resolution patterns. The correction $\Delta A_{HR}^{(i)}$ is based on a simple function C parameterizing the cloud effect as function of the values observed for a selection of input variables $B_{HR}^{(i)}$, so it can be expressed as

$$\Delta \tilde{A}_{HR}^{(i)} = C(B_{HR}^{(i)}). \quad (5)$$

The F statistical relationship remains unchanged, so the new correction $\Delta \tilde{T}_{HR}^{(i)}$ for a cloudy scene is

$$\Delta \tilde{T}_{HR}^{(i)} = F(A_{HR}^{(i)} + C(B_{HR}^{(i)})) - \frac{1}{N} \sum_{j=1}^N F(A_{HR}^{(j)} + C(B_{HR}^{(j)})). \quad (6)$$

3.2.2. Input Selection

ERA5 $TCLW$ and LAI are selected as inputs to drive the parameterization of function C . Clouds limit the amount of solar radiation on the Earth surface during daytime, with this effect depending upon the cloud optical thickness. Clouds also limit the outgoing longwave flux with a magnitude depending on cloud base height and cloud amount [37]. At the surface level, the resulting T changes will essentially depend upon the type of surface, with an expected damping of the T spatial variability compared with clear-sky conditions. This is illustrated in Figure 4, where the impact of cloudiness for two surfaces with contrasting vegetation is displayed. The distribution of ΔT , i.e., the surface T difference between clear and thin ($TCLW < 50 \text{ g/m}^2$), or clear and thick ($TCLW > 50 \text{ g/m}^2$) clouds are presented over evergreen forests and savanna, selected using the International Geosphere Biosphere Program (IGBP) land cover [38], at different local times. The selected times are close to the overpass time of the SSMIS and AMSR sensors and include times of day close to the minimum and maximum of the T diurnal cycle. Nighttime (01:00 and 06:00) ΔT are mostly negative, which would indicate that cloud cover tends to increase the T , due to the trapping of the thermal radiation emitted from the surface. The 06:00 negative ΔT appears to be slightly larger than the 01:00 one, which is consistent with the declining surface temperatures until the sun starts illuminating the surface again. On the other hand, daytime T seems to decrease when a cloud is present, with an effect stronger near the middle of the day (13:00), when the T diurnal cycle has its maximum. Thicker clouds can also be associated with a stronger effect, either positive or negative on ΔT . The distributions for savanna (with sparse vegetation) and evergreen forest (with a dense vegetation) are shown in Figure 4 and this also seems to have an impact on ΔT : vegetation appears to reduce the magnitude of the effect on T .

Although additional effects can also affect the T distributions, Figure 4 distributions show that the variations of $TCLW$ and vegetation can be informative enough for a first order description of the cloud impact on T . To have continuous variation on vegetation content, LAI from ERA5 replaces the static IGBP classification.

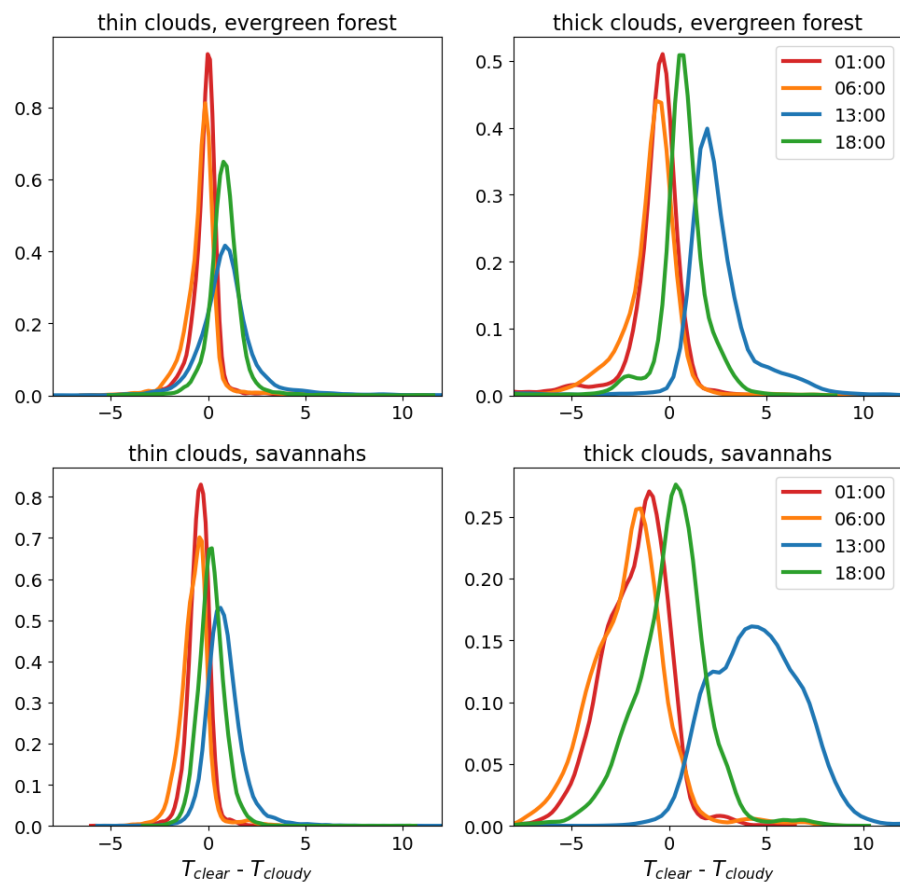


Figure 4. Distribution of ΔT , i.e., surface-temperature difference, from ERA5 data in January, May, and September 2016. The difference is computed between clear and thin clouds ($TCLW < 50 \text{ g/m}^2$) on the first columns, and between clear and thick clouds ($TCLW > 50 \text{ g/m}^2$) on the second row over evergreen forest and the second row over savannas. Each distribution is computed at different time of the day: 01:00, 06:00, 13:00, 18:00.

3.2.3. Function Implementation

Figure 4 shows that the effect of clouds on nighttime ΔT is negligible compared to the one on daytime ΔT . The effect is stronger near the peak of the diurnal cycle, when the warmest T is reached. Therefore, a correction on the maximum temperature of the day driven by the $TCLW$ and LAI variations, which translates into a correction of the T_a input to the F function, is proposed as an implementation for the function C .

The C function parameterization is based on the distributions shown in Figure 5. The observed difference between clear and cloudy T_a (computed from the ERA5 dataset) for different LAI (left panel) and $TCLW$ (right) values is displayed. For each bin the standard deviation around the mean of the computed ΔT_a is shown in the solid line. As with the observations from Figure 4, the thicker the clouds and the less dense the vegetation, the larger ΔT_a . A polynomial parameterization derived from the distributions is also displayed as dashed lines, and it is used to practically implement the function C . The 1σ uncertainty on the fitted parameters is given as a colored area around the polynomial fits. The uncertainty increases for larger $TCLW$ and smaller FVC , related to the smallest number of samples in the database for these extremes resulting in a more uncertain fit.

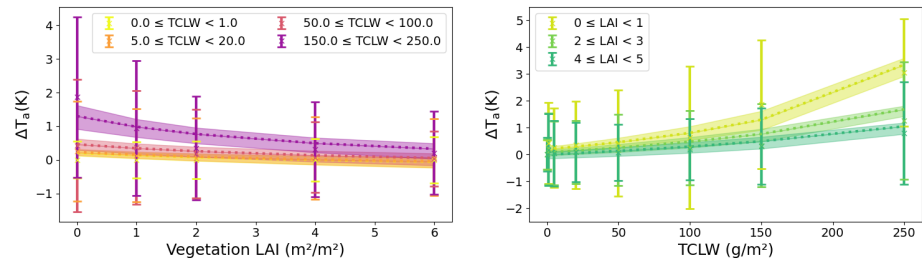


Figure 5. Average difference of T_a from ERA5 data from the month of January, May and September 2016 as solid lines (with 1 standard deviation) as a function of the leaf area index (LAI) (left) and total column liquid water (TCLW) (right). On each plot different colors are used for bins of TCLW (left) and LAI (right). The dotted lines show the predicted values of ΔT_a by the parameterization with the same input, and the shaded area shows the 1σ uncertainty on the fitted parameters.

In essence, the parameterization acts as a damping coefficient reducing the amplitude of the diurnal T cycle when clouds are present. It is rather simple, but is considered efficient enough for a first order attempt to include cloud information in the downscaling scheme.

3.3. Processing

The overall processing of the downscaling scheme is illustrated in Figure 6. In a first training phase, the statistical function F is trained on a representative ensemble of LSA SAF T_{HR} and corresponding A_{HR} inputs: FVC , T_0 , T_a , and $\sin(\alpha_S)$. Next to this, a simple correction function C is parameterized to damp the T_a under cloudy conditions using ERA5 $TCLW$ and LAI . Once these functions are implemented, they can be used to downscale the T_{LR} in the application phase. Here, for a selected scene with a given number of T_{LR} estimates, the corresponding A_{HR} inputs FVC , T_0 , and T_a are extracted, passed through the F function to derive the $\Delta\tilde{T}_{HR}$ high-resolution patterns, which are then combined with the original T_{LR} as described by Equation (3). If the scene contains T_{LR} estimates considered to be cloudy by ERA5, for those estimates the ERA5 $TCLW$ and LAI are extracted, passed through the C function to correct the T_a input to the F function, and the processing continues as before. Finally, previous experience with downscaling of geophysical variables showed that when applied separately to individual low-resolution estimates, this type of methods tends to produce spurious discontinuities at the edges of each low-resolution pixel [39]. Following that study, a smoothing procedure is applied considering that the temperature of a given HR pixel results from a contribution of neighboring low-resolution T . So, each LR pixel contribution is spatially averaged with a weight proportional to its distance to the considered high-resolution pixel. The partial contribution of neighboring LR pixels is relevant considering that the real measurements come from an antenna pattern with contributions larger than a single pixel.

The different parts of the methodology are evaluated in the following manner: the statistical model F is trained and tested on a different part of a large database of training samples as described in Section 3.1.3, the cloud correction is built using an independent set of data (Section 3.2). Then, the following section evaluates the performance of the methodology in an artificial clear-sky set-up only before illustrating the combined cloud correction and downscaling with a real scene.

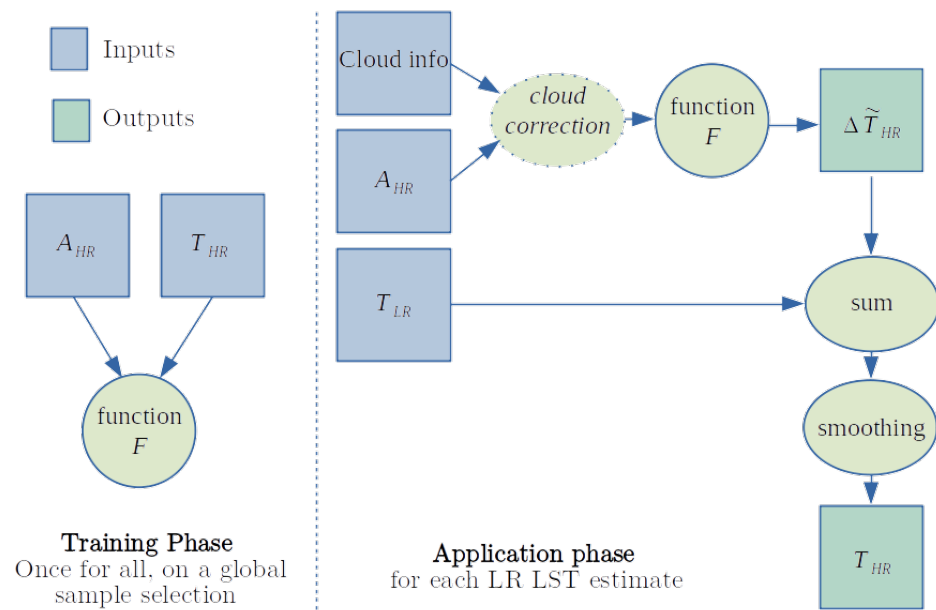


Figure 6. Schematic representation of the downscaling methodology, on the left for the training phase, on the right the application phase.

4. Results

The downscaling methodology is tested with two different low-resolution T . In a first test, synthetic low-resolution T estimates at ~ 25 km is produced by averaging the original SEVIRI IR T at ~ 5 km over 5×5 -pixel windows. This allows to evaluate the downscaling scheme independently of potential differences between the MW and IR temperatures, as it would be the case when comparing downscaled MW T with the corresponding high-resolution IR T . In addition, it makes it possible to test the methodology at different time of day as the full diurnal cycle is observed by the geostationary IR instrument (and not by a single MW polar satellite with only 2 overpass times). Illustrating scenes are shown and a global evaluation of the performance over Africa is computed. In a second test, SSMIS MW T is downscaled from its original ~ 25 km to the ~ 5 km of the IR ancillary information. Unlike with the previous IR clear-sky scene, the downscaling for cloudy conditions can now be illustrated as the MW T estimates are available for all weather conditions.

4.1. Downscaling of Synthetic IR Temperature

A SEVIRI T scene is selected in the south of Africa (16.7°S 21.0°E to 19.8°S 24.7°E), in the Okavango region that shows high surface heterogeneities. Sharp spatial structures are caused by the Okavango Delta with its water presence and vegetation variability also impacting the T patterns. Results are presented for two contrasted seasons and times of the day, in February (11 February 2016) at 05:00 (early morning) and in August (13 August 2016) at 12:00 (noon) respectively in Figures 7 and 8 illustrating the good performance independently of the time of day of the scene.

For both figures, the three left panels (A, D, G) show the values of the ancillary variables T_0 , T_a and FVC . Differences in the spatial structures of the three variables are visible, with an apparent opposite behavior between T_a and T_0 around the inner delta. This is related to the difference in heat capacity between these areas: colder minimal temperatures are associated with higher amplitude during the day. The difference in vegetation cover between the 2 selected scenes (at different seasons) is also strong.

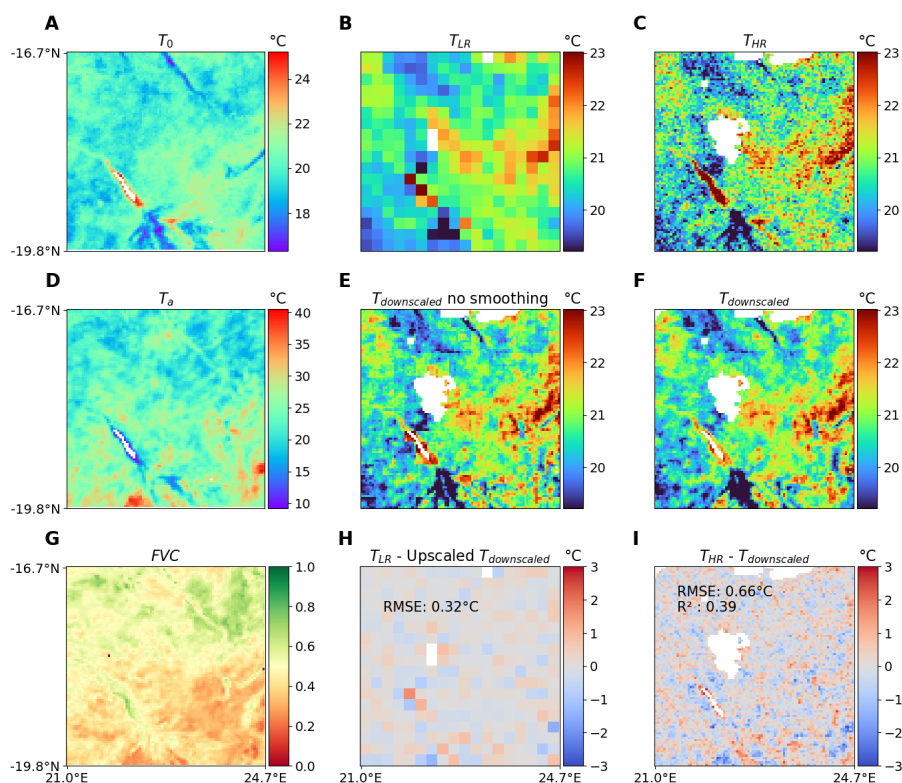


Figure 7. Different steps of the methodology applied to a scene from 11 February 2016 at 05:00 over the Okavango Delta (16.7°S 21.0°E to 19.8°S 24.7°E). Input ancillary variables (T_0 , T_a , FVC) are in the (A,D,G) panels. T at low resolution and original high resolution in the panels (B,C). The T obtained by the downscaling method, before and after smoothing in the second row (E, F). The (H) panel contains the difference between the input T and 5×5 average $T_{downscaled}$ at low resolution with the RMSD overlaid over the scene. Finally, the bottom-right panel (I) shows the difference between the original high resolution and the $T_{downscaled}$ produced by the methodology. The RMSD and the coefficient of determination (R^2) are overlaid on the scene.

Panel B shows the low-resolution aggregates (T_{LR}), with each pixel representing around $25 \times 25 \text{ km}^2$, whereas panel C displays the original T from SEVIRI (T_{HR}), i.e., the target T to be reproduced by the downscaling. The temperature range differs between the 2 selected times: around $20 \text{ }^\circ\text{C}$ in the 05:00 scene, and around $40 \text{ }^\circ\text{C}$ in the 12:00 scene. As the low-resolution values are synthetically constructed from the high-resolution ones, some are missing because of clouds contamination or invalid values (e.g., no T retrievals over water bodies). Panel E shows T after downscaling but before the smoothing step, with the final downscaled T ($T_{downscaled}$) after smoothing displayed in panel F. The smoothing effect is apparent from the absence of block-like artifacts on the final product. Compared to the original T_{HR} , $T_{downscaled}$ appears smoother with less contrast and some missing spatial structures. This can be explained by the method (averaging of contributions), F model errors, or missing data in the ancillary datasets used (e.g., around the water bodies for the February scene).

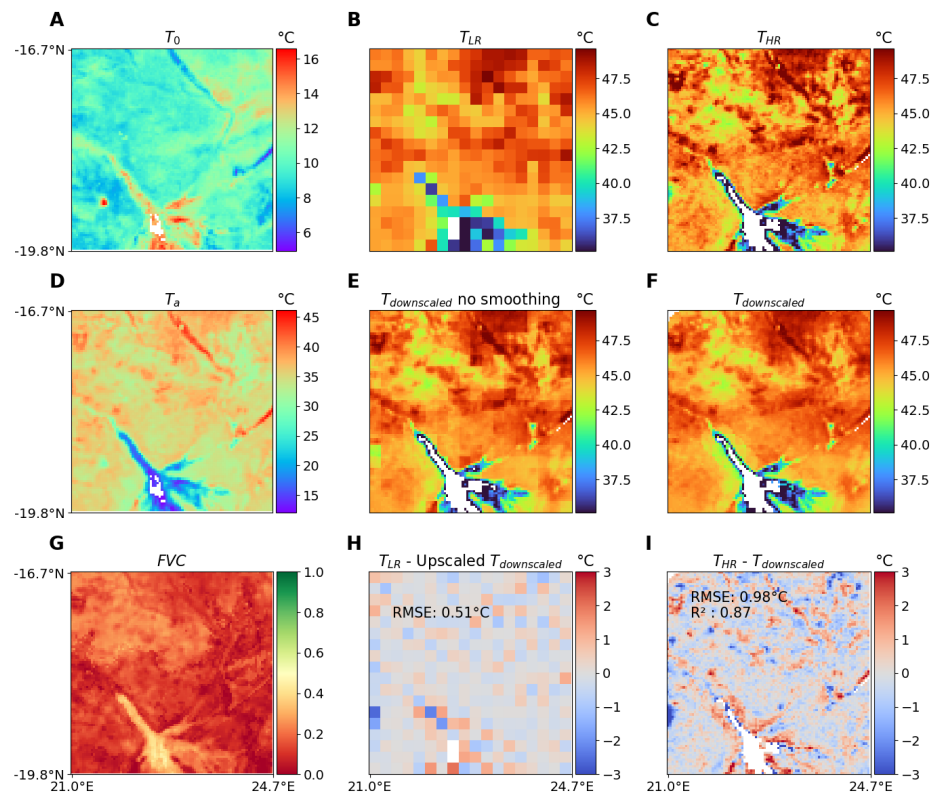


Figure 8. Same as Figure 7, but for 13 August 2016 at 12:00.

Finally, the bottom-row central and right panels (H, I) show the departure between $T_{downscaled}$ and original T_{HR} . In the central panel, $T_{downscaled}$ is spatially averaged to get a low-resolution product that can be compared to the synthetic T_{LR} used as input to the downscaling scheme. This shows where the downscaling procedure degrades the input temperature information. The only areas with large departures are those around places with missing data or sharp contrast between nearby pixels. The right panel displays the difference between the original T_{HR} and $T_{downscaled}$. The dark blue (red) spots are mostly around local extreme T that are regularly over(under)-estimated. This lower performance on extreme values is inherent to the neural network training scheme used to implement the F function. This scheme minimizes the squared deviation error, which leads to an underestimation of variance in the output that can sometimes be addressed with variance inflation methods [40]. Some strong errors also appear near missing values, for instance close to the water extent in the delta.

The downscaling method is further illustrated in Figure 9 where a transect along the latitude of the estimated $T_{downscaled}$, the targeted original T_{HR} , and the spatially averaged input to be downscaled T_{LR} are displayed. The transect integrates spatially between 12.6°N and 12.7°N, extending from 20°E to 22°E for the 18th of February/May/August/November in 2016 at 10:00. The variability shown in T_{HR} is clearly damped on T_{LR} as expected, while it is well captured in $T_{downscaled}$, with relatively large correlation values (coefficients of determination R^2 between 0.65 and 0.88). The applied downscaling performs equally well for a variety of temperatures between 30 °C and 60 °C. The already mentioned underestimation of extrema is also visible on the transect.

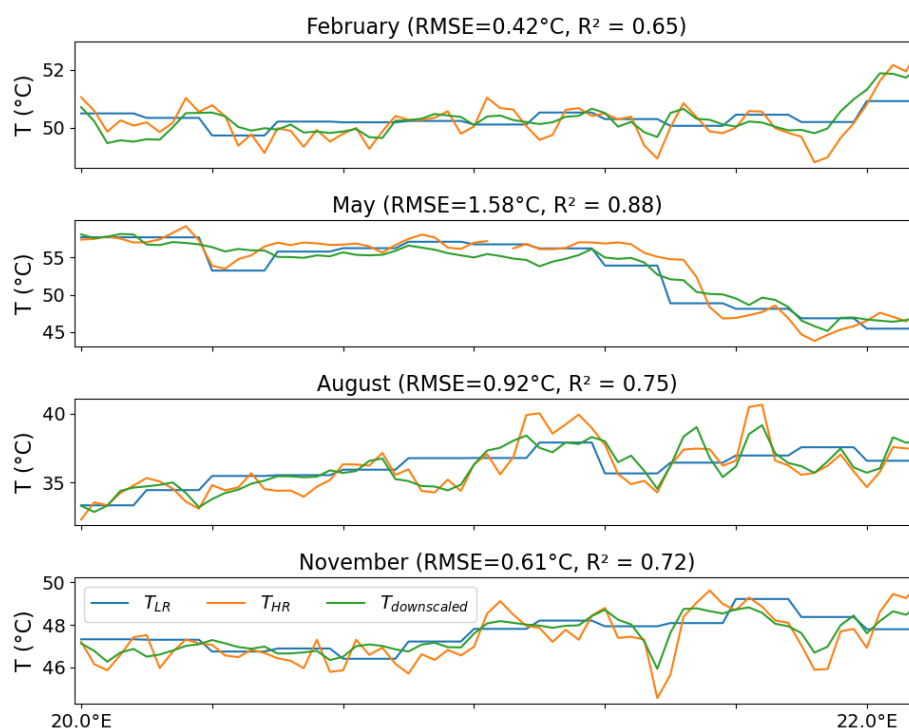


Figure 9. Transect of the T around 12.5°N between 20°E and 22°E at 10:00 on the 18th of the months of February, May, August and November (top to bottom). For each month, the SEVIRI T_{HR} in orange, the 5×5 average T_{LR} in blue and the $T_{downscaled}$ in green are shown. Also, the RMSD and coefficient of determination R^2 between the $T_{downscaled}$ and T_{HR} is displayed in each subplot title.

The previous examples show that the F function correctly reproduces the T variations at the fine scale, although it performs less well for the more extreme T values within the scenes. Overall, the downscaling produces a good approximation of the high-resolution temperatures. It also adapts well to different seasons, times of the day, and surface conditions. To quantify the performance, the Root Mean Square Deviation (RMSD) between T_{HR} and $T_{downscaled}$ was computed for the 18th of the February/May/August and November months in 2016 at all hours between 35°S – 35°N and 22 – 35°E . The RMSD values range between 0.62°C in May and 1.07°C in November. The RMSD but calculated between T_{HR} and the spatially averaged T_{HR} lies between 0.74°C and 1.16°C , confirming the effectiveness of the downscaling method in producing a T estimate closer to the true T .

4.2. Downscaling of Observed SSMIS Temperature

The downscaling methodology is now applied to SSMIS MW T from the same area in the Okavango Delta. The results are presented in Figure 10. 26 November 2016 at 05:00 is selected as it presents both clear (the southern part of the scene) and cloudy conditions (the northern part, with no IR T corresponds to the cloud presence). The two panels on the left (A, D) show the ancillary variables T_a and FVC . Their values are different from the ones in Figures 7 and 8, reflecting the seasonal changes of these variables. Please note that even if the northern part of the scene is cloudy, IR-derived T_a and FVC estimates still exist due to their 10-day temporal integration. Consequently, the downscaling can be applied everywhere in the scene. Panel B shows the SSMIS T_{LR} , i.e., the low-resolution T input to be downscaled. In panel C, the cloud correction, aggregation, and smoothing steps are applied to produce $T_{downscaled}$. Panel E shows the 5×5 averaged T_{HR} from SEVIRI. Finally, the original SEVIRI T_{HR} is shown in panel F.

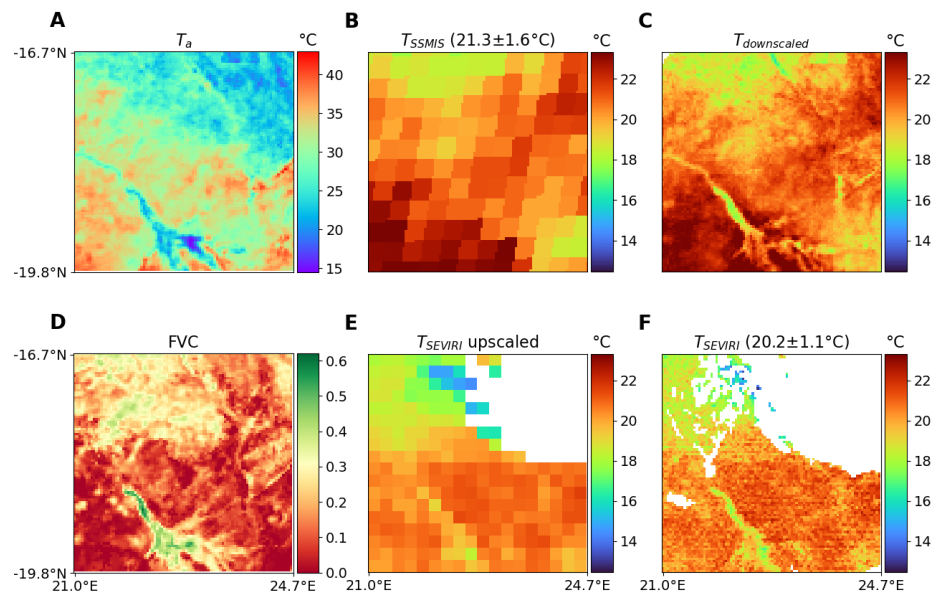


Figure 10. For a scene from the 26th of November 2016 at 05:00 over the Okavango Delta (16.7°S 21.0°E to 19.8°S 24.7°E): the first column shows 2 of the input ancillary variables: T_a (A) and FVC (D). Panel (B) shows the original SSMIS (T_{LR}) and panel (C) the $T_{downscaled}$, finally panel (F) shows the T_{SEVIRI} at high resolution and panel (E) the IR T after a 5×5 spatial average. The average temperature and standard deviation for the clear-sky part of the scene are indicated for T_{SSMIS} and T_{SEVIRI} .

Comparison of original SSMIS T_{LR} , $T_{downscaled}$, and SEVIRI T_{HR} shows that the SEVIRI T_{HR} spatial structures are well captured by $T_{downscaled}$ estimates. In the clear-sky part of the scene, $T_{downscaled}$ correctly reproduces the north-to-south T gradient and the colder T in the riverbed, as observed on SEVIRI T_{HR} . The average value and standard deviation of the scene clear pixels are given for the SSMIS T_{LR} and SEVIRI T_{HR} in Figure 10, showing a warm bias of ~ 1 °C for SSMIS. This bias is likely related to an initial difference between the MW and IR T in this area, not to an artifact of the downscaling. For the clear-sky part of the scene, RMSD computed between the SEVIRI T_{HR} and SSMIS T gives values of 1.72 °C with the original resolution that is improved to 1.60 °C after the downscaling.

The downscaling for the cloudy part of the scene is reasonable, without any obvious artifacts compared with the clear part. The cloud correction on the T_a variable was of the order of 3 °C for this particular scene, but its impact on $T_{downscaled}$ was not very large compared with a clear situation as the effect of T_a on the F function is lower around the early morning hours than around midday. Comparing $T_{downscaled}$ with SEVIRI T_{HR} close to the cloud edges reveals some inconsistencies. Some SEVIRI T estimates appear much colder (~ 10 °C) than the surrounding average T of the scene (~ 20 °C) and than the corresponding SSMIS $T_{downscaled}$ estimates, suggesting a likely cloud contamination in the IR T estimates.

To better highlight the correction provided by the parameterization of the cloud effect Figure 11 shows the downscaling scheme applied to 2 consecutive days over a scene in northern Tanzania (2.4°S 32°E to 4.9°S 35.4°E). For each day, in 2 columns, the first row (panels A, B) shows the original SSMIS T and indicates the average temperature and standard deviation over the clear-sky parts of the scene. Then the second row (C, D) shows the result after the downscaling procedure is applied. T_{SEVIRI} is displayed in the third row (E, F) with the average temperature and standard deviation over the scene. Finally, the last row (G, H) shows the correction on the T_a variable that is applied based on the cloud parameterization and ERA5 data, for the cloud structure from the south-west to the north-east of the image.

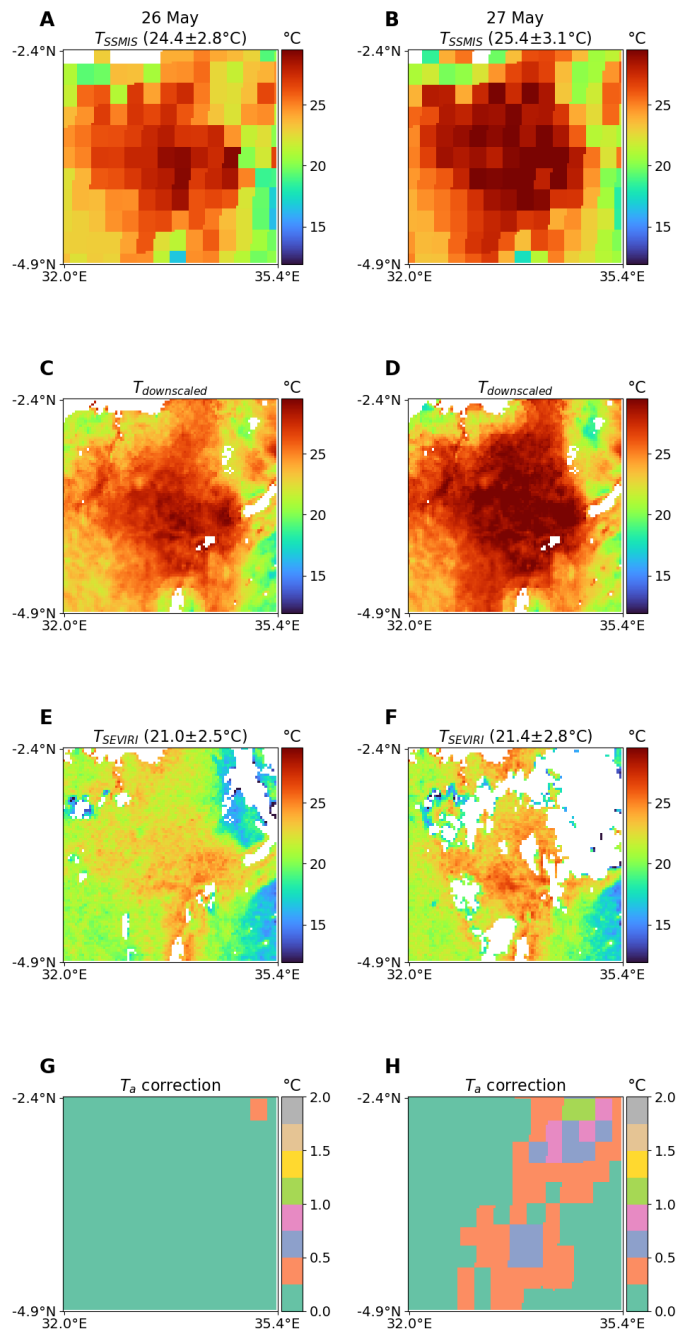


Figure 11. Scenes from the afternoon overpass during 2 consecutive days, 26 and 27th of May 2016 over northern Tanzania (2.4°S 32°E to 4.9°S 35.4°E) : On the first line (A,B) is the SSMIS observation at low resolution, then panels (C,D) show the result after the downscaling scheme is applied, (E,F) show the T_{SEVIRI} at the same time. Finally, the last row (G,H) shows the correction on the T_a parameter derived from the ERA5 reanalysis. The average and standard deviation of T_{SSMIS} and T_{SEVIRI} for the clear-sky parts of the scene are indicated in the sub-panel titles.

Most of the cloud presence is visible in the north-east of the scene for the second day. Over the clear parts of the scene the second day is warmer than the first day, both for SSMIS and SEVIRI. The SSMIS T_{LR} clear-sky part of the scenes are ~ 3 K warmer than the SEVIRI T_{HR} ones. This bias is larger than the one found in Figure 10 (~ 1 K), and can be partly caused by a larger cloud contamination in T_{SEVIRI} due to the challenges of filtering a more broken cloud. Over a large part of the cloudy area in the second day T_{SSMIS} is a bit colder, compared to the first day, possibly reflecting the impact of the cloud presence on the surface T . The illustrated T_a correction, at the same resolution as ERA5 it is derived from, in the cloudy part of the scene reduces the T_a of the day, with the largest correction not exceeding 2.0 °C. The $T_{downscaled}$ for both clear-sky and cloudy scenes show comparable structures, with very realistic high-resolution patterns, in agreement with the geophysical structure in this area such as the delta leading to Lake Victoria in the north-west, or the crater highland in the north-east with lower average temperatures.

5. Discussion

The previous section illustrates the performance of the methodology for the downscaling of synthetic data but also the application of the procedure to real world land-surface temperature derived from SSMIS observations. Several points arise from these results that should be taken into account when implementing such a downscaling scheme.

First, the methodology developed here is independent of the overpass time of the platform carrying the MW instruments used to derive the T . The different overpass times of each instrument (06:00 for SSMIS and 13:30 for AMSR-E for instance) are handled by the solar zenith angle factor dependence introduced in the model. The same model can therefore be applied to a low-resolution land-surface temperature product at any time of the day. This is a great benefit toward a coherent high-resolution MW-derived T product given the changing overpass times depending on the satellite missions. However, the performance of this method relies a lot on the performance of the underlying MW T , for which evaluation have been conducted for instance in [16,17,34].

Also, the ancillary variables used in the training of the F model, especially the diurnal cycle derived parameter T_a and T_0 , offer new information sources to describe the high-resolution T patterns with a finer accuracy. This performance benefit is shown in Figure 3, and is understandable as these variables inherently summarize the thermal behavior of different surfaces. The 10-day period used in this study is a conveniently available timescale but these values could be replaced with longer time scale climatologies. For future developments outside the regions covered by geostationary satellites, MODIS T and vegetation indexes that are globally available could be used. MODIS being on board a polar orbiter with overpassing times at 01:30 and 13:30, it will not provide information over the full T diurnal cycle, but T at 01:30 is close to the minimum T and its difference with T at 13:30 can be a proxy to the amplitude of the T diurnal cycle.

The known difficulties in comparing IR and MW T over some surfaces are mitigated in this study as only the relative increase in temperatures from IR is used to infer the correction to the MW T . The training of the F model is independent from the sensor on which the methodology is applied. Of course, known limitations of the MW (over arid areas for instance) are not corrected by this methodology, but it can nevertheless improve the MW T resolution in most situations.

Finally, the correction for ancillary values under clouds developed here is not only a major building block of this methodology but is also a valuable contribution to the understanding of the expected impact of clouds on surface T . As noted in the work of [26], the assumption that clear-sky derived coefficient can also be used for cloudy scenes should be further investigated. The parameterization of the cloud effect proposed here is a first attempt to better understand what remains a high source of uncertainty in global T measurements.

6. Conclusions

To provide high-resolution ‘all-weather’ land-surface temperature, this study develops a methodology to perform a downscaling of MW-derived T , from their original spatial resolution of ~ 25 km to a resolution of ~ 5 km, similar to IR geostationary observations.

To that end, a statistical relationship is first established between high-resolution ancillary information and high-resolution T , to estimate the T fine spatial resolution patterns under clear-sky conditions. The model is trained using IR T and ancillary products from IR SEVIRI geostationary observations at ~ 5 km. Different combinations of variables are tested and the model error is lowest over the whole range of temperature, using the minimal value and the amplitude of the median clear-sky temperature diurnal cycle (T_0 and T_a), along with the fractional vegetation cover FVC , and the solar elevation angle ($\sin(\alpha_S)$). In an application phase, the statistical function is then used to produce a high-resolution temperature pattern that is added to the coarse resolution T and provides spatially finer T estimates.

To test the methodology under clear-sky conditions, the original ~ 5 km T estimates are used as the fine resolution T , with the coarse resolution T synthetically produced from aggregation of the original T to a ~ 25 km resolution. The RMSD of the downscaled T against the high-resolution one remains below 1 °C for a variety of conditions (spanning multiple months and all hours of day), showing the possibility to correctly reproduce the original T with this method. For cloudy scenes, the same statistical function is used, but with a prior reduction of the T_a input to account for the cloudy conditions. This correction is based on an analysis of the average impact of clouds on T using cloud information from ECMWF ERA5 reanalysis. Finally, the downscaling and cloud correction performance are tested on an original SSMIS-derived land-surface temperature scene at ~ 25 km resolution, and the resulting downscaled scene at ~ 5 km is compared to the nearly coincident SEVIRI-derived estimates. The RMSD over the clear-sky part of the scene and the original low-resolution is 1.72 °C and is reduced to 1.60 °C after downscaling. The downscaling could not be tested under cloudy conditions as no high-resolution satellite product is yet available for cloudy scenes.

The results demonstrate the relevance of the methodology to perform the downscaling of MW land-surface temperature, with an adaptability to seasons and times of the day, along with a specific handling of the cloudy areas. The required inputs are limited, keeping the implementation of the method rather simple. The method can already be used for all areas for which the geostationary satellites data is available, including the often-cloudy tropical areas. Additional study could be performed to evaluate the possibility to generalize the predictors for a global downscaling, for instance using polar orbiting high-resolution IR sensors. These data sources would also provide an opportunity to downscale the MW T to a higher resolution, up to ~ 1 km.

This is the first step toward the application of a generic methodology to downscale the ~ 40 years of MW-derived land-surface temperature currently under production [29]. It will provide an ‘all-weather’ climatology of land-surface temperature at global scale with a high spatial resolution, to complement the clear-sky only IR estimates.

Author Contributions: Data collection and preliminary work, conception and data analysis have been conducted by S.F. C.P., C.J. and S.F. wrote and edited the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Space Agency under the Climate Change Initiative, Essential Climate Variables Land-Surface Temperature project with contract n° 4000123553/18/I-NB. Funding was also granted by the Agence Nationale Recherche Technologie, with the CIFRE n° 2018/0637.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in this study is freely available from the sources mentioned in the Data section.

Acknowledgments: The authors are grateful to Sofia Ermida from Instituto Portugues do Mar e da Atmosfera, Lisbon, for the information on the cloud effect on temperatures, as well as Filipe Aires from Observatoire de Paris, for the discussions regarding the formalism in the downscaling methodology. The various reviewers contributions was also very helpful in crafting the final manuscript version.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bojinski, S.; Verstraete, M.; Peterson, T.C.; Richter, C.; Simmons, A.; Zemp, M. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 1431–1443. [\[CrossRef\]](#)
- Bechtel, B. A new global climatology of annual land surface temperature. *Remote Sens.* **2015**, *7*, 2850–2870. [\[CrossRef\]](#)
- Wan, Z.; Dozier, J. Land-Surface Temperature Measurement from Space: Physical Principles and Inverse Modeling. *IEEE Trans. Geosci. Remote Sens.* **1989**, *27*, 268–278. [\[CrossRef\]](#)
- Trigo, I.F.; Peres, L.F.; DaCamara, C.C.; Freitas, S.C. Thermal land surface emissivity retrieved from SEVIRI/Meteosat. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 307–315. [\[CrossRef\]](#)
- Wan, Z.; Li, Z.L. A physics-based algorithm for retrieving land-surface emissivity and temperature from eos/modis data. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 980–996. [\[CrossRef\]](#)
- Ermida, S.L.; Trigo, I.F.; DaCamara, C.C.; Jiménez, C.; Prigent, C. Quantifying the Clear-Sky Bias of Satellite Land Surface Temperature Using Microwave-Based Estimates. *J. Geophys. Res. Atmos.* **2019**, *124*, 844–857. [\[CrossRef\]](#)
- Jin, M. Interpolation of surface radiative temperature measured from polar orbiting satellites to a diurnal cycle: 2. Cloudy-pixel treatment. *J. Geophys. Res. Atmos.* **2000**, *105*, 4061–4076. [\[CrossRef\]](#)
- Martins, J.P.; Trigo, I.F.; Ghilain, N.; Jimenez, C.; Göttsche, F.M.; Ermida, S.L.; Olesen, F.S.; Gellens-Meulenberghs, F.; Arboleda, A. An all-weather land surface temperature product based on MSG/SEVIRI observations. *Remote Sens.* **2019**, *11*, 3044. [\[CrossRef\]](#)
- Wu, P.; Yin, Z.; Zeng, C.; Duan, S.B.; Gottsche, F.M.; Li, X.; Ma, X.; Yang, H.; Shen, H. Spatially Continuous and High-Resolution Land Surface Temperature Product Generation: A Review of Reconstruction and Spatiotemporal Fusion Techniques. *IEEE Geosci. Remote Sens. Mag.* **2021**. [\[CrossRef\]](#)
- Xu, S.; Cheng, J. A new land surface temperature fusion strategy based on cumulative distribution function matching and multiresolution Kalman filtering. *Remote Sens. Environ.* **2021**, *254*, 112256. [\[CrossRef\]](#)
- Wu, P.; Yin, Z.; Yang, H.; Wu, Y.; Ma, X. Reconstructing geostationary satellite land surface temperature imagery based on a multiscale feature connected convolutional neural network. *Remote Sens.* **2019**, *11*, 300. [\[CrossRef\]](#)
- McFarland, M.; Miller, R.; Neale, C. Land surface temperature derived from the SSM/I passive microwave brightness temperatures. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 839–845. [\[CrossRef\]](#)
- Weng, F.; Grody, N.C. Physical retrieval of land surface temperature using the special sensor microwave imager. *J. Geophys. Res. Atmos.* **1998**, *103*, 8839–8848. [\[CrossRef\]](#)
- Aires, F.; Prigent, C.; Rossow, W.B.; Rothstein, M. 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. Atmos.* **2001**, *106*, 14887–14907. [\[CrossRef\]](#)
- Holmes, T.R.H.; De Jeu, R.A.M.; Owe, M.; Dolman, A.J. Land surface temperature from Ka band (37 GHz) passive microwave observations. *J. Geophys. Res. Atmos.* **2009**, *114*, 1–15. [\[CrossRef\]](#)
- Jiménez, C.; Prigent, C.; Ermida, S.L.; Moncet, J.L. Inversion of AMSR-E observations for land surface temperature estimation: 1. Methodology and evaluation with station temperature. *J. Geophys. Res.* **2017**, *122*, 3330–3347. [\[CrossRef\]](#)
- Ermida, S.L.; Jiménez, C.; Prigent, C.; Trigo, I.F.; DaCamara, C.C. Inversion of AMSR-E observations for land surface temperature estimation: 2. Global comparison with infrared satellite temperature. *J. Geophys. Res.* **2017**, *122*, 3348–3360. [\[CrossRef\]](#)
- Kustas, W.P.; Norman, J.M.; Anderson, M.C.; French, A.N. Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sens. Environ.* **2003**, *85*, 429–440. [\[CrossRef\]](#)
- Zhan, W.; Chen, Y.; Zhou, J.; Wang, J.; Liu, W.; Voogt, J.; Zhu, X.; Quan, J.; Li, J. Disaggregation of remotely sensed land surface temperature: Literature survey, taxonomy, issues, and caveats. *Remote Sens. Environ.* **2013**, *131*, 119–139. [\[CrossRef\]](#)
- Merlin, O.; Walker, J.P.; Chehbouni, A.; Kerr, Y. Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sens. Environ.* **2008**, *112*, 3935–3946. [\[CrossRef\]](#)
- Peng, J.; Loew, A.; Merlin, O.; Verhoest, N.E. A review of spatial downscaling of satellite remotely sensed soil moisture. *Rev. Geophys.* **2017**, *55*, 341–366. [\[CrossRef\]](#)
- Aires, F.; Prigent, C.; Rossow, W.B. Temporal interpolation of global surface skin temperature diurnal cycle overland under clear and cloudy conditions. *J. Geophys. Res. Atmos.* **2003**, *109*. [\[CrossRef\]](#)
- Göttsche, F.M.; Olesen, F.S. Modelling the effect of optical thickness on diurnal cycles of land surface temperature. *Remote Sens. Environ.* **2009**, *113*, 2306–2316. [\[CrossRef\]](#)
- Duan, S.B.; Li, Z.L.; Leng, P. A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data. *Remote Sens. Environ.* **2017**, *195*, 107–117. [\[CrossRef\]](#)

25. Sun, D.; Li, Y.; Zhan, X.; Houser, P.; Yang, C.; Chiu, L.; Yang, R. Land surface temperature derivation under all sky conditions through integrating AMSR-E/AMSR-2 and MODIS/GOES observations. *Remote Sens.* **2019**, *11*, 1–20. [[CrossRef](#)]
26. Zhang, X.; Zhou, J.; Gottsche, F.M.; Zhan, W.; Liu, S.; Cao, R. A Method Based on Temporal Component Decomposition for Estimating 1-km All-Weather Land Surface Temperature by Merging Satellite Thermal Infrared and Passive Microwave Observations. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4670–4691. [[CrossRef](#)]
27. Schmid, J. The SEVIRI instrument. In Proceedings of the 2000 EUMETSAT Meteorological Satellite Data User’s Conference, Bologna, Italy, 29 May–2 June 2000.
28. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
29. Consortium CCI LST. *Algorithm Theoretical Basis Document WP2.1 – DEL-LST-CCI-D2.2-ATBD*; 2019. Available online: <https://climate.esa.int/sites/default/files/LST-CCI-D2.2-ATBD%20-%20i1r1%20-%20Algorithm%20Theoretical%20Basis%20Document.pdf> (accessed on 31 March 2021).
30. Perry, M.; Ghent, D.; Jimenez, C.; Dodd, E.; Ermida, S.L.; Isabel, F. Multi-Sensor thermal infrared and microwave land surface temperature algorithm intercomparison. *Remote Sens.* **2020**, *12*, 4164. [[CrossRef](#)]
31. Trigo, I.F.; Dacamara, C.C.; Viterbo, P.; Roujean, J.L.; Olesen, F.; Barroso, C.; Camacho-De-Coca, F.; Carrer, D.; Freitas, S.C.; García-Haro, J.; et al. The satellite application facility for land surface analysis. *Int. J. Remote Sens.* **2011**, *32*, 2725–2744. [[CrossRef](#)]
32. García-Haro, F.; Camacho, F.; Meliá, J. The EUMETSAT Satellite Application Facility on Land Surface Analysis Product User Manual Vegetation Parameters (VEGA). *EUMETSAT Prod. User Man.* **2013**, *401*, 1–46.
33. Johannsen, F.; Ermida, S.; Martins, J.P.; Trigo, I.F.; Nogueira, M.; Dutra, E. Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian Peninsula. *Remote Sens.* **2019**, *11*, 2570. [[CrossRef](#)]
34. Prigent, C.; Jimenez, C.; Aires, F. Toward “all weather,” long record, and real-time land surface temperature retrievals from microwave satellite observations. *J. Geophys. Res.* **2016**, *121*, 5699–5717. [[CrossRef](#)]
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
36. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [[CrossRef](#)]
37. Chen, T.; Rossow, W.B.; Zhang, Y. Radiative effects of cloud-type variations. *J. Clim.* **2000**, *13*, 264–286. [[CrossRef](#)]
38. Friedl, M.; Strahler, A.; Hodges, J. *ISLSCP II MODIS (Collection 4) IGBP Land Cover, 2000–2001*; ORNL Distributed Active Archive Center: Oak Ridge, TN, USA, 2010. [[CrossRef](#)]
39. Aires, F.; Miolane, L.; Prigent, C.; Pham, B.; Fluet-Chouinard, E.; Lehner, B.; Papa, F. A global dynamic long-term inundation extent dataset at high spatial resolution derived through downscaling of satellite observations. *J. Hydrometeorol.* **2017**, *18*, 1305–1325. [[CrossRef](#)]
40. Von Storch, H. On the use of ‘inflation’ in statistical downscaling. *J. Clim.* **1999**, *12*, 3505–3506. [[CrossRef](#)]

Bibliographie

- Abrams, M., Crippen, R., and Fujisada, H. (2020). ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). *Remote Sensing*, 12(7) :1–12.
- Aires, F., Marquisseau, F., Prigent, C., and Sèze, G. (2011a). A Land and Ocean Microwave Cloud Classification Algorithm Derived from AMSU-A and -B, Trained Using MSG-SEVIRI Infrared and Visible Observations. *Monthly Weather Review*, 139(8) :2347–2366.
- Aires, F., Miolane, L., Prigent, C., Pham, B., Fluet-Chouinard, E., Lehner, B., and Papa, F. (2017). A global dynamic long-term inundation extent dataset at high spatial resolution derived through downscaling of satellite observations. *Journal of Hydrometeorology*, 18(5) :1305–1325.
- Aires, F., Paul, M., Prigent, C., Rommen, B., and Bouvet, M. (2011b). Measure and exploitation of multisensor and multiwavelength synergy for remote sensing : 2. application to the retrieval of atmospheric temperature and water vapor from metop. *Journal of Geophysical Research : Atmospheres*, 116(D2).
- Aires, F., Prigent, C., Bernardo, F., Jiménez, C., Saunders, R., and Brunel, P. (2011c). A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 137(656) :690–699.
- Aires, F., Prigent, C., Rossow, W. B., and Rothstein, M. (2001). 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. *Journal of Geophysical Research Atmospheres*, 106(D14) :14887–14907.
- Allison, L. J. (1977). Geological applications of nimbus radiation data in the middle east. *NASA technical note*, (April 1977).
- Amari, S. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5) :185–196.
- Anderson, M. C., Kustas, W. P., and Norman, J. M. (2003). Soil plant atmosphere continuum. pages 1408–1423.
- Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., and Schulz, J. (2010). The Hamburg ocean atmosphere parameters and fluxes from satellite data-HOAPS-3. *Earth System Science Data*, 2(2) :215–234.
- Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., and Van Bogaert, E. (2012). Global Land Cover Map for 2009 (GlobCover 2009).

- Atkinson, N. (2015). NEDT specification and monitoring for microwave sounders. *NWP SAF*, (December).
- Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22(1) :106–114.
- Balsamo, G., Dutra, E., Beljaars, A., and Viterbo, P. (2011). Evolution of land-surface processes in the IFS. *ECMWF Newsletter*, 127(January) :6.
- Bauer, P., Moreau, E., Chevallier, F., and O’Keeffe, U. (2006). Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society*, 132(617) :1259–1281.
- Beaudoin, L., Nicolas, J.-M., Tupin, F., and Hueckel, M. (2001). Introducing spatial information in k-means algorithm for cloud detection in optical satellite images. (33) :67–77.
- Bechtel, B. (2015). A new global climatology of annual land surface temperature. *Remote Sensing*, 7(3) :2850–2870.
- Becker, F. and Li, Z. (1995). Surface temperature and emissivity at various scales : definition, measurement and related problems. *Remote Sensing Reviews*, 12(3-4) :225–253.
- Beersma, J. J., van Dorland, R., and Opsteegh, J. D. (2002). *Shortwave radiation and cloud parameterizations for intermediate complexity models*. KNMI.
- Berg, W., Kroodsmas, R., Kummerow, C., and McKague, D. (2018). Fundamental Climate Data Records of Microwave Brightness Temperatures. *Remote Sensing*, 10(8) :1306.
- Berg, W., Sapiano, M. R., Horsman, J., and Kummerow, C. (2013). Improved geolocation and earth incidence angle information for a fundamental climate data record of the SSM/I sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3) :1504–1513.
- Blinn, J. C., Conel, J. E., and Quade, J. G. (1972). Microwave emission from geological materials : Observations of interference effects. *Journal of Geophysical Research*, 77(23) :4366–4378.
- Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society*, 95(9) :1431–1443.
- Bonnefoy, L. (2020). *Emission thermique micro-ondes des lunes glacées de Saturne Thermal micro-wave emission from Saturn ’ s icy moons Léa Bonnefoy*.
- Boone, A. A., Calvet, J.-C., and Noilhan, J. (1999). Inclusion of a Third Soil Layer in a Land Surface Scheme Using the Force–Restore Method. *Journal of Applied Meteorology*, 38(11) :1611–1630.
- Bottou, L., Cortes, C., Denker, J., Drucker, H., Guyon, I., Jackel, L., LeCun, Y., Muller, U., Sackinger, E., Simard, P., and Vapnik, V. (2002). Comparison of classifier methods : a case study in handwritten digit recognition. pages 77–82.
- Brandt, M., Wigner, J. P., Chave, J., Tagesson, T., Penuelas, J., Ciais, P., Rasmussen, K., Tian, F., Mbow, C., Al-Yaari, A., Rodriguez-Fernandez, N., Schurgers, G., Zhang, W., Chang, J., Kerr, Y., Verger, A., Tucker, C., Mialon, A., Rasmussen, L. V., Fan,

- L., and Fensholt, R. (2018). Satellite passive microwaves reveal recent climate-induced carbon losses in African drylands. *Nature Ecology and Evolution*, 2(5) :827–835.
- Bridle, J. S. (2011). Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. *Neurocomputing*, (C) :227–236.
- Briggs, B. (1980). Radar observations of atmospheric winds and turbulence : a comparison of techniques. *Journal of Atmospheric and Terrestrial Physics*, 42(9) :823–833.
- Brodzik, M. J., Billingsley, B., Haran, T., Raup, B., and Savoie, M. H. (2012). EASE-Grid 2.0 : Incremental but significant improvements for earth-gridded data sets. *ISPRS International Journal of Geo-Information*, 1(1) :32–45.
- Brown, S. T. and Ruf, C. S. (2005). Determination of an Amazon hot reference target for the on-orbit calibration of microwave radiometers. *Journal of Atmospheric and Oceanic Technology*, 22(9) :1340–1352.
- Buehler, S. A., Kuvatov, M., Sreerekha, T. R., John, V. O., Rydberg, B., Eriksson, P., and Notholt, J. (2007). A cloud filtering method for microwave upper tropospheric humidity measurements. *Atmospheric Chemistry and Physics*, 7(21) :5531–5542.
- Campbell, M. J. and Ulrichs, J. (1969). Electrical properties of rocks and their significance for lunar radar observations. *Journal of Geophysical Research*, 74(25) :5867–5881.
- Cao, B., Gruber, S., Zheng, D., and Li, X. (2020). The era5-land soil temperature bias in permafrost regions. *The Cryosphere*, 14(8) :2581–2595.
- Carver, K. R., Elachi, C., and Ulaby, F. T. (1985). Microwave Remote Sensing from Space. *Proceedings of the IEEE*, 73(6) :970–996.
- Cavalieri, D. J., Parkinson, C. L., Digirolamo, N., and Ivanoff, A. (2012). Intersensor calibration between F13 SSMI and F17 SSMIS for global sea ice data records. *IEEE Geoscience and Remote Sensing Letters*, 9(2) :233–236.
- Cavalieri, D. J., Parkinson, C. L., Gloersen, P., Comiso, J. C., and Zwally, H. J. (1999). Deriving long-term time series of sea ice cover from satellite passive-microwave multisensor data sets. *Journal of Geophysical Research : Oceans*, 104(C7) :15803–15814.
- Chander, G., Hewison, T. J., Fox, N., Wu, X., Xiong, X., and Blackwell, W. J. (2013). Overview of intercalibration of satellite instruments. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3) :1056–1080.
- Chen, T. and Guestrin, C. (2016). XGBoost : A Scalable Tree Boosting System.
- Chen, T., Rossow, W. B., and Zhang, Y. (2000). Radiative effects of cloud-type variations. *Journal of Climate*, 13(1) :264–286.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Choudhury, B. J., Schmugge, T. J., and Mo, T. (1981). A parameterization of effective soil temperature for microwave emissions. 87 :1301–1304.
- Christopher M. Bishop (1994). Mixture density networks.
- Cohn, S. A. (2017). *A New Edition of the International Cloud Atlas*. Vol 66 (1) edition.

- Comiso, J. C., Cavalieri, D. J., Parkinson, C. L., and Gloersen, P. (1997). Passive microwave algorithms for sea ice concentration : A comparison of two techniques. *Remote Sensing of Environment*, 60(3) :357–384.
- Comiso, J. C. and Nishio, F. (2008). Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research : Oceans*, 113(2) :1–22.
- Dai, L. and Che, T. (2009). Cross-platform calibration of SMMR, SSM/I and AMSR-E passive microwave brightness temperature. *Data Processing*, 7841 :784103–784103–10.
- Dai, Y., Wei, N., Yuan, H., Zhang, S., Shangguan, W., Liu, S., Lu, X., and Xin, Y. (2019). Evaluation of Soil Thermal Conductivity Schemes for Use in Land Surface Modeling. *Journal of Advances in Modeling Earth Systems*, 11(11) :3454–3473.
- De Jeu, R., DE NIJS, A., and Michel, V. K. (2017). Method and system for improving resolution of sensor data.
- De Jeu, R. A. M. (2003). Retrieval of Land Surface Parameters Using Passive Microwave Remote Sensing. *Thesis*, page 144.
- de Rosnay, P., Gruhier, C., Timouk, F., Baup, F., Mougin, E., Hiernaux, P., Kergoat, L., and LeDantec, V. (2009). Multi-scale soil moisture measurements at the Gourma meso-scale site in Mali. *Journal of Hydrology*, 375(1-2) :241–252.
- Derrien, M. and Le Gléau, H. (2005). MSG/SEVIRI cloud mask and type from SAFNWC. *International Journal of Remote Sensing*, 26(21) :4707–4732.
- Derrien, M. and le Gléau, H. (2010). Improvement of cloud detection near sunrise and sunset by temporal-differencing and region-growing techniques with real-time SEVIRI. *International Journal of Remote Sensing*, 31(7) :1765–1780.
- Dicke, R. H. (1946). The measurement of thermal radiation at microwave frequencies. *Review of Scientific Instruments*, 17(7) :268–275.
- Dirmeyer, P. A., Gentine, P., Ek, M. B., and Balsamo, G. (2019). Chapter 8 - land surface processes relevant to sub-seasonal to seasonal (s2s) prediction. In Robertson, A. W. and Vitart, F., editors, *Sub-Seasonal to Seasonal Prediction*, pages 165–181. Elsevier.
- Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-Rayes, M. A. (1985). Microwave Dielectric Behavior of Wet Soil-Part II : Dielectric Mixing Models. *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(1) :35–46.
- Dreiseitl, S. and Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models : a methodology review. *Journal of Biomedical Informatics*, 35(5-6) :352–359.
- Duan, S. B., Han, X. J., Huang, C., Li, Z. L., Wu, H., Qian, Y., Gao, M., and Leng, P. (2020). Land surface temperature retrieval from passive microwave satellite observations : State-of-the-art and future directions. *Remote Sensing*, 12(16).
- Duan, S. B., Li, Z. L., and Leng, P. (2017). A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data. *Remote Sensing of Environment*, 195 :107–117.
- ECMWF (2017). IFS DOCUMENTATION Cy43r3 Operational implementation 11 July 2017 PART IV : PHYSICAL PROCESSES. (July) :221.

- English, S., McNally, A., Borman, N., Salonen, K., Matricardi, M., Horányi, A., Rennie, M., Janisková, M., Di Michele, S., Geer, A., Di Tomaso, E., Cardinali, C., de Rosnay, P., Sabater, J., Bonavita, M., Albergel, C., Engelen, R., and Thépaut, J.-N. (2013). Impact of Satellite Data. *Technical Memorandum ECMWF*, (October) :46.
- Ermida, S. L., Jiménez, C., Prigent, C., Trigo, I. F., and DaCamara, C. C. (2017). Inversion of AMSR-E observations for land surface temperature estimation : 2. Global comparison with infrared satellite temperature. *Journal of Geophysical Research*, 122(6) :3348–3360.
- Ermida, S. L., Trigo, I. F., DaCamara, C. C., Göttsche, F. M., Olesen, F. S., and Hulley, G. (2014). 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.
- Ermida, S. L., Trigo, I. F., DaCamara, C. C., Jiménez, C., and Prigent, C. (2019). Quantifying the Clear-Sky Bias of Satellite Land Surface Temperature Using Microwave-Based Estimates. *Journal of Geophysical Research : Atmospheres*, 124(2) :844–857.
- FAO (1957). Soil map of the world. *Nature*, 179(4571) :1168.
- Favrichon, S., Jimenez, C., and Prigent, C. (2020). Inter-calibrating SMMR brightness temperatures over continental surfaces. *Atmospheric Measurement Techniques*, 13(10) :5481–5490.
- Favrichon, S., Prigent, C., and Jiménez, C. (2021). A method to downscale satellite microwave land-surface temperature. *Remote Sensing*, 13(7) :1–20.
- Favrichon, S., Prigent, C., Jimenez, C., and Aires, F. (2019). Detecting cloud contamination in passive microwave satellite measurements over land. *Atmospheric Measurement Techniques*, 12(3) :1531–1543.
- Feltz, M., Borbas, E., Knuteson, R., Hulley, G., and Hook, S. (2018). The combined ASTER and MODIS emissivity over land (CAMEL) global broadband infrared emissivity product. *Remote Sensing*, 10(7) :1–16.
- Fennig, K., Schröder, M., Andersson, A., and Hollmann, R. (2019). A Fundamental climate data record of SMMR, SSM/I, and SSMIS brightness temperatures. *Earth System Science Data Discussions*, 15(September) :1–54.
- Ferraro, R. R. (2004). Special sensor microwave imager derived global rainfall estimates for climatological applications. *Journal of Geophysical Research : Atmospheres*, 102(D14) :16715–16735.
- Fouquart, Y. (1987). Observations of Saharan aerosols : results of ECLATS field experiment. Part II : broadband radiative characteristics of the aerosols and vertical radiative flux divergence.
- Freitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., and Perdigão, R. (2013). Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, 34(9-10) :3051–3068.
- Gaiser, P. W., St. Germain, K. M., Twarog, E. M., Poe, G. A., Purdy, W., Richardson, D., Grossman, W., Jones, W. L., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R. M., and Chang, P. S. (2004). The windSat spaceborne polarimetric microwave radiometer : Sensor description and early orbit performance. *IEEE Transactions on Geoscience and Remote Sensing*, 42(11) :2347–2361.

- García-Haro, F., Camacho, F., and Meliá, J. (2013). The EUMETSAT Satellite Application Facility on Land Surface Analysis Product User Manual Vegetation Parameters (VEGA). 401(December) :1–46.
- Gloersen, P. and Barath, F. T. (1977). A Scanning Multichannel Microwave Radiometer for Nimbus-G and SeaSat-A. *IEEE Journal of Oceanic Engineering*, 2(2) :172–178.
- Gottsche, F. and Olesen, F. (2009). A simple physically based model of diurnal cycles of land surface temperature. *Proceedings 2009 EUMETSAT Meteorological Satellite Conference Bath, United Kingdom, EUM P.55*.
- Göttsche, F. M. and Olesen, F. S. (2009). Modelling the effect of optical thickness on diurnal cycles of land surface temperature. *Remote Sensing of Environment*, 113(11) :2306–2316.
- Greenwald, T. J., Stephens, G. L., Vonder Haar, T. H., and Jackson, D. L. (1993). A physical retrieval of cloud liquid water over the global oceans using special sensor microwave/imager (SSM/I) observations. *Journal of Geophysical Research*, 98(D10) :18471.
- Grody, N. C. (1991). Classification of snow cover and precipitation using the special sensor microwave imager. *Journal of Geophysical Research*, 96(D4) :7423–7435.
- Grody, N. C., Vinnikov, K. Y., Goldberg, M. D., Sullivan, J. T., and Tarpley, J. D. (2004). Calibration of multisatellite observations for climatic studies : Microwave Sounding Unit (MSU). *Journal of Geophysical Research D : Atmospheres*, 109(24) :1–12.
- Grody, N. C. and Weng, F. (2008). Microwave emission and scattering from deserts : Theory compared with satellite measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2) :361–375.
- Guillevic, P., Göttsche, F., Nickeson, J., Hulley, G., Ghent, D., Yu, Y., Trigo, I., Hook, S., Sobrino, J. A., Remedios, J., Román, M., and Camacho, F. (2018). Land surface temperature product validation best practice protocol version 1.1. *Best Practice for Satellite-Derived Land Product Validation : Land Product Validation Subgroup (WGCV/CEOS)*, (January) :58.
- Guilloteau, C., Foufoula-Georgiou, E., and Kummerow, C. D. (2017). Global multiscale evaluation of satellite passive microwave retrieval of precipitation during the TRMM and GPM eras : Effective resolution and regional diagnostics for future algorithm development. *Journal of Hydrometeorology*, 18(11) :3051–3070.
- Hansen, J., Ruedy, R., Sato, M., and Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, 48(4) :1–29.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730) :1999–2049.
- Hillel, D. (1998). *Environmental soil physics : Fundamentals, applications, and environmental considerations*. Elsevier.

- Hollinger, J. et al. (1990). SSM/I instrument evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, 28(5) :781–790.
- Holmes, T., de Rosnay, P., De Jeu, R., Wigneron, R.-P., Kerr, Y., Calvet, J.-C., Escorihuela, M., Saleh, K., and Lemaitre, F. (2006). A new parameterization of the effective temperature for l band radiometry. *Geophysical Research Letters*, 33(7).
- Holmes, T. R., Owe, M., De Jeu, R. A., and Kooi, H. (2008). Estimating the soil temperature profile from a single depth observation : A simple empirical heatflow solution. *Water Resources Research*, 44(2) :1–11.
- Holmes, T. R. H., De Jeu, R. A. M., Owe, M., and Dolman, A. J. (2009). Land surface temperature from Ka band (37 GHz) passive microwave observations. *Journal of Geophysical Research Atmospheres*, 114(4) :1–15.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2) :251–257.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5) :359–366.
- Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., and Iguchi, T. (2014). The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, 95(5) :701–722.
- Howard, L. (1803). *Essay on the Modification of Clouds*.
- Hughes, C. G. and Ramsey, M. S. (2010). Super-resolution of THEMIS thermal infrared data : Compositional relationships of surface units below the 100 meter scale on Mars. *Icarus*, 208(2) :704–720.
- Hyvärinen, O. (2014). A probabilistic derivation of heidke skill score. *Weather and Forecasting*, 29(1) :177–181.
- Islam, T., Rico-Ramirez, M. A., Srivastava, P. K., Dai, Q., Han, D., Gupta, M., and Zhuo, L. (2015). CLOUDET : A Cloud Detection and Estimation Algorithm for Passive Microwave Imagers and Sounders Aided by Naïve Bayes Classifier and Multilayer Perceptron. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(9) :4296–4301.
- Jezeq, K., Merry, C., Cavalieri, D., Grace, S., Bedner, J., Wilson, D., and Lampkin, D. (1991). Comparison Between SMMR and SSM/I Passive Microwave Data Collected over the Antarctic Ice Sheet. *BPRC Technical Report Number 91-03*.
- Jiménez, C., Prigent, C., Ermida, S. L., and Moncet, J. L. (2017). Inversion of AMSR-E observations for land surface temperature estimation : 1. Methodology and evaluation with station temperature. *Journal of Geophysical Research*, 122(6) :3330–3347.
- Jin, M. and Dickinson, R. E. (2010). Land surface skin temperature climatology : benefiting from the strengths of satellite observations. *Environmental Research Letters*, 5(4) :044004.
- Jinkai, X. (1990). Dielectric properties of minerals and their applications in microwave remote sensing. *Chinese Journal of Geochemistry*, 9(2) :169–177.
- Johannsen, F., Ermida, S., Martins, J. P., Trigo, I. F., Nogueira, M., and Dutra, E. (2019). Cold bias of ERA5 summertime daily maximum land surface temperature over Iberian Peninsula. *Remote Sensing*, 11(21) :1–21.

- Journet, E., Balkanski, Y., and Harrison, S. P. (2014). A new data set of soil mineralogy for dust-cycle modeling. *Atmospheric Chemistry and Physics*, 14(8) :3801–3816.
- Kalma, J. D., McVicar, T. R., and McCabe, M. F. (2008). Estimating land surface evaporation : A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, 29(4-5) :421–469.
- Karstens, U., Simmer, C., and Ruprecht, E. (1994). Remote sensing of cloud liquid water. *Meteorology and Atmospheric Physics*, 54(1-4) :157–171.
- Kasahara, M., Kachi, M., Inaoka, K., Fujii, H., Kubota, T., Shimada, R., and Kojima, Y. (2020). Overview and current status of GOSAT-GW mission and AMSR3 instrument. In Neeck, S. P., Hélière, A., and Kimura, T., editors, *Sensors, Systems, and Next-Generation Satellites XXIV*, volume 11530. International Society for Optics and Photonics, SPIE.
- Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., and Spencer, R. W. (2003). The Advanced Microwave Scanning Radiometer for the Earth observing system (AMSR-E), NASDA’s contribution to the EOS for global energy and water cycle studies. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2 PART 1) :184–193.
- Kemker, R., Salvaggio, C., and Kanan, C. (2017). Algorithms for Semantic Segmentation of Multispectral Remote Sensing Imagery using Deep Learning. pages 1–45.
- Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. M., Font, J., and Berger, M. (2001). Soil moisture retrieval from space : The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8) :1729–1735.
- Kilic, L., Prigent, C., Aires, F., Boutin, J., Heygster, G., Tonboe, R. T., Roquet, H., Jimenez, C., and Donlon, C. (2018). Expected Performances of the Copernicus Imaging Microwave Radiometer (CIMR) for an All-Weather and High Spatial Resolution Estimation of Ocean and Sea Ice Parameters. *Journal of Geophysical Research : Oceans*, 123(10) :7564–7580.
- Kingma, D. P. and Ba, J. L. (2015). Adam : A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15.
- Krasnopolsky, V. (2020). Using Machine Learning for Model Physics : an Overview. *NCEP*, (January).
- Krasnopolsky, V. M. and Schiller, H. (2003). Some neural network applications in environmental sciences. Part I : Forward and inverse problems in geophysical remote measurements. *Neural Networks*, 16(3-4) :321–334.
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutionalneural networks. *Proc. Advances in Neural Information Processing Systems 25*.
- Kroodasma, R. A., McKague, D. S., and Ruf, C. S. (2012). Inter-calibration of microwave radiometers using the vicarious cold calibration double difference method. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(3) :1006–1013.

- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J. (1998). The Tropical Rainfall Measuring Mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15(3) :809–817.
- Kustas, W. P., Norman, J. M., Anderson, M. C., and French, A. N. (2003). Estimating subpixel surface temperatures and energy fluxes from the vegetation index-radiometric temperature relationship. *Remote Sensing of Environment*, 85(4) :429–440.
- Lean (2017). Assimilation of Global Precipitation Mission (GPM) Microwave Imager (GMI) in all-sky conditions. *Boletín Técnico/Technical Bulletin*, 55(19) :255–262.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2323.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the Frog’s Eye Tells the Frog’s Brain. Proceedings of the. *Proceedings of the IRE*, 47(11) :1940–1959.
- Li, Z. L., Tang, B. H., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I. F., and Sobrino, J. A. (2013). Satellite-derived land surface temperature : Current status and perspectives. *Remote Sensing of Environment*, 131 :14–37.
- Liu, J. G. (2000). Smoothing Filter-based Intensity Modulation : A spectral preserve image fusion technique for improving spatial details. *International Journal of Remote Sensing*, 21(18) :3461–3472.
- Long, D. G., Remund, Q. P., and Daum, D. L. (1999). A cloud-removal algorithm for SSM/I data. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1 PART 1) :54–62.
- LST CCI (2019). Algorithm Theoretical Basis Document WP2.1 – DEL-LST-CCI-D2.2-ATBD.
- Lv, S., Zeng, Y., Wen, J., Zhao, H., and Su, Z. (2018). Estimation of penetration depth from soil effective temperature in microwave radiometry. *Remote Sensing*, 10(4) :1–19.
- Ma, X. L., Wan, Z., Moeller, C. C., Menzel, W. P., Gumley, L. E., and Zhang, Y. (2000). Retrieval of geophysical parameters from moderate resolution imaging spectroradiometer thermal infrared data : evaluation of a two-step physical algorithm. *Appl. Opt.*, 39(20) :3537–3550.
- Martin, M. A., Ghent, D., Pires, A. C., Göttsche, F. M., Cermak, J., and Remedios, J. J. (2019). Comprehensive in situ validation of five satellite land surface temperature data sets over multiple stations and years. *Remote Sensing*, 11(5).
- Martins, J. P., Trigo, I. F., Ghilain, N., Jimenez, C., Göttsche, F. M., Ermida, S. L., Olesen, F. S., Gellens-Meulenberghs, F., and Arboleda, A. (2019). An all-weather land surface temperature product based on MSG/SEVIRI observations. *Remote Sensing*, 11(24).
- Mätzler, C. (1998). Microwave Permittivity of Dry Sand. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1) :317–319.
- McFarland, M., Miller, R., and Neale, C. (1990). Land surface temperature derived from the SSM/I passive microwave brightness temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, 28(5) :839–845.

- McVicar, T. R. and Jupp, D. L. (2002). Using covariates to spatially interpolate moisture availability in the murray-darling basin a novel use of remotely sensed data. *Remote Sensing of Environment*, 79(2-3) :199–212.
- Mechri, R., Ottlé, C., Pannekoucke, O., Kallel, A., Maignan, F., Courault, D., and Trigo, I. F. (2016). Downscaling Meteosat land surface temperature over a heterogeneous landscape using a data assimilation approach. *Remote Sensing*, 8(7).
- Meirink, J. F., Roebeling, R. A., and Stammes, P. (2013). Inter-calibration of polar imager solar channels using SEVIRI. *Atmospheric Measurement Techniques*, 6(9) :2495–2508.
- Merlin, O., Walker, J. P., Chehbouni, A., and Kerr, Y. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment*, 112(10) :3935–3946.
- Minder, J. R., Mote, P. W., and Lundquist, J. D. (2010). Surface temperature lapse rates over complex terrain : Lessons from the cascade mountains. *Journal of Geophysical Research : Atmospheres*, 115(D14).
- Mo, Y., Xu, Y., Chen, H., and Zhu, S. (2021). A Review of Reconstructing Remotely Sensed Land Surface Temperature under Cloudy Conditions. *Remote Sensing*, 13(14) :2838.
- Moncet, J. L., Liang, P., Lipton, A. E., Galantowicz, J. F., and Prigent, C. (2011). Discrepancies between MODIS and ISCCP land surface temperature products analyzed with microwave measurements. *Journal of Geophysical Research Atmospheres*, 116(21) :1–13.
- Morcrette, J. J., Barker, H. W., Cole, J. N., Iacono, M. J., and Pincus, R. (2008). Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Monthly Weather Review*, 136(12) :4773–4798.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D., Piles, M., Rodríguez-Fernández, N., Zsoter, E., Buontempo, C., and Thépaut, J.-N. (2021). ERA5-Land : A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, (January 1981) :1–50.
- Nachtergaele, F., van Velthuizen, H., and Verelst, L. (2009). Harmonized world soil database.
- NASA (2015). GPM GMI 1C-constellation ATBD. (March).
- Njoku, E. G. and Kong, J. A. (1977). Theory for Passive Microwave Remote Sensing of Near-Surface Soil Moisture. *J Geophys Res*, 82(20) :3108–3118.
- Njoku, E. G., Stacey, J. M., and Barath, F. T. (1980). The Seasat Scanning Multichannel Microwave Radiometer (SMMR) : Instrument Description and Performance. *IEEE Journal of Oceanic Engineering*, 5(2) :100–115.
- Nogueira, M., Albergel, C., Boussetta, S., Johannsen, F., Trigo, I., Ermida, S., Martins, J. P., and Dutra, E. (2020). Role of vegetation in representing land surface temperature in the CHTESSEL (CY45R1) and SURFEX-ISBA (v8.1) land surface models : a case study over Iberia. *Geoscientific Model Development Discussions*, (March) :1–29.
- O’Dell, C. W., Wentz, F. J., and Bennartz, R. (2008). Cloud liquid water path from satellite-based passive microwave observations : A new climatology over the global oceans. *Journal of Climate*, 21(8) :1721–1739.

- Paul, M., Aires, F., Prigent, C., Trigo, I. F., and Bernardo, F. (2012). An innovative physical scheme to retrieve simultaneously surface temperature and emissivities using high spectral infrared observations from IASI. *Journal of Geophysical Research Atmospheres*, 117(11).
- Peng, J., Loew, A., Merlin, O., and Verhoest, N. E. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, 55(2) :341–366.
- Peters-Lidard, C. D., Blackburn, E., Liang, X., and Wood, E. F. (1998). The effect of soil thermal conductivity parameterization on surface energy fluxes and temperatures. *Journal of the Atmospheric Sciences*, 55(7) :1209–1224.
- Petty, G. W. (1994). Physical retrievals of over-ocean rain rate from multichannel microwave imagery. Part I : Theoretical characteristics of normalized polarization and scattering indices. *Meteorology and Atmospheric Physics*, 54(1-4) :79–99.
- Petty, G. W. (2001). Physical and microwave radiative properties of precipitating clouds. Part I : Principal component analysis of observed multichannel microwave radiances in tropical stratiform rainfall. *Journal of Applied Meteorology*, 40(12) :2105–2114.
- Petty, G. W. (2006). *A first course in atmospheric radiation*.
- Polcher, J., McAvaney, B., Viterbo, P., Gaertner, M. A., Hahmann, A., Mahfouf, J. F., Noilhan, J., Phillips, T., Pitman, A., Schlosser, C. A., Schulz, J. P., Timbal, B., Verseghy, D., and Xue, Y. (1998). A proposal for a general interface between land surface schemes and general circulation models. *Global and Planetary Change*, 19(1-4) :261–276.
- Prata, A. J., V. Casellescoll, C., Sobrino, J. A., and Oettle, C. (1995). Thermal remote sensing of land surface temperature from satellites : current status and future prospects. *Remote Sensing Reviews*, 12(3-4) :175–224.
- Prigent, C., Aires, F., and Rossow, W. B. (2006). Land surface microwave emissivities over the global for a decade. *Bulletin of the American Meteorological Society*, 87(11) :1573–1584.
- Prigent, C. and Jimenez, C. (2021). An evaluation of the synergy of satellite passive microwave observations between 1.4 and 36 GHz, for vegetation characterization over the Tropics. *Remote Sensing of Environment*, 257(September 2020) :112346.
- Prigent, C., Jimenez, C., and Aires, F. (2016). Toward “all weather,” long record, and real-time land surface temperature retrievals from microwave satellite observations. *Journal of Geophysical Research*, 121(10) :5699–5717.
- Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E. (2007). Global inundation dynamics inferred from multiple satellite observations, 1993-2000. *Journal of Geophysical Research Atmospheres*, 112(12) :1–13.
- Prigent, C., Phalippou, L., and English, S. (1997). Variational inversion of the SSM/I observations during the ASTEX campaign. *Journal of Applied Meteorology*, 36(5) :493–508.
- Prigent, C., Rochetin, N., Aires, F., Defer, E., Grandpeix, J.-Y. Y., Jimenez, C., and Papa, F. (2011). Impact of the inundation occurrence on the deep convection at continental scale from satellite observations and modeling experiments. *Journal of Geophysical Research Atmospheres*, 116(24) :n/a–n/a.

- Prigent, C., Rossow, W. B., and Matthews, E. (1998a). Global maps of microwave land surface emissivities : Potential for land surface characterization. *Radio Science*, 33(3) :745–751.
- Prigent, C., Rossow, W. B., and Matthews, E. (1998b). Global maps of microwave land surface emissivities : Potential for land surface characterization. *Radio Science*, 33(3) :745–751.
- Prigent, C., Rossow, W. B., Matthews, E., and Marticorena, B. (1999). Microwave radiometric signatures of different surface types in deserts. *Journal of Geophysical Research Atmospheres*, 104(D10) :12147–12158.
- Prigent, C. and Rossow, W. R. (1999). Retrieval of surface and atmospheric parameters over land from SSM/I : Potential and limitations. *Quarterly Journal of the Royal Meteorological Society*, 125(559) :2379–2400.
- Prigent, C., Wigneron, J. P., Rossow, W. B., and Pardo-Carrion, J. R. (2000). Frequency and angular variations of land surface microwave emissivities : Can we estimate SSM/T and AMSU emissivities from SSM/I emissivities? *IEEE Transactions on Geoscience and Remote Sensing*, 38(5 II) :2373–2386.
- Pulliainen, J. (2006). Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sensing of Environment*, 101(2) :257–269.
- Pulliainen, J. T., Grandell, J., and Hallikainen, M. T. (1997). Retrieval of surface temperature in boreal forest zone from SSM/I data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5) :1188–1200.
- Rapinel, S., Hubert-Moy, L., and Clement, B. (2015). Combined use of lidar data and multispectral earth observation imagery for wetland habitat mapping. *International Journal of Applied Earth Observation and Geoinformation*, 37 :56–64.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351 :234–241.
- Rosenkranz, P. W. (1998). Water vapor microwave continuum absorption : A comparison of measurements and models. *Radio Science*, 33(4) :919–928.
- Rossow, W. B., Garder, L., Lu, P.-J., and Walker, A. (1991). International satellite cloud climatology project (isccp) documentation of cloud data.
- Rossow, W. B. and Schiffer, R. A. (1999). Advances in Understanding Clouds from ISCCP. *Bulletin of the American Meteorological Society*, 80(11) :2261–2287.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. pages 1–14.
- Ruf, C. S. (2000). Detection of calibration drifts in spaceborne microwave radiometers using a vicarious cold reference. *IEEE Transactions on Geoscience and Remote Sensing*, 38(1 I) :44–52.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088) :533–536.

- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108(24) :9899–9904.
- Sapiano, M. R., Berg, W. K., McKague, D. S., and Kummerow, C. D. (2013). Toward an intercalibrated fundamental climate data record of the SSM/I sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 51(3) :1492–1503.
- Schmid, J. (2000). The SEVIRI instrument. *Proceedings of the 2000 EUMETSAT meteorological satellite data user’s conference, Bologna, Italy*, 29.
- Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., and Lehair, W. J. (2017). A closer look at the ABI on the goes-r series. *Bulletin of the American Meteorological Society*, 98(4) :681–698.
- Segal-Rozenhaimer, M., Li, A., Das, K., and Chirayath, V. (2020). Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (cnn). *Remote Sensing of Environment*, 237 :111446.
- Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H. (2014). A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems*, 6(1) :249–263.
- Sippe, S. J., Hamilton, S. K., Melack, J. M., and Novo, E. M. (1998). Passive microwave observations of inundation area and the area/stage relation in the amazon river floodplain. *International Journal of Remote Sensing*, 19(16) :3055–3074.
- Smith, G. D., Smith, G. D., and Smith, G. D. S. (1985). *Numerical solution of partial differential equations : finite difference methods*. Oxford university press.
- Spencer, R. W., Goodman, H. M., and Hood, R. E. (1989). Precipitation Retrieval over Land and Ocean with the SSM/I : Identification and Characteristics of the Scattering Signal. *Journal of Atmospheric and Oceanic Technology*, 6(2) :254–273.
- Sun, D., Li, Y., Zhan, X., Houser, P., Yang, C., Chiu, L., and Yang, R. (2019). Land surface temperature derivation under all sky conditions through integrating AMSR-E/AMSR-2 and MODIS/GOES observations. *Remote Sensing*, 11(14) :1–20.
- Sun, P. and Chawla, S. (2004). On local spatial outliers. In *Fourth IEEE International Conference on Data Mining (ICDM’04)*, pages 209–216. IEEE.
- Tesauro, G. (1994). TD-Gammon, a Self-Teaching Backgammon Program, Achieves Master-Level Play. *Neural Computation*, 6(2) :215–219.
- Thomas, B. (2004). *Etude et réalisation d ’ une tête de réception hétérodyne en ondes submillimétriques pour l ’ étude des atmosphères et surfaces de planètes*.
- Trigo, I. F., Peres, L. F., DaCamara, C. C., and Freitas, S. C. (2008). Thermal land surface emissivity retrieved from SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2) :307–315.
- Turk, J., Vivekanandan, J., Marzano, F. S., Hood, R. E., Spencer, R. W., and LaFontaine, F. J. (1994). Active and passive microwave remote sensing of precipitating storms during CaPE. Part I : Advanced microwave precipitation radiometer and polarimetric radar measurements and models. *Meteorology and Atmospheric Physics*, 54(1-4) :3–27.

- Ulaby, F. T., Bengal, T. H., Dobson, M. C., East, J. R., Garvin, J. B., and Evans, D. L. (1990). Microwave Dielectric Properties of Dry Rocks. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3) :325–336.
- Ulaby, F. T., Long, D. G., and Fung, A. (2014). *Microwave Radar and Radiometric Remote Sensing*. Ann Arbor : The University of Michigan Press.
- van de Griend, A. A. and Owe, M. (1994). Microwave vegetation optical depth and inverse modelling of soil emissivity using Nimbus/SMMR satellite observations. *Meteorology and Atmospheric Physics*, 54(1-4) :225–239.
- Van Den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J. (2016). LS3MIP (v1.0) contribution to CMIP6 : The Land Surface, Snow and Soil moisture Model Intercomparison Project - Aims, setup and expected outcome. *Geoscientific Model Development*, 9(8) :2809–2832.
- Viterbo, P., Beljaars, A., Mahfouf, J.-F., and Teixeira, J. (1999). The representation of soil moisture freezing and its impact on the stable boundary layer. *Quarterly Journal of the Royal Meteorological Society*, 125(559) :2401–2426.
- Von Storch, H. (1999). On the use of 'inflation' in statistical downscaling. *Journal of Climate*, 12(12) :3505–3506.
- Vries, D. d. (1963). Physics of plant environment. *Environmental Control of Plant Growth. Edited.*
- Wan, Z. and Dozier, J. (1989). Land-Surface Temperature Measurement from Space : Physical Principles and Inverse Modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 27(3) :268–278.
- Wan, Z. and Li, Z. L. (1997). A physics-based algorithm for retrieving land-surface emissivity and temperature from eos/modis data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(4) :980–996.
- Weng, F. and Grody, N. C. (1998). Physical retrieval of land surface temperature using the special sensor microwave imager. *Journal of Geophysical Research Atmospheres*, 103(D8) :8839–8848.
- Wentz, F. J. (1997). A well-calibrated ocean algorithm for special sensor microwave / imager. *Journal of Geophysical Research : Oceans*, 102(C4) :8703–8718.
- Wentz, F. J. (2013). SSM/I version-7 calibration report. *Remote Sensing Systems Tech. Rep*, 11012 :46.
- Wentz, F. J. and Draper, D. (2016). On-orbit absolute calibration of the global precipitation measurement microwave imager. *Journal of Atmospheric and Oceanic Technology*, 33(7) :1393–1412.
- Wigneron, J. P., Chanzy, A., De Rosnay, P., Rüdiger, C., and Calvet, J. C. (2008). Estimating the effective soil temperature at L-band as a function of soil properties. *IEEE Transactions on Geoscience and Remote Sensing*, 46(3) :797–807.
- Wilheit, T. T. (1978). Radiative Transfer in a Plane Stratified Dielectric. *IEEE Transactions on Geoscience Electronics*, 16(2) :138–143.

- Windsor, E. P. and Charlton, J. E. (1991). European development in microwave radiometry : 1978-1988. *International Journal of Remote Sensing*, 12(8) :1741–1751.
- Wu, P., Yin, Z., Yang, H., Wu, Y., and Ma, X. (2019). Reconstructing geostationary satellite land surface temperature imagery based on a multiscale feature connected convolutional neural network. *Remote Sensing*, 11(3).
- Wu, P., Yin, Z., Zeng, C., Duan, S.-B., Gottsche, F.-M., Li, X., Ma, X., Yang, H., and Shen, H. (2021). Spatially Continuous and High-Resolution Land Surface Temperature Product Generation : A Review of Reconstruction and Spatiotemporal Fusion Techniques. *IEEE Geoscience and Remote Sensing Magazine*, pages 0–0.
- Wyser, K., Van Noije, T., Yang, S., Von Hardenberg, J., O’Donnell, D., and Döscher, R. (2020). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geoscientific Model Development*, 13(8) :3465–3474.
- Xu, S. and Cheng, J. (2021). A new land surface temperature fusion strategy based on cumulative distribution function matching and multiresolution Kalman filtering. *Remote Sensing of Environment*, 254(February 2020) :112256.
- Yang, H., Lv, L., Xu, H., He, J., and Wu, S. (2011). Evaluation of FY3B-MWRI instrument on-orbit calibration accuracy. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2252–2254. IEEE.
- Yang, J. X., McKague, D. S., and Ruf, C. S. (2016). Boreal, Temperate, and Tropical Forests as Vicarious Calibration Sites for Spaceborne Microwave Radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2) :1035–1051.
- Yao, W., Zeng, Z., Lian, C., and Tang, H. (2018). Pixel-wise regression using U-Net and its application on pansharpening. *Neurocomputing*, 312 :364–371.
- Ye, L., Cao, Z., and Xiao, Y. (2017). Deepcloud : Ground-based cloud image categorization using deep convolutional features. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1) :5729–5740.
- Yue, L., Shen, H., Li, J., Yuan, Q., and Zhang, H. (2016). Image super-resolution : The techniques , applications , and future. *Signal Processing*, 128 :389–408.
- Zhan, W., Chen, Y., Wang, J., Zhou, J., Quan, J., Liu, W., and Li, J. (2012). Downscaling land surface temperatures with multi-spectral and multi-resolution images. *International Journal of Applied Earth Observation and Geoinformation*, 18(1) :23–36.
- Zhan, W., Chen, Y., Zhou, J., Wang, J., Liu, W., Voogt, J., Zhu, X., Quan, J., and Li, J. (2013). Disaggregation of remotely sensed land surface temperature : Literature survey, taxonomy, issues, and caveats. *Remote Sensing of Environment*, 131(19) :119–139.
- Zhang, Q., Wang, N., Cheng, J., and Xu, S. (2020). A Stepwise Downscaling Method for Generating High-Resolution Land Surface Temperature from AMSR-E Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13 :5669–5681.
- Zhang, T., Zhang, L., Zhao, S., Jiang, L., and Chai, L. (2012). A statistic model developed to estimate the penetration depth using passive microwave remote sensing. *International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 666–669.

- Zhang, X., Zhou, J., Gottsche, F. M., Zhan, W., Liu, S., and Cao, R. (2019). A Method Based on Temporal Component Decomposition for Estimating 1-km All-Weather Land Surface Temperature by Merging Satellite Thermal Infrared and Passive Microwave Observations. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7) :4670–4691.
- Zhao, T. J., Zhang, L. X., Shi, J. C., and Jiang, L. M. (2011). A physically based statistical methodology for surface soil moisture retrieval in the Tibet Plateau using microwave vegetation indices. *Journal of Geophysical Research Atmospheres*, 116(8) :1–11.
- Zhao, Y., Norouzi, H., Azarderakhsh, M., and AghaKouchak, A. (2021). Global Patterns of Hottest, Coldest and Extreme Diurnal Variability on Earth. *Bulletin of the American Meteorological Society*, pages 1–23.
- Zheng, Y., Wang, S., Feng, J., Ouyang, Z., and Li, X. (2005). Measurement of the complex permittivity of dry rocks and minerals : Application of polythene dilution method and Lichtenecker’s mixture formulae. *Geophysical Journal International*, 163(3) :1195–1202.
- Zhou, F. C., Li, Z. L., Wu, H., Duan, S. B., Song, X., and Yan, G. (2018). A practical two-stage algorithm for retrieving land surface temperature from AMSR-E Data - A Case Study over China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(6) :1939–1948.
- Zhou, J., Zhang, X., Zhan, W., Göttsche, F. M., Liu, S., Olesen, F. S., Hu, W., and Dai, F. (2017). A Thermal Sampling Depth Correction Method for Land Surface Temperature Estimation From Satellite Passive Microwave Observation Over Barren Land. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8) :4743–4756.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F. (2017). Deep Learning in Remote Sensing : A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4) :8–36.