



HAL
open science

Prise en compte des incertitudes du transfert radiatif par une approche ensembliste pour l'assimilation des données satellitaires micro-ondes en ciel nuageux

Marylise Barreyat

► **To cite this version:**

Marylise Barreyat. Prise en compte des incertitudes du transfert radiatif par une approche ensembliste pour l'assimilation des données satellitaires micro-ondes en ciel nuageux. Océan, Atmosphère. Université Paul Sabatier - Toulouse III, 2022. Français. NNT : 2022TOU30262 . tel-04073259

HAL Id: tel-04073259

<https://theses.hal.science/tel-04073259v1>

Submitted on 18 Apr 2023

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

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



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
MARYLIS BARREYAT**

Le 28 novembre 2022

**Prise en compte des incertitudes du transfert radiatif par une
approche ensembliste pour l'assimilation des données
satellitaires micro-ondes en ciel nuageux**

Ecole doctorale : **SDU2E - Sciences de l'Univers, de l'Environnement et de
l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :

CNRM - Centre National de Recherches Météorologiques

Thèse dirigée par

Philippe CHAMBON et Jean-François MAHFOUF

Jury

Mme CÉCILE MALLET, Rapporteur

M. FREDERIC SZCZAP, Rapporteur

M. CHRISTOPHE ACCADIA, Examineur

M. JOËL BEDARD, Examineur

M. PHILIPPE CHAMBON, Directeur de thèse

M. JEAN-FRANÇOIS MAHFOUF, Co-directeur de thèse

M. GHISLAIN FAURE, Co-encadrant de thèse

M. JEAN-PIERRE CHABOUREAU, Président

À Martine, Philippe et Roland.

Remerciements

Nous voici à la fin de ces 3 années de thèse qui auront mené à l'écriture de ce manuscrit. Je tiens à exprimer ma reconnaissance envers toutes les personnes qui ont rendu ce travail possible. Je remercie premièrement mes directeurs de thèse. Philippe, tu es la personne qui m'a fait grandir professionnellement parlant et grâce à qui tout ça a été possible. Depuis mon stage de M1 jusqu'à la fin de cette thèse, tu m'as transmis énormément de connaissances tout en m'apportant un soutien constant. Je serai toujours impressionnée par ton large spectre de compétences te permettant de t'impliquer dans nombre de projets scientifiques. Jean-François, tu as guidé ce travail de thèse grâce à ton remarquable bagage scientifique que j'ai découvert lors de nos longues discussions sur l'assimilation de données. Je retiendrai aussi tes qualités humaines qui auront permis à tout un chacun de passer des moments dans une ambiance conviviale et agréable. Ghislain, depuis mon stage de M1 et jusqu'à maintenant, tu as toujours porté un regard pertinent sur mes diverses présentations. Je te remercie particulièrement d'avoir toujours décelé les moments où j'avais besoin de soutien, nos discussions ont été précieuses.

Je remercie tous les collègues du GMAP avec qui j'ai eu le plaisir de manger pendant ces 3 années : Christophe, Eric, Etienne, Fabrice, François, Loïk et Valérie. Je remercie en particulier Loïk grâce à qui la dernière partie de cette thèse a été possible. Tu m'as apporté de précieux conseils et m'a permis d'appréhender l'assimilation d'ensemble grâce à tes nombreuses connaissances sur le sujet.

En outre, je souhaite remercier les personnes avec qui j'ai partagé de beaux moments lors des pauses à la tisanerie, des repas ou que j'ai croisé pendant mon aventure à Météo-France : Adrien, Charles, Clément, Danaé, Elisa, Louis, Mary, Nicolas, Oliver, Olivier, Robin, Sarah et Thomas. Je remercie particulièrement plusieurs personnes : Guillaume qui m'a apporté un grand soutien lors des nombreux confinements, Michael qui a toujours été d'une très grande gentillesse, Rohit grâce à qui j'ai progressé en anglais et qui m'a appris que toute une vie pouvait être planifiée, Sophie avec qui j'ai eu des conversations très enrichissantes sur le plan humain et Thomas pour nos discussions très enjouées depuis le stage de M1.

Je remercie également des personnes que j'ai croisées plus souvent à la tisanerie : Claude pour toutes les fois où j'ai interrompu ses repas pour discuter, Eric pour toutes nos discussions sur le foot, la musique, les voitures mais aussi pour ton aide technique, et Thibaut avec qui j'ai souvent ri et qui m'a donné de nombreux conseils pour ma présentation finale.

Je remercie de même les personnes de mon bureau : Olivier, Salomé et en particulier Antoine avec qui j'ai beaucoup échangé sur nos fins de thèse respectives et Ethel avec qui j'ai eu le plaisir de partager une conférence aux États-Unis.

Je suis également reconnaissante envers les membres de mon comité et mon jury de thèse pour leur temps et leur contribution à l'amélioration de ce travail.

Enfin, je souhaite remercier ma famille et mes amis pour leur soutien inconditionnel sans lequel ce travail n'aurait été possible. Je pense particulièrement à mes frères, ma grand-mère, ma belle-mère et mon père ainsi qu'à Antoine, Benjamin, Cédric, Emelyne, Juliette, Nathan, Suzanne et Vanessa.

Résumé

Les observations satellitaires micro-ondes sont une composante essentielle du système d'observation global ; elles contribuent fortement à la qualité des prévisions météorologiques. Sensibles à la vapeur d'eau, la température et aux hydrométéores sous toutes leurs formes, le domaine spectral des micro-ondes est caractérisé par un contenu riche en information sur la structure verticale de l'atmosphère. Dans les années à venir, plusieurs instruments micro-ondes vont être embarqués à bord de satellites et vont permettre d'enrichir le système d'observation déjà disponible. Afin de bénéficier au mieux de ces futurs instruments, il est important de préparer leur assimilation au sein des modèles de Prévision Numérique du Temps (PNT). Un des éléments limitant la pleine exploitation des observations micro-ondes réside dans la spécification des propriétés de diffusion simple (SSP, 'single scattering properties') à l'opérateur d'observations (Geer et Baordo, 2014[1], Geer et al. 2017[2]). Cette spécification se retrouve au sein du processus d'assimilation de données lors de la simulation des observations à partir de profils atmosphériques du modèle. Ces spécifications sont particulièrement importantes pour des observations localisées en ciel nuageux car elles permettent de caractériser les effets de diffusion du rayonnement électromagnétique par les hydrométéores. À l'heure actuelle, la plupart des systèmes de PNT n'exploitent qu'une unique SSP pour la simulation de chaque hydrométéore. Les hydrométéores solides étant caractérisés par une grande variabilité morphologique, cela amène à une simplification importante des processus microphysiques présents dans la nature (Geer and Baordo 2014[1], Haddad et al. 2015[3], Guerbette et al. 2016[4], Mangla et al. 2021[5]). Ces dernières années, des méthodes traitant plusieurs SSPs ont commencé à apparaître (Baran and Labonnote 2007[6], Kulie et al. 2010[7], Ringerud et al. 2019[8]). Néanmoins, ces méthodes peuvent encore être améliorées afin de s'adapter aux diverses situations météorologiques et aux instruments de télédétection dans différents domaines spectraux.

L'objectif de cette thèse est d'explorer de nouvelles méthodes permettant de s'affranchir d'un choix de particule optimale au sein des simulations du transfert radiatif et tirant avantage de la diversité des bases de données de SSPs disponibles. Ces méthodes seront ici expérimentées pour l'hydrométéore de la neige pour lequel la simulation des hautes fréquences micro-ondes est très sensible. Nous avons dans un premier temps examiné si les observations possèdent un contenu en information suffisant pour définir de manière dynamique lors du processus d'assimilation de données des combinaisons de SSPs optimales selon la situation météorologique observée. Pour cela, nous avons considéré une inversion Bayésienne exploitée opérationnellement à Météo-France pour l'assimilation de données satellitaires micro-ondes en ciel nuageux.

Des études de sensibilité ont premièrement été conduites, puis la méthode a été modifiée pour prendre en compte un ensemble de SSPs au sein des nuages. Des expériences ont ensuite été menées avec la nouvelle méthode au sein des modèles AROME-Antilles (Faure et al. 2020[9]) et ARPEGE (Bouyssel et al. 2022[10]) à l'aide des observations du radiomètre GMI à bord du satellite Américano-Japonais GPM-Core. Pour simuler les radiances, différentes SSPs de la base de Liu et al. 2008[11] ont été spécifiées au modèle de transfert radiatif RTTOV-SCATT v12 (Saunders et al. 2018[12]). Les résultats ont montré la pertinence de chaque SSP au sein des inversions. Des expériences d'assimilation ont ensuite été réalisées au sein du modèle ARPEGE et ont montré des améliorations lors de l'utilisation de quatre SSPs sur les scores de prévision du vent. Deuxièmement, nous avons voulu exploiter la diversité des SSPs disponibles comme moyen de propagation des erreurs de l'opérateur d'observation lors de l'assimilation de données. Un système d'assimilation d'ensemble (Berre et al. 2015[13], Berre et al. 2019[14]) construisant des matrices de covariances d'erreurs d'ébauche a été utilisé pour caractériser les incertitudes liées aux propriétés radiatives. Une assimilation 4D-Var étant prévue pour remplacer la méthode Bayésienne actuelle, des expériences ont été menées à l'aide d'une approche 4D-Var 'all-sky'. Ces expériences ont été réalisées à l'aide du modèle ARPEGE et des observations du radiomètre GMI en exploitant la base de Liu et al. 2008[11]. Des scores objectifs ont mis en évidence des améliorations significatives apportées par une sélection aléatoire de SSPs sur la qualité des prévisions de vents, de l'humidité spécifique et de la température.

Abstract

Microwave satellite observations are an essential component of the global observing system ; they strongly contribute to the quality of weather forecasts. Sensitive to water vapour, temperature and hydrometeors in all their forms, the microwave spectral domain is characterized by a rich information content on the vertical structure of the atmosphere. In the coming years, several microwave instruments will be carried on board satellites and will enhance the already available constellation. In order to fully benefit from these future instruments, it is important to prepare for their assimilation within Numerical Weather Prediction (NWP) models. One of the limitations preventing the full exploitation of microwave observations lies in the specification of the radiative properties to the observation operator (Geer and Baordo, 2014[1], Geer et al. 2017[2]). This specification is required within the data assimilation process during the simulation of observations from atmospheric profiles of the model. These specifications are particularly important for observations localized in cloudy sky because they enable to characterize the effects of scattering of electromagnetic radiation by hydrometeors. Currently, most NWP systems seek to define a unique radiative property for the simulation of each hydrometeor. Since solid hydrometeors are characterized by high morphological variability, this leads to a significant simplification of the microphysical processes present in nature (Geer and Baordo 2014[1], Haddad et al. 2015[3], Guerbette et al. 2016[4], Mangla et al. 2021[5]). In recent years, methods addressing multiple properties started to emerge (Baran and Labonnote 2007[6], Kulie et al. 2010[7], Ringerud et al. 2019[8]). Nevertheless, these methods can still be improved in order to be adapted to various meteorological situations and remote sensing instruments in different spectral domains.

The objective of this thesis is to explore new methods that can overcome the need of an optimal particle selection within radiative transfer simulations and take advantage of the diversity of the available radiative property databases. These methods will be tested here for the snow hydrometeor for which the high frequency microwave simulation is very sensitive. We first examined whether the observations have enough information content to dynamically define during the data assimilation process optimal combinations of radiative properties according to the observed meteorological situation. For this, we considered a Bayesian inversion operationally used at Météo-France for the assimilation of microwave satellite data in cloudy sky.

Sensitivity studies were first conducted, then the method was modified to take into account a set of radiative properties within clouds. Experiments were then conducted with the new method within the AROME-Antilles (Faure et al. 2020[9]) and the ARPEGE (Bouyssel et al. 2022[10]) models using observations from the GMI radiometer on board the American-Japanese GPM-Core satellite. To simulate radiances, different properties from the Liu et al. 2008[11] database were specified to the radiative transfer model RTTOV-SCATT v12 (Saunders et al. 2018[12]). The results showed the relevance of each property within the inversions. Assimilation experiments were then carried out within the ARPEGE model and showed improvements on forecast wind scores when using four properties. Secondly, we wanted to exploit the diversity of available radiative properties as a way of propagation of the observation operator uncertainties within data assimilation systems. An ensemble assimilation system (Berre et al. 2015[13], Berre et al. 2019[14]) building background error covariance matrices was used to characterize the uncertainties related to radiative properties. Since a 4D-Var assimilation is planned to replace the current Bayesian method, experiments were carried using an all-sky 4D-Var approach. These experiments were performed using the ARPEGE model and GMI radiometer observations by using the Liu et al. 2008[11] database. Objective skill scores showed significant improvements by a random selection of radiative properties on the quality of wind, specific humidity and temperature forecasts.

Table des matières

Introduction	2
1 La prévision numérique du temps	7
1.1 Les modèles numériques	7
1.1.1 Le modèle ARPEGE	8
1.1.2 Le modèle AROME	10
1.2 Les observations	14
1.2.1 Les observations depuis la Terre	14
1.2.2 Les observations depuis l'espace	16
1.3 L'assimilation de données	18
1.3.1 L'Interpolation Optimale	18
1.3.2 Les méthodes variationnelles	22
2 Modélisation du transfert radiatif micro-ondes dans les nuages	29
2.1 Le transfert radiatif	29
2.1.1 Principe général	29
2.1.2 La théorie du corps noir	34
2.1.3 L'équation du transfert radiatif	35
2.1.4 Un code de transfert radiatif : RTTOV-SCATT	39
2.2 Calcul des propriétés radiatives des hydrométéores	45
2.2.1 Les régimes de diffusion	45
2.2.2 Historique des bases de données	51
3 Méthodes d'assimilation pour les observations micro-ondes nuageuses et précipitantes	53
3.1 L'initialisation physique	53
3.2 Les méthodes d'assimilation indirectes	54

3.2.1	1D-Var+4D-Var	54
3.2.2	1D-Bay+4D-Var	57
3.3	Les méthodes d'assimilation directe	58
3.3.1	4D-Var avec le modèle global IFS	58
3.3.2	Filtre ensembliste avec le modèle régional WRF	61
3.4	Avantages et inconvénients des différentes méthodes	62
4	Problématique de la thèse	63
4.1	Contexte	63
4.2	Outils	66
4.3	Méthodologie	71
4.4	Plan de la suite du manuscrit	72
5	Étude de sensibilité de l'inversion Bayésienne	75
5.1	Introduction	75
5.2	Publication	77
5.3	Conclusions	104
6	Étude d'une méthode d'inversion considérant les incertitudes du transfert radiatif	107
6.1	Introduction	107
6.2	Publication en cours de révision	109
6.3	Résultats complémentaires	127
6.3.1	Étude statistique avec les prédicteurs GPROF	127
6.3.2	Résultats d'expériences d'assimilation avec le modèle ARPEGE	127
6.4	Conclusions	134
7	Prise en compte des incertitudes du transfert radiatif dans une assimilation d'ensemble	137
7.1	Introduction	137
7.2	Méthodologie	138
7.3	Effets potentiels des modifications ajoutées dans l'AEARP	140
7.3.1	Expansion linéarisée des erreurs d'analyse et d'ébauche	142
7.3.2	Expansion linéarisée des perturbations d'analyse et d'ébauche	142
7.3.3	Effets potentiels de l'ajout de nouvelles observations et des perturbations associées	144
7.4	Résultats d'expériences à l'aide du système d'assimilation d'ensemble AEARP	144

TABLE DES MATIÈRES

7.5	Résultats d'expériences d'assimilation avec le modèle ARPEGE	153
7.6	Conclusions	154
	Conclusions et perspectives	161
	Annexe	177

Introduction

La Prévision Numérique du Temps : importance des observations

La Prévision Numérique du Temps (PNT) est une discipline météorologique qui permet de prévoir l'état atmosphérique à l'horizon de quelques jours. Pour cela, l'évolution temporelle de champs géophysiques est simulée en prenant en compte de nombreux processus physiques et en résolvant des équations basées sur les lois de la dynamique de l'atmosphère au moyen d'une discrétisation spatio-temporelle. Chacune de ces simulations, ou prévisions, requièrent des conditions initiales reflétant le plus fidèlement possible l'état atmosphérique réel. Les différentes observations de l'atmosphère jouent alors un rôle primordial, car elles sont combinées de manière optimale à une prévision précédente afin d'obtenir le nouvel état atmosphérique duquel démarrera la prévision suivante. Cet état atmosphérique est appelé *analyse* et s'obtient au moyen d'un algorithme mathématiques dit d'*assimilation de données*.

Dans le but de construire une analyse, plusieurs types d'observations renseignant sur l'état atmosphérique et sur l'état de la surface et provenant de différents instruments sont exploitées. On peut par exemple citer les stations de surface, les radio-sondages ou encore les radars au sol. Depuis quelques années, les modèles de PNT sont également initialisés à l'aide d'observations fournies par les satellites météorologiques dont les premiers ont été lancés il y a plus de 60 ans. Parmi les instruments embarqués à bord des satellites, on trouve les sondeurs ou encore les imageurs. Deux catégories principales de satellites ont été définies en fonction de leur orbite. Les premiers se trouvent à une très haute altitude de l'ordre de 36 000 km et sont appelés satellites géostationnaires. Les deuxièmes, caractérisés souvent par une orbite gardant la même orientation par rapport au soleil sont situés à une altitude de l'ordre de plusieurs centaines de kilomètres et sont appelés satellites défilants. Le premier satellite météorologique mis en orbite fut le satellite défilant TIROS-1 (Television InfraRed Observation Satellite) lancé en 1960 et conçu par la NASA (National Aeronautics and Space Administration). Ce satellite avait à son bord des caméras de télévision permettant d'acquérir des images de la Terre depuis l'espace. On retrouve aussi d'autres types d'orbites, comme par exemple le 'Medium Earth Orbit' pour les GNSS (Géolocalisation et Navigation par un Système de Satellites) ou l'orbite de Molnia capable de couvrir les pôles.

De nos jours, la plupart des instruments embarqués opèrent dans les domaines du visible, de l'infrarouge, et des micro-ondes. Ce dernier domaine est particulièrement intéressant car le rayonnement micro-ondes traverse les nuages et permet de sonder l'atmosphère par tout type de temps. De plus, le spectre micro-ondes possède des fréquences particulièrement sensibles à la vapeur d'eau, la température et les hydrométéores présents dans l'atmosphère. Ce type d'observations peut donc fournir d'importantes informations lors d'événements fortement précipitants, comme les tempêtes et les cyclones. La prise en compte des observations micro-ondes lors de la construction des analyses météorologiques est ainsi d'une importance fondamentale. Historiquement, l'assimilation des observations satellitaires micro-ondes a démarré en conditions de ciel clair, autrement dit sans nuages ou précipitations, puis plus récemment en conditions nuageuses avec à la clé des améliorations significatives sur la qualité des prévisions.

L'assimilation des observations micro-ondes nuageuses

Un des premiers instruments micro-ondes embarqué fut l'imageur SSM/I (Special Sensor Microwave/Imager) à bord d'un des satellites DMSP (Defense Meteorological Satellite Program) en 1987. C'est avec l'arrivée

de cet instrument que les recherches sur l'assimilation de ce type d'observations en milieu nuageux ont débutées, puis avec celle de l'imageur TMI (Tropical Rainfall Measuring Mission Microwave Imager) mis en orbite en 1997. Aussitôt, des problématiques concernant l'assimilation de telles observations sont apparues. En effet, les processus microphysiques et les effets de la convection décrivant le comportement des scènes météorologiques nuageuses sont très complexes à modéliser. De ce fait, il était impossible à l'époque d'assimiler les observations micro-ondes en ciel nuageux sans traitements préliminaires. Les premières méthodes ont ainsi consisté à assimiler des taux de pluie dérivés de radiances micro-ondes (Krishnamurti 1991[15]). Ces méthodes ont progressivement été remplacées par des méthodes indirectes, basées sur un formalisme Bayésien ou variationnel, consistant à assimiler des taux de pluie dérivés en deux étapes (Marécal et Mahfouf 2000[16], 2002[17], 2003[18]). Par la suite, de nombreux travaux ont permis d'améliorer le traitement des interactions entre les rayonnements électromagnétiques et l'atmosphère au sein des processus d'assimilation de données. Les méthodes indirectes, aidées d'un nouvel opérateur permettant une simulation réaliste des observations micro-ondes en zones nuageuses, ont alors permis l'assimilation des observations micro-ondes et non plus des produits dérivés de celles-ci. Une des méthodes reposant sur une inversion Bayésienne est exploitée opérationnellement à Météo-France depuis 2010 pour assimiler les observations provenant des radars sols (Caumont et al. 2010[19], Wattrelot et al. 2014[20], Borderies et al. 2019[21]) et à l'agence météorologique japonaise JMA (Japan Meteorological Agency) pour assimiler des observations de radars spatiaux (Ikuta et al. 2021[22]).

Malgré des résultats encourageants amenant à poursuivre les efforts, de nombreuses problématiques liées à l'utilisation des méthodes indirectes subsistaient à la fin des années 2000 (Errico et al. 2007[23], Geer et al. 2008[24]). Une d'entre elles reposait sur l'assimilation différenciée des observations micro-ondes en ciel nuageux et en ciel clair. Ces traitements différents amenaient à des biais d'échantillonnage importants. Des efforts ont par la suite été réalisés afin de pouvoir remplacer ces méthodes par une assimilation directe au sein des systèmes de PNT à l'aide d'une approche dite 'all-sky', i.e. ne différenciant pas les observations en ciel nuageux de celles en ciel clair lors du processus d'assimilation de données. Ces dernières améliorations ont ainsi permis d'assimiler, pour la première fois, les observations micro-ondes en ciel nuageux et en ciel clair simultanément au sein du système d'assimilation opérationnel au Centre Européen de Prévision Météorologique à Moyen Terme (CEPMNT) en 2009 (Bauer et al. 2010[25], Geer et al. 2010[26]). Cette méthode a permis d'apporter de bons résultats sur la qualité des prévisions. Cependant, et en dépit des derniers progrès réalisés, l'assimilation des observations micro-ondes nuageuses présente encore de nombreux enjeux. Un d'entre eux concerne la modélisation des processus microphysiques décrivant les interactions entre la matière et le rayonnement au sein de l'atmosphère (Geer et Baordo 2014[1]).

Motivations de l'étude et problématiques abordées

L'ensemble des méthodes permettant l'assimilation des observations micro-ondes nécessite la simulation du rayonnement émis par la surface et l'atmosphère, et reçu par un radiomètre à bord d'un satellite, nommé *température de brillance*. Les fréquences micro-ondes étant très sensibles aux différents hydrométéores, un point particulièrement important concerne la spécification, au sein de l'assimilation de données, des propriétés de diffusion simple (SSP, 'single scattering property') des hydrométéores. Ces SSPs permettent de décrire la diffusion d'un rayonnement lors de son interaction avec un hydrométéore. La description des hydrométéores liquides est plus facile que celle des hydrométéores solides de par leur forme, généralement représentée par une sphère ou un sphéroïde aplati (Brandes et al. 2002[27], Augros et al. 2016[28], Ekelund et al. 2020[29]). Le calcul des SSPs est assez simple pour une telle forme de particule, grâce à des formules analytiques proposées par G. Mie en 1908 pour une sphère et grâce à la méthode de la T-matrice pour les sphéroïdes (Waterman, 1971[30]). Les hydrométéores solides, au contraire, ont des formes très diverses au sein de l'atmosphère, pouvant varier entre des structures étoilées, des colonnes, des plaquettes ou des agrégats constitués d'ensembles de formes simples (Bailey et Hallett, 2009[31]). Les hautes fréquences micro-ondes sont particulièrement sensibles aux hydrométéores solides (car la longueur d'onde du rayonnement est du même ordre de grandeur que la taille des particules) et nécessitent donc une spécification adéquate des SSPs lors des simulations de températures de brillance.

Les méthodes de calcul des SSPs considèrent généralement une forme et une densité de particule pour chaque hydrométéore solide ainsi qu'une distribution en taille associée. Les SSPs associées à des particules de forme et de densité contrastées sont répertoriées dans de nombreuses bases de données (Liu

2008[11], Hong 2007[32], Petty et al. 2010[33], Tyynelä et al. 2009[34], Ori et al. 2014[35], Leinonen et al. 2015[36][37], Kuo et al. 2016[38], Johnson et al. 2018, Lu et al. 2016[39], Ding et al. 2017[40], Eriksson et al. 2018[41]). Le nombre de bases disponibles augmente de manière significative depuis plusieurs années (Kneifel et al. 2018[42]).

Organisation du mémoire de thèse

Dans cette thèse, nous proposons de réfléchir à de nouvelles méthodes pour prendre en compte la diversité des SSPs au sein des systèmes d'assimilation de données utilisant les observations micro-ondes satellitaires en ciel nuageux. L'objectif principal de nos travaux sera de déterminer si :

- une combinaison optimale de SSPs caractérisant les hydrométéores solides peut être définie pour chaque scène nuageuse et/ou pluvieuse à partir des seules observations micro-ondes.
- une caractérisation des incertitudes liées aux SSPs des hydrométéores solides est pertinente et exploitable pour l'assimilation de données.

Pour répondre à ces problématiques, nous allons premièrement dans le chapitre 1 présenter les notions fondamentales de la PNT. Les modèles de PNT opérationnels à Météo-France, ARPEGE (modèle global) et AROME (modèles à aire limitée), seront introduits. Les principales méthodes d'assimilation de données exploitées ces dernières années pour initialiser ARPEGE et AROME seront également décrites dans ce chapitre. Le chapitre 2 introduira les notions de transfert radiatif essentielles à la compréhension de la simulation des observations micro-ondes en zones nuageuses. Différentes méthodes existantes pour les calculs de SSPs seront détaillées et nous présenterons les principales bases de données disponibles à ce jour. Le chapitre 3 abordera les différentes méthodes permettant l'assimilation d'observations satellitaires micro-ondes nuageuses ainsi que leurs avantages et inconvénients. Le chapitre 4 exposera les différentes études amorcées pour traiter la diversité des SSPs et détaillera de manière plus précise la problématique de cette thèse. Les trois chapitres suivants rassembleront les principaux résultats des études entreprises au cours de ce travail. Ces études seront toutes réalisées à l'aide des observations du radiomètre GMI (GPM Microwave Imager), du code de transfert radiatif RTTOV-SCATT (J.Eyre, 1991[43]) et de la base de données de SSPs de Liu (2008[11]). Au chapitre 5, une étude de sensibilité à la méthode d'assimilation indirecte reposant sur le formalisme Bayésien sera réalisée. Un aspect original sera d'étendre les travaux entrepris pour l'assimilation des observations satellitaires micro-ondes en ciel nuageux aux basses fréquences et d'optimiser la méthode. En effet, les premières études réalisées à Météo-France avaient considéré des observations du sondeur micro-ondes SAPHIR (Sondeur Atmosphérique du Profil d'Humidité Intertropicale par Radiométrie) possédant 6 fréquences toutes centrées à 183 GHz (Guerbette et al. 2016[4], Duruisseau et al. 2019[44]). Une adaptation originale de la méthode Bayésienne permettant de simuler les observations micro-ondes à l'aide de combinaison de SSPs sera alors présentée et étudiée au sein du chapitre 6. Cette étude sera réalisée successivement à l'aide du modèle à aire limitée AROME-Antilles puis du modèle global ARPEGE. Elle nous permettra de répondre à la première problématique abordée dans ce manuscrit. Les impacts d'une approche combinant plusieurs SSPs pour l'assimilation de données micro-ondes en zones nuageuses sur les scores de prévisions jusqu'à 4 jours d'échéance du modèle ARPEGE seront ici présentés. Nous traiterons la deuxième partie de la problématique au sein du chapitre 7 où une étude réalisée à l'aide du modèle ARPEGE et d'un système d'ensemble d'assimilations nommé AEARP (Assimilation d'Ensemble ARPEGE) sera présentée. Cette étude consistera à évaluer l'impact des incertitudes liées à la spécification des SSPs des hydrométéores solides sur l'estimation de la matrice de covariances d'erreurs des prévisions à court terme calculée par l'AEARP. Pour cela chaque membre de l'ensemble se verra attribuer des SSPs sélectionnées de manière aléatoire au sein de la base de données de Liu (2008[11]). La matrice de covariances d'erreurs obtenue sera alors utilisée au sein d'une assimilation directe d'observations micro-ondes. Pour finir, les conclusions de ces travaux de thèse et les perspectives qu'ils offrent seront proposées dans le dernier chapitre.

Chapitre 1

La prévision numérique du temps

Sommaire

1.1 Les modèles numériques	7
1.1.1 Le modèle ARPEGE	8
1.1.2 Le modèle AROME	10
1.2 Les observations	14
1.2.1 Les observations depuis la Terre	14
1.2.2 Les observations depuis l'espace	16
1.3 L'assimilation de données	18
1.3.1 L'Interpolation Optimale	18
1.3.2 Les méthodes variationnelles	22
Le 3D-Var	22
Le 4D-Var	24
L'assimilation d'ensemble AEARP	26

La connaissance de la situation météorologique à des échelles temporelles variant de quelques heures à plusieurs semaines est nécessaire dans de nombreux secteurs économiques (transports, agriculture, énergie, tourisme, ...) et est indispensable pour assurer la sécurité des personnes et des biens. La prévision du temps repose principalement sur la modélisation numérique de l'évolution de l'atmosphère à partir d'une discrétisation spatio-temporelle des équations de la mécanique des fluides et de conditions initiales les plus précises possibles. C'est ce que l'on nomme la Prévision Numérique du Temps (PNT), imaginée dès 1922 par le météorologiste britannique L. F. Richardson, bien avant l'avènement des premiers ordinateurs qui rendirent ce rêve possible dans les années 1950.

Afin de fournir des prévisions, l'évolution de l'atmosphère est décrite par des modèles de PNT préalablement initialisés. Les conditions initiales de ces modèles sont obtenues en combinant de manière optimale au sein d'un algorithme numérique, appelé 'assimilation de données', les observations de l'atmosphère disponibles avec une information a priori provenant d'une prévision numérique à courte échéance (appelée ébauche). L'état ainsi obtenu est la meilleure connaissance de l'atmosphère (en termes de vents, température, humidité et pression) à un instant donné et est appelé 'analyse'.

Dans ce chapitre nous allons présenter les deux modèles numériques développés et utilisés opérationnellement à Météo-France. Nous introduirons ensuite les différents types d'observations disponibles pour initialiser ces modèles. Finalement, après avoir décrit les principes généraux de l'assimilation de données, les algorithmes utilisés à Météo-France seront résumés.

1.1 Les modèles numériques

Les méthodes numériques utilisées dans les modèles de PNT nécessitent une discrétisation tridimensionnelle de l'atmosphère ainsi qu'une discrétisation temporelle. La discrétisation en espace consiste à

découper l’atmosphère en volumes finis sur la verticale et l’horizontale afin d’obtenir un maillage 3D de l’atmosphère. La discrétisation en temps consiste à fixer un intervalle de temps séparant deux estimations successives de l’état de l’atmosphère. On distingue deux grands types de modèles de PNT :

- Les modèles globaux, pour lesquels la totalité du globe est discrétisée,
- Les modèles à aire limitée où seulement une partie du globe est discrétisée.

Ils présentent chacun des avantages spécifiques. Les modèles à aire limitée sont généralement caractérisés par une discrétisation très fine, par exemple de l’échelle kilométrique. Grâce à cette discrétisation, ils sont capables de représenter explicitement des phénomènes météorologiques de petite échelle, telle que la convection profonde. Les modèles globaux sont quant à eux caractérisés par une discrétisation moins fine mais couvrant la totalité du globe. Cette couverture globale permet de modéliser des phénomènes météorologiques se déplaçant sur de longues trajectoires, comme les tempêtes et les cyclones, de leur formation à leur fin de vie. Ces modèles doivent également être efficaces numériquement, car leurs prévisions doivent être disponibles très rapidement. Ainsi, la configuration d’un modèle de PNT résulte d’un compromis fort entre emprise géographique, discrétisation spatio-temporelle, et degré de réalisme des différents composants.

1.1.1 Le modèle ARPEGE

Depuis 1992, le modèle numérique global de prévision opérationnel à Météo-France est le modèle Action de Recherche Petite Échelle Grande Échelle (ARPEGE, Courtier et al. 1991)[45]. Il s’agit d’un modèle spectral¹ sur l’horizontale issu d’une collaboration entre Météo-France et le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT ou en anglais ECMWF pour European Centre for Medium-Range Weather Forecasts) initiée en 1987. Ce code commun permet à la fois de réaliser les prévisions du modèle ARPEGE, du modèle IFS (Integrated Forecasting System) du CEPMMT et également du modèle à aire limitée AROME comme nous le verrons dans la section suivante.

La version opérationnelle d’ARPEGE en 2022 possède une grille verticale composée de 105 niveaux, allant de 10 m au dessus de la surface jusqu’à environ 70 km d’altitude. Il produit des prévisions jusqu’à 102 h d’échéance, effectuées toutes les 6 heures : 0 h, 6 h, 12 h et 18 h UTC. Sa résolution, de 5 km sur la France et de 25 km aux antipodes, comme le montre la figure 1.1, lui permet de fournir des prévisions plus précises sur la métropole par rapport à un autre modèle à la même troncature spectrale sans grille étirée. Sa résolution actuelle traitant des phénomènes à l’échelle synoptique, il prévoit des phénomènes tels que les dépressions et les anticyclones des moyennes latitudes. Actuellement ce modèle repose sur l’hypothèse hydrostatique qui suppose un équilibre entre la force de gravité et le gradient de pression vertical, conduisant ainsi à une estimation diagnostique de la vitesse verticale.

Chaque modèle numérique possède un certain nombre de variables dites pronostiques que l’on fait évoluer dans le temps grâce à des schémas numériques. Les variables pronostiques pour la partie atmosphérique du modèle ARPEGE sont les suivantes :

- les composantes horizontales du vent \vec{u} et \vec{v} ,
- la température T ,
- la pression de surface P ,
- l’énergie cinétique turbulente TKE ,
- l’humidité spécifique Q ,
- le contenu spécifique de quatre hydrométéores : les gouttelettes d’eau nuageuse Q_l , les cristaux de glace nuageuse Q_i , la pluie stratiforme Q_r et la neige stratiforme Q_s .

Un certain nombre de processus physiques d’échelle inférieure à la maille mais impactant les échelles résolues par le modèle sont décrits au moyen de paramétrisations physiques. Ces processus sont :

1. décomposition des champs météorologiques en séries limitées d’harmoniques sphériques

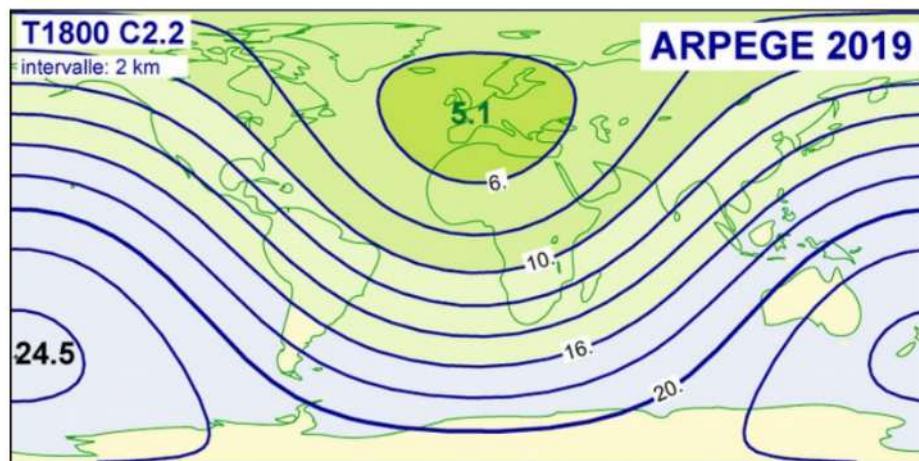


FIGURE 1.1 – Résolution horizontale de la grille basculée/étirée de la version du modèle ARPEGE opérationnelle depuis 2019.

- **le rayonnement** : les rayonnements de grandes longueurs d’ondes, représentant la composante infrarouge du rayonnement, exploitent le schéma de transfert radiatif RRTM (pour ‘Rapid and accurate Radiative Transfer Model’ en anglais) défini dans Mlawer et al. 1997[46]. Les rayonnements de courtes longueurs d’ondes, représentant la composante du rayonnement solaire ainsi qu’une partie du rayonnement infrarouge, sont quant à eux définis dans Fouquart et Bonnel 1980[47].
- **les échanges avec les surfaces continentales et océaniques** : ces échanges d’énergie, d’eau et de quantité de mouvement avec l’atmosphère sont représentés avec le modèle SURFEX (SURFace EXternalisée), défini dans Masson et al. 2013[48].
- **les transferts turbulents** : ils sont définis dans Cuxart et al. 2000[49] au moyen d’une équation de diffusion et de coefficients dépendant de l’évolution de l’énergie cinétique turbulente (TKE pour ‘turbulent kinetic energy’ en anglais). Ici la longueur de mélange utilisée est non locale et basée sur la paramétrisation décrite dans Bougeault et Lacarrere (1989[50]).
- **les effets de la convection** : les effets des convections profonde (précipitante) et peu profonde (non précipitante) nécessitent d’être définis au sein des modèles. Pour le modèle ARPEGE, la convection profonde est défini grâce au schéma de Tiedtke–Bechtold (Tiedtke 1989[51], Bechtold et al. 2008[52], Bechtold et al. 2014[53]) et la convection peu profonde grâce au schéma de Bechtold (2001[54]). Dans la mesure où les nuages ne sont pas explicitement résolus, la représentation des processus microphysiques (changements de phases de l’eau et évolution des hydrométéores) est, dans ces schémas, assez empirique et définie comme la moyenne des processus au sein de la maille.
- **l’effet des ondes de gravité orographiques** : est décrit dans Catry et al. 2008[55].
- **Les nuages et les précipitations d’échelles résolues** : le schéma micro-physique pronostique pour décrire l’évolution des espèces nuageuses et précipitantes est celui de Lopez et al 2002[56]. Le pas de temps du modèle ARPEGE de 240 secondes nécessite la prise en compte la sédimentation des précipitations de manière statistique décrite dans Bouteloup et al. 2011[57].

Les principales caractéristiques du modèle ARPEGE opérationnel en 2022 sont présentées dans le tableau 1.1.

ARPEGE	
Type de modèle	global
Equations	équations primitives
Hydrostatique	oui
Niveaux verticaux	105 niveaux de 10 m au dessus de la surface à un niveau supérieur à 70 km
Résolution	5 km sur la France jusqu'à 24 km aux antipodes
Heures d'initialisation des prévisions	0 h, 6 h, 12 h et 18 h UTC
Échéance	jusqu'à 102 h
Pas de temps	240 s
Schéma temporel	Semi-Lagrangien à 2 niveaux Semi-implicite Ritchie et al. (1995[58])
Représentation sur la verticale	en éléments finis
Schéma microphysique	Lopez (2002[56])
Convection profonde	schéma de Tiedtke–Bechtold (Tiedtke 1989[51], Bechtold et al. 2008[52], Bechtold et al. 2014[53])

TABLE 1.1 – Caractéristiques du modèle ARPEGE

1.1.2 Le modèle AROME

Le modèle Application de la Recherche à l'Opérationnel à Méso-Echelle (AROME) est une version à aire limitée utilisée opérationnellement en France métropolitaine depuis 2008 (Seity et al. 2011[59]) et en Outre-Mer depuis 2016 (Faure et al. 2020[9]). Actuellement sa grille verticale est composée de 90 niveaux s'étalant de 5 m au dessus de la surface jusqu'à environ 34 km d'altitude. Cette discrétisation conduit à une meilleure résolution que le modèle ARPEGE dans la troposphère (notamment près de la surface). Sa résolution horizontale de 1.3 km permet quant à elle de représenter explicitement certains phénomènes convectifs comme les orages, contribuant ainsi à la mission de Météo-France de protection des personnes et des biens.

Les conditions initiales du modèle AROME-France proviennent d'un système d'assimilation décrit plus loin et ce modèle est forcé aux limites latérales par des prévisions horaires du modèle global ARPEGE. Il réalise des prévisions à 0 h, 3 h, 6 h, 12 h et 18 h UTC jusqu'à 48 heures d'échéance. Les différentes caractéristiques de ce modèle sont listées dans le tableau 1.2. Pour les modèles AROME-Outre-mer, les conditions initiales sont fournies par une analyse du modèle global IFS du CEPMMT et les limites latérales sont forcées par des prévisions de ce même modèle.

AROME

Type de modèle	à aire limitée
Equations	équations d'Euler compressibles
hydrostatique	non
Niveaux verticaux	90 niveaux de 5 m au dessus de la surface à environ 34 km
Résolution	1.3 km
Heures d'initialisation des prévisions	0 h, 3 h, 6 h, 12 h et 18 h UTC
Échéance	jusqu'à 48 h
Pas de temps	50s
schéma temporel	Semi-Lagrangien à 2 niveaux Semi-implicite
Représentation sur la verticale	en différences finies
schéma microphysique	ICE3
Convection profonde	Résolue explicitement

TABLE 1.2 – Caractéristiques du modèle AROME-France

Le modèle AROME possède les mêmes variables pronostiques que le modèle ARPEGE plus les suivantes :

- la vitesse verticale,
- l'écart à la pression hydrostatique,
- et le contenu spécifique d'une cinquième catégorie d'hydrométéore : le grésil² Q_g .

Pour le modèle AROME, comme pour le modèle ARPEGE, certains processus physiques restent inférieurs à la taille de la maille spatiale, leur impact doit donc être paramétré. Ces processus sont :

- **le rayonnement** : les rayonnements solaire et infrarouge sont décrits de la même manière que dans le modèle ARPEGE.
- **les processus de surface** : les échanges énergétiques et hydriques avec les surfaces (continentales et océaniques) : les échanges énergétiques et hydriques avec les surfaces sont décrits de la même manière que dans le modèle ARPEGE.
- **les transferts turbulents** : les transferts turbulents sont décrits de la même manière que dans le modèle ARPEGE

- **les effets de la convection** : contrairement au modèle ARPEGE, le modèle AROME résout explicitement la convection profonde. La convection peu profonde est quant à elle définie à l'aide du schéma de Pergaud et al. (2009[60]).

- **Les nuages et les précipitations** : Les processus microphysiques permettant de représenter les conversions des espèces condensées (hydrométéores) entre elles et avec la vapeur d'eau utilisent le schéma ICE3 (Pinty et Jabouille, 1998[61]) récemment amélioré par Riette (2020[62]) afin de réduire la dépendance au pas de temps du modèle. Les processus microphysiques pris en compte dans ICE3 sont représentés sur la figure 1.2. Parmi ces processus, on retrouve deux catégories : les processus froids et les processus chauds respectivement en bleu et en rouge.

Pour les processus froids traitant la phase solide, les hydrométéores sont paramétrés à l'aide du schéma de Caniaux (1993[63]). Les hydrométéores solides sont représentés par trois variables pronostiques : les cristaux de glace, la neige et le grésil. Pour les processus chauds, traitant la phase liquide, les hydrométéores sont représentés à l'aide du schéma de Kessler (1969[64]). Les hydrométéores liquides sont représentés par deux variables pronostiques : le contenu en gouttelettes nuageuses et le contenu en gouttes de pluie. Chaque espèce condensée possède une seule variable pronostique : sa concentration massique (schéma microphysique à un moment). Comme ceci, la distribution en taille des hydrométéores dépend seulement de la concentration massique. Pour chaque type d'hydrométéores j , on définit la distribution en taille N_j à l'aide d'une loi Gamma généralisée Γ multipliée par la densité volumique de la particule N_{oj} (m^{-3}) telle que :

$$N_j(D) = N_{oj} \frac{\alpha_j}{\Gamma(\mu_j)} \lambda_j^{\alpha_j \mu_j} D^{\alpha_j \mu_j - 1} \exp[-(\lambda_j D)^{\alpha_j}]$$

Avec D la dimension maximale de la particule, μ_j et α_j des paramètres de la forme de la distribution pour l'hydrométéore j décrivant des lois de puissance provenant de Locatelli et Hobbs (1974[65]), λ_j le paramètre de pente pour l'hydrométéore j . La densité volumique N_{oj} de chaque hydrométéore est fonction d'une puissance de λ_j :

$$N_{oj} = C_j \lambda_j^{X_j}$$

Avec C_j qui vaut ici m^{X_j-3} . Pour les particules de glace et les gouttelettes de pluie, certaines distribution en taille des particules, ou Particle Size Distribution (PSD), sont décrites à l'aide du diamètre de sphère équivalent (Petty et Huang, 2011[66]). Ce diamètre s'exprime tel que :

$$D_{veq} = \left(\frac{6m_j}{\pi \rho_i} \right)^{1/3}$$

2. petites particules de neige avec une croûte fragile de glace

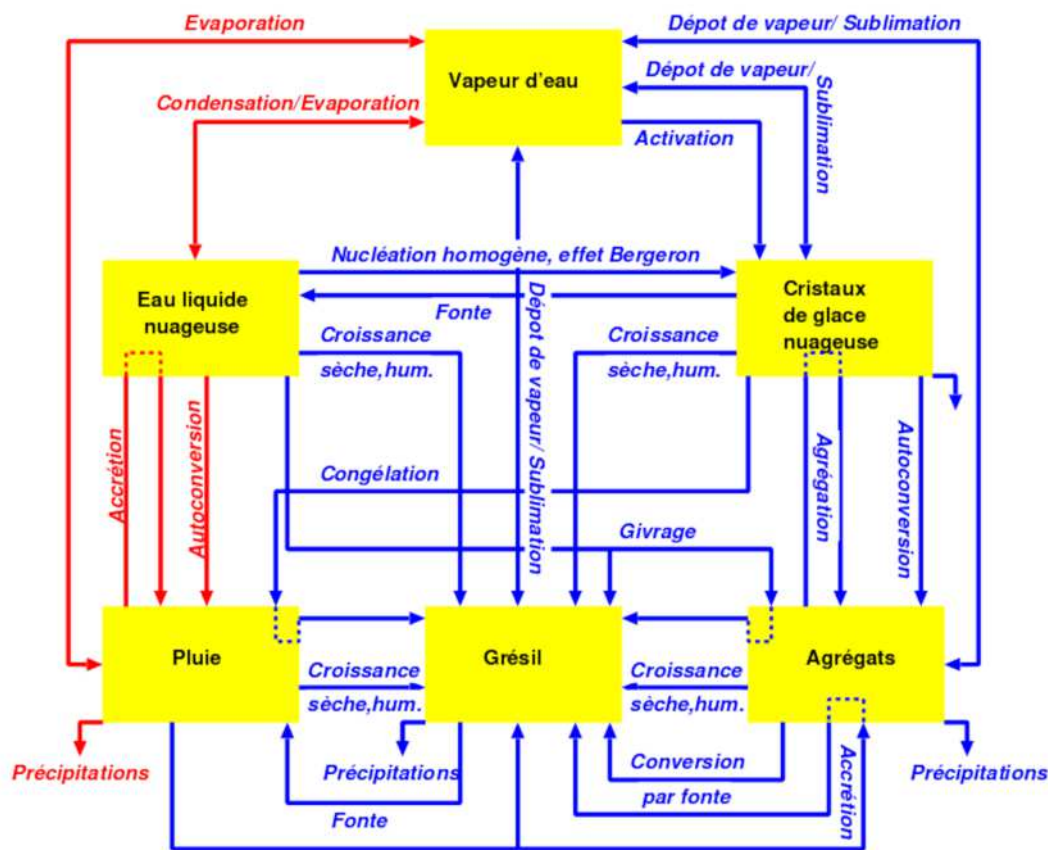


FIGURE 1.2 – Processus microphysiques considérés dans le schéma ICE3. Les processus froids sont représentés en bleus et les processus chauds en rouge.

Où m_j est la masse de l'hydrométéore j et ρ_i est la densité de la glace (ou de l'eau liquide). Il est possible de retrouver la dimension maximale D_{max} d'une particule à partir de D_{veq} par la formule :

$$D_{max} = \left(\frac{\pi \rho_i D_{veq}^3}{6a_j} \right)^{1/b_j}$$

où $m_j(D)$ vaut $a_j D^{b_j}$ avec a_j et b_j des paramètres liés à la densité et à la forme de l'hydrométéore j . La comparaison de D_{max} et D_{veq} peut être trouvée dans Eriksson et al. 2015[67]. Le contenu en hydrométéore M_j , s'exprimant en $kg.m^{-3}$, est calculé en intégrant la distribution en tailles (ou en anglais PSD pour Particle Size Distribution) multipliée par sa masse m_j :

$$M_j = \int_0^{\text{inf}} m_j(D) N_j(D) dD$$

Les valeurs des différents paramètres liés aux hydrométéores j définis dans ICE3 sont répertoriées au Tableau 1.3.

La documentation du modèle à mésoéchelle Méso-NH, développé conjointement par le Laboratoire d'Aérodologie et le CNRM (Lafore et al. 1998[68]), décrit de manière plus approfondie le schéma microphysique ICE3.

Météo-France réalise de plus des prévisions pour les régions Outre-Mer dans 5 zones géographiques (les Antilles, la Guyane, le sud-ouest de l'océan Indien, la Nouvelle Calédonie, la Polynésie) depuis 2016

Type	α_j	μ_j	C_j (m^{X_j-3})	X_j	a_j ($kg.m^{b_j}$)	b_j
Pluie	1	1	8.10^6	-1	524	3
Glace primaire	3	3	C (diagnostique)	0	0.82	2.5
Graupel	1	1	5.10^5	-0.5	19.6	2.8
Neige	1	1	5	1	0.02	1.9
Eau liquide nuageuse sur terre	1	3	3.10^8	0	524	3
Eau liquide nuageuse sur mer	3	1	1.10^8	0	$\frac{\pi \rho_w}{6} = 524$	3

TABLE 1.3 – Caractéristiques liées aux hydrométéores définis dans le modèle ICE3. Ici p_w représente la densité de l'eau.

(Faure et al. 2020)[9]. Ces modèles appelés par la suite AROME-OM sont représentés sur la figure 1.3. Compte-tenu de leurs situations géographiques, comme notamment le bassin de l'Atlantique Nord avec les Antilles et le bassin du Sud-Ouest de l'océan Indien avec l'île de la Réunion et Mayotte, plusieurs de ces domaines sont sujets à des événements convectifs et connaissent des épisodes cycloniques. Ainsi les prévisions de ces modèles sont importantes pour anticiper ces événements météorologiques violents.

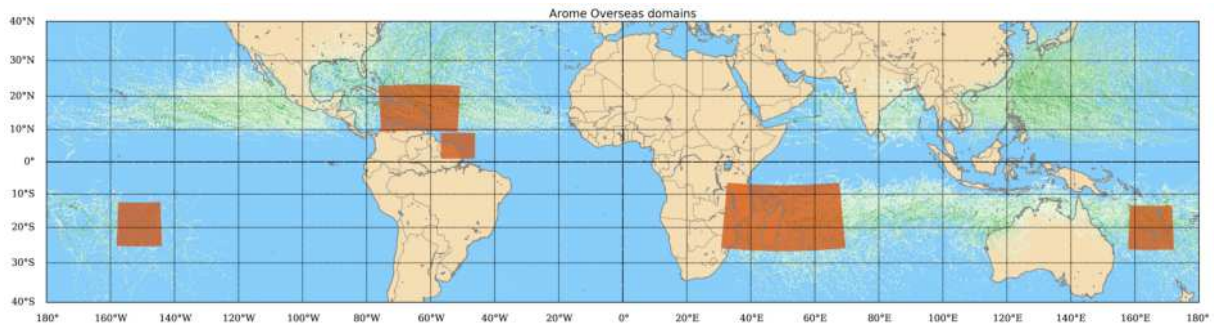


FIGURE 1.3 – Carte des domaines des modèles AROME-OM (boîtes oranges) et climatologie entre 1980 et 2018 des trajectoires de cyclones (traits verts)

Source: Faure et al. 2020[9]

1.2 Les observations

Pour définir l'état initial des modèles numériques permettant de démarrer une prévision, les systèmes d'assimilation utilisent des mesures plus ou moins directes de paramètres météorologiques (vent, température, humidité, pression) appelées observations. Ces observations permettent de caractériser l'état atmosphérique à un instant donné, et sont donc indispensables pour obtenir des prévisions numériques de qualité. Parmi ces observations, on retrouve deux grandes catégories : les observations effectuées depuis la Terre et les observations effectuées depuis l'espace. Ces deux types d'observations échantillonnent l'atmosphère de manière très différentes comme on peut déjà le voir sur la figure 1.4, nous allons le voir dans les paragraphes suivants.

1.2.1 Les observations depuis la Terre

Les observations effectuées depuis la Terre comprennent les observations in-situ (dites conventionnelles) et les observations effectuées par télédétection terrestre. Les observations dites conventionnelles sont obtenues grâce :

- aux ballons sondes,

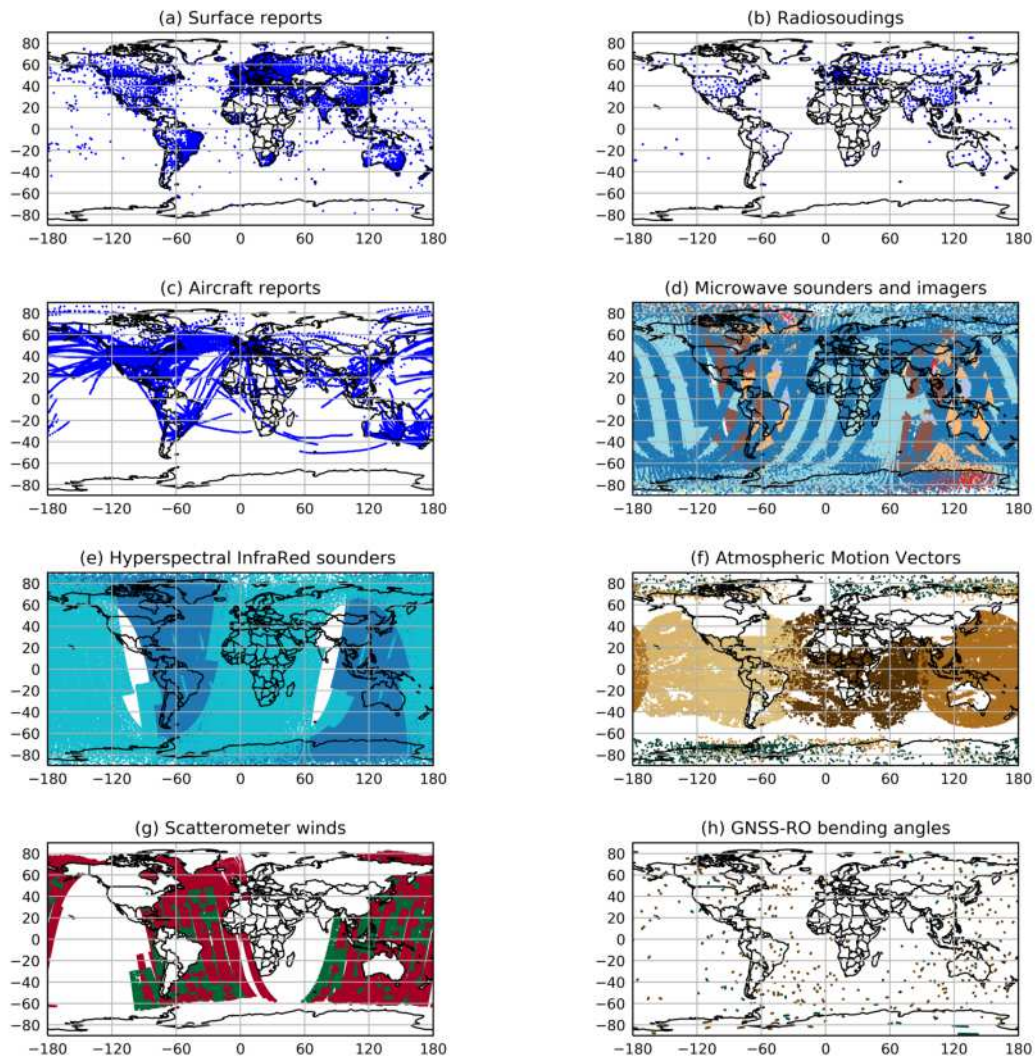


FIGURE 1.4 – Densité du réseau d’observations (a) en surface, (b) mesurées par radiosondages, (c) par les avions, (d) par les imageurs et sondeurs micro-ondes, (e) par les sondeurs infrarouges hyperspectraux, (f) par les vecteurs de mouvement atmosphériques déduits des imageurs infrarouges, (g) de vents mesurés par les diffusionomètres et (h) et d’angles de courbures mesurés par radio occultation : Global Navigation Satellite System (GNSS)-based radio occultation (RO)

Source: Chambon et al. 2022[69]

- aux avions,
- aux bateaux,
- aux bouées,
- aux stations terrestres ...

Et les observations mesurées par télédétection terrestre sont obtenues grâce :

- aux lidars,
- aux radars.

La pandémie en 2020 causa une importante baisse des observations conventionnelles mesurées par les avions pour le vent et la température. Pour tenter de diminuer cet effet, certains centres de Météo-France ont augmenté leurs lâchers de ballon-sondes illustré sur la figure 1.5 permettant la mesure de profils de température, pression, vent et humidité.

Pour un modèle global, le plus grand nombre de ces observations est fourni par les avions à cause de la faible densité des réseaux d'observations terrestres dans certaines régions (zones tropicales et dans l'hémisphère sud notamment). Ainsi pour le modèle ARPEGE, qui utilise un peu moins de 10^8 de ces observations chaque mois, 4×10^7 proviennent des mesures d'avions. Le modèle AROME-France utilise chaque mois un peu plus de 4×10^7 observations in situ et mesurées par télédétection terrestre pour construire ses analyses. La plus grande partie provient des radars du réseau ARAMIS (Application Radar À la Météorologie Infra-Synoptique, Tabary 2007[70]) mesurant la réflectivité et le vent radial avec environ 2×10^7 observations par mois.



FIGURE 1.5 – Exemple de ballon-sonde, lancé automatiquement par Météo-France. Les différents instruments mesurant les composantes de l'atmosphère sont situés dans la nacelle à l'extrémité inférieure du dispositif.

Source: <http://www.meteomodem.com>

1.2.2 Les observations depuis l'espace

La deuxième catégorie d'observations correspond aux observations mesurées depuis l'espace, ou "satellites". Elles proviennent de deux types de satellites. Les premiers qui sont les satellites dits 'défilants' sont placés sur une orbite basse entre 400 et 850 km d'altitude; ils ont pour avantage de réaliser un tour de la Terre rapidement, en environ 2 heures. Ils sont, pour la plupart, placés sur une orbite dite 'héliosynchrone' qui permet d'observer une zone donnée de la Terre à la même heure solaire à chaque passage.

Le deuxième grand type correspond aux satellites dits 'géostationnaires'. Ils se situent à environ 36 000 km d'altitude et voient en permanence la même partie du globe terrestre, ce qui leur permet d'avoir une forte répétitivité temporelle. Par exemple, le satellite géostationnaire Météosat-Second Generation 4 (MSG-4) observe un disque couvrant l'Afrique et l'Europe de l'Ouest toutes les 15 minutes.

Deux grandes catégories d'instruments sont embarqués à bord de ces satellites : les instruments actifs et les instruments passifs. Les instruments passifs détectent les rayonnements émis ou réfléchis par l'atmosphère et la surface terrestre, et opèrent dans des gammes de longueurs d'ondes allant du visible aux micro-ondes en passant par le spectre infrarouge. Parmi ces instruments on trouve des radiomètres, ou encore des interféromètres. Les instruments actifs, contrairement aux précédents, envoient un rayonnement électromagnétique et mesurent le rayonnement réfléchi associé. Les instruments actifs opérant dans le visible et l'infra-rouge sont des lidars, alors que dans le domaine des micro-ondes ce sont des radars (les radars utilisés pour caractériser les surfaces sont des altimètres ou des diffusiomètres). Les satellites géostationnaires embarquent uniquement des instruments passifs, opérant dans le visible et l'infrarouge. En effet, les instruments micro-ondes ne peuvent pour l'instant être emportés à bord de ce type de satellite. Afin d'obtenir une résolution des pixels raisonnable, il serait nécessaire de construire une antenne de taille conséquente qui n'est pour l'instant pas réalisable techniquement. En revanche les satellites défilants embarquent des instruments passifs et actifs, opérant dans le visible, l'infrarouge et les micro-ondes.

Contrairement aux observations conventionnelles, les observations satellitaires ne mesurent pas directement des grandeurs physiques décrites dans les modèles de PNT telles que la température ou l'humidité. Pour pouvoir les assimiler, il faut donc relier le rayonnement électromagnétique dans un certain domaine de longueur d'onde (visible, l'infrarouge et les micro-ondes) avec les variables météorologiques d'un modèle de PNT. Les spectres utilisés possèdent des bandes d'absorption correspondant à des intervalles de longueur d'onde pour lesquels certaines molécules absorbent le rayonnement. Ainsi, selon les spectres d'absorption de certains gaz atmosphériques (oxygène, vapeur d'eau, gaz carbonique, ozone), il est alors possible d'extraire des informations sur la verticale pour plusieurs grandeurs physiques dont l'humidité et la température.

En 2022, le modèle AROME utilise environ 5×10^6 observations satellitaires par mois dont les principales sont fournies par le sondeur infra-rouge IASI (Infrared Atmospheric Sounding Interferometer, environ 1/4) et par le satellite géostationnaire Meteosat observant dans l'infrarouge (environ 1/4 également), et ce malgré que 80% des observations d'AROME proviennent des mesures depuis la surface, principalement de radars sols (figure 1.6). Le modèle ARPEGE, quant à lui, utilise un plus grand nombre d'observations satellitaires que le modèle AROME grâce à son emprise géographique globale. Ainsi, il assimile environ 10^9 observations satellitaires par mois dont la majorité sont aussi des observations infrarouges, avec notamment la moitié des observations totales provenant des seuls instruments IASI à bord des satellites Metop-B et Metop-C. Le nombre imposant d'observations infrarouges par rapport à celles dans les micro-ondes s'explique par la meilleure résolution spectrale des sondeurs dits 'hyperspectraux' comme IASI et CrIS (Cross-track Infrared Sounder) permettant de résoudre le spectre infrarouge de manière continue entre 4 et 15 μm au moyen de plusieurs milliers de canaux fournissant ainsi un nombre très élevé d'observations. Ces observations sont toutefois très sensibles à la présence de nuages et nécessitent des contrôles de qualité importants pour l'identification des scènes claires. La différence du nombre d'observations assimilées entre les modèles AROME et ARPEGE s'explique par le nombre relatif d'observations satellitaires disponibles au dessus de l'Europe de l'Ouest en comparaison de nombre d'autres types d'observations.

Les observations micro-ondes sont devenues très importantes au cours des dernières années dans les systèmes de PNT. Cette importance grandissante est principalement reliée à deux facteurs : (i) la mise en orbite progressive d'une constellation de satellites embarquant à leur bord des sondeurs et imageurs micro-ondes comme on peut le voir sur la frise temporelle de la Figure 1.7, (ii) l'amélioration de l'assimilation de ces observations, par exemple en les utilisant sur de plus en plus de types de surfaces (Karbou et al. 2006[71]).

Les impacts FSOi (Forecast Sensitivity to Observations impacts), qui sont une mesure des réductions d'erreurs de prévision à 24 h d'échéance par système d'observations (e.g. Langland and Baker, 2004[73]) à partir d'une norme énergie, montrent de très bons résultats pour ce type d'observations : à l'automne 2019 ce sont les radiances micro-ondes qui ont eues le plus d'impact dans le modèle global (Unified Model) du centre PNT du Royaume-Uni, le Met Office, avec une valeur relative de 28% (Candy et al. 2021[74]). Pour celui du CEPMMT (Centre Européen de Prévision Météorologique à Moyen Terme) en 2020, les prévisions étaient aussi plus sensibles aux radiances micro-ondes qu'aux autres observations avec une valeur FSOi de 35% (Ingleby 2022[75]). Enfin, pour Météo-France, on peut voir sur la figure 1.8 montrant la valeur FSOi pour chaque type d'observations assimilées par le modèle global ARPEGE en 2020 et 2021, que les radiances micro-ondes sont également les observations ayant eu le plus d'impact avec une valeur FSOi oscillant entre 20 et 35%.

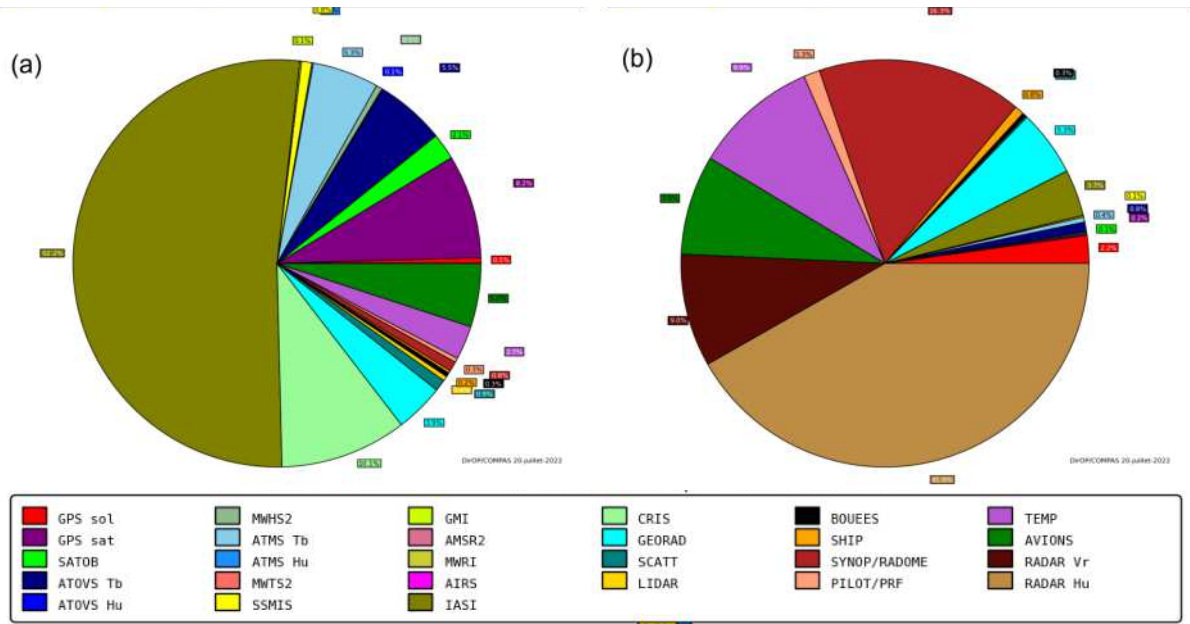


FIGURE 1.6 – Proportions de nombres d’observations disponibles et utilisées par type d’observations pour l’analyse pour les modèles (a) ARPEGE et (b) AROME pour le mois de juin 2022.

1.3 L’assimilation de données

Afin d’intégrer les informations provenant des observations aux prévisions, des méthodes combinant de manière optimale les observations y à une prévision antérieure x_b , aussi appelée *ébauche*, sont utilisées. Ce processus s’appelle l’assimilation de données et permet de générer un état atmosphérique x_a , appelé *analyse*, considéré comme l’état le plus représentatif de l’atmosphère ‘réelle’ x_{true} . Cette analyse est utilisée comme condition initiale dans les modèles de PNT pour estimer l’état futur de l’atmosphère.

1.3.1 L’Interpolation Optimale

En 1966, les premiers modèles de PNT basés sur les équations primitives apparaissent. Ces nouveaux modèles prennent alors en compte l’aspect multivarié de l’atmosphère et imposent aux conditions initiales de considérer les équilibres physiques comme le géostrophisme³ ou l’hydrostatisme⁴. Le filtrage des ondes de gravité, phénomènes non réalistes pouvant apparaître dans les équations primitives des modèles, est réalisé au moyen d’un algorithme spécifique nommé "initialisation" et appliqué après l’étape d’analyse. Toutes les méthodes sont alors basées sur l’estimateur mathématique appelé ‘BLUE’ pour Best Linear Unbiased Estimate.

La première méthode adoptée comme méthode d’assimilation de données dans les années 70 par la plupart des centres de PNT est la méthode de l’interpolation optimale (Gandin 1963[76]). Cette méthode repose sur une interpolation statistique cherchant une combinaison optimale entre des observations y et un état atmosphérique du modèle x_b comme information a priori (avant la connaissance des observations), ou ébauche, afin de trouver l’état atmosphérique analysé x_a . L’interpolation optimale est une simplification du BLUE consistant à ne considérer qu’un nombre limité d’observations pour les variables à analyser en chaque point de grille et exprimant l’analyse x_a comme une interpolation linéaire entre l’ébauche et les observations :

$$x_a = \mathbf{L}x_b + \mathbf{K}y$$

3. Équilibre théorique entre la force de Coriolis et la force de pression horizontale

4. État atteint par un système lorsque les forces de gravitation sont contrebalancées par un gradient de pression de direction opposée

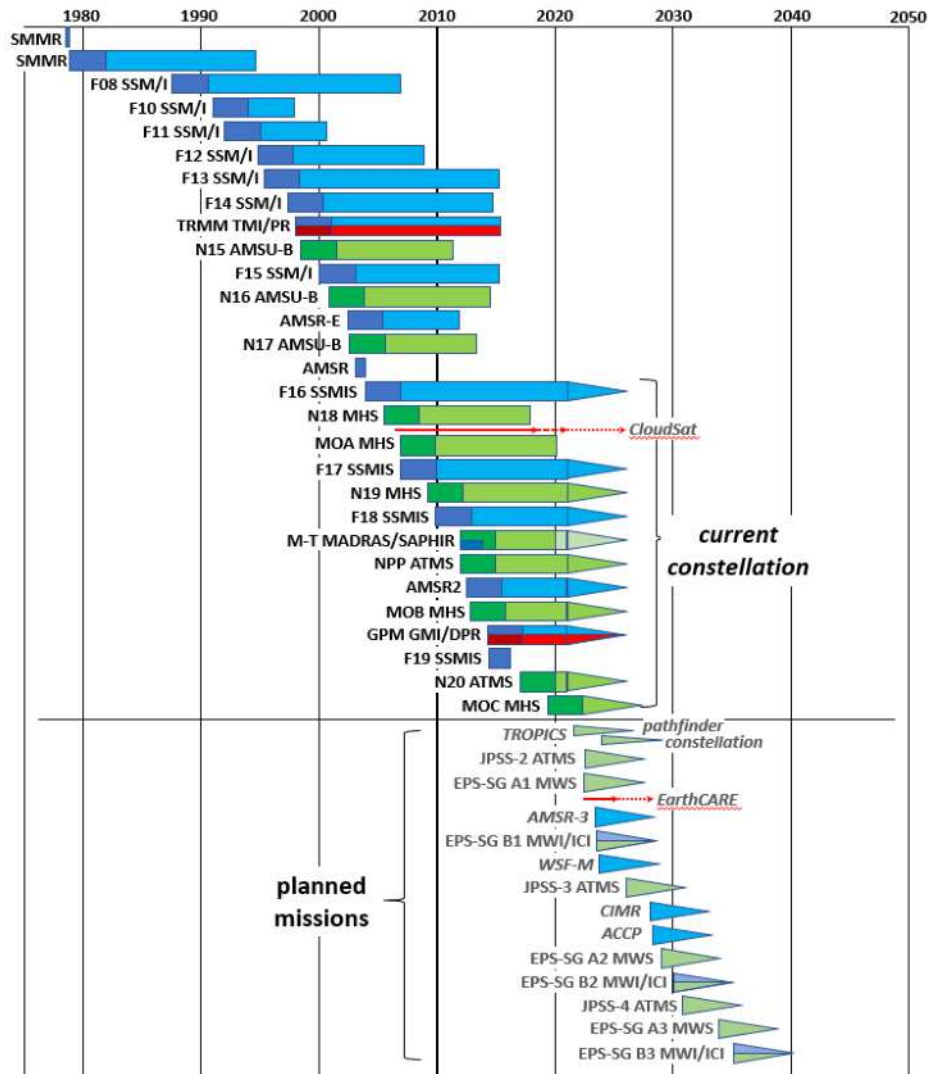


FIGURE 1.7 – Chronologie des satellites et capteurs micro-ondes fournissant des données prêtes à être analysées. Les barres et les flèches indiquent respectivement les observations mesurées à partir d’une fauchée et au nadir. Les barres bleues représentent les radiomètres d’imagerie/à balayage conique, les barres vertes représentent les radiomètres de sondage/cross-track et les barres rouges indiquent les capteurs actifs (radars). Les triangles indiquent les capteurs qui fournissent actuellement des données (au 2021-02-09) et peuvent continuer à le faire, ainsi que les missions futures. (Source des données : basée sur la base de données OSCAR (Observing Systems Capability Analysis and Review Tool) de l’Office météorologique mondial (OMM) et EUMETSAT (European organisation for the exploitation of METeorological SATellites))

Source: Kidd et al. 2021[72]

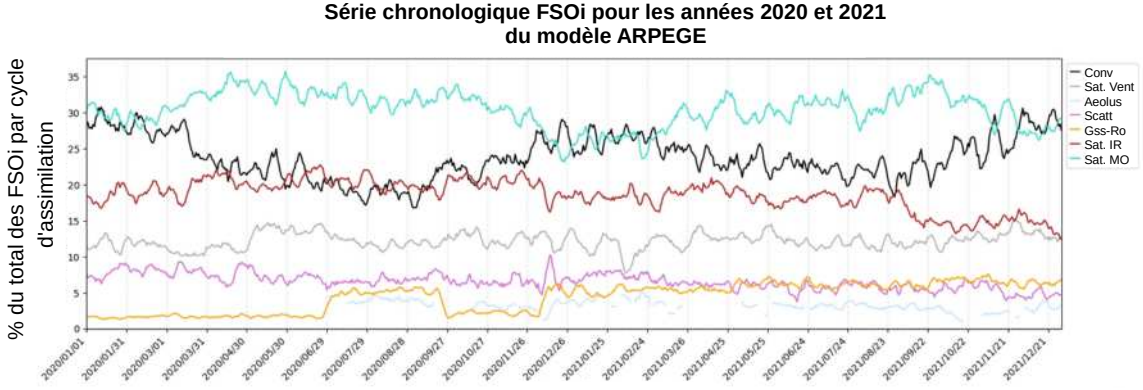


FIGURE 1.8 – Scores FSOi des observations assimilées par le modèle ARPEGE pour les années 2020 et 2021

Source: D. Raspaud, DESR/CNRM/GMAP/OBS

En assimilation de données on considère premièrement que les observations y s'expriment comme :

$$y = \mathbf{H}x_{true} + \epsilon_o$$

où \mathbf{H} est un opérateur d'observation linéaire, x_{true} l'état 'vrai' de l'atmosphère, et ϵ_o est une erreur non biaisée associée aux observations. On cherche à estimer les matrices \mathbf{L} et \mathbf{K} à l'aide des hypothèses H1⁵, H2⁶ et H3⁷ selon lesquelles les estimateurs sont non biaisés et optimaux pour minimiser la variance d'erreur d'analyse (on veut un état le plus précis possible d'un point de vue statistique). L'erreur d'analyse ϵ_a peut alors s'exprimer comme :

$$\begin{aligned} \epsilon_a &= x_a - x_{true} \\ &= \mathbf{L}x_b + \mathbf{K}y - x_{true} \\ &= \mathbf{L}(x_b - x_{true}) + \mathbf{L}x_{true} + \mathbf{K}(\mathbf{H}x_{true} + \epsilon_o) - x_{true} \\ &= \mathbf{L}\epsilon_b + \mathbf{K}\epsilon_o + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})x_{true} \end{aligned}$$

En s'aidant des hypothèses spécifiant que ϵ_b , ϵ_o sont non biaisées et en imposant un biais d'analyse nul i.e. $E[\epsilon_a] = 0$, on obtient :

$$\begin{aligned} E[\epsilon_a] &= E[\mathbf{L}\epsilon_b] + E[\mathbf{K}\epsilon_o] + E[(\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})x_{true}] \\ 0 &= \mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I} \\ \Leftrightarrow \mathbf{L} &= \mathbf{I} - \mathbf{K}\mathbf{H} \end{aligned}$$

5. **H1 - Opérateur d'observation linéaire** : Au voisinage de l'ébauche, l'opérateur \mathcal{H} est supposé linéaire si $\forall x$ suffisamment proche de x_b on a :

$$\mathcal{H}(x) - \mathcal{H}(x_b) = \mathbf{H}(x - x_b)$$

6. **H2 - Erreurs non biaisées** : L'ébauche et les observations sont non biaisées :

$$\mathbb{E}[x_b - x_{true}] = \mathbb{E}[y - \mathcal{H}(x)] = 0$$

7. **H3 - Analyse optimale** : L'estimateur x_a doit être non biaisé i.e. $\mathbb{E}[x_a - x_{true}] = 0$ avec une variance d'erreur minimale $\mathbb{E}[(x_a - x_{true})(x_a - x_{true})^T]$.

On considère ensuite deux nouvelles hypothèses. La première est l'hypothèse H4⁸ selon laquelle les matrices \mathbf{R} et \mathbf{B} , qui sont respectivement les matrices de covariances des erreurs d'observation ($\epsilon^o = y^o - \mathcal{H}(x^{true})$) et d'ébauche ($\epsilon^b = x^b - x^{true}$), sont connues et définies positives. Et finalement l'hypothèse H5⁹ énonce que les erreurs d'ébauche et des observations sont décorrelées. Elle permet d'utiliser l'expression de \mathbf{L} pour exprimer la matrice de covariances d'erreurs d'analyse \mathbf{A} et conduire à :

$$\begin{aligned}\mathbf{A} &= E[\epsilon_a \epsilon_a^T] = \mathbf{K}E[\epsilon_b \epsilon_b^T]\mathbf{K}^T + \mathbf{L}E[\epsilon_o \epsilon_o^T]\mathbf{L}^T \\ &= \mathbf{K}\mathbf{B}\mathbf{K}^T + \mathbf{L}\mathbf{R}\mathbf{L}^T \\ &= \mathbf{K}\mathbf{B}\mathbf{K}^T + (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{R}(\mathbf{I} - \mathbf{K}\mathbf{H})^T\end{aligned}$$

Ensuite, on cherche une estimation de la matrice \mathbf{K} qui minimisera la variance de \mathbf{A} , ce qui revient à minimiser sa trace. On écrit donc \mathbf{K} comme :

$$\begin{aligned}\mathbf{K} &= \underset{\mathbf{K}'}{\operatorname{argmin}} \operatorname{Tr}(\mathbf{A}) \\ &= \underset{\mathbf{K}'}{\operatorname{argmin}} \operatorname{Tr} \left[\mathbf{K}'\mathbf{B}\mathbf{K}'^T + (\mathbf{I} - \mathbf{K}'\mathbf{H})\mathbf{R}(\mathbf{I} - \mathbf{K}'\mathbf{H})^T \right]\end{aligned}$$

Pour cela on s'aide d'un calcul de variations et des identités suivantes :

- $\frac{\partial \operatorname{Tr}(\mathbf{K}\mathbf{A}\mathbf{K}^T)}{\partial \mathbf{K}} = \mathbf{K}(\mathbf{A} + \mathbf{A}^T)$
- $\frac{\partial \operatorname{Tr}(\mathbf{K}\mathbf{A})}{\partial \mathbf{K}} = \mathbf{A}^T$
- $\frac{\partial \operatorname{Tr}(\mathbf{A}\mathbf{K}^T)}{\partial \mathbf{K}} = \mathbf{A}$

On calcule :

$$\begin{aligned}\frac{\partial \operatorname{Tr}(\mathbf{A})}{\partial \mathbf{K}} &= \frac{\partial \operatorname{Tr}((\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T)}{\partial \mathbf{K}} \\ &= \frac{\operatorname{Tr}(\mathbf{I}\mathbf{B}\mathbf{I}^T - \mathbf{B}\mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\mathbf{B} + \mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K} + \mathbf{K}\mathbf{R}\mathbf{K}^T)}{\partial \mathbf{K}} \\ &= -\mathbf{B}\mathbf{H}^T - \mathbf{B}\mathbf{H}^T + \mathbf{K}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} + \mathbf{R}) \\ &= -2\mathbf{B}\mathbf{H}^T + 2\mathbf{K}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})\end{aligned}$$

Ainsi pour obtenir $\frac{\partial \operatorname{Tr}(\mathbf{A})}{\partial \mathbf{K}} = 0$, la matrice \mathbf{K} doit s'écrire :

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

Finalement, en réécrivant tous les termes à l'aide de \mathbf{K} , la méthode de l'Interpolation Optimale s'exprime grâce aux hypothèses H1 jusqu'à H5 et aux formules :

-
- 8. **H4 - Erreurs non triviales** : Les matrices B et R sont connues et définies positives
 - 9. **H5 - Erreurs décorrelées** : Les erreurs d'ébauche et des observations sont décorrelées :

$$\mathbb{E}[(x_b - x_{true})^T (y - \mathcal{H}(x))] = \mathbb{E}[(y - \mathcal{H}(x))^T (x_b - x_{true})] = 0$$

$$x_a = x_b + \mathbf{K}(y - \mathbf{H}x_b) \quad (1.1)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (1.2)$$

Le principal problème de cette méthode est qu'elle requiert le calcul de $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$. Ces matrices étant de très grandes dimensions, une sélection d'observations est faite pour les réduire. Le modèle est divisé en boîtes i ce qui permet de calculer une analyse au sein de chacune d'entre elles de manière indépendante. Pour cela on considère les observations se trouvant dans chaque boîte i ainsi que dans son voisinage proche. On peut voir sur la figure 1.9 un exemple de sélection d'observations : ici pour chaque boîte i du modèle, on considère les observations étant dans une boîte un peu plus grande, en bleue sur la figure, et on calcule : $[(\mathbf{H}\mathbf{B}\mathbf{H}^T)^{(i)} + \mathbf{R}^{(i)}]^{-1}$, grandeur qui possède une dimension égale au nombre d'observations contenues dans cette boîte. Cette méthode a été implémentée dans de nombreux centres de PNT, notamment au CEPMMT où elle était la méthode d'assimilation opérationnelle entre 1979 et 1996. Nous allons maintenant décrire des méthodes plus récentes qui sont mieux adaptées à l'utilisation d'observations de télédétection spatiale. En effet, la méthode qui vient d'être présentée repose sur l'existence d'une relation linéaire entre les observations y et l'état x que l'on cherche à analyser à travers l'opérateur \mathbf{H} ce qui n'est pas le cas pour le domaine des micro-ondes et les scènes météorologiques nuageuses.

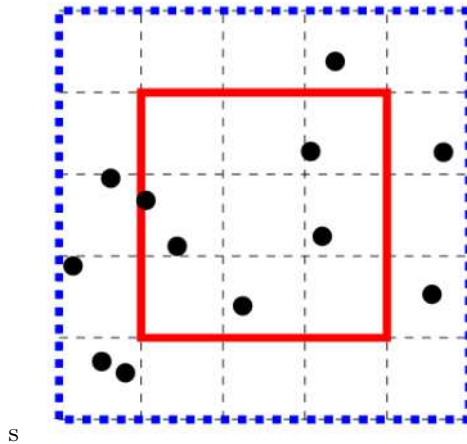


FIGURE 1.9 – Version de l'interpolation optimale (OI) consistant à utiliser, pour tous les points d'une boîte d'analyse (en rouge), toutes les observations situées dans une boîte plus grande (en bleue) de sorte que la plupart des observations sélectionnées dans deux boîtes voisines soient identiques. Ici les observations sont représentées par des points noirs.

Source: www.cerfacs.fr/~ricci/documents/Total_mai_2008_general_ricci.ppt

1.3.2 Les méthodes variationnelles

Le 3D-Var

Les équations du BLUE, présentées précédemment dans un cadre linéaire pour l'interpolation optimale, peuvent s'écrire sous forme d'un problème de minimisation. C'est ainsi qu'une nouvelle méthode d'assimilation de données dite "variationnelle tridimensionnelle" (3D-Var) a conduit à des progrès significatifs en PNT à partir du milieu des années 1990 (Le Dimet et Talagrand, 1986[77]). Cette méthode a tout d'abord été implémentée en 1992 au NCEP (National Center of Environmental Prediction, Parrish et Derber 1992[78]), puis en 1996 au CEPMMT (Courtier et al 1998[79]). Elle permet de se passer du calcul

de la matrice \mathbf{K} en cherchant plutôt à minimiser l'écart entre l'ébauche et les observations à l'aide d'une fonctionnelle J . Les statistiques d'erreurs pour l'ébauche et les observations sont généralement définies à l'aide de distributions de probabilité gaussiennes, qui sont ainsi entièrement définies par les moyennes et les covariances spécifiées. Dans ce cas, la solution du problème d'assimilation des données est égale au maximum de la densité de probabilité de l'estimation bayésienne de l'état atmosphérique a posteriori. On cherche donc l'analyse x_a qui maximise la densité de probabilité p de x connaissant les observations y et une information a priori l'ébauche x_b , soit :

$$x_a = \underset{x}{\operatorname{argmax}}(p(x|y))$$

Grâce à cette équation, nous pouvons nous servir de la fonction logarithme et d'une constante arbitraire pour définir une fonction de coût J :

$$J = -\log(p(x|y)) + cste$$

et ainsi le problème de minimisation revient à trouver la solution suivante :

$$x_a = \underset{x}{\operatorname{argmin}}(J(x))$$

Il est possible de redéfinir cette fonction coût à l'aide de plusieurs éléments. Premièrement, le théorème de Bayes permet de réécrire la probabilité de x sachant y :

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Ici $p(y)$ est la fonction de densité de probabilité marginale des observations aussi considérée et représente la constante du théorème de Bayes. Ainsi on obtient :

$$p(x|y) \propto p(y|x)p(x)$$

Afin de trouver la variable x minimisant la probabilité $p(y|x)$, on applique la fonction logarithme. Ainsi, on cherche à minimiser la fonction coût :

$$J(x) = -\log[p(y|x)] - \log[p(x)] + cste$$

Le 3D-Var reposant sur certaines hypothèses du BLUE, les probabilités $p(x)$ et $p(y|x)$ sont supposées être des densités de probabilité gaussiennes. En particulier $p(x)$ s'écrit comme la probabilité a priori d'un écart à l'ébauche x_b , i.e. :

$$p(x) = \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} e^{[-\frac{1}{2}(x_b-x)^T \mathbf{B}^{-1}(x_b-x)]}$$

$$p(y|x) = \frac{1}{(2\pi)^{n/2}|\mathbf{R}|^{1/2}} e^{[-\frac{1}{2}(y-\mathcal{H}(x))^T \mathbf{R}^{-1}(y-\mathcal{H}(x))]}$$

Avec n la taille du vecteur x . En choisissant une constante valant 0, il est possible de réécrire J telle que :

$$J(x) \propto J_b(x) + J_o(x)$$

$$J(x) \propto \frac{1}{2}(x_b - x)^T \mathbf{B}^{-1}(x_b - x) + \frac{1}{2} [y - \mathcal{H}(x)]^T \mathbf{R}^{-1} [y - \mathcal{H}(x)]$$

Avec J_b qui représente l'écart à l'ébauche et J_o l'écart aux observations. Ensuite, on utilise une méthode itérative grâce à un algorithme de minimisation de type gradient conjugué (modèle ARPEGE) ou Quasi-Newton (modèle AROME) pour trouver le minimum de J en prenant x_b comme point de départ. L'analyse x_a vaut alors :

$$x_a = \underset{x}{\operatorname{argmin}} J(x)$$

Cette méthode possède plusieurs avantages par rapport à l'interpolation optimale. Le 3D-Var effectue une analyse dans sa globalité, en évitant les discontinuités possibles de l'interpolation optimale aux limites des sous-domaines et permet d'utiliser toutes les observations pour chaque point de grille du modèle. Il est aussi possible d'utiliser des opérateurs d'observations non-linéaires autorisant ainsi l'assimilation de luminances satellitaires au lieu de produits géophysiques inversés. Par conséquent, la spécification des erreurs d'observations est plus facile, ainsi que l'identification et la correction de biais éventuels dans les mesures. L'assimilation variationnelle permet également un filtrage des ondes de gravité par ajout d'un terme supplémentaire dans la fonction coût. On s'affranchit de ce fait de l'étape "d'initialisation" nécessaire après une analyse par interpolation optimale (Simmons, 1999[80]).

Les assimilations de type interpolation optimale et 3D-Var fournissent des analyses à un instant donné. On considère dans la pratique une fenêtre temporelle autour de l'heure d'analyse (qui est de 60 min pour le modèle AROME) dans laquelle toutes les observations disponibles sont utilisées en les supposant valides au même instant. Nous allons à présent décrire une méthode d'assimilation permettant de prendre en compte explicitement la dimension temporelle des observations.

Le 4D-Var

La méthode d'assimilation variationnelle quadridimensionnelle (le 4D-Var, Talagrand et Courtier 1987[81], Courtier et Talagrand 1987[82], Courtier et al 1994[83], Courtier 1997[84]) est l'équivalent du 3D-Var en prenant en compte l'aspect temporel pour la modélisation et l'utilisation des observations. La nécessité d'utiliser le 4D-Var plutôt que le 3D-Var réside dans le fait que les observations asynoptiques, comme les observations satellitaires, ne sont prises en compte qu'aux heures synoptiques avec un 3D-Var ce qui conduit à une utilisation sous-optimale des informations qu'elles contiennent. Le 4D-Var, au contraire, assimile ce type d'observation de manière continue grâce à sa fenêtre d'assimilation étendue. Le 4D-Var ne calcule donc pas l'état optimal à un instant donné mais une trajectoire optimale sur une fenêtre de temps donnée. Ainsi on compare des observations à un instant donné avec un état modèle au même instant. L'état initial en début de fenêtre temporelle est alors ajusté (dans un algorithme de minimisation) de façon à produire une évolution du modèle (trajectoire) qui s'approche au mieux des observations disponibles tout en ne s'éloignant pas trop de l'ébauche (état du modèle au début de la fenêtre) comme représenté figure 1.10. Pour cela on se sert des états successifs x_t définis à l'aide du modèle décrivant l'évolution des variables atmosphériques \mathcal{M}_t depuis l'instant $t - 1$ jusqu'à l'instant t tel que :

$$x_t = \mathcal{M}(x_{t-1})$$

Ici le modèle \mathcal{M} est supposé parfait et est imposé comme une contrainte forte, permettant de réduire la minimisation à l'état 3D du modèle au début de la fenêtre d'assimilation. En reprenant les hypothèses du 3D-Var et en prenant en compte différents instants t , où m est le nombre d'instants où des observations sont présentes, la fonction coût à minimiser s'écrit :

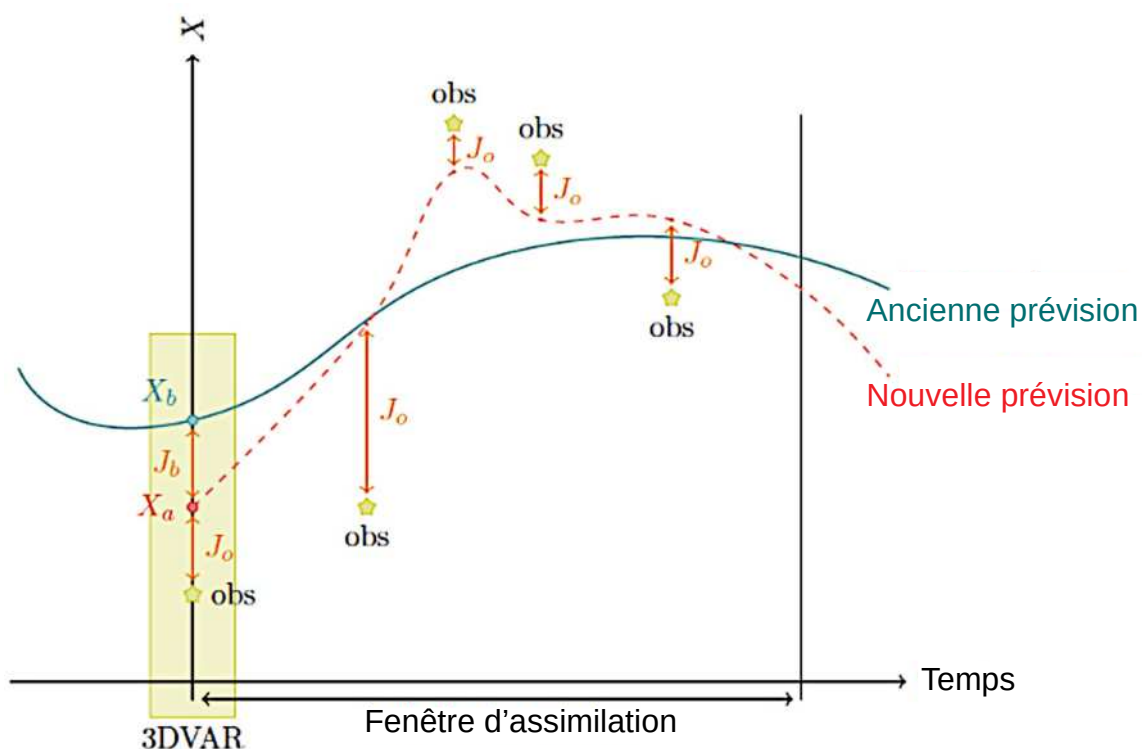


FIGURE 1.10 – Schéma de la méthode du 4D-Var : la prévision commence à partir de l'ébauche X_b et suit la trajectoire en rouge. La prévision corrigée en bleu part de l'analyse X_a qui a été modifiée pour ajuster la nouvelle trajectoire en fonction des observations disponibles dans la fenêtre d'assimilation.

Source: <https://asynch.readthedocs.io/en/latest/assim.html>

$$J(x_0) = \frac{1}{2}(x_0 - x_b)^T \mathbf{B}^{-1}(x_0 - x_b) + \frac{1}{2} \sum_{t=0}^m [y_t - \mathcal{H}_t(x_t)]^T \mathbf{R}^{-1} [y_t - \mathcal{H}_t(x_t)]$$

Le 4D-Var a été implémenté en novembre 1997 au CEPMMT puis à Météo-France en 2000. Des études réalisées à cette époque ont en effet montré les améliorations apportées à l'analyse et aux prévisions grâce au 4D-Var par rapport au 3D-Var (Rabier et al. 2000[85], Mahfouf et Rabier 2000[86], Klinker et al. 2000[87]). Le 4D-Var demeure à ce jour le système d'assimilation de données utilisé opérationnellement dans ces deux centres. À Météo-France, la matrice de covariances des erreurs d'ébauche \mathbf{B} exploitée et fournie toutes les 6 h au 4D-Var provient d'un ensemble d'assimilation (Pereira et Berre, 2006[88]).

Le 4D-Var nécessite l'opérateur adjoint (transposée de l'opérateur tangent-linéaire) du modèle de prévision ainsi que des opérateurs d'observations pour résoudre efficacement le problème de minimisation (calcul du gradient de la fonction coût) (Le Dimet et Talagrand 1986). Les méthodes utilisant l'adjoint et son tangent linéaire associé sont souvent appelées modèle linéaire tangent et modèle adjoint (TL/AD). Ainsi l'analyse à haute résolution (HR) s'écrit comme la somme d'une ébauche à haute résolution et d'un incrément δx à basse résolution (BR) :

$$x_a(HR) = x_b(HR) + \delta x(BR)$$

On justifie aussi cette approche incrémentale par le fait que le réseau d'observations ne va pas apporter d'informations utiles sur les plus petites échelles spatiales résolues par le modèle.

Aussi il est possible de réécrire J en introduisant d , le vecteur d'innovation, qui vaut $y - \mathcal{H}(x_b)$ et δx qui vaut $x - x_b$. La reformulation pour le 3D-Var s'écrit :

$$J(\delta x) = \frac{1}{2}(\delta x)^T \mathbf{B}^{-1}(\delta x) + \frac{1}{2}(d - \mathbf{H}\delta x)^T \mathbf{R}^{-1}(d - \mathbf{H}\delta x)$$

$$\nabla_{\delta x} J(\delta x) = \mathbf{B}^{-1}(\delta x) - \mathbf{H}^T \mathbf{R}^{-1}(d - \mathbf{H}\delta x)$$

Pour le 4D-Var le processus est le même sauf que l'incrément d'analyse vaut à chaque pas de temps :

$$\delta x_i = \mathbf{M}_i(\delta x)$$

où \mathbf{M}_i est le modèle de prévision (linéaire tangent) jusqu'à l'instant i et on obtient :

$$J(\delta x) = \frac{1}{2}(\delta x)^T \mathbf{B}^{-1}(\delta x) + \sum_{i=1}^k (\mathbf{H}_i[\mathbf{M}_i(\delta x)] - d_i)^T \mathbf{R}^{-1}(\mathbf{H}_i[\mathbf{M}_i(\delta x)] - d_i)$$

$$\nabla_{\delta x} J(\delta x) = \mathbf{B}^{-1}(\delta x) + \sum_{i=1}^k \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{R}^{-1}(\mathbf{H}_i[\mathbf{M}_i(\delta x)] - d_i)$$

Cette formulation permet ainsi d'exprimer le problème à l'aide d'une fonction coût quadratique dont le minimum est plus simple à calculer. D'un point de vue pratique, la minimisation du 4D-Var résout un problème incrémental à plus basse résolution, permettant de limiter la dimension du vecteur de contrôle à estimer et de réaliser certaines simplifications au niveau des modèles linéarisés (notamment concernant la description des processus physiques). Ces différents aspects permettent de diminuer significativement le coût de calcul du problème d'analyse. Finalement, la minimisation et les prévisions sont réalisées à différentes résolutions définies tableau 1.4 pour le 4D-Var opérationnel à Météo-France.

4D-Var ARPEGE	
Résolution 1 ^{ere} minimisation	T224 ¹⁰ (~90 km)
Résolution 2 ^{ime} minimisation	T499 (~40 km)
Résolution de la prévision	T1798 (~5 km)

TABLE 1.4 – Les différentes résolutions du 4D-Var opérationnel à Météo-France

L'assimilation d'ensemble AEARP

L'estimation de la matrice \mathbf{B} de covariances d'erreurs d'ébauche apparaissant dans les équations décrites précédemment est très importante en assimilation de données pour plusieurs raisons. En effet, elle :

- propage spatialement et/ou temporellement les informations provenant des observations et celles venant de l'ébauche,
- propage les informations de manière multivariée en assurant l'équilibre entre les différentes variables du modèle,

- fournit des incréments statistiquement cohérents avec le modèle : dans les zones avec un grand nombre d'observations un lissage est effectué afin que les échelles des incréments soient compatibles avec le modèle et l'équilibre de l'incrément est géré au sein de blocs diagonaux reliant les différentes variables atmosphériques,
- donne davantage de poids aux observations dans les zones où l'ébauche est incertaine et réciproquement, peu de poids aux observations dans des zones où l'ébauche est précise.

Ainsi la matrice \mathbf{B} doit être définie de manière la plus précise possible. Cependant elle reste très difficile à estimer à cause de sa très grande dimension, de l'ordre de $10^7 \times 10^7$. L'approche incrémentale permet de réduire la taille du vecteur de contrôle mais d'autres méthodes sont nécessaires. La plupart d'entre elles utilisent des réductions de rang pour calculer cette matrice et des approximations dans la mesure où l'état vrai de l'atmosphère n'est pas connu. Plusieurs méthodes existent à ce jour, dont la méthode s'appuyant sur les statistiques d'innovations (Hollingsworth et Lönnberg, 1986[89]), la méthode dite du "NMC" (National Meteorological Center, basée sur l'utilisation de prévisions à des échéances différentes mais valides à la même date, Parrish et Derber 1992[78]) et les méthodes d'ensembles. Ces dernières s'appuient sur des approches ensemblistes des méthodes d'assimilation de données initialement proposées pour le filtre de Kalman¹¹ (Evensen, 1994[90]).

Depuis quelques années à Météo-France, la méthode d'ensemble est exploitée pour estimer la matrice \mathbf{B} (Berre et al. 2015[13], Berre et al. 2019[14]). Contrairement aux premières méthodes d'ensemble qui utilisaient le filtre de Kalman, cette méthode exploite le 4D-Var déterministe opérationnel à Météo-France au sein du modèle global ARPEGE. Elle se nomme ainsi Assimilation d'ensemble ARPege (AEARP). À l'heure actuelle l'AEARP possède 50 membres fournissant 50 analyses provenant des ébauches perturbées du cycle précédent (Bouyssel et al. 2022[10]). Chaque membre de l'ensemble est calculé en considérant les observations disponibles sur une période d'assimilation et qui vont être perturbées en y ajoutant un bruit gaussien. Ces observations perturbées sont ensuite assimilées au sein d'un ensemble de 50 assimilations 4D-Var. Lors des différentes assimilations, les perturbations vont évoluer de la même manière que les erreurs du système d'assimilation ce qui va permettre d'obtenir les matrices nécessaires au calcul de la matrice de covariances d'erreurs d'ébauche \mathbf{B} .

En assimilation de données variationnelle, la matrice \mathbf{B} est souvent décomposée à l'aide de matrices creuses¹². Pour se faire, la matrice \mathbf{B} est exprimée à l'aide d'opérateurs en fonction des variables de la vortacité, de la divergence, de la température, du logartihme de surface et de l'humidité spécifique. Le premier opérateur est la matrice \mathbf{L} d'équilibre, tenant compte des covariances croisées entre les variables de masse (température et pression de surface) et de vents à partir d'une combinaison d'équations d'équilibre dépendantes des flux (Fisher, 2003[91]) et de régressions multivariées (Derber et Bouttier, 1999[92]). Le deuxième opérateur est la matrice diagonale Σ des écarts-types d'erreurs d'ébauche dans l'espace des points de grille¹³. Elle peut être directement exprimée grâce aux N membres de l'ensemble telle que :

$$\Sigma = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (x_b^k - \bar{x}_b)^2}$$

Avec x_b^k les ébauches perturbées et \bar{x}_b la moyenne de l'ensemble.

Le dernier opérateur est la matrice \mathbf{C} des corrélations 3D de chaque variable. Afin de relaxer l'hypothèse d'homogénéité, cette matrice est construite dans l'espace des ondelettes présentant l'avantage de contenir des informations sur l'échelle et la position horizontale. Pour décomposer la matrice dans l'espace des ondelettes, les erreurs d'ébauche e_b sont convoluées avec des fonctions passe-bandes¹⁴ radiales¹⁵ Φ_j pour différentes échelles $\hat{e}_{b,j}$. On considère alors deux équations :

11. le filtre de Kalman utilise les mêmes équations que celles présentées pour l'interpolation optimale avec en plus la propagation dans le temps d'une matrice de covariances d'erreurs de l'ébauche et l'estimation de l'erreur d'analyse

12. Matrices composées majoritairement d'éléments nuls

13. Espace dans lequel on réalise les calculs de la physique, du semi-lagrangien, par opposition à l'espace spectral dans lequel on calcule les gradients, la diffusion et le problème implicite

14. Filtre consistant à atténuer les fréquences hors de la bande passante

15. fonction dépendant uniquement de la distance radiale entre son entrée et un point fixe

$$\hat{e}_{b,j} = e_b \otimes \Psi_j$$
$$e_b = \sum_j \hat{e}_{b,j} \otimes \Psi_j$$

Ainsi, la matrice \mathbf{B} s'exprime à l'aide des différents opérateurs telle que :

$$\mathbf{B} = \mathbf{L}\Sigma\mathbf{C}\Sigma^T\mathbf{L}^T$$

Finalement, la matrice \mathbf{B} de covariances d'erreurs d'ébauches est calculée à partir des résultats de ces 50 assimilations et fournie au modèle déterministe opérationnel toutes les 6h. Les résolutions du modèle de prévision utilisée lors du processus de cyclage sont cependant moins fines que le 4D-Var déterministe opérationnel à Météo-France. Les sorties de l'AEARP sont également utilisées à Météo-France dans le cadre de la Prévision d'Ensemble ARPege (PEARP, Descamps et al. 2015[93]). La PEARP fournit un ensemble de prévisions perturbées, et utilise les analyses provenant de l'AEARP comme états initiaux.

Chapitre 2

Modélisation du transfert radiatif micro-ondes dans les nuages

Sommaire

2.1	Le transfert radiatif	29
2.1.1	Principe général	29
2.1.2	La théorie du corps noir	34
2.1.3	L'équation du transfert radiatif	35
2.1.4	Un code de transfert radiatif : RTTOV-SCATT	39
2.2	Calcul des propriétés radiatives des hydrométéores	45
2.2.1	Les régimes de diffusion	45
	Les formes des hydrométéores solides	47
	Les particules de formes non sphériques	49
2.2.2	Historique des bases de données	51

Après avoir défini les principes fondamentaux de la PNT, nous allons décrire dans ce chapitre le principe du transfert radiatif en vue de la simulation des observations micro-ondes en zones nuageuses au sein des processus d'assimilation de données. Nous allons tout d'abord introduire les notions de rayonnement électromagnétique et de corps noir, puis nous établirons l'Equation du Transfert Radiatif (ETR) quantifiant les échanges d'énergie du rayonnement électromagnétique lors de son interaction avec les molécules (e.g. oxygène, vapeur d'eau) et les particules atmosphériques (e.g. hydrométéores, aérosols). Ensuite nous examinerons toutes les contributions atmosphériques nécessaires à l'établissement de l'ETR afin de décrire les échanges. Nous verrons finalement comment la résoudre, tout en détaillant plus précisément les aspects microphysiques des hydrométéores solides qui nous intéressent particulièrement dans cette thèse.

2.1 Le transfert radiatif

2.1.1 Principe général

La théorie du transfert radiatif décrit les interactions du rayonnement électromagnétique avec les constituants de l'atmosphère terrestre (gaz et particules). Il est ainsi important de définir premièrement la notion de rayonnement électromagnétique. Le rayonnement électromagnétique peut être décrit d'un point de vue ondulatoire et d'un point de vue corpusculaire. Dans le premier cas, le rayonnement électromagnétique est décrit comme une onde combinant un champ magnétique (\vec{B}) et un champ électrique (\vec{E}). Les relations entre ces deux champs ont été formalisées mathématiquement, au *XIX^{ème}* siècle, par les équations de Maxwell décrivant les phénomènes électriques, magnétiques et lumineux. Le rayonnement électromagnétique possède différentes caractéristiques dont la principale est sa **longueur d'onde** λ , représentant la

longueur d'un cycle d'une onde et s'exprimant en mètres.

La longueur d'onde du rayonnement est liée à d'autres caractéristiques d'un rayonnement électromagnétique. Par exemple, **la fréquence** ν d'un rayonnement électromagnétique représente le nombre de cycles d'une onde par unité de temps et s'exprime en Hertz. Elle est liée à la longueur d'onde par la formule :

$$\lambda = c/\nu$$

où c est **la célérité**, autrement dit **la vitesse de la lumière dans le vide** et vaut $2.998 \times 10^8 \text{ m/s}^{-1}$. **La période**, autre caractéristique du rayonnement électromagnétique, représente la durée séparant deux cycles d'une onde. Elle s'exprime en seconde et est liée à la fréquence par la formule :

$$T = \frac{1}{\nu}$$

La longueur d'onde, la fréquence et la période d'une onde électromagnétique sont invariables quelles que soient les conditions de propagation de l'onde.

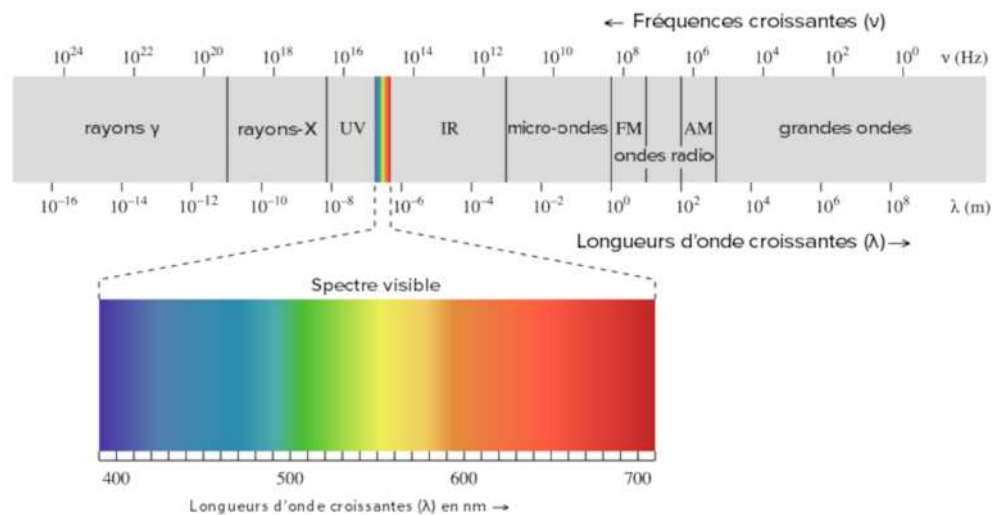


FIGURE 2.1 – Le spectre électromagnétique

Source: UC Davis ChemWiki

La longueur d'onde permet de classer les rayonnements électromagnétiques en plusieurs catégories comme on le voit sur la figure 2.1. Cette classification se nomme le *spectre électromagnétique*. Parmi les différents types de longueurs d'ondes on distingue les rayonnements infrarouge, visible et micro-ondes souvent utilisés en télédétection comme vu dans le chapitre 2.

Décrits d'un point de vue corpusculaire, le rayonnement électromagnétique représente une forme d'énergie quantifiée par des photons¹. Pour décrire l'évolution du rayonnement lors de son interaction avec des molécules, il est nécessaire d'introduire les états d'énergies de ces molécules. On trouve premièrement l'état fondamental où l'énergie de la molécule est la plus basse et ensuite les états dits "d'excitations" i.e. des états d'énergie supérieurs où la molécule est stable. On distingue :

- Un niveau d'énergie électronique E_e ;

1. Quantum d'énergie

Transition	ΔE (eV)	$\Delta E/h$ (Hz)	$\lambda = c/\nu$	Région spectrale
électronique	10	2.4×10^{15}	83.3 nm	ultraviolet et visible
vibrations	10^{-1}	2.4×10^{13}	8.33 μm	infrarouge moyen et thermique
rotations	10^{-3}	2.4×10^{11}	0.833 mm	micro-ondes

TABLE 2.1 – Niveaux d'énergie d'une molécule selon différentes caractéristiques. ΔE représente la quantité d'énergie pour passer d'une énergie E_1 à une énergie de niveau supérieure E_2 en électronvolt (eV, $1 \text{ eV} = 1.602176634 \times 10^{-19} \text{ J}$)

Source: <http://step.ipgp.fr/images/0/04/TeledetectionL3cours1.pdf>

- Un niveau d'énergie vibrationnelle E_v qui correspond au mouvement des atomes de la molécule autour de leur position d'équilibre ;
- Et un niveau d'énergie rotationnelle E_r qui correspond aux mouvements de rotation de la molécule autour d'un axe passant par son centre d'inertie.

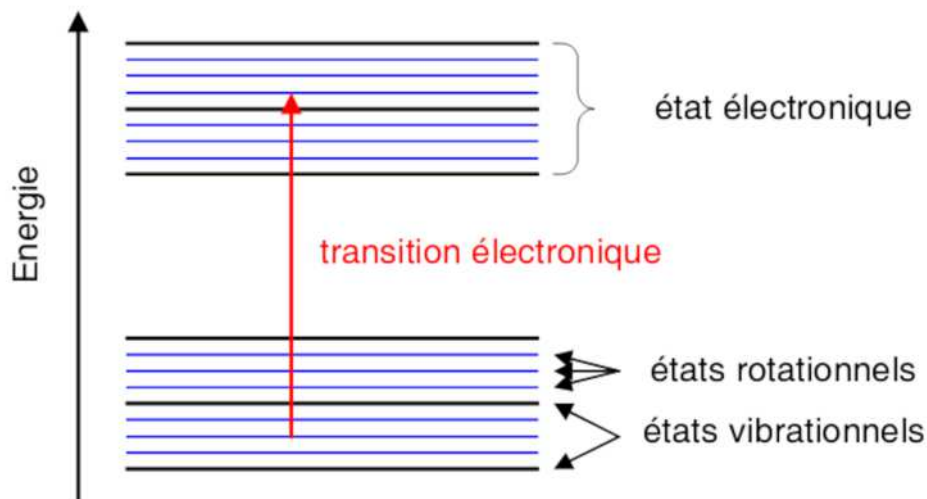


FIGURE 2.2 – Représentation schématique des différents niveaux d'énergie d'une molécule

Source: <https://phychim.ac-versailles.fr/>

Ces niveaux d'énergie suivent la relation suivante :

$$E_e \gg E_v \gg E_r$$

Un schéma de ces différents niveaux d'énergie est représenté sur la figure 2.2. Lors de leur interaction avec un rayonnement électromagnétique incident, ces molécules entrent dans des états d'excitation dépendants de la longueur d'onde de ce dernier. Le tableau 2.1 présente un récapitulatif des différents états de transition de la molécule selon le domaine de longueur d'onde du rayonnement électromagnétique.

Les différents états générés par l'onde électromagnétique incidente entraînent une réaction des particules :

- **L'émission** : lorsque la particule est dans un état d'énergie excité instable, celle-ci va relâcher un photon afin de revenir à un état stable. C'est ce que l'on appelle le

processus d'émission.

- **L'absorption** : sous certaines conditions, l'énergie incidente est transmise à la particule qui rentre dans un état excité. Ceci correspond à l'absorption du rayonnement. Pour répondre à ces conditions, il est nécessaire que le photon du rayonnement électromagnétique ait une énergie similaire à la différence d'énergie entre deux états stables donnés de la molécule :

$$E' - E = h\nu_o = \frac{hc}{\lambda_o}$$

Avec E' et E des états stables de la molécule, et h la constante de Planck.

- **La diffusion** : le rayonnement peut être diffusé par la particule s'il n'est pas absorbé par celle-ci. Ainsi, des photons du rayonnement vont être redirigés par la particule dans une autre direction sans que la molécule ne subisse un changement d'énergie

La théorie du transfert radiatif permet de caractériser et calculer les échanges radiatifs subit par le rayonnement atmosphérique lors de sa propagation dans l'atmosphère. Cette théorie est très intéressante, car elle permet d'obtenir des informations sur le contenu de l'atmosphère grâce à l'intensité du rayonnement électromagnétique dans des bandes de fréquences bien choisies et mesurées par instruments embarqués à bord des satellites. Par exemple, si une longueur d'onde est absorbée par la vapeur d'eau, on pourra en déduire des informations sur ce gaz.

Il est possible de mesurer un spectre d'absorption, associé à des longueurs d'ondes données, pour connaître le pourcentage d'absorption de ce rayonnement par les composants atmosphériques. Si on examine plus spécifiquement le spectre d'absorption des micro-ondes (figure 2.3) qui va nous intéresser dans cette thèse, on voit qu'il possède des fenêtres atmosphériques et des raies d'absorption à différentes fréquences. Sur ce graphique, les *fenêtres atmosphériques* sont représentées par les zones possédant un fort pourcentage de transmission. Cela signifie que le rayonnement électromagnétique n'est pas absorbé par les composantes atmosphériques ou l'est très faiblement. Sur le spectre micro-ondes, ces zones sont situées entre 1 et 40 GHz, à 90 GHz et à 135 GHz. Au contraire, les parties du spectre où l'absorption intervient en raison d'une transition entre deux états sont appelées raies d'absorption et sont représentées par les zones avec un pourcentage de transmission à 0%. Pour les fréquences comprises entre 50 et 70 GHz, les raies d'absorption se trouvent autour du complexe de raies d'absorption de l'oxygène, ainsi l'effet de ce gaz à concentration uniforme dans l'atmosphère permet une bonne restitution de la température à ces fréquences. Ceci est aussi le cas pour la fréquence 118 GHz qui possède une raie isolée d'absorption de l'oxygène. Les fréquences 22 et 183 GHz se situent quant à elles dans les raies d'absorption de la vapeur d'eau. Ainsi en PNT, la fréquence 22 GHz sert habituellement à restituer le contenu en vapeur d'eau intégrée tandis que la fréquence 183 GHz est utilisée pour le sondage de l'humidité atmosphérique. Finalement, les autres fréquences se situent dans des canaux fenêtres éloignés des raies d'absorption et on observe les effets du continuum de la vapeur d'eau qui contribue à l'opacité de l'atmosphère à mesure que les fréquences micro-ondes augmentent.

Nous allons maintenant voir comment calculer l'énergie émise par un rayonnement électromagnétique suite à sa propagation dans l'atmosphère. Pour cela la notion *d'angle solide* doit être introduite. Sur la figure 2.4, on considère un rayonnement se propageant par rapport au centre de la Terre avec un angle azimutal ϕ et un angle zénithal θ sur une surface de la sphère dA . L'angle solide $d\Omega$ est le rapport entre la surface dA interceptée par un cône et le carré du rayon r de la sphère, soit :

$$d\Omega = \frac{dA}{r^2} \quad (2.1)$$

Nous allons à présent définir la *luminance spectrale* L_λ , aussi appelée *radiance*. Cette quantité représente l'énergie émise ou reçue par unité de temps, de longueur d'onde, et d'angle solide. Elle s'exprime en $W.m^{-2}.sr^{-1}$. En reprenant l'illustration figure 2.4, on calcule la radiançe L_λ par unité d'angle solide $d\Omega$, par unité de surface $ds \cos(\theta)$, dans une direction donnée (θ, Φ) grâce à la formule :

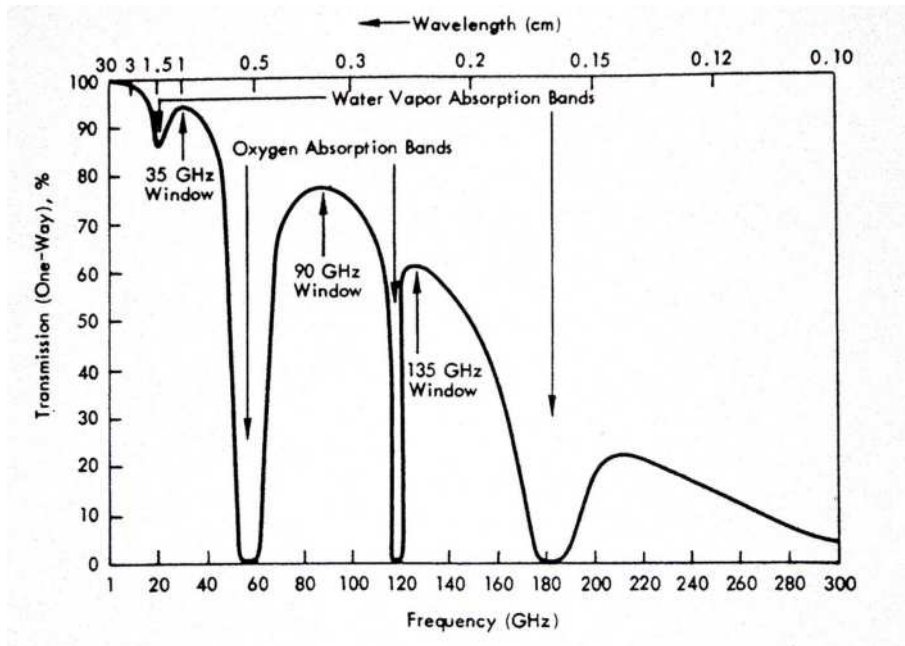


FIGURE 2.3 – Pourcentage de transmission à travers l’atmosphère terrestre le long de la verticale en conditions de ciel clair pour un rayonnement micro-ondes dont la fréquence varie entre 1 et 300 GHz

Source: Ulaby et al. 1981[94]

$$L_{\lambda} = \frac{d^2\Phi}{d\Omega dscos\theta} = \frac{1}{d\Omega cos\theta} dM(\theta, \phi) \quad (2.2)$$

Où $M(\theta, \phi)$ est l’**exitance** i.e. la densité de l’énergie par unité de surface et qui s’exprime en $W.m^{-2}$. Sa formule s’obtient en intégrant la luminance spectrale sur une demi-sphère :

$$M(\theta, \phi) = \int_0^{2\pi} \int_0^1 L_{\lambda}(\mu, \phi) \mu d\mu d\phi \quad (2.3)$$

Dans le cas d’une luminance spectrale isotrope (invariante suivant les directions), on obtient la *loi de Lambert* où l’exitance est proportionnelle à la luminance spectrale, soit :

$$M = \pi L_{\lambda} \quad (2.4)$$

Dans ses interactions avec la matière, l’énergie d’un rayonnement électromagnétique peut être modifiée de différentes manières :

- Une fraction τ_{λ} de l’énergie est transmise :

$$\tau_{\lambda} = \frac{\text{radiation transmise à } \lambda}{\text{radiation émise à } \lambda}$$

- Une fraction a_{λ} d’énergie est absorbée :

$$a_{\lambda} = \frac{\text{radiation absorbée à } \lambda}{\text{radiation émise à } \lambda}$$

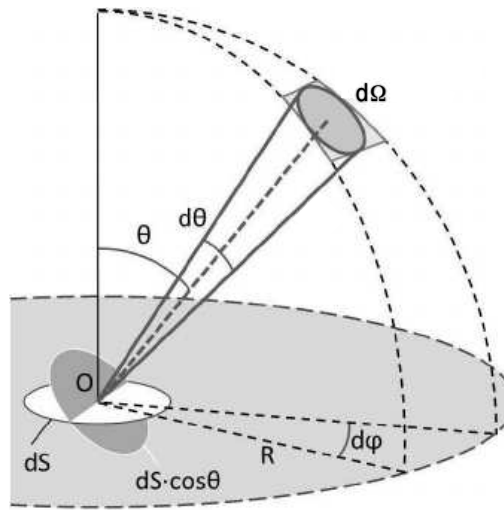


FIGURE 2.4 – Représentation de l’angle solide $d\Omega$, avec θ l’angle zénithal et ϕ l’angle azimutal du rayonnement incident.

Source: Matsapey et al. 2013[95]

- Une fraction ρ_λ d’énergie est réfléchie ou diffusée :

$$\rho_\lambda = \frac{\text{radiation réfléchie à } \lambda}{\text{radiation émise à } \lambda}$$

Ces trois phénomènes sont les seules possibilités du devenir d’un rayonnement de par le principe de la conservation d’énergie. La première *loi de Kirchhoff* exprime ce principe par la formule :

$$\rho_\lambda + a_\lambda + \tau_\lambda = 1 \quad \text{où} \quad \rho_\lambda, a_\lambda \text{ et } \tau_\lambda \leq 1$$

2.1.2 La théorie du corps noir

Nous allons maintenant voir la théorie du corps noir. Un corps absorbant complètement le rayonnement quelle que soit sa longueur d’onde (i.e. avec $a_\lambda = 1$) est appelé **corps noir**. En 1901, Max Planck montra que la luminance énergétique d’un corps noir notée B_λ dépend seulement de sa température T et de sa longueur d’onde λ suivant une loi dite de Planck :

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5 \left(\exp \left[\frac{hc}{k\lambda T} \right] - 1 \right)}$$

- $h = 6.626 \times 10^{-34} \text{ m}^2.\text{kg}.\text{s}^{-1}$ est la constante de Planck
- $c = 2.998 \times 10^8 \text{ m}.\text{s}^{-1}$ est la vitesse de la lumière dans le vide
- $k = 1.38 \times 10^{-23} \text{ J}.\text{K}^{-1}$ est la constante de Boltzmann

Les longueurs d'onde micro-ondes vérifient de plus la relation :

$$\frac{hc}{k\lambda T} \ll 1$$

Un développement limité à l'ordre 1 de $\exp(x) - 1$ lorsque x tend vers 0 étant x , on obtient de ce fait l'importante simplification :

$$\exp\left[\frac{hc}{k\lambda T}\right] - 1 \simeq \frac{hc}{k\lambda T}$$

Et la fonction de Planck devient finalement dans le domaine des micro-ondes :

$$B_\lambda(T) = \frac{2ck}{\lambda^4} \times T$$

Cette simplification, appelée *approximation de Rayleigh-Jeans*, permet de noter que la luminance d'un corps noir est proportionnelle à la température T du corps noir dans les micro-ondes.

Le corps noir permet d'introduire une nouvelle variable : **la température de brillance**. La température de brillance est la température d'un corps noir avec la même luminance pour une longueur d'onde donnée. Ainsi pour toute luminance L_λ , on obtient à l'aide de la fonction de Planck (simplifiée pour les micro-ondes) la température de brillance TB_λ :

$$TB_\lambda = \frac{\lambda^4}{2ck} \times L_\lambda$$

Pour tout corps qui n'est pas un absorbant parfait, i.e. qui n'est pas un corps noir, la luminance est fonction du **coefficient d'émissivité spectrale** ϵ_λ :

$$\epsilon_\lambda = \frac{L_\lambda}{B_\lambda(T)}$$

Les observations micro-ondes utilisées plus loin dans ce manuscrit seront exprimées à l'aide des températures de brillance et nécessiteront la connaissance de ce coefficient qui caractérise la capacité d'un corps à émettre et à absorber l'énergie radiative. La loi de Kirchhoff permet d'affirmer que tout corps ayant un facteur d'émissivité ϵ_λ à une température, une longueur d'onde et une direction données a un facteur d'absorption a_λ égal, i.e. :

$$a_\lambda = \epsilon_\lambda$$

Pour un milieu opaque, i.e. $\tau_\lambda = 0$, on obtient de plus la relation :

$$a_\lambda + \rho_\lambda = 1 \text{ ou encore } \rho_\lambda = 1 - \epsilon_\lambda$$

Les notions caractérisant les comportements d'un rayonnement ayant été abordées, nous allons présenter les calculs nécessaires pour décrire les luminances à un point donné en prenant en compte les contributions des différents couches d'atmosphère avec leurs constituants (gaz et particules).

2.1.3 L'équation du transfert radiatif

Pour construire l'Equation du Transfert Radiatif (ETR), il faut tout d'abord étudier le comportement d'un rayonnement électromagnétique lorsqu'il traverse un volume de matière. Ceci permettra par la suite

d'exprimer les échanges de flux radiatifs dans un volume d'atmosphère. Quatre processus sont à prendre en compte : l'absorption, la diffusion en dehors du volume, l'émission par le volume de matière et enfin la diffusion à l'intérieur du volume d'une source externe. L'absorption et la diffusion sortante sont des termes de déplétion, c'est-à-dire qu'ils affaiblissent le rayonnement lors de sa traversée dans le volume. L'émission et la diffusion entrante sont quant à elles des termes sources, c'est-à-dire qu'ils participent à l'augmentation de la luminance du rayonnement sortant. En considérant une luminance au point x allant dans la direction w et passant par un milieu d'une distance ds , quatre termes s'expriment :

- l'absorption :

$$-K_a(x)L_\lambda(x \rightarrow w)$$

avec K_a le **coefficient d'absorption**.

- l'émission :

$$E(x \rightarrow w)$$

où $E(x \rightarrow w)$ peut être exprimé $K_a(x)B_\lambda(T)$.

- la diffusion sortante :

$$-K_s(x)L_\lambda(x \rightarrow w)$$

avec K_s le **coefficient de diffusion**.

- la diffusion entrante :

$$K_s(x)L_{\lambda,i}(x \rightarrow w)$$

Où

$$\begin{aligned} L_{\lambda,i}(x \rightarrow w) &= \frac{1}{4\pi} \int_{\Omega} L_\lambda(\theta', \phi') \rho(\Omega', \Omega) d\Omega' \\ &= \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi L_\lambda(\theta', \phi') \rho(\Omega', \Omega) \sin\theta' d\theta' d\phi' \end{aligned} \quad (2.5)$$

Avec $p(\Omega, \Omega')$ qui correspond à **la fonction de phase** décrivant la probabilité de diffusion d'un rayonnement provenant de la direction (θ', ϕ') dans la direction (θ, ϕ) . Comme ceci, $L_i(x \rightarrow w)$ représente les contributions d'autres rayonnements parvenant jusqu'au volume et participant à l'augmentation de l'intensité du rayonnement étudié.

L'addition de ces quatre termes donne l'ETR. Cette équation, interprétée comme un bilan d'énergie, s'exprime alors :

$$\frac{dL_\lambda(\theta, \phi)}{ds} = K_a(\lambda)(-L_\lambda(\theta, \phi) + B_\lambda(T)) - K_s(\lambda)L_\lambda + \frac{K_s(\lambda)}{4\pi} \int_0^{2\pi} \int_0^\pi L_\lambda(\theta', \phi') p(\Omega, \Omega') \sin\theta' d\theta' d\phi' \quad (2.6)$$

Pour estimer une luminance spectrale au sommet de l'atmosphère dans une direction donnée (telle qu'elle serait mesurée par un radiomètre à bord d'un satellite), il est commun en météorologie d'appliquer plusieurs transformations (Kidder Vonder Haar 1995[96]). Ces transformations font intervenir plusieurs définitions, la première étant le **coefficient d'extinction de volume** K_e . Ce coefficient mesure la perte d'énergie du rayonnement électromagnétique lors de sa propagation dans un milieu. Il prend en compte les contributions venant de l'absorption et de la diffusion. Il est défini comme :

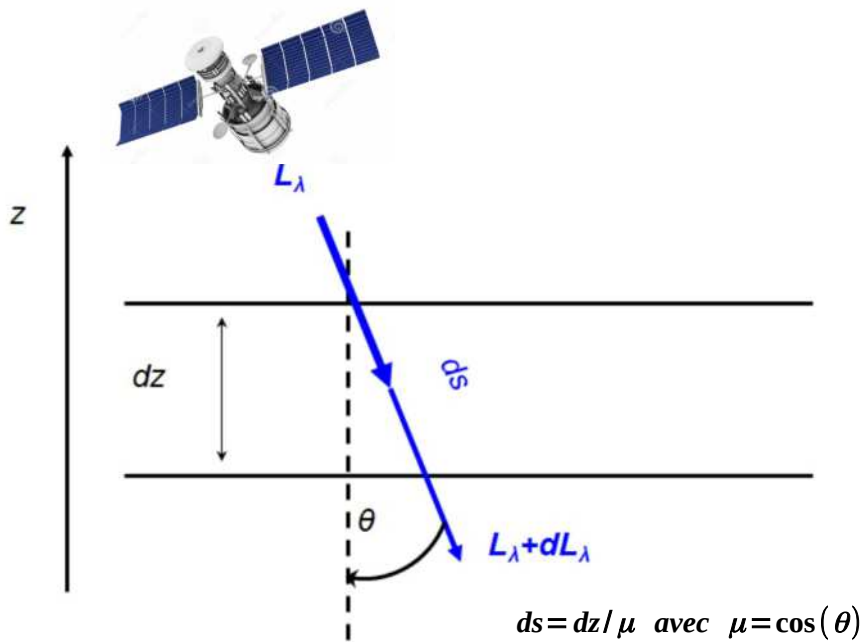


FIGURE 2.5 – Illustration du changement de variable entre ds et dz .

$$K_e(\lambda) = K_a(\lambda) + K_s(\lambda) \quad (2.7)$$

En remplaçant K_a et K_s par le coefficient d'extinction, l'ETR se réécrit :

$$\frac{dL_\lambda(\theta, \phi)}{ds} = -K_e(\lambda)L_\lambda(\theta, \phi) + K_a B_\lambda(T) + \frac{K_s(\lambda)}{4\pi} \int_0^{2\pi} \int_0^\pi L_\lambda(\theta', \phi') p(\Omega, \Omega') \sin\theta' d\theta' d\phi \quad (2.8)$$

Une autre définition est l'**albédo de diffusion simple** ω_λ . Il représente l'importance relative de la diffusion d'un milieu par rapport à son absorption. La valeur de l'albédo est comprise entre 0 (milieu totalement absorbant) et 1 (milieu totalement diffusant) et il s'exprime en fonction du coefficient d'extinction et du coefficient de diffusion :

$$\omega_\lambda = \frac{K_s(\lambda)}{K_e(\lambda)}$$

En multipliant l'ETR par $\frac{1}{K_e}$, l'albédo de diffusion simple est introduit dans celle-ci :

$$\frac{1}{K_e(\lambda)} \frac{dL_\lambda(\theta, \phi)}{ds} = -L_\lambda(\theta, \phi) + (1 - \omega_\lambda) B_\lambda(T) + \frac{\omega_\lambda}{4\pi} \int_0^{2\pi} \int_0^\pi L_\lambda(\theta', \phi') p(\Omega, \Omega') \sin\theta' d\theta' d\phi \quad (2.9)$$

On considère fréquemment l'approximation de 'plan-parallèle' qui consiste à négliger la courbure de la Terre. Cette approximation décompose l'atmosphère en différentes couches supposées homogènes sur l'horizontale en introduisant l'**épaisseur optique verticale**. Pour se faire, la distance oblique ds d'une couche d'atmosphère est remplacée par dz/μ avec $\mu = \cos(\theta)$ comme illustré sur la figure 2.5. Pour $\mu < 0$, le rayonnement est montant et pour $\mu > 0$, le rayonnement est descendant. Ainsi en notant $\delta_\lambda = \delta_\lambda(0, z)$ l'épaisseur optique verticale allant de 0 à l'altitude z , on obtient :

$$d\delta_\lambda(0, z) = K_e(\lambda) dz = K_e(\lambda) \cos\theta ds = K_e(\lambda) \mu ds$$

De ce fait, ds devient $\frac{d\delta_\lambda}{K_e(\lambda)\mu}$ et l'ETR s'écrit alors :

$$\mu \frac{dL_\lambda}{d\delta_\lambda} = -L_\lambda(\theta, \phi) + (1 - \omega_\lambda)B_\lambda(T) + \frac{\omega_\lambda}{4\pi} \int_0^{2\pi} \int_{-1}^1 L_\lambda(\mu', \phi') p(\Omega, \Omega') d\mu' d\phi \quad (2.10)$$

Bien que l'atmosphère ne soit pas homogène horizontalement, l'utilisation de cette approximation pour résoudre l'ETR, se justifie pour les raisons suivantes (Petty 2006[97]) : elle est la seule géométrie semi-réaliste permettant une résolution 'simple' de l'ETR, elle est de plus une approximation très raisonnable dans de nombreuses situations météorologiques de par la stratification verticale de l'atmosphère (gradient verticaux très supérieurs aux gradients horizontaux), et lorsque ce n'est pas le cas, elle est tout de même utilisée car elle est un bon compromis compte tenu des difficultés pour traiter les inhomogénéités tridimensionnelles.

Finalement, en remplaçant toutes les contributions de la diffusion des autres directions vers la direction θ examinée ainsi que l'émission propre du volume d'intérêt par un **terme source** J , on obtient l'équation :

$$\mu \frac{dL_\lambda}{d\delta_\lambda} = -L_\lambda(\theta, \phi) + J(\theta, \phi)$$

Avec J qui vaut :

$$J(\theta, \phi) = (1 - \omega_\lambda)B_\lambda(T) + \frac{\omega_\lambda}{4\pi} \int_0^{2\pi} \int_{-1}^1 L_\lambda(\mu', \phi') p(\Omega, \Omega') d\mu' d\phi \quad (2.11)$$

Pour résoudre l'ETR, il est important de connaître différentes propriétés sur les nuages. Ces propriétés peuvent être classées en deux catégories : les **propriétés macrophysiques** et les **propriétés microphysiques**. Parmi les propriétés macrophysiques, on trouve par exemple l'altitude du sommet des nuages, leur épaisseur ou encore leur variabilité sous maille. Les propriétés microphysiques du nuage regroupent quant à elles les formes des hydrométéores solides et liquides, leurs dimensions, leurs densités, leurs concentrations ainsi que leurs propriétés diélectriques. Ces propriétés permettent par la suite, à l'aide d'un opérateur d'observation basé sur le transfert radiatif, de simuler à une fréquence donnée des températures de brillance à partir d'un profil atmosphérique comme illustré sur la figure 2.6. Ainsi, il devient possible de comparer les modèles aux observations.

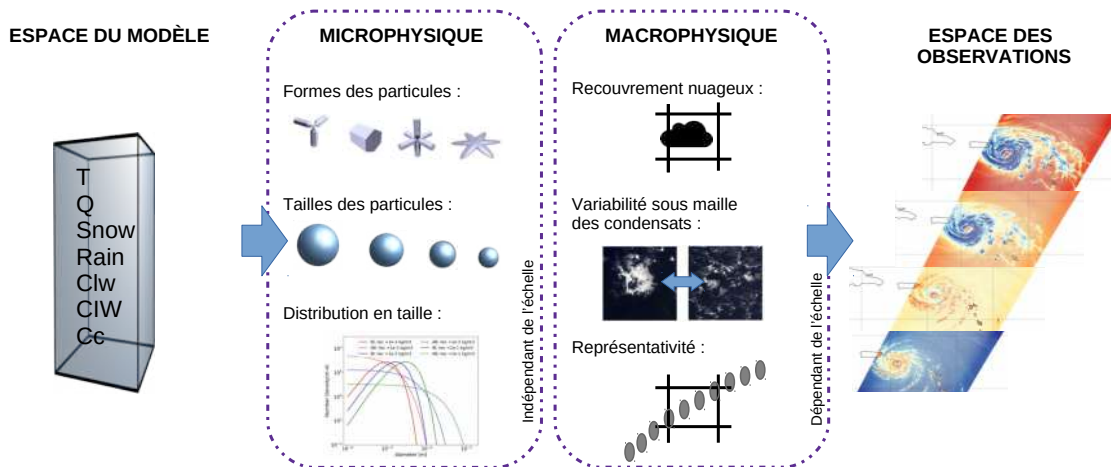


FIGURE 2.6 – Schéma illustrant le passage de l'espace du modèle à l'espace des observations à l'aide des propriétés microphysiques et macrophysiques adaptées aux luminances micro-ondes.

Source: Schéma adapté d'un schéma présenté par Marc Fielding

2.1.4 Un code de transfert radiatif : RTTOV-SCATT

Plusieurs codes de transfert radiatif ont été développés pour résoudre numériquement l'ETR. Chacun de ces codes utilise différentes hypothèses et méthodes de résolution afin de pouvoir calculer les luminances. La plupart d'entre eux applique l'approximation de plan parallèle qui ne considère pas les gradients horizontaux des variables atmosphériques ni la courbure de la Terre. D'autres, au contraire, considèrent l'aspect 3D des différentes quantités présentes dans l'atmosphère comme les méthodes de Monte-Carlo (Petty 1994[98], Liu et al. 1996[99], Szczap et al. 2021[100]). Cette technique, qui exige des calculs beaucoup plus longs et complexes, n'est pas privilégiée pour les besoins de la PNT, qui demande des temps de calcul rapides pour pouvoir être appliquée à des millions de profils. Parmi les méthodes les plus connues on recense les méthodes d'Eddington (Joseph et al. 1976[101], Kummerow 1993[102] et Bauer et al. 2006[103]), les méthodes des ordonnées discrètes (Liu 1998[104], Liu et Weng 2002[105]) ou encore les méthodes d'ordre successif de diffusion (SOS en anglais pour Successive Order of Scattering, Weinman et Guetter 1977[106], Wendisch et von Hoyningen-Huene 1991[107]).

À Météo-France, le code utilisé opérationnellement pour résoudre l'ETR est un code de transfert radiatif rapide se nommant RTTOV. La première version de RTTOV utilisée dans les modèles de PNT a été développée par J.Eyre (1991[43]) et est maintenant un des logiciels dont le développement est financé par EUMETSAT, l'organisme Européen de gestion des satellites météorologiques opérationnels. Depuis, de nombreuses mises à jour ont été effectuées par des organismes tels que le UK Met Office, Météo-France et le CEPMMT. La version 12.3 est utilisée en opérationnel à Météo-France. Elle est présentée dans Saunders et al (2018)[12]. RTTOV est un code numérique écrit en langage Fortran 90 et qui possède des interfaces en Python, C et C++. Une version dérivée de RTTOV est utilisée afin de traiter les micro-ondes en ciel diffusant : RTTOV-SCATT. Le nom de cette version provient de l'utilisation d'un solveur exploitant la diffusion (scattering en anglais) pour calculer les luminances micro-ondes en présence de nuages et de précipitations. Nous allons à présent expliquer comment RTTOV-SCATT résout l'ETR.

Revenons d'abord à l'écriture de l'ETR :

$$\mu \frac{dL_\lambda}{d\delta_\lambda} = -L_\lambda(\theta, \phi) + J(\theta, \phi)$$

Dans RTTOV, la luminance spectrale observée depuis un satellite est simulée à l'aide de deux contributions : une partie en ciel clair et une partie en ciel nuageux. L'atmosphère est ainsi décomposée en deux colonnes pondérées par **la fraction nuageuse effective** C_{frac} caractérisant le taux de nuages au sein d'une maille. Le calcul pour la colonne en ciel clair néglige les effets de la diffusion et ne considère que les gaz, cette colonne est représentée par la luminance L_λ^{clear} . Le calcul pour la colonne en ciel nuageux prend en compte les effets de diffusion par les nuages et les précipitations, cette colonne est représentée par la luminance L_λ^{cloud} . La luminance totale simulée parvenant au satellite prend alors l'erreur de représentativité en compte et est égale à :

$$L_\lambda = C^{frac} L_\lambda^{cloud} + (1 - C^{frac}) L_\lambda^{clr}$$

Il existe communément quatre manières de calculer la fraction nuageuse effective comme on peut voir sur la figure 2.7 :

- **La méthode de recouvrement maximal :** C_{max}^{frac}

Les nuages du profil atmosphérique vertical sont alignés sur une seule colonne. La fraction nuageuse de la couche ayant la valeur la plus élevée sera considérée comme celle vue au niveau du satellite (Morcrette et Fouquart 1986[108]).

$$C_{max}^{frac} = \max(C_1, \dots, C_N)$$

- **La méthode de recouvrement aléatoire :** C_{rand}^{frac}

Les nuages du profil atmosphérique vertical sont disposés de manière aléatoire entre

chaque couche (Manabe et Stricker 1964[109]).

$$C_{rand}^{frac} = 1 - \prod_{i=1}^N (1 - C_i)$$

• **La méthode de recouvrement maximal aléatoire :** C_{ma}^{frac}

Les nuages du profil vertical se recouvrent au maximum pour les couches nuageuses adjacentes sur la verticale et sont disposés aléatoirement pour les nuages séparés par une zone de ciel clair (Geleyn et al. 1979[110]).

$$C_{ma}^{frac} = 1 - \prod_{i=1}^{N-1} \frac{1 - \max(C_i, C_{i+1})}{1 - C_i}$$

• **La méthode de recouvrement exponentiel aléatoire :** C_{exp}^{frac}

Cette méthode suppose que le recouvrement des nuages qui se chevauchent diminue de manière exponentielle d'une fraction nuageuse maximale vers une fraction nuageuse aléatoire. La méthode de recouvrement exponentiel aléatoire a été élaborée par Hogan and Illingworth (2000[111]) afin de remplacer l'hypothèse de recouvrement maximal faite dans la plupart des modèles qui ne correspond pas aux recouvrements observés par les radars. La fraction nuageuse exponentielle aléatoire est ainsi exprimée à l'aide d'une pondération entre C_{max}^{frac} et C_{rand}^{frac} tel que :

$$C_{exp}^{frac} = \alpha_{i,j} C_{max}^{frac} + (1 - \alpha_{i,j}) C_{rand}^{frac}$$

Avec $\alpha_{i,j}$ le paramètre de recouvrement entre deux couches atmosphériques i et j valant :

$$\alpha_{i,j} = \exp\left(\frac{\Delta z}{\Delta z_o}\right)$$

Avec δz la distance en kilomètres entre les couches atmosphériques i et j et δz_o la longueur de décorrélation provenant de diagnostics a priori calculés en fonction des résolutions temporelle et verticale d'un modèle. Ainsi plus $\alpha_{i,j}$ est proche de 1 plus C_{exp}^{frac} va tendre vers une fraction nuageuse maximale et plus il est proche de 0 plus celui-ci va tendre vers une fraction nuageuse aléatoire.

La première fraction nuageuse à avoir été implémentée dans RTTOV-SCATT est celle considérant l'hypothèse de recouvrement maximal, soit C_{max}^{frac} (Bauer et al. 2006[103]). Ainsi, la fraction nuageuse maximale trouvée au sein des niveaux verticaux était attribuée à l'ensemble du profil, amenant à des poids trop forts pour certains profils, notamment ceux en zones précipitantes et/ou nuageuses ce qui pouvait conduire à des biais dans les simulations. Elle a été remplacée par la fraction nuageuse proposée dans Geer et al. (2009[112]). Cette fraction consiste en une moyenne pondérée par les hydrométéores du profil vertical et s'exprime comme :

$$C^{frac} = \frac{\sum_{i=0}^N (r_i + ciw_i + clw_i + sp_i) \Delta z_i C_i}{\sum_{i=0}^N (r_i + ciw_i + clw_i + sp_i) \Delta z_i}$$

Avec r_i , ciw_i , clw_i et sp_i respectivement les concentrations massiques en $kg.m^{-3}$ de pluie, de glace nuageuse, d'eau liquide nuageuse et de neige, C_i la fraction nuageuse et Δz_i l'épaisseur de la couche atmosphérique au niveau i .

La luminance négligeant les effets de diffusion L_{λ}^{clr} est estimée à l'aide du code RTTOV. Quant à la luminance prenant en compte les effets de diffusion L_{λ}^{cloud} , elle est estimée avec le module RTTOV-SCATT. Nous présentons maintenant brièvement la résolution de la luminance L_{λ}^{clear} et ensuite celle de L_{λ}^{cloud} .

Afin de simuler les luminances sans effet de diffusion, le code de transfert radiatif RTTOV nécessite la paramétrisation de l'absorption gazeuse. Cette paramétrisation renseigne l'épaisseur optique d'absorption δ pour un canal d'instrument i et une couche atmosphérique j , dont l'expression est définie telle que :

Quatre hypothèses de recouvrement :

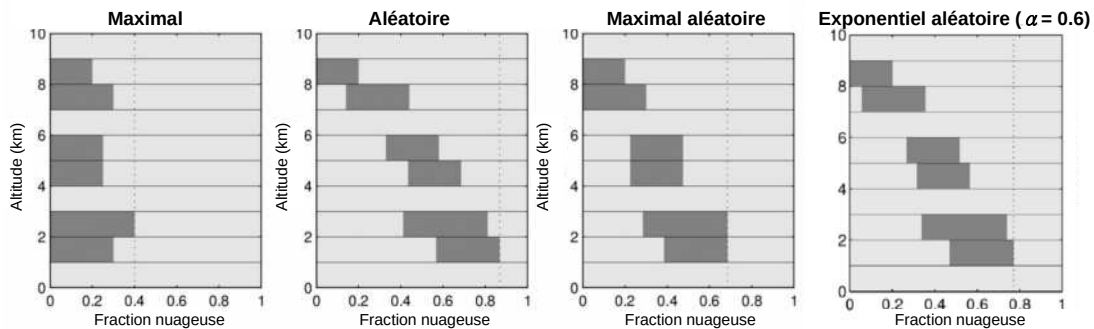


FIGURE 2.7 – Exemples de différentes méthodes pour calculer la fraction nuageuse totale effective.

Source: Présentation de Hogan et al. 2014, <http://www.met.reading.ac.uk/~swrhgnrj/presentations/>

$$\delta_{abs,i,j} = \delta_{abs,i,j-1} + \sum_{k=1}^K a_{i,j,k} X_{i,k}$$

Avec $a_{i,j,k}$ calculés grâce à des régressions linéaires à coefficients multiples sur un jeu d'entraînement à l'aide d'un modèle de transfert radiatif raie par raie et d'un jeu de profils X représentant la variabilité de l'atmosphère. Pour les micro-ondes, on exploite le modèle AMSUTRAN prenant ses paramètres spectroscopiques de la version 1989 du modèle de propagation des ondes millimétriques de Liebe et al. 1989[113]. Ainsi que 52 profils atmosphériques fournis par le CEPMMT sur 54 niveaux verticaux entre 0.05 et 1050 hPa (Chevallier et al. 2006[114]). Chaque gaz considère un nombre K de prédicteurs. La version 7 des prédicteurs est utilisée dans ce manuscrit, avec K valant 10, 15 et 11 respectivement pour les gaz fixes, l'eau et l'ozone. Les coefficients sont uniques pour chaque instrument considéré car ils dépendent de la réponse spectrale de ces derniers.

RTTOV nécessite également des paramétrisations de surface. Celles-ci sont définies grâce à des modèles de surface ou des atlas pour chaque domaine de longueurs d'ondes si elles ne sont pas renseignées par l'utilisateur. Pour les micro-ondes, les modèles FASTEM (FAST microwave ocean Emissivity Model) versions 4, 5 et 6 sont exploités au dessus des surfaces océaniques (Liu et al. 2010[115], Bormann et al. 2012[116], Kazumori et English 2015[117]). Au dessus des surfaces continentales, le modèle TELSEM (Tool to Estimate Land Surface Emissivities at Microwave frequencies, Aires et al. 2011[118]) et l'atlas CNRM (Karbou et al. 2006[71], 2010[119]) peuvent être utilisés.

Ces paramétrisations permettent de linéariser l'opérateur d'observation et ainsi d'exploiter les modèles linéaire-tangent, adjoint et jacobien de ce dernier, nécessaires à l'assimilation déterministe variationnelle.

Ces paramétrisations renseignées, la luminance L_{λ}^{clear} peut maintenant être calculée. Pour cela, on note $\delta^* = \delta_{\lambda}(0, sat)$ l'épaisseur optique verticale allant de 0 (surface terrestre) à l'altitude du satellite, l'expression de la luminance en ciel clair L_{λ}^{clear} est :

$$L_{\lambda}^{clear}(\delta^*; \theta, \phi) = \epsilon_0 B(T_0) \tau(\delta^*) + \int_0^{\delta^*} J(\delta; \theta, \phi) e^{-\frac{\delta}{\mu}} \frac{d\delta}{\mu} + (1 - \epsilon_0) \tau(\delta^*) \int_{\delta^*}^{\infty} J(\delta; \theta, \phi) e^{-\frac{(\delta^* - \delta)}{\mu}} \frac{d\delta}{\mu} \quad (2.12)$$

Avec :

- $\epsilon_0 B(T_0) \tau(\delta^*)$ la contribution de la surface,

- $\int_{\delta^*}^0 J(\delta; \theta, \phi) e^{-\frac{\delta}{\mu}} \frac{d\delta}{\mu}$ la contribution du rayonnement ascendant émis par l'atmosphère,
- et $(1 - \epsilon_0) \tau(\delta^*; \theta, \phi) \int_{\delta^*}^{\infty} J(\delta; \theta, \phi) e^{-(\delta^* - \delta')/\mu} \frac{d\delta'}{\mu}$ la contribution du rayonnement descendant émis par l'atmosphère.

Ici, ϵ_0 représente l'émissivité de surface et τ la transmittance atmosphérique totale s'exprimant :

$$\tau(\delta^*; \theta, \phi) = \int_0^{\delta^*} e^{-\delta'/\mu} \frac{d\delta'}{\mu}$$

Dans la version de RTTOV-SCATT exploitée au sein de cette thèse, seuls les effets de polarisation provenant de l'émissivité de surface sont pris en compte. L'émissivité de surface joue un rôle important dans la résolution de cette équation. C'est pourquoi il est essentiel d'estimer sa valeur précisément. Elle varie selon certains critères : la longueur d'onde du rayonnement, la rugosité et le type de surface. À Météo-France, l'émissivité micro-ondes pour les continents et la glace de mer est estimée avec la méthode de Karbou et al. (2006)[71]. L'émissivité des surfaces océaniques est paramétrisée en fonction principalement de l'intensité du vent par le modèle FASTEM (Liu et al. 2010[115]). La luminance est ici calculée à l'aide de la colonne du modèle la plus proche du pixel au sol tout en prenant en compte l'angle de visée de l'instrument simulé au sein de RTTOV-SCATT par le biais d'une dépendance géométrique de la profondeur optique. L'ébauche est ainsi supposée comme localement uniforme horizontalement et sphériquement symétrique. Des études réalisées à Environnement et Changement Climatique Canada (ECCC) ont montré des améliorations des prévisions à courtes et longues échéances lorsque cette hypothèse n'était pas réalisée et que la trajectoire de l'angle de visée était prise en compte (Shahabadi et al. 2018[120]). De plus, une orientation aléatoire des hydrométéores est supposée afin de négliger les effets de polarisation qui résulteraient d'orientation spécifique des hydrométéores. Néanmoins une modélisation simplifiée des effets de polarisation est disponible dans la version 13 de RTTOV-SCATT et pourra être utilisée dans de futures études.

Passons maintenant à la méthode de résolution utilisée par RTTOV-SCATT pour calculer L_λ^{cloud} . Ce code de transfert radiatif utilise la méthode d'Eddington consistant à approcher la luminance sous forme d'un développement en polynômes de Legendre à un ordre k (pour la prise en compte des anisotropies). Ici la luminance est approchée à l'ordre 1, Kummerow (1993[102]) ayant montré qu'un ordre supérieur pour la simulation des luminances micro-ondes au sommet de l'atmosphère n'améliorait pas significativement la précision des calculs. La luminance nuageuse est décomposée en une luminance isotrope² L_0 et une luminance anisotrope³ μL_1 où, pour rappel, $\mu = \cos(\theta)$ avec θ l'angle zénithal du rayonnement incident :

$$L(\mu) = L_0 + \mu L_1 \quad (2.13)$$

En injectant cette formule dans $J(\theta, \phi)$ on obtient :

$$J(\theta, \phi) = (1 - \omega_\lambda) B_\lambda(T) + \omega_\lambda (L_0 + g \mu L_1) \quad (2.14)$$

Le terme g représentant le **coefficient d'asymétrie** apparaît alors et nous renseigne sur la direction de la diffusion du rayonnement. Cette variable est donnée par :

$$g = \frac{1}{4\pi} \int_{4\pi} p(\Theta) \cos(\Theta) d\Omega$$

- Si $g = 0$ la diffusion est isotrope,
- si $g < 0$ la diffusion se fait principalement vers l'arrière,
- si $g > 0$ la diffusion se fait principalement vers l'avant.

2. qui présente les mêmes caractéristiques dans toutes les directions

3. qui présente des propriétés physiques variant selon la direction considérée

L'angle de diffusion $\cos(\Theta)$ représente l'angle entre les directions incidente et de diffusion, et vaut :

$$\cos(\Theta) = \cos(\theta') \cos(\theta) + \sqrt{(1 - \mu^2)(1 - \mu'^2)} \cos(\phi' - \phi)$$

La fonction de phase est exprimée en fonction de l'angle de diffusion $\cos(\Theta)$ au moyen d'un développement en polynômes de Legendre qui s'écrit à l'ordre N :

$$p(\cos \Theta) = \sum_{i=0}^N \beta_i P_i(\cos \Theta)$$

Avec $P_i(\cdot)$ des polynômes de Legendre et β_i les coefficients d'expansions valant :

$$\beta_i = \frac{2i + 1}{2} \int_{-1}^1 P_i(\cos \Theta) p(\cos \Theta) d \cos \Theta$$

Comme indiqué précédemment pour la luminance, l'approximation de la fonction de phase est suffisante à l'ordre 1. L'approximation plan parallèle est ensuite réalisée afin de considérer l'atmosphère comme homogène horizontalement et de négliger la dépendance azimutale. Et on obtient finalement :

$$p(\cos \Theta) = 1 + 3g \cos \Theta = 1 + 3g\mu\mu' \quad (2.15)$$

En exprimant les valeurs des variables de l'équation 2.11 en fonction de l'altitude z , on obtient alors un système d'équations tel que :

$$\begin{cases} \frac{dL_0(z)}{dz} = K_{e\lambda}(1 - \omega_\lambda g)L_1(z) \\ \frac{dL_1(z)}{dz} = 3K_{e\lambda}(1 - \omega_\lambda)(L_0(z) - B_\lambda(T(z))) \end{cases} \quad (2.16)$$

La diffusion multiple au sein de l'atmosphère, correspondant à la propagation des rayonnements lors de leurs rencontres avec les différents diffuseurs, est très complexe et n'est pour l'instant pas envisagée au sein de RTTOV-SCATT. Cette diffusion est donc prise en compte de manière très simplifiée dans ce code grâce à l'approximation "two-stream". Cette approximation suppose que le rayonnement est la combinaison de deux flux : un flux ascendant et un flux descendant. Des travaux proposant une modélisation plus aboutie de la diffusion multiple existent néanmoins et pourront être exploitées dans les prochaines années (Battaglia et al. 2006[121], Battaglia et al. 2015[122]).

Dans l'approximation "two-stream", le système d'équations (2.16) peut être discrétisé et résolu numériquement à partir d'une condition à la limite supérieure (rayonnement cosmique à 2,7 K) et à la limite inférieure (fournie par la température et l'émissivité de surface). Cette résolution nécessite toutefois de renseigner chaque couche atmosphérique avec les quantités suivantes :

- ω_λ : l'albédo de diffusion simple,
- $K_e(\lambda)$: le coefficient d'extinction,
- g : le facteur d'asymétrie.

Ces paramètres définissent les propriétés radiatives des gaz et hydrométéores. Afin de définir ces paramètres pour un ensemble de particules, il est nécessaire de connaître leurs propriétés individuelles. Pour les hydrométéores, ces propriétés dépendent principalement de leur taille D , qui correspond au diamètre dans le cas de particules sphériques. Les propriétés radiatives pour un ensemble de particules sont calculées en intégrant les propriétés individuelles sur leur PSD $N(D)$, qui est pour rappel la distribution

en taille. La littérature en propose de nombreuses formes en fonction des types d'hydrométéores et de nuages. On peut par exemple citer la distribution de Marshall et Palmer (1948[123]) qui a été définie à l'origine pour les gouttes de pluie tout comme la distribution d'Abel et Boutle (2012[124]). En illustration, la figure 2.8 montre les deux distributions pour un contenu en pluie allant de 1×10^{-3} à $1 \times 10^{-1} \text{ kg/m}^3$, correspondantes de Marshall et Palmer et d'Abel et Boutle. On peut voir que la distribution de Marshall et Palmer propose davantage de gouttes de petit diamètre que celle d'Abel et Boutle pour ce contenu. La distribution Gamma modifiée (voir Petty et Huang, 2011[66]) est quant à elle souvent utilisée pour la représentation des hydrométéores présents dans les nuages, tels que l'eau nuageuse liquide ou l'eau nuageuse glacée. Les distributions de McFarquhar et Heymsfield (1997[125]), Field et al. (2005[126]), Field et al. (2007[127]), Heymsfield et al. (2013[128]) sont davantage utilisées pour la représentation des hydrométéores solides. La distribution de McFarquhar et Heymsfield (1997[125]) est particulièrement exploitée dans la communauté des mesures micro-ondes au limbe et convient pour la simulation des hydrométéores solides présents dans les nuages tels que les cumulonimbus et les cirrus (Liu et al. 2008[11], Eriksson et al. 2007[129]). Néanmoins, la distribution de Field et al. (2007[127]) est plus largement utilisée, en particulier par la communauté travaillant sur l'assimilation de données micro-ondes (Kulie et al. 2010[7], Geer et Baordo 2014[1]).

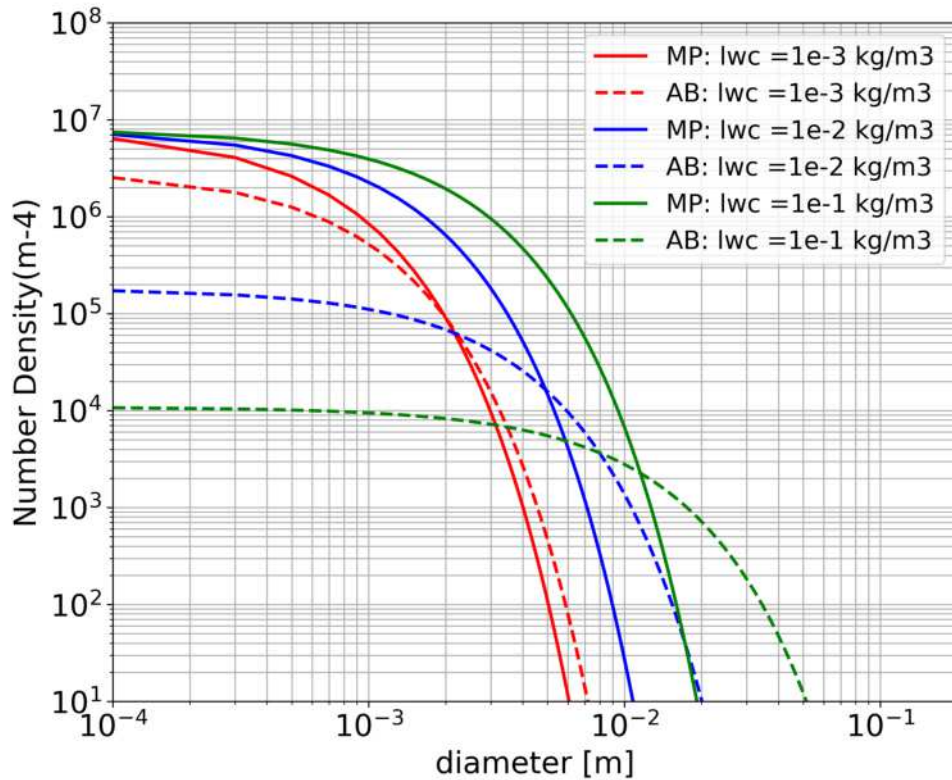


FIGURE 2.8 – Exemples de distributions en goutte pour des contenus de pluie allant de 1×10^{-3} à $1 \times 10^{-1} \text{ kg/m}^3$. Les lignes pleines représentent des exemples de distributions de Marshall et Palmer[123] et les lignes pointillées de distributions de Abel et Boutle[124].

Source: Mangla et al. 2023[130]

L'albédo de diffusion simple, le coefficient d'extinction et le facteur d'asymétrie sont généralement calculés à partir des **sections efficaces d'extinction** C_e , de **diffusion** C_s et d'**absorption** C_a qui représentent leurs propriétés individuelles exprimées comme une surface (Bauer, 2004[131]) :

$$\omega = \frac{\pi}{4} \int_0^{\infty} \frac{C_s(D)}{C_e(D)} N(D) D^2 dD$$

$$K_e = \frac{\pi}{4} \int_0^\infty C_e(D) N(D) D^2 dD$$

$$g = \frac{\pi}{4} \int_0^\infty \frac{\overline{\cos(\theta)}}{C_s(D)} N(D) D^2 dD$$

Où $\overline{\cos(\theta)}$ l'angle de diffusion moyenné à la température T et à la fréquence ν . Les sections efficaces introduites ci-dessus dépendent des propriétés diélectriques des particules que l'on peut exprimer sous la forme d'un **indice de réfraction complexe**. Cet indice, noté m , se compose d'une partie réelle et d'une partie imaginaire :

$$m = n + ik'$$

où n est l'**indice de réfraction**, qui exprime le changement de vitesse de propagation du rayonnement dans un milieu par rapport au vide (e.g. cette valeur est de d'ordre de 1,33 pour des gouttes d'eau), et k' décrit l'absorption du rayonnement par le milieu.

2.2 Calcul des propriétés radiatives des hydrométéores

Afin de calculer les propriétés radiatives indispensables à la résolution de l'ETR, il est nécessaire de spécifier des paramètres tels que la distribution en taille, la densité, la composition (e.g. eau, air, glace) et la forme des hydrométéores. Les hydrométéores liquides sont généralement représentés comme des sphères ou comme des formes assez simples telles que des sphéroïdes aplatis, afin de ressembler à la forme d'une goutte d'eau (Brandes et al. 2002[27], Augros et al. 2016[28], Ekelund et al. 2020[29]). En ce qui concerne les hydrométéores solides, une grande diversité de formes existe et il est donc plus complexe d'en calculer les propriétés radiatives. Nous allons tout d'abord expliquer comment les propriétés radiatives sont calculées selon les régimes de diffusion que nous introduirons au préalable. Ensuite nous présenterons la diversité des formes possibles pour les hydrométéores solides ainsi que les méthodes appliquées aux formes complexes pour déterminer leurs propriétés radiatives. Enfin nous proposerons un inventaire des bases de données les plus connues répertoriant les propriétés radiatives des hydrométéores solides.

2.2.1 Les régimes de diffusion

Comme vu précédemment, la résolution de l'ETR exige de connaître les propriétés radiatives traduisant la diffusion et l'absorption du rayonnement. Dans la suite du manuscrit nous remplacerons le nom de propriétés radiatives par l'acronyme SSP de par leur appellation en anglais 'Single Scattering Properties'. Ces SSP sont : *le coefficient d'absorption K_a , le coefficient de diffusion K_s et la fonction de phase $P(\Omega, \Omega')$* . Grâce à ces variables il est alors possible de déterminer l'albédo de diffusion simple w_λ ainsi que le facteur d'asymétrie g (résumant les effets de la fonction de phase dans l'approximation d'Eddington). Ces valeurs dépendent du régime de diffusion qui est déterminé par la taille de particules et la longueur d'onde du rayonnement.

C'est ainsi que l'on définit un critère appelé **paramètre de taille**, noté x , valant :

$$x = \frac{2\pi r}{\lambda}$$

où λ est la longueur d'onde du rayonnement et r le rayon de la particule. On distingue alors quatre régimes principaux :

- $x \ll 1$, les effets de diffusion sont négligeables,
- $x < 1$, le régime de diffusion de Rayleigh est appliqué,

- $x \simeq 1$, le régime de diffusion de Mie est appliqué,
- $x \gg 1$, le régime de diffusion de l'optique géométrique est appliqué.

Plus spécifiquement, les fréquences micro-ondes sont sensibles aux différents types d'hydrométéores présents dans l'atmosphère. Les basses fréquences micro-ondes, i.e. en dessous de 50 GHz, sont particulièrement sensibles aux phénomènes d'absorption et d'émission par les hydrométéores liquides. Plus on se rapproche des hautes fréquences, plus les micro-ondes vont être sensibles au phénomène de diffusion par les hydrométéores solides. Le schéma présenté sur la figure 2.9 résume le choix du régime en fonction de la longueur d'onde λ du rayonnement et du rayon r de la particule. Le rayonnement adopte un comportement particulier lors de la rencontre avec une particule selon le régime de diffusion : par exemple, si le rayonnement suit le régime de Rayleigh, il est diffusé de manière isotrope contrairement au régime de Mie où le rayonnement est davantage diffusé vers l'avant. Pour les fréquences micro-ondes, les hydrométéores suivent soit le régime de Mie soit le régime de Rayleigh comme on peut le voir figure 2.9. Nous allons donc maintenant décrire ces deux régimes.

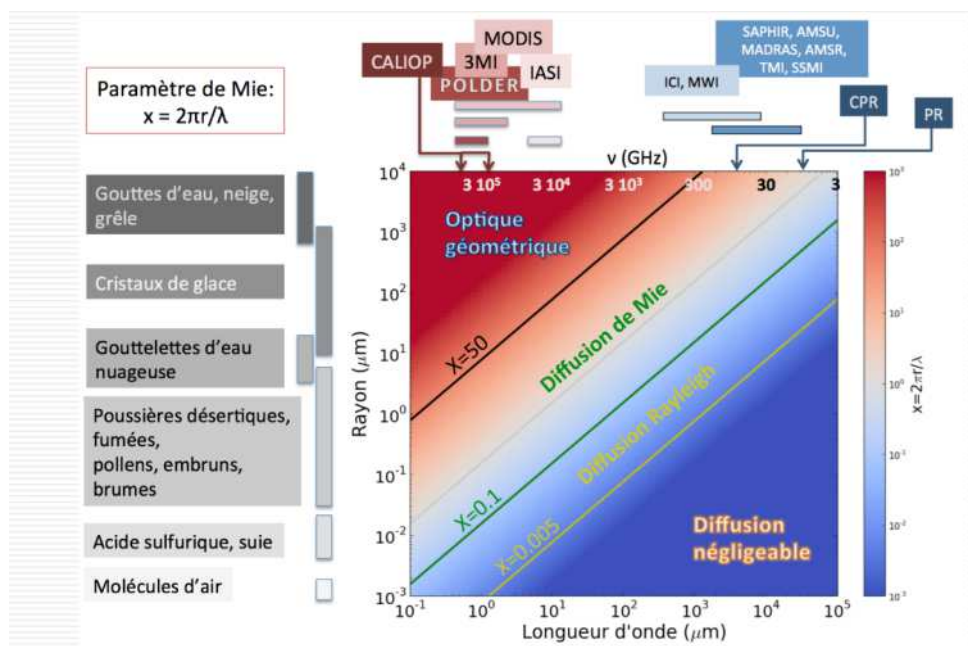


FIGURE 2.9 – Graphique représentant les différents régimes de diffusion en fonction de la longueur d'onde du rayonnement et de la taille des particules considérées provenant d'une présentation de Dubuisson (2018)

Source: [http://www.meteo.fr/cic/meetings/2018/AMA/presentations/AMA_1302/7-\\$\\$\\$20AMA_2018-Dubuisson.pdf](http://www.meteo.fr/cic/meetings/2018/AMA/presentations/AMA_1302/7-$$$20AMA_2018-Dubuisson.pdf)

La théorie de Mie (Mie, 1908[132]), établie par Gustav Mie en 1908, s'applique lorsque le paramètre de taille x est compris entre 0.1 et 50 pour des particules de forme sphérique. Cette théorie s'appuie sur la résolution des équations de Maxwell réduites à des équations d'Helmholtz pour la diffraction du rayonnement électromagnétique au sein de particules sphériques :

$$\begin{cases} \nabla^2 E + k^2 E = 0 \\ \nabla^2 H + k^2 H = 0 \end{cases} \quad (2.17)$$

Avec pour rappel E et H respectivement le champ électrique et le champ magnétique, et k le nombre d'onde du rayonnement (i.e. $k = 2\pi/\lambda$). En réécrivant cette équation en coordonnées sphériques, on obtient l'équation suivante dont la résolution s'effectue à l'aide d'une méthode de séparation de variables :

$$\nabla^2 \Psi + \left(\frac{2\pi m}{\lambda}\right)^2 \Psi = 0$$

Hydrométéores	Rayon r	Paramètre de taille x	Régime
Gouttelette d'eau nuageuse	5-50 μm	0.003 - 0.03	Rayleigh
Pluie fine	$\sim 100 \mu m$	0.06	Rayleigh
Goutte de pluie	0.1 - 3 mm	0.06 - 1.8	Mie
Cristaux de glace	10 - 100 μm	0.006-0.06	Rayleigh
Neige	1-10 mm	0.6 - 6	Mie
Grêlon	$\sim 10 mm$	6	Mie

TABLE 2.2 – Régime de diffusion selon plusieurs hydrométéores à la fréquence 30 GHz.

Source: A. Geer ECMWF Training Course (2020).

Avec Ψ la fonction scalaire décrivant l'équation 2.17. Dans cette théorie, les sections efficaces C_e , C_s et C_a ici sont exprimées par des sommes de séries.

Pour des tailles de particules petites devant la longueur d'onde, le régime de **diffusion de Rayleigh** (Rayleigh, 1914[133]; Van De Hulst, 1957[134]) est prescrit. Elle peut être considérée comme une approximation de la théorie de Mie quand x est inférieur à 0.1. Les différentes sections efficaces pour le régime de diffusion de Rayleigh s'écrivent :

$$\begin{aligned}
 C_s &= \frac{8x^4}{3} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 \\
 C_a &= 4xIm \left[\frac{m^2 - 1}{m^2 + 2} \right] \\
 C_e &= C_s + C_a
 \end{aligned}
 \tag{2.18}$$

Finalement la fonction de phase s'exprime comme :

$$P(\Theta) = \frac{3}{4} [1 + \cos^2(\Theta)]$$

Maintenant que ces régimes de diffusion ont été présentés, le tableau 2.2 illustre les régimes de diffusion suivis selon le type d'hydrométéore à la fréquence micro-ondes 30 GHz. Il apparaît que pour les hydrométéores liquides le régime de Rayleigh peut être appliqué, contrairement aux hydrométéores solides où le régime de Mie doit être considéré. Cependant, la théorie de Mie est valable seulement pour des particules de forme sphérique, et les cristaux de glace et la neige possèdent de très grandes variétés de formes très éloignées des sphères. Nous allons premièrement voir quels types de formes existent dans la nature et comment celles-ci sont répertoriés. Ensuite nous listerons quelques méthodes utilisées pour traiter les particules dont les SSPs ne peuvent être calculées de manière analytique à cause de leur complexité.

Les formes des hydrométéores solides

Afin de répertorier les formes de particules existantes dans la réalité, différents moyens sont exploités. Le premier consiste à effectuer des campagnes de mesures in situ grâce à des avions de recherche (Lawson et al. 1998[135], Lawson et al. 2001[136], Heymsfield et al. 2002[137]). Ce type de mesures permet d'obtenir des informations sur l'évolution des particules au sein de l'atmosphère et ainsi d'en étudier leurs tailles et leurs formes. Un autre moyen, moins coûteux que le premier, consiste à utiliser les mesures au sol : comme des caméras à angles multiples (Garrett et al. 2012[138]) ou des imageurs vidéo (Newman et al. 2009[139]). Certains de ces instruments, en plus de mesurer la taille et la forme des particules, calculent la vitesse de chute des particules. Cependant, dû à leur position au sol, ces instruments permettent seulement d'obtenir des informations sur les hydrométéores précipitants. Une dernière approche consiste à réaliser des mesures en laboratoire. Il est ainsi possible de faire croître des particules de glace dans diverses conditions et d'obtenir des informations détaillées sur celles-ci.

Grâce aux campagnes de mesure et aux mesures réalisées en laboratoire, des dépendances ont été établies entre des formes de particules et des conditions atmosphériques (température et sursaturation par rapport à la glace). Une des classifications les plus complètes réalisées à l'aide d'observations in-situ et de mesures en laboratoire a été proposée par Bailey et Hallett (2009[31], figure 2.10). Dans cette classification, on retrouve plusieurs formes de particules telles que les plaquettes (sector snowflake, dendrite snowflake, thin plate et thick plate), les colonnes (solid column, block column et needle) et les étoiles (rosette bullet) par exemple. On remarque que pour des taux élevés de sursaturation par rapport à la glace et des températures comprises entre 0 et -40°C , il existe davantage de particules de type 'plaques' tandis que pour les autres cas les particules proches de colonnes sont plus nombreuses.

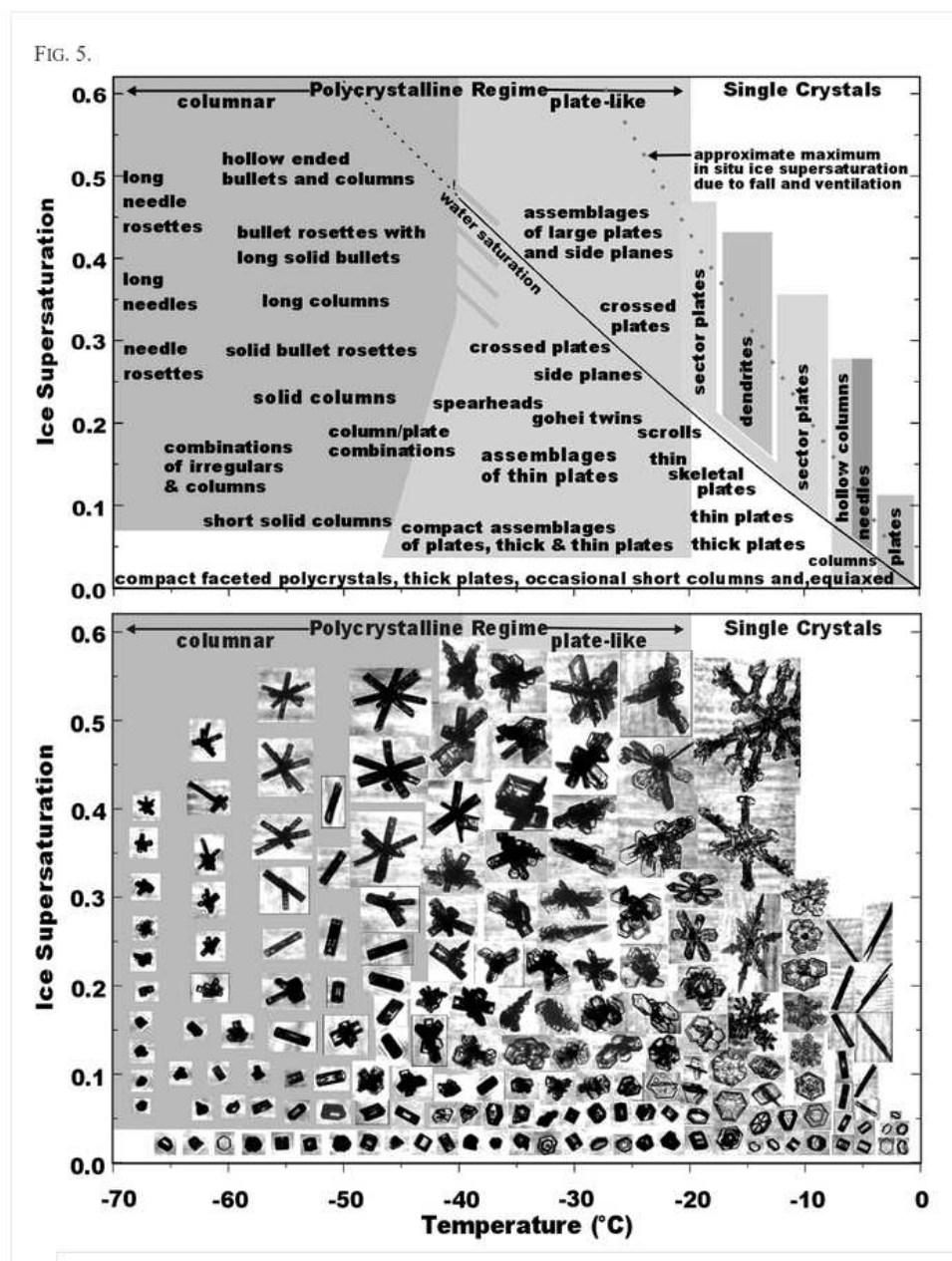


FIGURE 2.10 – Différents hydrométéores solides en fonction de la sursaturation par rapport à la glace et de la température.

Source: Bailey et Hallett 2009[31]

Grâce aux différentes classifications, les formes les plus répandues et leurs différentes caractéristiques ont été répertoriées. Il est important, toutefois, de garder en tête que ces classifications ne représentent qu'un

sous-ensemble des hydrométéores solides présents dans l’atmosphère.

Les particules de formes non sphériques

Les régimes de Mie et Rayleigh, précédemment présentés, permettent de calculer les SSP de particules sphériques mais ne peuvent être étendus à d’autres formes sans conduire à des résultats moins précis. Historiquement, le modèle RTTOV-SCATT considérait les particules de neige comme des sphères poreuses constituées de glace avec des inclusions d’air. Des méthodes numériques prenant en compte la variabilité morphologique des hydrométéores solides ont par la suite été développées. Il existe deux catégories principales : les méthodes avec résolution exacte et les méthodes approximatives. La première catégorie peut être à nouveau divisée en deux sous-catégories : les méthodes basées sur les équations différentielles et celles basées sur les intégrales de volume ou de surface. Nous décrivons tout d’abord les méthodes basées sur les équations différentielles :

- **La méthode de séparation de variable (SVM)** consiste à chercher la solution exacte à l’équation d’Helmholtz qui est une réduction des équations de Maxwell lorsqu’on recherche la solution analytique de la diffusion du rayonnement par des formes non sphériques. Pour cela, tous les champs sont développés en termes d’ondes sphéroïdales. Premièrement utilisée pour des sphères dans l’approche de Lorenz-Mie (Mie 1908[132], Van de Hulst 1957[134]), elle a ensuite été considérablement améliorée par Voshchinnikov et Farafonov (1993[140]) et étendue à des formes moins triviales telles que des cylindres et des sphéroïdes (Oguchi 1973[141], Asano et Yamamoto 1975[142]). L’avantage de cette méthode est de produire une solution exacte avec un temps de calcul rapide, cependant elle reste restreinte à des formes de particules simples. Une méthode améliorée de la SVM se basant sur la théorie de Mie se nomme la GMM (Generalized Multiparticle Mie-solution, Xu 1995[143]). Celle-ci consiste en une formulation simple des conditions aux limites pour une forme sphérique grâce à une expansion multipolaire⁴ des champs diffusés de chaque particule. Cette méthode permet de traiter des agrégats de sphères lisses sans aucune limite de taille.
- **La méthode temporelle par différences finies (FDTD)** (Yee, 1966[144]) consiste à calculer la diffusion électromagnétique en discrétisant en différences finies les coordonnées spatiales et temporelles des équations de Maxwell. Il est ainsi possible de résoudre les équations dans le domaine temporel. Cette méthode à l’avantage d’être simple à mettre en oeuvre et peut traiter des particules de formes arbitraires. Néanmoins elle est limitée à des paramètres de taille inférieurs à 50 (Yang et al. 2019[145]) : les calculs sont complexes, doivent être répétés pour chaque orientation de la particule et peuvent mener à des problèmes de convergence.
- **La méthode de mise en correspondance des points (PMM)** (Oguchi 1973[141]) fait correspondre point par point les champs incident, interne et diffusé en tenant compte des conditions limites à la surface de la particule. Le nombre de points correspond alors au nombre de coefficients d’expansion inconnus. Cette méthode, simple à implémenter, est limitée aux particules de formes proches de la sphère. Pour des particules s’éloignant de cette forme, des problèmes de convergence sont rencontrés.

Nous allons maintenant introduire les méthodes numériques basées sur les intégrales de surface ou de volume :

- Tout d’abord **la méthode de la matrice T** (Waterman, 1971[30]) consiste à calculer la matrice qui relie les coefficients d’expansion du champ incident aux coefficients d’expansion du champ diffusé, autrement appelée la matrice T. Cette matrice contient alors la solution du problème de diffusion et peut être à nouveau utilisée

4. Une expansion multipolaire est une série mathématique représentant une fonction dépendante d’angles définis. Généralement ces angles sont ceux utilisés dans le système de coordonnées sphériques et la fonction est décomposée en somme d’harmoniques sphériques

pour plusieurs orientations d'une même particule. L'avantage de cette méthode est qu'elle est très rapide pour des particules de formes simples ou axi-symétriques. En effet, les éléments de la matrice T étant stockés et pouvant être réutilisés pour toutes les orientations d'une même particule, un gain en temps de calcul considérable est obtenu pour ce type de formes. Elle peut être utilisée pour traiter des particules de formes plus complexes, mais devient relativement coûteuse pour des particules s'éloignant de formes simples et perd en précision. Cette méthode, communément calculée à l'aide des conditions limites étendues, est souvent désignée par **EBCM (Extended Boundaries Condition Method)** ou **NFM (Null Field Method)**. Néanmoins, plutôt qu'une méthode numérique de résolution à proprement parler, la matrice T décrit un type de formulation du problème de diffusion et peut donc être associée aux autres méthodes numériques résolvant les équations de Maxwell. Parmi les méthodes exploitant la matrice T on trouve the Invariant Imbedding T-matrix Method (II-TM, Johnson 1988[146]) ou encore the Discrete Dipole Moment Method (DDMM, Mackowski 2002[147]).

- Une autre méthode, se nommant la méthode **d'approximation par dipôles discrets** (DDA, Purcell et Pennypacker 1973[148]; Draine et Flatau 1994[149]) est principalement utilisée aujourd'hui pour simuler les SSP des hydrométéores solides. Cette méthode considère les particules comme étant des points polarisables définis sur une grille et résout itérativement le système résultant d'équations de Maxwell. Grâce à cette méthode, de nombreuses bases de données d'hydrométéores solides ont été produites ces dernières années. L'avantage de cette méthode est qu'elle peut traiter toutes les formes de particules. Cependant, elle devient coûteuse lorsque le nombre de dipôles considérés augmente. La plupart des SSPs calculées à l'aide de cette méthode proviennent des codes DDSCAT (Discrete Dipole SCATtering, Draine et Flatau, 2013[150]) codé en Fortran 77 et ADDA (Amsterdam Discrete Dipole Approximation, Yurkin et al. 2011[151]) codé en C.

La gamme de paramètres de tailles et de formes des particules étant très large, les méthodes numériques exactes ne sont pas suffisantes pour traiter toutes les particules. Pour pallier cela, des techniques dites d'approximations ont été développées. Nous allons décrire brièvement deux d'entre elles.

- Une première approximation est **l'approximation de Rayleigh-Gans-Debye** (RGA, Rayleigh 1914[133], Gans 1925[152], Debye 1915[153] et Van de Hulst 1957[134]), aussi appelée approximation de Rayleigh-Gans-Born (Born 1926[154]). Cette approximation consiste à négliger les interactions électromagnétiques internes d'un diffuseur si elles sont faibles. Elle s'applique à des particules de formes arbitraires dont l'indice de réfraction propre est proche de l'indice de réfraction du milieu environnant. Pour résumer, la méthode peut être appliquée si les conditions suivantes sont respectées :

$$|m - 1| \ll 1$$

$$2x|m - 1| \ll 1$$

- La deuxième méthode introduite est basée sur la précédente : **l'approximation autosimilaire de Rayleigh-Gans (SSRGA, the Self-Similar Rayleigh-Gans Approximation)** (Hogan et Westbrook 2014[155], Hogan et al. 2017[156]). Dans l'approximation RGA, la description 1D de la structure de la particule dans la direction de l'onde incidente possède une structure dite 'auto-similaire' pour des particules de type agrégats. Cette structure permet de définir une fonction analytique pour la diffusion moyenne grâce à quatre variables seulement, simplifiant ainsi les calculs pour ce type de particule. La méthode de la SSRGA exploite cette simplification afin de traiter les agrégats. Malgré la rapidité de cette méthode en temps de calcul, elle est limitée aux agrégats satisfaisant les conditions de l'approximation RGA définis plus haut.

2.2.2 Historique des bases de données

Afin de simuler les luminances micro-ondes, plusieurs bases de données, répertoriant les formes de particules les plus fréquemment rencontrées dans la nature et leurs SSPs associées, ont été produites. Il est important de préciser que ces bases ne représentent qu'une moyenne des possibles compte tenu de la diversité des hydrométéores solides dans l'atmosphère. Une première base de données parmi les plus connues est celle de Liu2008 (Liu, 2008[11]). Cette base a été calculée à l'aide de la méthode numérique DDA dans une gamme de fréquences micro-ondes allant de 15 à 340 GHz pour 11 particules solides contrastées (forme, densité, taille, figure 2.11). Parmi cet ensemble, la particule Sector snowflake a été estimée, lors d'une étude réalisée dans Geer et Baordo 2014[1], comme la plus optimale pour simuler les luminances micro-ondes au sein du modèle global de PNT du CEPMMT.

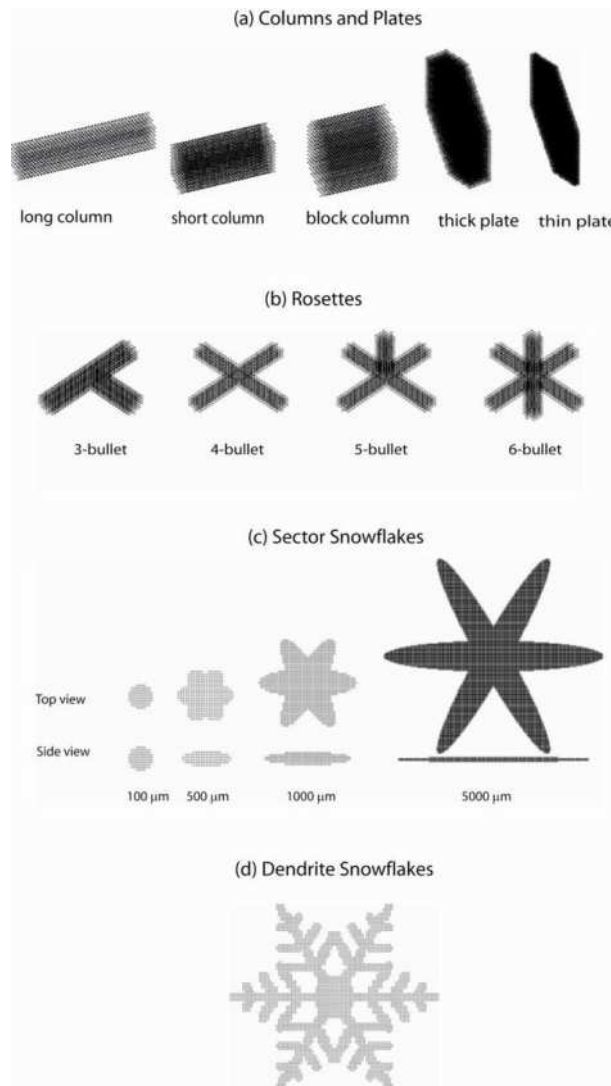


FIGURE 2.11 – Base de données Liu2008 comprenant 11 particules solides contrastées (Liu 2008)

Afin de répondre à de nouveaux besoins pour simuler les luminances micro-ondes, plusieurs nouvelles bases ont été produites. Pour la plupart, les SSPs ont été calculées à l'aide de la méthode numérique DDA : Hong2007[32], Petty2010[33], Tyynelä2009[34], Ori2014[35], Leinonen2015[36][37], Kuo2016[38] et enfin la Johnson2018. Une autre base utilise quant à elle à la fois la méthode DDA et la méthode GMM : Lu2016[39]. Celle de Ding2017[40], a été développée à l'aide des méthodes II-TM et Improved Geometric Optics Method (IGOM) afin de calculer des SSPs couvrant les fréquences micro-ondes et sub-millimétriques (de 300 GHz à 3 THz). Elle répertorie 12 SSPs calculées pour des fréquences allant de 1 à 874 GHz.

TABLE 2.3 – Tableau récapitulatif des caractéristiques d’un ensemble de bases de données de SSPs dans le domaine des micro-ondes pour divers hydrométéores solides

Base de données	Type de particule	Taille des particules	Fréquences (GHz)	Méthode de calcul
Hong (2007)	hexagonal column, hollow column, rosette, droxtal, aggregate, and plate colonne	2 μm à 5.5 mm	89 à 340	DDA (DDSCAT)
Liu (2008)	pristine crystals of various shapes and aggregates of rosettes	125 μm à 10 mm	13.4 à 340	DDA (DDSCAT)
Tynnelä (2009)	pristine, aggregates of dendrites, fractal aggregates	100 μm à 24 mm	2.7 à 220	DDA (ADDA)
Petty (2010)	aggregates of dendrites and needles	125 μm à 10 mm	13.4 à 89	DDA
Ori (2014)	aggregates of columns and partially melted aggregates	0.4 à 15 mm	5.6 à 157	DDA (ADDA)
Leinonen (2015)	unrimed and rimed aggregates	0.2 à 20 mm	9.7 à 94	DDA (ADDA)
Kuo (2016)	pristine and aggregates	0.26 à 14 mm	3 à 190	DDA (DDSCAT)
Lu (2016)	pristine and aggregates	0.1 à 62 mm	9.4 à 94	GMM, DDA
Ding (2017)	pristine and aggregates	2 μm à 10 mm	1 à 886	II-TM et IGOM
Johnson (2018)	melting aggregates of dendritess	100 μm à 14 mm	13.6 à 183.31	DDA (DDSCAT)
Eriksson (2018)	pristine plates and columns, aggregates of plates and columns, bullet rosettes, fractal aggregates, graupel	100 μm à 22 mm	1 à 884.6	DDA(ADDA)

La dernière base de données que nous citerons est celle de Eriksson2018[41], plus connue sous le nom de base de données d’ARTS (Atmospheric Radiative Transfer Simulator). Cette base, comme celle de Ding2017, a été étendue aux fréquences sub-millimétriques afin de préparer la simulation de trois nouveaux instruments : le radiomètre MicroWave Imager (MWI), l’imageur Ice Cloud Imager (ICI) et le sondeur MicroWave Sounder (MWS). Elle contient un ensemble de 34 particules dont les SSPs calculées à l’aide de la méthode numérique DDA pour 34 fréquences allant de 1 à 886.4 GHz. Les caractéristiques des bases de données décrites précédemment sont rassemblées dans le tableau 2.3.

Chapitre 3

Méthodes d'assimilation pour les observations micro-ondes nuageuses et précipitantes

Sommaire

3.1	L'initialisation physique	53
3.2	Les méthodes d'assimilation indirectes	54
3.2.1	1D-Var+4D-Var	54
3.2.2	1D-Bay+4D-Var	57
3.3	Les méthodes d'assimilation directe	58
3.3.1	4D-Var avec le modèle global IFS	58
3.3.2	Filtre ensembliste avec le modèle régional WRF	61
3.4	Avantages et inconvénients des différentes méthodes	62

Les recherches sur l'assimilation des observations satellitaires micro-ondes nuageuses et/ou précipitantes ont commencé avec les arrivées de l'imageur micro-ondes passif Special Sensor Microwave / Imager (SSM/I) en 1987 et du radiomètre micro-ondes passif Tropical Rainfall Measuring Mission Microwave Imager (TMI) en 1997. Ces instruments possédaient respectivement 7 et 9 fréquences particulièrement intéressantes pour étudier la vapeur d'eau, l'eau nuageuse et les précipitations.

En complément des chapitres précédents, dans lesquels nous avons présenté les concepts généraux d'assimilation de données et ceux du transfert radiatif, nous allons dans ce chapitre nous focaliser sur l'assimilation des observations micro-ondes en milieux nuageux et précipitant, en détaillant les différentes méthodes qui ont été élaborées au cours du temps, ainsi que les problèmes rencontrés au cours de leur développement.

3.1 L'initialisation physique

Nous avons vu dans les chapitres précédents que l'assimilation de radiances satellitaires nécessitait de relier les mesures du rayonnement électromagnétique à des variables "météorologiques". Concernant les zones pluvieuses, ce fut réalisé initialement via des taux de pluie, qui étaient dérivés de pluviomètres et de flux infra-rouges puis de radiances micro-ondes. Avant les années 2000, et l'arrivée des méthodes d'assimilation variationnelle (cf. Le Dimet et Talagrand 1986[157], Talagrand et Courtier 1987[81], Courtier et Talagrand 1987[82]), divers algorithmes dépendant des instruments choisis étaient nécessaires pour calculer des taux de pluie. Ces méthodes reposaient sur une initialisation physique ou sur un 'nudging'. L'initialisation physique consistait à réaliser une relaxation newtonienne pour forcer les variables du modèle à se rapprocher des observations en ajoutant un terme de forçage aux équations pronostiques. Ceci se

faisait après l'inversion empirique de schémas de convection simples (de type Kuo, Kuo et al. 1974[158]) permettant de modifier les profils d'humidité et de température afin de simuler des taux de pluie proches des observations. La méthode la plus utilisée à l'époque était celle décrite dans Krishnamurti (1991)[15]. Elle avait pour avantage de réduire les déséquilibres initiaux pour les composantes du cycle hydrologique (problème dit de "spin-up") et ainsi d'améliorer la prévision des précipitations principalement dans les régions tropicales.

La méthode nommée 'nudging' utilisait quant à elle le schéma dit du 'latent heat nudging', que l'on peut traduire par 'relaxation par libération de chaleur latente'. Cette méthode consistait à modifier l'équation thermodynamique du modèle afin de conduire à un meilleur accord avec les taux de précipitations observés. Pour cela, les taux de précipitations cumulés obtenus à partir d'observations d'instruments infra-rouges, micro-ondes et des pluviomètres étaient comparés aux taux de précipitations cumulés du modèle. Un incrément ΔR était alors défini comme :

$$\Delta R = R_o - R_b$$

avec R_o le taux de pluie observée et R_b le taux de pluie de l'ébauche. À l'aide d'hypothèse sur la structure verticale de l'incrément, ce dernier était par la suite converti en un incrément de température au sein du modèle. La libération de chaleur latente par condensation était alors modifiée à chaque pas de temps permettant au modèle de se rapprocher des précipitations observées. Jones et Macpherson (1997[159]) ont montré la capacité de cette méthode à améliorer la prévision des précipitations jusqu'à 9 heures d'échéance dans le modèle à aire limitée du UK MetOffice.

3.2 Les méthodes d'assimilation indirectes

A partir de la fin de années 1990, les méthodes variationnelles ont permis d'explorer l'assimilation directe de taux de pluie. Plusieurs expérimentations ont été tentées dans des modèles régionaux (Zupanski et Mesinger, 1995[160]; Zou and Kuo, 1996[161]) et globaux (Tsuyuki, 1996[162], 1997[163]). Toutefois, ce n'est qu'après la première implémentation opérationnelle de l'assimilation 4D-Var au CEPMMT fin 1997 que des expérimentations dans un cadre compatible avec la prévision numérique opérationnelle ont été réalisées. L'approche proposée alors par Marécal et Mahfouf (2000[16], 2002[17]) était une méthode indirecte, car les opérateurs d'observations linéarisés (transfert radiatif et paramétrisations des nuages et des précipitations) n'existaient pas encore dans le 4D-Var et des problèmes de non-linéarités avaient été identifiés dans l'approche incrémentale de l'assimilation (Marécal et Mahfouf, 2003[18]). Nous allons décrire deux de ces méthodes dans les paragraphes ci-dessous : la méthode 1D-Var+4D-Var et la méthode 1D-Bay+4D-Var.

3.2.1 1D-Var+4D-Var

La méthode 1D-Var+4D-Var est proposée dans les années 2000 par Marécal et Mahfouf (2000[16]). À l'époque, cette méthode est déjà opérationnelle pour assimiler les températures de brillance micro-ondes en ciel clair au CEPMMT (Gérard et Saunders 1999[164]). C'est donc naturellement que celle-ci a été utilisée pour assimiler des observations dans les zones pluvieuses. Comme l'initialisation physique, cette méthode nécessite de convertir au préalable les radiances en taux de pluie afin d'être assimilées. Cette méthode est composée de deux étapes ; la première consiste à minimiser une fonctionnelle J telle que :

$$J(x) = \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) + \frac{1}{2} \left[\frac{R(x) - R_o}{\sigma_o} \right]^2$$

Où $R(\cdot)$ est l'opérateur d'observation constitué par deux paramétrisations physiques : un schéma de condensation de grande échelle (pour le taux de pluie résolue) et un schéma de convection profonde (pour le taux de pluie sous-maille), R_o le taux de pluie observé, x_b l'ébauche et σ_o l'écart type d'erreur d'observation. Cette minimisation permet ainsi d'obtenir un profil atmosphérique x modifié. Pour l'assimilation de taux de pluie dérivés de radiances micro-ondes, les variables de contrôle sont l'humidité et la température. Il est important de signaler que la minimisation de la fonction coût J nécessite la connaissance de son gradient qui est fonction de l'adjoint de l'opérateur d'observation. Cette linéarisation de $R(\cdot)$ a été réalisée par Marécal et Mahfouf (2000[16]) en différences finies avant le développement des versions linéarisées de ces paramétrisations au CEPMMT (Lopez et Moreau, 2005[165]; Tompkins et Janisková, 2004[166]).

Les corrections apportées aux profils atmosphériques viennent principalement de l'humidité spécifique mais leur structure verticale est fortement conditionnée par la matrice \mathbf{B} de covariances d'erreurs de l'ébauche. Il a donc été choisi d'assimiler dans le 4D-Var l'intégrale sur la verticale du profil d'humidité spécifique : le contenu intégré en vapeur d'eau. Ce choix permet de réduire les corrélations éventuelles des erreurs entre ces "pseudo-observations" et l'ébauche, et il est cohérent avec l'approche utilisée pour l'assimilation des températures de brillance micro-ondes en ciel clair (Gérard et Saunders, 1999) comme indiqué précédemment.

Des études utilisant cette méthode avec les instruments TMI et SSM/I ont permis de montrer l'intérêt de ces informations en zones pluvieuses sur la qualité des prévisions du modèle IFS (e.g. Mahfouf et al., 2005). Ces études ont également mis en évidence qu'il y aurait un intérêt à utiliser directement les températures de brillances micro-ondes : spécification plus aisée des erreurs d'observations, moins de problèmes de discontinuités entre les zones claires et nuageuses grâce à une sensibilité des observations à la vapeur d'eau, aux nuages et à la pluie (cf. Moreau et al, 2003[167]).

Dans les années suivantes, des modifications sont apportées à la méthode 1D-Var+4D-Var, dont les principales ont permis l'assimilation des radiances micro-ondes. L'analyse de l'humidité a été améliorée en utilisant comme nouvelle variable de contrôle : l'humidité relative normalisée, afin de simplifier les caractéristiques des erreurs de l'ébauche par rapport à l'humidité spécifique q (Holm 2002[168]). Le schéma de convection de Tiedtke (Tiedtke 1989[51]) a été modifié par Lopez et Moreau 2005[165] pour rendre son comportement plus linéaire et un schéma diagnostique de nuages et de précipitations stratiformes a été développé pour les besoins de l'assimilation variationnelle (Tompkins et Janiskova 2004[166]). Finalement un modèle de transfert radiatif rapide pour les fréquences micro-ondes en ciel nuageux et précipitant ainsi que ses versions linéarisées ont été inclus dans le code RTTOV (Bauer et al., 2006[169]). Des filtrages des données et des corrections dynamiques ont été aussi ajoutés afin de rejeter les mauvaises observations et de corriger celles restantes : le pré-screening, la correction de biais et le post-screening. Nous détaillons maintenant ces différentes étapes :

- **(1) La correction du biais géométrique** : les biais dépendant de la position du pixel sur la ligne de balayage (scan) de l'instrument doivent être calculés afin de corriger les observations. Ces biais sont propres à chaque instrument et ne varient pas dans le temps.

- **(2) Le Pré-screening** : cette étape consiste à écarter certaines observations selon des critères prédéfinis. Tout d'abord sont rejetées les observations micro-ondes sur continents et glace de mer à cause de la difficulté à estimer à l'époque l'émissivité de surface dans cette gamme de fréquences. Ensuite, on supprime les observations en ciel clair si le modèle, à la même localisation, présente aussi des conditions atmosphériques claires (ces observations sont alors traitées dans une autre filière pour l'assimilation). Finalement, on supprime aussi les cas de forts vents en surface dans le modèle, car la modélisation de

l'émissivité sur océans est alors moins précise.

- **(3) Correction du biais** : un calcul de biais est réalisé entre les observations et les simulations des températures de brillance des profils du modèle. Ensuite, une correction de biais est appliquée de manière à corriger les valeurs des colonnes totales de vapeur d'eau.
- **(4) 1D-Var** : ensuite par application d'un 1D-Var on cherche le profil atmosphérique x , composé des variables de contrôle de l'humidité spécifique et de la température, minimisant une fonctionnelle J exprimée à l'aide des températures de brillance observées.
- **(5) Le post-screening** : un post-screening est appliqué aux résultats du 1D-Var si la convergence de la minimisation n'est pas satisfaisante ou si les incréments pour la colonne totale de vapeur d'eau sont trop élevés.
- **(6) Le 4D-VAR** : pour finir, les restitutions de vapeur d'eau intégrée ayant passé toutes ces étapes sont finalement assimilées au sein d'un 4D-Var avec les autres observations.

Au milieu des années 2000, et avec l'exploitation des méthodes variationnelles pour l'assimilation des observations micro-ondes nuageuses et précipitantes, de nouvelles problématiques se dessinent. Les principales ont été décrites par Errico et al (2007)[23]. En 2008, Geer et al (2008)[24] font aussi état des différentes problématiques de l'approche 1D-Var+4D-Var et proposent des pistes pour les résoudre. Les non-linéarités au sein de l'algorithme de minimisation sont de nouveau évoquées et une meilleure spécification de la matrice d'erreur du modèle est suggérée. Nous allons à présent mentionner quelques-unes de ces difficultés :

- Les erreurs de représentativité peuvent dépasser les erreurs instrumentales ; elles doivent de ce fait être prises en compte.
- Les statistiques d'erreurs d'ébauche de l'écoulement sont souvent ignorées.
- Les schémas de paramétrisation des nuages et des précipitations utilisés dans l'opérateur d'observation présentent diverses faiblesses (e.g. absence de condensats nuageux et précipitants dans certains schémas de convection).
- Les SSPs des hydrométéores (notamment solides) devant être spécifiées dans l'opérateur d'observation sont mal connues.
- Les importantes non-linéarités des phénomènes simulés peuvent poser des difficultés dans la minimisation des assimilations variationnelles et obliger à réduire les fenêtres temporelles d'assimilation.

Parallèlement aux activités du CEPMMT sur l'assimilation des observations micro-ondes nuageuses et/ou précipitantes, la National Aeronautics and Space Administration (NASA) a mené des recherches similaires en lien avec l'exploitation des données de la mission spatiale TRMM et la préparation de la future mission GPM. Une nouvelle méthode d'assimilation appelée 1+1D (Hou et al. 2000a[170], Hou et al. 2000b[171], Hou et al. 2004[172]), puis '1D+VCA' (Variational Continuous Assimilation, Derber 1989[173], Hou et al. 2004[172]) a été proposée. Le système d'assimilation de la NASA GEOS-DAS (Goddard Earth Observing System - Data Assimilation System) utilise une méthode nommée IAU¹ (Incremental Update Analysis, Bloom et al. 1996[174]). La méthode 1D+VCA s'est inspirée de cette approche pour assimiler

1. Méthode IAU : les incréments d'analyse estimés à un instant donné sont injectés progressivement dans les équations du modèle sur une certaine période de temps afin de réduire les chocs initiaux (problèmes de spin-up)

les observations micro-ondes nuageuses et/ou précipitantes. Pour cela les observations micro-ondes sont au préalable converties en taux de pluie en surface, instantanés ou moyennés dans le temps. Ensuite, à l'aide de ces taux de pluie, un incrément en température potentielle et en humidité spécifique est estimé puis appliqué aux équations pronostiques du modèle de prévision. Cette méthode a été tout d'abord testée avec des taux de pluie de surface et des estimations d'eau précipitable totale dérivés des observations de l'imageur SSM/I. Celle-ci a montré de bons résultats : par exemple les précipitations et l'eau précipitable ont été améliorées dans l'analyse ainsi que les prévisions à courte échéance dans une moindre mesure (Hou et al. 2000a[170]). Par la suite, cette méthode a utilisé des taux de pluie issus de l'imageur TMI avec des résultats similaires. La représentation du cycle de l'eau a été améliorée en modifiant certains paramètres comme les nuages ou la circulation à grande échelle (Hou et al. 2001[175]).

3.2.2 1D-Bay+4D-Var

Une méthode alternative d'assimilation de températures de brillance micro-ondes en deux étapes, nommée '1D-Bay+3D/4D-Var' et basée sur une inversion Bayésienne à l'aide des profils du modèle dans le voisinage de l'observation, a été initialement proposée par Guerbet et al. (2016[4]), en s'inspirant d'une méthodologie développée pour l'assimilation de réflectivités radar (Wattrelot et al., 2014[20]). Cette méthode permet d'obtenir des profils atmosphériques inversés qui sont par la suite assimilés dans un 3D ou un 4D-Var à différents niveaux de pression. La variable humidité relative est préférée à l'assimilation des profils d'humidité spécifique et de température afin d'impacter directement la saturation de l'atmosphère et de moduler la nébulosité dans la prévision, ainsi que pour faciliter la spécification des erreurs d'observation. Nous présentons maintenant le formalisme de la première étape de cette méthode : l'inversion Bayésienne.

Dans cette première étape, on cherche le profil atmosphérique x le plus probable connaissant une observation y_o à la même localisation. Pour cela on cherche l'espérance du profil atmosphérique réel ' x_{true} ' compte tenu de l'observation y_o :

$$E[x = x_{true}|y = y_o] = \int_V x \mathbf{P}(x = x_{true}|y = y_o) dx$$

Avec V un voisinage de l'ébauche autour l'observation y_o . En utilisant successivement le théorème de Bayes permettant de réécrire la probabilité de l'état x connaissant y_o :

$$\mathbf{P}(x = x_{true}|y = y_o) = \frac{\mathbf{P}(y = y_o|x = x_{true})\mathbf{P}(x = x_{true})}{\mathbf{P}(y = y_o)}$$

et en remarquant que la probabilité $\mathbf{P}(y = y_o)$ vaut 1 du fait que l'on traite l'observation y_o , la formule de l'espérance peut se réécrire :

$$E[x = x_{true}|y = y_o] = \int_V x \mathbf{P}(y = y_o|x = x_{true})\mathbf{P}(x = x_{true}) dx$$

On suppose par la suite que les erreurs d'observations suivent une distribution gaussienne telle que :

$$\mathbf{P}(y = y_o|x = x_{true}) \propto e^{-\frac{1}{2}(y_o - H(x))^t \mathbf{R}^{-1}(y_o - H(x))}$$

avec H l'opérateur d'observations et \mathbf{R} la matrice d'erreur d'observations. Ce qui amène à :

$$E[x = x_{true}|y = y_o] = \frac{1}{A} \int_V x e^{-\frac{1}{2}(y_o - H(x))^t \mathbf{R}^{-1}(y_o - H(x))} \mathbf{P}(x = x_{true}) dx$$

où A est une constante de normalisation. On suppose, de plus, que l'occurrence et l'intensité des profils de nuages sont similaires dans l'ébauche et dans la nature, en particulier au voisinage de y_o . Ainsi, la

probabilité $P(x = x_{true})$ peut être représentée par le nombre relatif d'occurrences d'un type de profil donné x_i représentant les profils se trouvant dans le voisinage de y_o . On obtient finalement :

$$x_{inv} := E[x = x_{true}|y = y_o] \propto \frac{\sum_{i \in Ind_V} x_i w_i}{\sum_{i \in Ind_V} w_i} \quad (3.1)$$

où la constante de normalisation $A = \sum_{i \in Ind_V} w_i$ avec $w_i = e^{(-\frac{1}{2}[y_o - H(x_i)]^t \mathbf{R}^{-1} [y_o - H(x_i)])}$.

De cette manière, nous obtenons la moyenne pondérée x_{inv} de la probabilité conditionnelle de x étant donnée l'observation y_o . Il est possible de fixer un seuil sur la somme des poids afin de définir le succès d'une inversion : si le poids total est plus fort que le seuil, l'inversion sera réalisée, dans le cas contraire, celle-ci ne sera pas réalisée. Cette méthode est utilisée opérationnellement à Météo-France depuis juillet 2022 pour assimiler les radiances micro-ondes nuageuses des sondeurs ATMS (Atmospheric Technology Microwave Sounder) et MHS (Microwave Humidity Sounder) dans le modèle global ARPEGE (Duruisseau et al. 2019[44]). L'information issue du profil est assimilée à n niveaux de pression correspondant ici au nombre de canaux de l'instrument. Ces niveaux de pression sont ici définis à l'aide des sensibilités moyennes de chaque canal. La figure 3.1 présente un exemple d'une inversion Bayésienne avec le modèle ARPEGE et les observations à la fréquence 183.31 ± 11.0 du Sondeur Atmosphérique du Profil d'Humidité Intertropicale par Radiométrie (SAPHIR). On remarque que les structures nuageuses sont mal représentées dans l'ébauche du modèle ARPEGE. Après inversion avec une erreur d'observation valant 1 K, on remarque que beaucoup d'inversions ont échoué. Cependant lorsqu'on change l'erreur d'inversion par un modèle d'erreur empirique (mais plus réaliste qu'une valeur unique), on obtient des résultats où les structures nuageuses sont correctement repositionnées par rapport aux observations malgré une surestimation des valeurs des températures de brillance (sous estimation de la diffusion).

L'Agence Météorologique Japonaise (JMA) assimile opérationnellement des réflectivités radars à l'aide de cette méthode mais en considérant un estimateur basé sur le maximum de vraisemblance (Ikuta et al. 2021[22]). Le profil résultant de cette formulation est alors celui qui maximise la fonctionnelle :

$$L(x = x_{true}|y = y_o) = \prod_{j=1}^d e^{-\frac{1}{2}(y_{o,j} - H_j(x))^t \mathbf{R}^{-1} (y_{o,j} - H_j(x))}$$

Avec d le nombre de fréquences. On cherche donc ici :

$$x_{inv} = \underset{x}{\operatorname{Argmax}} L(x = x_{true}|y = y_o)$$

L'estimateur basé sur le maximum de vraisemblance revient ainsi à chercher le profil le plus ressemblant à l'observation ; l'estimateur basé sur la moyenne pondérée revient lui à lisser le résultat avec la moyenne des profils. Les différences entre les deux approches de moyenne pondérée et de maximum de vraisemblance seront étudiées au chapitre 5.

3.3 Les méthodes d'assimilation directe

3.3.1 4D-Var avec le modèle global IFS

En mars 2009, les radiances micro-ondes nuageuses des instruments SSM/I et Advanced Microwave Scanning Radiometer for Earth Observation System (AMSR-E), provenant des fréquences de 19 à 89 GHz, ont été assimilées directement au sein du 4D-Var au CEPMMT (Bauer et al. 2010[25], Geer et al. 2010[26]). Ce passage à l'assimilation directe a été motivé par certaines faiblesses de l'approche 1D+4D-Var citées précédemment. En effet, il est apparu que la sélection des observations entre ciel clair et ciel nuageux introduisait un biais d'échantillonnage important. Ces biais conduisaient à une mauvaise correction de

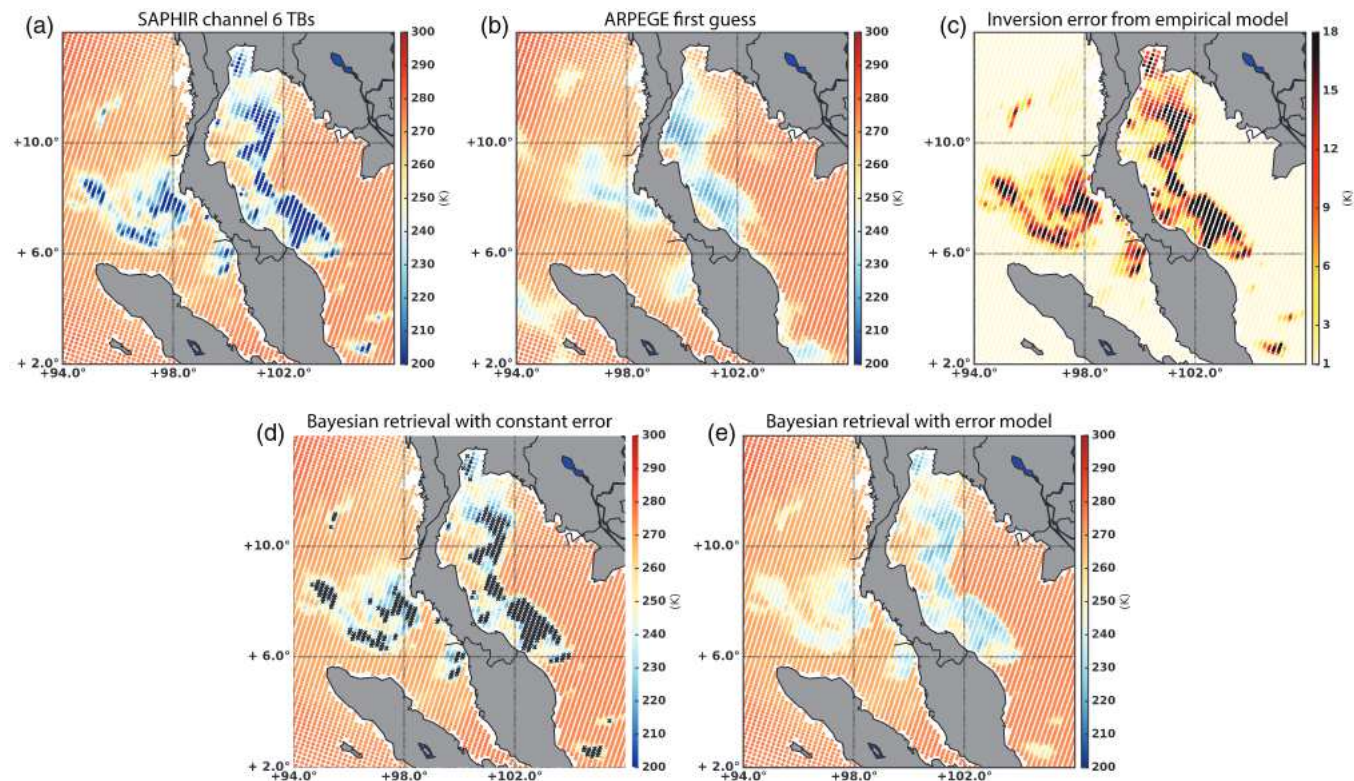


FIGURE 3.1 – Exemple d’inversion Bayésienne sur les côtes Thaïlandaises. (a) Observations du canal 183.31 ± 11.0 du sondeur SAPHIR le 8 Juin 2017 à 00:00 UTC ; (b) estimation du modèle correspondant à la même heure et au même endroit ; (c) erreur d’observation du modèle empirique ; (d) Résultats de l’inversion en considérant une erreur d’observation de 1 K ; (e) Résultats de l’inversion en considérant le modèle d’erreur d’inversion adapté aux scènes observées. Les petites croix noires représentent les endroits où l’inversion Bayésienne a échoué.

Source: Duruisseau et al. 2019[44]

l’ébauche pour les cas d’observations claires et de profils du modèle nuageux. Dans un bilan des leçons à tirer de l’approche 1D-Var+4D-Var, Geer et al. 2008[24] insistaient ainsi sur l’importance de ne pas traiter les radiances nuageuses différemment des radiances claires. Une nouvelle approche, appelée assimilation ‘all-sky’ permettant de traiter toutes les situations météorologiques par la même filière a ainsi été proposée. Les premiers résultats de cette nouvelle approche n’ont toutefois pas eu le succès escompté : en effet, pour certaines variables, les résultats étaient de moins bonne qualité qu’avec la méthode 1D-Var+4D-Var. Pour pallier ces problèmes, de nouvelles modifications ont été réalisées.

La première modification concernait un problème déjà énoncé par Errico et al. 2007[23] : la définition des erreurs d’observations. Avant 2010, la matrice de covariances d’erreurs d’observations était définie selon une fonction calculant la distance entre les observations et les points de grille du modèle. Cependant, cette définition était peu appropriée : elle dégradait les cas où les nuages et les systèmes convectifs étaient mal localisés par le modèle. Un autre problème venait d’un effet de saturation des températures de brillance en ciel nuageux, pouvant induire des écarts très larges et non gaussiens entre les observations et l’équivalent modèle. Pour remédier à ce problème, un modèle d’erreur exploitant un aspect dit ‘symétrique’, basé sur la valeur des erreurs d’innovation et la fraction nuageuse de l’ébauche, a été défini. Un premier modèle d’erreur a utilisé un indice de diffusion défini comme la différence normalisée entre les polarisations horizontale et verticale pour un canal sensible aux nuages, ici le canal 37 GHz du radiomètre AMSR-E (Petty 1994[98]). Ce modèle d’erreur a alors permis d’obtenir un échantillonnage symétrique avec des statistiques d’innovations proches d’une fonction gaussienne (Geer et Bauer, 2011[176]).

La deuxième modification principale concerne le filtrage de certaines régions du globe pour lesquelles les

biais du modèle étaient trop importants. Cela concerne d'abord les intrusions extrêmes d'air froid ('Cold Air Breakout' en anglais), fréquentes dans les océans en saison hivernale et aux hautes latitudes, qui provoquent des nuages de convection peu profonde à cause de l'air froid et sec advecté au dessus des océans. Le modèle a alors tendance à produire trop de glace, augmentant alors les biais du modèle. On peut trouver les critères permettant de filtrer ces régions dans Geer et Bauer (2010)[26]. Les autres régions à filtrer, situées en zone tropicale, correspondent à la création de nuages, avec d'importantes quantités de neige et de glace qui sont encore mal simulées et créent de ce fait une diffusion excessive dans le modèle IFS. Ces zones ont été filtrées à l'aide de la fraction nuageuse modélisée.

Ces deux améliorations ont par la suite été implémentées au sein du 4D-Var, permettant finalement d'améliorer les scores de prévisions. La figure 3.2 illustre l'impact de l'assimilation d'une observation du sondeur MHS au sein d'une fenêtre d'assimilation du 4D-Var (entre 21 UTC et 09 UTC). L'observation étant située en fin de fenêtre (8 UTC), le 4D-Var est capable de modifier les champs dynamiques en début de fenêtre à 21 UTC en amont du système dépressionnaire (incrément de pression de surface à l'ouest de la croix localisant l'observation) pour permettre la création à l'heure et au point d'observation d'un incrément significatif de contenu intégré en neige afin d'améliorer la simulation de la température de brillance pour ce canal sensible à ces hydrométéores.

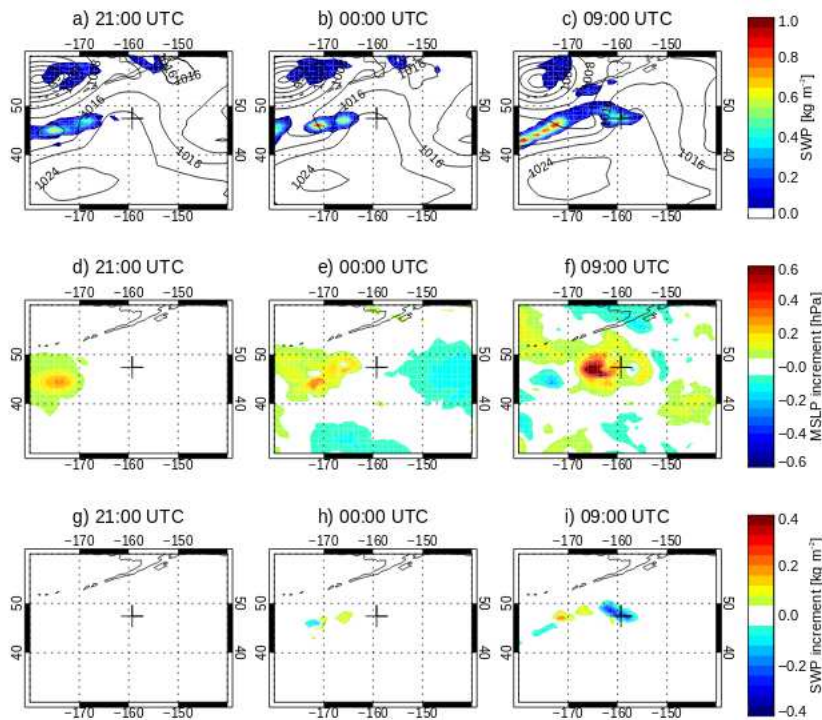


FIGURE 3.2 – Illustration des effets de l'assimilation d'une température de brillance du sondeur MHS à la fréquence 183 GHz, représentée par une croix noire, durant une fenêtre d'assimilation (21:00 UTC jusqu'à 09:00 UTC). La première colonne montre les effets à 21:00 UTC, la deuxième à 00:00 UTC et la troisième à 09:00 UTC. La première ligne montre les contenus en taux neige en kg/m^2 , la deuxième les incréments de pression au niveau moyen de la mer en hPa et la troisième les incréments de contenus en neige.

Source: Geer et al. 2017[2]

Dans la continuité de ces études sur l'assimilation des observations en ciel nuageux, les principales recherches concernent l'amélioration des propriétés de diffusion dans la modélisation du transfert radiatif, afin de pouvoir mieux exploiter l'information contenue dans les hautes fréquences micro-ondes. Dans la première version du code RTTOV-SCATT (Bauer et al., 2006) les particules de neige étaient considérées comme des sphères de glace avec des inclusions d'air pour utiliser les propriétés de diffusion issues de la théorie de Mie. Toutefois, ces propriétés amenaient à une diffusion excessive pour les basses fréquences entre 30 et 50 GHz et à une diffusion insuffisante pour les hautes fréquences entre 150 et 183 GHz. Geer et

Baordo (2014)[1] ont mené des expériences exploitant les SSPs de la base de Liu2008[11], précédemment vues dans le chapitre 2, avec une distribution en taille des particules proposée par Field et al. 2007[127] afin de trouver une particule 'optimale'. Ces expériences ont permis d'assimiler les hautes fréquences micro-ondes avec des impacts positifs en utilisant les SSPs de la particule *Sector snowflake* de la base de Liu2008. Ces résultats ont ainsi amené le CEPMMT à utiliser ce choix de particule au sein de leur modèle opérationnel pour assimiler les radiances micro-ondes nuageuses en 2013. Les améliorations ont ensuite permis d'assimiler pour la première fois le canal 183 GHz de l'instrument SSMI/S (the Special Sensor Microwave Imager/Sounder) en 2015.

3.3.2 Filtre ensembliste avec le modèle régional WRF

Une autre méthode d'assimilation directe considère les observations micro-ondes nuageuses et/ou précipitantes à l'aide d'une approche ensembliste. Ce type de méthode estime la matrice de covariances d'erreurs d'ébauches à l'aide d'un ensemble de prévisions. Basée sur ce principe, la NASA développe ainsi le système d'assimilation de données d'ensemble du modèle Goddard Weather Research Forecast (WRF) (Goddard WRF-EDAS) pour assimiler des observations micro-ondes en ciel nuageux au sein du modèle WRF (Zupanski et al. 2011[177], Zhang et al. 2013[178] et Chambon et al. 2014[179]). Ce système repose sur une méthode de résolution exploitant un filtre ensembliste à maximum de vraisemblance (Zupanski 2005[180]; Zupanski et al. 2008[181]), combinant la propagation des covariances d'erreurs de prévision basées sur l'ensemble à l'estimation du maximum de vraisemblance pour obtenir une solution d'analyse optimale.

Avec cette méthode, on cherche à minimiser la fonctionnelle J à l'aide d'un estimateur à maximum de vraisemblance :

$$J(x) = \frac{1}{2} [x - x_b]^T \mathbf{B}_f^{-1} [x - x_b] + \frac{1}{2} [y - H(x)]^T \mathbf{R}^{-1} [y - H(x)]$$

Avec x la meilleure estimation de l'état atmosphérique définie à l'aide des variables de contrôle du vent, de la température, de l'humidité et cinq types d'hydrométéores (rapport de mélange pour la pluie, la neige, le graupel, l'eau nuageuse et la glace nuageuse à partir de schémas microphysiques), x_b l'ébauche, y l'observation, \mathbf{B}_f est la matrice de covariances d'erreur de prévision, \mathbf{R} la matrice de covariances d'erreurs d'observations et H désigne l'opérateur d'observation. Le filtre à maximum de vraisemblance consiste alors à appliquer une transformation sur les variables de contrôle pour résoudre l'équation dans un espace ensembliste tel que :

$$x - x_b = \mathbf{B}_f^{1/2} [\mathbf{I} - \mathbf{Z}(\mathbf{x}_b)^T \mathbf{Z}(\mathbf{x}_b)]^{-1/2} \xi$$

Avec ξ la variable de contrôle dans l'espace de l'ensemble, \mathbf{Z} la matrice d'information des observations dont les colonnes valent : $z_i = \mathbf{R}^{-1/2} [H(x_i) - H(x_b)]$ et i un membre de l'ensemble.

Cette méthode a été évaluée par Zhang et al.(2013)[178] avec l'imageur micro-ondes TMI et le radiomètre micro-ondes AMSR-E au sein du système Goddard WRF-EDAS et a montré de bons résultats notamment sur les prévisions à court terme de précipitations. Plus tard, Chambon et al. (2014)[179] évalueront à nouveau cette méthode à l'aide de sept instruments micro-ondes, dont trois imageurs SSMI/S, le radiomètre AMSR-E et trois sondeurs MHS. Les résultats de cette étude ont permis de réduire l'erreur quadratique moyenne des prévisions de précipitations de pluie cumulées sur un cas d'étude.

3.4 Avantages et inconvénients des différentes méthodes

Dans cette partie nous allons décrire les avantages et les inconvénients des différentes méthodes d'assimilation d'observations nuageuses et/ou précipitantes qui sont pour rappel : l'initialisation physique, la méthode 1D-Var+4D-Var, la méthode 1D-Bay+4D-Var, l'assimilation directe par deux méthodes différentes : le 4D-Var et le filtre ensembliste. Nous allons tout d'abord détailler les avantages et inconvénients de la méthode de l'initialisation physique, puis nous comparerons les méthodes indirectes aux méthodes directes.

L'initialisation physique, contrairement aux autres méthodes, ne satisfait pas un critère de minimisation entre les observations et le modèle, et de cette manière ne fournit pas un résultat optimal. Néanmoins, l'initialisation physique ne requiert pas le développement des modèles linéaires tangents et d'adjoints de la microphysique ou de calculs complexes de matrices de covariances comme dans les méthodes directes. L'initialisation physique représente de plus une bonne alternative de par son implémentation simple, ne nécessitant que l'ajout d'un incrément dans les équations pronostiques du modèle. De cela résulte un coût de calcul relativement faible comparé aux méthodes variationnelles qui nécessitent l'inversion complexe de processus non linéaires.

Le passage progressif des méthodes indirectes aux méthodes directes dans les centres de PNT, s'est fait pour plusieurs raisons. Tout d'abord les méthodes indirectes considèrent comme étape préliminaire un premier algorithme d'inversion qui utilise l'ébauche du modèle : 1D-Var ou un 1D-Bay. Ensuite, les méthodes indirectes créent un biais d'échantillonnage en assimilant les observations claires directement au sein d'un 4D-Var, différemment des observations nuageuses (Geer et al. 2008[24]). L'assimilation directe, en adoptant une approche "all-sky" pour traiter à la fois les observations claires et nuageuses dans un 4D-Var, évite de créer ce biais qui impacte la qualité des prévisions. Néanmoins les méthodes indirectes ont des avantages par rapport à l'assimilation directe. Premièrement, la méthode indirecte 1D-Bay+4D-Var ne souffre pas du problème du "gradient nul" et peut donc produire de la pluie là où le modèle n'en possède pas. Enfin, les méthodes indirectes peuvent être également exploitées pour des inversions en paramètres géophysiques et des diagnostics a posteriori.

Si l'on compare maintenant les méthodes indirectes entre elles, un avantage pour la méthode 1D-Bay sur la méthode 1D-Var est qu'elle ne nécessite pas la linéarisation du schéma de convection. Le problème principal est la validité de la linéarisation, car les processus physiques possèdent des seuils et des non-linéarités, qui nécessitent beaucoup de précautions quand on développe les versions TL/AD (Janisková et Lopez 2013[182]). Une autre différence réside dans la corrélation des deux méthodes au modèle. En effet, la méthode 1D-Bay+4D-Var est moins corrélée au modèle que la méthode 1D+4D-Var grâce à deux points :

- Au sein de la méthode 1D-Bay, une hypothèse sur la base de données d'inversion est réalisée : chaque profil peut être obtenu de manière équiprobable. La méthode 1D-Var, qui utilise une matrice \mathbf{B} pour l'étape d'inversion, attribue quant à elle plus de poids aux profils, augmentant de ce fait la corrélation à l'ébauche.
- Le profil du modèle à la localisation de l'observation est retiré de la base de données d'inversion.

Enfin, si l'on compare les méthodes directes entre elles, la méthode directe par filtre ensembliste a l'avantage de ne pas nécessiter des modèles linéaires tangents et d'adjoints de la microphysique (car la convection est explicitement résolue et les hydrométéores font partie de la variable de contrôle) et du transfert radiatif. Pour cela, le système utilise un ensemble de simulations du modèle non linéaires pour relier l'espace du modèle à l'espace des observations. L'inconvénient principal réside dans l'exploitation d'un ensemble de prévision qui nécessite des calculs numériques conséquents.

Toutes les méthodes possèdent donc des inconvénients et des avantages qu'il est important d'identifier afin de les implémenter de manière opérationnelle au sein des systèmes de PNT. Dans cette thèse, nous examinerons deux de ces méthodes, la méthode 1D-Bay+4D-Var aux chapitres 5 et 6 puis la méthode d'assimilation directe au chapitre 7.

Chapitre 4

Problématique de la thèse

Sommaire

4.1	Contexte	63
4.2	Outils	66
4.3	Méthodologie	71
4.4	Plan de la suite du manuscrit	72

Nous avons vu dans les chapitres 2 et 3 que l’assimilation des hautes fréquences micro-ondes dans les systèmes de PNT soulevait plusieurs problématiques, notamment par rapport à la sélection des SSPs des hydrométéores solides au sein de l’opérateur d’observations. Cette thèse se concentre sur les incertitudes apportées par ces SSPs lors de l’assimilation d’observations satellitaires micro-ondes, le but final étant de proposer une méthode permettant d’en tenir compte au sein des systèmes de PNT de Météo-France. Nous allons tout d’abord expliquer les différentes méthodes déjà développées dans d’autres centres de PNT pour prendre en compte la diversité des SSPs disponibles dans les bases de données. Ensuite nous introduirons les outils utilisés pour explorer cette problématique au CNRM (Centre National de Recherches Météorologiques) à Météo-France, ainsi que les différentes méthodes que nous développerons au cours de ce travail. Nous finirons ce chapitre en présentant l’organisation des chapitres suivants.

4.1 Contexte

Nous allons voir dans cette section ce qui nous a amené à construire des méthodes permettant de traiter un ensemble de particules afin de prendre en compte les incertitudes liées au transfert radiatif. L’assimilation des observations satellitaires micro-ondes montre des résultats très positifs dans la plupart des centres de PNT opérationnels, comme cela a été présenté au chapitre 1. En effet, elle conduit généralement à des améliorations des prévisions plus significatives que d’autres types d’observations (Bormann et al. 2019, Chambon et al. 2022). Cela s’explique par le nombre d’instruments micro-ondes actuellement en vol mais aussi par la sensibilité moindre des rayonnements micro-ondes à la présence des nuages, contrairement aux rayonnements infrarouges, ainsi que par leur sensibilité à la vapeur d’eau, à la température et aux différents hydrométéores. Dans les années à venir, plusieurs instruments micro-ondes vont être embarqués à bord des futurs satellites MetOp-SG (Meteorological Operational-Second Generation) qui seront lancés à partir de 2024 dans le cadre du programme EPS-SG (Eumetsat Polar System Second Generation) : le sondeur MWS, l’imageur MWI et l’imageur ICI qui ont respectivement des fréquences s’étalant de 23.8 GHz à 229 GHz, de 18.7 GHz à 183.31 GHz et de 183.31 GHz jusqu’aux fréquences sub-millimétriques à 664 GHz. Il est de ce fait important de préparer l’assimilation de ces instruments qui vont restituer un nombre important d’informations sur la structure verticale de l’atmosphère en ciel clair et dans les zones nuageuses.

Ces instruments possèdent des fréquences sensibles à la vapeur d’eau qui vont permettre de renseigner l’atmosphère en présence de nuages. Pour assimiler les observations localisées en zones nuageuses, les

SSPs des hydrométéores, ainsi que leur distribution en tailles, doivent être spécifiées dans l'opérateur d'observations. Ainsi, les paramètres nécessaires à la résolution de l'équation du transfert radiatif, vue au chapitre 2, seront disponibles pour la simulation des températures de brillances. Les hydrométéores liquides (gouttes de pluie, gouttelettes de nuages) sont simulés à l'aide de formes simples généralement modélisées par des sphères ou des sphéroïdes aplatis correspondants à celles qu'ils prennent lors de leur chute dans l'atmosphère (Brandes et al. 2002[27], Augros et al. 2016[28], Ekelund et al. 2020[29]). Au contraire, les hydrométéores solides (neige, grêle, cristaux) sont caractérisés par une grande variabilité morphologique dans la nature conduisant ainsi à une grande variabilité de leurs SSPs.

Dans le chapitre 2, plusieurs bases de données répertoriant un ensemble de SSPs ont été présentées (Liu2008[11], Hong2007[32], Ding2017[40], ...). Parmi ces SSPs, on trouve des formes telles que les étoiles, les dendrites, les colonnes, les aiguilles, les plaquettes ou encore les agrégats. Par souci de clarté, nous appellerons dans la suite de ce manuscrit les SSPs par le nom anglais de la forme de particule qui leur est associée. Ces dernières années, le nombre de telles bases de données a connu une augmentation majeure grâce à une meilleure connaissance de la modélisation des processus de formation des particules de neige et à l'augmentation des moyens de calcul. La figure 4.1, tirée de Kneifel et al. (2018)[42], présente le nombre de bases de données répertoriant les monocristaux, les agrégats, les particules fondues et les particules givrées des années 1995 jusqu'à 2016. Cependant, malgré le nombre de SSPs disponibles, l'ensemble des systèmes de PNT n'utilise à ce jour qu'une unique configuration du transfert radiatif. Ainsi un seul type de forme et un seul type de distribution de particules sont spécifiés pour modéliser les SSPs de chaque d'hydrométéore, et ceci pour tous types de nuages et sur la totalité du globe. Cette utilisation simplifiée provient d'une limitation de certains systèmes d'assimilation qui nécessitent de linéariser l'opérateur d'observations nuageux autour d'une seule réalisation du transfert radiatif non linéaire. Cette limitation est commune à plusieurs domaines spectraux dont l'infrarouge. Pour pallier cette limitation, diverses méthodes ont été élaborées, et nous allons maintenant en décrire quelques unes.

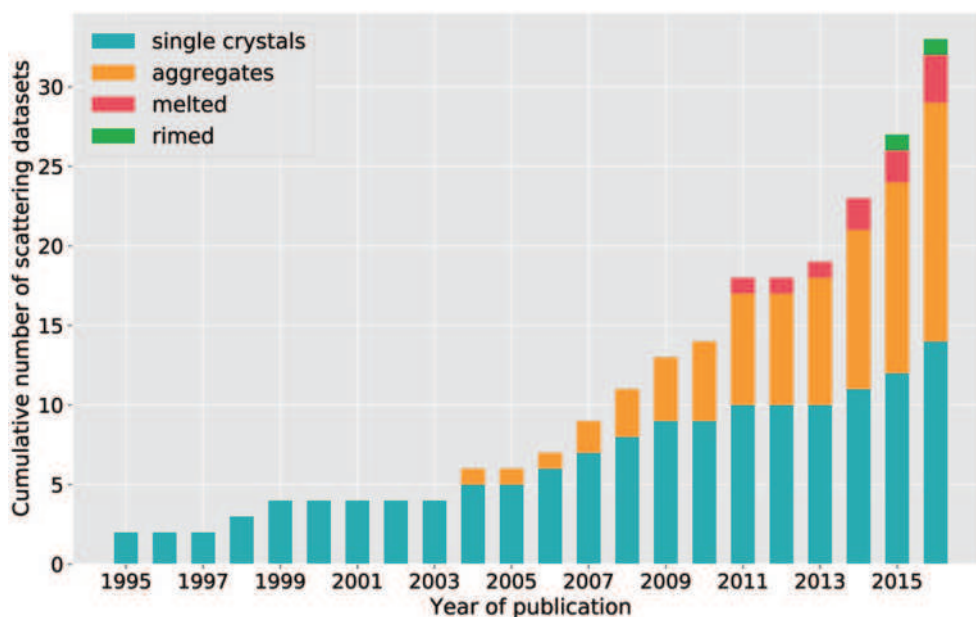


FIGURE 4.1 – Nombre cumulé de bases de données de SSPs pour le spectre micro-ondes publiées entre 1995 et 2016. La couleur indique quelles classes de particules sont incluses dans les ensembles de données : les monocristaux sont en bleu, les agrégats en jaune, les particules fondues en rouge et les particules givrées en vert.

Source: Kneifel et al. 2018[42]

Les premières méthodes cherchaient à définir une manière optimale d'exploiter les SSPs des hydrométéores solides. Baran et Labonnote (2007[6]) ont par exemple développé un modèle pour les cirrus exploitant un ensemble de cristaux pour simuler des températures de brillance dans le domaine infrarouge. Dans cette étude, les SSPs sont paramétrées en fonction du contenu en glace et de la température. Les plus petites particules sont représentées par des colonnes de glace jusqu'à être représentées par des particules de plus

en plus complexes équivalentes à des chaînes de particules lorsque leur taille augmente. Cet ensemble, représenté figure 4.2, a permis d’estimer des quantités physiques telles que l’extinction volumique, le contenu en glace ainsi que l’épaisseur optique de nuages cirrifformes.

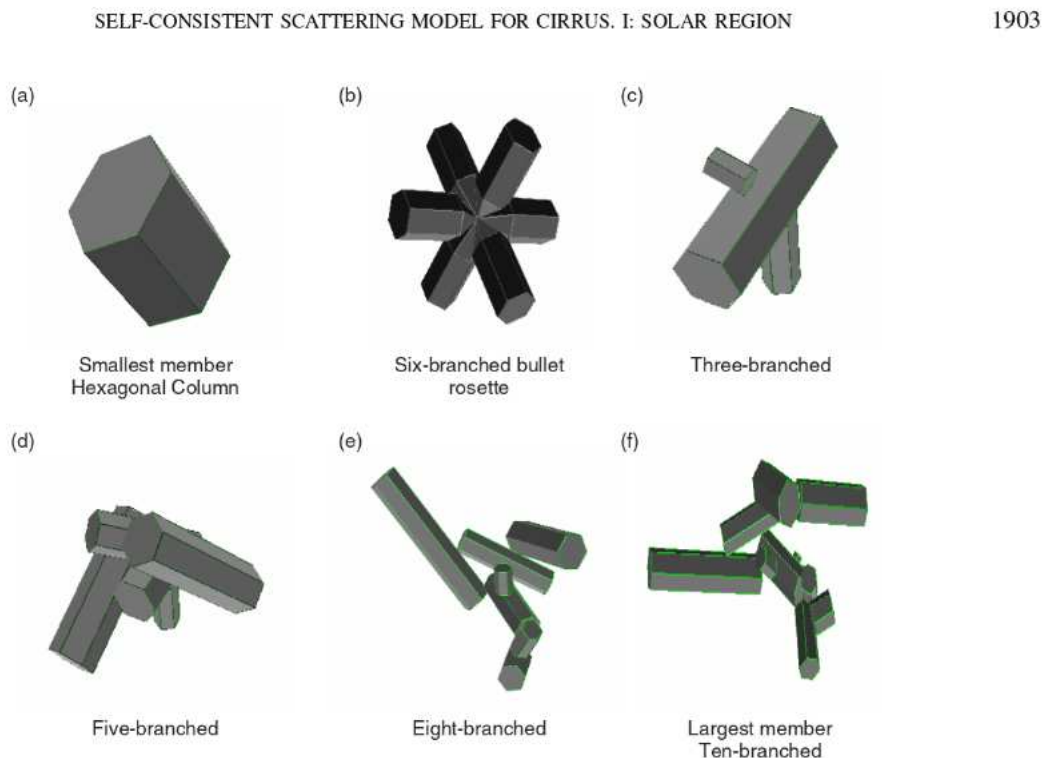


FIGURE 4.2 – Ensemble de formes de particules solides utilisé dans l’étude de Baran et Labonnote (2007)[6].

En 2010, Kulie et al.[7] ont réalisé une moyenne des simulations du transfert radiatif de l’ensemble des SSPs des bases de données de Liu2008[11] et Hong2007[32]. Cet ensemble avait pour but de mettre en évidence les incertitudes résultant de la spécification des SSPs pour simuler des températures de brillance dans le domaine des micro-ondes. Cette étude a ainsi montré des incertitudes importantes selon les fréquences micro-ondes et les régions du globe. De ce fait, ils ont suggéré de construire un ensemble basé sur des critères tels que la latitude, le type de précipitations et le contenu en glace. Ringerud et al. (2019[8]) ont repris cette méthode avec la base de données de Liu2008[11] au sein de l’algorithme de restitution des précipitations exploitant des observations satellitaires micro-ondes Goddard PROFiling algorithm (GPROF, Kummerow et al. 2001[183], Randel et al. 2020[184]). Cet algorithme est basé sur une méthode d’inversion bayésienne simulant le transfert radiatif et nécessite, comme en assimilation de données, une spécification précise des processus micro-physiques. Les résultats de cette étude ont montré que l’utilisation de l’ensemble améliore la restitution des précipitations de l’algorithme ce qui a amené à sa mise en opérationnel en mai 2017.

D’autres études ont au contraire eu pour objectif de trouver une forme de particule optimale pour simuler les SSPs de la neige (aussi appelée ‘one-shape-fits-all’). Geer et Baordo en 2014[1] ont mené cette étude à l’aide de la base de données de Liu2008[11] et des sphères de Mie avec la version 11 de RTTOV-SCATT. Les résultats de cette étude ont montré que la forme *Sector snowflake* représentait un bon compromis pour la simulation des températures de brillance à diverses fréquences, amenant à son utilisation en opérationnel dans le système du CEPMMT à partir de 2013. Ce comportement se retrouve sur la figure 4.3 qui montre la distribution globale des moyennes des innovations¹ réalisées avec plusieurs SSPs à la fréquence 183.31 ± 7 GHz. On remarque que les moyennes sont plus faibles lorsque les températures de brillance sont simulées à l’aide de la SSP *Sector snowflake*. Des résultats moins satisfaisants mais meilleurs qu’avec les sphères de Mie sont également obtenus avec la particule *Rosette 3-bullets*. Geer

1. Le vecteur d’innovation vaut $y_o - H(x_b)$ avec y_o les observations, x_b l’ébauche et H l’opérateur d’observation

2021[185] a par la suite réalisé une étude plus générale pour adapter la représentation de 6 paramètres (le recouvrement nuageux, la distribution en taille de la neige convective, la forme des particules de neige convective et de grande échelle, ainsi que la forme et la distribution en taille des particules de glace) au sein de la version 13 de RTTOV-SCATT. Pour cela, il a réalisé une estimation de paramètres basée sur une technique d'apprentissage des paramètres du modèle à l'aide d'assimilations. Dans cette étude, plusieurs combinaisons de distributions en taille et de SSPs sont testées. La configuration finale choisie au sein de RTTOV-SCATT est une combinaison des SSPs de la base de Eriksson2018[41] et de différentes distributions :

- Pour la glace nuageuse, la particule *large column aggregate* est combinée à une distribution de type Gamma[66].
- Pour la neige stratiforme, la particule *large plate aggregate* est combinée à la distribution de Field et al. 2007[127].
- Et pour la neige convective, la particule *column* est combinée à la distribution de Field et al. 2007[127].

Ainsi, on voit à travers ces différentes tentatives qu'il est possible d'exploiter différentes SSPs au sein des modèles de transfert radiatif à l'aide de moyennes d'ensembles ou d'opter pour une SSP supposée optimale.

Néanmoins, la plupart des méthodes utilisées opérationnellement en assimilation de données conduisent à spécifier une unique forme de particule ou moyenne de formes par hydrométéore amenant ainsi à une unique réalisation du transfert radiatif par scène météorologique. Ceci s'explique par le fait que les systèmes d'assimilation actuels ne peuvent spécifier qu'une seule SSP pour chaque hydrométéore à l'opérateur d'observation pour simuler les températures de brillance en un point d'observation. Compte tenu de la diversité des formes dans la nature, cela représente une très grande simplification de l'atmosphère et l'utilisation de plusieurs formes directement dans l'assimilation de données n'a pas encore été explorée. Le nombre de base de données de SSPs grandissant, il devient de plus en plus complexe de choisir les meilleures spécifications pour le transfert radiatif, les méthodes visant à estimer une particule 'optimale' semblent ainsi proches de leurs limites. Une solution serait donc en effet de ne plus avoir à spécifier une unique forme dans l'assimilation mais d'en exploiter plusieurs. Au cours de cette thèse, nous nous sommes attachés à l'élaboration de telles méthodes, traitant plusieurs SSPs et propageant les incertitudes induites par leur spécification au sein des simulations. Nous allons maintenant présenter les différents outils mises en oeuvre au sein de cette thèse.

4.2 Outils

Modèles numériques

Nous avons mené les études à l'aide de deux modèles de PNT : le modèle AROME-Antilles et le modèle ARPEGE. Le modèle AROME-Antilles ne possède pas de système d'assimilation de données mais constitue une bonne base pour obtenir des résultats préliminaires dans des régions caractérisées par des phénomènes météorologiques majeurs, tels que les cyclones. En effet, ce modèle résout explicitement la convection profonde, et possède une microphysique détaillée des systèmes convectifs précipitants. Les prévisions de ce modèle seront exploitées au sein d'une méthode d'assimilation pour laquelle une étude de sensibilité a préalablement été menée et a montré une robustesse aux échéances exploitées. Ainsi, l'ensemble des études exploitant ce modèle ont été menées dans ce manuscrit avec des prévisions 3 H. Par la suite, les études seront généralisées avec le modèle ARPEGE qui, grâce à sa couverture globale, permet l'assimilation d'un grand nombre d'observations satellitaires micro-ondes, mais décrit les phénomènes convectifs de manière moins réaliste (paramétrisation de la convection profonde notamment).

Observations spatiales

Nous avons choisi d'exploiter les observations satellitaires micro-ondes du radiomètre à bord du satellite GPM Core Observatory (Global Precipitation Measurement Core Observatory) pour réaliser les diffé-

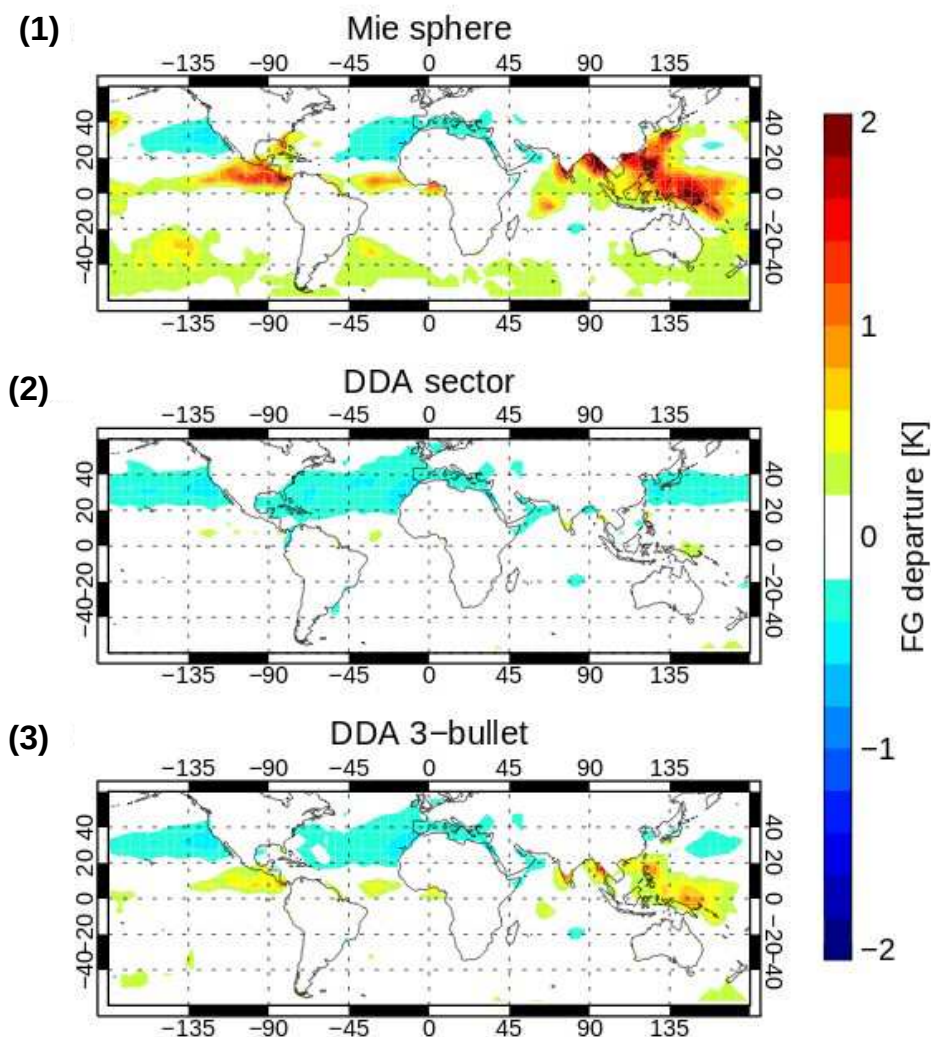


FIGURE 4.3 – Moyenne des innovations du modèle IFS et des observations du radiomètre SSMI/S pour le mois de Juin 2012 à la fréquence 183.31 ± 7 pour (1) les sphères de Mie, (2) la SSP *Sector snowflake* et (3) la SSP *Rosette 3-bullets* de la base de Liu2008[11]. Les moyennes sont calculées sur une grille $5^\circ \times 5^\circ$.

Source: Geer et Baordo et al. 2014[1]

rentes expériences qui seront décrites dans cette thèse. Le GPM Core Observatory est un satellite défilant situé à une altitude de 407 km et ayant une orbite inclinée à 65° par rapport à l'équateur. Il a été développé par les agences spatiales japonaise (JAXA, Japan Aerospace Exploration Agency) et américaine (NASA) puis lancé le 27 février 2014 dans le cadre du programme GPM. Cette mission vise à améliorer les prévisions météorologiques des événements extrêmes et la compréhension du cycle de l'eau et de l'énergie de la Terre. Ce satellite possède à son bord des instruments micro-ondes actif et passif permettant de mesurer les précipitations au niveau des Tropiques jusqu'aux moyennes latitudes : le radar DPR (Dual-frequency Precipitation Radar), qui est pourvu d'une bande Ku à 13.6 GHz et d'une bande Ka à 35.5 GHz, ainsi que le radiomètre micro-ondes GPM Microwave Imager (GMI, figure 4.4) qui sera l'instrument utilisé tout au long de ce manuscrit.

Ce dernier possède plusieurs canaux dont les caractéristiques sont résumées dans le tableau 4.1 : 9 canaux basses fréquences s'étalant de 10.65 GHz à 89 GHz et 4 canaux hautes fréquences de 166 à 183 GHz.

Chacune de ces fréquences permet d'acquérir des informations complémentaires sur l'atmosphère (Hou et al. 2014[186]) :

- les canaux à 10 GHz sont sensibles aux précipitations liquides ;
- les canaux à 18 GHz et 37 GHz sont sensibles aux précipitations modérées ;
- le canal à 23 GHz (bande d'absorption faible de la vapeur d'eau) est utile pour l'estimation de la colonne de vapeur d'eau intégrée ;
- les canaux à 89 GHz sont sensibles aux particules de glace ;
- les canaux à 166 GHz sont sensibles aux précipitations solides et à la glace ;
- les canaux centrés autour de la fréquence 183 GHz (bande d'absorption forte de la vapeur d'eau) sont sensibles aux précipitations solides et à la glace, ainsi qu'à la vapeur d'eau.

La diversité dans les fréquences de GMI va permettre d'étudier les profils atmosphériques sur toute la verticale de l'atmosphère, mais aussi pour différents types d'hydrométéores. Les hautes fréquences vont être particulièrement utiles pour étudier les incertitudes liées au choix des SSPs au sein de l'opérateur d'observation simulant les radiances micro-ondes. Enfin, il permettra de préparer l'assimilation des instruments MWS, MWI et ICI en approfondissant les recherches sur l'assimilation des hautes fréquences micro-ondes.

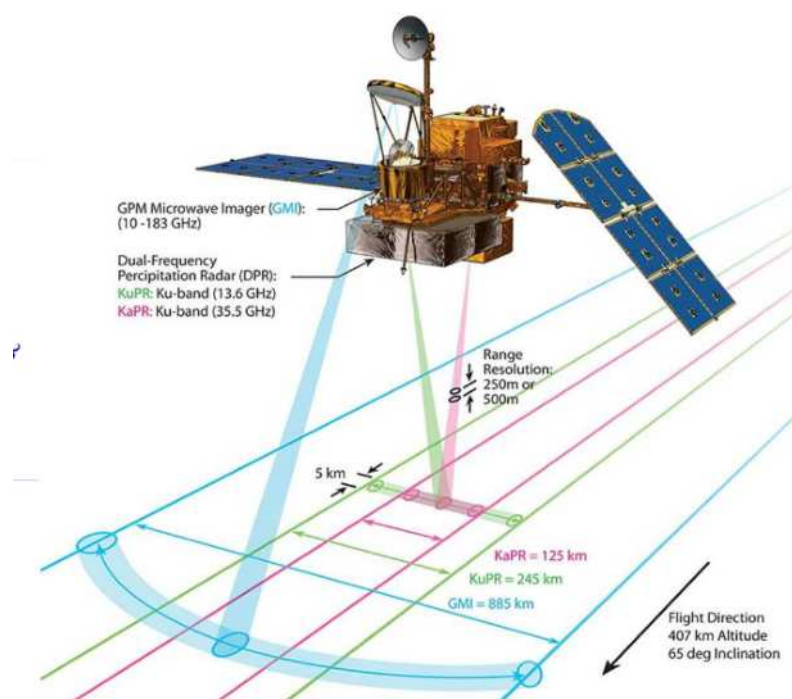


FIGURE 4.4 – Satellite GPM Core Observatory avec à son bord le radar DPR et le radiomètre micro-ondes GMI. Les fauchées observées par le radar DPR sont représentées respectivement en vert et en rose pour la bande Ku et la bande Ka et la fauchée observée par GMI en bleue.

Source: <http://radiometrie.free.fr>

Du point de vue de la géométrie des observations, deux antennes différentes sont utilisées pour les deux groupes de fréquences. Celles-ci sont caractérisées par deux angles de visée différents et n'observent donc pas la Terre exactement au même endroit et au même moment. Ainsi, lors d'un début d'acquisition des données, un écart de quelques kilomètres existe entre la fauchée des basses et moyennes fréquences et celle des hautes. Une colocalisation entre les basses et les hautes fréquences a donc été réalisée avant le début des études. Pour cela, les observations ont été projetées sur une grille régulière en latitude et en longitude

ID	Fréquences centrale (GHz)	Largeur de bande (MHz)	Polarisation	NE Δ T (°)	Angle (K)	IFOV (°)	Résolution (°)
1	10.65	100	V	0.96	52.8	19×32	/
2	10.65	100	H	0.96	52.8	19×32	/
3	18.7	200	V	0.84	52.8	11× 18	0.1 × 0.1
4	18.7	200	H	0.84	52.8	11× 18	0.1 × 0.1
5	23.8	400	V	1.05	52.8	9.2 × 15	0.1 × 0.1
6	36.64	1000	V	0.65	52.8	8.6 × 14	0.1 × 0.1
7	36.64	1000	H	0.65	52.8	8.6 × 14	0.1 × 0.1
8	89.0	6000	V	0.57	52.8	4.4 × 7.2	0.1 × 0.1
9	89.0	6000	H	0.57	52.8	4.4 × 7.2	0.1 × 0.1
10	166.0	4000	V	1.5	49.1	4.4 × 7.2	0.1 × 0.1
11	166.0	4000	H	1.5	49.1	4.4 × 7.2	0.1 × 0.1
12	183.3±3	2000	V	1.5	49.1	4.4 × 7.2	0.1 × 0.1
13	183.3±7	2000	V	1.5	49.1	4.4 × 7.2	0.1 × 0.1

TABLE 4.1 – Liste des caractéristiques des canaux du radiomètre GMI. IFOV (the Instantaneous Field of View) désigne les champs de visée instantanés et NE Δ T (the Noise Equivalent Delta Temperature) la température équivalente au bruit.

Source: <https://www.wmo-sat.info/oscar/instruments/view/158>

à une résolution de 0.1° (resp. 0.25°), compatible avec la résolution effective du modèle AROME (resp. ARPEGE).

Les fréquences micro-ondes étant très sensibles à l'émissivité de surface, seules les observations sur océan du radiomètre micro-ondes GMI seront assimilées au sein des différentes expériences détaillées dans les prochains chapitres. De plus, nous considérerons ici une matrice de covariances d'erreurs d'observations \mathbf{R} diagonale qui est une hypothèse forte réalisée en assimilation de données. Bormann et al. (2003[187]) montra que cette hypothèse était raisonnable lorsqu'un écrémage des observations et qu'une inflation de la variance d'erreur associée était réalisés. En pratique dans le modèle ARPEGE, les observations du radiomètre GMI sont écrémées de sorte à avoir une observation tous les 100 km. Les corrélations inter-canaux ont été négligées au sein des différentes études mais il est possible de les estimer à l'aide des diagnostics de Desroziers (2005[188], Geer 2019[189] et Coopmann et al. 2020[190]).

Code de transfert radiatif et base de données

Nous allons utiliser la version 12.3 du code de transfert radiatif RTTOV-SCATT opérationnel au sein des modèles de PNT de Météo-France AROME et ARPEGE pour simuler les températures de brillance. Ce code de transfert radiatif a fait l'objet de nombreuses améliorations notamment grâce à l'utilisation de la SSP *Sector snowflake* pour la simulation de la neige (Geer et Baordo, 2014[1]) au CEPMMT. Cet opérateur a donc été choisi pour simuler les températures de brillance du radiomètre GMI. Dans cette version de RTTOV-SCATT, il est possible de spécifier les distributions en taille de Marshall et Palmer (1948), de Field et al. (2005), de Field et al. (avec un régime tropical ou des moyennes latitudes, 2007) ou encore une loi Gamma modifiée. La configuration retenue pour décrire les hydrométéores avec RTTOV-SCATT est repris de l'étude préliminaire réalisée avec le sondeur SAPHIR dans Duruisseau et al. (2019[44]) et est résumée dans le tableau 4.2.

Nous avons de plus choisi d'exploiter les SSPs de la base de Liu2008[11] déjà disponible au sein de la version RTTOV-SCATT 12. La base de Eriksson2018[41] était aussi disponible mais nous avons choisi de privilégier une base de données avec moins de SSPs pour réaliser les études afin de faciliter l'interprétation des résultats.

Afin de montrer la pertinence de la problématique que nous souhaitons étudier, les distributions en températures de brillance simulées pour chaque SSP de la base de Liu2008[11] sont tracées sur les figures 4.5 et 4.6. Les données proviennent de deux modèles : le modèle global ARPEGE (sur une durée de 10 jours) et le modèle AROME-Antilles (sur une durée de 2 mois). Sur ces deux graphiques, on observe première-

Type d'hydrométéore	Distribution en taille des particules	SSP pour l'absorption et la diffusion
Pluie	Marshall-Parlmer[123]	Sphère de Mie
Neige	Field et al. 2007[127] régime tropical	Les 11 SSPs de Liu2008[11] calculées avec la méthode DDA
Eau liquide nuageuse	Gamma modifiée[66]	Sphère de Mie
Eau glacée nuageuse	Gamma modifiée[66]	Sphère de Mie

TABLE 4.2 – Configuration de RTTOV-SCATT choisie pour les études présentées dans le manuscrit. Grâce à l'équilibre thermodynamique local, l'absorption est ici égale à l'émission.

ment que les distributions en températures de brillance simulées sont très variables selon la SSP spécifiée dans l'opérateur d'observation. Par exemple, on remarque que les simulations avec la SSP *Block column* conduisent à des températures beaucoup plus froides qu'avec les simulations réalisées avec la SSP *Dendrite snowflake*. On remarque ensuite que la forme des distributions en températures de brillance simulées à la fréquence 183.31 ± 7 GHz par AROME-Antilles sont plus proches de la distribution observée que celles simulées par le modèle ARPEGE. En effet, les distributions du modèle AROME-Antilles sont caractérisées par une inflexion coïncidant avec l'aspect de la distribution observées contrairement aux distributions du modèle ARPEGE. Ainsi, l'allure des distributions du modèle ARPEGE (figure 4.5) ne ressemblent pas à celles du modèle AROME-Antilles (figure 4.6). Ceci s'explique par les différences entre ces deux modèles. Le modèle AROME-Antilles est un modèle à aire limitée qui résout la convection profonde explicitement alors qu'elle est paramétrisée dans le modèle ARPEGE, conduisant à une représentation très simplifiée des processus microphysiques dans ces zones. Ces différentes caractéristiques permettent au modèle AROME de mieux prévoir les phénomènes météorologiques locaux et convectifs, mais aussi de simuler des distributions en températures de brillance plus réalistes qu'avec le modèle ARPEGE. On remarque, de plus, dans les deux graphiques, qu'aucune distribution ne s'ajuste parfaitement aux températures de brillance observées malgré des distributions plus réalistes avec le modèle AROME-Antilles.

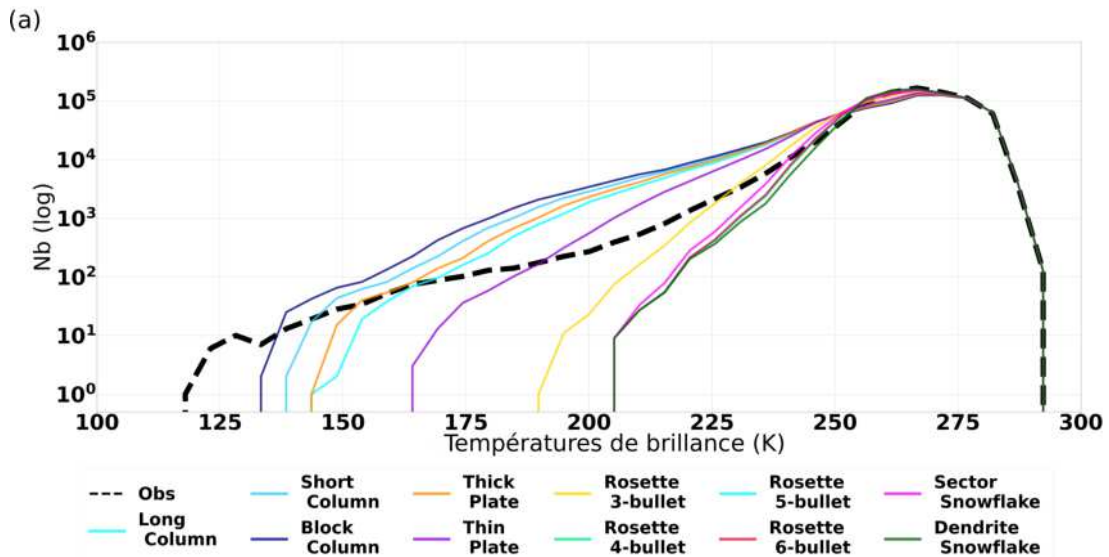


FIGURE 4.5 – Distributions des températures de brillance sur océan simulées à la fréquence 183.31 ± 7 GHz avec les particules de la base Liu2008[11] et le modèle ARPEGE sur une période de 10 jours en avril 2021 pour le domaine géographique entre -60° et 60° de latitude. Les distributions des observations du radiomètre GMI sont tracées en noires.

Ces grandes différences dans les températures de brillance simulées sont susceptibles de modifier significativement les résultats d'expériences d'assimilation. Ainsi il est important de définir une méthode tenant compte des incertitudes du transfert radiatif tout en exploitant la diversité des bases de données. La méthode doit également être indépendante des paramétrisations physiques des modèles de PNT. Dans cette thèse nous allons nous concentrer uniquement sur les incertitudes liées à la spécification des SSPs de

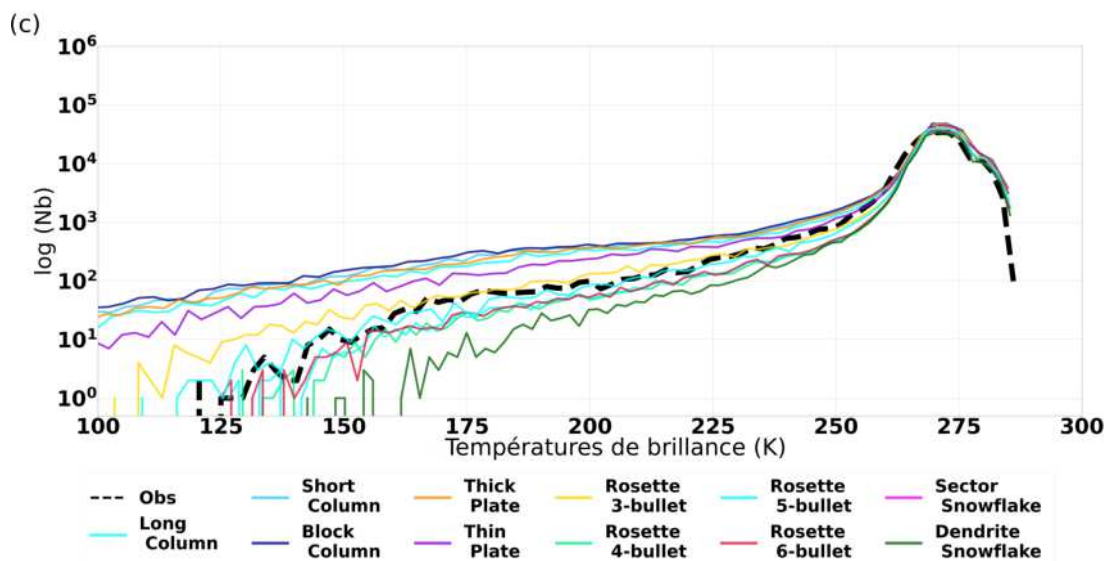


FIGURE 4.6 – Distributions des températures de brillance sur océan simulées à la fréquence 183.31 ± 7 GHz avec les particules de la base Liu2008[11] et le modèle AROME-Antilles sur une période de 2 mois en septembre et octobre 2017. Les distributions des observations du radiomètre GMI sont tracées en noires.

par un consensus général dans la communauté d’assimilation des micro-ondes passives en ciel nuageux qui pointait cette spécification comme étant une des priorités actuelles. Ce n’est pas la seule source d’incertitude du transfert radiatif, la distribution en taille des particules ou encore la spécification des fractions précipitantes dans la maille seraient également à considérer. Cependant, la démarche entreprise dans ce manuscrit est toutefois suffisamment générale pour être appliquée ultérieurement à d’autres sources d’incertitudes du transfert radiatif.

4.3 Méthodologie

L’objectif de cette thèse est donc d’explorer de nouvelles méthodes permettant de ne plus faire de choix de particule optimale a priori au sein des simulations du transfert radiatif. Pour cela nous allons examiner deux questions :

- Les observations ont-elles un contenu en information suffisant pour définir de manière dynamique des combinaisons de SSPs optimales selon la situation météorologique ?
- Peut-on utiliser la diversité des SSPs disponibles comme levier de propagation d’erreurs au sein des systèmes d’assimilation de données ?

Pour répondre à la première question nous allons explorer une méthode à l’aide de la technique d’assimilation de données étudiée à Météo-France, au début de la thèse en 2019, pour les observations satellitaires micro-ondes en ciel nuageux : la technique 1D-Bay+3D/4D-Var (Guerbette et al. 2016[4], Duruisseau et al. 2019[44]). Comme vu dans le chapitre 3, cette méthode est opérationnelle depuis 2010 à Météo-France pour l’assimilation des observations radars (Caumont et al. 2010[19], Wattrelot et al. 2014[20]). Depuis 2022 et grâce à toutes les recherches menées, elle l’est également pour certains radiomètres micro-ondes. Elle a donc été logiquement considérée pour traiter cette problématique. Afin d’adapter cette méthode à plusieurs formes de particules, la formule (4.1) de la moyenne pondérée des profils de l’ébauche au voisinage de l’observation au sein de l’inversion Bayésienne, vue au chapitre 3, a été modifiée. La nouvelle formule s’exprime avec une somme additionnelle sur les SSPs telle que :

$$x_{inv} := \frac{\sum_{j=1}^p \sum_{i=1}^n x_i w_{i,j}}{\sum_{j=1}^p \sum_{i=1}^n w_{i,j}}$$

Avec $w_{i,j} = e^{(-\frac{1}{2}[y_o - H_j(x_i)]^t R^{-1}[y_o - H_j(x_i)])}$, p le nombre de SSPs prise en compte au sein de l'inversion Bayésienne, n le nombre de profils dans le voisinage, y_o l'observation, x_i les profils du voisinage, \mathbf{R} la matrice de covariances d'erreurs d'observations et H_j l'opérateur d'observation simulant les températures de brillance à l'aide de la SSP j . Cette formule permet ainsi d'obtenir un profil en humidité relative inversé où les incertitudes liées au transfert radiatif auront été propagées. Chaque profil est ici pondéré par sa ressemblance à l'observation, ce qui permet de donner un poids plus fort aux SSPs simulant des températures de brillance proches de l'observation.

Pour répondre à la deuxième question, nous allons exploiter la technique de l'assimilation directe introduite dans le chapitre 3. Météo-France envisageait au début de cette thèse de remplacer dans les années à venir la technique 1D-Bay+4D-Var par une assimilation directe, d'où notre choix d'abandonner la méthode 1D-Bay+4D-Var et d'explorer une méthode dans ce cadre. Néanmoins il est important de préciser que les résultats obtenus ici avec la méthode 1D-Bay+4D-Var seront utiles à la fois pour la communauté d'assimilation de données mais aussi pour celle d'inversion de produits géophysiques. Dans la nouvelle méthode nous allons définir une erreur caractérisant les incertitudes liées à la spécification des hydrométéores solides au sein des simulations du transfert radiatif. Cette erreur sera par la suite propagée tout du long du processus d'assimilation de données. Pour calculer cette erreur, nous allons nous servir de la technique d'assimilation d'ensemble AEARP abordée lors du chapitre 1. Pour rappel, cet ensemble permet de construire une matrice \mathbf{B} de covariances d'erreurs d'ébauche utilisée au sein de l'assimilation directe à l'aide d'un ensemble à 50 membres. Pour étudier l'intérêt d'une erreur propageant les incertitudes liées au transfert radiatif, nous allons définir une expérience où chaque membre de l'ensemble se verra attribuer aléatoirement des SSPs parmi celles de la base de Liu2008. Les erreurs d'observations liées aux innovations seront aussi adaptées à chaque SSP. La matrice \mathbf{B} résultante sera couplée à une assimilation directe, ce qui permettra de propager les incertitudes liées au transfert radiatif.

L'évaluation des résultats sera faite ici en calculant des erreurs de prévision jusqu'à 96h d'échéance. Pour cela, nous comparerons premièrement les différentes expériences menées à une référence à atteindre. Les références choisies dans ce manuscrit sont les prévisions provenant du modèle IFS exploité par le CEPMMT. Comme ceci, il sera possible de calculer des erreurs de prévision entre les expériences et la référence. Les différents résultats seront finalement comparés entre eux afin d'évaluer quelles expériences ce sont rapprochées le plus de la référence. Ces scores sont généralement calculés à Météo-France pour différentes variables telles que le vent, l'humidité ou la température, pour différents niveaux de pression et pour différents domaines géographiques : NORD20, SUD20, TROPICS, EUROPE, AMÉRIQUE DU NORD et AUSTRALIE. Ici les scores seront seulement calculés sur les domaines NORD20, SUD20 et TROPICS dont on peut voir les délimitations figure 4.7. De plus, on examinera seulement les résultats à des niveaux de pression compris entre 400 hPa et la surface, les autres niveaux étant moins sensibles à l'assimilation du profil d'humidité relative aux niveaux définis dans ARPEGE.

4.4 Plan de la suite du manuscrit

Dans les prochains chapitres de ce manuscrit nous allons réaliser des études préliminaires sur l'inversion Bayésienne, première étape de la technique 1D-Bay+3D/4D-Var, avec le modèle AROME. Puis nous mènerons des expériences d'assimilation complètes 1D-Bay+4D-Var à l'aide du modèle ARPEGE. Nous allons ainsi premièrement réaliser une étude de sensibilité à la méthode de l'inversion Bayésienne (chapitre 5). Cela permettra d'adapter et d'optimiser cette étape pour la suite des études portant sur la propagation des incertitudes provenant des SSPs. Dans le chapitre 6 nous allons étudier la première méthode exploitant les observations comme information a priori pour définir des combinaisons de SSPs exploitées au sein de la technique 1D-Bay+3D/4D-Var. Nous allons tout d'abord examiner des résultats d'inversions réalisées avec

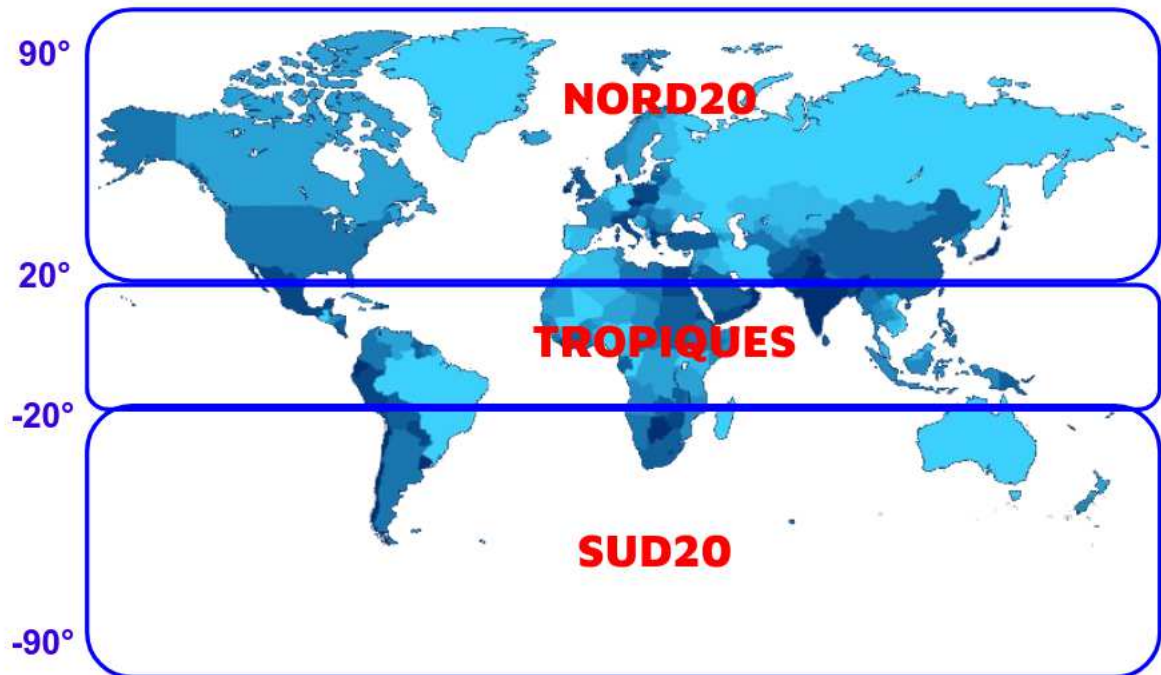


FIGURE 4.7 – Ensemble des domaines sur lesquels les calculs de score ont été réalisés au sein de ce manuscrit. Le domaine NORD20, TROPICS et SUD20 sont ici respectivement délimité par des latitudes minimales de 20°N, 20°S et 90°S et maximales de 90°N, 20°N et 20°S.

le modèle à aire limitée AROME-Antilles et ensuite des expériences d'assimilation avec la technique 1D-Bay+4D-Var seront réalisées avec le modèle global ARPEGE sur une période de deux mois. Finalement, lors du chapitre 7, nous examinerons la deuxième méthode considérant un ensemble d'assimilation de données de type 4D-Var (AEARP) pour construire une matrice \mathbf{B} de covariances d'erreurs d'ébauches. Les matrices construites seront par la suite utilisées dans une assimilation directe dans un 4D-Var, afin de propager les incertitudes introduites par les SSPs. Ces expériences seront menées sur une période de trois mois avec le modèle ARPEGE.

Chapitre 5

Étude de sensibilité de l'inversion Bayésienne

Sommaire

5.1	Introduction	75
5.2	Publication	77
5.3	Conclusions	104

5.1 Introduction

Dans ce chapitre, nous allons examiner la sensibilité à divers paramètres à la première étape de la méthode d'assimilation des observations satellitaires micro-ondes nuageuses et/ou précipitantes 1D-Bay+3D/4D-Var développée à Météo-France : l'inversion Bayésienne. Pour rappel, la première étape de cette méthode, basée sur le théorème de Bayes, consiste à inverser un ensemble de températures de brillance micro-ondes pour restituer un profil vertical d'humidité relative, en s'appuyant sur une information a priori constituée par les profils atmosphériques, dans le voisinage de l'observation, d'une prévision numérique à courte échéance. Dans une deuxième étape, les profils d'humidité relative restitués sont introduits comme de nouvelles observations au sein d'un système d'assimilation de type 3D-Var ou d'un 4D-Var.

Comme vu au chapitre 3, cette méthode est utilisée de manière opérationnelle à Météo-France depuis 2010 pour l'assimilation des réflectivités radar en conditions pluvieuses au sein du modèle à échelle convective AROME (Caumont et al. 2010[19], Wattrelot et al. 2014[20]). Cette implémentation a permis d'améliorer les prévisions à courte échéance, et en particulier les prévisions des précipitations. Plus récemment, des études sur l'assimilation des températures de brillance aux hautes fréquences micro-ondes en condition nuageuse ont été menées à Météo-France à l'aide de cette méthode (Guerbette et al. 2016[4]). Le premier instrument exploité a été le sondeur SAPHIR (Sondeur Atmosphérique du Profil d'Humidité Intertropical par Radiométrie) se trouvant à bord du satellite franco-indien Megha-Tropiques (Brogniez et al. 2015[191]). Ce sondeur possède 6 canaux centrés sur la fréquence 183.31 GHz qui ont été assimilés en ciel clair à Météo-France entre 2015 et 2021 (Chambon et al. 2015[192]). Cette bande de fréquences est intéressante car, en plus de fournir des informations sur le profil de vapeur d'eau, elle est particulièrement sensible aux hydrométéores solides. Les premières assimilations dans le modèle global ARPEGE des données de SAPHIR en ciel nuageux ont considéré les 4 fréquences les plus éloignées du centre de la raie d'absorption pour restituer une information sur l'humidité relative à 4 niveaux de pression. Duruisseau et al. 2019[44] ont ainsi montré une amélioration de la qualité des prévisions de vent, de température et d'humidité relative. Cette méthode a été étendue à d'autres instruments et fréquences micro-ondes dans les systèmes d'assimilation de données de Météo-France, et d'autres pistes sont explorées afin d'améliorer encore les premiers impacts obtenus. Dans ce but, le radiomètre micro-ondes GMI a été choisi pour réaliser des études de sensibilité à l'inversion Bayésienne en raison de sa grande diversité de fréquences.

La méthode 1D-Bay+3D/4D-Var a également été implémentée dans le modèle méso-échelle du centre météorologique Japonais JMA pour assimiler les observations du radar DPR (Dual-frequency Precipitation Radar) à bord du satellite GPM-Core. Afin de répondre à la problématique de distributions de probabilités multimodales des poids de l'inversion Bayésienne (Ikuta 2021[22]), un estimateur basé sur un maximum de vraisemblance a été comparé à la moyenne pondérée utilisée à Météo-France. En utilisant un maximum de vraisemblance, le profil inversé est le profil de l'ébauche dans le voisinage du point observé ayant le poids le plus fort, autrement dit, dont les températures de brillance simulées sont les plus proches de celles observées. Cette méthode, nommée '1D-MLE+4D-Var', est opérationnelle depuis 2016 à la JMA (Ikuta et Honda 2011[193], Ikuta 2021[22]). Elle a également été testée à Météo-France avec les observations du radar aéroporté RASTA (Radar Airbone System Tool for Atmosphere) et a montré de bons résultats (Borderies et al. 2018[194]).

L'inversion Bayésienne possède plusieurs degrés de liberté qu'il convient d'étudier afin d'optimiser les résultats des inversions. Dans l'étude présentée ici, sont testées les sensibilités de la méthode à quatre paramètres différents dont voici une brève description :

La matrice de covariances d'erreurs d'observations Le premier degré de liberté que nous allons étudier concerne les erreurs d'observations, spécifiées dans la matrice \mathbf{R} , qui pour rappel est la matrice de covariances d'erreurs d'observations. Les corrélations des erreurs d'observations inter-canaux n'étant ici pas prises en compte, la matrice \mathbf{R} est de ce fait diagonale. L'amplitude des termes diagonaux permet de moduler la précision requise sur le résultat de l'inversion. Afin de limiter le nombre d'expériences de sensibilités, la matrice \mathbf{R} diagonale présente la même valeur σ_R^2 pour tous les canaux. La matrice \mathbf{R} combine ici les erreurs d'observations (instrumentales et représentativité) et celles liées à l'opérateur d'observation (code de transfert radiatif). Notre étude de sensibilité évaluera 5 valeurs différentes de σ_R^2 : 1 K, 2 K, 5 K, 10 K et 20 K.

La sélection des canaux Chaque instrument possède un certain nombre de canaux centrés sur des fréquences choisies afin de fournir des mesures informatives sur différents paramètres et altitudes dans l'atmosphère. Néanmoins tous les canaux de chaque instrument ne sont pas systématiquement exploités, pour des raisons de coût de calcul ou de contenu en information. L'étude de la sensibilité au nombre de canaux utilisés dans l'inversion va nous permettre d'étudier l'apport des différentes fréquences et en particulier celui des basses fréquences. Nous réaliserons cela par ajout progressif des canaux de GMI.

Les SSPs Comme nous l'avons vu dans le chapitre 3, plusieurs SSPs peuvent être spécifiées dans un code de transfert radiatif pour simuler la diffusion par les particules. Ceci est particulièrement vrai pour les hydrométéores solides qui revêtent des formes et des densités très variables dans la nature. Cette spécification impacte les températures de brillance simulées et par voie de conséquence les résultats des inversions. Nous allons mener une étude de sensibilité en réalisant des expériences d'inversions Bayésiennes avec les différents SSPs disponibles dans la base de Liu2008[11].

L'estimateur Dans la méthode utilisée à Météo-France, une moyenne pondérée est utilisée comme estimateur pour les profils inversés x_{inv} . Néanmoins d'autres estimateurs existent, notamment l'Estimateur du Maximum de Vraisemblance (EMV) utilisé à la JMA comme évoqué plus haut. Nous allons mener une étude de sensibilité en réalisant des expériences d'inversions Bayésiennes avec ces deux estimateurs.

Au sein de ce chapitre nous avons réalisé des études de sensibilités aux différents paramètres présentés. Pour cela, nous avons exploités des prévisions AROME 3 H ainsi que les observations du radiomètre GMI disponibles sur une période de deux mois entre début septembre et fin octobre 2017, période connue pour la passages successifs des cyclones Irma, José et Maria. Le seuil des inversions réussies a été ici fixé à 10^{-200} afin de traiter un grand nombre d'observations. L'étude réalisée illustrant les sensibilités de l'inversion Bayésienne est détaillée dans un article paru dans le *Journal of the Meteorological Society of Japan* pour un numéro spécial consacré à la mission satellite GPM. L'article est inclus dans la prochaine

section de ce chapitre et dans la dernière section les principales conclusions de l'article sont rappelées et des perspectives proposées.

5.2 Publication

A 1D Bayesian Inversion Applied to GPM Microwave Imager Observations: Sensitivity Studies

Marylis BARREYAT, Philippe CHAMBON, Jean-François MAHFOUF, Ghislain FAURE

CNRM, Université de Toulouse, Météo-France, CNRS, France

and

Yasutaka IKUTA

*Department of Observation and Data Assimilation, Meteorological Research Institute,
Japan Meteorological Agency, Tsukuba, Japan*

Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

(Manuscript received 24 September 2020, in final form 7 April 2021)

Abstract

The assimilation of cloudy and rainy microwave observations is under investigation at Météo-France with a method called “1D-Bay+3D/4D-Var”. This method comprises two steps: (i) a Bayesian inversion of microwave observations and (ii) the assimilation of the retrieved relative humidity profiles in a 3D/4D-Var framework. In this paper, two estimators for the Bayesian inversion are used: either a weighted average (WA) or the maximum likelihood (ML) of a kernel density function. Sensitivity studies over the first step of the method are conducted for different degrees of freedom: the observation error, the channel selection and the scattering properties of frozen hydrometeors in the observation operator. Observations over a 2 month period of the Global Precipitation Measurement (GPM) Microwave Imager (GMI) onboard the Global Precipitation Measurement Core Observatory satellite and forecasts of the convective scale model Application of Research to Operations at Mesoscale (AROME) have been chosen to conduct these studies. Two different meteorological situations are analyzed: those predicted cloudy in AROME but clear in the observations and, those predicted clear in AROME but cloudy in the observations. The main conclusions are as follows: First, low observational errors tend to be associated with the profiles with the highest consistency with the observations. Second, the validity of the retrieved profiles varies vertically with the set of channels used. Third, the radiative properties used in the radiative transfer simulations have a strong influence on the retrieved atmospheric profiles. Finally, the ML estimator has the advantage of being independent of the observation error but is less constrained than the WA estimator when few frequencies are considered. Although the presented sensitivities have been conducted to incorporate the scheme in a data assimilation system, the results may be generalized for geophysical retrieval purposes.

Keywords data assimilation; microwave; cloud; precipitation

Citation Barreyat, M., P. Chambon, J.-F. Mahfouf, G. Faure, and Y. Ikuta, 2021: A 1D Bayesian inversion applied to GPM Microwave Imager observations: Sensitivity studies. *J. Meteor. Soc. Japan*, **99**, 1045–1070, doi:10.2151/jmsj.2021-050.

Corresponding author: Philippe Chambon, CNRM, Météo-France and CNRS, Université de Toulouse, 42 Avenue Coriolis, 31057 Toulouse Cedex, France

E-mail: philippe.chambon@meteo.fr

J-stage Advance Published Date: 30 April 2021



1. Introduction

In the mid-1990s, satellite observations began to be assimilated into Numerical Weather Prediction (NWP) models using radiative transfer codes. However, only clear-sky observations could be processed because of the nonlinearity of the required observation operators, the nonnormality and large variances of representativeness retrievals and other issues listed in Errico et al. (2007). The next step has been to assimilate information from cloudy and rainy microwave radiances. Early studies focused on the assimilation of satellite-derived rainfall rates and showed significant improvements in weather forecasting (Treadon et al. 2002; Marécal and Mahfouf 2002, 2003; Hou et al. 2004; Mahfouf et al. 2005; Hou and Zhang 2007). These methods required different algorithms to convert satellite radiances in rain amounts for each instrument and were therefore abandoned in favor of a gradual transition toward the use of raw brightness temperatures. Then, in 2005, the European Centre for Medium-Range Weather Forecasts (ECMWF) started to assimilate both cloudy and rainy microwave radiances of the Special Sensor Microwave/Imager instrument thanks to a 1D+4D-Var method (Bauer et al. 2006a, b; Geer et al. 2010, 2017). Such an approach required the development of fast radiative transfer modeling with scattering processes, together with their linearized versions, for variational data assimilation. In 2009 this methodology was replaced by a direct assimilation method accounting for model and representativeness errors and also led to better results in their 4D-Var system (Bauer et al. 2010; Geer and Bauer 2010, 2011).

Bayesian inversions have a long history of being used for extracting information on atmospheric hydrometeors from microwave observations (e.g., Kummerow et al. 1996). Building on this heritage, a method named “1D-Bay+3D-Var” to assimilate cloudy and rainy microwave observations, was first used at Météo-France with ground-based weather radar reflectivities (Caumont et al. 2010; Wattrelot et al. 2014). This method seeks to retrieve an optimal relative humidity profile, from an observation and short-term forecast profiles in its vicinity, thanks to a Bayesian inversion. Then, the retrieved profiles of relative humidity are assimilated into a 3D-Var system. This variable was preferred over the assimilation of the specific humidity and temperature profiles to directly impact the saturation of the atmosphere and modulate cloudiness in the forecast, as well as to facilitate the specification of observation errors. Among the advantages of this

method with respect to direct assimilation of cloudy and rainy radiances, which is now more commonly used in the NWP context (Geer et al. 2017), one can mention that (i) it does not rely on tangent linear and adjoint versions of the observation operator, and (ii) it does not suffer from the zero-gradient problem and can generate a cloud where no cloud is present in the first guess by saturating the atmosphere. A potential drawback of the method is that the retrieval to be assimilated can be correlated with the first guess but to a smaller extent than with a 1D-Var+4D-Var (Geer et al. 2008), thanks to a mitigation strategy further explained in Section 3. The Japan Meteorological Agency (JMA) successfully implemented a similar method, with a 4D-Var system, to operationally assimilate observations from spaceborne radars (Ikuta and Honda 2011; Ikuta 2016). More recently, studies have been undertaken at Météo-France with the use of the 1D-Bay+4D-Var method with satellite radiances from the Sondeur Atmosphérique du Profil d’Humidité Intertropicale par Radiométrie (SAPHIR) instrument onboard the Megha-Tropiques satellite (Roca et al. 2020). This water vapor sounder has six frequencies all centered around the 183 GHz and is currently assimilated in the Action de Recherche Petite Echelle Grande Echelle (ARPEGE) global operational NWP model at Météo-France (Courtier et al. 1991) only in clear-sky areas (Chambon et al. 2015). Preliminary results show that cloudy and rainy SAPHIR observations improve wind, temperature, and relative humidity forecasts in the ARPEGE model (Duruiseau et al. 2019). These improvements result from the assimilation of retrieved relative humidities at four different pressure levels. These encouraging studies lead to a desire of extending the method to all microwave instruments. Since the 183 GHz frequency only probes solid hydrometeors, the use of a wider range of microwave channels will allow us to observe liquid hydrometeors as well. The present paper therefore focuses on generalizing the results of Duruiseau et al. (2019) to a larger set of frequencies, as well as investigating various sensitivities of the Bayesian inversion.

In February 2014, the Global Precipitation Measurement Core (GPM-Core) Observatory satellite mission was launched to provide a new standard for nearly global measurements of liquid and solid precipitation from the Tropics to the Midlatitudes (Hou et al. 2014). To that end, it was equipped with two instruments: the Dual-frequency Precipitation Radar (DPR) with two frequencies (13.6 GHz and 35.55 GHz) and the GMI radiometer. The latter is a microwave radiometer with

Table 1. List of GMI channels with several instrument characteristics relevant for the study including the radiometric accuracy and the selected superobbing resolution (Skofronick-Jackson et al. 2017). IFOV denotes the instantaneous fields of view and $Ne\Delta T$ the noise equivalent delta temperature.

ID	Central Frequency (GHz)	Bandwidth (MHz)	Polarization	$Ne\Delta T$ (K)	Viewing Angle ($^{\circ}$)	IFOV (km)	Superobbed Resolution ($^{\circ}$)	Clear Sky Bias of FG departures (K)
1	10.65	100	V	0.96	52.8	19×32	/	/
2	10.65	100	H	0.96	52.8	19×32	/	/
3	18.7	200	V	0.84	52.8	11×18	0.1×0.1	-0.65
4	18.7	200	H	0.84	52.8	11×18	0.1×0.1	0.09
5	23.8	400	V	1.05	52.8	9.2×15	0.1×0.1	-0.92
6	36.64	1000	V	0.65	52.8	8.6×14	0.1×0.1	0.02
7	36.64	1000	H	0.65	52.8	8.6×14	0.1×0.1	-0.96
8	89.0	6000	V	0.57	52.8	4.4×7.2	0.1×0.1	0.33
9	89.0	6000	H	0.57	52.8	4.4×7.2	0.1×0.1	-0.29
10	166.0	4000	V	1.5	49.1	4.4×7.2	0.1×0.1	-0.4
11	166.0	4000	H	1.5	49.1	4.4×7.2	0.1×0.1	-0.2
12	183.31 ± 3	2000	V	1.5	49.1	4.4×7.2	0.1×0.1	0.39
13	183.31 ± 7	2000	V	1.5	49.1	4.4×7.2	0.1×0.1	-0.66

a wide range of frequencies sounding water vapor as well as solid and liquid hydrometeors. The ECMWF started to directly assimilate GMI clear and cloudy radiances into the 4D-Var system in 2017 (Lean et al. 2017; Geer et al. 2017). The Global Modeling and Assimilation Office of the National Aeronautics and Space Administration started the all-sky assimilation as well in July 2018 (Kim et al. 2020). Météo-France assimilates GMI radiances as well but in clear-sky conditions only.

Because of its frequency diversity, the GMI instrument is well suited to study the extension of the “1D-Bay+3D/4D-Var” method to all microwave instruments and complement the preliminary studies conducted with SAPHIR. The present paper describes sensitivity studies regarding the first step of this method, the Bayesian inversion, with the GMI instrument. This inversion involves several parameters for which the impact of their prescribed values must be evaluated to improve the method.

The paper is organized as follows: Section 2 describes the NWP data and the GMI observations selected to perform the sensitivity studies, and Section 3 presents the Bayesian inversion method and the metrics used to evaluate its retrieved results. Section 4 shows the sensitivity results of the inversion on three degrees of freedom: (i) the specification of the observation error, (ii) the channel selection, and (iii) the specification of radiative transfer scattering properties within the observation operator. Finally, Section 5 provides several conclusions from the sensitivity studies with recommendations on future activities.

2. GMI data and NWP system

2.1 Satellite radiances

Level 1B products of the GMI instrument (GPM GMI L1B ATBD 2016) have been chosen for conducting the sensitivity studies to be described hereafter. This radiometer is onboard the Low Earth Orbiting satellite GPM-Core Observatory, which is orbiting around the Earth at an altitude of 407 km on an inclined orbit of 65° with respect to the Equator. The GMI instrument is characterized by a comprehensive set of channels, summarized in Table 1, from 10.65 to 183.31 ± 7 GHz. The dataset of interest spans a 2 month period from September 1st to October 31st, 2017. From this dataset, only a geographical domain over the North Atlantic Ocean is selected, corresponding to the domain of the NWP limited area model described in the next section.

The channels have different footprint sizes that have been remapped on a common grid to simplify the inversion. Hence, the raw Level 1B data have been superobbed onto a regular latitude/longitude grid at 0.1° resolution, compatible with the effective resolution of the NWP model used in the present study.

2.2 Numerical Weather Prediction (NWP) model

At Météo-France, the convective-scale NWP model Application of Research to Operations at Mesoscale (AROME) is used operationally over numerous geographical domains with forecasts ranging up to 48 h (Seity et al. 2011). This model has a nonhydrostatic spectral dynamical core, explicitly resolving deep

Table 2. RTTOV-SCATT configurations selected for simulating GMI within the AROME model.

Hydrometeors	Particle size distribution	Single emission/absorption/scattering Radiative properties
Rain	Marshall-Palmer	Mie sphere
Snow	Field et al. (2007), Tropical regions	Discrete Dipole Approximations (DDA) 11 different shapes from Liu (2008)
Cloud liquid water	Modified Gamma	Mie sphere
Cloud ice water	Modified Gamma	Mie sphere

moist convection; and is characterized by a finite difference representation on the vertical with discretization on 90 levels from the surface to 10 hPa. In the midlatitudes, Météo-France operates AROME over Western Europe at 1.3 km horizontal resolution; this version is initialized with a 3D-Var data assimilation system (Brousseau et al. 2016) and forced on its lateral boundaries with forecasts from the Météo-France global model ARPEGE. In the Tropics, Météo-France operates AROME at 2.5 km horizontal resolution over five geographical domains covering several French Overseas territories. This tropical version, noted afterward AROME-OM, does not have a dedicated data assimilation system in operations yet. Instead, the AROME-OM versions are initialized with ECMWF analyses every 6 h. The full description of the AROME-OM system can be found in Faure et al. (2020).

In this paper, the AROME-Antilles version, one of the AROME-OM models has been chosen, it covers the geographical domain [10.4°N to 22.45°N; 67.8°W to 52.2°W]. Although the horizontal resolution of computation is on a 2.5 km grid, numerical models typically have an effective resolution 3–4 times larger than the original grid (Ricard et al. 2013). This leads to a typical effective resolution of 10 km, which roughly corresponds to the chosen 0.1° superobbing resolution for GMI observations. The selected GMI superobbed observations are collocated with a 3 h range AROME forecast, using a ± 90 min time tolerance, to build up a 3 h 3D-Var for AROME in a future study. Note that the results presented below are robust of a change of the first guess forecast range (not shown).

2.3 Observation operator

The observation operator used in the NWP model AROME to simulate brightness temperatures is the Radiative Transfer for TIROS Operational Vertical sounder code (RTTOV, Saunders et al. 2018). Within RTTOV, a module named RTTOV-SCATT, considering the scattering effect via hydrometeors on the basis of the Delta-Eddington approximation allows the simulations of observations from microwave sensors

within clouds (Bauer et al. 2006c; Geer and Baordo 2014). This study considers version 12 of RTTOV-SCATT to simulate GMI's Brightness Temperatures. Table 2 lists the radiative properties necessary to simulate scattering effects from hydrometeors.

Regarding the radiative properties of snowfall, Geer and Baordo (2014) showed that within the context of the ECMWF model, the best compromise across the different microwave frequencies affected by scattering was to choose the particle size distribution proposed by Field et al. (2007) together with the single-scattering properties of a sector snowflake. This configuration is selected to be the baseline for the present study. The other particle shapes from the scattering database of Liu (2008) are also considered in the sensitivity studies described below.

3. Method

3.1 Bayesian inversion

The first step of the “1D-Bay+4D-Var”, based on the ‘Bayes’ theorem, performs a one-dimensional retrieval of the atmospheric state, based on observations and prior information from a short-term forecast named First-Guess (FG). The retrieved state includes hydrometeors as well as humidity and temperature profiles.

Considering a set of observations y_o (GMI radiances), one wants to compute the probability to estimate the true state of the atmosphere: $P(x = x_{true} | y = y_o)$. The x_{true} vector is the “real” atmospheric profile. Thanks to the Bayes theorem, this probability can be rewritten as follows:

$$P(x = x_{true} | y = y_o) \propto P(y = y_o | x = x_{true})P(x = x_{true}).$$

In practice, it is assumed that observation errors and FG (prior) errors are not correlated and follow a Gaussian distribution, i.e.,

$$P(y = y_o | x = x_{true}) \propto e^{-\frac{1}{2}[y_o - H(x)]^T R^{-1}[y_o - H(x)]}$$

where H is the observation operator, which simulates brightness temperatures from an atmospheric profile,

i.e., $H(x) = y$ and \mathbf{R} is the covariance matrix of observation errors. Several assumptions are made to simplify the \mathbf{R} matrix: (i) a unique value is considered for the variances and noted σ^2 in the following, and (ii) the off-diagonal terms are set to 0, neglecting the interchannel correlations. This experimental \mathbf{R} matrix is the first approach to be further improved in the future to consider better the specifications of each frequency.

To estimate x_{true} , one must define a database of atmospheric profiles for which $P(x = x_{true} | y = y_o)$ values can be computed. The database used in the Bayesian inversion is made of FG profiles within the vicinity of a given observation. To limit error correlations between the retrieved profile and the first guess, which could be problematic for assimilation (Errico et al. 2007; Geer et al. 2008), the model profile at the observation location is removed from the inversion database. It is assumed that these profiles have high $P(x = x_{true})$ values because they are consistent in terms of weather regime with the meteorological situation of interest. Additionally, the occurrence and intensity of the cloud and precipitation profiles are assumed to be similar in the FG database and in nature. Therefore, the probability $P(x = x_{true})$ can be represented by the relative number of occurrences of profiles x_i in the database, i.e., by $1/n$, where n is the size of the selected database. Several mathematical estimators of x_{true} can be considered. In this study, we focus on two of them: either by selecting the x atmospheric state, which is associated with the highest value of $P(x = x_{true} | y = y_o)$. Such an estimator is used at JMA (Ikuta and Honda 2011) and was used as well at Météo-France with airborne cloud radar observations and provided good results (Borderies et al. 2018). This estimator is known as the maximum likelihood estimator and is noted hereafter as ML.

Another estimator is the expected value (mean) using the full database as a source of information (Caumont et al. 2010), leading to the following expression: $x_{ret} = \int x P(x = x_{true} | y = y_o) dx$, where x_{ret} is the retrieved atmospheric profile from which we then extract the relative humidity profile. In that case, the mean estimator is a weighted average (WA) of the FG profiles available in the database:

$$x_{ret} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}.$$

With

$$w_i = e^{\sum_{j=1}^d \{-\frac{1}{2}[y_{o,j} - H_j(x_i) - b_j]^T R^{-1} [y_{o,j} - H_j(x_i) - b_j]\}},$$

where i corresponds to a given profile in the inversion

database (therefore not including the first guess as explained above), j is a given channel within the selected ones, d is the number of channels selected for the inversion, and b_j is a clear-sky bias correction between the FG and the observations previously computed for each GMI frequencies over a 2 month period (not shown). The values can be found in Table 1. This estimator is noted hereafter WA for weighted average.

3.2 Definition of sensitivity studies

The mathematical expressions of the probabilities in the Bayesian inversion include several tunable parameters: (i) the specification of observation errors, (ii) the channel selection, (iii) the specification of radiative properties within the observation operator and (iv) the definition of the database. Additionally, for each parameter, the choice of the estimator (WA or ML) may lead to a different result as well.

The database for each inversion consists of 625 atmospheric columns within a 250×250 km² square surrounding each observation (25×25 profiles); the atmospheric columns are taken from 3 h forecasts of the AROME-Antilles model. Enlarging the database or changing the forecast range does not significantly modify the retrieval results (not shown); thus, the sensitivity experiments hereafter focus only on the first three parameters.

Table 3 summarizes the range of values for the three different degrees of freedom of interest. The experiments have been designed to change a single parameter at a time from a reference configuration and examine the impact of each of them. The set up of the reference configuration has been chosen to be close to the GMI noise equivalent delta temperature (Ne ΔT) and as well as to be consistent with previous studies for the observation error specification of the Bayesian inversion (Guerbette et al. 2016; Duruisseau et al. 2019). A single value of observation error was selected for all channels for simplicity, and to limit the numbers of sensitivity experiments. However, optimal observation errors would likely vary from one channel to another and could be derived with dedicated diagnostics (e.g., Desroziers et al. 2005). Desroziers et al. 2005 first diagnostics performed (not shown) indicate that observation errors explored for the sensitivity study range between 1 K and 20 K (see Table 3). The setup is also based on the willingness to use the full capability of the GMI sensor for the channel selection, and on the literature for the specification of the radiative transfer model (Geer and Baordo 2014).

Note that the 10 GHz channels have been excluded from the study because the 0.1° superobbing resolu-

Table 3. Parameter values for the experiments and the reference.

Degrees of freedom	Reference	Experiments
The observation error σ	1 K	2 K
		5 K
		10 K
		20 K
GMI channels	All channels from 18.7 to 183.31 \pm 7 GHz	Adding channels from 183.31 \pm 7 to 18.7 GHz
Radiative properties for snow particle	Sector Snowflake	Long Column
		Short Column
		Block Column
		Thick Plate
		Thin Plate
		Rosette 3-bullet
		Rosette 4-bullet
		Rosette 5-bullet
Dendrite Snowflake		

tion is not consistent with its coarser resolution (20 \times 30 km). Furthermore, radiative transfer modeling of this particular frequency is known to suffer from deficiencies (Lean et al. 2017).

3.3 Metrics used for evaluating the inversions

The Bayesian inversion described before is designed to assimilate microwave radiances in clouds and precipitating systems. To identify these regions, a scattering index based on the normalized difference of brightness temperatures (Bt) between the two polarizations of the 37 GHz frequency is chosen (Petty and Katsaros 1990; Petty 1994; Geer and Bauer 2010). Its expression is given by the following:

$$P_{37} = 1 - \frac{Bt^{37v} - Bt^{37h}}{Bt_{Clear}^{37v} - Bt_{Clear}^{37h}},$$

with Bt_{Clear} corresponding to the brightness temperatures simulated without considering the hydrometeors. This index can be computed both for the observations and the FG. In the case of observations, the denominator is taken the same as for the FG (Geer and Bauer 2010). A clear scene is assumed when P_{37} is below 0.3 (Geer and Bauer 2010) and a cloudy scene is assumed when it exceeds this threshold. In this study, two situations are considered: a situation named $FG_{clear}O_{cloud}$ where the profiles considered are clear in the FG but cloudy in the observations and the opposite situation named $FG_{cloud}O_{clear}$ containing the cloudy profiles in the FG but clear in the observations. These two situations are crucial in all-sky assimilation because they have the potential to correct for large prediction errors.

Table 4. Number of cases of the meteorological scenes comparing FG profiles and GMI observations over a 2 month period based on the P37 index and brightness temperatures simulated with the sector snowflake particle in the Liu database. FG_{clear} and FG_{cloud} denote respectively a clear profile and a cloudy profile in the FG. O_{clear} and O_{cloud} denote respectively a clear and a cloudy GMI observation.

	O_{clear}	O_{cloud}	Total
FG_{clear}	207308	8440	215748
FG_{cloud}	3218	2291	5509
Total	210526	10731	221257

They will also allow the examination of how the Bayesian inversion can moisten or dry the FG when the retrievals are assimilated. For the 2 month period, each category contains 3218 and 8440 observations, respectively, for the $FG_{cloud}O_{clear}$ and $FG_{clear}O_{cloud}$ (see Table 4). The number of meteorological scenes not treated in this paper is given in Table 4.

The results are compared in two different spaces: the model space and the observation space. Hence, two different metrics are defined: The first one compares the brightness temperatures simulated with the retrieved profiles against GMI observations. The differences in brightness temperatures are examined in terms of the correlation coefficient, bias, and standard deviation of their distributions. Note that when the same channels are used within the inversion and in the comparisons, the derived statistics should not be seen as an independent validation. This evaluation can

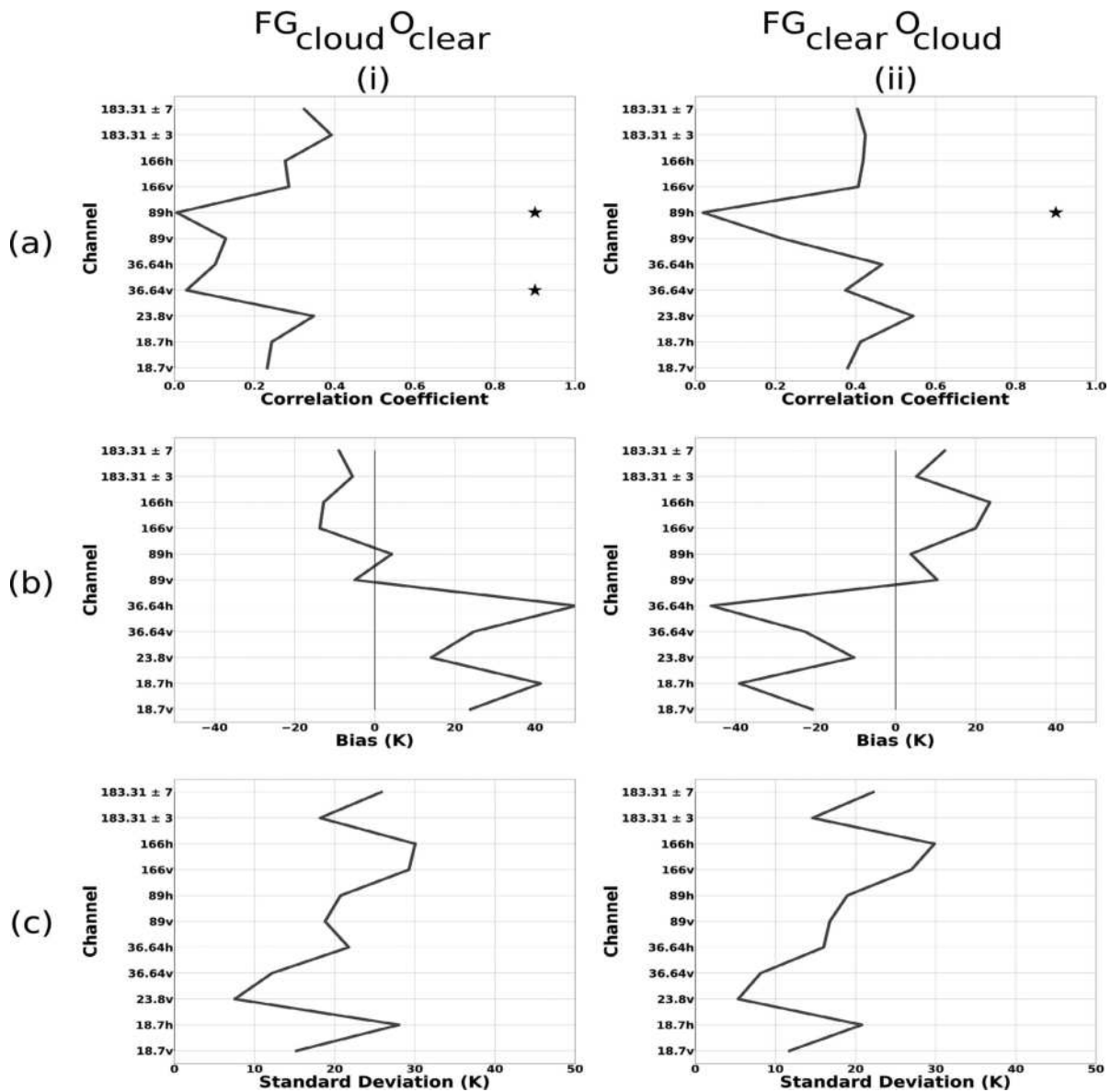


Fig. 1. (a) Correlation coefficients, (b) bias, and (c) standard deviation distributions of GMI brightness temperature differences between the observations and the FG, uncorrected of the clear-sky bias. Column (i) represents the results for experiment $FG_{cloud}O_{clear}$ and column (ii), those for $FG_{clear}O_{cloud}$. Stars indicate results that are not significant at 95 % level. Statistics computed over the 2 month period from September to October 2017.

be considered as a sanity check similar to the examination of analysis departures in a variational context that should be reduced with respect to background departures.

The second one compares the retrieved relative humidity profiles to the FG relative humidity profiles. The differences in relative humidities are examined in terms of the mean and standard deviation of their distributions.

4. Results from sensitivity studies

4.1 First-Guess statistics

To correctly interpret the results of the Bayesian inversion, the diagnostics between the simulated FG brightness temperatures and the GMI observations are calculated over the 2 month period for the $FG_{cloud}O_{clear}$ scenes [Fig. 1(i)] and the $FG_{clear}O_{cloud}$ scenes [Fig. 1(ii)]. Figure 1a shows the correlation coefficients between the simulated FG brightness temperatures and

the GMI observations. These correlations range from 0 to 0.4 for the $FG_{cloud}O_{clear}$ scenes. The 23.8 V and 183.31 ± 7 GHz channels have the highest values, whereas the 36.64 V and 89H GHz channels have the lowest coefficients. For $FG_{clear}O_{cloud}$ scenes, the coefficients never exceed 0.6. The 23.8 V GHz channel has the highest value, and the 89H GHz channel has the lowest one. This may be explained by the sensitivity of the 23.8 GHz and 89 GHz frequencies to water vapour besides cloud, precipitation, and the surface.

The bias of the distribution for the scenes $FG_{cloud}O_{clear}$ is positive for the low frequencies and negative for the high frequencies and ranges from -18 K to 50 K. For scenes $FG_{clear}O_{cloud}$, the bias is negative for the high frequencies and positive for the low frequencies and ranges from -45 K to 23 K. The sign of the bias changing between low and high frequencies is the result of the scattering effects, which becomes stronger than the absorption/emission effects as the frequencies increase and are therefore in agreement with the scenes studied. For both scenes, the biases of frequencies with horizontal polarization are degraded with respect to the vertical polarization. This problem is likely linked to the difficulty to simulate surface emissivity and the impossibility to simulate oriented particle emissivities for both polarizations within the RTTOV radiative transfer code.

The standard deviation for scenes $FG_{cloud}O_{clear}$ ranges from 7 K to 30 K and is generally higher for horizontally polarized channels. For scenes $FG_{clear}O_{cloud}$ the same behavior is observed with a standard deviation ranging between 5 K and 30 K.

The following section examines the results of sensitivity studies on observation errors, channel selection, and radiative properties used within the observation operator.

4.2 Sensitivity to observation error

a. Case study

The GMI instrument observed Hurricane Maria in the Atlantic Ocean on September 18th, 2017. Figure 2 shows the weights of several Bayesian inversions for a cloudy observation located in the core of this Hurricane and represented by the cross in Fig. 2f. Because of the errors in the prediction of its track and its structure by AROME, the model state at the same location corresponds to an atmosphere with much less clouds and precipitation. The weights of a neighborhood defined by 625 profiles in a 250×250 km² box have been calculated and represented by black boxes for different values of the observation error σ : 1, 2, 5, 10 and 20 K, respectively, in Figs. 2a–e. In this figure,

the black boxes are filled according to the magnitude of their normalized weight. It appears that a large weight is given to a single profile for the 1, 2 and 5 K observation errors (Fig. 2a to Fig. 2c), but when larger errors are considered, other profiles have a significant weight (Fig. 2d to Fig. 2e).

For this particular case, the differences between the two estimators WA and ML are illustrated by the cumulative probability distribution, plotted as a function of relative humidity (Fig. 3). The distributions are shown for a single level at 713 hPa (corresponding to AROME level 51). The impact of specifying a larger observation error can be noticed: the larger the observation error is, the more skewed the distribution is, toward the dryer values of relative humidity. The change in this distribution does not affect the ML estimator, which always leads to a saturated atmosphere. But this affects the WA estimator, which decreases toward less saturated amounts because of the long-tailed distribution of probabilities.

Figures 4 and 5 show the retrieved atmospheric profiles with the WA and ML estimators, respectively, for 1 K and 20 K errors. In both methods and for both errors, a low precipitating cloud layer is added to the first guess profile and the upper layer cloud is thickened with increased falling snow. Consequently, the brightness temperatures simulated with the retrieved profiles are much closer to those observed in both cases. Similar to relative humidity, the other atmospheric profiles are almost identical for the 1 K error, but strongly differ for the 20 K error. Particularly, the retrieved cloud fraction for 20 K slowly decreases from 1 to 0 from 100 hPa to 1000 hPa. This retrieved cloud fraction profile is not representative of AROME cloud fraction climatology, which simulates frequently sharp binary profiles.

b. General results

In this section, five experiments with the observation errors σ set to 1, 2, 5, 10 and 20 K are intercompared over the 2-month period. Statistics between retrieved and observed brightness temperatures for both $FG_{cloud}O_{clear}$ and $FG_{clear}O_{cloud}$ scenes are shown Fig. 6.

Regarding the correlation coefficients between retrieved and observed brightness temperatures, the $FG_{cloud}O_{clear}$ scenes lead to correlation values always higher than 0.7 [Fig. 6a(i)]. They are also higher for the smallest observation errors. In the case of 1 K and 2 K, the correlations are above 0.9 for all channels. For the $FG_{clear}O_{cloud}$ scenes, one can notice that the correlation coefficients are characterized by

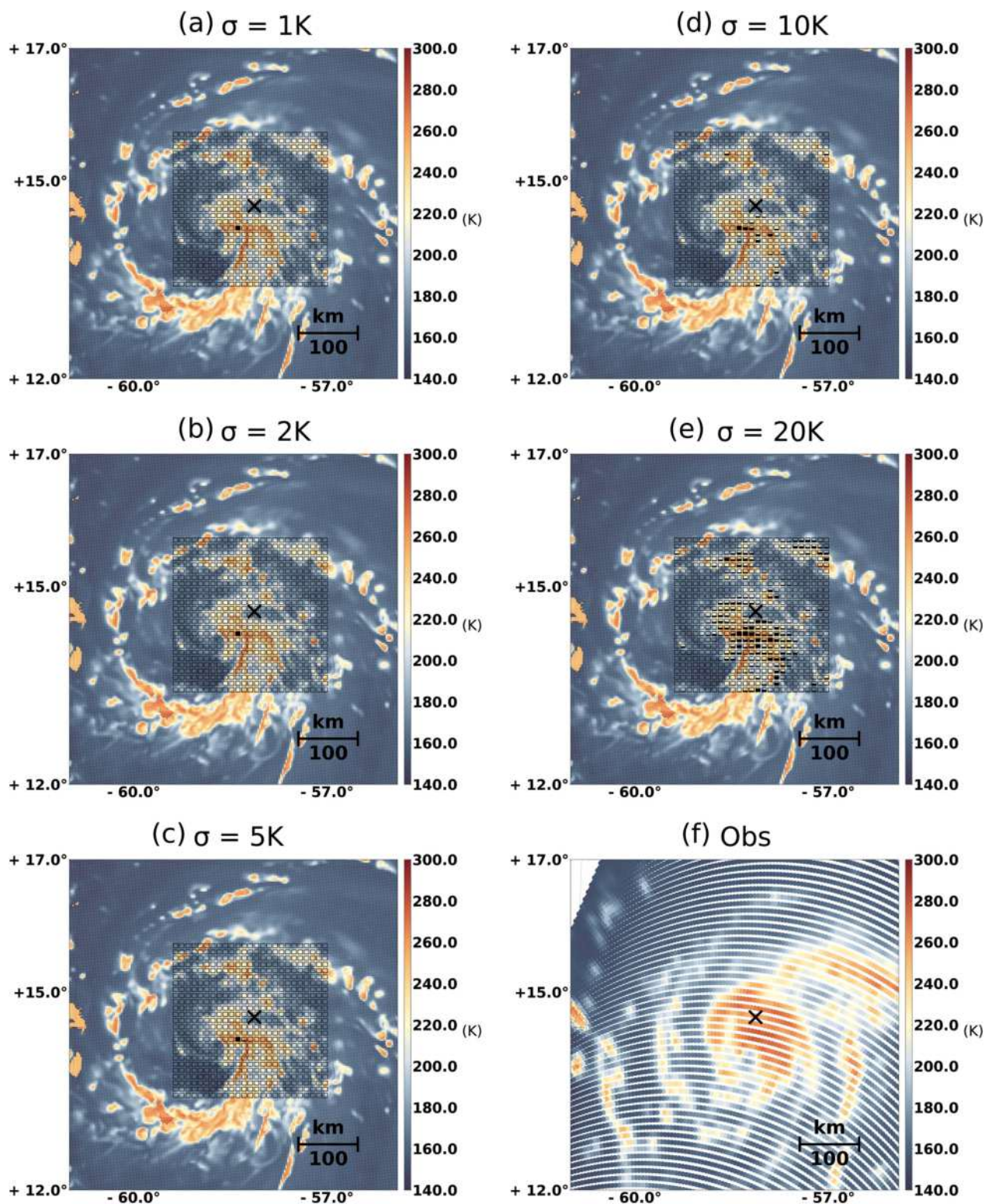


Fig. 2. (a), (b), (c), (d), and (e): FG brightness temperature simulations of the Maria hurricane reaching the West Indies on September 18th, 2017 for the 18 H GHz frequency. Inversion weights are represented as the filling fraction of the black boxes for an observation error that is equal to 1, 2, 5, 10 and 20 K, respectively. (f): GMI observations of the Hurricane with the 18.7 H GHz frequency. The selected observations are represented by the black cross. In this figure, the black boxes are filled according to the magnitude of their normalized weight.

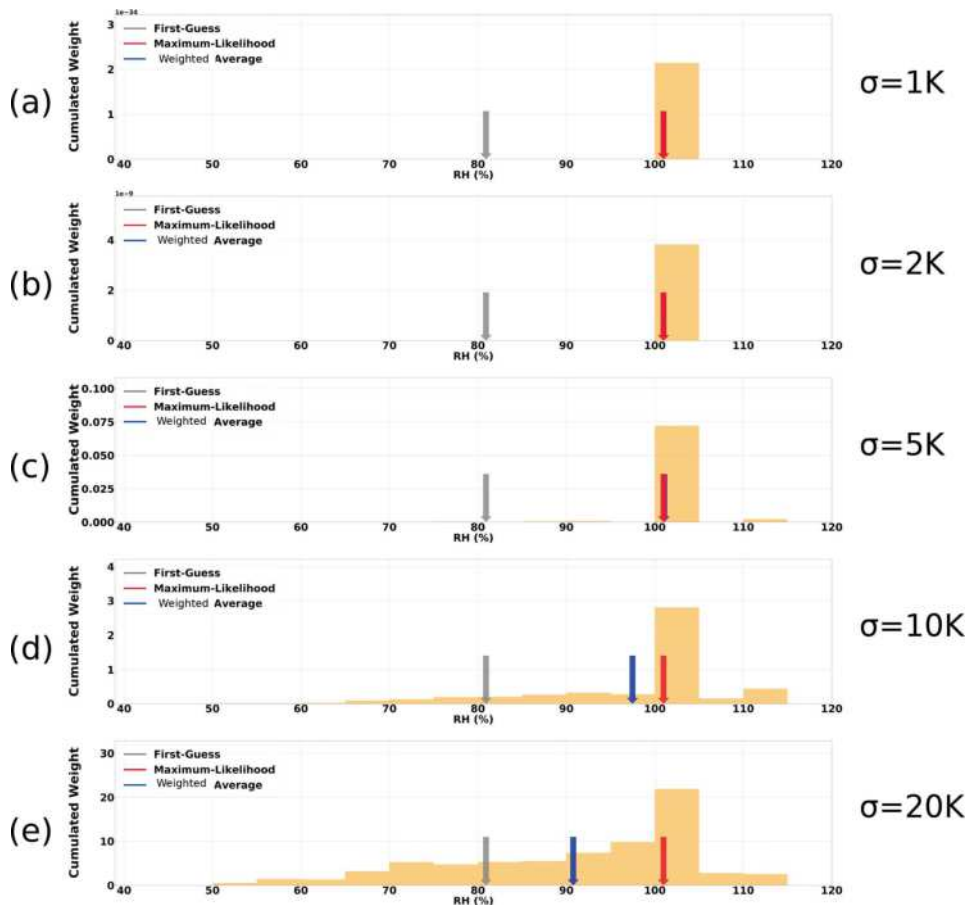


Fig. 3. (a), (b), (c), (d), and (e): Cumulated histograms of the weights from the Bayesian inversion for the case study by bins of relative humidity of 5 % at 713 hPa for an observation error of respectively 1, 2, 5, 10 and 20 K. The gray, blue and red arrows respectively indicate the values of the FG relative humidity and the WA and the ML retrieved relative humidity. (a–c) the blue and red arrows are identical.

larger variations across channels compared with the $FG_{cloud}O_{clear}$ scenes. This makes sense because in the $FG_{cloud}O_{clear}$ case, no hydrometeor content is observed; thus, in this case, the window channels are directly sensitive to the surface. In the case of $FG_{clear}O_{cloud}$, the expected retrieval is cloudy and precipitating and the correlations obtained with the different channels reflect the quality of the retrieval either for the liquid or the ice phase. Overall, the correlations are always smaller for the low frequencies than for the higher ones. This indicates that the retrievals are of lower quality for liquid precipitation than for solid precipitation. This may be a result of the physics of the model, which may have difficulty in propagating realistic information about the liquid hydrometeor content when model profiles contain solid hydrometeors. Regarding the variations of the correlation with the observation error specification, the conclusions are similar to the

$FG_{cloud}O_{clear}$ scenes: Larger values are found with smaller observation errors. Overall, the correlation coefficients are significantly improved compared with the correlation coefficients of the FG (Fig. 1).

Regarding the bias of retrieved brightness temperatures, the values are rather small for the $FG_{cloud}O_{clear}$ scenes [Fig. 6b(i)], typically below 2 K. This means that the retrieval technique is well constrained for those scenes and does not create large biases. By contrast, for the $FG_{clear}O_{cloud}$ biases, Fig. 6b(ii) shows that the bias can reach -40 K for $\sigma = 20$ K. For smaller observation errors, like 1 K or 2 K, the biases are limited to values ranging between 5 K and -10 K. If these results are compared with the initial innovations of Fig. 1, it can be seen that the biases are reduced regardless of the observation error used. The improvements are quite large for $FG_{cloud}O_{clear}$ scenes since there is almost no bias for each observation error. For

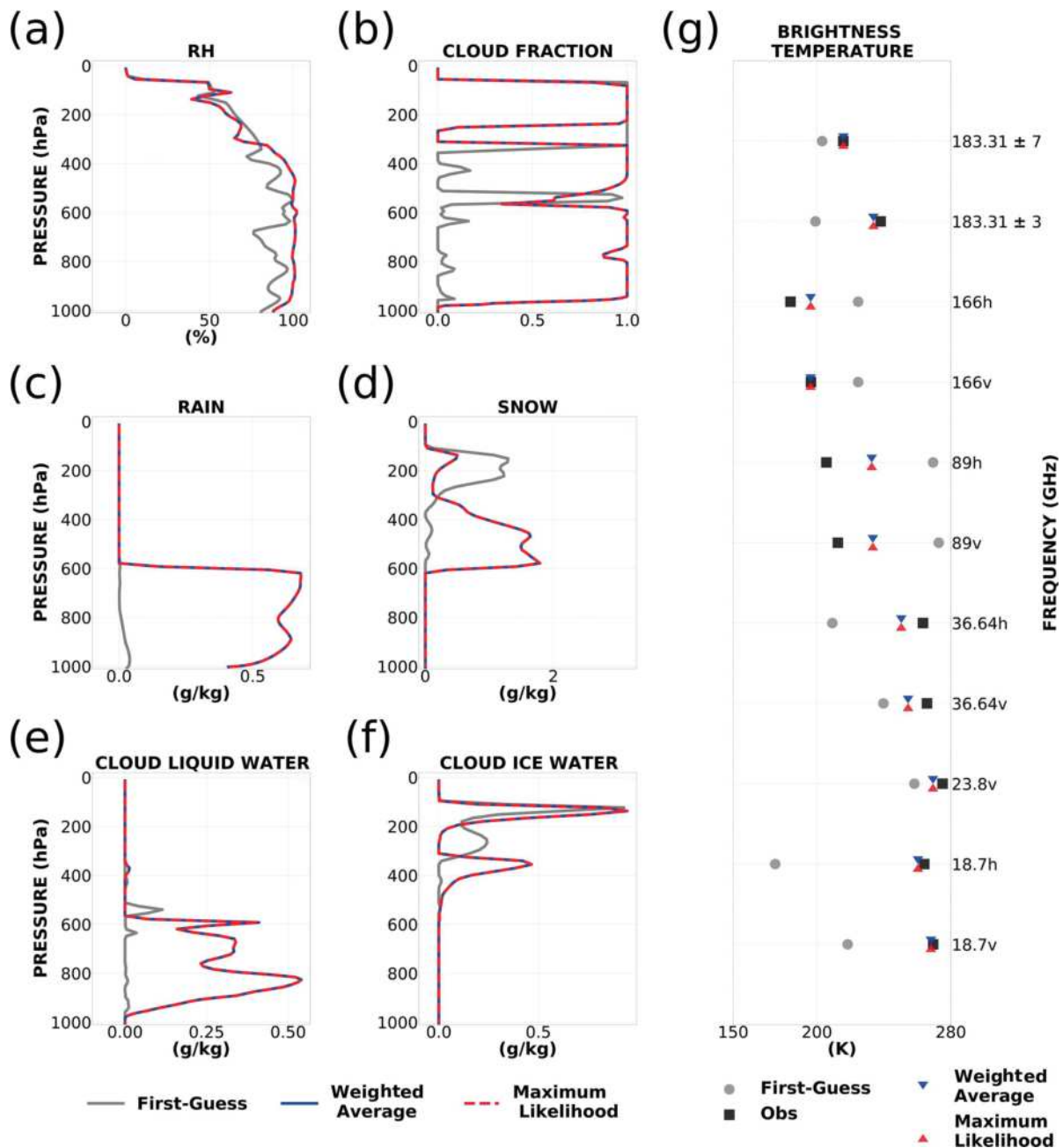


Fig. 4. (a), (b), (c), (d), (e), and (f): Case study atmospheric profiles of the relative humidity, cloud fraction, rain, snow, cloud liquid water and cloud ice water for the FG in gray; the WA retrieval in blue; and the ML retrieval in dotted red with an observation error that is equal to 1 K. (g): GMI brightness temperatures are represented by a black square for the observations, by a gray circle for the FG, by a blue triangle for the WA retrievals, and by a red triangle for the ML retrievals with the same observation error.

$FG_{clear} O_{cloud}$ scenes, the improvements are less important but still significant.

Regarding the standard deviation of the brightness temperature differences for $FG_{cloud} O_{clear}$ scenes shown in Fig. 6c(i), it appears that the standard deviation

increases with the observation error. The standard deviation never exceeds 5 K with an observation error of 1 K but can exceed 10 K for an observation error of 20 K. For the $FG_{clear} O_{cloud}$ scenes, the standard deviations have very similar values, except for the high

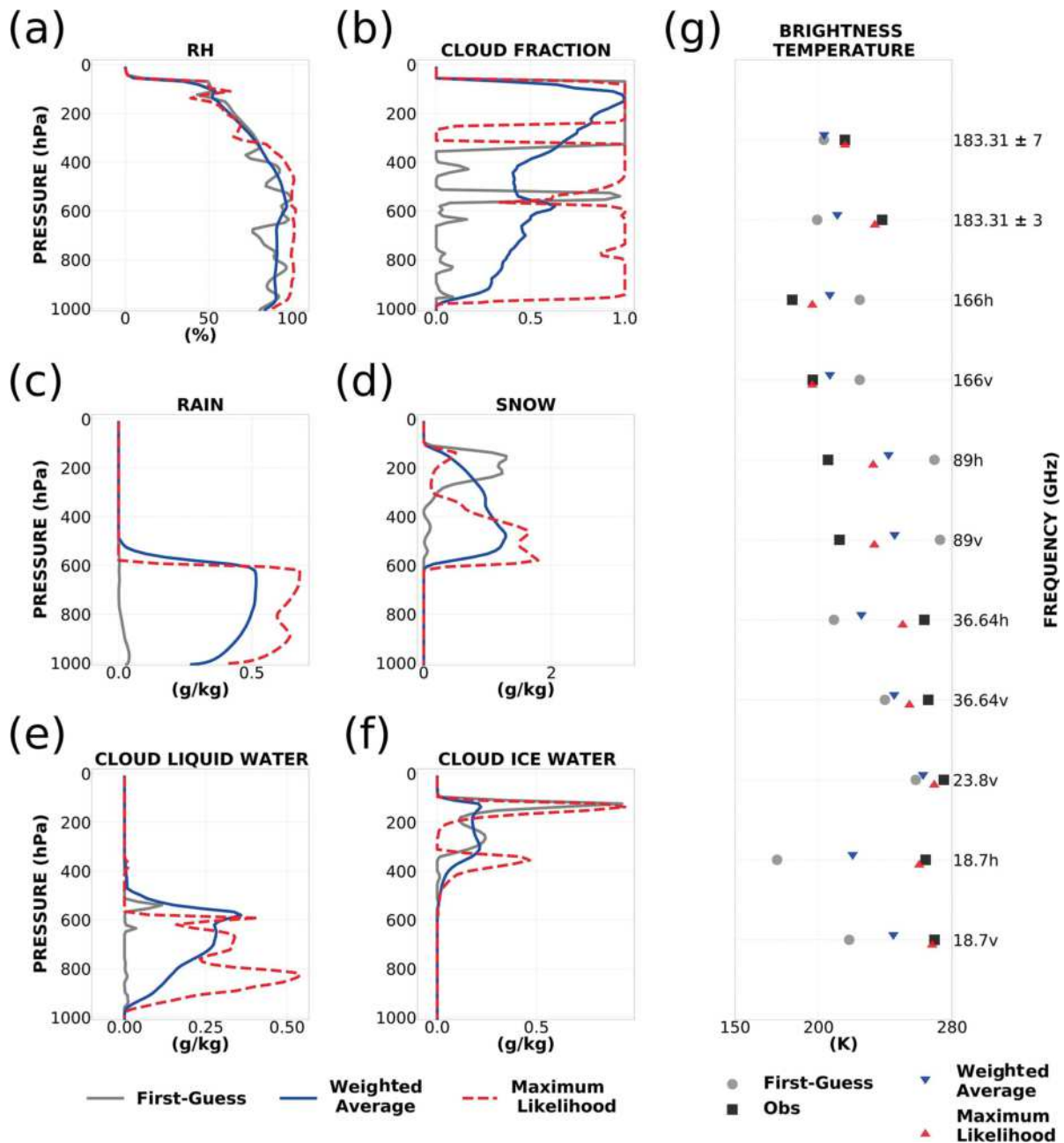


Fig. 5. Same as in Fig. 4, but with an observation error of 20 K.

frequencies and the 20 K observation error case.

The differences in results for observation errors can be further explained. Larger observation errors allow the retrievals to remain close to the FG values (i.e., far from the observed ones) and thus to produce a WA of profiles with either too low or too much hydrometeor contents. The vicinity of observations being mostly composed of clear atmospheric profiles (see Table 4), the profile retrieved from the WA will be further influ-

enced by clear profiles. The result will be to retrieve an atmospheric profile with a simulated brightness temperature that is clearer than the observed one. The example in Fig. 5 illustrates such behavior. By contrast, a low observation error will consider atmospheric profiles that are rather close to the observations and the retrieved profile will be less influenced by the difference in number between clear and cloudy scenes in its vicinity.

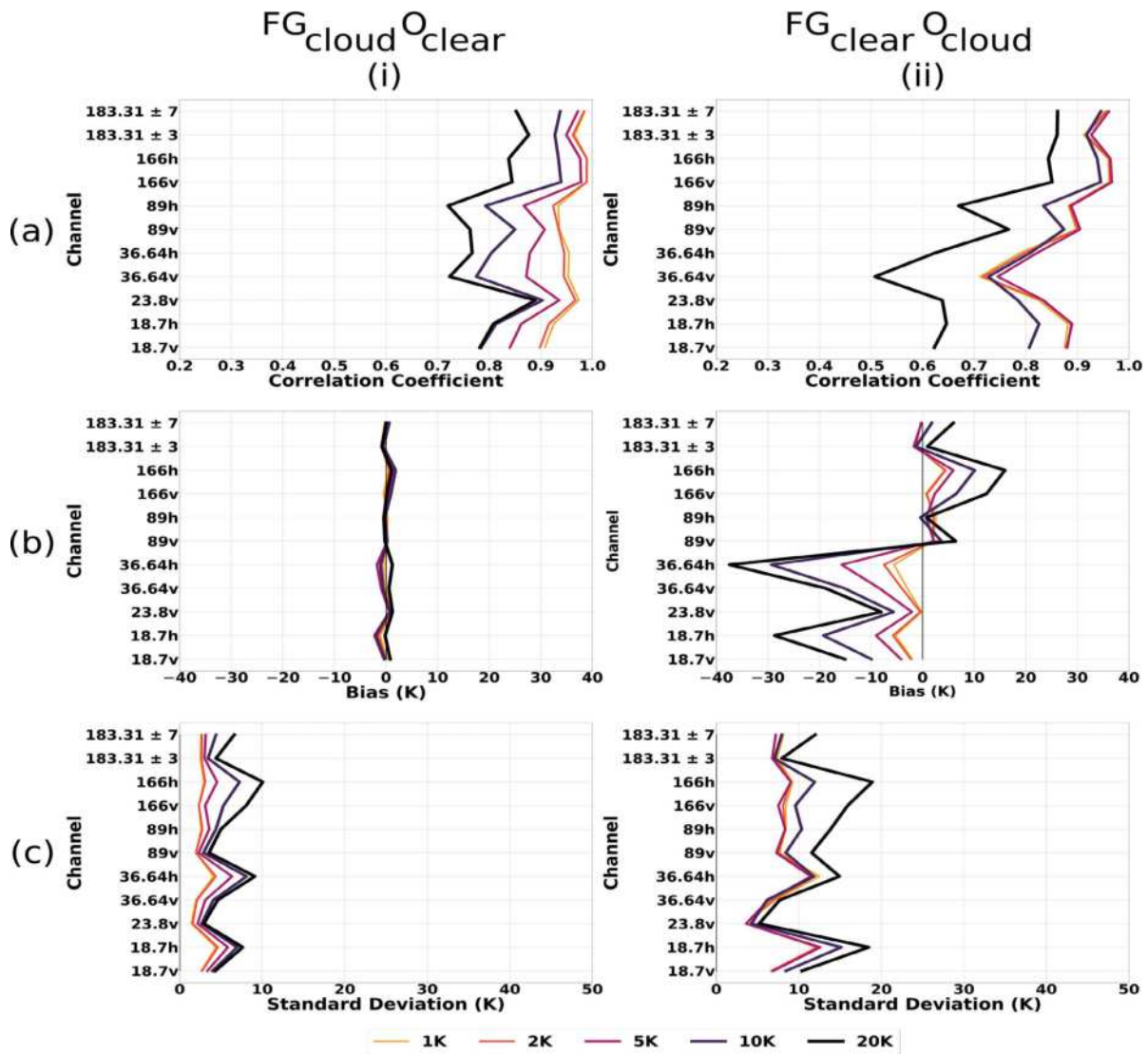


Fig. 6. (a): Correlation coefficient, (b) bias, and (c) standard deviation distributions of GMI brightness temperature differences between the observations and the WA experiments, uncorrected of the clear-sky bias, for observation errors that are equal to 1, 2, 5, 10, and 20 K. Column (i) represents the results for experiment $FG_{cloud} O_{clear}$ and column (ii), those for $FG_{clear} O_{cloud}$. Statistics are computed over the 2 month period from September to October 2017.

The retrieved profiles of relative humidity are compared with the FG profiles (analysis increments) in terms of bias and standard deviation statistics and displayed in Fig. 7. The mean values for $FG_{cloud} O_{clear}$ scenes, Fig. 7a(i), are negative for all observation errors, which means that the assimilation of retrieved profiles would tend to dry the model FG. The mean differences can reach values of up-to -15% between 600 hPa and 800 hPa for the largest observation errors. For the $FG_{clear} O_{cloud}$ scenes, the mean values are always positive which means that the assimilation of the retrieved profiles will tend to moisten the atmo-

sphere. The smaller the observation error is, the larger the moistening will be: The 1 K and 2 K curves can reach values of up to 12 % of mean relative humidity FG departure. By contrast, the mean values for the 20 K observation error are close to 0, which means that the assimilation of these retrievals will have very little effect on the analysis. Overall, these results indicate that a small observation error will enhance the FG moistening in case of cloudy observations.

When examining the standard deviations corresponding to the $FG_{cloud} O_{clear}$ scenes [Fig. 7b(i)], the curves show almost no dependency with observation

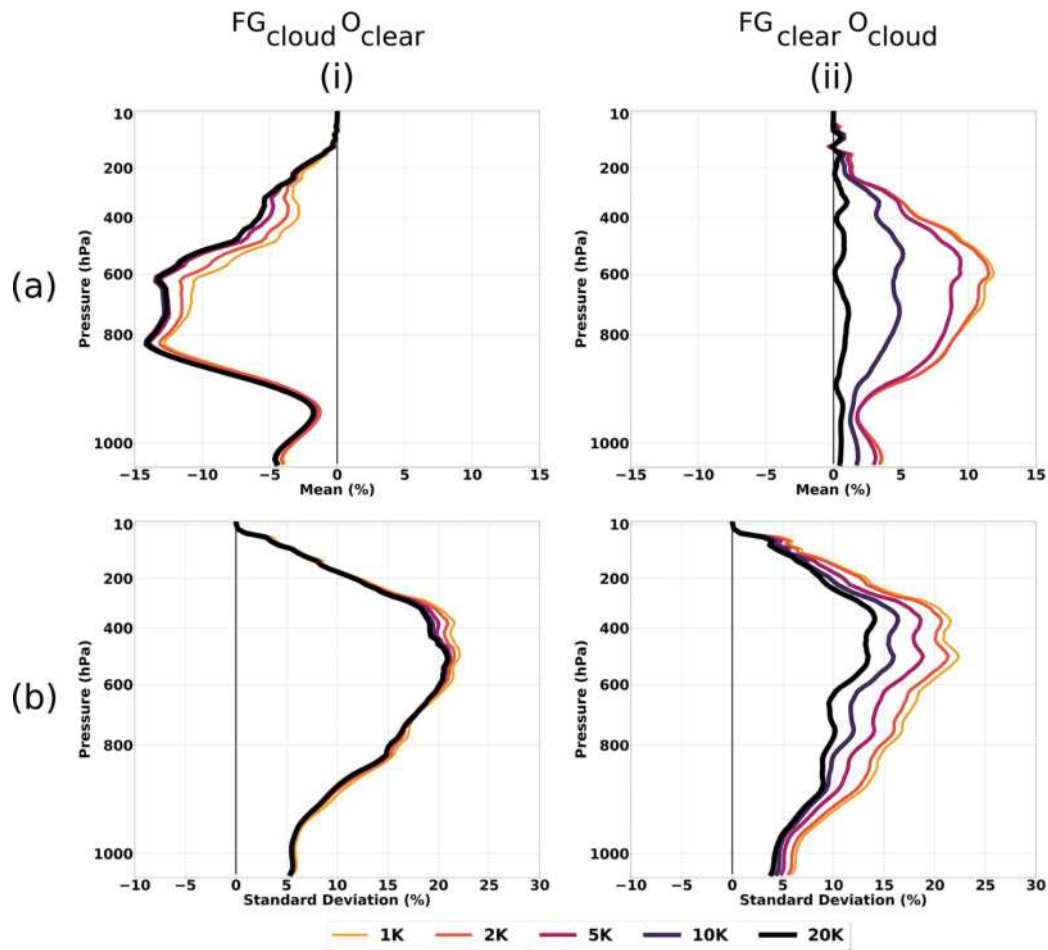


Fig. 7. (a) and (b): Mean and standard deviation distributions of the relative humidity differences between the WA experiments and the FG for observation errors that are equal to 1, 2, 5, 10 and 20 K. Column (i) represents the results for the experience $FG_{cloud} O_{clear}$ and column (ii), those for $FG_{clear} O_{cloud}$. Statistics are computed over the 2 month period from September to October 2017.

error and reach values of up to 22 % at 500 hPa. By contrast, for the $FG_{clear} O_{cloud}$ scenes, there is a clear decrease in the standard deviations when larger observation errors are prescribed. Indeed, with an observation error of 20 K, the maximum value is 15 %, whereas with a 1 K error, the standard deviation can reach the value of 23 %.

From the above results, several conclusions can be drawn: small observation errors lead to the best fits between Bayesian retrievals and observations. Compared with the FG relative humidity profiles, retrievals with a small observation error either dry ($FG_{cloud} O_{clear}$) or moisten ($FG_{clear} O_{cloud}$) the atmosphere with a similar magnitude in an average of $\pm 10\%$ to 15% between 500 hPa and 900 hPa.

By contrast, larger observation errors tend to smooth the retrievals, with an overall degraded fit to

observations.

Differences in statistics between the WA and the ML estimators are not shown because (i) the ML estimator results do not vary with the observation error and (ii) the ML estimator results are almost identical to the WA result with a 1 K error.

4.3 Sensitivity to the channel selection

a. Case study

In this section, sensitivity studies to the channel selection have been undertaken. When considering the same case study as in Section 4.1, the FG simulation (Fig. 8a) is very different from the observations (Fig. 8b). The hurricane structure of the FG appears more scattered with smaller scale structures than in the observations and the spiral bands are not well located. Three different Bayesian inversions with either one

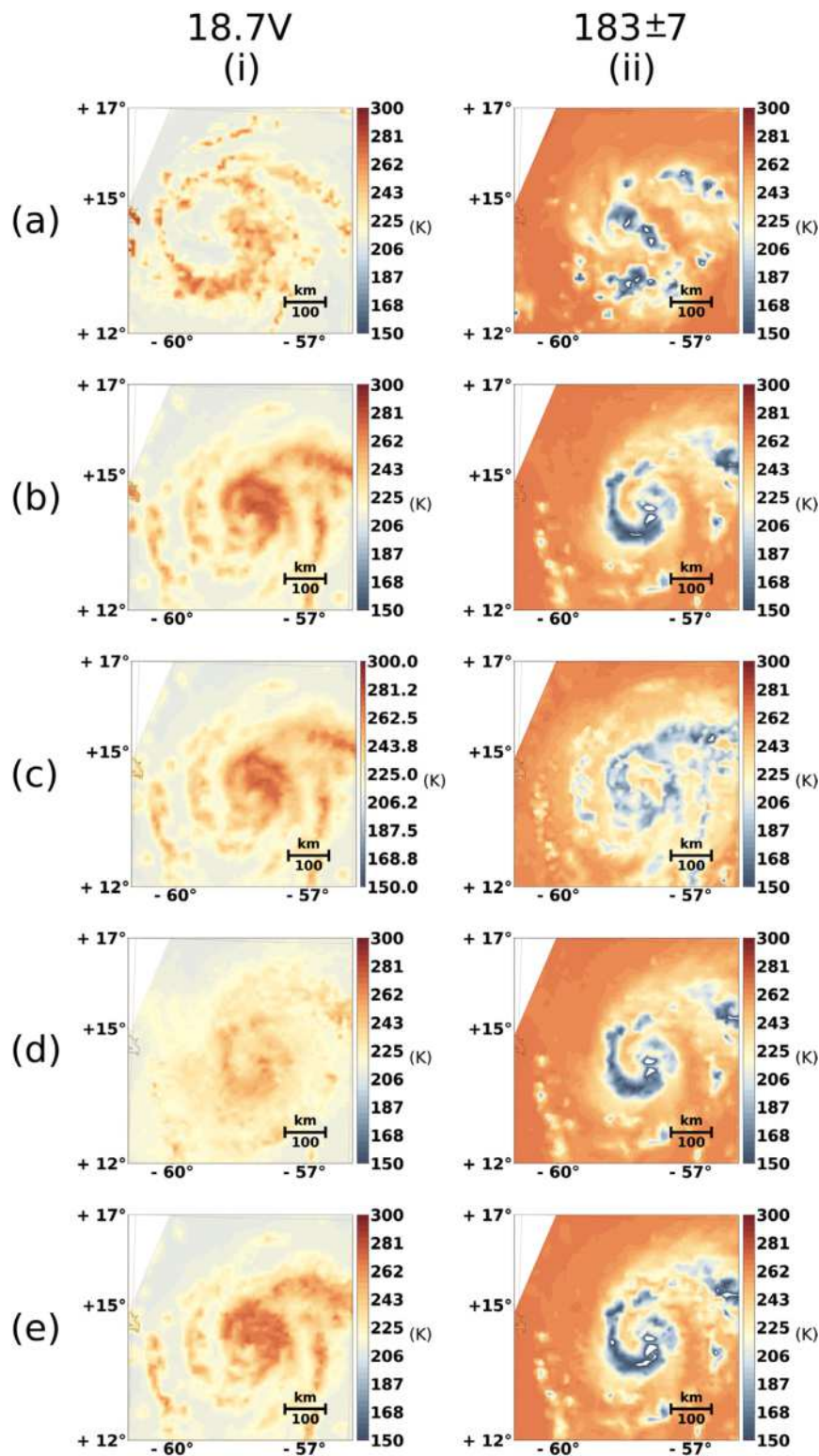


Fig. 8. GMI brightness temperatures of Hurricane Maria on September 18th, 2017 (a) FG simulations, (b) observations, (c) WA-retrieved simulations with the use of only the 18.7 V GHz frequency in the inversion, (d) WA-retrieved brightness temperature simulations with the use of only the 183.31 ± 7 GHz frequency in the inversion and (e): WA-retrieved simulations with the use of all GMI frequencies in the inversion, for channel 18.7 V GHz column (i) and for channel 183.31 ± 7 GHz column (ii).

GMI frequency (18.7 V or 183.31 ± 7 GHz) or all GMI frequencies are performed, the corresponding simulated brightness temperatures over Hurricane Maria are respectively shown in Figs. 8c–e.

In the first case (Fig. 8c), the Hurricane structure retrieved with the 18.7 V GHz GMI frequency is in good agreement with the observed one. Hence, the retrievals are well constrained for the low-frequency 18.7 V [Fig. 8c(i)], but it is not the case for the 183.31 ± 7 GHz frequency [Fig. 8c(ii)] exhibiting a rather different brightness temperature, structure. The spiral bands are characterized by too warm brightness temperatures whereas a strong underestimation is noticeable at the South of the core of the system.

Regarding the experiment using the 183.31 ± 7 GHz GMI frequency in the inversion, one finds the opposite results (Fig. 8d). The retrieved brightness temperatures have a structure in agreement with the observed ones at 183.31 ± 7 GHz. By contrast, when examining the retrieved structure at 18.7 V GHz, the strong observed contrast between warm brightness temperatures in the cloudy regions and colder ones in clear-sky regions is extremely blurred in the retrieval.

Finally, using all frequencies in the inversion leads to good results when compared with both frequencies (Fig. 8e).

From this case study, the retrievals appear to be well constrained in terms of brightness temperatures for the frequencies used within the Bayesian inversion. However, they lead to rather poor simulations of other frequencies. This means that this retrieval process can hardly infer information over the whole atmospheric profile. Nevertheless, it has been noticed that when using all channels in the inversion the retrieved profiles are better constrained. In the following, we examine if these conclusions are also valid over a larger sample.

b. General results

A set of 11 experiments, which consists of introducing progressively the high to low frequencies in the Bayesian inversion were performed over the 2 month period.

Statistics between retrieved and observed brightness temperatures for all experiments are shown in Fig. 9. The correlations for the $FG_{clear}O_{cloud}$ scenes [Fig. 9(ii)] are, like in the previous sections, worse than those for the $FG_{cloud}O_{clear}$ scenes. Furthermore, they are low for the channels not used in the inversion and high for the other ones. For example, when using only high frequencies, the correlation coefficients are highly degraded for low frequencies with values down

to 0.6 for the $FG_{cloud}O_{clear}$ scenes and down to 0.35 $FG_{clear}O_{cloud}$ scenes. When adding the low frequencies within the inversion the correlations increase progressively for both scenes. Only the 23.8 GHz frequency coefficients are relatively high for all experiments on Fig. 9a(i), but it was already the case for the initial innovations (Fig. 1a).

The biases of retrieved brightness temperatures are well constrained for the $FG_{cloud}O_{clear}$ scenes [Fig. 9b(i)]: below 4 K. For the opposite weather scenes [Fig. 9b(ii)], the biases can reach -35 K for low frequencies when only high frequencies are used in the inversion. Finally, by using all frequencies, all biases are between -5 K and 5 K.

The standard deviations of retrieved brightness temperatures (Fig. 9c) for the $FG_{cloud}O_{clear}$ scenes display low values with all frequencies and increased ones for low-frequency channels when only high frequencies are selected. For the $FG_{clear}O_{cloud}$ scenes, the results are similar with larger standard deviation values. The lowest value is 6 K, whereas for the $FG_{cloud}O_{clear}$ it is only 2 K. Furthermore, the standard deviation of the 18.7 H GHz frequency channel reaches 31 K when only high frequencies are used.

In summary, the channels can be split into two sets having contrasted behaviors: The frequencies below 37 GHz and those above. The best results for the high frequencies are obtained when using only high frequencies in the inversion. When adding low frequencies, correlations for these channels increase at the expense of slightly reduced correlations for high-frequency channels.

If one now considers the statistics of RH departures in terms of bias and standard deviations shown in Fig. 10, results for the $FG_{cloud}O_{clear}$ scenes show almost no dependency with the number of channels considered in the inversion. The mean values are negative (leading to a model drying) and the standard deviation values can reach up to 22 % between 300 hPa and 600 hPa.

For the $FG_{clear}O_{cloud}$ scenes the results are significantly different [Fig. 10(ii)]. The mean values are positive, so the retrieved profiles will tend to moisten the atmosphere. However, the maximum value of the mean gradually decreases with height as more channels are added to the inversion. Hence, with only the 183.31 ± 3 GHz and 183.31 ± 7 GHz frequencies, the maximum values are located at approximately 400hPa and are respectively 9 % and 6 %. However, the experiment using all channels reaches a maximum of 12 % at 600 hPa. Moreover, when adding channels in the inversion, the mean relative humidity FG departure becomes progressively higher at several pressure

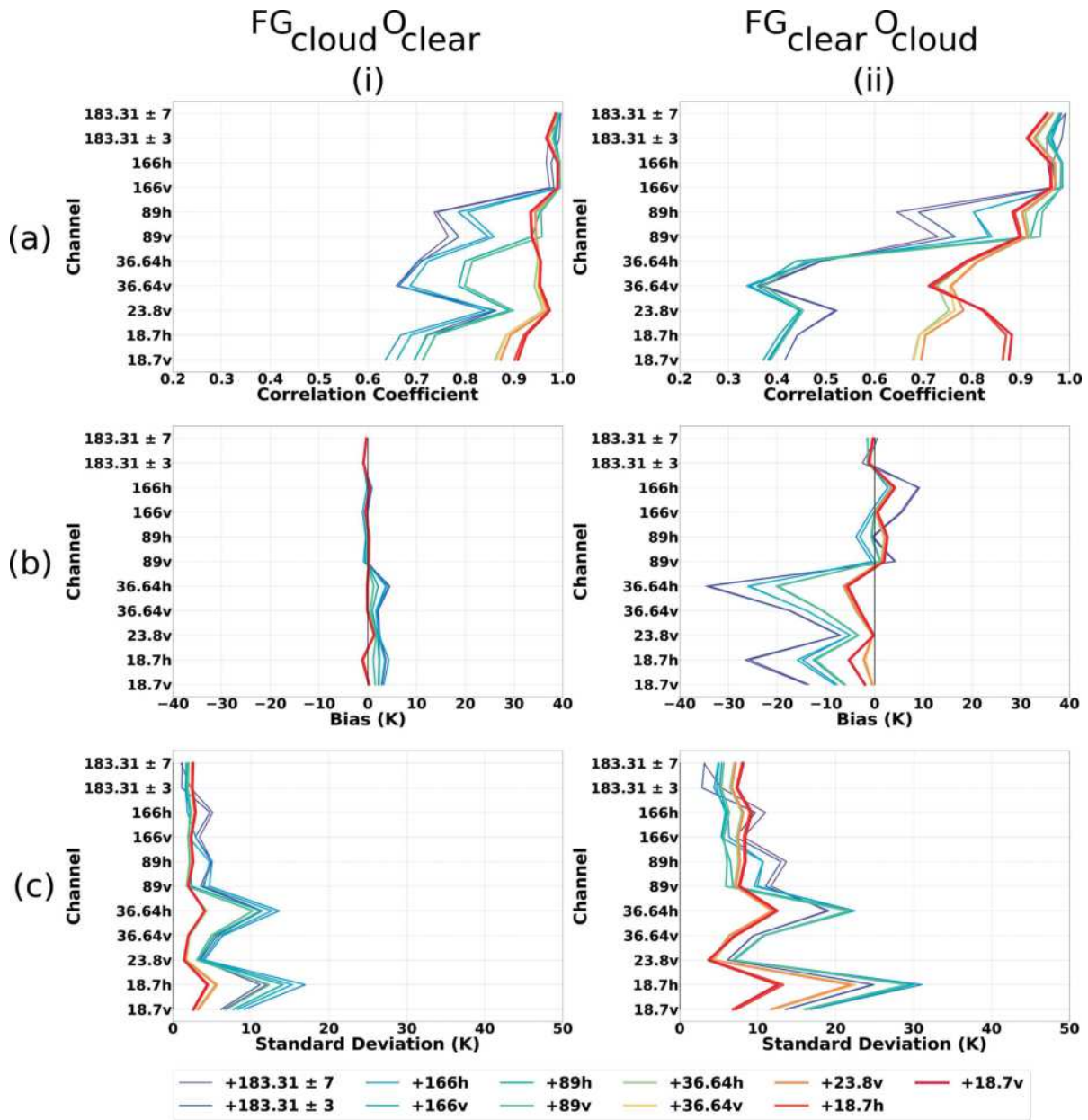


Fig. 9. (a): Correlation coefficient, (b) bias, and (c) standard deviation distributions of GMI brightness temperature differences between the WA experiments and the observations, uncorrected of the clear-sky bias, adding channels from 183.31 ± 7 GHz to 18.7 V GHz. Column (i) represents the results for experiment $FG_{cloud}O_{clear}$ and column (ii), those for $FG_{clear}O_{cloud}$. Each color corresponds to an experiment with adding channels, and the curves are also thicker with the increase in the number of channels. Statistics are computed over the 2 month period from September to October 2017.

levels. Thus, the addition of low-frequency channels in the inversion allows relative humidity to be more modified closer to the surface.

Figure 11 shows, with the ML estimator, the correlation coefficients between retrieved and observed brightness temperatures for the experiment set adding

channels from 183.31 ± 7 GHz to 18.7 GHz (Fig. 11). The variations of the correlation coefficients shown on Fig. 11 are similar to the variations with the WA estimator (Fig. 9a). However, it can be noticed that the coefficients for the low-frequency channel of experiments using only frequencies ranging from 183.31

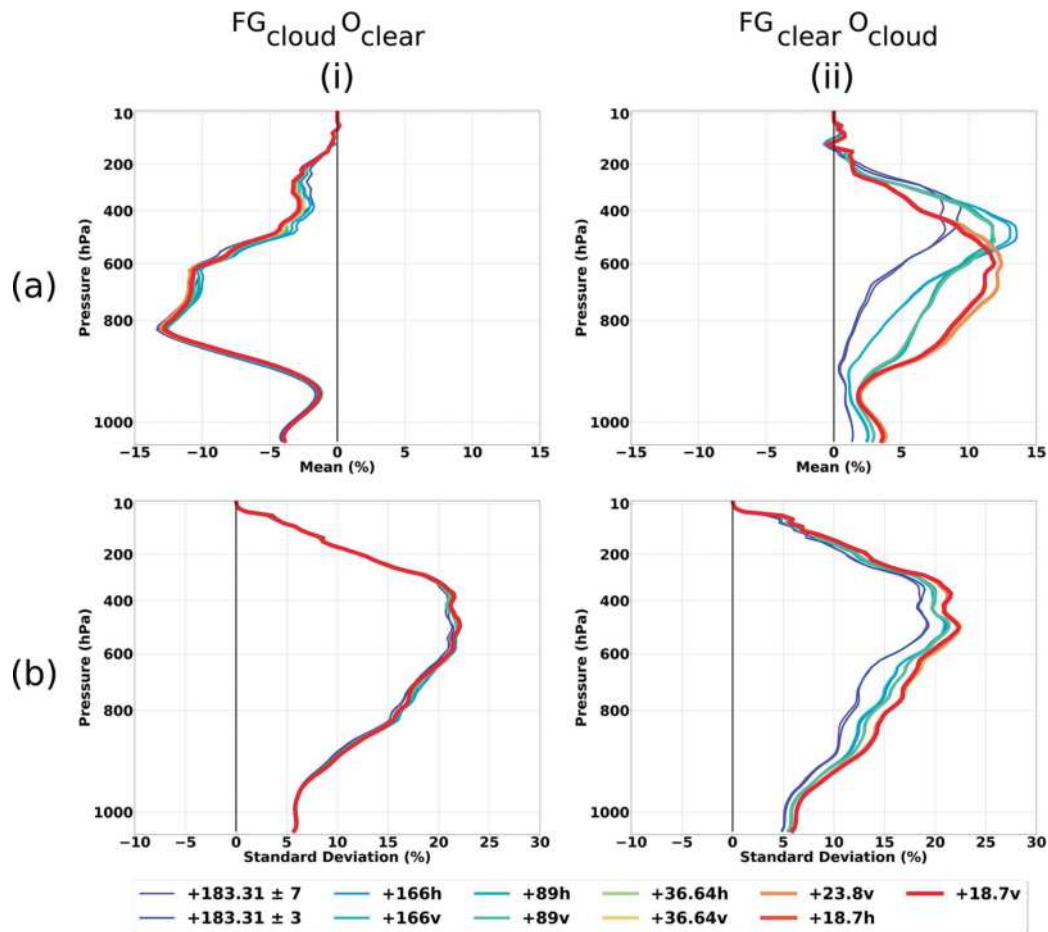


Fig. 10. (a) and (b): Mean and standard deviation distributions of the relative humidity differences between the WA experiments and the FG adding channels from 183.31 ± 7 GHz to 18.7 V GHz. Column (i) represents the results for experiment $FG_{cloud}O_{clear}$ and column (ii), those for $FG_{clear}O_{cloud}$. As in Fig. 9, curves are drawn thicker with the number of channels increasing. Statistics are computed over the 2 month period from September to October 2017.

± 7 GHz to 89 V GHz have smaller correlations with the ML estimator than with the WA estimator, for both categories of scenes. The other statistics (biases and standard deviations) are very close to the WA estimator (not shown). Thus, the Bayesian inversion using the ML estimator seems to be less well constrained with a limited number of channels used than using the WA estimator.

Several conclusions can be drawn from the above results. A large number of channels within the inversion allows to better constrain the retrievals on the vertical unlike the experiments with a reduced set [as done by Duruisseau et al. (2019) only with 183 GHz channels]. Indeed, the various channel help to provide information at pressure levels where they are sensitive to either solid or liquid hydrometeors. Furthermore, the FG will moisten more homogeneously on the

vertical with a full set of channels for the $FG_{clear}O_{cloud}$ scenes. For the other scenes, the FG will be dried out in the same way: The fact that no hydrometeors are present in the observed scenes leads to rather similar retrieved atmospheric profiles. Finally, the ML estimator has degraded performances with a limited set of channels.

4.4 Sensitivity to scattering properties of hydrometeors

Over the last decade, the radiative properties of frozen hydrometeors are a subject of research in the microwave assimilation community. Indeed, the scattering properties in the microwave depend strongly upon particle shape, density, and size, which are extremely variable for solid hydrometeors within clouds. For the time being, the Météo-France data assimila-

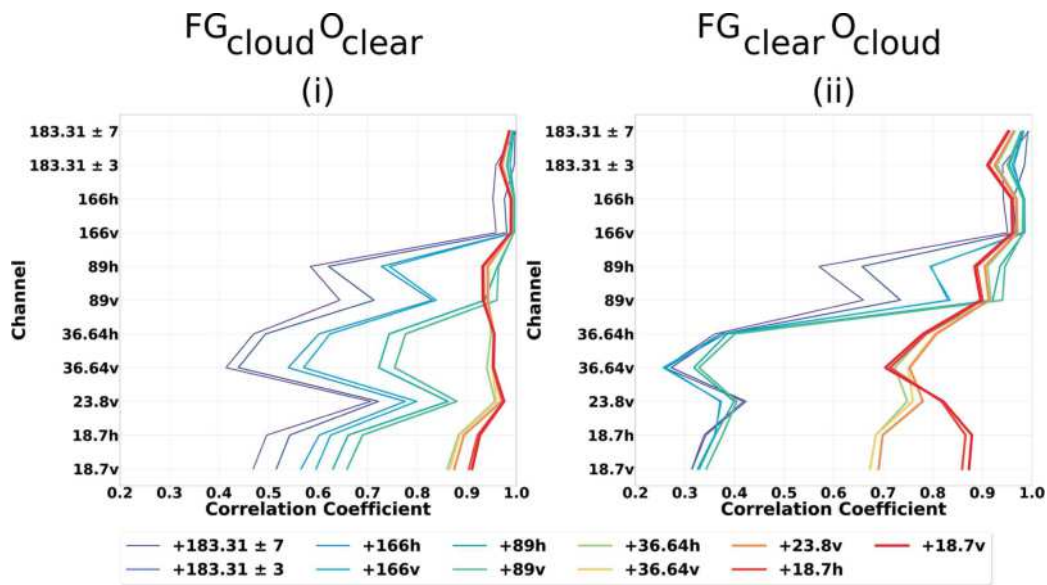


Fig. 11. Correlation coefficient between the observations (GMI brightness temperatures) and the ML experiments adding channels from 183.31 ± 7 GHz to 18.7 V GHz. Column (i) represents the results for experiment $FG_{cloud} O_{clear}$ and column (ii) those for $FG_{clear} O_{cloud}$. As in Fig. 9, curves are drawn thicker with the number of channels increasing. Statistics are computed over the 2 month period from September to October 2017.

tion systems used with the RTTOV-SCATT model can only handle one particle shape and one particle size distribution per hydrometeor type, for all weather situations. Knowing the actual solid hydrometeor diversity, this represents a huge simplification of the atmosphere behavior (e.g., Schmitt et al. 2016). Hence, several studies have been undertaken, and nowadays, a number of databases provide radiative properties for a wide range of snow particles with various shapes and densities (e.g., Eriksson et al. 2018; Kneifel et al. 2020). Additionally, the retrieval community is studying several methods to obtain information on hydrometeor diversity in the atmosphere from microwave observations and the assimilation community is just beginning to consider this diversity in numerical models (e.g., Haddad et al. 2015). In this section, we examine the impact of the selection of radiative properties for hydrometeors on the Bayesian inversion results. The Liu database (Liu 2008) has been chosen because of its availability in the RTTOV-SCATT model and wide use in the community.

The FG and observed brightness temperature distributions over the 2 month period are displayed in Fig. 12 for each particle of the Liu database. In this figure, one can note that the longest tails of distributions toward cold brightness temperatures are associated with particle shapes having the strongest scattering efficiency. For frequencies 18.7 V GHz and 23.8 V

GHz (Figs. 12a, b), the distributions have a mode at approximately 210 K and 250 K, respectively, corresponding to clear-sky scenes and are skewed toward warmer temperatures of up to 280 K, corresponding to the microwave emission of liquid hydrometeors. The occurrences of simulated brightness temperatures between 210 K and 255 K for the 18.7 V frequency and between 255 K and 280 K for the 23.8 V frequency are smaller than the observed occurrences. These channels being sensitive to rain, thus seem to indicate a lack of liquid precipitation in the AROME model forecasts, which was already highlighted in Faure et al. (2020). Additionally, the distributions for each particle remain the same, which is consistent with the fact that low frequencies are not sensitive to snow particles.

For the 36.64 GHz frequency (Fig. 12c), the distributions have a mode at approximately 220 K with a significant positive skewness as noticed for the lower frequencies. Conversely, for several particles (the most efficient scatterers) the distributions are skewed toward colder temperatures down to 170 K. This behavior is the signature of scattering by solid particles which seems to overcome the emission signature in these cases.

For higher frequencies (89, 166, and 183.31 GHz) the distributions shown in Figs. 12d, e, and f have a mode at warm temperatures (clear-sky scenes) and

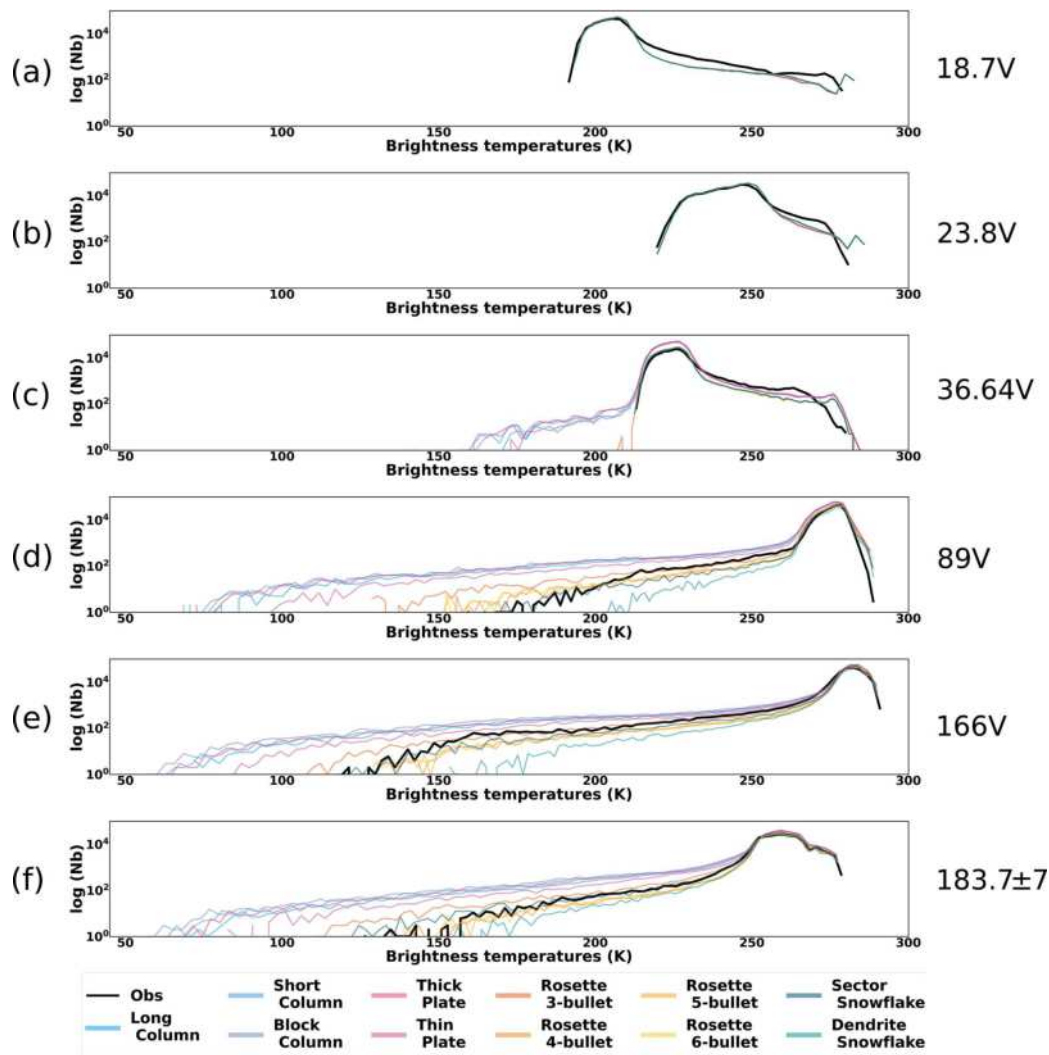


Fig. 12. (a–f): Distributions of GMI FG brightness temperatures for each particle at frequency 18.7 V GHz (a), 23.8 V GHz (b), 36.64 V GHz (c), 89 V GHz (d), 166 V GHz (e), and 183.31 ± 7 GHz (f). Statistics computed over the 2 month period from September to October 2017.

a negative skewness associated with the scattering processes by frozen particles. Additionally, the observation distributions displayed in Figs. 12d, e, and f appear to reasonably match the sector snowflake distributions as well as the bullet rosettes distributions. These results are consistent with the use of the sector snowflake particle as the reference in RTTOV-SCATT as suggested by Geer and Baordo (2014).

Furthermore, depending upon the choice of particle, the coldest simulated temperatures vary between 70 K and 150 K. We then decide to rank particles according to their scattering efficiency to aid the interpretation of the inversion results presented afterward.

The Bayesian inversion has been run over the 2

month period with the 11 particle shapes. It was found that the retrieved brightness temperatures are characterized by highly similar statistics (not shown). The retrieval algorithm always finds a combination of FG profiles that does match the observations in the brightness temperature space. Nevertheless, the underlying retrieved atmospheric profiles are rather different. Figure 13 displays the mean distributions of retrieved hydrometeor profiles for the $FG_{clear}O_{cloud}$ scenes. The less efficient scattering particles tend to retrieve atmospheric profiles with the largest rain, snow and cloud ice water contents, as shown respectively in Figs. 13a, 13b, and 13d.

For example, the maximum of snow content means

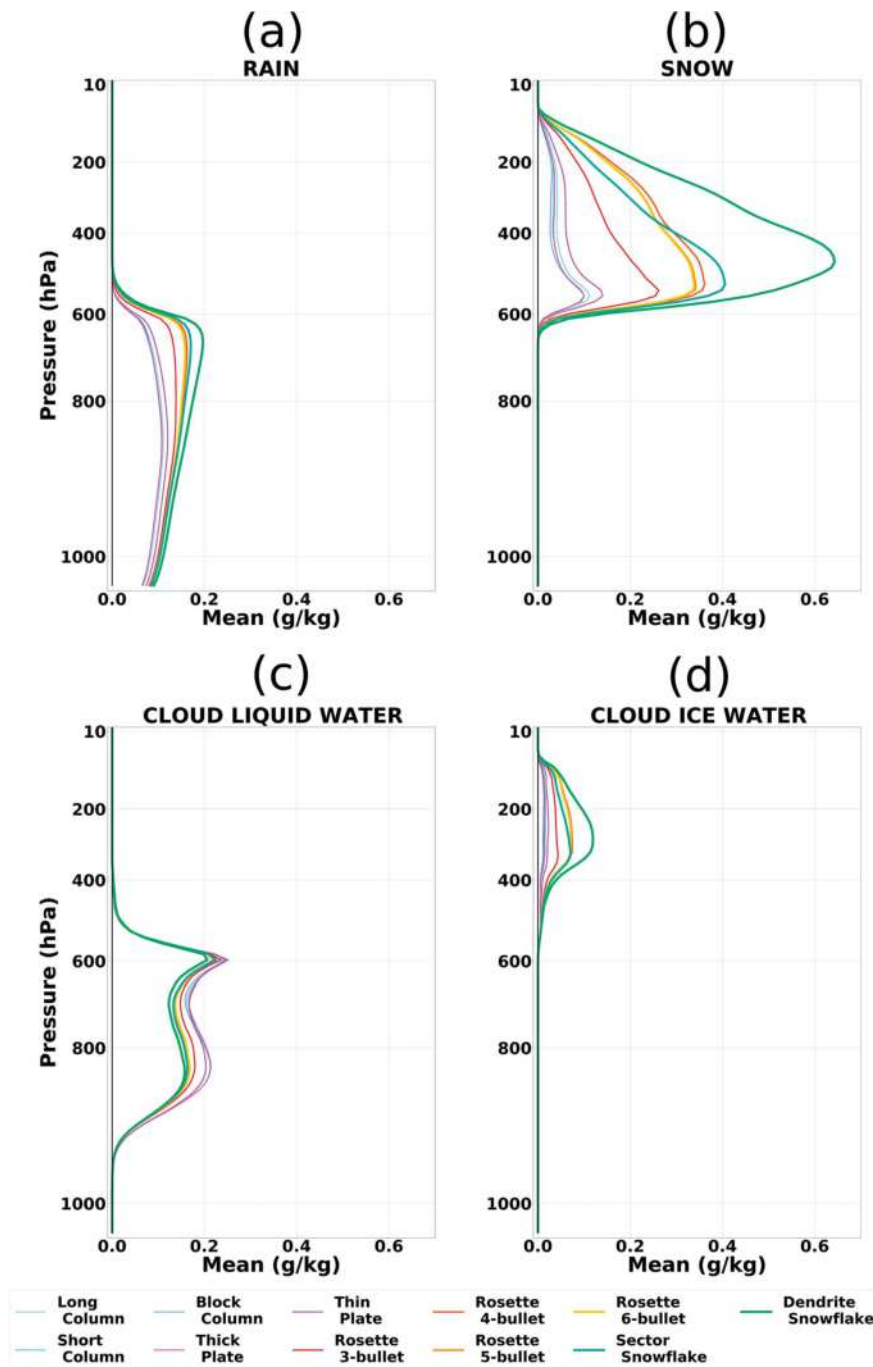


Fig. 13. Mean retrieved profiles of (a) rain, (b) snow, (c) cloud liquid water and (d) cloud ice water for the various particles in the Liu table (2008) for $FG_{clear}O_{cloud}$ scenes. Statistics are computed over the 2 month period from September to October 2017.

is 0.1 g kg^{-1} for one of the most efficient scattering particle shape, such as a block column, and 0.65 g kg^{-1} for one of the least efficient scattering particle shape, such as that of the dendrite snowflake. For the mean retrieved cloud liquid water profiles shown in Fig.

13c, the behavior is the opposite. Indeed, the retrievals with the more efficient scattering particles are those containing the highest cloud liquid water content: up to 0.25 g kg^{-1} for the block column versus 0.2 g kg^{-1} for dendrite snowflake particles. This suggests that

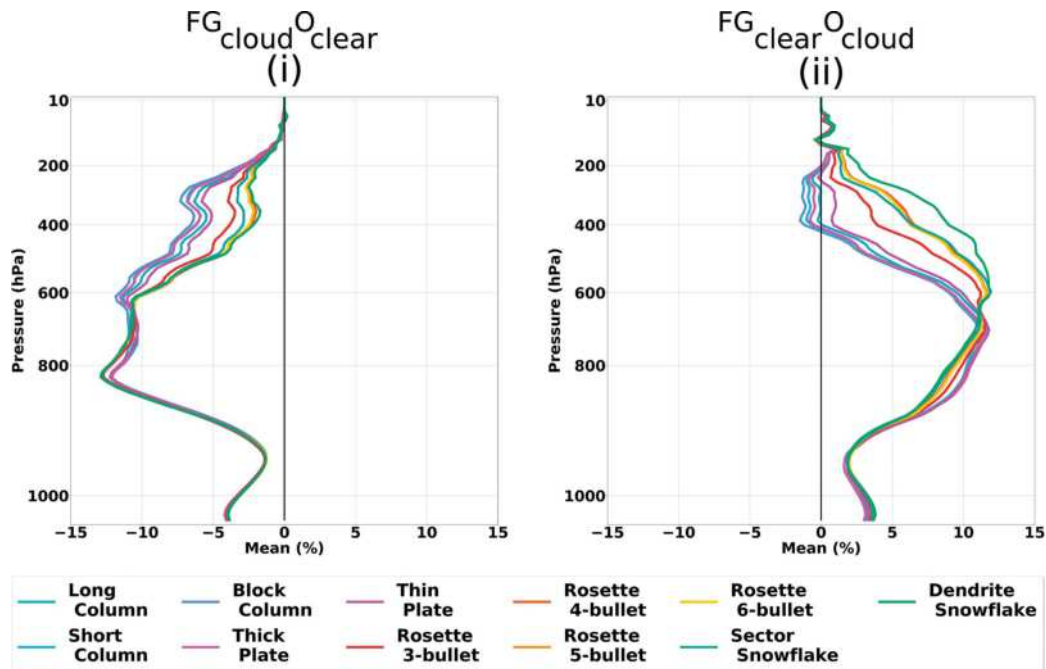


Fig. 14. Mean distributions of relative humidity differences between the FG and the WA experiments for particles from the Liu database (2008). Column (i) represents the results for the experiment $FG_{cloud}O_{clear}$ and column (ii), those for $FG_{clear}O_{cloud}$. Statistics are computed over the 2 month period from September to October 2017.

the retrieval algorithm seeks to generate profiles with more rain, snow, and cloud ice water but less cloud water for the less efficient scattering particles to match a cloudy observation. This might come from a lack of constrain on these microphysics species or a compensation effect, which will need to be further investigated.

Figure 14 shows the statistics of relative humidity FG departures in terms of means and standard deviations. The mean profiles for $FG_{cloud}O_{clear}$ scenes exhibit negative values leading to the drying of the atmosphere with variations in magnitude between 200 hPa and 600 hPa. The most effective scattering particles lead to the largest atmospheric drying: for example, at 300 hPa, the mean values for the block column and the dendrite snowflake particles are respectively of -7% and -3% . For this category of scenes, one could have expected the particle shapes to have no effect on the retrieved profiles since clear-sky profiles do not or contain very little snow amounts. The fact that this is not the case indicates that the filtering of meteorological scenes, on the basis of the 36.64 GHz frequencies, does not fully discard all cloudy scenes, especially the nonprecipitating ice clouds. Moreover, the different meteorological weather scenes have been defined thanks to the brightness temperatures simulat-

ed with the sector snowflake particle. Another particle would have led to a different categorization, but the filter based on the sector snowflake particle has been kept in all the experiments for the rest of the study to compare identical samples.

For the $FG_{clear}O_{cloud}$ scenes, the mean profiles are distributed toward positive values leading to an atmospheric moistening. Additionally, two behaviors can be noticed:

- Between approximately 200 hPa and 600 hPa, the less effective scattering particles, like the dendrite snowflake particles, tend to further moisten the atmosphere than do other particles. This behavior is consistent with the results shown in Fig. 13: the less effective scattering particles will retrieve a larger amount of hydrometeors to reach the observed brightness temperatures.
- Between approximately 600 hPa and 900 hPa, the opposite behavior can be seen. The retrieval algorithm seeks to create a profile with further relative humidity in the lower layer for the more effective scattering particles. This is a consistent behavior with the mean distributions of retrieved cloud liquid water profiles displayed in Fig. 13.

In conclusion, the choice of frozen particles within the observation operator has an essential impact on

retrieved relative humidity profiles. A most efficient scatterer will tend to reduce the profile moistening in the $FG_{clear}O_{cloud}$ scenes. Opposite behavior has been found over the mean distributions of the cloud liquid water profiles and over the mean distributions of relative humidity profiles in the lower layers. This will need further investigations for a deeper understanding of this effect. Finally, because a 1 K observation error was selected in the reference configuration, the results of statistics with the ML estimator (not shown) are similar to those with the WA estimator.

5. Summary and conclusions

Several sensitivity studies have been performed on various specifications of the “1D-Bay+3D/4D-Var” method over a 2 month period in 2017 with the convective scale model AROME and GMI brightness temperatures from 18.7 GHz to 183.31 GHz. The following aspects have been examined: (i) the observation errors, (ii) the channel selection, and (iii) the scattering radiative properties of frozen hydrometeors in the observation operator. On top of assessing the impact of these specifications, two estimators have been compared: the mean of likelihood distribution (the so-called weighted average: WA) and the maximum value of this distribution (the so-called maximum likelihood: ML).

Observation errors ranging from 1 K to 20 K have been tested for the inversion with the WA estimator. Large errors tend to smooth and dry the retrieved profiles, as well as producing less realistic hydrometeor profiles. By contrast, small observation errors tend to retrieve profiles more consistent with observed brightness temperatures. Moreover, with small errors, the retrieved profiles are similar with both estimators, the ML one being invariant to this parameter when using a covariance error matrix with a single value along its diagonal. In a future study, dedicated diagnostics could be applied to estimate optimal parameters for the assimilation (e.g., Desroziers et al. 2005).

Eleven experiments have been conducted regarding the channel selection. Compared with that in previous studies (Guerbette et al. 2016; Duruisseau et al. 2018), it was shown that the Bayesian inversion does work with a larger set of channels. Adding low frequency channels to the inversion brings information on the low atmospheric layers of the atmosphere and leads to good-quality retrievals as shown in the observation space. Moreover, the results highlight the usefulness of two groups of frequencies, the lower ones (from 18.7 V to 36.64 H GHz) and the higher ones (from 89 V to 183.31 ± 7 GHz). Using low and high frequen-

cies allows to constrain the retrievals below and above the freezing level, respectively.

The inversion experiments were performed for each frozen hydrometeor shape from Liu’s database of scattering properties (Liu 2008). The conclusions drawn from this study are that the simulated brightness temperatures of the retrieved profiles remain consistent with observed brightness temperatures with any of the scattering properties selected. However, depending on the choice of particle shape for snowfall representation, the retrieved atmospheric profiles can vary significantly at high levels (moister profiles for less efficient scatterers). Indeed, a weakly scattering particle leads to a retrieval with larger hydrometeor and relative humidity contents. The opposite behavior was also found for the cloud liquid water content and the relative humidity content at low levels and need further investigations.

Finally, the ML and the WA estimators show differences over two cases: (i) ML does not depend on the observation error which is one strength of this estimator, and (ii) the fit to observations is degraded with ML when using a reduced number of channels.

This study brings several perspectives for testing the GMI retrieval assimilation as well as retrievals from other instruments of the GPM constellation.

First, the sensitivity study on the observation error highlighted the benefit of considering low values. Then, depending on the instrument frequencies selected in the Bayesian inversion, the resulting relative humidity profiles should be filtered out vertically. This will allow the assimilation of relative humidity in the 3D-Var or in the 4D-Var only at relevant pressure levels. To assess the quality of the retrievals, an independent validation would be to compare them with reflectivity profiles from the DPR onboard the GPM-Core satellite. Then, the Bayesian inversion could be enhanced by additional channels. Particularly, the use of temperature sounding channels will be investigated in the future to better constrain the retrievals. The MicroMas instruments onboard Time-Resolved Observations of Precipitation structure and storm Intensity with a Constellation of Smallsats could be used to conduct such study (Blackwell et al. 2012).

The sensitivity study on the radiative properties highlighted that, as in the direct assimilation methods, it is essential to optimize the radiative properties within the observation operator. One interesting feature of the Bayesian inversion over the direct assimilation of brightness temperatures is that it is possible to use several different radiative properties within the inversion thus extending the database of profiles. This

possibility will be tested in a future study.

Finally, the ML estimator shows both advantages and disadvantages over the weighted mean estimator.

Therefore, further comparisons will be performed and evaluated within assimilation experiments. One interesting feature of the Bayesian inversion is that it is possible to use different radiative properties of hydrometeors within the inversion by extending the database of simulated profiles. This option can be used both with the WA and ML estimators. This will be the subject of future research to further explore how the variability of hydrometeor shapes and distributions observed in nature can be represented within a data assimilation system.

Acknowledgments

This research is funded by Météo-France and Région Occitanie (PhD grant for Marylis Barreyat). The authors acknowledge the Centre National d'Études Spatiales (CNES) for the financial support of this scientific research activity part of the Infrarouge, Micro-Ondes et Transfert radiatif ensembliste pour la prévision des Extrêmes de Précipitations (IMOTEP) project. The two anonymous reviewers are acknowledged for their useful comments which helped improve the manuscript.

References

- Bauer, P., P. Lopez, A. Benedetti, D. Salmond, and E. Moreau, 2006a: Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Quart. J. Roy. Meteor. Soc.*, **132**, 2277–2306.
- Bauer, P., P. Lopez, D. Salmond, A. Benedetti, S. Saarinen, and M. Bonazzola, 2006b: Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. II: 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **132**, 2307–2332.
- Bauer, P., E. Moreau, F. Chevallier, and U. O'keeffe, 2006c: Multiple-scattering microwave radiative transfer for data assimilation applications. *Quart. J. Roy. Meteor. Soc.*, **132**, 1259–1281.
- Bauer, P., A. J. Geer, P. Lopez, and D. Salmond, 2010: Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quart. J. Roy. Meteor. Soc.*, **136**, 1868–1885.
- Blackwell, W., G. Allen, C. Galbraith, T. Hancock, R. Leslie, I. Osaretin, L. Retherford, M. Scarito, C. Semisch, M. Shields, M. Silver, D. Toher, K. Wight, D. Miller, K. Cahoy, and N. Erickson, 2012: Nanosatellites for earth environmental monitoring: The MicroMAS project. *12th Specialist Meeting on Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*, IEEE, 1–4.
- Borderies, M., O. Caumont, C. Augros, É. Bresson, J. Delanoë, V. Ducrocq, N. Fourrié, T. Le Bastard, and M. Nuret, 2018: Simulation of W-band radar reflectivity for model validation and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **144**, 391–403.
- Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the AROME-France system. *Quart. J. Roy. Meteor. Soc.*, **142**, 2231–2243.
- Caumont, O., V. Ducrocq, É. Wattrelot, G. Jaubert, and S. Pradier-Vabre, 2010: 1D+3DVar assimilation of radar reflectivity data: A proof of concept. *Tellus A*, **62**, 173–187.
- Chambon, P., L.-F. Meunier, F. Guillaume, J.-M. Piriou, R. Roca, and J.-F. Mahfouf, 2015: Investigating the impact of the water-vapour sounding observations from SAPHIR on board Megha-Tropiques for the ARPEGE global model. *Quart. J. Roy. Meteor. Soc.*, **141**, 1769–1779.
- Courtier, P., C. Freydier, J.-F. Geleyn, F. Rabier, and M. Rochas, 1991: The ARPEGE project at Météo-France. *Proceeding of the ECMWF Seminar on Numerical Methods in Atmospheric Models*, 9–13 September 1991, Shinfield Park, Reading, 193–231. [Available at <https://www.ecmwf.int/en/elibrary/8798-arpege-project-meteo-france>.]
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396.
- Duruissseau, F., P. Chambon, E. Wattrelot, M. Barreyat, and J.-F. Mahfouf, 2019: Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Quart. J. Roy. Meteor. Soc.*, **145**, 620–641.
- Eriksson, P., R. Ekelund, J. Mendrok, M. Brath, O. Lemke, and S. A. Buehler, 2018: A general database of hydrometeor single scattering properties at microwave and sub-millimetre wavelengths. *Earth Syst.*, **10**, 1301–1326.
- Errico, R. M., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798.
- Faure, G., P. Chambon, and P. Brousseau, 2020: Operational Implementation of the AROME model in the tropics: Multiscale validation of rainfall forecasts. *Wea. Forecasting*, **35**, 691–710.
- Field, P. R., A. J. Heymsfield, and A. Bansemmer, 2007: Snow size distribution parameterization for midlatitude and tropical ice clouds. *J. Atmos. Sci.*, **64**, 4346–4365.
- Geer, A. J., and P. Bauer, 2010: *Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation*. ECMWF Reading, UK, Technical Memorandum, No. 620, 43 pp. [Available at <https://www.ecmwf.int/node/9518>.]

- Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor. Soc.*, **137**, 2024–2037.
- Geer, A. J., and F. Baordo, 2014: Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmos. Meas. Tech.*, **7**, 1839–1860.
- Geer, A. J., P. Bauer, and P. Lopez, 2008: Lessons learnt from the operational 1D + 4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **134**, 1513–1525.
- Geer, A. J., P. Bauer, and P. Lopez, 2010: Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment. *Quart. J. Roy. Meteor. Soc.*, **136**, 1886–1905.
- Geer, A. J., F. Baordo, N. Bormann, P. Chambon, S. English, M. Kazumori, H. Lawrence, P. Lean, K. Lonitz, and C. Lupu, 2017: The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteor. Soc.*, **143**, 3189–3206.
- Guerbette, J., J.-F. Mahfouf, and M. Plu, 2016: Towards the assimilation of all-sky microwave radiances from the SAPHIR humidity sounder in a limited area NWP model over tropical regions. *Tellus A*, **68**, 28620, doi: 10.3402/tellusa.v68.28620.
- Haddad, Z. S., J. L. Steward, H.-C. Tseng, T. Vukicevic, S.-H. Chen, and S. Hristova-Veleva, 2015: A data assimilation technique to account for the nonlinear dependence of scattering microwave observations of precipitation. *J. Geophys. Res: Atmos.*, **120**, 5548–5563.
- Hou, A. Y., and S. Q. Zhang, 2007: Assimilation of precipitation information using column model physics as a weak constraint. *J. Atmos. Sci.*, **64**, 3865–3878.
- Hou, A. Y., S. Q. Zhang, and O. Reale, 2004: Variational continuous assimilation of TMI and SSM/I rain rates: Impact on GEOS-3 hurricane analyses and forecasts. *Mon. Wea. Rev.*, **132**, 2094–2109.
- Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722.
- Ikuta, Y., 2016: Data assimilation using GPM/DPR at JMA. *CAS/JSC WGNE Res. Activ. Atmos. Oceanic Modell.*, **46**, 01–11.
- Ikuta, Y., and Y. Honda, 2011: Development of 1D+ 4DVAR data assimilation of radar reflectivity in JNoVA. *CAS/JSC WGNE Res. Activ. Atmos. Oceanic Modell.*, **41**, 01–09.
- Kim, M.-J., J. Jin, A. El Akkraoui, W. McCarty, R. Todling, W. Gu, and R. Gelaro, 2020: The framework for assimilating all-sky GPM Microwave Imager brightness temperature data in the NASA GEOS data assimilation system. *Mon. Wea. Rev.*, **148**, 2433–2455.
- Kneifel, S., J. Leinonen, J. Tyynelä, D. Ori, and A. Battaglia, 2020: Scattering of hydrometeors. *Satellite Precipitation Measurement*. Levizzani, V., C. Kidd, D. Kirschbaum, C. Kummerow, K. Nakamura, and F. Turk (eds.), *Advances in Global Change Research*, **67**, Springer, Cham, 249–276.
- Kummerow, C., W. S. Olson, and L. Giglio, 1996: A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sens.*, **34**, 1213–1232.
- Lean, P., A. J. Geer, and K. Lonitz, 2017: *Assimilation of global precipitation mission (GPM) microwave imager (GMI) in all-sky conditions*. ECMWF Technical Memorandum, No.799, ECMWF, Reading, UK, 30 pp. [Available at <https://www.ecmwf.int/node/17174/>.]
- Liu, G., 2008: A database of microwave single-scattering properties for nonspherical ice particles. *Bull. Amer. Meteor. Soc.*, **89**, 1563–1570.
- Mahfouf, J.-F., P. Bauer, and V. Marécal, 2005: The assimilation of SSM/I and TMI rainfall rates in the ECMWF 4D-Var system. *Quart. J. Roy. Meteor. Soc.*, **131**, 437–458.
- Marécal, V., and J.-F. Mahfouf, 2002: Four-dimensional variational assimilation of total column water vapor in rainy areas. *Mon. Wea. Rev.*, **130**, 43–58.
- Marécal, V., and J.-F. Mahfouf, 2003: Experiments on 4D-Var assimilation of rainfall data using an incremental formulation. *Quart. J. Roy. Meteor. Soc.*, **129**, 3137–3160.
- 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. *Meteor. Atmos. Phys.*, **54**, 79–99.
- Petty, G. W., and K. B. Katsaros, 1990: Precipitation observed over the South China Sea by the Nimbus-7 Scanning Multichannel Microwave Radiometer during Winter MONEX. *J. Appl. Meteor.*, **29**, 273–287.
- Ricard, D., C. Lac, S. Riette, R. Legrand, and A. Mary, 2013: Kinetic energy spectra characteristics of two convection-permitting limited-area models AROME and Meso-NH. *Quart. J. Roy. Meteor. Soc.*, **139**, 1327–1341.
- Roca, R., M. 981 Dejus, P. Chambon, S. Cloché, and M. Capderou, 2020: The Megha-Tropiques mission after seven years in space. *Satellite Precipitation Measurement*. Levizzani, V., C. Kidd, D. Kirschbaum, C. Kummerow, K. Nakamura, and F. Turk (eds.), *Advances in Global Change Research*, **67**, Springer, Cham, 45–62.
- Saunders, R., J. Hocking, E. Turner, P. Rayer, D. Rundle, P. Brunel, J. Vidot, P. Roquet, M. Matricardi, A. Geer, N. Bormann, and C. Lupu, 2018: An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci. Model. Dev.*, **11**, 2717–2737.
- Schmitt, C. G., A. J. Heymsfield, P. Connolly, E. Järvinen, and M. Schnaiter, 2016: A global view of atmospheric ice particle complexity. *Geophys. Res. Lett.*, **43**, 11913–11920.

- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991.
- Skofronick-Jackson, G., W. A. Petersen, W. Berg, C. Kidd, Stocker, E. F., D. B. Kirschbaum, R. Kakar, S. A. Braun, G. J. Huffman, T. Iguchi, P. E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W. S. Olson, Y. N. Takayabu, K. Furukawa, and T. Wilheit, 2017: The Global Precipitation Measurement (GPM) mission for science and society. *Bull. Amer. Meteor. Soc.*, **98**, 1679–1695.
- Treadon, R. E., H.-L. Pan, W.-S. Wu, Y. Lin, W. S. Olson, and R. J. Kuligowski, 2002: Global and regional moisture analyses at NCEP. *Proceeding of the ECMWF/GEWEX Workshop on Humidity Analysis*, 33–47. [Available at <https://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/2002-humidity-analysis>.]
- Wattrelot, E., O. Caumont, and J.-F. Mahfouf, 2014: Operational implementation of the 1D+3D-Var assimilation method of radar reflectivity data in the AROME model. *Mon. Wea. Rev.*, **142**, 1852–1873.

5.3 Conclusions

Une étude de sensibilité à la première étape de la méthode d'assimilation de températures de brillance micro-ondes en ciel nuageux (de type 1D-Bay+3D/4D-Var) a été réalisée sur différents aspects : (i) les erreurs d'observations, (ii) la sélection des canaux (iii) les SSPs et (iv) le choix d'estimateur optimal. Pour cela, nous avons exploité les observations des fréquences entre 18.7 et 183.31 GHz du radiomètre micro-ondes GMI disponibles sur une période de deux mois en 2017, ainsi que le modèle de PNT à échelle convective AROME-Antilles. Pour rappel, ce type de modèle résout la convection profonde de manière explicite contrairement au modèle global ARPEGE. Ces études ont permis de mieux comprendre le fonctionnement de cette méthode et d'évaluer sa sensibilité aux SSPs des hydrométéores solides. Nous résumons ci-dessous les principales conclusions de cette étude.

Nous avons premièrement évalué la sensibilité de la méthode à la spécification des erreurs d'observation sur une plage de variation comprise entre 1 K et 20 K. Il apparaît que les erreurs les plus grandes ont tendance à lisser et à assécher les profils inversés d'humidité relative, ainsi qu'à produire des profils d'hydrométéores moins réalistes. Ces résultats s'expliquent de par le fait qu'une forte erreur d'observation a tendance à faire se rapprocher la moyenne pondérée d'une moyenne arithmétique, amenant ainsi à un mélange entre scènes météorologiques en ciel clair et en ciel nuageux. Au contraire, de faibles valeurs amènent la méthode à produire des profils inversés plus cohérents avec les températures de brillance observées, en privilégiant les profils atmosphériques de l'ébauche dont les températures de brillance simulées sont proches des observations. Notons toutefois que l'erreur de représentativité, définie comme la différence entre les observations et les profils simulés, demeurera et qu'il n'est pas possible de la diminuer. En complément, des diagnostics a posteriori de type Desroziers ont été réalisés pour spécifier les termes diagonaux d'erreurs d'observations de manière plus objectives (Desroziers et al. 2005[188]). Cette étude complémentaire, dont les résultats ne sont pas montrés ici, aboutit à des conclusions très proches de celles obtenues à partir de l'étude de sensibilité, à savoir une faible erreur d'observation identique pour tous les canaux. Il est toutefois important de noter ici que les diagnostics de Desroziers peuvent également être exploités pour calculer les termes d'erreur d'observations non-diagonaux, négligés dans cette étude, comme il est actuellement fait pour les instruments infrarouges ([195]).

Ensuite, nous avons évalué la sensibilité de la méthode à la sélection des canaux. Les études précédentes à Météo-France ayant été seulement réalisées avec des fréquences micro-ondes centrées autour de la bande d'absorption de la vapeur d'eau à 183.31 GHz, il s'agissait d'évaluer l'apport des basses fréquences. Pour cela des inversions Bayésiennes avec ajout de fréquences ont été réalisées. La faisabilité de l'inversion Bayésienne avec un plus grand ensemble de fréquences a d'abord été montrée. L'ajout de canaux basses fréquences apporte des informations sur les basses couches atmosphériques et conduit à des restitutions de meilleure qualité, tout en mettant en évidence l'utilité des deux groupes de fréquences. Ainsi, les basses fréquences amènent principalement de l'information en-dessous de l'isotherme 0°C et les hautes au-dessus de celle-ci. Comme ceci, une assimilation sur toute la verticale du profil d'humidité relative ne serait pas adaptée pour cette méthode. Une perspective serait d'adapter de manière dynamique l'assimilation des profils d'humidité relative inversés au sein d'un 3D ou 4D-Var, à des niveaux de pression pertinents plutôt qu'à des niveaux fixés à l'avance. Une étude que nous avons réalisé dans le cadre de mon stage de Master 1 donne des pistes de méthodologie pour cette sélection dynamique ; elle a été publiée en annexe de l'article Duruisseau et al. 2019[44]. Cette annexe est disponible dans l'annexe 1 de ce manuscrit.

L'évaluation suivante a concerné la sensibilité de la méthode à la spécification des SSPs des hydrométéores solides. Pour cela nous avons réalisé des inversions avec chaque SSP de la base de Liu2008[11]. Nous avons vu que le choix de la SSP influençait fortement la forme des profils d'humidité relative inversés : une SSP générant peu de diffusion conduit à une inversion avec des contenus en hydrométéores et en humidité relative plus importants qu'avec une SSP générant beaucoup de diffusion où les contenus sont plus faibles.

Finalement, nous avons examiné la sensibilité de la méthode au choix de l'estimateur optimal défini pour inverser les profils d'humidité relative. Pour cela nous avons comparé la moyenne pondérée utilisée à Météo-France à un maximum de vraisemblance utilisé à la JMA. La méthode du maximum de vraisemblance permet de supprimer le degré de liberté concernant l'erreur d'observation. En effet la méthode ne tient pas compte de cette information pour inverser les profils atmosphériques lorsque les erreurs d'observation sont identiques pour tous les canaux. Néanmoins, cet aspect peut être également vu comme un avantage pour la moyenne pondérée car cela permet à cette méthode d'être plus flexible. Ensuite la

méthode du maximum de vraisemblance donne des résultats dégradés par rapport à la moyenne pondérée lorsqu'un petit nombre de canaux est utilisé et des résultats similaires sont trouvés pour la sensibilité à la spécification des SSPs des hydrométéores solides.

Cette étude ouvre ainsi plusieurs perspectives pour l'assimilation des observations satellitaires micro-ondes par une méthode Bayésienne. L'apport positif ajouté par les basses fréquences permet d'envisager l'utilisation d'autres instruments de la constellation micro-ondes pourvus de celles-ci (instruments imageurs) en plus des canaux vapeur d'eau à 183.31 GHz. Nous avons mis en évidence l'intérêt de considérer des valeurs relativement faibles d'erreurs d'observation, tout comme la sélection des basses fréquences a permis de démontrer leur utilité avec les hautes fréquences pour restituer une information cohérente sur la verticale. La nécessité de filtrer sur la verticale les profils d'humidité relative inversés selon les canaux disponibles est apparue, pour que leur assimilation intervienne uniquement à des altitudes pertinentes. Nous avons démontré que l'estimateur basé sur une moyenne pondérée amène à une meilleure restitution sur toute la verticale par rapport à un maximum de vraisemblance, même si ce dernier possède l'avantage de ne pas être sensible à la spécification de l'erreur d'observation. Enfin nous avons montré la grande sensibilité des inversions à la spécification des SSPs. Cette conclusion démontre l'importance de bien spécifier les SSPs au sein des systèmes d'assimilation. Elle illustre aussi les limites à ne sélectionner qu'un seul jeu de SSPs lors de l'assimilation de données, ce qui peut conduire à une mauvaise interprétation du signal radiométrique. Dans le chapitre suivant nous allons voir une adaptation de l'inversion Bayésienne permettant de spécifier plusieurs SSPs distinctes afin de limiter ce risque.

Chapitre 6

Étude d’une méthode d’inversion considérant les incertitudes du transfert radiatif

Sommaire

6.1 Introduction	107
6.2 Publication en cours de révision	109
6.3 Résultats complémentaires	127
6.3.1 Étude statistique avec les prédicteurs GPROF	127
6.3.2 Résultats d’expériences d’assimilation avec le modèle ARPEGE	127
6.4 Conclusions	134

6.1 Introduction

Comme annoncé au chapitre 4, une des difficultés pour simuler les températures de brillance micro-ondes en zones nuageuses réside dans les SSPs pour les hydrométéores (notamment pour les particules solides précipitantes) qui doivent être spécifiées dans l’opérateur d’observation. Le chapitre 5 a mis en évidence l’importance de cette spécification des SSPs pour les hydrométéores solides au sein des systèmes d’assimilation de données. Dans la suite de ce manuscrit nous allons développer des méthodes afin de pouvoir considérer plusieurs SSPs par type d’hydrométéore, et non une seule comme dans la plupart des systèmes actuels de PNT. Ces méthodes permettront également de prendre en compte la diversité des formes présentes dans la nature (figure 2.10) et de pouvoir tirer à termes meilleur profit des nombreuses bases de données des hydrométéores solides.

Dans ce chapitre nous allons étudier une méthode exploitant le contenu en information des observations afin de définir une combinaison optimale de SSPs. Pour cela nous allons nous servir de la méthode 1D-Bay+3D/4D-Var introduite au chapitre 3. Une modification permettant de considérer plusieurs SSPs est ajoutée à la formulation mathématique de l’inversion Bayésienne. Cette modification est réalisée au niveau de la moyenne pondérée des profils atmosphériques de l’ébauche : une somme sur les SSPs est ajoutée à celle-ci afin de pondérer chaque profil par autant de températures de brillance préalablement simulées qu’il y a de SSPs (et de fréquences) considérées. Cette nouvelle formule est donnée par :

$$x_{inv} := \frac{\sum_{j=1}^p \sum_{i=1}^n x_i w_{i,j}}{\sum_{j=1}^p \sum_{i=1}^n w_{i,j}}$$

Avec $w_{i,j} = e^{(-\frac{1}{2}[y_o - H_j(x_i)]^t R^{-1}[y_o - H_j(x_i)])}$, p le nombre de SSPs prises en compte au sein de l'inversion Bayésienne, n le nombre de profils dans le voisinage, y_o l'observation, x_i les profils de variables atmosphériques (température, contenus en humidité et en hydrométéores) du voisinage, \mathbf{R} la matrice de covariances d'erreurs d'observations et H_j l'opérateur d'observation simulant les températures de brillance à l'aide de la SSP j . Cette méthode, que nous appellerons méthode 'multi-particules' par la suite, permet de prendre en compte autant de SSPs que souhaitées en tirant profit de la richesse des bases de données disponibles. Ainsi, la méthode peut attribuer un poids significatif à une seule SSP et négliger les autres. Elle offre donc la possibilité d'examiner dans un premier temps la pertinence de chaque SSP puis de réaliser de possibles combinaisons de ces dernières. Dans un deuxième temps, nous analyserons ces résultats en les catégorisant à l'aide d'informations pertinentes sur l'atmosphère (prédicteurs) afin d'étudier si des combinaisons de SSPs seraient plus fréquemment utilisées dans certaines situations météorologiques.

Afin de tester cette méthode nous avons réalisé une étude avec le modèle AROME-Antilles et les observations du radiomètre GMI sur une période de 2 mois allant de septembre à octobre 2017. Cette période a été particulièrement convective sur les Antilles cette année-là, avec le passage en 15 jours de temps de deux cyclones de catégorie 5 (*Irma* et *Maria*) et un de catégorie 4 (*Jose*). Nous avons défini deux expériences utilisant différentes configurations de SSPs. La première utilise une configuration composée de 3 SSPs. Celles-ci sont caractérisées par des distributions en températures de brillance simulées encadrant la distribution des températures de brillance observées (voir figure 1 de l'article dans ce chapitre). Ces trois SSPs sont : *Rosette 3-bullet*, *Rosette 6-bullet* et *Sector snowflake*. Cette expérience est notée dans la suite du manuscrit "3SSP". La deuxième expérience utilise une configuration composée des trois mêmes SSPs et d'une SSP additionnelle caractérisée par une distribution atteignant des températures de brillance jusqu'à 50K : la SSP *Block column* (voir figure 1 de l'article). Cette expérience est notée dans la suite du manuscrit "4SSP". La distribution en températures de brillance simulées de la SSP *Block column* est sensiblement différente de celles observées. Même si a priori *Block column* ne semble pas être un choix judicieux au sein des expériences d'assimilation, cela permettra d'évaluer la pertinence d'une telle SSP pour des situations météorologiques ayant une activité convective très intense.

Pour chaque scène météorologique, la nouvelle méthode utilise une combinaison différente de SSPs. Afin d'analyser de manière statistique ces combinaisons et proposer des interprétations physiques, les résultats d'inversion ont été catégorisés à l'aide de trois prédicteurs :

- Les deux premiers sont **les précipitations de surface** (mm/hr) et **le contenu intégré en glace** (kg/m^2) provenant des produits GPROF (Goddard PROFiling, Kummerow, 2016[196]). Les produits GPROF sont construits à l'aide d'une base de données d'observations du satellite GPM Core. Grâce à un formalisme Bayésien, cet algorithme fournit des taux de pluie moyennés en surface et des structures verticales d'hydrométéores ainsi que leurs incertitudes d'estimation associées. Les produits utilisés dans cette étude proviennent de la version 5 de GPROF exploitant les observations de plusieurs instruments micro-ondes, comme le radiomètre GMI (voir documentation). Afin d'exploiter les produits GPROF, dont la résolution spatiale est de $13\text{ km} \times 13\text{ km}$, les observations de GMI ont été colocalisées avec ceux-ci sur une grille régulière $0.1^\circ \times 0.1^\circ$. Enfin, nous avons utilisé un indice fournit par les produits GPROF afin de traiter uniquement les données jugées de bonne qualité.

- Le troisième et dernier prédicteur est **un indice de diffusion** calculé à partir des températures de brillance observées et permettant de caractériser la diffusion d'une scène météorologique. Cet indice exprimé en Kelvin a été initialement conçu pour détecter les précipitations. Il est calculé à partir des fréquences 18.7v, 23.8v et 89v GHz (Grody 1991[197], Wilheit et al. 2003[198]) comme :

$$P_{SI} = 451.9 - 0.44TB_{Obs}^{18.7v} - 1.775TB_{Obs}^{23.8v} + 0.00575(TB_{Obs}^{23.8v})^2 - TB_{Obs}^{89v}$$

Plus l'indice P_{SI} sera élevé, plus celui-ci détectera une scène caractérisée par une forte diffusion, c'est-à-dire avec un contenu important en hydrométéores solides.

Finalement, pour ne traiter que les inversions réalisées à partir d'observations nuageuses/pluvieuses, nous avons conservé uniquement les observations répondant aux critères suivants :

- un indice $P_{SI} > 5\text{ K}$;
- un contenu en précipitation de surface $> 0.1\text{ mm/hr}$;
- et avec un contenu intégré en glace $> 0.1\text{ kg/m}^2$.

Les principaux résultats de cette étude ont été soumis pour publication dans le journal *Atmospheric Science Letters* (ASL). La première section de ce chapitre comporte la version de l'article re-soumise après les premières révisions proposées par le comité de lecture, la deuxième section comporte des résultats complémentaires sur cette étude, notamment avec des résultats d'expériences d'assimilation au sein du modèle ARPEGE. La dernière section rassemble les principales conclusions de cette étude et propose des perspectives à celle-ci.

6.2 Publication en cours de révision

3 **A 1D Bayesian inversion of microwave radiances**
4 **using several radiative properties of solid**
5 **hydrometeors**

6 **Marylis Barreyat¹ | Philippe Chambon^{1*} |**
Jean-François Mahfouf¹ | Ghislain Faure¹

7 ¹CNRM, Université de Toulouse,
Météo-France, CNRS, Toulouse, France.

NWP centres increasingly make use of cloudy and rainy mi-

Correspondence

*Philippe Chambon, CNRM, Université de
Toulouse, Météo-France, CNRS, Toulouse,
France

Email: philippe.chambon@meteo.fr

Funding information

This research is funded by Météo-France
and Région Occitanie (PhD grant for Marylis
Barreyat). The authors acknowledge the
Centre National d'Études Spatiales (CNES)
for the financial support of this scientific
research activity part of the Infrarouge,
Micro-Ondes et Transfert radiatif
ensembliste pour la prévision des Extrêmes
de Précipitations (IMOTEP) project.

*Equally contributing authors.

crowave radiances. Currently, the high microwave frequencies are simulated using simplified assumptions regarding radiative properties of frozen hydrometeors. In particular, one single particle shape is often used for all precipitating frozen particle, all over the globe and for all cloud types. In this paper, a multi-SSP (Single Scattering Properties) approach for 1D Bayesian inversions is examined. Two experiments were set up : (1) one with three SSPs and (2) one with the previous SSPs plus one which leads to very cold brightness temperature distributions. For that purpose, we used observations from the GMI radiometer over a two month period and forecasts from the Météo-France convective scale AROME model. The results showed that mixtures of SSP are chosen by the inversion method for meteorological conditions with low scattering and that a single particle is chosen for those with high scattering to perform the inversions. Despite the fact that no specific weather scenes was found to be associated to a particular SSP the most efficient scattering particles can be favoured for some of them.

KEYWORDS

microwave, cloud, snow, hydrometeors

8

9 1 | INTRODUCTION

10 In recent years, the exploitation of high microwave frequencies in cloudy sky conditions has significantly progressed
11 both for Numerical Weather Prediction (NWP) applications (e.g. Geer and Baordo 2014 [1], Geer et al 2017 [2]) and
12 for surface rain retrieval applications (e.g. Kidd et al., 2016 [3], Kidd et al., 2021 [4]). To improve the quality of the
13 retrievals, many aspects of their treatment processes could be ameliorated like radiative transfer simulations. Similarly
14 for NWP models, this type of observations also requires the use of radiative transfer codes within which scattering
15 properties parameterizations could also be improved in case of cloudy and/or rainy observations (Geer et al. 2017[2]).
16 Thanks to those codes, the NWP model brightness temperatures (Bts) are simulated according to a chosen frequency.
17 Several micro-physical assumptions on solid hydrometeors, regarding the single scattering properties (SSPs), and the
18 size distribution, are indeed required. Within current radiative transfer models, one single set of radiative properties
19 has to be selected in advance for each hydrometeor type. The resulting bulk properties (single scattering albedo,
20 extinction, asymmetry, backscattering) are then stored into look-up-tables as function of hydrometeor content and
21 temperature ; these look-up-tables are then to be further used within assimilation systems. Since high frequency
22 microwave radiances are sensitive to snowfall which can have a wide range of shapes, sizes and densities for example,
23 an accurate SSP specification is crucial for an optimal exploitation (Kulie et al. 2010[5], Geer and Baordo, 2014[1]),

24 Ringerud et al. 2019[6], Barreyat et al. 2021[7]).

25 Currently, in most NWP centres exploiting cloudy and rainy radiances, the high microwave frequencies are simulated
26 with a single "optimal" particle shape for snowfall. Several studies focused on searching for optimal particle radiative
27 properties to fit the microwave observation distributions (e.g. Geer and Baordo 2014 [1], Haddad et al. 2015 [8],
28 Guerbette et al. 2016 [9], Mangla et al. 2021 [10]). Optimisations including different solid hydrometeor properties
29 have been investigated : in Geer et al. 2021[11] a parameter estimation method is used in order to select an optimal
30 shape for several species at the same time. Knowing the diversity of the possible solid hydrometeor characteristics
31 (Nakaya 1954 [12], J Hallett and BJ Mason 1958 [13], Magono and Lee 1966 [14]), this represents a huge simplification
32 of the atmosphere despite the availability of databases providing scattering properties for many solid hydrometeors
33 (Liu 2008 [15], Ding et al 2017 [16], ARTS Eriksson 2018 [17], Brath, M. et al. 2020 [18], Kneifel et al. 2018 [19]).
34 Several studies have investigated the use of ensembles of radiative properties, for different part of the spectrum.
35 Baran and Labonnote (2007[20]) developed a model for cirrus clouds using a set of crystals to simulate brightness
36 temperatures in the infrared range. Kulie et al. (2010[5]) averaged radiative transfer simulations with the properties
37 available in the databases of Liu (2008[15]) and Hong et al. (2007[21]) to derive microwave brightness temperatures.
38 Using this latter method within a precipitation retrieval algorithm (Kummerow et al. 2001[22], Randel et al.2020[23]),
39 Ringerud et al. (2019[6]) obtained positive improvements with the database of Liu (2008[15]).

40 At Météo-France, the '1D-Bay+3D/4D-Var' scheme is operationnaly used to assimilate ground based radar reflectivi-
41 ties since 2010 (Caumont et al. 2010 [24], Wattrelot et al. 2014 [25]) and cloudy and/or rainy microwave radiances
42 from the ATMS and MHS sounders since 2021 (Duruiseau et al. 2019 [26]). In order to consider the variability of solid
43 hydrometeor, a method based on the first step of the '1D-Bay+3D/4D-Var' scheme, and choosing between mixtures
44 of SSPs, has been developed to retrieve vertical profiles of relative humidity from cloudy and/or rainy observations.
45 Unlike previous explored methods, the present one dynamically defines optimal radiative property ensembles using
46 the microwave observations as an a priori information. The aim of this study is to examine the relevance of such a
47 method to perform inversions of cloud and/or rainy microwave radiances. It is part of a research project aiming to
48 take into account within data assimilation the variability of hydrometeor properties. Before performing an evaluation
49 through data assimilation experiments, the particle choices within the revised 1D-Bayesian inversion scheme need to
50 be documented. Therefore, our main objective is not yet to quantify the benefits of such novel approach but to exam-
51 ine its potential in order to exploit it to the full afterwards. The outcome from this study will be to provide guidance
52 on how to exploit scattering property databases more efficiently in a data assimilation framework.

53 Section 2 describes the NWP system and the observations used in this paper to conduct experiments with the 1D
54 Bayesian inversion scheme. The experiments and the products selected to evaluate the results are introduced in
55 section 3. In section 4 results from the experiments are presented and discussed. Finally conclusions drawn from this
56 study are summarized in Section 5.

57 2 | GMI DATA AND NWP SYSTEM

58 The experiments with different mixtures of SSPs have been conducted over a 2-month period using Bts from level 1B
59 products from the GPM Microwave Imager (GMI L1B ATBD, 2016) within a ± 90 min time window around the time
60 of validity of 3-hour forecasts of the Application of Research to Operations at Mesoscale (AROME)-Antilles NWP
61 model (Faure et al. 2020 [27]). This convective scale model, with a resolution of 2.5 km, is used operationally at
62 Météo-France over several geographical domains in the Tropics including the Caribbean geographical area [10.4° N
63 to 22.45° N ; 67.8° W to 52.2° W]. The microphysical scheme of this model generates prognostic cloud liquid water,

cloud ice water, rain, snow and graupel mixing ratios. We selected GMI Bts over a period of high convective activity from 1 September 2017 to 31 October 2017. This instrument was chosen because of its wide range of frequencies from 10.65 GHz to 183.31±7 GHz. Due to different channel footprint sizes across frequencies, the raw Level 1B data were superobbed onto a regular lat/lon grid at 0.1° deg resolution. Considering the fact that numerical models have an effective resolution three to four times larger than their actual grid (Ricard et al, 2013 [28]), this resolution for superobbing is therefore compatible with the one of the AROME-Antilles model. Then, the GMI Bts were simulated with the version 12 of the Radiative Transfer for TIROS Operational Vertical sounder code RTTOV-SCATT (Saunders et al. 2018 [29]). This code, based on the Delta-Eddington approximation, needs, among other parameters, the bulk scattering properties of the hydrometeors. These bulk properties are computed by integrating the SSPs over particle sizes using a particle size distribution (PSD). An important feature is that the particle mass needs to be defined as a function of the maximum dimension of the particle. This is done thanks to the particle mass-size function which depends on two coefficients. In this study both coefficients are taken from Table 1 of Kulie et al. (2010[5]). Additionally, the RTTOV-SCATT version used in this study does not consider the graupel hydrometeor. Therefore, we summed the graupel content with the snow content generated by the forecast model. Finally, the chosen characteristics to describe the hydrometeors are :

79

- A Marshall-Parlmer[30] PSD and a Mie sphere to simulate the rain ;
- A Field et al. 2007[31] tropical PSD and the 11 different shapes from Liu (2008)[15] to simulate the snow ;
- A modified gamma[32] PSD and a Mie sphere to simulate the cloud liquid water ;
- A modified gamma[32] PSD and a Mie sphere to simulate the cloud ice water.

3 | METHOD

3-a Experiment definition

A 1D Bayesian inversion has been developed at Météo-France to retrieve atmospheric profiles from cloudy microwave Bts (Guerbette et al. 2016 [9], Duruisseau et al. 2019 [26], Barreyat et al. 2021 [7]). The retrieved profiles are derived from a weighted average of profiles x_i in the neighborhood of a given observation y . A 250 × 250 km domain is chosen around each observation with profiles taken from an AROME short-range forecast (First-Guess: FG). The largest weights are given to the simulated Bts closest to the observed ones. The retrieved profile is given by :

$$x_{ret} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

With

$$w_j = e^{\sum_{j=1}^d (-\frac{1}{2} [y_{o,j} - H_j(x_i) - b_j]^t R^{-1} [y_{o,j} - H_j(x_i) - b_j])}$$

where i corresponds to a given profile in the inversion database (excluding the first guess), j is a given channel, d is the number of channels selected for the inversion, $H(\cdot)$ the observation operator, R the covariance matrix of observation errors and b_j is a clear-sky bias correction between the FG and the observations previously computed for all GMI

94 frequencies over a 2-month period. In order to process the largest possible number of inversions, the accepted weight
 95 threshold has been set to 10^{-200} . If this threshold is exceeded, it leads to a successful inversion.

96 In the 1D Bayesian inversion developed for the present study, instead of simulating a set of d channels for each
 97 observation, we consider additional Bts, where m is the number of radiative configurations selected in the radiative
 98 transfer model RTTOV-SCATT. Each specification of SSP leads to a change of the PSD free parameter to compute the
 99 bulk scattering properties. However for simplicity, the following results coming from bulk scattering property changes
 100 are commented as coming from SSP changes. Therefore, each retrieved profile is written as:

$$x_{ret} = \frac{\sum_{i=1}^n \sum_{k=1}^m w_{i,k} x_i}{\sum_{i=1}^n \sum_{k=1}^m w_{i,k}}$$

With

$$w_{i,k} = e^{\sum_{j=1}^d (-\frac{1}{2} [y_{o,j} - H_{j,k}(x_i) - b_j]^t R^{-1} [y_{o,j} - H_{j,k}(x_i) - b_j])}$$

101

102

103 For each inversion, we computed the normalized inversion weights for a given SSP. In this paper, the inversions having
 104 a normalized inversion weight greater than 0.95 for one SSP are considered as using a 'single SSP' to perform the
 105 inversions. In the other cases, the inversions are considered as using a mixture of SSPs. To asses the impact of using
 106 various SSPs within the Bayesian inversion, the method is first tested by changing only snowfall radiative properties,
 107 since this hydrometeor has a wide diversity of shapes and densities that strongly impact Bts at high frequencies. Two
 108 experiments using different sets of SSPs have been set up to examine (i) if the method favours the use of mixtures
 109 against the use of a single SSP with a set of SSPs generating moderate scattering and (ii) if the introduction of an
 110 'outlier' SSP, generating a lot of scattering on top of the latter set, has an influence on the results.

111 The first experiment is performed with 3 SSPs from the Liu (2008 [15]) database. These SSPs are Sector snowflake,
 112 Rosette 6-bullet and Rosette 3-bullet. This choice was motivated by their Bt distributions simulated with the AROME-
 113 Antilles model which are close to the observed ones, as shown in Figure 1-a-b-c. In order to highlight the variability of
 114 the simulated Bts depending on particles for the two experiments, the distributions of their differences of simulations
 115 are plotted at the frequency 183.31 ± 7 GHz (Figure 1-d and 1-e). This frequency is interesting due to its sensitivity to
 116 scattering signals produced by ice particles. However, despite similarities in Bt distributions for the three SSPs (Figure
 117 1-a-b-c), Bts simulated with the Sector snowflake can be up to 50 K different from the two other particles (e.g. Figure
 118 1-d).

119

120 For the second experiment, the Bayesian inversion is performed with 4 SSPs. On top of three SSPs from the first
 121 experiment, the block column SSP is added. This particle, generating more scattering than the other ones, very low
 122 Bt values can be simulated (down to 50 K) as shown in Figure 1-a-b-c. Given the differences with observations the
 123 Block column particle is obviously not a particle which one would use a priori as a single particle shape within a data
 124 assimilation experiment. The motivation here is to check if this particle shape could still be useful in some specific
 125 meteorological situations. Figure 1-e shows the spread between the 4 SSPs, and it can be seen that the differences
 126 are much larger than in Figure 1-d with only 3 SSPs. Indeed, the spread reaches up to 140K since with the Block

127 Column particle Bts can be very low.

128 | 3-b Result evaluation metrics

129 The multi-SSP inversion allows to identify, for each retrieved profile, which particles are used with their associated
 130 weights. This aspect is examined in the following to understand the behaviour of the method. Objectively assessing
 131 the quality of the retrievals with multiple particles with respect to a single particle retrieval is a difficult task. One
 132 possibility is to examine the number of successful inversions in each case, a larger number indicating a better fit
 133 to observations. Another possibility is to run the radiative transfer with retrieved profiles to examine the fit of the
 134 retrievals in the observation space. But this means to specify a particular SSP within the RTTOV-SCATT model to
 135 simulate the Bts. We will see in the following the pros and the cons of this method.

136 The results of both experiments have been categorized according to three predictors. The aim of this investigation is
 137 to check whether a given particle shape is more frequently used depending on the weather scene and on its scattering
 138 properties. The first two predictors are the surface precipitation and the ice water path from the GPM GMI (GPROF)
 139 Radiometer Precipitation Profiling L2A 1.5 hours 13 km V05 products collocated with GMI Bts on the regular lat/lon
 140 0.1° grid. The last one is an index characterizing the scattering intensity of the weather scenes. This index, originally
 141 designed to detect precipitation, is calculated from the 18.7v, the 23.8v and the 89v GHz frequencies (Grody 1991[33],
 142 Wilheit et al. 2003[34]). Its expression is given by :

$$P_{SI} = 451.9 - 0.44Bt^{18.7v} - 1.775Bt^{23.8v} + 0.00575(Bt^{23.8v})^2 - Bt^{89v}$$

143 Finally, in order to select only the cloudy and rainy profiles in the observations (so called "cloudy observations" in the
 144 following) to perform the inversions, the sample profiles were selected according to :

- 145 - a P_{SI} index $> 5K$;
- 146 - a surface precipitation amount $> 0.1\text{mm/hr}$;
- 147 - an ice water path amount $> 0.1\text{kg/m}^2$.

148

149 | 4 | EXPERIMENTS

150 In this section a case study is first chosen to examine the behaviour of the multi-SSP 1D Bayesian inversion and to
 151 illustrate it in details. Then, the method is investigated thanks to a statistical study over a 2-month convective period.
 152 The selected case study is associated with the tropical cyclone Maria which developed over the Carribean sea in
 153 September 2017.

154 | 4-a Case study

155 Figure 2-a and 2-b display respectively the cyclone Maria observed at 183.31 ± 7 GHz by GMI and simulated by the
 156 AROME-Antilles model at 3 UTC on 22 September 2017. Despite a good agreement in terms of shape for the main
 157 patterns, the cold Bts around the eye of the cyclone are not located at the same place in the observations and in the
 158 FG.

159 First, we performed the Bayesian inversion allowing a mixture of SSPs with 3 and 4 particles and verified that the
160 weather structures found using the simulated Bts of the retrieved profiles were consistent with the observed ones
161 (not shown). The corresponding weights are shown on Figure 2-c and 2-d. It appears that either a mixture of SSPs or
162 a single SSP was used (normalized inversion weights of one SSP greater than 0.95). The Rosette 3-bullet and Rosette
163 6-bullet are chosen over specific areas (leading to geographically consistent patterns) whereas the Sector snowflake
164 selection is more patchy. However, the particle choice cannot always be interpreted as a function of observation
165 location within the cyclone (e.g. close to the core or within a spiral band). The addition of the Block column particle
166 reduces the use of the Rosette 3-bullet particle in particular inside the south spiral band of the cyclone.

167 | 4-b Statistical study

168 After this qualitative assessment of the multi-SSP approach, results over two months (from September to October
169 2017) are examined using statistical diagnostics. The 3SSP and 4SSP experiments are both compared with another
170 experiment (1SSP). The latter experiment corresponds to a 1D-Bayesian inversion using only the Sector snowflake
171 particle to perform the inversions, which means that no mixtures can be used. The particle choice was motivated by its
172 use in the Météo-France operational system following studies at ECMWF (Geer and Baordo, 2014[1]). Out of a total
173 number of 8537 cloudy observations over the domain and during the period of the study, the number of successful
174 inversions for each experiment are respectively 8238, 8366 and 8425 for the 1SSP, 3SSP and 4SSP experiments. It
175 can be seen that this number increases with the number of SSPs taken into account.

176

177 The standard deviations of differences between analyzed Bts and observations have been computed for the three
178 experiments over a 2-month period and for each GMI frequency used in this study. The relative differences between
179 the standard deviations of the 3 and 4SSPs experiments with respect to the 1SSP experiment are displayed in Figure
180 3. One can see that for the channels 10 to 13, which are the most affected by the specification of snow radiative prop-
181 erties, there is an apparent degradation of the analysis fit to the observations for both the 3 and 4SSPs experiments.
182 However, for the channels 3 to 9, which are the channels the less sensitive to this specification, the results show
183 significant improvements. Indeed, as mentioned above, rerunning the radiative transfer onto the analysed profiles
184 is inevitable to take into account the non linearities of simulations. In this study, the Sector Snowflake was used to
185 rerun the simulations for the three experiments. Hence one interpretation of the degradation of channels 10 to 13
186 is that the latter assumption artificially favours the 1SSP experiment. Therefore, even though the improvements of
187 channels 3 to 9 analysis fits seem encouraging, this method of evaluation should be taken with cautious in particular
188 for the high frequency results.

189

190 The distributions of weights of the cloud and/or precipitation profiles are displayed using violin plot diagrams (Thrun
191 et al. 2020 [35], Barton et al. 2021 [36], Riccardi et al. [37]) as shown in Figure 4. Thanks to their representation
192 of the probability density for each value of the distribution, these diagrams have the advantage of being able to
193 represent multimodal distributions unlike box plot diagrams. In this statistical study, we considered only cloudy and
194 rainy observations defined by the three predictors in section 3-b.

195 In Figures 4-a and 4-f, the particles weights are characterized by multimodal distributions with one mode towards 0
196 and another one towards 1. The Rosette 3-bullet, and more particularly the Rosette 6-bullet and the Sector snowflake
197 particles have their distribution widening around 33% which means that they are selected within mixtures. The ad-
198 dition of the Block column particle in Figure 4-f reduces the weights close to 1 of the Rosette 3-bullet particle in
199 agreement with Figure 3-d. The Rosette 3-bullet being the SSP generating the strongest scattering after the Block

200 Column among the four selected particles, it appears that some weather scenes required increased diffusion in RTTOV-
201 SCATT to match observed Bts. Indeed, the Block Column particle weights which are different from 0 are all close to
202 one. This indicates that this particle is not used within mixtures but individually for weather scenes that cannot be
203 well simulated with other particles. Indeed, some SSPs do not allow model simulations to reach sufficiently low Bts,
204 which can be achieved using the Block column particle.

205 It could also come from a lack of hydrometeor contents within the model forecast and therefore compensating for a
206 model bias. A related bias on surface precipitation of the AROME-Antilles model has been brought out in Faure et
207 al. 2020[27] thanks to an independent evaluation of precipitation forecast with respect to satellite rainfall products.
208 Further diagnostics, like cross-validations with other instruments, would then need to be performed to understand if
209 this behaviour is the result of model biases or observation operator biases.

210 It also appears in Figure 4 that the weights are increasingly close to 0 or 1 as the P_{SI} index increases which means
211 that the mixtures are gradually decreasing as the weather situation leads to more scattering.

212 The means of the distributions give information about the percentage of use of each particle as function of the P_{SI}
213 index. We observe that the means are not changing a lot across the range of scattering indices for both experiments,
214 except for the Sector snowflake particle : the mean slightly decreases with the increase of the latter for the two
215 experiments. Nonetheless, no single SSP is clearly predominant over the others according to the P_{SI} intervals. Thus,
216 it seems difficult here to choose a preferred SSP according to the P_{SI} index.

217 In addition, we computed the percentage of usage for each particle within the 1D-Bayesian inversion for the 2-month
218 period for both experiments. It appeared that for the 3SSP experiment the most used particle was the Rosette 3-
219 bullet one with a percentage of 43%. The other percentages are equal to 33% and 24% respectively for the Rosette
220 6-bullet and the Sector snowflake particles. The results for the 4SSP experiment are 25%, 29%, 27% and 19%, re-
221 spectively for the Block column, the Rosette 3-bullet, the Rosette 6-bullet and the Sector snowflake particles. In
222 both experiments, the Rosette 3-bullet is therefore selected more frequently than the others and the Block column
223 ends up being used a significant number of times despite being an outlier particle in the sense of Figure 1 distributions.

224
225 Table 1 reveals that the percentage of mixtures is reduced as the P_{SI} index increasing, in agreement with our previous
226 findings. If we look at the 3SSP results, from 78% for P_{SI} index between 5 K and 10 K, the fraction of mixture
227 decreases to 26% for a P_{SI} index above 25 K. We then examined this behaviour with the two other predictors. A
228 similar behaviour is noticed for the other predictors. From 55% of mixtures for surface precipitation between 0.1
229 mm/hr and 2 mm/hr, this fraction is reduced to 24% for surface precipitation above 25 mm/hr. Similarly from 68%
230 for an ice water path amounts ranging between 0.1 kg/m² and 0.5 kg/m², the percentage decreases to 17% for ice
231 water path amounts above 2 kg/m². The percentages with the 4SSP experiment lead to similar conclusions : the use
232 of a single particle increases along with the surface precipitation, the P_{SI} index and the ice water path. This indicates
233 that highly scattering weather scenes are frequently well simulated by a single SSP in the Bayesian inversion. This is
234 in contrast to weakly scattering scenes which can be often simulated accurately by a mixture of particles. This can
235 be explained by the simulated Bt distributions of the SSPs which have their largest differences for highly scattering
236 weather scenes. This leads to important differences in simulated of the neighbourhood Bt profiles. Thus, for these
237 cases, only one SSP (associated with a lot of scattering) is useful to reach values close to observations. On the contrary,
238 for an observation that contains few hydrometeors, the method simulates Bts that differ only slightly among various
239 SSPs. Thus, the inversion can assign an almost identical weight to each SSP and create a mixture.

240 Overall, both results reveal that the mixtures are as frequently chosen as a single particle within the Bayesian inversion.
241 However, they are dominant for low scattering situations whereas a single particle is more frequently selected in
242 highest scattering situations.

243 5 | CONCLUSIONS

244 This study examined the interest of using several radiative properties for solid hydrometeors within a 1D Bayesian
245 inversion of microwave brightness temperatures. Our motivations came from the strong sensitivity of radiative trans-
246 fer simulations at high frequencies to scattering properties of hydrometeors and from the need to overcome some
247 limitations of the choice of a single 'optimal' particle. For that purpose, atmospheric profiles from the convective
248 scale model AROME Antilles and brightness temperatures between 18 and 183 GHz from the microwave radiometer
249 GMI onboard the GPM-Core satellite have been considered. Results have been examined for a case study (hurricane
250 Maria) and over a two month period (September-October 2019). A revised Bayesian inversion has been assessed us-
251 ing a number of SSPs from the Liu (2008 [15]) database with two experiments: one with 3 SSPs leading to simulated
252 Bt distributions with the AROME model close to the GMI one ; and another one with the 3 previous SSPs plus one
253 with a distribution characterized by a larger tail of low Bts compared to observations. As the inversion can choose
254 among a mixture of SSPs to perform the inversion, we examined carefully how the weights for each particle were
255 distributed.

256 The case study highlighted the usefulness of each SSP in the inversions by excluding none of them. In the second
257 experiment, adding the Block Column SSP on top of the 3 SSPs, showed the usefulness of a SSP generating further
258 scattering leading to a larger number of successful inversions. Indeed, this particle was selected for some inversions
259 at the expense of the Rosette 3-bullet.

260 The statistical study further highlighted the previous findings and displayed promising results. As a finding, a larger
261 number of SSPs within the 1D Bayesian inversion increases the number of successful inversions. Next, using the
262 scattering index P_{SI} to categorize the 2-month results, it was found that mixtures are more frequent for low values.
263 On the contrary, for high P_{SI} index, the method tends to select only one particle. In agreement with P_{SI} index,
264 mixtures are preferred for low surface precipitation and low ice water path contents. Individual particles are favored
265 for high values of surface precipitation and ice water path contents. One perspective of this study would be to
266 categorize the inversion results with different parameters. The convective/stratiform parameters and environmental
267 fields being closely related to microphysical properties, they could help to gain in understanding on the SSP choices.
268 This study demonstrated the feasibility and the interest of using several SSPs for the inversion of cloudy microwave
269 radiances at larger frequencies. Compared to the other methods (Baran and Labonnote 2007[20], Kulie et al. 2010[5],
270 Ringerud et al. 2019[6]), this approach has the advantage to dynamically build a mixture of SSP by using the microwave
271 observations as an a priori information. Experiments using this revised 1D Bayesian method will be undertaken to
272 assimilate cloudy and/or precipitating Bts from the GMI microwave radiometer.

273 Acknowledgements

274 This research is funded by Météo-France and Région Occitanie (PhD grant for Marylis Barreyat). The authors acknowl-
275 edge the Centre National d'Études Spatiales (CNES) for the financial support of this scientific research activity part
276 of «the Infrarouge, Micro-Ondes et Transfert radiatif ensembliste pour la prévision des Extrêmes de Précipitations»
277 (IMOTEP) project. The authors also want to thank Christophe Accadia and Yves Bouteloup for useful advices.

278 references

- 279 [1] Geer A, Baordo F. Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmo-*
280 *spheric Measurement Techniques* 2014;7(6):1839–1860.
- 281 [2] Geer A, Baordo F, Bormann N, Chambon P, English S, Kazumori M, et al. The growing impact of satellite observations sen-
282 sitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society* 2017;143(709):3189–
283 3206.
- 284 [3] Kidd C, Matsui T, Chern J, Mohr K, Kummerow C, Randel D. Global precipitation estimates from cross-track passive
285 microwave observations using a physically based retrieval scheme. *Journal of Hydrometeorology* 2016;17(1):383–400.
- 286 [4] Kidd C, Matsui T, Ringerud S. Precipitation Retrievals from Passive Microwave Cross-Track Sensors: The Precipitation
287 Retrieval and Profiling Scheme. *Remote Sensing* 2021;13(5):947.
- 288 [5] Kulie MS, Bennartz R, Greenwald TJ, Chen Y, Weng F. Uncertainties in microwave properties of frozen precipitation:
289 Implications for remote sensing and data assimilation. *Journal of the Atmospheric Sciences* 2010;67(11):3471–3487.
- 290 [6] Ringerud S, Kulie MS, Randel DL, Skofronick-Jackson GM, Kummerow CD. Effects of ice particle representation on
291 passive microwave precipitation retrieval in a Bayesian scheme. *IEEE Transactions on Geoscience and Remote Sensing*
292 2019;57(6):3619–3632.
- 293 [7] Barreyat M, Chambon P, Mahfouf JF, Faure G, Ikuta Y. A 1D Bayesian Inversion Applied to GPM Microwave Imager
294 Observations: Sensitivity Studies. *Journal of the Meteorological Society of Japan Ser II* 2021;.
- 295 [8] Haddad Z, Steward J, Tseng HC, Vukicevic T, Chen SH, Hristova-Veleva S. A data assimilation technique to account
296 for the nonlinear dependence of scattering microwave observations of precipitation. *Journal of Geophysical Research:*
297 *Atmospheres* 2015;120(11):5548–5563.
- 298 [9] Guerbette J, Mahfouf JF, Plu M. Towards the assimilation of all-sky microwave radiances from the SAPHIR humid-
299 ity sounder in a limited area NWP model over tropical regions. *Tellus A: Dynamic Meteorology and Oceanography*
300 2016;68(1):28620.
- 301 [10] Mangla R, Indu J, Chambon P, Mahfouf JF. First steps towards an all-sky assimilation framework for tropical cyclone event
302 over Bay of Bengal region: Evaluation and assessment of GMI radiances. *Atmospheric Research* 2021;257:105564.
- 303 [11] Geer AJ. Physical characteristics of frozen hydrometeors inferred with parameter estimation. *Atmospheric Measurement*
304 *Techniques* 2021;14(8):5369–5395. <https://amt.copernicus.org/articles/14/5369/2021/>.
- 305 [12] Nakaya U. Formation of snow crystals, vol. 3. Snow, Ice and Permafrost Research Establishment, Corps of Engineers,
306 US Army; 1954.
- 307 [13] Hallett J, Mason BJ. The influence of temperature and supersaturation on the habit of ice crystals grown from the vapour.
308 *Proceedings of the Royal Society of London Series A Mathematical and Physical Sciences* 1958;247(1251):440–453.
- 309 [14] Magono C, Lee CW. Meteorological classification of natural snow crystals. *Journal of the Faculty of Science, Hokkaido*
310 *University Series 7, Geophysics* 1966;2(4):321–335.
- 311 [15] Liu G. A database of microwave single-scattering properties for nonspherical ice particles. *Bulletin of the American*
312 *Meteorological Society* 2008;89(10):1563–1570.
- 313 [16] Ding J, Bi L, Yang P, Kattawar GW, Weng F, Liu Q, et al. Single-scattering properties of ice particles in the microwave
314 regime: Temperature effect on the ice refractive index with implications in remote sensing. *Journal of Quantitative*
315 *Spectroscopy and Radiative Transfer* 2017;190:26–37.
- 316 [17] Eriksson P, Ekelund R, Mendrok J, Brath M, Lemke O, Buehler SA. A general database of hydrometeor single scattering
317 properties at microwave and sub-millimetre wavelengths. *Earth System Science Data* 2018;10(3):1301–1326.

- 318 [18] Brath M, Ekelund R, Eriksson P, Lemke O, Buehler SA. Microwave and submillimeter wave scattering of oriented ice
319 particles. *Atmospheric Measurement Techniques* 2020;13(5):2309–2333.
- 320 [19] Kneifel S, Neto JD, Ori D, Moisseev D, Tyynelä J, Adams IS, et al. Summer Snowfall Workshop: Scattering Properties of
321 Realistic Frozen Hydrometeors from Simulations and Observations, as well as Defining a New Standard for Scattering
322 Databases. *Bulletin of the American Meteorological Society* 2018;99(3):ES55 – ES58. [https://journals.ametsoc.org/
323 view/journals/bams/99/3/bams-d-17-0208.1.xml](https://journals.ametsoc.org/view/journals/bams/99/3/bams-d-17-0208.1.xml).
- 324 [20] Baran AJ, Labonnote LC. A self-consistent scattering model for cirrus. I: The solar region. *Quarterly Journal of the
325 Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*
326 2007;133(629):1899–1912.
- 327 [21] Hong G. Parameterization of scattering and absorption properties of nonspherical ice crystals at microwave frequencies.
328 *Journal of Geophysical Research: Atmospheres* 2007;112(D11).
- 329 [22] Kummerow C, Hong Y, Olson W, Yang S, Adler R, McCollum J, et al. The evolution of the Goddard Profiling Algorithm
330 (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology* 2001;40(11):1801–
331 1820.
- 332 [23] Randel DL, Kummerow CD, Ringerud S. The Goddard Profiling (GPROF) precipitation retrieval algorithm. In: *Satellite
333 precipitation measurement* Springer; 2020.p. 141–152.
- 334 [24] Caumont O, Ducrocq V, Wattrelot É, Jaubert G, Pradier-Vabre S. 1D+ 3DVar assimilation of radar reflectivity data: A
335 proof of concept. *Tellus A: Dynamic Meteorology and Oceanography* 2010;62(2):173–187.
- 336 [25] Wattrelot E, Caumont O, Mahfouf JF. Operational implementation of the 1D+ 3D-Var assimilation method of radar
337 reflectivity data in the AROME model. *Monthly Weather Review* 2014;142(5):1852–1873.
- 338 [26] Duruisseau F, Chambon P, Wattrelot E, Barreyat M, Mahfouf JF. Assimilating cloudy and rainy microwave observations
339 from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Quarterly Journal of the Royal Meteorological
340 Society* 2019;145(719):620–641.
- 341 [27] Faure G, Chambon P, Brousseau P. Operational Implementation of the AROME Model in the Tropics: Multiscale Validat-
342 ion of Rainfall Forecasts. *Weather and Forecasting* 2020;35(2):691–710.
- 343 [28] Ricard D, Lac C, Riette S, Legrand R, Mary A. Kinetic energy spectra characteristics of two convection-permitting limited-
344 area models AROME and Meso-NH. *Quarterly Journal of the Royal Meteorological Society* 2013;139(674):1327–1341.
- 345 [29] Saunders R, Hocking J, Turner E, Rayer P, Rundle D, Brunel P, et al. An update on the RTTOV fast radiative transfer
346 model (currently at version 12). *Geoscientific Model Development* 2018;11(7):2717–2737.
- 347 [30] Marshall JS, Palmer WMK. THE DISTRIBUTION OF RAINDROPS WITH SIZE. *Journal of Atmospheric Sciences*
348 1948;5(4):165 – 166. [https://journals.ametsoc.org/view/journals/atsc/5/4/1520-0469_1948_005_0165_tdorws_
349 2_0_co_2.xml](https://journals.ametsoc.org/view/journals/atsc/5/4/1520-0469_1948_005_0165_tdorws_2_0_co_2.xml).
- 350 [31] Field PR, Heymsfield AJ, Bansemer A. Snow size distribution parameterization for midlatitude and tropical ice clouds.
351 *Journal of the Atmospheric Sciences* 2007;64(12):4346–4365.
- 352 [32] Petty GW, Huang W. The modified gamma size distribution applied to inhomogeneous and nonspherical particles: Key
353 relationships and conversions. *Journal of the Atmospheric Sciences* 2011;68(7):1460–1473.
- 354 [33] Grody NC. Classification of snow cover and precipitation using the Special Sensor Microwave Imager. *Journal of Geo-
355 physical Research: Atmospheres* 1991;96(D4):7423–7435.
- 356 [34] Wilheit T, Kummerow CD, Ferraro R. NASDARainfall algorithms for AMSR-E. *IEEE Transactions on Geoscience and
357 Remote Sensing* 2003;41(2):204–214.

-
- 358 [35] Thrun MC, Gehlert T, Ultsch A. Analyzing the fine structure of distributions. *PloS one* 2020;15(10):e0238835.
- 359 [36] Barton N, Metzger EJ, Reynolds CA, Ruston B, Rowley C, Smedstad OM, et al. The Navy's Earth System Prediction Capa-
360 bility: A new global coupled atmosphere-ocean-sea ice prediction system designed for daily to subseasonal forecasting.
361 *Earth and Space science* 2021;8(4):e2020EA001199.
- 362 [37] Riccardi U, Tamaro U, Capuano P. Tropospheric Delay in the Neapolitan and Vesuvius Areas (Italy) by Means of a Dense
363 GPS Array: A Contribution for Weather Forecasting and Climate Monitoring. *Atmosphere* 2021;12(9):1225.

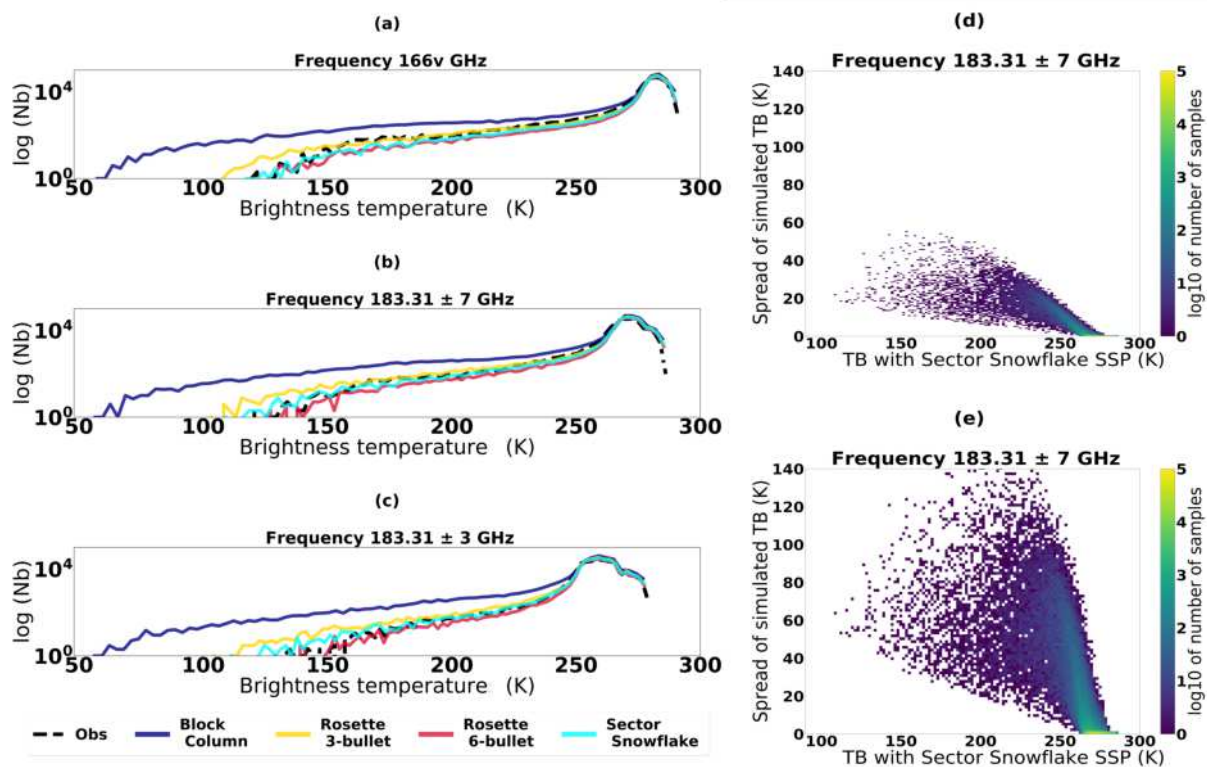


FIGURE 1 a, b, c - FG simulated Bt distributions respectively at frequencies 166v, 183.31 ± 7 and 183.31 ± 3 GHz for a 2-month period. d, e - Scatter plots of the spread between the minimum and the maximum simulated Bts respectively for 3 and 4 SSPs at 183.31 ± 7 GHz frequency for a 2-month period.

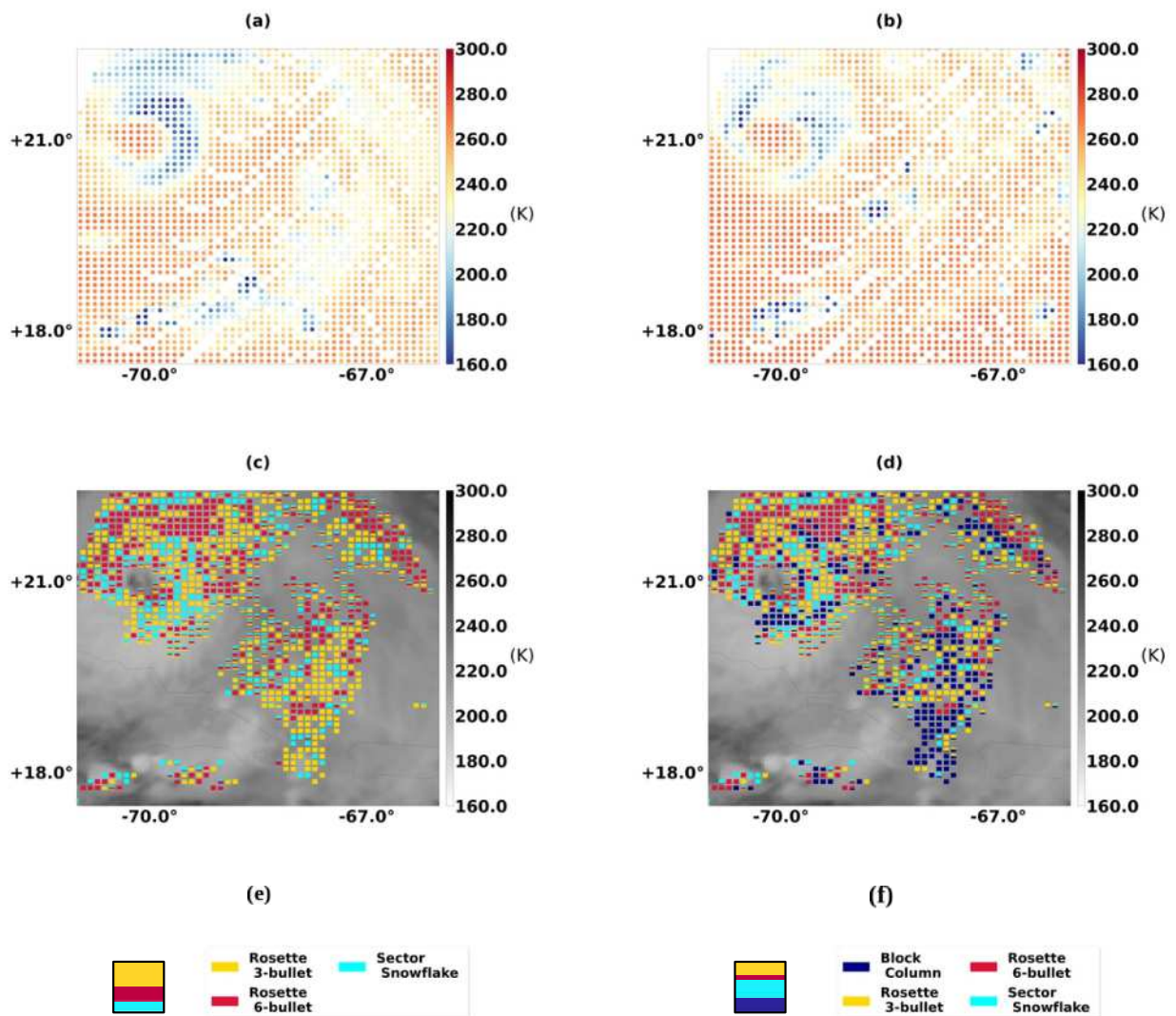


FIGURE 2 a - GMI observations of the Maria cyclone reaching the West Indies on 22 September 2017 for the 183.31 ± 7 GHz frequency. b - FG simulated Bt simulations of the cyclone for the same frequency with the Sector snowflake SSP. c, d - Normalized inversion weights for profiles characterized as cloudy and rainy by the criteria defined section 3-b with respectively the 3 SSPs (Rosette 3-bullet in yellow, Rosette 6-bullet in red and Sector snowflake in cyan) and the 4 SSPs (the 3 SSPs plus block column in dark blue) with observed infrared imagery at frequency $10.3 \mu\text{m}$ as background. e and f - Example of normalized inversion weights respectively for the 3 SSPs and 4 SSPs experiments.

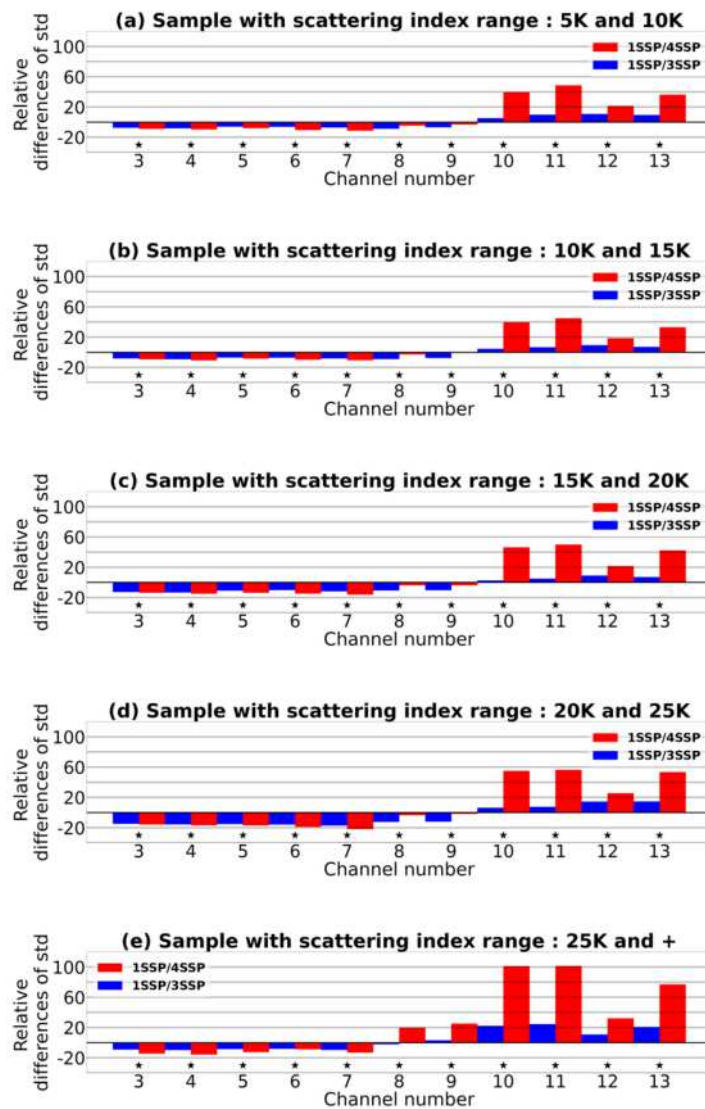


FIGURE 3 Relative differences of standard deviations of the analyzed Bts minus observations between the 3SSP and 1SSP (blue) or between the 4SSP and 1SSP (red) experiments over a 2-month period for each GMI channel used. Stars indicate that the differences of standard deviations are significant at the 95% level.

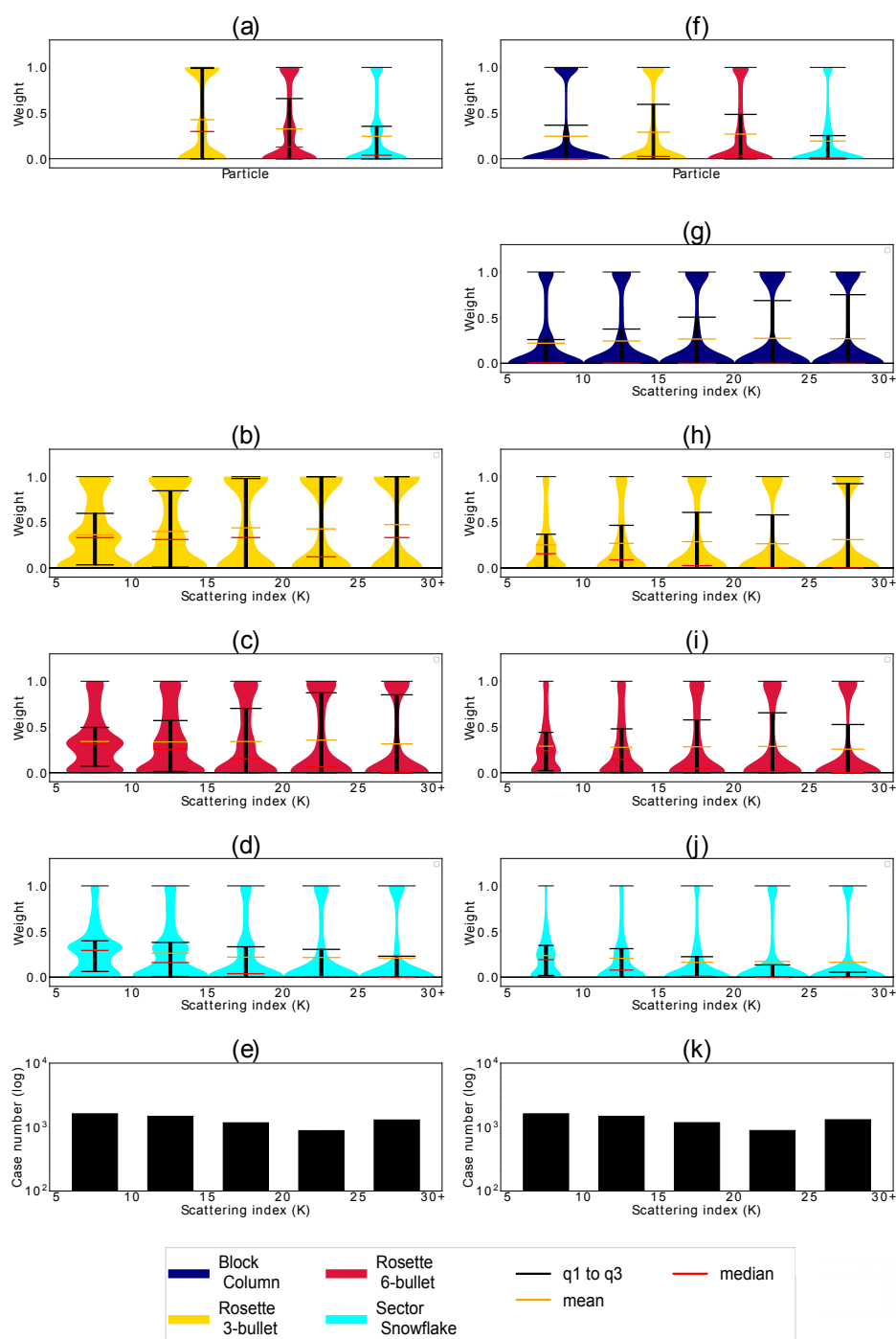


FIGURE 4 a - Violinplots of the cloudy observation distributions, according to criteria defined in section 3-b, for each SSP over the 2-month period with the 3 SSPs experiment. f - Same with the 4 SSPs experiment. b, c, d - Violinplots of the cloudy observation distributions over the 2-month period with respect to the P_{SI} index of the 3 SSPs experiment respectively for Rosette 3-bullet, Rosette 6-bullet and Sector snowflake. g, h, i, j - Same with the 4 SSPs experiment respectively for Block column, Rosette 3-bullet, Rosette 6-bullet and Sector snowflake. e - Histogram of the number of cloudy observations per bins of P_{SI} index for the 3 SSPs experiment. k - Same with the 4 SSPs experiment.

TABLE 1 Percentage of use between mixtures and single SSP within the 1D-Bayesian inversion over a 2-month period as function of 3 predictors respectively surface precipitation, P_{SI} scattering index and ice water path for the 3SSP (4SSP) experiments. The use of a single SSP is defined when one of the SSPs has a normalized weight larger than 0.95.

Surface precipitation				
Bin (mm/hr)	[0.1;2]	[2;10]	[10;25]	[25;+]
Mixture (%)	55 (55)	38 (37)	24 (25)	24 (22)
Single SSP (%)	45 (45)	62 (63)	76 (75)	76 (78)

P_{SI}					
Bin (K)	[5;10]	[10;15]	[15;20]	[20;25]	[25;+]
Mixture (%)	78 (76)	68 (67)	51 (51)	37 (38)	26 (27)
Single SSP (%)	22 (24)	32 (33)	49 (49)	63 (62)	74 (73)

Ice water path				
Bin (kg/m^2)	[0.1;0.5]	[0.5;1]	[1;2]	[2;+]
Mixture (%)	68 (67)	42 (43)	22 (22)	17 (18)
Single SSP (%)	32 (33)	58 (57)	78 (78)	83 (82)

6.3 Résultats complémentaires

Dans cette section, nous allons approfondir l'étude présentée dans l'article qui précède. Pour cela nous allons comparer la catégorisation des résultats des expériences de l'inversion Bayésienne modifiée avec les différentes variables présentées en introduction à ce chapitre (indice caractérisant la diffusion des scènes météorologiques plus deux autres variables atmosphériques provenant de produits GPROF). Dans cette section, nous allons également décrire les résultats d'expériences d'assimilation ARPEGE utilisant cette nouvelle méthode d'inversion.

6.3.1 Étude statistique avec les prédicteurs GPROF

Nous examinons ici les résultats présentés dans l'article à l'aide d'autres prédicteurs provenant des produits GPROF : les précipitations de surface et le contenu intégré en glace. La figure 6.1 illustre les distributions des poids pour chaque SSP de la même manière que la figure 3 de l'article précédent. Les résultats montrent un comportement identique à celui obtenu avec l'indice de diffusion. En effet, pour des précipitations faibles, on observe avec les deux expériences 3SSP et 4SSP que la majorité des poids sont soit proches de 0, soit proches de 1 pour chaque forme de particule. Ceci signifie que la plupart des inversions sont réalisées à l'aide d'une seule SSP plutôt qu'à l'aide de mixtures. On peut cependant remarquer sur les figures 6.1e-f-g-h que les distributions de poids de l'expérience 4SSP s'élargissent légèrement aux alentours de 0.25. Ce comportement est cependant davantage marqué avec l'expérience 3SSP où les poids pour des précipitations faibles montrent un pic à 0.33 (figures 6.1-a-b-c). Il apparaît également sur la figure 6.1 que les poids sont de plus en plus proches de 0 ou 1 au fur et à mesure que les précipitations de surface augmentent, ce qui signifie que l'utilisation des mixtures de SSPs diminuent progressivement au fur et à mesure que la situation météorologique devient fortement précipitante. Néanmoins, ici non plus nous ne trouvons aucune SSP réellement dominante sur les autres selon les intervalles de précipitations de surface.

Si on s'intéresse maintenant aux résultats avec le prédicteur de contenu intégré en glace (figure 6.2) on retrouve une nouvelle fois le même comportement. Pour des contenus intégrés en glace faibles, de nombreux poids sont distribués aux alentours de 0 quelle que soit la SSP. On trouve de nouveau des valeurs proches de 0.25 confirmant l'utilisation de mixtures de SSPs pour des scènes caractérisées par peu de diffusion. Plus les contenus intégrés en glace augmentent, plus les poids sont distribués autour de 0 et 1, confirmant la prédominance de l'utilisation d'une seule SSP pour des scènes caractérisées par une diffusion importante. Finalement, aucune SSP n'est dominante sur les autres selon les intervalles de contenus intégrés en glace.

Les résultats des catégorisations réalisées avec les deux prédicteurs issus des produits GPROF ont montré un comportement identique à celui obtenu avec l'indice de diffusion calculé à partir des températures de brillance observées de GMI. Aucun de ces prédicteurs ne permet de privilégier une SSP par rapport à une autre, ni même une combinaison de SSPs. Aucune préférence de mixtures de particules ne semble être priorisée par rapport à ces trois prédicteurs. Néanmoins d'autres prédicteurs de la scène météorologique pourraient être intéressants à examiner, comme par exemple le type de précipitations (convectives/stratiformes).

6.3.2 Résultats d'expériences d'assimilation avec le modèle ARPEGE

Par la suite nous avons voulu tester l'assimilation des produits inversés avec une approche "multi-particules" et quantifier leur impact sur la qualité des prévisions. Le système de prévision du temps AROME-Antilles ne bénéficiant pas encore de son propre système d'assimilation, nous avons effectué cette étude au sein du modèle de prévision global ARPEGE et de son système d'assimilation 4D-Var décrit dans le chapitre 1. Pour ce faire, une nouvelle période d'étude de deux mois s'étalant entre mai et juin 2021 a été définie afin de bénéficier des dernières évolutions du modèle. Le seuil des inversions réussies a été cette fois-ci fixé avec le seuil opérationnel au sein du modèle ARPEGE qui correspond au nombre de profils du voisinage qui est ici de 49 multiplié par 10^{-12} . Présentons les trois expériences prolongeant l'étude précédente :

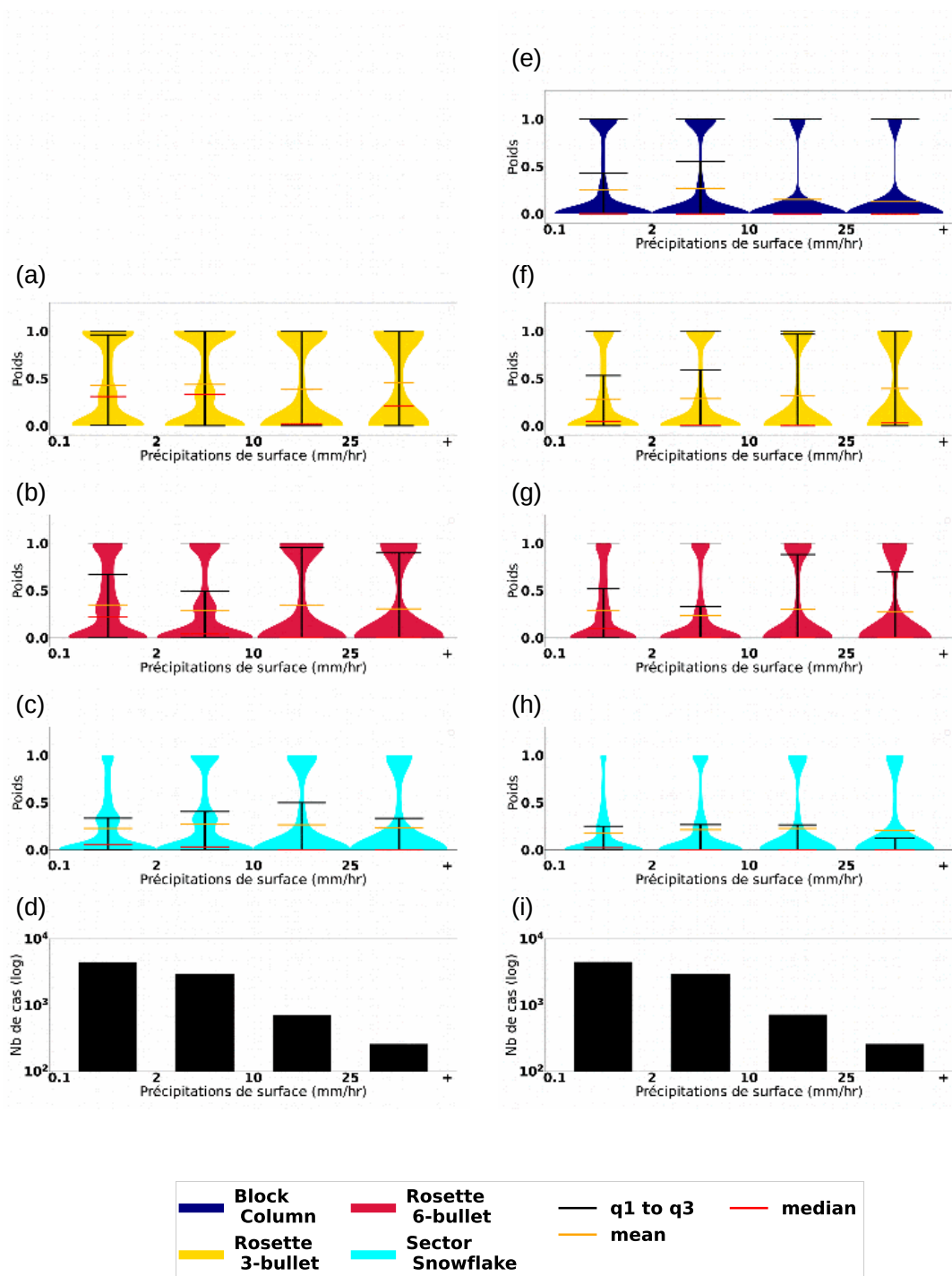


FIGURE 6.1 – a,b c - Diagrammes en violon des distributions des observations nuageuses (définies Section 7.1) en fonction des précipitations de surface GPROF pour chaque SSP de l'expérience 3SSP. e,f,g,h - Idem pour les 4 SSPs de l'expérience 4SSP. d, i - Histogrammes du nombre d'observations nuageuses par intervalle de précipitations de surface respectivement pour les expériences 3SSP et 4SSP.

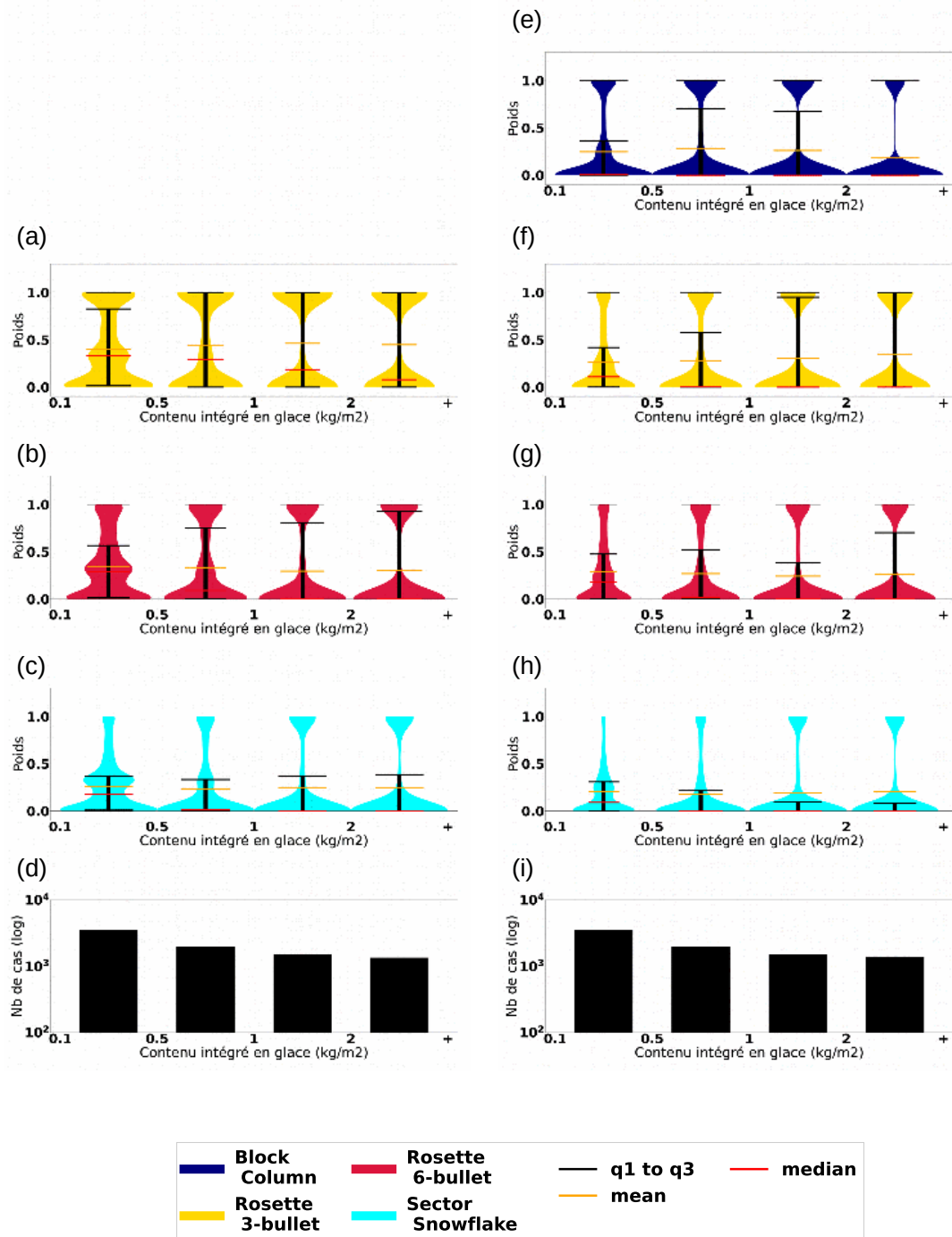


FIGURE 6.2 – a,b,c - Diagrammes en violon des distributions des observations nuageuses (définies Section 7.1) en fonction du contenu intégré en glace GPROF pour chaque SSP de l'expérience 3SSP. e,f,g,h - Idem pour les 4 SSPs de l'expérience 4SSP. d, i - Histogrammes du nombre d'observations nuageuses par intervalle de contenu intégré en glace respectivement pour les expériences 3SSP et 4SSP.

Mesure	Type d'observations
Surface	Stations de surface, bateaux, bouées
	Données terrestres GNSS
	Profileurs de vents
Altitude	Radiosondes
	Mesures d'avions
Satellite	Vecteurs de mouvement atmosphérique de : SEVIRI, ABI, AHI, MODIS, AVHRR
	Vents du diffusiomètre ASCAT
	Données de sondage micro-ondes de : AMSU-A, MHS, SAPHIR, MWHS2, ATMS, SSMI/S, GMI
	Données d'hyperspectraux infrarouges : IASI, CrIS
	Données infrarouges des géostationnaires : Meteosat, GOES, Himawari
	Données de radio occultation radio GNSS

TABLE 6.1 – Observations assimilées au sein du 4D-Var de l'étude

- La première expérience est notée **GMI-BAYRAD 1SSP**. En plus de toutes les autres observations assimilées opérationnellement, cette expérience assimile les observations de GMI en ciel nuageux, à l'aide d'une inversion Bayésienne et en choisissant la SSP *Sector snowflake* pour l'opérateur d'observation. Le choix de cette particule est en accord avec les travaux de Geer et Baordo (2014[1]) avec le modèle IFS du CEPMMT et de Duruisseau et al. (2019[44]) avec le modèle ARPEGE, où cette SSP fournit le meilleur compromis au sein de la base de données de Liu2008[11] pour la simulation de températures de brillance micro-ondes en ciel nuageux au niveau global.

- La deuxième expérience est notée **GMI-BAYRAD 3SSP**. En plus de toutes les autres observations assimilées opérationnellement, cette expérience assimile les observations de GMI en ciel nuageux à l'aide d'une inversion Bayésienne multi-particules. La même configuration de SSPs que l'expérience précédente 3SSP est utilisée, soient les SSPs : *Rosette 3-bullet*, *Rosette 6-bullet* et *Sector snowflake*.

- La troisième expérience est notée **GMI-BAYRAD 4SSP**. En plus de toutes les autres observations assimilées opérationnellement, cette expérience assimile les observations de GMI en ciel nuageux à l'aide d'une inversion Bayésienne avec une somme supplémentaire sur les SSPs. La même configuration de SSPs que l'expérience précédente 4SSP est utilisée, soient les SSPs : *Block column*, *Rosette 3-bullet*, *Rosette 6-bullet* et *Sector snowflake*.

Pour chacune de ces expériences, nous initialisons le modèle ARPEGE avec un 4D-Var incrémental possédant une fenêtre d'assimilation de 6 h. Les observations assimilées en complément des observations utilisées dans les différentes expériences sont disponibles dans le tableau 6.1. Le 4D-Var est cyclé sur la période complète de deux mois, une prévision longue jusqu'à 102 heures d'échéance est effectuée à partir de chaque analyse de 0 h UTC. Des scores de prévisions sont ensuite calculés par rapport à une référence. La référence choisie dans notre étude est l'analyse du modèle du CEPMMT. Le score consiste en un écart-type des différences entre prévision et analyse. Pour les variables vents, humidité, température, géopotentiel, ces écart-types peuvent être calculés à chaque échéance de prévision jusqu'à 102 h et pour différents niveaux de pression. On peut alors calculer des différences relatives des écart-types entre deux expériences. Une différence négative indique alors une réduction de l'erreur de prévision par rapport à la référence.

Le nombre total d'observations assimilées en ciel nuageux et clair pour chaque expérience est résumé tableau 6.2. On voit dans ce tableau que le nombre d'observations assimilées en ciel clair pour le radiomètre GMI est environ 5 fois supérieur aux nombres d'observations assimilées en ciel nuageux à l'aide de la méthode Bayésienne multi-particules. On remarque de plus, que le nombre d'observations assimilées au sein de la méthode Bayésienne augmente avec le nombre de SSP considérées au sein de l'expérience : 17121, 17167 et 17359 respectivement pour les expériences 1SSP, 3SSP et 4SSP. Les différences entre les nombres d'observation assimilées peuvent ici paraître faibles, cependant les nouvelles observations viennent contraindre l'analyse à des endroits où les autres observations ne le permettent pas. La figure 6.3 montre les différences relatives de scores de prévision du vent sur la période de deux mois entre mai et juin 2021. Ces scores sont illustrés ici pour 4 niveaux de pression dans la basse troposphère (entre 1000 et 700 hPa) sur le domaine NORD20 couvrant toutes les longitudes entre 20 degrés de latitude et le pôle

Nord. On remarque premièrement que l'expérience 3SSP ne semble pas montrer ici d'améliorations par rapport à l'expérience GMI-BAYRAD 1SSP. Au contraire, l'expérience 4SSP montre des impacts positifs sur les prévisions de vent dans les basses couches d'ARPEGE. L'impact est statistiquement significatif à 95% pour des échéances allant jusqu'à +36 h de prévision pour les niveaux de pression 700, 925 et 1000 hPa. Une explication possible pour ces améliorations viendrait de l'effet combiné des résultats d'inversion et du nombre d'inversions réalisées plus important dans l'expérience avec 4SSP par rapport à celui avec 3SSP et 1 SSP (tableau 6.2). Les scores ont été calculés pour deux autres domaines géographiques : le domaine des Tropiques entre -20 et 20 degré de latitude et le domaine SUD20 entre -20 degrés de latitude et le pôle Sud. Pour ces autres domaines et les autres niveaux de pression, les différences de scores entre les expériences ne dépassent pas le seuil de significativité à 95 % réalisé à l'aide d'un test de Student. Ceci n'est pas complètement surprenant car nous nous intéressons ici à des raffinements de la représentation de la microphysique dans l'assimilation de données. La projection de ces raffinements sur l'amélioration des prévisions de vents, température et humidité n'est donc pas directe. Nous avons donc complété cette évaluation en analysant la qualité des précipitations prévues.

Expérience	Nb obs assimilées en ciel nuageux	Nb obs assimilées en ciel clair
GMI-BAYRAD 1SSP	17121	90785
GMI-BAYRAD 3SSP	17167	90785
GMI-BAYRAD 4SSP	17359	90367

TABLE 6.2 – Nombre total d'observations du radiomètre GMI assimilées pour les différentes expériences sur la période d'étude de deux mois. Les observations en ciel clair sont assimilées directement au sein du 4D-Var déterministe tandis que les observations en ciel nuageux sont assimilées à l'aide de la méthode Bayésienne.

Des scores de précipitations ont été calculés à différentes échéances. Pour cela, nous nous sommes servis des produits GSMaP (Global Satellite Mapping of Precipitation, Kubota et al. 2007[199]) fournis par la JAXA. Ces produits fournissent des cartes mondiales de précipitations sur une grille $0.1^\circ \times 0.1^\circ$, calculées toutes les heures à partir d'observations de plusieurs instruments satellitaires micro-ondes et infra-rouges. Les produits utilisés dans ce manuscrit correspondent à sa version 04 (algorithme version 7, Kubota et al. 2020[200]).

Les prévisions d'ARPEGE ont été colocalisées avec les produits GSMaP sur une grille $0.5^\circ \times 0.5^\circ$ avec une délimitation des latitudes entre 60°S et 60°N . Pour évaluer objectivement les prévisions de précipitations parmi les différentes expériences, le score à voisinage nommé FSS (Fraction Skill Score), introduit par Roberts et Lean (2008[201]), a été utilisé. Ce score réalise une correspondance entre la fréquence des précipitations prévues et celles observées avec une certaine tolérance spatiale. Une correspondance parfaite entre les observations et les champs du modèle correspond à un score FSS de 1, tandis que des données en totale contradiction correspondent à un score FSS de 0. Nous décrivons maintenant le formalisme permettant d'obtenir un score FSS.

Pour calculer ce score il faut tout d'abord définir un seuil de précipitations q . Ensuite on calcule sur une grille le champ binaire I_o relatif aux observations (resp. le champ M_o relatif au modèle), qui vaut 1 aux points où les observations (resp. les champs modèles) dépassent le seuil de pluie q , et 0 ailleurs. Ensuite on calcule pour chaque point du domaine une fraction correspondant au pourcentage de profils ayant dépassés le seuil de précipitations dans un voisinage n , i.e. les profils dont les champs binaires valent 1. Ces champs de scores, respectivement $O_{(n)}$ et $M_{(n)}$ pour les observations et le modèle, s'expriment tels que :

$$O_{(n)(i,j)} = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n I_o \left[i+k-1 - \frac{(n-1)}{2}, j+l-1 - \frac{(n-1)}{2} \right]$$

$$M_{(n)(i,j)} = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n M_o \left[i+k-1 - \frac{(n-1)}{2}, j+l-1 - \frac{(n-1)}{2} \right]$$

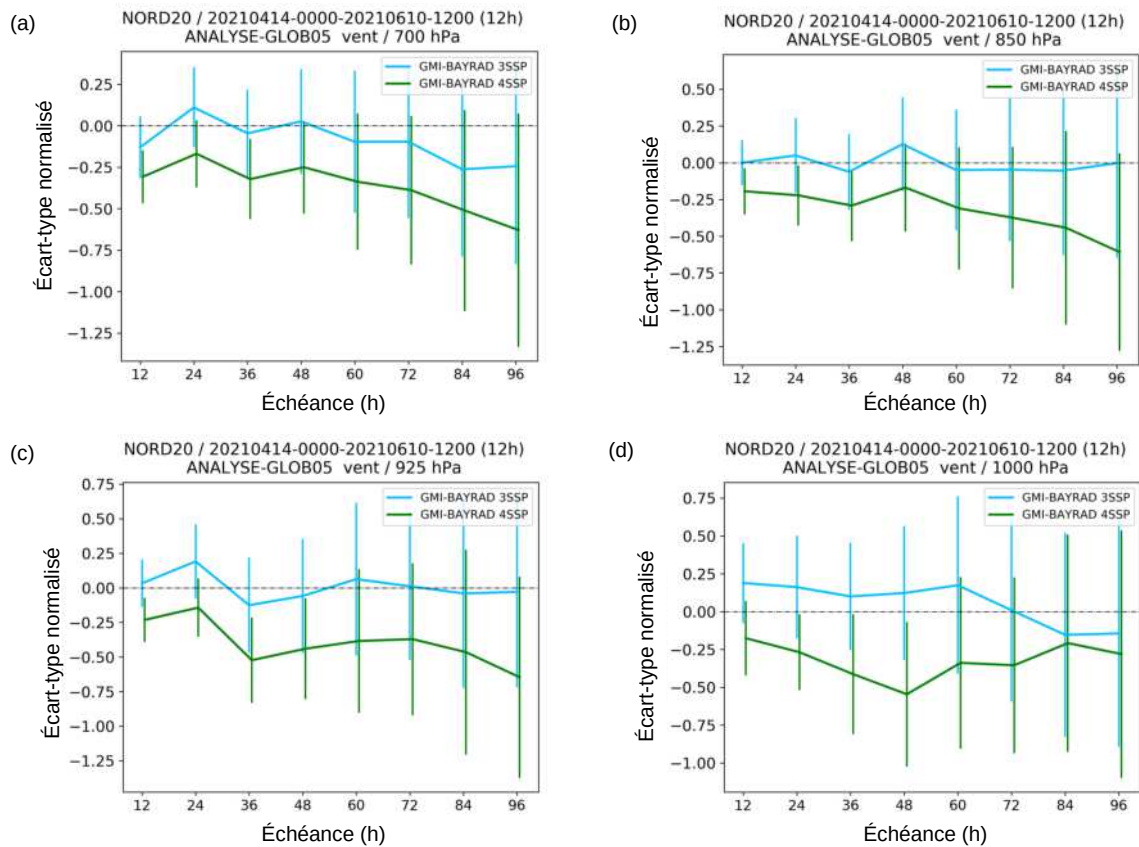


FIGURE 6.3 – Différences relatives de scores de prévisions entre trois expériences, pour le vent, le domaine géographique NORD20 et pour 4 niveaux pressions : 700 hPa, 850 hPa, 925 hPa, 1000 hPa. Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du CEPMMT, pour les échéances d'ARPEGE jusqu'à +102h sur une période de 2 mois. L'expérience de référence, notée 1SSP, assimile les données de GMI en ciel nuageux avec une seule SSP pour simuler les effets de la présence de neige dans l'atmosphère. Les deux autres expériences, nommées 3SSP et 4SSP, assimilent les données de GMI en ciel nuageux avec respectivement 3 ou 4 SSPs pour la neige.

Un exemple de calcul des champs $O_{(n)(i,j)}$ et $M_{(n)(i,j)}$ est représenté figure 6.4. Finalement le score FSS est calculé à partir de tous les profils du domaine considérés :

$$FSS_{(n)} = 1 - \frac{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)(i,j)} - M_{(n)(i,j)}]^2}{\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [O_{(n)(i,j)}^2 - M_{(n)(i,j)}^2]}$$

Avec N_x et N_y respectivement les nombres de points suivant un axe Est-Ouest et un axe Nord-Sud, pour un domaine géographique.

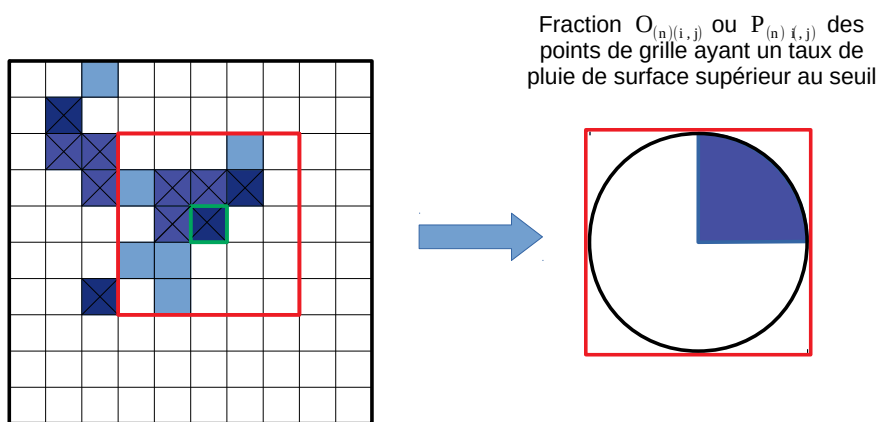


FIGURE 6.4 – Exemple de calcul des fractions $O_{(n)(i,j)}$ ou $M_{(n)(i,j)}$ pour un point (i,j) représenté par le carré vert. Pour un voisinage n donné, représenté par un carré rouge ici, on calcule la fraction de profils dépassant le seuil q de pluie. Le diagramme en secteur à droite représente cette fraction.

Source: T. Weusthoff COSMO (COnsortium for Small-scale MOdelling) General Meeting (2019)

Dans cette étude, nous avons calculé les précipitations de surface cumulées sur 24 h pour les produits GSMaP et pour le modèle ARPEGE sur la période deux mois entre avril et mai 2021. Les cumuls de pluie 24 h modélisés peuvent correspondre à plusieurs échéances maximales : +24 h, +48 h, +72 h et +96 h. Les scores FSS ont ensuite été calculés pour plusieurs seuils q de précipitations de surface allant de 0.1 à 50 mm de pluie par jour, et pour plusieurs tailles de voisinage allant de 25 km jusqu'à 200 km sur les Tropiques. Les résultats sont illustrés sur la figure 6.5 pour deux seuils d'intensité de précipitations (3 mm par jour et 40 mm par jour) et un voisinage de 200 km. La figure 6.5-a et b compare les résultats entre les expériences GMI-BAYRAD 1SSP et 3SSP. La figure 6.5-a montre le score FSS entre 24 h et 96 h d'échéance et on remarque que pour des précipitations de surface de plus de 3 mm par jour, les valeurs pour les deux expériences à 24 h d'échéance atteignent presque 0.8 et décroissent légèrement jusqu'à 0.7 en fin d'échéance. Toutefois, ces bons scores FSS sur ce seuil de précipitation ne varient pas entre les expériences GMI-BAYRAD 1SSP et 3SSP. Pour des précipitations dont le seuil est supérieur à 40 mm par jour, les scores FSS des deux expériences sont plus faibles, oscillant entre 0.45 à 24 h d'échéance et 0.2 à 96 h d'échéance. On remarque cependant une légère différence entre les courbes des deux expériences à 24 h d'échéance et à 96 h d'échéance, en faveur de l'expérience GMI-BAYRAD 3SSP. La figure 6.5-b présente les différences normalisées des scores de FSS entre les expériences GMI-BAYRAD 1SSP et 3SSP pour les deux seuils de précipitations. On remarque tout d'abord des scores neutres à toutes les échéances pour le seuil de pluie à 3 mm par jour. Au contraire, pour les scores associés au seuil de pluie à 40 mm par jour, on note une amélioration significative de l'expérience GMI-BAYRAD 3SSP par rapport à l'expérience GMI-BAYRAD 1SSP à 24 h d'échéance, et des résultats neutres aux autres échéances. Ainsi l'utilisation

d'une inversion Bayésienne traitant 3 SSPs pour l'assimilation des températures de brillance micro-ondes en ciel nuageux semble améliorer la prévision des précipitations à 24 h d'échéance sur la période d'étude concernée. Une période d'étude plus longue devra être étudiée afin de confirmer les résultats.

Sur les figures 6.5-c et d les mêmes comportements peuvent être notés entre les expériences 1SSP et 4SSP. Des similarités apparaissent sur les deux graphiques, avec néanmoins une différence moins marquée entre les deux expériences pour un seuil de pluie de 40 mm par jour à 96 h d'échéance. On retrouve également une amélioration significative des prévisions de fortes précipitations à 24 h d'échéance pour l'expérience GMI-BAYRAD 4SSP par rapport à 1SSP. Cette amélioration du FSS est, comme pour l'expérience 3SSP, de l'ordre de 1 à 2 %. Les différences obtenues sur les scores de précipitations, qui pointent légèrement l'expérience à 3SSP comme amenant à de meilleurs scores comparé à celle à 4SSP, pourraient éventuellement être le résultat de l'assimilation de quantités en humidité relative moins fortes lors de l'introduction de la particule *Block column*. Cela pourrait avoir comme effet de ne pas engendrer assez de précipitations, toutefois les différences dans les résultats obtenus ici ne sont pas significatives. Il semble donc ici prématuré de tirer des conclusions.

6.4 Conclusions

Dans ce chapitre, nous nous sommes intéressés à la première partie de notre problématique. Le but était de déterminer si les observations avaient un contenu en information suffisant pour définir une combinaison optimale de SSPs dans le cadre de l'assimilation des observations micro-ondes nuageuses.

Pour cela nous avons modifié la première étape de la méthode 1D-Bay+3D/4D-Var opérationnelle à Météo-France dans le modèle à échelle convective AROME pour l'assimilation des radars sols (Caumont et al. 2010[19], Wattrelot et al. (2014)). Grâce à cette modification originale, la méthode est maintenant capable d'inverser un profil d'humidité relative à l'aide de plusieurs SSPs pour la modélisation des températures de brillance micro-ondes en ciel nuageux. Nous avons choisi d'étudier cette méthode à l'aide des observations du radiomètre GMI et successivement des modèles AROME-Antilles puis ARPEGE, avec lequel des scores de prévisions ont été calculés. Les températures de brillance des profils atmosphériques des ébauches ont ici été simulées grâce au code de transfert radiatif pour les micro-ondes RTTOV-SCATT ainsi que 4 SSPs de la table Liu2008.

Deux expériences ont été réalisées en utilisant les prévisions 1 h du modèle AROME-Antilles sur une période de deux mois : (i) une expérience réalisant des inversions avec 3 SSPs dont les distributions en températures de brillance sont proches de celle observée et (ii) une expérience réalisant des inversions avec ces trois mêmes SSPs plus une autre SSP capable de générer davantage de diffusion. L'utilité de chaque SSP a été mise en évidence : aucune d'entre elles n'a été exclue des inversions. L'ajout de la particule *Block column* a montré que celle-ci était sélectionnée pour certaines inversions au détriment de la particule *Rosette 3-bullet*. Les résultats ont ensuite été catégorisés à l'aide de trois prédicteurs : un indice de diffusion calculé à l'aide des observations de GMI, ainsi qu'avec les précipitations de surface et le contenu intégré en glace issus des produits GPROF. Grâce à ces catégorisations, il a été constaté que les combinaisons de SSPs étaient sélectionnées pour des faibles valeurs de ces prédicteurs et de ce fait par les scènes météorologiques peu diffusantes. Au contraire, pour des scènes météorologique très diffusantes, et des quantités conséquentes en précipitations de surface et en contenu intégré en glace, la méthode tend à n'attribuer de poids significatif qu'à une seule SSP. Néanmoins, il n'a pas été possible de déterminer un facteur météorologique précis permettant de mieux comprendre pourquoi l'inversion choisit telle particule au détriment d'une autre. Notre analyse est restée mono-variée et il est possible qu'une analyse multi-variée et multi-paramètres puisse permettre d'améliorer l'interprétation de ces résultats.

Nous avons poursuivi cette étude en l'implémentant au sein du modèle ARPEGE. Trois expériences différentes sur une période de deux mois ont été mises en place : (i) l'expérience **GMI-BAYRAD 3SSP** réalisant des inversions avec les 3 SSPs présentées précédemment, (ii) l'expérience **GMI-BAYRAD 4SSP** réalisant des inversions avec les 4 SSPs présentées précédemment et (iii) l'expérience **GMI-BAYRAD 1SSP** réalisant des inversions en utilisant uniquement les SSPs de la particule *Sector snowflake* pour la simulation des températures de brillance. L'expérience GMI-BAYRAD 4SSP apporte des améliorations des scores de prévisions de vents à courte échéance aux niveaux de pression 700, 850, 925 et 1000 hPa pour l'hémisphère Nord, mais les résultats sur les autres domaines sont neutres. Les résultats ont été

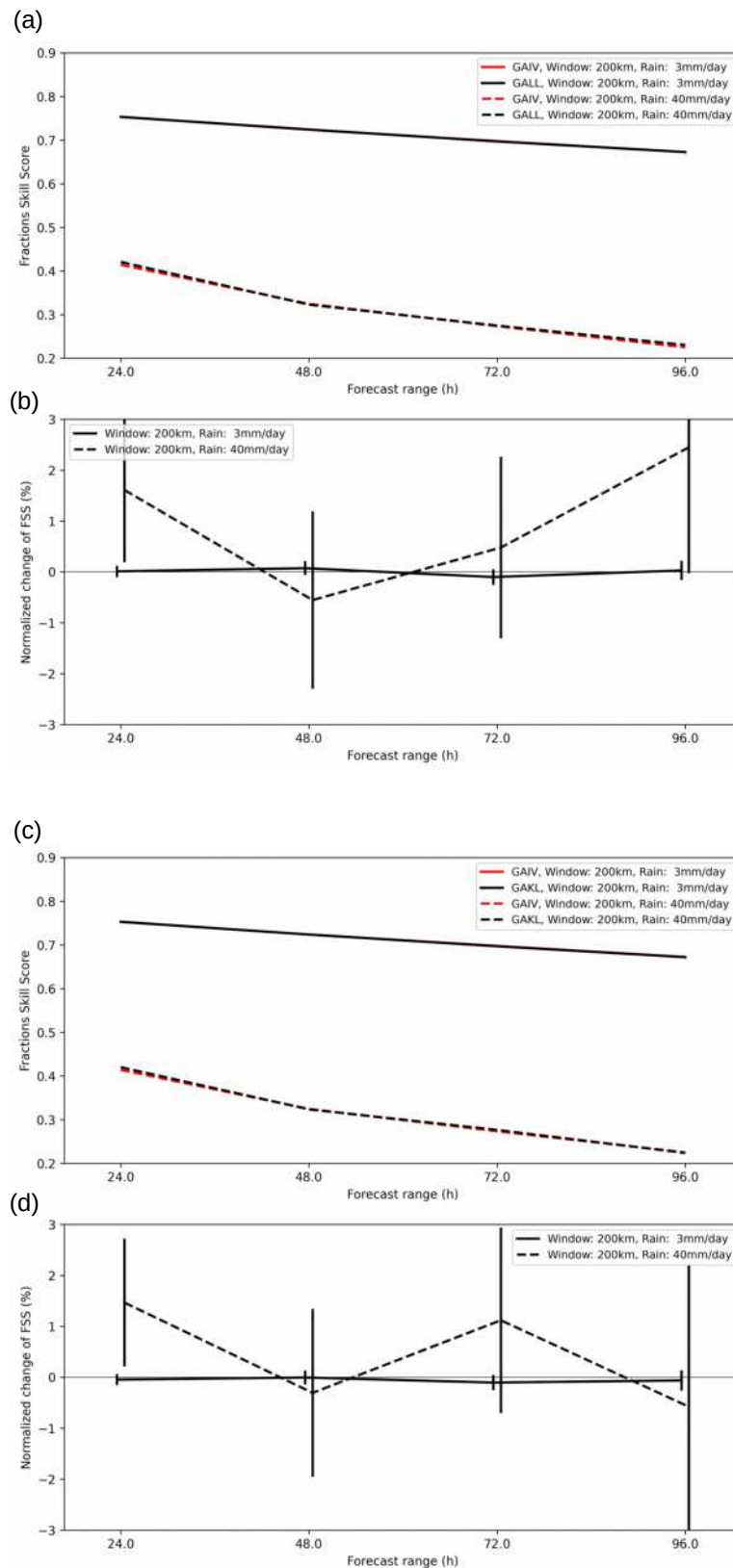


FIGURE 6.5 – a - Scores FSS des expériences GMI-BAYRAD 1SSP (en rouge) et 3SSP (en noir) sur les Tropiques. b - Différences relatives des scores FSS entre les expériences GMI-BAYRAD 1SSP et 3SSP. c - Idem à (a) pour les expériences GMI-BAYRAD 1SSP (en rouge) et 3SSP (en noir). d - Idem à (c) pour les expériences GMI-BAYRAD 1SSP et 3SSP. Tous les résultats ont été calculés pour des échéances comprises entre 24 et 96 h et des seuils de pluie de 3 (ligne pleine) et 40 mm/hr (ligne pointillée).

approfondis grâce à des scores de précipitations cumulées. Les scores ont été calculés à l'aide des produits GSMaP fournissant des cartes globales de précipitations horaires. Ces scores ont mis en évidence une amélioration des prévisions des fortes précipitations cumulées sur 24 h à l'échéance 24 h dans les expériences GMI-BAYRAD 3SSP et 4SSP face à l'expérience GMI-BAYRAD 1SSP.

Ces études ont démontré la faisabilité et l'intérêt d'utiliser plusieurs SSPs pour l'inversion des températures de brillance micro-ondes en ciel nuageux, via l'attribution dynamique de poids à ces différentes SSPs. Une analyse des distributions de poids a permis de dégager deux tendances d'utilisation de ces SSPs. En régime de faible diffusion, le contenu en information des observations se révèle être trop faible pour faire un choix entre elles, des mixtures sont donc majoritairement privilégiées. En revanche, en régime de forte diffusion, le système a tendance à faire un choix entre les SSPs mises à sa disposition, et donc à ne plus les mélanger. Une analyse monovariée à l'aide de trois prédicteurs n'a cependant pas permis d'identifier clairement les raisons des choix de particules effectués. L'assimilation de profils inversés dans le modèle global ARPEGE à l'aide de plusieurs SSPs a permis de mettre en évidence des impacts positifs, à courte échéance, sur les scores de vents et de précipitations. De telles différences entre les expériences n'étaient pas nécessairement attendues car nous nous intéressons ici à des raffinements de la représentation de la microphysique au sein du processus d'assimilation de données. Cette nouvelle méthode pour l'inversion Bayésienne a toutefois conduit à des résultats positifs qui encouragent à poursuivre les recherches sur la prise en compte des incertitudes liées au transfert radiatif dans les algorithmes d'assimilation. Au lieu d'exploiter les observations comme information à priori, nous allons dans le chapitre suivant étudier l'intérêt d'une propagation des erreurs liées au transfert radiatif. Cette méthode présente l'avantage de ne pas modifier le processus d'assimilation tout en prenant en compte ces incertitudes au sein d'un système d'assimilation d'ensemble qui permet de quantifier les erreurs des prévisions à courte échéance en considérant leurs différentes origines (observations, modèle, opérateurs d'observations).

Chapitre 7

Prise en compte des incertitudes du transfert radiatif dans une assimilation d'ensemble

7.1 Introduction

Dans ce chapitre nous allons nous attacher à répondre à la deuxième question introduite au chapitre 4, à savoir : peut-on utiliser la diversité des SSPs disponibles pour simuler les incertitudes du transfert radiatif qui leurs sont associées et les propager au sein des systèmes d'assimilation de données ? Pour cela, nous allons étudier une méthode exploitant l'assimilation directe des observations nuageuses micro-ondes au sein d'un 4D-Var. Il est en effet prévu à Météo-France d'utiliser cette technique en remplacement de la technique 1D-Bay+4D-Var discutée aux chapitres 5 et 6. Nous allons définir au sein de cette méthode une composante de l'erreur d'observation caractérisant les incertitudes liées à la spécification des hydrométéores solides au sein des simulations du transfert radiatif. La propagation des incertitudes sera par la suite simulée tout au long du processus d'assimilation de données.

Afin d'aller plus loin qu'au chapitre précédent, nous allons exploiter cette méthode dans le système d'assimilation complet d'ARPEGE qui se compose de deux éléments comme abordé au chapitre 1 : (i) un ensemble d'assimilation 4D-Var, nommé AEARP¹, qui permet de calculer des matrices \mathbf{B} de covariances des erreurs d'ébauche, (ii) un système d'assimilation déterministe 4D-Var qui utilise ces matrices \mathbf{B} . Pour rappel, dans le chapitre 6, nous avons seulement modifié le 4D-Var déterministe avec la méthode "multi-particules", mais pas la matrice \mathbf{B} . Duncan et al. (2021[202]) ont étudié l'effet de la mise à jour de la matrice \mathbf{B} sur des scores de prévisions et ont en effet montré que les impacts sur la prévision déterministe étaient d'une importance secondaire par rapport à ceux des observations elles-mêmes. Ici, nous allons cependant tenter de représenter une source d'erreur non représentée jusqu'à présent dans les systèmes d'assimilation d'ensemble.

Pour étudier l'intérêt d'une simulation des incertitudes liées au transfert radiatif, nous allons, dans la première partie de ce chapitre, définir quatre expériences d'assimilation d'ensemble. Une première expérience de contrôle, sans observations du radiomètre GMI, va permettre de calculer une matrice \mathbf{B} de référence. Deux autres expériences seront ensuite définies avec l'ajout progressif des observations claires puis nuageuses du radiomètre micro-ondes GMI pour construire d'autres matrices \mathbf{B} de covariances d'erreurs d'ébauche. La dernière expérience consistera à attribuer aléatoirement des SSPs choisies dans la base de Liu2008 pour chaque membre de l'ensemble. Comme nous le verrons par la suite, les erreurs d'observations spécifiées à l'aide des innovations seront elles aussi adaptées à chaque SSP. La troisième partie de ce chapitre est consacrée à l'évaluation des résultats des expériences d'assimilation d'ensemble. Finalement dans la dernière partie, les matrices \mathbf{B} issues des expériences d'assimilation d'ensemble seront

1. Par analogie avec la prévision d'ensemble, qui vise à simuler les erreurs de prévision, on parle aussi d'assimilation d'ensemble, qui elle cherche à simuler les erreurs d'analyse et d'ébauche.

couplées à un 4D-Var déterministe afin d'évaluer l'impact des nouvelles covariances sur ce dernier. Des scores d'humidité spécifique, de température et de vent seront ici présentés. L'ensemble de cette étude sera réalisée sur une période de trois mois à l'aide du modèle de PNT global ARPEGE.

7.2 Méthodologie

Nous définissons ici les différentes expériences mises en place pour étudier la méthode propageant les incertitudes du transfert radiatif à l'aide d'un système d'assimilation d'ensemble. Ces expériences ont toutes été cyclées (i.e. l'ébauche du réseau d'intérêt est une prévision à courte échéance issue de l'analyse du réseau précédent) sur une période de trois mois allant d'août 2021 à novembre 2021 à l'aide du système d'ensemble AEARP. Comme décrit au chapitre 1, l'AEARP est un ensemble de 4D-Var tournant en parallèle. Ces 4D-Var ont des caractéristiques communes avec le 4D-Var déterministe : ils possèdent des fenêtres d'assimilation de 6 h et ils assimilent les mêmes observations. Les observations et l'ébauche sont cependant perturbés de façon à représenter les erreurs associées comme détaillé dans la section 7.3. Comme indiqué dans le tableau 7.1, l'AEARP utilise une seule minimisation T224, ainsi qu'une résolution T499 pour la prévision, pour des raisons de coûts numériques. Les différentes observations assimilées au sein de l'AEARP sont de même répertoriées dans le tableau 6.1 introduit au chapitre 6. Détaillons maintenant les caractéristiques des quatre expériences définies :

AEARP	
Résolution 1 ^{ère} minimisation	T224 (~90 km)
Résolution de la prévision	T499 (~40 km)

TABLE 7.1 – Les différentes résolutions de l'AEARP

- **EDA-NOGMI** : Dans cette expérience, l'AEARP opérationnelle n'est que peu modifiée. Ainsi, alors qu'en opérationnel les observations du radiomètre GMI sont assimilées, ici seules les observations répertoriées tableau 6.1 sont assimilées. Il s'agit d'une expérience de référence.

- **EDA-CLEARSKY** : Dans cette expérience, les observations en ciel clair du radiomètre GMI sont assimilées en plus des observations rassemblées au tableau 6.1.

- **EDA-ALLSKY** : Dans cette expérience, les observations en ciel clair et en ciel nuageux du radiomètre GMI sont assimilées en plus des observations tableau 6.1. La SSP *Sector snowflake* de la base de données de Liu2008[11] est spécifiée à l'opérateur d'observation pour simuler les températures de brillance de l'ébauche (Geer et Baordo, 2014[1]).

- **EDA-MULTI** : Afin de décrire cette expérience, nous devons introduire la notion de modèle d'erreur. Dans la plupart des cas, les erreurs d'observations sont fixes dans l'assimilation de données. Pour les observations micro-ondes nuageuses et pluvieuses, la forte non-gaussianité des statistiques d'innovations a conduit à adopter une stratégie différente. Comme décrit ci-dessous, un modèle d'erreur est défini afin d'augmenter les erreurs d'observations en présence de nuages, réduisant ainsi les non-gaussianités.

Pour le radiomètre micro-ondes GMI, Météo-France exploite le modèle d'erreur d'observations développé par Geer et Bauer (2010[26]). Ce modèle est calculé en catégorisant les innovations par intervalles d'un prédicteur caractérisant la quantité de nuage au sein des observations et de l'ébauche. En effet, comme les innovations sont d'autant plus élevées que les observations ou les profils du modèle sont nuageux ou précipitants, la nécessité d'élaborer un modèle d'erreur dépendant de la quantité de nuages s'est rapidement révélée indispensable. Le modèle de Geer et Bauer (2010[26]) est défini à l'aide d'un prédicteur dit "symétrique". Ce prédicteur est calculé pour le radiomètre GMI à l'aide des fréquences à 37 GHz en polarisation horizontale et verticale (Petty et Katsaros 1990[203], Petty 1994[98]). Pour définir ce prédicteur, il convient en premier lieu de calculer l'indice P_{37} pour les observations et pour les ébauches :

$$P_{37} = \frac{TB^{37v} - TB^{37h}}{TB_{clr}^{37v} - TB_{clr}^{37h}}$$

où TB^{37v} et TB^{37h} sont respectivement les températures de brillance observées ou simulées à la fréquence 37 GHz polarisées verticalement et horizontalement. Et TB_{clr}^{37v} et TB_{clr}^{37h} les températures de brillance simulées en ciel clair, i.e. ne tenant pas compte des contributions nuageuses, à la fréquence 37 GHz polarisées verticalement et horizontalement. Afin que cet indice augmente avec la quantité de nuages, on applique la transformation :

$$C_{37} = 1 - P_{37}$$

Lorsque C_{37} , associé à une observation (resp. un profil modèle) vaut 1, cela signifie que l'observation (resp. le profil modèle) est située en ciel fortement nuageux et/ou précipitant (car les hydrométéores ayant généralement une orientation aléatoire dépolarisent le signal radiométrique). Au contraire, si cet indice vaut 0, cela signifie que l'observation (resp. le profil modèle) est située en ciel clair.

Finalement le **prédicteur nuageux symétrique** qui permet de définir les modèles d'erreurs des observations s'exprime comme la moyenne de la quantité de nuages contenue dans les profils atmosphériques de l'ébauche C_{37}^B et des observations C_{37}^O , à savoir :

$$\bar{C}_{37} = \frac{C_{37}^B + C_{37}^O}{2}$$

De la même manière que précédemment, lorsque \bar{C}_{37} vaut 1, l'observation et le profil modèle se situent en ciel nuageux. Lorsque \bar{C}_{37} vaut 0, l'observation et le profil modèle se situent en ciel clair. Et lorsque que cet indice est situé entre 0 et 1 cela signifie que l'on se trouve dans une situation intermédiaire : l'observation et/ou le profil modèle ne sont pas tout à fait nuageux ou tout à fait en ciel clair.

En catégorisant les écarts-types des innovations par intervalles de valeurs du prédicteur nuageux, il est possible d'ajuster une courbe analytique définissant le modèle d'erreur (dans la pratique on choisit souvent une régression linéaire par morceaux). Finalement, lors du processus d'assimilation de données, chaque observation se voit attribuer une valeur d'écart-type des innovations comme terme d'erreur d'observation. Cette valeur est définie dans le modèle d'erreur en fonction de la fréquence à laquelle on observe et de la valeur du prédicteur nuageux symétrique. En théorie cette erreur reflète une contribution des erreurs d'observations et de l'ébauche, mais en pratique cette erreur est seulement attribuée à l'erreur d'observation (Geer et al. 2010[26]). Finalement, ces modèles d'erreurs ont été calculés pour chaque SSP et chaque fréquence du radiomètre GMI.

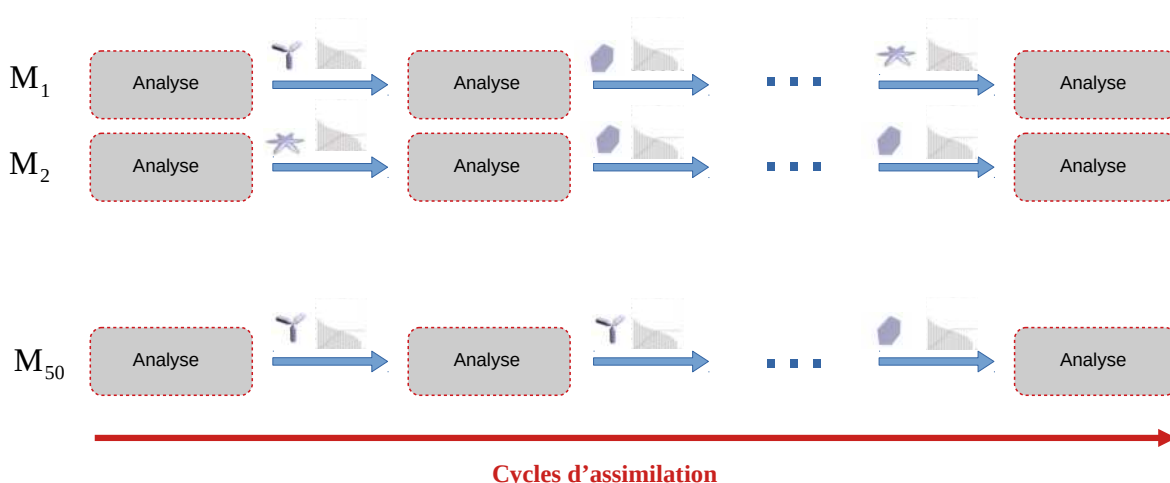


FIGURE 7.1 – Schéma explicatif de l'expérience EDA-MULTI. Pour chaque cycle d'assimilation et chaque membre de l'ensemble, une SSP est tirée aléatoirement et spécifiée à l'opérateur d'observation. Son modèle d'erreur associé est également exploité. Les SSPs pouvant être choisies aléatoirement sont celles des particules suivantes : *Thin plate*, *Rosette 3-bullet* et *Sector snowflake* de la base de données de Liu2008[11].

Ainsi, dans l'expérience EDA-MULTI, nous avons voulu évaluer l'impact des incertitudes du transfert radiatif dans une assimilation d'ensemble. Pour cela un tirage aléatoire des SSPs est réalisé pour chaque membre de l'ensemble et chaque cycle d'assimilation. La SSP ainsi que son modèle d'erreur associé sont alors spécifiés lors du processus d'assimilation de données, comme illustré sur la figure 7.1. Nous avons effectué un tirage aléatoire entre trois SSPs de la base de Liu2008 encadrant la distribution des températures de brillance observées : la *Thin plate*, la *Rosette 3-bullet* et la *Sector snowflake*. Leurs modèles d'erreurs associés ont été calculés à l'aide d'observations du radiomètre GMI et de prévisions à 6 h d'échéance du modèle ARPEGE, valides à l'instant des observations, sur une période de 10 jours pour les 13 fréquences de GMI. Des exemples de ces modèles sont représentés à la fréquence 183.31 ± 7 sur la figure 7.2. On observe que pour cette fréquence, particulièrement sensible aux hydrométéores solides, les écarts-types des innovations (en pointillé bleu) varient en fonction de la SSP spécifiée. Les modèles utilisés en pratique dans l'assimilation correspondent aux courbes rouges linéaires par morceaux avec les caractéristiques suivantes : une valeur de plateau basse pour le ciel clair, une rampe croissante afin d'augmenter les erreurs d'observations en ciel de plus en plus diffusant, une valeur plateau afin d'utiliser une erreur maximale en ciel diffusant (la décroissance des points bleus étant plus reliée à un manque d'échantillons qu'à une réelle erreur plus faible en ciel très diffusant). En comparant les différentes courbes rouges, on constate que la valeur du plateau est plus basse pour *Sector Snowflake* que pour les deux autres particules. Ces résultats confirment ainsi la nécessité d'adapter les modèles d'erreurs selon la SSP au sein des expériences d'assimilation.

Au cours des différentes études, les résultats seront affichés en fonction des niveaux verticaux du modèle ARPEGE. Le tableau 7.2 permet d'associer les valeurs de pression à certains niveaux du modèle ARPEGE afin d'avoir une meilleure compréhension des résultats. Pour rappel, seules les observations du radiomètre GMI localisées sur océans sont traitées dans cette étude et nous avons également choisi d'assimiler les fréquences utilisées opérationnellement au CEPMMT (i.e. 18.7v, 18.7h, 23.8v, 36.64v, 89v, 166v, 166h, $183.31 \pm 3v$, et $183.31 \pm 7v$).

Niveau	Pression (hPa)
10	3
20	40
30	105
40	175
50	277
60	424
70	621
80	812
90	914
100	995

TABLE 7.2 – Pressions atmosphériques sur océans associées aux niveaux verticaux du modèle ARPEGE

7.3 Effets potentiels des modifications ajoutées dans l'AEARP

Dans cette section, nous allons définir une métrique évaluant les effets produits par les modifications réalisées au sein des différentes expériences d'assimilation d'ensemble. Cette métrique va être définie comme le rapport de la dispersion de l'ensemble sur des valeurs de référence pour les moyennes spatiales des erreurs d'ébauche calculées à partir de diagnostics. Ces différents termes seront exprimés à l'aide d'expansions linéarisées : (i) des perturbations d'ébauche et d'analyse calculées à partir de l'ensemble et (ii) des erreurs d'ébauche et d'analyse calculées à partir de diagnostics. Dans la suite de ce manuscrit, les valeurs de référence pour les moyennes spatiales des erreurs d'ébauche seront considérées comme la référence à atteindre par la dispersion de l'ensemble. Nous allons maintenant détailler les différents termes d'expansions qui vont nous permettre d'exprimer la métrique utilisée au sein de cette étude pour évaluer les résultats, puis discuter les effets potentiels produits par l'ajout de nouvelles observations et de leurs perturbations associées.

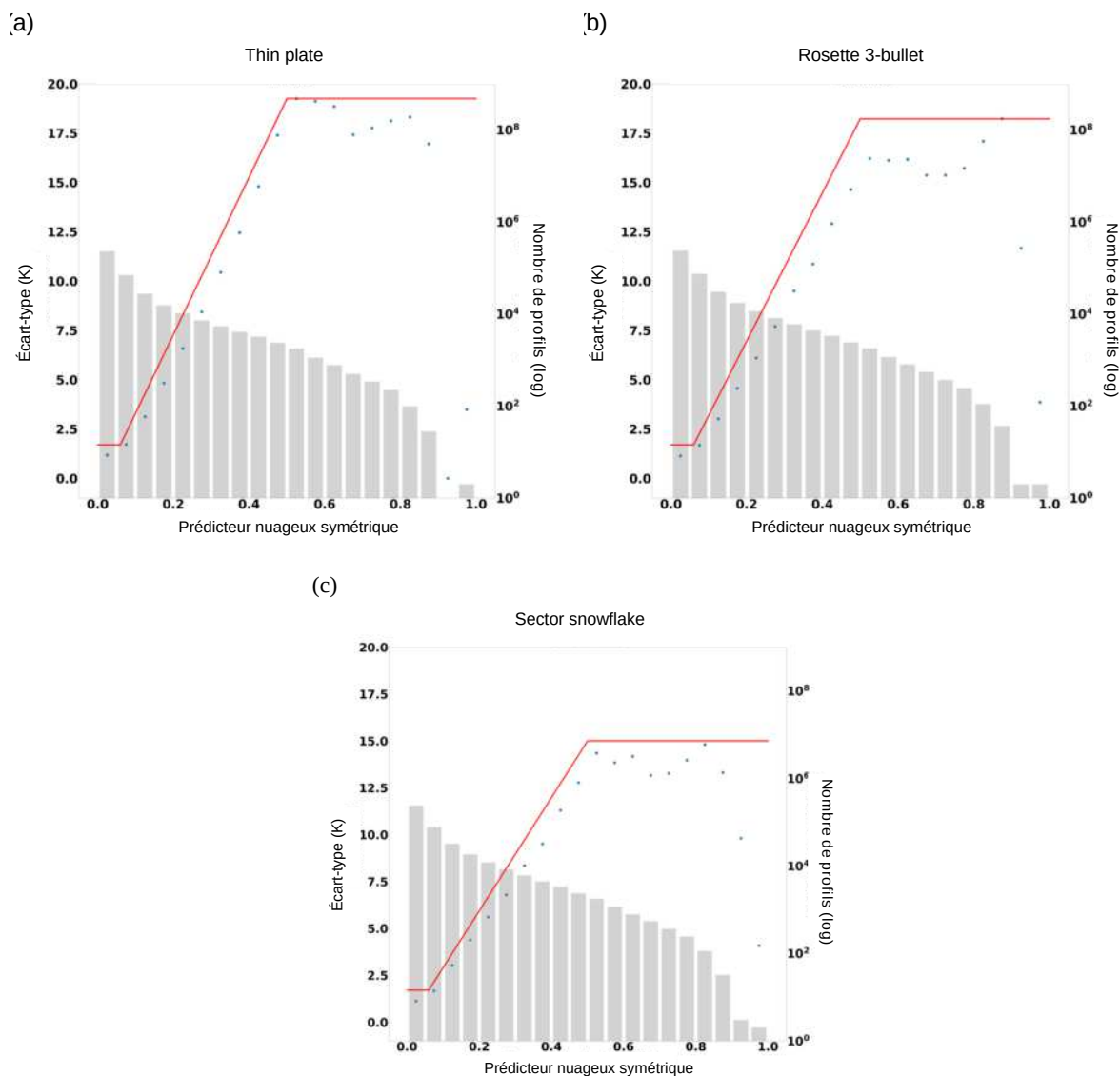


FIGURE 7.2 – (a) Écarts-types des innovations calculés à partir de prévisions 6 h du modèle ARPEGE et d’observations du radiomètre GMI à la fréquence 183.31 ± 7 GHz en fonction du prédicteur nuageux symétrique et pour la SSP *Thin plate* (en pointillé bleu). (b) Même représentation mais pour la SSP *Rosette 3-bullet*. (c) Même représentation mais pour la SSP *Sector snowflake*. Le nombre de profils de l’ébauche par intervalle du prédicteur nuageux est représenté par l’histogramme en gris. Les courbes rouges représentent les modèles d’erreurs définis pour chaque particule. Ils ont été choisis de manière à coïncider avec les écarts-types des innovations et avec celui défini dans Geer et Bauer (2010[26]).

7.3.1 Expansion linéarisée des erreurs d'analyse et d'ébauche

En assimilation de données, il est possible de relier les erreurs d'analyse ϵ_a aux termes des erreurs d'ébauche ϵ_b et d'observations ϵ_o à l'aide d'une expansion linéaire (e.g. Berre et al. 2006[204], Desroziers et al. 2009[205]). Cette manipulation permet de distinguer les erreurs d'analyse dues aux contributions de celles dues à l'ébauche. En reprenant l'équation 1.1 et en injectant les formules des erreurs d'ébauches $\epsilon_b = \mathbf{x}_b - \mathbf{x}_{true}$ et d'observations $\epsilon_o = \mathbf{y} - \mathbf{H}\mathbf{x}_{true}$, l'erreur d'analyse ϵ_a est obtenue à l'aide des arrangements suivants :

$$\begin{aligned} x_a &= x_b + K(y - \mathbf{H}\mathbf{x}_b) \\ \epsilon_a + \mathbf{x}_{true} &= \epsilon_b + \mathbf{x}_{true} + \mathbf{K}(\epsilon_o + \mathbf{H}\mathbf{x}_{true} - \mathbf{H}\epsilon_b - \mathbf{H}\mathbf{x}_{true}) \\ \epsilon_a &= \epsilon_b + \mathbf{K}(\epsilon_o - \mathbf{H}\epsilon_b) \\ \epsilon_a &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \epsilon_b + \mathbf{K}\epsilon_o \end{aligned} \tag{7.1}$$

où \mathbf{I} est la matrice identité, \mathbf{K} est la matrice de gain implicitement spécifiée dans le système d'assimilation, et \mathbf{H} est la version linéarisée de l'opérateur d'observation. Nous allons maintenant détailler les deux termes de contribution :

- le premier terme, $(\mathbf{I} - \mathbf{K}\mathbf{H}) \epsilon_b$, correspond à la contribution des erreurs d'ébauche. L'opérateur $(\mathbf{I} - \mathbf{K}\mathbf{H})$ traduit l'effet d'atténuation des erreurs d'ébauche que permet l'assimilation des observations \mathbf{y} . On peut cependant noter que les valeurs propres de l'opérateur \mathbf{K} sont inférieures à 1 et vont ainsi permettre d'atténuer les erreurs d'observations. En outre, cet opérateur se rapproche de la matrice identité lorsque les observations deviennent plus précises et plus nombreuses. Ainsi, plus les observations seront nombreuses et précises, plus les valeurs propres de l'opérateur $(\mathbf{I} - \mathbf{K}\mathbf{H})$ seront proches de 0 (e.g. Daley 1991[206]). Cela aura pour effet de réduire la contribution des erreurs d'ébauche aux erreurs d'analyse.
- le deuxième terme, $\mathbf{K}\epsilon_o$, correspond à la contribution des erreurs d'observation. Les observations étant entachées d'erreurs, le terme ϵ_o va inévitablement contribuer aux erreurs d'analyse. Si les observations sont précises, les variances d'erreur d'observation ϵ_o vont tendre vers 0, ce qui aura pour effet de réduire la contribution des erreurs d'observations aux erreurs d'analyses.

Le terme des erreurs d'ébauche ϵ_b est ici exprimé à partir des erreurs de l'analyse précédente. En effet, à chaque cycle d'assimilation, une analyse x_a est calculée à l'instant t et intégrée au modèle de prévision sur 6h d'échéance afin de fournir un état prévu \mathbf{x}_b^{t+} , valide à l'instant $t^+ = t + 6h$. Cet état est utilisé par la suite comme ébauche à l'analyse suivante, calculée à l'instant $t + 6h$. Les erreurs d'ébauche associées, notées ϵ_b^{t+} , peuvent ainsi être exprimées telles que :

$$\epsilon_b^{t+} = \mathbf{M}\epsilon_a + \epsilon_m \tag{7.2}$$

où \mathbf{M} est la version linéarisée du modèle de prévision, et $\epsilon^m = M(\mathbf{x}^{true}) - M^{true}(\mathbf{x}_{true})$ correspond à l'erreur modèle accumulée sur 6h d'échéance, résultant des différences entre le modèle de prévision M (ici, le modèle Arpege) et le modèle exact M^{true} . Ainsi, les erreurs d'ébauche ϵ_b^{t+} résultent d'une part de la propagation des erreurs d'analyse ϵ_a par le modèle \mathbf{M} , et d'autre part de l'accumulation des erreurs modèle ϵ_m au cours de l'étape de prévision. Il est important de noter que ces processus d'atténuation, d'accumulation et de propagation des erreurs sont répétés au cours des étapes successives du cyclage de l'assimilation, qui correspond aux étapes d'analyse et de prévision qui s'enchaînent toutes les 6 heures.

7.3.2 Expansion linéarisée des perturbations d'analyse et d'ébauche

L'assimilation d'ensemble AEARP cherche à construire des matrices \mathbf{B} de covariance d'erreurs d'ébauche tout en simulant les contributions des différentes erreurs et leur propagation au cours des cycles d'as-

simulation. Pour chaque analyse perturbée \mathbf{x}_a' , associée à un membre donné de l'ensemble AEARP, les erreurs d'analyse \mathbf{e}_a sont définies comme les écarts à l'analyse non perturbée telles que :

$$\mathbf{e}_a = \mathbf{x}_a' - \mathbf{x}_a \quad (7.3)$$

De façon analogue aux erreurs, ces perturbations d'analyse résultent des contributions des perturbations d'ébauche \mathbf{e}_b et des contributions des perturbations des observations \mathbf{e}_o :

$$\mathbf{e}_a = (\mathbf{I} - \mathbf{KH}) \mathbf{e}_b + \mathbf{K}\mathbf{e}_o \quad (7.4)$$

Les perturbations d'observation \mathbf{e}_o sont produites sous la forme de tirages aléatoires de la matrice \mathbf{R} de covariance des erreurs d'observation. Ces perturbations peuvent être également produites par d'autres procédés : afin de traiter les radiances satellitaires de GMI au sein de l'expérience EDA-MULTI, nous avons ici produit les perturbations à l'aide d'un ensemble de différentes versions du traitement radiatif. Les perturbations d'ébauche \mathbf{e}_b résultent, quant à elles, d'une part de la propagation des perturbations de l'analyse précédente par le modèle de prévision, et d'autre part des perturbations de modèle. Ainsi, en considérant des perturbations d'ébauche à l'instant $t^+ = t + 6h$ et par analogie avec l'équation (7.2), celles si s'expriment :

$$\mathbf{e}_b^{t^+} = \mathbf{M}\mathbf{e}_a + \mathbf{e}_m \quad (7.5)$$

Dans le cas de l'AEARP, la représentation des erreurs modèle est obtenue à l'aide d'une inflation multiplicative (Raynaud et al. 2012[207]). Les perturbations propagées d'analyse $\mathbf{M}\mathbf{e}_a$ sont ainsi amplifiées à l'aide de coefficients multiplicatifs α généralement supérieurs à 1 :

$$\begin{aligned} \mathbf{e}_b^{t^+} &= \alpha \mathbf{M}\mathbf{e}_a \\ &= \mathbf{M}\mathbf{e}_a + (\alpha - 1)\mathbf{M}\mathbf{e}_a \end{aligned} \quad (7.6)$$

En partant de l'équation 7.6 et afin d'exprimer les erreurs modèles on obtient finalement :

$$\mathbf{e}_m = (\alpha - 1)\mathbf{M}\mathbf{e}_a$$

Les coefficients α sont exprimés à partir des valeurs de référence pour les moyennes spatiales des erreurs d'ébauche $V_{moy}^{ref}[\epsilon_b]$ et des moyennes spatiales des variances des perturbations propagées d'analyse $V_{moy}^{ens}[\mathbf{M}\mathbf{e}_a]$ tels que :

$$\alpha = \sqrt{\frac{V_{moy}^{ref}[\epsilon_b]}{V_{moy}^{ens}[\mathbf{M}\mathbf{e}_a]}} \quad (7.7)$$

Généralement, les valeurs de référence $V_{moy}^{ref}[\epsilon_b]$ sont obtenues à l'aide de diagnostics a posteriori (Desroziers et al. 2005[188]) appliqués à l'estimation des variances d'erreur d'ébauche. Ces diagnostics sont calculés à partir de périodes précédentes et à l'aide de moyennes spatiales sur les variances. Dans le cas de l'humidité spécifique, les valeurs de références $V_{moy}^{ref}[\epsilon_b]$ sont estimées sous la forme d'une paramétrisation empirique (Rabier et al 1998[208], basée sur une régression linéaire) en fonction des valeurs locales de l'ébauche \mathbf{x}_b pour la température et l'humidité. Le réalisme de ces valeurs a été confirmé par des diagnostics a posteriori (Desroziers et al. 2005[188]).

7.3.3 Effets potentiels de l'ajout de nouvelles observations et des perturbations associées

Les éléments présentés dans les sections précédentes vont ici nous aider à évaluer les effets potentiels de l'ajout de nouvelles observations dans l'AEARP et de leurs perturbations associées. On a vu dans l'équation (7.4) que plus les observations seront nombreuses et précises, plus les valeurs propres de l'opérateur $(\mathbf{I} - \mathbf{KH})$ seront proches de 0, ceci ayant pour effet de réduire la contribution des erreurs d'ébauche aux erreurs d'analyse. Dans le cas où la dispersion de l'ensemble des ébauches est surestimée, on suppose ici que le terme représentant la contribution de l'ébauche a un effet plus important que l'augmentation potentielle du terme représentant la contribution des observations \mathbf{Ke}_o . Ainsi, on s'attend plutôt à la réduction des amplitudes des erreurs d'ébauche, et donc d'analyse, lors de l'ajout des observations de GMI dans les différentes expériences présentées précédemment. En effet, on a vu que l'ajout d'observations dans l'AEARP contribue à la réduction du terme de l'équation (7.4) relié aux sources d'erreurs de l'ébauche.

À l'inverse, dans le cas où les perturbations d'ébauche sont plus ou moins fortement sous-estimées, l'amplitude du deuxième terme de l'équation (7.4), correspondant à la contribution des observations, augmentera avec le nombre d'observations assimilées.

Ainsi, il sera intéressant par la suite d'examiner les variations de dispersion liées à l'ajout de nouvelles observations, en fonction des valeurs du coefficient $\beta = \frac{1}{\alpha}$ exprimé tels que :

$$\beta = \frac{1}{\alpha} = \sqrt{\frac{V_{moy}^{ens}[\mathbf{Me}_a]}{V_{moy}^{ref}[\epsilon_b]}} = \frac{\sigma_{moy}^{ens}[\mathbf{Me}_a]}{\sigma_{moy}^{ref}[\epsilon_b]} \quad (7.8)$$

Pour rappel, ici les valeurs de référence $V_{moy}^{ref}[\epsilon_b]$ sont obtenues à l'aide de diagnostics a posteriori (Desroziers et al. 2005[188], Rabier et al 1998[208]) appliqués à l'estimation des variances d'erreur d'ébauche. Le coefficient β va nous aider dans les différentes expériences à distinguer :

- les cas de dispersion relativement faible correspondant à des valeurs de β inférieures à 0.7 environ ;
- les cas de dispersion proches des écarts types de référence correspondant à des valeurs de β proches de 1 (à $\pm 30\%$ près) ;
- et les cas de dispersion relativement forte correspondant à des valeurs de β supérieures à 1.3 environ.

De cette manière, on s'attend à ce que l'ajout de perturbations représentatives du transfert radiatif dans l'expérience EDA-MULTI contribue en partie à une augmentation de l'amplitude des perturbations d'analyse et d'ébauche. Cependant, il faut aussi noter que l'amplitude moyenne des perturbations d'ébauche reste en grande partie contrôlée par la procédure d'inflation. Celle-ci permet une représentation des effets d'erreur modèle au cours de la prévision et également de contrôler les amplitudes moyennes des perturbations en bornant ces dernières. En conséquence, il est possible que la simulation des incertitudes liées au transfert radiatif conduise plutôt à des modifications relativement locales des variances d'erreur d'ébauche ainsi que des covariances spatiales associées.

7.4 Résultats d'expériences à l'aide du système d'assimilation d'ensemble AEARP

Dans cette partie, nous évaluons les bénéfices apportés par l'ajout successif des observations de GMI et du tirage aléatoire des SSPs au sein d'expériences d'assimilation d'ensemble. Nous allons étudier les variations de dispersion dues aux modifications de l'AEARP, y compris en fonction du ratio β défini dans l'équation 7.8. Pour rappel, cette référence peut être vue comme une cible à atteindre par la dispersion de l'ensemble. Si le rapport β est plus petit (resp. grand) que 1 dans une expérience de référence, on va plutôt chercher à augmenter (resp. diminuer) la dispersion de l'ensemble dans la nouvelle expérience.

L'apport des observations en ciel clair est tout d'abord évalué sur la figure 7.3-a, où les différences des écarts-types en humidité spécifique moyennées entre les expériences EDA-NOGMI et EDA-CLEARSKY sont présentées en fonction du rapport β de dispersion de l'expérience de référence (EDA-NOGMI) pour chaque niveau modèle. Les différences sont moyennées sur la période de 3 mois. La figure 7.3-b présente le nombre de profils du modèle ARPEGE par intervalle de ce même rapport. On observe sur la figure 7.3-a que pour les hautes couches du modèle (niveaux de pression compris entre les niveaux 1 et 60), les différences ne sont pas assez fortes pour être visibles. Cela s'explique par l'altitude de ces niveaux, pour la plupart compris dans la stratosphère et la haute troposphère et pour lesquels le radiomètre GMI est peu sensible mais aussi par les très faibles valeurs de l'humidité spécifique dans ces régions. Si l'on s'intéresse aux niveaux du modèle compris entre 60 et 95, on observe une forte diminution de la dispersion en humidité spécifique lorsque $\beta \in [0.8;1.0]$ et cette diminution de la dispersion devient ensuite plus marquée pour des valeurs du rapport β plus grand que 1. Ce résultat indique que la dispersion en humidité spécifique de l'ensemble a été diminuée, grâce à l'ajout des observations en ciel clair de GMI, notamment aux localisations où la dispersion en humidité spécifique de l'ensemble de l'expérience EDA-NOGMI était trop élevée. L'ajout des observations en ciel clair de GMI permet donc de se rapprocher des valeurs cibles pour la dispersion.

Les différences des écarts-types en humidité spécifique et les valeurs du rapport de dispersion moyenné sur la période de 3 mois vers 800 hPa (niveau modèle 80) sont présentées respectivement figures 7.3-c et d. Sur la figure 7.3-d, on observe que pour les zones caractérisées par un fort rapport (en rouge), la dispersion de l'ensemble en humidité spécifique (figure 7.3-c) a été diminuée dans l'expérience EDA-CLEARSKY par rapport à l'expérience de contrôle. Au contraire, pour les zones caractérisées par un faible rapport (en bleu), la dispersion de l'ensemble en humidité spécifique a été augmentée par l'ajout des observations en ciel clair de GMI. Si l'on regarde le nombre de profils modèles catégorisé à l'aide du rapport β de l'expérience EDA-CLEARSKY figure 7.3-b, on remarque une majorité de cas sous-dispersifs dans cette même expérience. Les différences des écarts-types en humidité spécifique résultantes ont ensuite été représentées pour tous les niveaux modèles ARPEGE figure 7.3-a. Les résultats sont cohérents avec ceux des figures 7.3-c et d : on note une légère augmentation de la dispersion pour les valeurs du rapport inférieures à 0.6 dans le contrôle. Ainsi, au niveau 80, l'ajout des observations en ciel clair de GMI a permis d'ajuster la dispersion de l'ensemble vers les valeurs cibles de référence de l'expérience EDA-NOGMI comme souhaité. On observe de même sur la figure 7.3-a une augmentation de la dispersion de l'ensemble dans l'expérience EDA-CLEARSKY pour des valeurs du rapport plus grand que 1, associée à des niveaux modèles proches de la surface (entre 95 et 105 ou entre 971 et 1010 hPa). On peut aussi noter que la dispersion de l'ensemble diminue pour des valeurs du rapport comprises entre 0.4 et 1. Bien que les variations des résultats sont ici décrites de manière factuelle, la projection des résultats par rapport à une dispersion référence peut être considérée comme subjective. Néanmoins la dispersion référence exploitée ici a été validée au cours des années à l'aide de diagnostics de Desroziers (Desroziers et al. 2005[188]).

Les différences des écarts-types en humidité spécifique et les valeurs du rapport moyennées sur la période de 3 mois vers 990 hPa (niveau modèle 100) sont présentées respectivement sur les figures 7.4-a et b. Sur le graphique 7.4-b, il apparaît que les fortes valeurs du rapport se situent sur les continents et que la dispersion en humidité spécifique (figure 7.4-a) a été augmentée dans ces zones dans l'expérience EDA-CLEARSKY. Les valeurs du rapport plus petites que 1 sont elles situées sur les océans et sont associées à une diminution de la dispersion en humidité spécifique. Ces résultats sont compliqués à interpréter notamment du fait de la complexité des effets de cyclage de l'assimilation. Une interprétation possible, qui pourrait être investiguée ultérieurement dans une étude plus approfondie, pourrait être que les perturbations d'analyse, résultant spécifiquement de l'ajout des observations GMI sur océan, aient tendance à se propager vers les continents de façon plus ou moins marquée au cours du cyclage de l'assimilation. En effet, il est possible que cette propagation se fasse de façon progressive, au cours des étapes successives d'analyse et/ou de prévision, et aussi en fonction du degré de contrainte exercée par les autres observations éventuellement disponibles sur continent. Il serait intéressant ici de calculer les longueurs de corrélation dans l'ensemble afin d'examiner si celles-ci sont assez importantes pour influencer les résultats obtenus sur les continents. Un autre axe ultérieur d'étude possible pourrait être celui du réalisme des écarts types de référence sur ces zones continentales, en recourant à des diagnostics a posteriori qui seraient menés de façon régionalisée sur ces domaines géographiques.

Ainsi, d'une manière globale, l'ajout des observations en ciel clair de GMI tend à améliorer la dispersion de l'ensemble en la faisant se rapprocher des valeurs cibles pour des niveaux de pression compris entre 60 (vers 420 hPa) et 95 (vers 970 hPa). En revanche, les effets proches de la surface, sur les surfaces continentales restent plus complexes à interpréter.

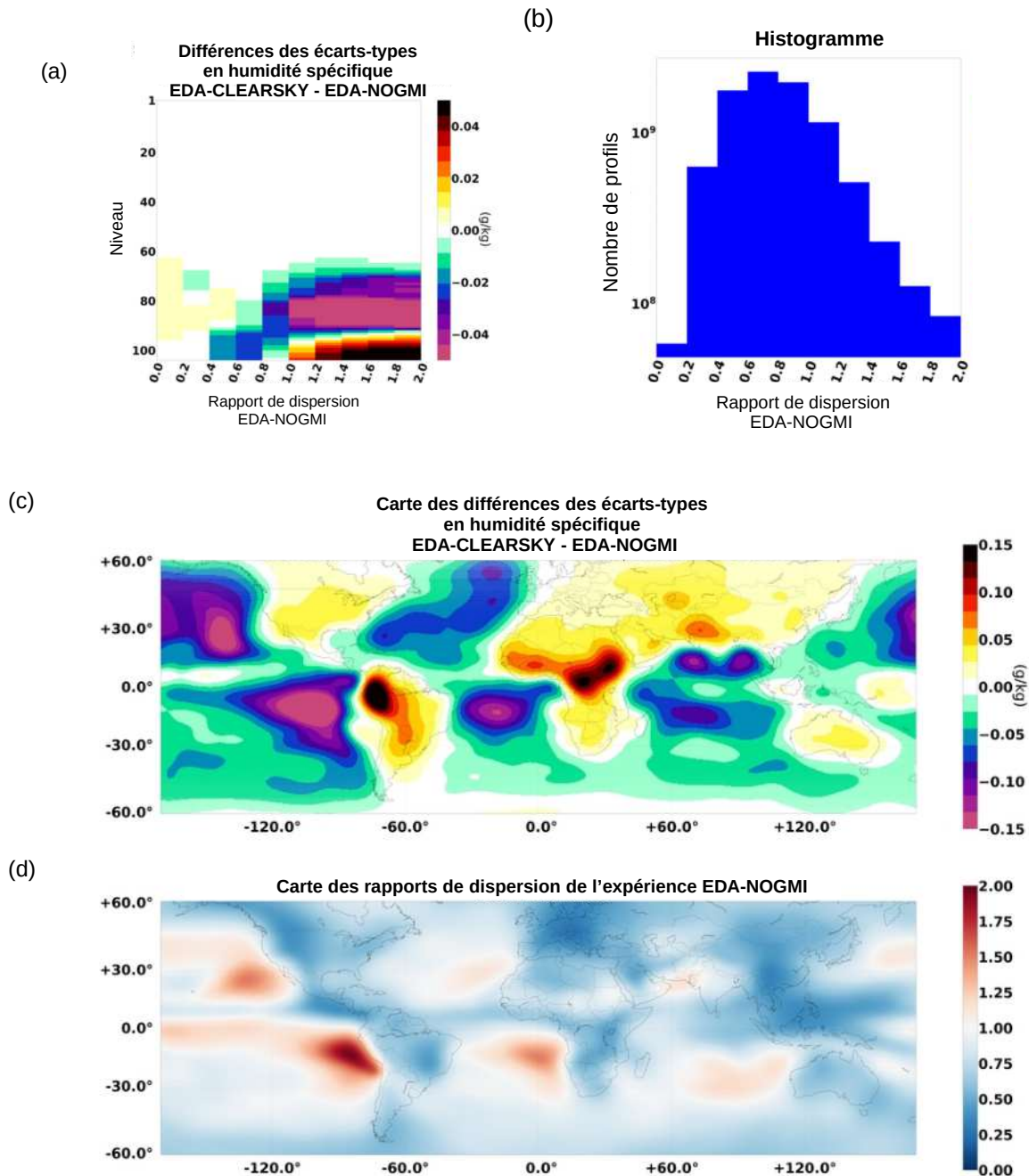


FIGURE 7.3 – (a) Différences des écarts-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI par intervalles du rapport de dispersion β de l'expérience EDA-NOGMI. (b) Histogramme des profils par intervalle du rapport de dispersion β pour l'expérience EDA-NOGMI. (c) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-NOGMI au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

Dans un deuxième temps, nous avons voulu évaluer l'apport des observations en ciel nuageux du radiomètre GMI. Les différences des écarts-types en humidité spécifique moyennées sur la période de 3

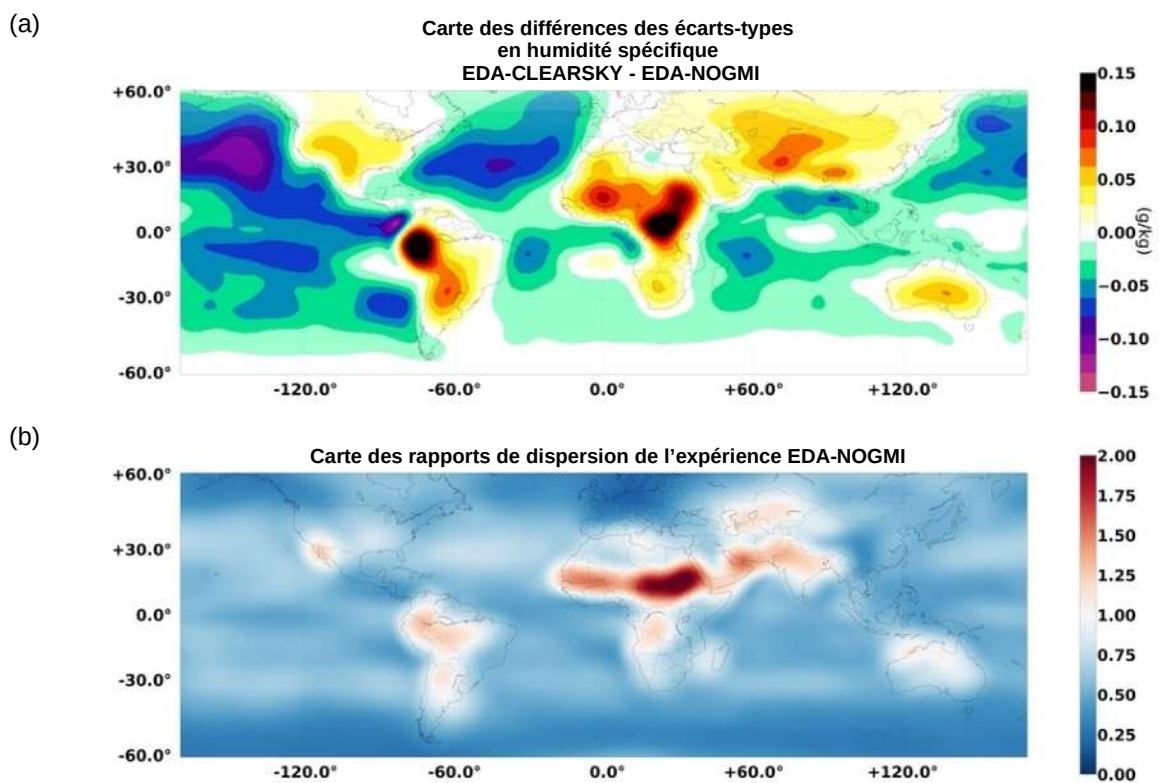


FIGURE 7.4 – (a) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-NOGMI au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

mois entre les expériences EDA-CLEARSKY et EDA-ALLSKY par intervalle du rapport de dispersion de l'expérience EDA-CLEARSKY sont présentés sur la figure 7.5-a alors que la figure 7.5-b indique le nombre de profils du modèle ARPEGE par intervalle de ce même rapport. On observe premièrement, aux niveaux modèles compris entre 60 et 105, un effet de balancier : les faibles valeurs du rapport β de dispersion dans EDA-CLEARSKY sont associées à une augmentation de la dispersion dans l'expérience EDA-ALLSKY alors que les fortes valeurs correspondent à une diminution de la dispersion.

Les différences des écarts-types en humidité spécifique et les valeurs du rapport moyennées sur la période de 3 mois au niveau modèle 80 sont présentées respectivement figures 7.5-c et d. Sur la figure 7.5-d, on constate que pour les zones caractérisées par un fort rapport (en rouge), la dispersion de l'ensemble en humidité spécifique figure 7.5-c est diminuée dans l'expérience EDA-ALLSKY. Au contraire, pour les zones caractérisées par un faible rapport (en bleu), la dispersion de l'ensemble en humidité spécifique est augmentée par l'ajout des observations en ciel nuageux de GMI. C'est par exemple le cas du golfe du Bengale, caractérisée par une forte activité convective (mousson indienne) durant les mois d'août et de septembre.

Les différences des écarts-types en humidité spécifique et les valeurs moyennées du rapport de dispersion sur la période de 3 mois vers 990 hPa (niveau modèle 100) sont présentées respectivement sur les figures 7.6-a et b. La dispersion de l'ensemble est ici augmentée et diminuée dans les zones souhaitées (figure 7.6-b), avec néanmoins quelques zones, comme sur l'océan Pacifique à l'Ouest de l'Amérique du Sud, où la dispersion de l'ensemble a diminué dans les zones où une augmentation était souhaitée. Finalement, l'ajout de l'assimilation des observations en ciel nuageux de GMI permet d'améliorer la dispersion en se rapprochant des valeurs cibles pour des niveaux de pression en dessous du niveau 60 (environ 420 hPa).

Finalement, nous avons voulu évaluer l'ajout d'une sélection aléatoire entre trois SSPs de la base de Liu2008[11] pour la simulation des températures de brillance du modèle ARPEGE. Les différences des écarts-types en humidité spécifique moyennées sur la période de 3 mois entre les expériences EDA-ALLSKY et EDA-MULTI par intervalle du rapport dispersion de l'expérience EDA-ALLSKY sont montrées sur la figure 7.7-a alors que la figure 7.7-b présente le nombre de profils du modèle ARPEGE par intervalles de ce même rapport. On note une nouvelle fois ici figure 7.7 qu'il y a une majorité de cas sous-dispersifs dans l'expérience EDA-ALLSKY. On observe figure 7.7 que pour les niveaux du modèle en dessous de 60, la dispersion de l'ensemble a diminué pour des valeurs du rapport plus grand que 1. Cela pourrait éventuellement venir de modifications de la matrice \mathbf{B} qui permettraient de mieux contrôler certaines valeurs fortes de dispersion d'ébauche. Ces différences de dispersion sont néanmoins environ dix fois plus faibles que lors des comparaisons entre les précédentes expériences (couleurs vertes claires contre couleurs violettes précédemment). Cela peut être en partie relié au rôle de stabilisation exercé pour l'inflation au cours du cyclage des assimilation perturbées. Pour les autres valeurs du rapport, les différences dans les résultats sont trop faibles pour être visibles sur le graphique.

Les différences des écarts-types en humidité spécifique et les valeurs du rapport de dispersion moyenné sur la période de 3 mois au niveau modèle 80 sont présentées respectivement sur les figures 7.7-c et d. La plage de variations sur la figure 7.7-c a été réduite de manière à mieux visualiser les différences. On note que la diminution de la dispersion représentée sur la figure 7.7-a semble provenir principalement d'une zone localisée à l'ouest de l'Amérique du Sud où le rapport de dispersion est plus grand que 1. En effet, on voit figure 7.7-c que la dispersion a surtout diminué dans cette zone. On observe aussi une augmentation de la dispersion à l'ouest de l'Amérique Centrale là où le rapport est inférieur à 1, alors qu'une diminution des écarts-types apparaît sur l'Afrique là où une augmentation serait voulue.

Les différences des écarts-types en humidité spécifique et les valeurs du rapport de dispersion moyennées sur la période de 3 mois vers 990 hPa (niveau modèle 100) sont présentées respectivement figures 7.8-a et b. On peut noter que la dispersion de l'ensemble est augmentée et diminuée autant dans les zones en ayant besoin que dans les autres. Par exemple, on observe une diminution et une augmentation de la dispersion sur l'Afrique là où on voudrait la diminuer. En conclusion, l'ajout d'une sélection aléatoire semble améliorer la dispersion de l'ensemble en la diminuant aux endroits où le rapport de dispersion est trop élevé. Néanmoins, cet effet d'amplitude doit être pris avec précaution car il n'apparaît pas distinctement lorsque l'on projette les résultats sur le globe en fonction des niveaux modèles, comme par exemple aux niveaux 80 et 100 (entre 810 et 990 hPa).

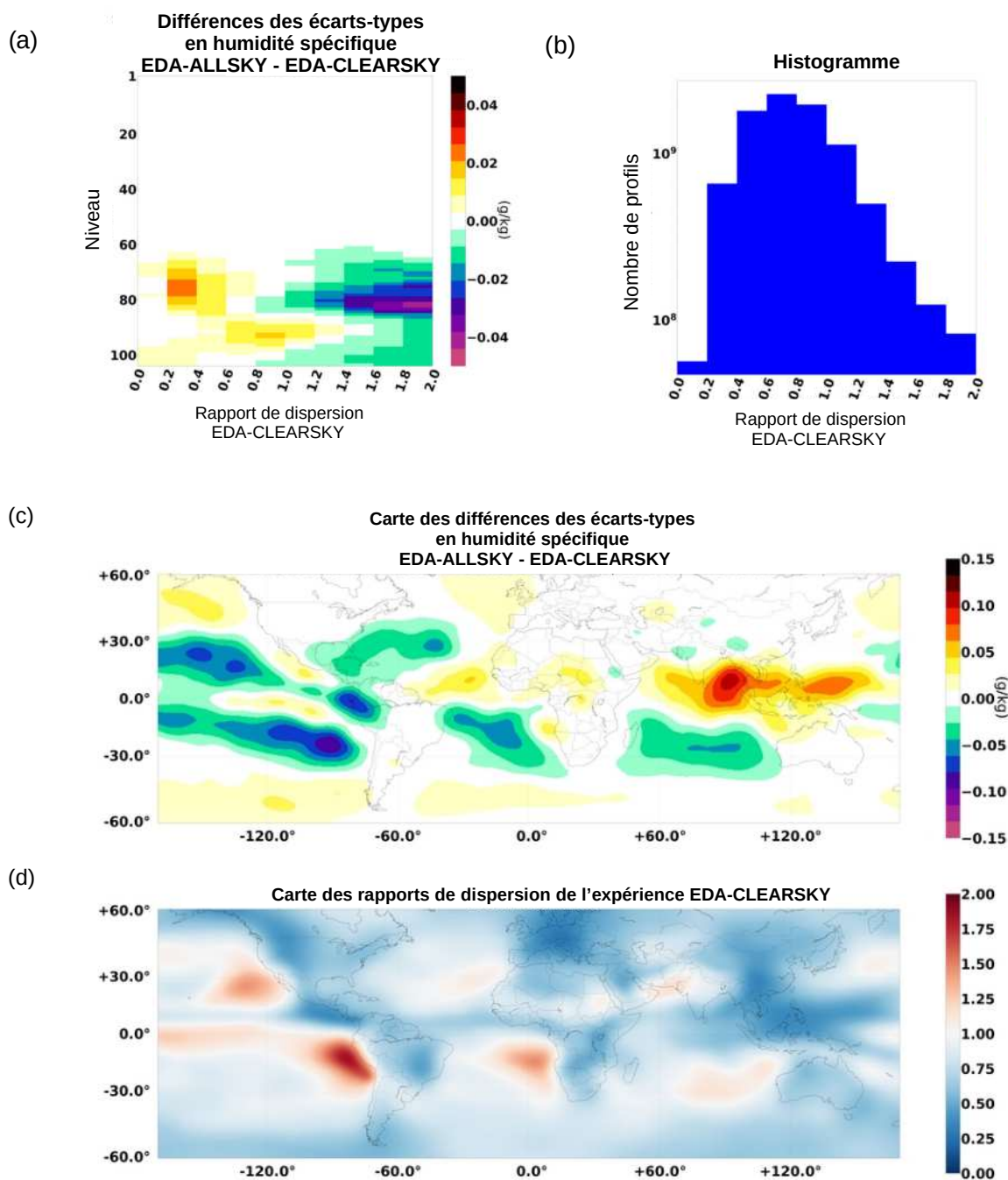


FIGURE 7.5 – (a) Différences des écarts-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY par intervalles du rapport de dispersion β de l'expérience EDA-CLEARSKY. (b) Histogramme des profils par intervalle du rapport de dispersion β pour l'expérience EDA-CLEARSKY. (c) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY au niveau 80 (vers 810 hPa) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-CLEARSKY au niveau 80 (vers 810 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

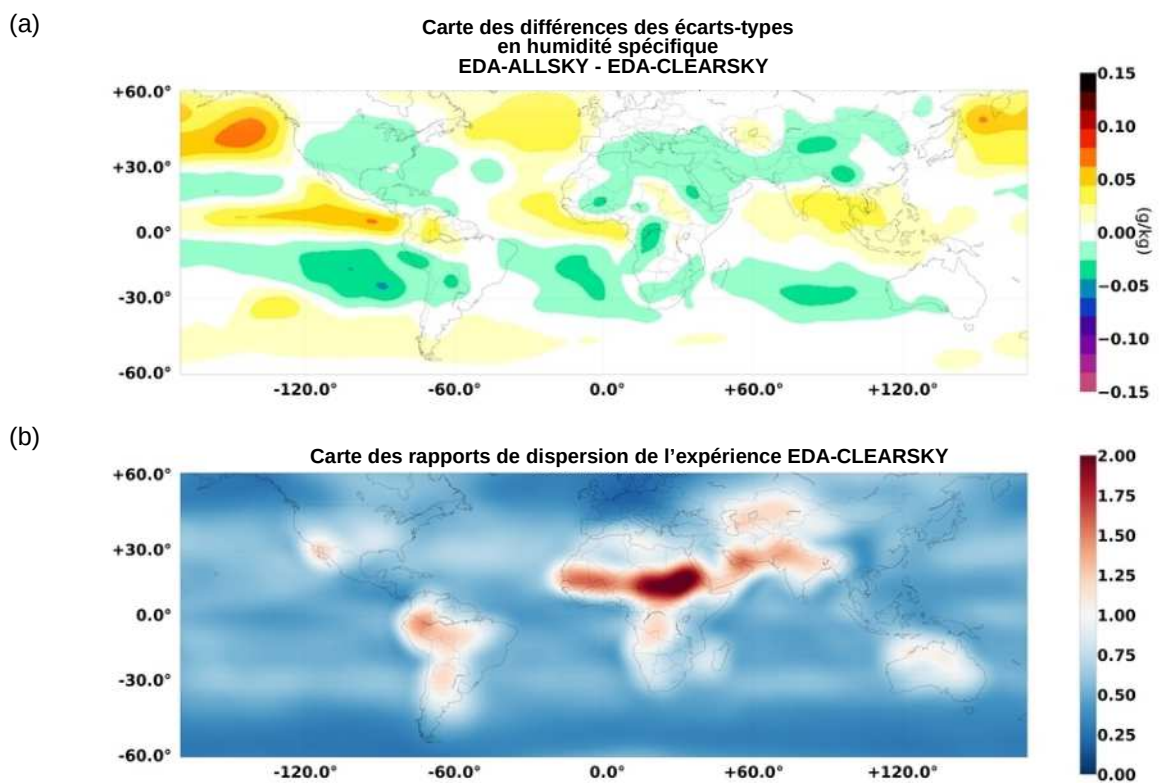


FIGURE 7.6 – (a) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-CLEARSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

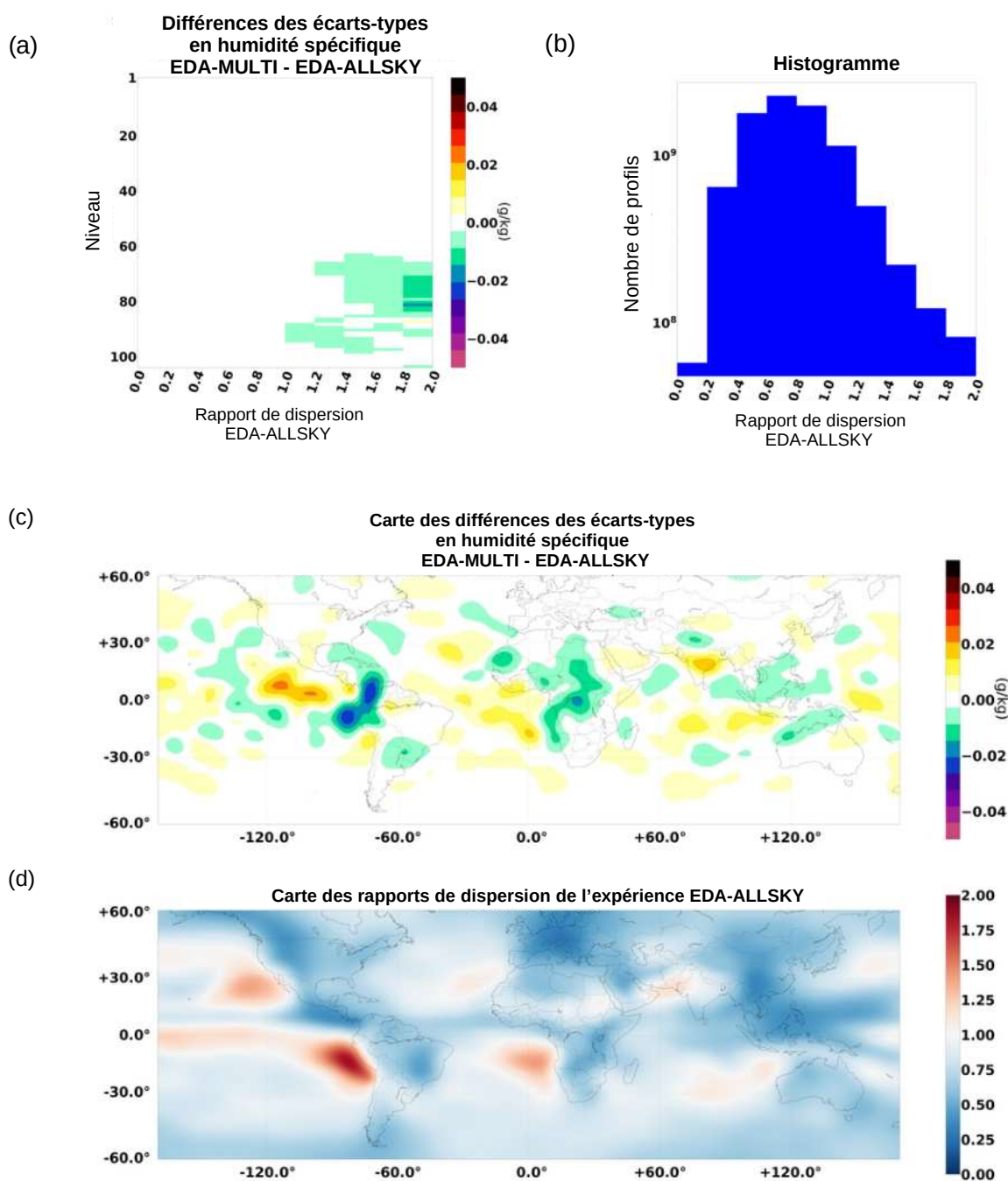


FIGURE 7.7 – (a) Différences des écarts-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY par intervalles du rapport de dispersion β de l'expérience EDA-ALLSKY. (b) Histogramme des profils par intervalles du rapport de dispersion β pour l'expérience EDA-ALLSKY. (c) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-ALLSKY au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

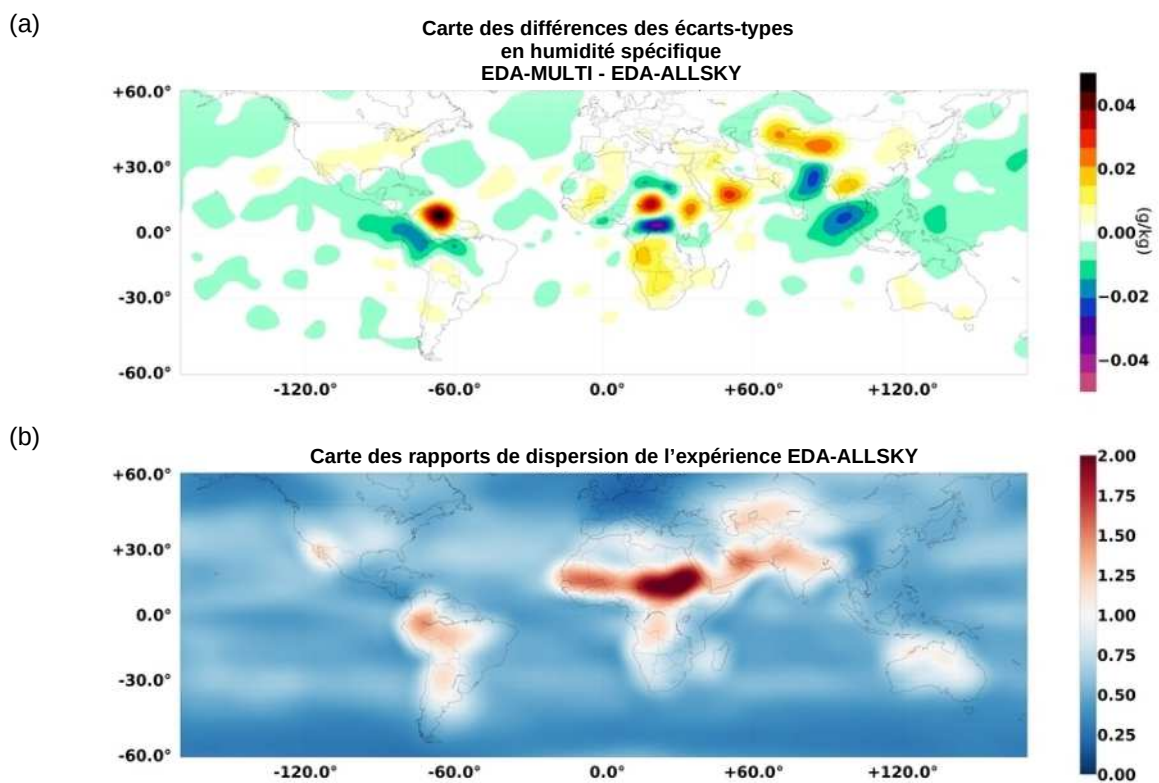


FIGURE 7.8 – (a) Carte des différences des écarts-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-ALLSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021.

7.5 Résultats d'expériences d'assimilation avec le modèle ARPEGE

Nous avons ici utilisé les matrices \mathbf{B} de covariances d'erreurs de l'ébauche issues des expériences précédentes dans des expériences avec un 4D-Var déterministe, permettant ainsi d'évaluer l'impact de ces matrices \mathbf{B} modifiées sur des scores de prévision. Le 4D-Var déterministe utilisé dans ce chapitre possède les mêmes caractéristiques que le système opérationnel à Météo-France. La seule exception est sa résolution horizontale plus basse et présentée dans le tableau 7.3. Trois expériences ont été définies pour prolonger les résultats de la section précédente :

4D-Var	
Résolution 1 ^{ere} minimisation	T224 (~90 km)
Résolution 2 ^{ime} minimisation	T224 (~90 km)
Résolution de la prévision	T798 (~11 km)

TABLE 7.3 – Les différentes résolutions du 4D-Var utilisé dans cette étude

- **4D-NOGMI** : Le 4D-Var déterministe est couplé aux matrices \mathbf{B} issues de l'expérience EDA-NOGMI et assimile les observations rassemblées dans le tableau 6.1.

- **4D-ALLSKY** : Le 4D-Var déterministe est couplé aux matrices \mathbf{B} issues de l'expérience EDA-ALLSKY et assimile, en plus des observations du tableau 6.1, l'ensemble des observations du radiomètre GMI. La SSP *Sector snowflake* de la base de données de Liu2008[11] est spécifiée à l'opérateur d'observations pour la simulation des températures de brillance de l'ébauche (Geer et Baordo, 2014[1]).

- **4D-MULTI** : On considère le même 4D-Var que dans l'expérience 4D-ALLSKY mais il est couplé aux matrices \mathbf{B} issues de l'expérience EDA-MULTI.

Les expériences d'assimilation ont été réalisées sur une période de 3 mois s'étalant d'août à novembre 2021. Au sein du système d'assimilation, en plus des prévisions à courte échéance (6 h), une prévision longue jusqu'à 102 heures d'échéance est effectuée à partir de l'analyse de 0 h UTC. Des scores de prévision ont par la suite été calculés par rapport aux analyses du modèle du CEPMMT sur la période de 3 mois pour les expériences 4D-ALLSKY et 4D-MULTI par rapport à l'expérience 4D-NOGMI sur les domaines NORD20, TROPIQUES et SUD20. Par la suite, des scores à deux niveaux de pression 400 et 925 hPa sont montrés sous forme de différences relatives entre expériences afin de faciliter l'analyse des résultats. Les systèmes d'assimilation de données étant capables d'extraire principalement des informations sur l'humidité spécifique et la température à partir de radiances micro-ondes (Geer et al. 2014[1], Lien et al. 2016[209]), les scores de prévision sur ces variables sont respectivement tracés sur les figures 7.9 et 7.10. Comme évoqué au chapitre précédent, les impacts sur les prévisions de vents sont également illustrés sur la figure 7.11. En effet, il a été montré que l'assimilation quadri-dimensionnelle permet d'inférer une information sur les champs dynamiques à partir de l'observation de l'humidité, ces impacts sont particulièrement prononcés pour l'assimilation nuageuse (Geer et al 2018[210]).

Les scores d'humidité à 400 hPa (figure 7.9-a, b, c) sont positifs (réduction de l'erreur de prévision) sur les domaines NORD20 et SUD20, et neutres sur les Tropiques. Pour l'expérience 4D-ALLSKY, les impacts sont significatifs à 95% jusqu'à 24 h sur le domaine NORD20 et jusqu'à 96 h sur le domaine SUD20. À 925 hPa (figure 7.9-d, e,f) les résultats des deux expériences sont positifs pour l'ensemble des échéances, excepté sur le domaine NORD20 où les impacts sont positifs jusqu'à 72 h pour l'expérience 4D-ALLSKY et jusqu'à 84 h pour l'expérience 4D-MULTI.

Les scores de température (figure 7.10) présentent des impacts neutres à 400 hPa sur l'ensemble des domaines, excepté sur le domaine NORD20 où l'expérience 4D-MULTI est caractérisée par des impacts positifs entre 48 h et 72 h d'échéance. À 925 hPa, on observe des impacts positifs jusqu'à 84 h d'échéance pour les deux expériences sur le domaine SUD20. Les impacts observés sur les Tropiques sont neutres pour l'expérience GMI-ALLSKY excepté entre 24 h et 48 h d'échéance où on note des impacts positifs. L'expérience 4D-MULTI est quant à elle caractérisée par des impacts positifs jusqu'à 60 h d'échéance. Sur

le domaine NORD20, les impacts sont positifs jusqu'à 36 h et 48 h respectivement pour les expériences 4D-ALLSKY et 4D-MULTI.

Les scores de vents à 400 hPa sont positifs pour le domaine NORD20 jusqu'à 36 h d'échéance pour l'expérience GMI-ALLSKY et jusqu'à 84 h d'échéance pour l'expérience 4D-MULTI (figure 7.11-a). Sur le domaine Tropiques, on constate à nouveau des impacts positifs. Ces impacts se maintiennent jusqu'à 60 h d'échéance pour l'expérience 4D-ALLSKY et entre 24 et 48 h d'échéance pour l'expérience 4D-MULTI. Sur le domaine SUD20, les impacts sont positifs jusqu'à 48 h et 84 h respectivement pour les expériences 4D-ALLSKY et 4D-MULTI. Si on examine les résultats à 925 hPa, on observe des impacts négatifs sur les Tropiques : jusqu'à 72 h d'échéance pour l'expérience 4D-ALLSKY et entre 24 et 60 h d'échéance pour l'expérience 4D-MULTI. Des impacts positifs sont cependant observés sur les domaines NORD20 et SUD20, jusqu'à 60 h d'échéance et 84 h d'échéance respectivement pour l'expérience 4D-ALLSKY et 4D-MULTI.

En résumé, les résultats présentés montrent des impacts positifs sur les scores d'humidité spécifique, de température et de vent lors de l'assimilation des observations en ciel clair et nuageux de GMI. Ces impacts sont davantage marqués pour le niveau de pression 925 hPa qui est proche de la surface. L'ajout d'une sélection aléatoire de SSPs pour construire la matrice \mathbf{B} se montre avantageux : en effet les scores de l'expérience 4D-MULTI sont significatifs pour davantage d'échéances que les scores de l'expérience 4D-ALLSKY. Ces résultats positifs apparaissent très encourageants pour la poursuite de cette première étude originale.

En plus des scores sur l'humidité, la température et le vent, nous avons également comparé les précipitations de surface cumulées sur 24 h prévues par le modèle ARPEGE avec le produit satellitaire GSMaP sur la période trois mois entre août et novembre 2021 dans la zone tropicale entre -20° et 20° de latitude. Comme au chapitre précédent, des scores FSS calculés (figure 7.12) sur le domaine Tropiques sont représentés pour les échéances maximales : +24 h, +48 h, +72 h et +96 h. Les seuils d'intensité de précipitations et le voisinage choisis sont respectivement de 3 mm par jour et 40 mm par jour, et de 200 km. La figure 7.12-a et b compare les résultats entre les expériences 4D-ALLSKY (en rouge) et 4D-MULTI (en noir). On remarque sur la figure 7.12-a, pour les scores FSS entre 24 h et 96 h d'échéance, associés aux précipitations de surface de plus de 3 mm par jour, les valeurs pour les deux expériences à 24 h d'échéance atteignent pratiquement la valeur 0.8 et décroissent légèrement vers 0.7 en fin d'échéance. Toutefois, ces bons scores FSS pour ce seuil de précipitations ne varient pas entre les expériences. Pour des précipitations dont le seuil est supérieur à 40 mm par jour, les scores FSS des deux expériences sont plus faibles se situant entre 0.25 à 24 h d'échéance et vers 0.1 à 96 h d'échéance. On remarque cependant une légère différence entre les courbes des deux expériences à 72 h d'échéance : l'expérience 4D-ALLSKY est très légèrement au dessus de la courbe de l'expérience 4D-MULTI. La figure 7.12-b présente les différences normalisées des scores de FSS entre les expériences 4D-ALLSKY et 4D-MULTI pour les deux seuils de précipitations. On remarque ici une dégradation significative des prévisions de fortes précipitations à 72 h d'échéance pour l'expérience 4D-MULTI par rapport à 4D-ALLSKY. Pour les scores de précipitations modérées aucune différence de scores n'est statistiquement significative. Ainsi l'ajout d'une sélection aléatoire sur les SSPs ne semble pas améliorer les prévisions de précipitations à l'échelle de la zone tropicale comme on avait pu le montrer dans le chapitre précédent avec la méthode "1D-Bay+4D-Var" multi-particules.

7.6 Conclusions

Dans ce chapitre, nous nous sommes intéressés à la deuxième partie de notre problématique. Le but était de déterminer s'il est possible d'utiliser la diversité des SSPs disponibles comme levier de propagation des erreurs du transfert radiatif au sein d'un système d'assimilation de données

Pour cela nous avons exploité le système d'assimilation d'ensemble AEARP qui fournit toutes les 6 h les matrices \mathbf{B} de covariances d'erreurs d'ébauches au 4D-Var déterministe opérationnel à Météo-France. Nous avons construit une méthode permettant à l'ensemble d'attribuer aléatoirement aux 50 membres une SSP ainsi que son modèle d'erreurs associé lors du processus d'assimilation de données. Nous avons

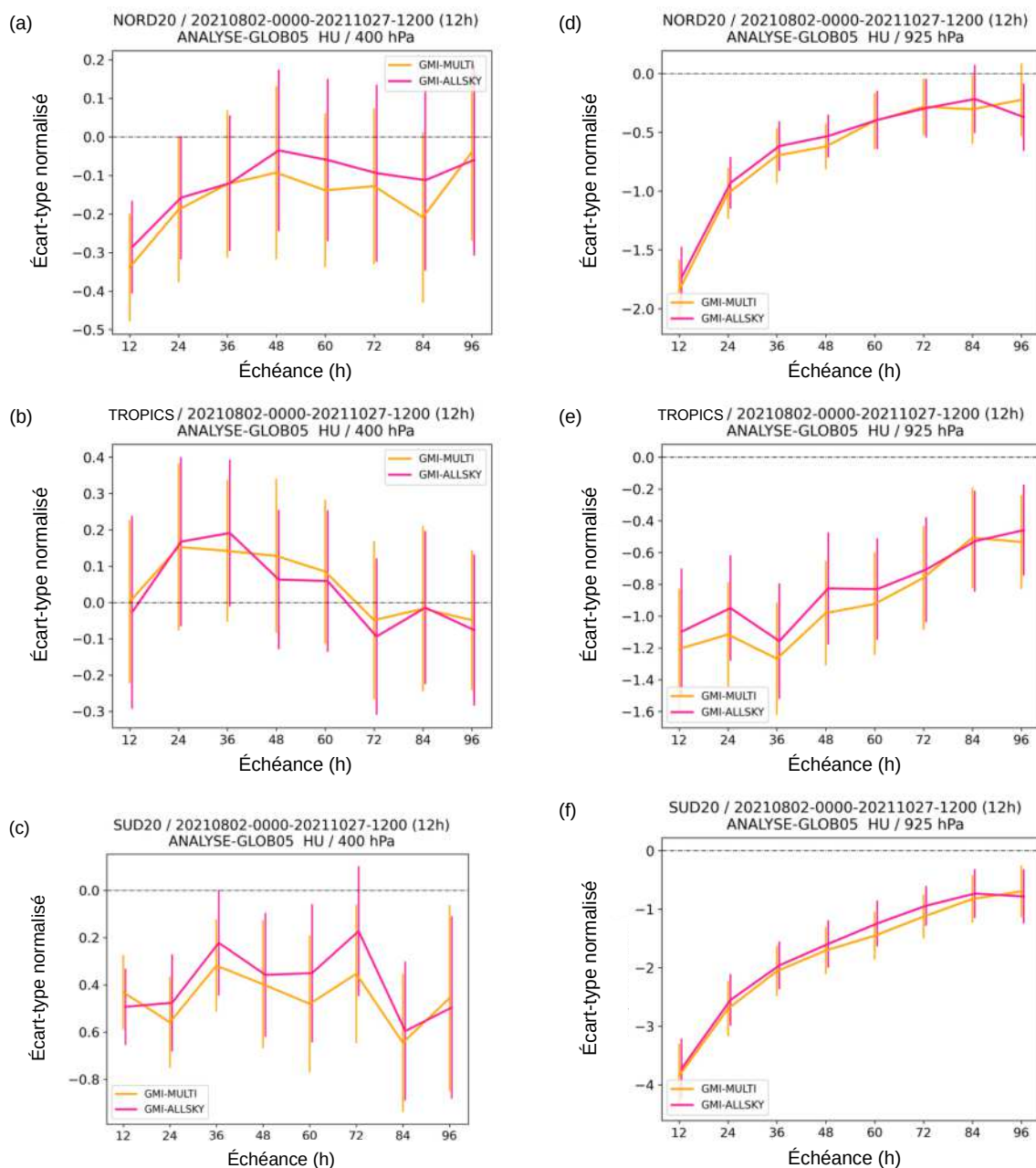


FIGURE 7.9 – Différences relatives de scores de prévisions entre trois expériences, pour l’humidité spécifique, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l’analyse du Centre Européen, pour les échéances d’ARPEGE jusqu’à +96h pour une période de 2 mois. L’expérience de référence est ici l’expérience 4D-NOGMI qui n’assimile ici aucune données de GMI et est couplée aux matrices \mathbf{B} de l’expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices \mathbf{B} des expériences EDA-ALLSKY et EDA-MULTI.

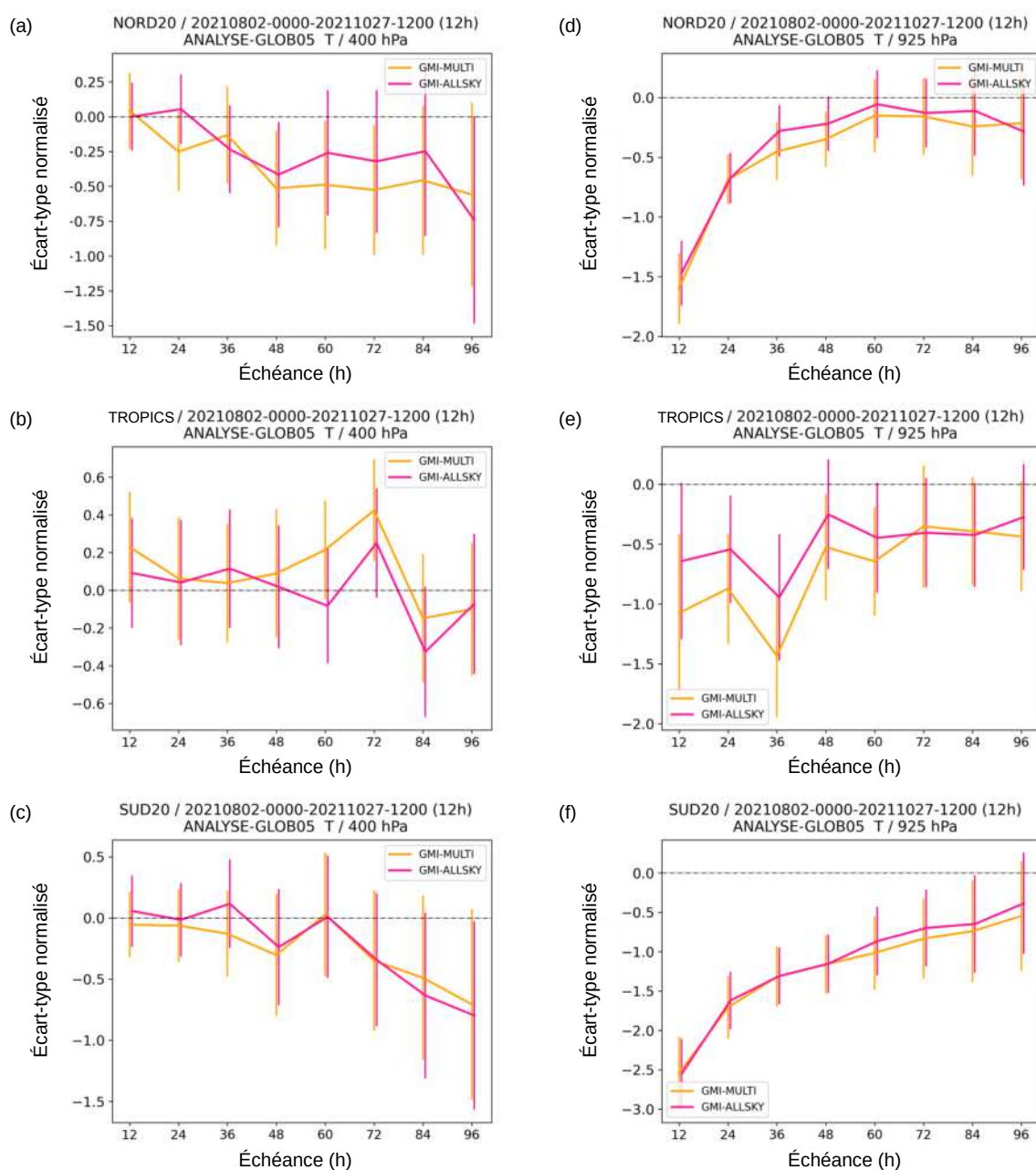


FIGURE 7.10 – Différences relatives de scores de prévisions entre trois expériences, pour la température, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du Centre Européen, pour les échéances d'ARPEGE jusqu'à +96h pour une période de 2 mois. L'expérience de référence est ici l'expérience 4D-NOGMI qui n'assimile ici aucune données de GMI et est couplée aux matrices \mathbf{B} de l'expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices \mathbf{B} des expériences EDA-ALLSKY et EDA-MULTI.

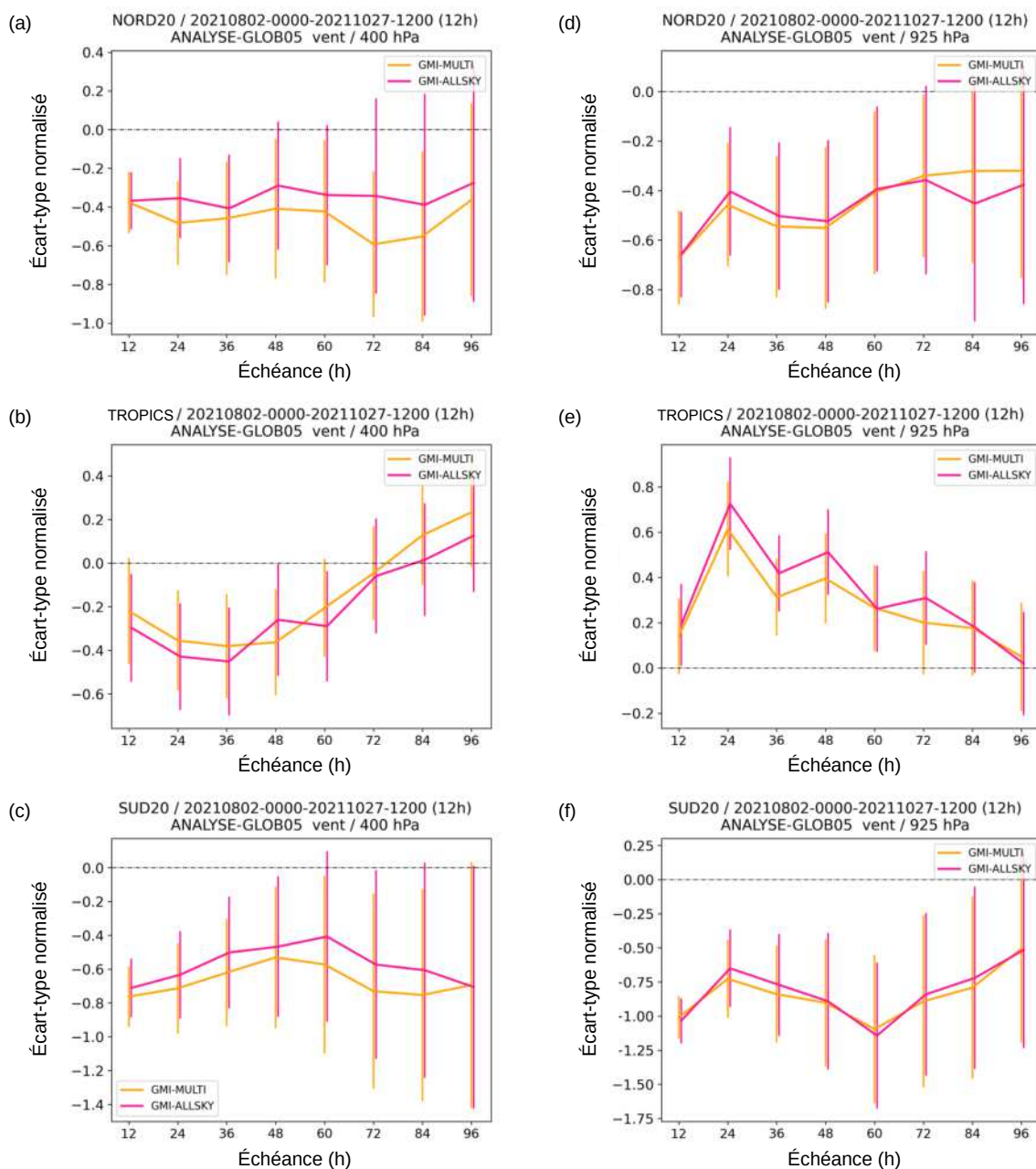


FIGURE 7.11 – Différences relatives de scores de prévisions entre trois expériences, pour le vent, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du Centre Européen, pour les échéances d'ARPEGE jusqu'à +96h pour une période de 2 mois. L'expérience de référence est ici l'expérience 4D-NOGMI qui n'assimile ici aucune donnée de GMI et est couplée aux matrices \mathbf{B} de l'expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices \mathbf{B} des expériences EDA-ALLSKY et EDA-MULTI.

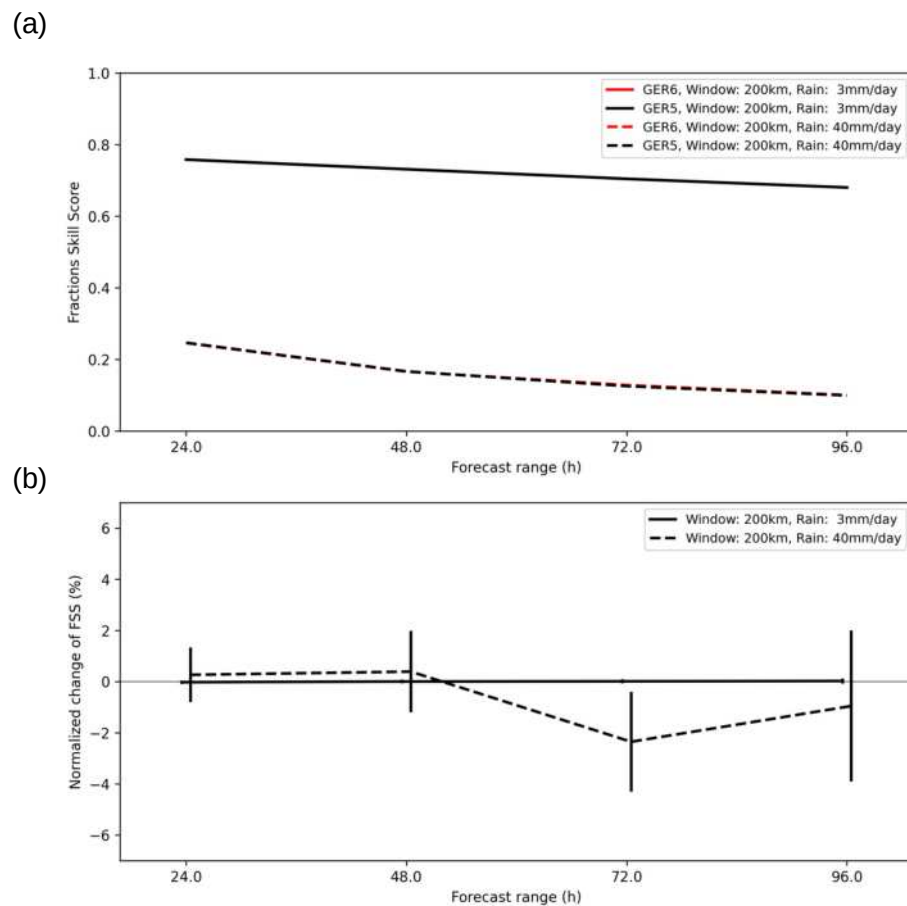


FIGURE 7.12 – a - Scores FSS des expériences 4D-ALLSKY (en rouge) et 4D-MULTI (en noir) sur les Tropiques. b - Différences relatives des scores FSS entre les expériences 4D-ALLSKY et 4D-MULTI. Tous les résultats ont été calculés pour des échéances comprises entre 24 et 96 h et des seuils de pluie de 3 (ligne pleine) et 40 mm/hr (ligne pointillée). Les barres d'erreurs représentent les différences significatives à 95%

choisi d'étudier cette méthode à l'aide des observations du radiomètre GMI assimilées dans le modèle ARPEGE avec lequel des scores de prévisions ont été calculés. Les températures de brillance des profils atmosphériques des ébauches ont ici été simulées grâce au code de transfert radiatif pour les micro-ondes RTTOV-SCATT ainsi que 3 SSPs de la table Liu2008.

Quatre expériences ont été réalisées sur une période de trois mois afin de construire la matrice \mathbf{B} à l'aide de l'ensemble de 50 4D-Var : (i) l'expérience **EDA-NOGMI** n'assimile aucune observations de GMI, (ii) l'expérience **EDA-CLEARSKY** assimile les observations en ciel clair de GMI, (iii) l'expérience **EDA-ALLSKY** assimile les observations en ciel clair et en ciel nuageux de GMI et (iv) l'expérience **EDA-MULTI** assimile les observations en ciel clair et en ciel nuageux de GMI et attribue aléatoirement à chaque membre de l'ensemble une SSP et son modèle d'erreur associé.

Après une étude des effets potentiels de ces changements, les modifications apportées par chaque expérience ont été évaluées en comparant la dispersion de l'ensemble en humidité spécifique à des valeurs de référence tirées de la littérature (Rabier et al. (1998)[208]). Nous avons fait l'hypothèse que si la dispersion de l'ensemble se rapprochait de cette dispersion diagnostique cela indiquait une amélioration du comportement de l'AEARP. Nous avons vu premièrement que l'ajout progressif des observations en ciel clair et des observations en ciel nuageux de GMI faisait tendre la dispersion de l'ensemble vers les valeurs cibles de référence à des niveaux modèles compris entre 60 et 105 (vers 420 et 1010 hPa). Lors de l'ajout des observations en ciel clair, la dispersion de l'ensemble, associée à des valeurs du rapport de dispersions fortes, s'éloignait néanmoins de la dispersion diagnostique pour des niveaux proches du sol (entre 95 et 105 i.e. entre environ 970 et 1010 hPa). Les cartes géographiques de différences d'écarts-types et de rapport de dispersion ont permis de détecter que ce comportement se situait principalement sur les continents. Cependant, les observations de GMI étant seulement assimilées sur océans dans nos expériences, il est possible que ce phénomène provienne des effets du cyclage de l'assimilation qui sont souvent complexes à interpréter. L'ajout de la sélection aléatoire sur les SSPs a aussi permis à la dispersion de l'ensemble de se rapprocher des valeurs cibles pour des valeurs élevées du rapport β et des niveaux entre 60 et 105 (vers 420 et 1010 hPa). Néanmoins, cet effet est peu visible sur les cartes aux niveaux modèles 80 et 100 (vers 810 et 990 hPa).

Nous avons prolongé cette étude en utilisant les matrices \mathbf{B} dans des expériences d'assimilation 4D-Var avec le modèle ARPEGE. Trois expériences sur une période de trois mois ont ainsi été mises en place : (i) l'expérience **4D-NOGMI** n'assimile aucune observations de GMI et utilise la matrice \mathbf{B} de l'expérience EDA-NOGMI, (ii) l'expérience **4D-ALLSKY** assimile les observations de GMI en spécifiant la SSP *Sector snowflake* à l'opérateur d'observations pour la simulation de température de brillance et utilise la matrice \mathbf{B} de l'expérience EDA-ALLSKY et (iii) l'expérience **4D-MULTI** assimile les observations de GMI en spécifiant la SSP *Sector snowflake* à l'opérateur d'observations pour la simulation de température de brillance et utilise la matrice \mathbf{B} de l'expérience EDA-MULTI.

Les expériences 4D-ALLSKY et 4D-MULTI apportent des améliorations des scores de prévisions d'humidité spécifique, de température et de vents à courte échéance aux niveaux de pression 400 et 925 hPa pour différents domaines géographiques par rapport à l'expérience 4D-NOGMI. Ce comportement est particulièrement marqué pour le niveau de pression 925 hPa. Les améliorations apportées par l'expérience 4D-MULTI se sont révélées plus fortes et plus souvent significatives que celles de l'expérience 4D-ALLSKY. Les résultats ont été approfondis grâce à des scores de précipitations cumulées. Les scores ont été calculés à l'aide des produits GSMaP fournissant des cartes sur les Tropiques de précipitations horaires. Ces scores n'ont en revanche pas permis de mettre en évidence d'amélioration des prévisions des fortes précipitations cumulées sur 24 h dans l'expérience 4D-MULTI face à l'expérience 4D-ALLSKY mais au contraire une dégradation à 72 h d'échéance.

Cette étude confirme, comme celle du chapitre précédent, l'intérêt de prendre en compte la diversité des SSPs pour réaliser des expériences d'assimilation d'observations dans des zones nuageuses et pluvieuses. Ici les incertitudes ont été considérées au sein des matrices \mathbf{B} de covariances d'erreurs de l'ébauche, pour ensuite être exploitées lors de l'assimilation dans un 4D-Var. Une analyse des dispersions en humidité spécifique des ensembles a permis de montrer l'intérêt d'assimiler les observations du radiomètre GMI pour la construction de ces matrices. La prise en compte supplémentaire d'une sélection aléatoire sur les SSPs a montré des résultats encourageants sur la dispersion de l'ensemble, qui sont cependant en amplitude d'un ordre de grandeur plus petit qu'avec les autres expériences. L'assimilation des observations de GMI au sein d'expériences 4D-Var couplées aux matrices \mathbf{B} issues des AEARP modifiées a mis en

évidence de nombreuses améliorations sur les scores d'humidité spécifique, de température et de vents. Ces améliorations sont d'autant plus marquées pour l'expérience couplée à la matrice \mathbf{B} où les incertitudes du transfert radiatif ont été prises en compte. Ces résultats positifs montrent l'intérêt de la prise en compte des incertitudes du transfert radiatif pour l'estimation des matrices de covariances d'erreurs de l'ébauche dans une assimilation d'ensemble et encouragent donc à poursuivre les études sur ce thème.

Conclusions et perspectives

Les avancées technologiques ont progressivement permis la mise en orbite de satellites embarquant des instruments micro-ondes avec une diversité de fréquences accrue, nécessitant des recherches spécifiques pour leur assimilation au sein des systèmes de prévision numérique du temps. Une difficulté importante pour l'assimilation d'observations en zones nuageuses réside dans la spécification des propriétés micro-physiques des particules constituant les nuages. Parmi ces spécifications, il est nécessaire de prescrire les SSPs de diffusion de chaque type d'hydrométéores, qui dépendent notamment de leur forme et de leur indice de réfraction. Cet aspect apparaît au sein des systèmes d'assimilation de données lors de la simulation des températures de brillance. Cette tâche est complexe pour les hydrométéores solides de par leur grande diversité morphologique présente dans la nature. Ces dernières années, plusieurs bases de données ont été produites afin de répertorier les formes des hydrométéores solides et leurs SSPs. Malgré ces nouvelles sources d'information, à ce jour une forme unique de particule (supposée *optimale*) est spécifiée pour chaque espèce d'hydrométéore aux opérateurs d'observations utilisés pour l'assimilation de données. Outre la grande simplification de l'atmosphère que cela implique, les températures de brillance simulées sont impactées et peuvent conduire à des résultats sous-optimaux dans certaines régions du globe ou pour certains systèmes nuageux lors de l'assimilation des observations micro-ondes.

L'objectif de cette thèse était d'investiguer des méthodes d'assimilation d'observations micro-ondes en ciel nuageux prenant en compte la diversité des SSPs disponibles. Pour cela deux problématiques ont été étudiées. La première était d'évaluer si les observations satellitaires micro-ondes représentaient une information a priori suffisante pour définir de manière dynamique des combinaisons de SSPs lors du processus d'assimilation. Et la deuxième était d'évaluer l'intérêt de propager les incertitudes liées à la diversité de ces SSPs à travers l'utilisation d'un système d'assimilation d'ensemble.

Conclusions et perspectives sur les sensibilités de l'inversion à divers paramètres

Pour explorer ces deux problématiques, une étude préliminaire de la première étape de la méthode "1D-Bay+4D-Var" a été réalisée. Cette méthode, déjà utilisée pour l'assimilation des observations satellitaires micro-ondes en ciel nuageux (Duruiseau et al. 2019[44]), consiste à estimer un profil d'humidité relative à partir des observations puis à l'assimiler au sein d'un système d'assimilation 4D-Var à des niveaux de pression prescrits. Cette étude a été conduite dans le but d'évaluer l'apport des basses fréquences micro-ondes jusque là non utilisées à Météo-France dans ce contexte. Le but était également d'approfondir notre connaissance sur les sensibilités de cette méthode à différents paramètres, notamment aux SSPs qui nous intéressent particulièrement dans le cadre de cette thèse. Cette étude a aussi été l'occasion d'optimiser la méthode. Quatre paramètres ont été étudiés : (i) l'erreur d'observations, (ii) la sélection des canaux, (iii) les SSPs et (iv) le choix de l'estimateur optimal. Pour cela nous avons exploité le modèle régional à échelle convective AROME-Antilles ainsi que les observations des fréquences allant de 18.7 GHz à 183.31 ± 7 GHz du radiomètre GMI à bord du satellite GPM-Core. Les températures de brillance ont été simulées avec le code de transfert radiatif RTTOV-SCATT et nous avons exploité les SSPs de la base de Liu2008. Les principales conclusions et perspectives de cette étude de sensibilité sont les suivantes :

- (i) la sensibilité à l'erreur d'observations a montré que les résultats des inversions simulées dans l'espace des observations concordent davantage avec les observations par la spécification de faibles erreurs. Au contraire, une erreur élevée amenait la moyenne pondérée des profils de l'ébauche à se rapprocher d'une moyenne arithmétique des profils du voisinage, moyennant ainsi des scènes météorologiques en ciel clair et en ciel nuageux. À la fin de cette étude, nous avons identifié la perspective d'utiliser des diagnostics a posteriori de type Des-

roziers (Desroziers et al. 2005[188]) comme erreur d'observations permettant d'attribuer une erreur différente pour chaque canal. Les essais dans ce sens ont montré que ce changement ne produisait pas des résultats significativement différents par rapport à l'utilisation d'une erreur de 1 K.

- (ii) la sensibilité à la sélection des canaux a mis en évidence les avantages de l'utilisation combinée des basses et hautes fréquences micro-ondes au sein de l'assimilation de données. Lors de l'utilisation unique des basses fréquences (resp. hautes fréquences), les résultats obtenus dans l'espace des observations ne sont pas en concordance avec les observations pour les hautes fréquences (resp. basses fréquences). Au contraire, l'utilisation combinée des basses et hautes fréquences a conduit à des résultats dans l'espace des observations en accord avec les observations pour l'ensemble des fréquences. Les profils d'humidité relative inversés résultants sont dans ce cas contraints sur toute la verticale. Cela a également mis en évidence le fait qu'il était important de se soucier des niveaux de pressions auxquels les profils d'humidité relative inversés peuvent être assimilés au sein du 4D-Var. Des études précédentes réalisées avec le sondeur SAPHIR avaient permis de définir des niveaux de pression à l'aide d'analyses sur des jacobiens d'humidité spécifique dérivés de simulations réalisées avec le code de transfert radiatif RTTOV-SCATT et de profils du modèle ARPEGE (Duruiseau et al. 2019). À la fin de cette étude, l'idée a été d'adopter une approche de calcul dynamique pour assimiler les profils d'humidité relative inversés à des niveaux de pression pertinents en fonction des canaux assimilés. Cette approche originale a été testée mais n'a pas montré d'améliorations claires sur les analyses.
- (iii) la sensibilité aux SSPs a permis de mettre en évidence les différences dans les quantités d'humidité relative des profils inversés. Les SSPs radiatives générant le plus de diffusion amènent à des profils inversés plus secs en humidité relative qu'avec des SSPs générant peu de diffusion. Cette disparité dans les résultats a confirmé l'importance d'étudier l'apport de la prise en compte de plusieurs SSPs investigués dans cette thèse.
- (iv) la sensibilité au choix de l'estimateur optimal a montré qu'un estimateur basé sur un maximum de vraisemblance avait l'avantage de ne pas être sensible à la spécification de l'erreur d'observation enlevant ainsi un degré de liberté à la méthode. Cependant, cet estimateur s'est avéré moins efficace que la moyenne pondérée pour l'assimilation d'un nombre réduit de fréquences. Les résultats dans l'espace des observations sont en effet davantage dégradés pour les fréquences micro-ondes non assimilées.

Conclusions et perspectives sur le multi-particules et l'assimilation déterministe

Après avoir étudié la méthode "1D-Bay+4D-Var", nous avons choisi de poursuivre les études suivantes à l'aide d'une erreur d'observation de 1 K, de l'ensemble des fréquences de GMI de 18.7 GHz à 183.31±7 GHz et avec comme estimateur optimal une moyenne pondérée. Nous avons ainsi dans un deuxième temps étudié l'intérêt d'exploiter les observations micro-ondes comme information a priori pour définir, pour chaque observation, une combinaison optimale de SSPs pour la simulation des températures de brillance associées. Pour cela, la formule de la moyenne pondérée a été modifiée. Une somme sur les SSPs a été introduite à celle-ci afin d'en considérer plusieurs lors des inversions. Cette étude a été menée à l'aide de la même configuration que l'étude précédente. Deux expériences ont été mises en place, une réalisant des inversions avec trois SSPs (3SSP) et une avec quatre (4SSP) sur une période de deux mois. Les résultats des inversions ont été catégorisés à l'aide de 3 prédicteurs : les précipitations de surface et le contenu en glace intégré des produits GPROF (Kummerow et al. 2016[196]) ainsi qu'à l'aide d'un indice de diffusion calculé à partir des observations de GMI (Grody 1991[197], Wilhelm et al. 2003[198]). Dans cette étude, seules les inversions réalisées à l'aide d'observations en ciel nuageux ont été considérées. Ces observations ont été classées suivant trois prédicteurs. Les résultats obtenus lors de cette étude ont mis en évidence la sélection par l'inversion Bayésienne de mixtures de SSPs pour les observations peu diffusantes. Au contraire, l'utilisation d'une unique SSP a été mise en évidence pour des observations localisées en ciel très diffusant. Toutefois, les prédicteurs sélectionnés n'ont pas permis d'identifier une combinaison de SSPs privilégiée en fonction de la scène météorologique examinée.

L'ensemble de la méthode "1D-Bay+4D-Var" modifiée a par la suite été implémentée au sein du modèle global ARPEGE et des expériences ont été réalisées et évaluées. Des scores de prévisions ont été calculés entre différentes expériences utilisant un nombre différent de SSP au sein des inversions (1SSP, 3SSP et 4SSP). Cette étude a montré des résultats encourageants notamment sur les scores de vents où l'expérience 4SSP a montré une amélioration par rapport à l'expérience 1SSP à plusieurs niveaux de pression sur le domaine NORD20 pour des échéances allant jusqu'à 48 h. Nous avons approfondi les résultats en calculant des scores de type FSS (fractions skill score) pour les précipitations en surface proposés par Roberts et Lean 2008[201]. Ces scores ont montré des améliorations des prévisions des fortes précipitations à 24 h d'échéance lors de l'utilisation de trois ou quatre SSPs par rapport à une seule SSP. Cette étude a ainsi permis de montrer la faisabilité et l'intérêt d'une méthode exploitant plusieurs SSPs pour réaliser des expériences d'assimilation. Une perspective intéressante à cette étude serait de catégoriser les résultats des inversions Bayésiennes à l'aide de nouveaux prédicteurs tels qu'une classification des scènes en nuages convectifs ou stratiformes ou encore en utilisant d'autres champs environnementaux afin de mieux comprendre ce qui pilote les choix de SSPs par l'inversion. Cette approche multi-particules a été testée dans le cadre "1D-Bay+4D-Var" et non dans l'assimilation directe car cette dernière aurait nécessité une implémentation technique plus complexe. Il aurait fallu introduire une étape de pré-calcul de la mixture optimale, mais aussi le calcul des SSPs correspondantes en ligne (calcul pour l'instant hors ligne) puis leur utilisation dans l'assimilation directe. Cette manipulation pourrait notamment être réalisée directement au sein du code de transfert radiatif RTTOV-SCATT en augmentant le nombre d'hydrométéores considérés. Chaque SSP correspondrait alors à un hydrométéore et il serait possible d'effectuer une pondération de ces derniers afin de créer des mixtures de particules comme cela est déjà possible pour certains hydrométéores.

Conclusions et perspectives sur le multi-particules et l'assimilation d'ensemble

Dans la dernière partie de cette thèse, nous avons évalué l'intérêt de propager les incertitudes du transfert radiatif dans un système d'assimilation d'ensemble. Cette propagation a été effectuée au sein de l'assimilation d'ensemble AEARP comptant 50 membres (systèmes 4D-Var) avec le modèle ARPEGE. L'assimilation d'ensemble a par la suite été exploitée en utilisant les matrices de covariances d'erreur de l'ébauche \mathbf{B} ainsi générées pour des assimilations 4D-Var déterministes afin de propager les incertitudes sur les SSPs des hydrométéores. Ces matrices de covariances d'erreurs de l'ébauche ont été construites en utilisant un système d'observations incluant progressivement les observations du radiomètre GMI sur une période de trois mois. La base de données de Liu2008 a de nouveau été exploitée. Plusieurs matrices ont ensuite été construites en exploitant : (i) les observations en ciel clair de GMI (en plus des autres observations du système global utilisées par ARPEGE), (ii) toutes les observations de GMI et (iii) toutes les observations de GMI et spécifiant aléatoirement pour chaque membre de l'ensemble des SSPs parmi trois choix possibles. À l'aide de diagnostics, nous avons montré que la prise en compte des observations nuageuses pour construire la matrice \mathbf{B} améliorerait les SSPs de dispersion de l'ensemble. L'ajout additionnel d'un choix aléatoire des SSPs pour chaque membre de l'ensemble a permis de diminuer la dispersion de l'ensemble lorsque celle-ci était trop forte dans la matrice \mathbf{B} . Par la suite chacune de ces matrices a été couplée à un 4D-Var assimilant : les observations en ciel clair de GMI pour la matrice (i) et toutes les observations de GMI pour les matrices (ii) et (iii). Des scores objectifs sur les prévisions de vents et d'humidité spécifique ont montré des améliorations pour les expériences couplées aux matrices \mathbf{B} assimilant l'ensemble des observations de GMI. Pour des niveaux de pression autour de 400 hPa, l'expérience exploitant une matrice \mathbf{B} construite à partir d'un choix aléatoire sur les SSPs montre de légères améliorations par rapport à celle spécifiant un choix fixe des SSPs. Ces niveaux pressions autour de 400 hPa correspondent à l'altitude des nuages de glace, il n'est donc pas choquant que les légères améliorations mises en évidence interviennent à ces niveaux.

Cette étude montre comme la précédente des résultats encourageants qui nécessitent toutefois d'être consolidés. La dispersion de l'ensemble est améliorée avec un choix aléatoire sur les SSPs et permet de propager les erreurs associées. L'amplitude de l'impact de ce choix aléatoire est d'un ordre de grandeur plus faible que celui induit par l'ajout des observations nuageuses elles-mêmes. Cette amplitude pourrait évoluer si cette approche était généralisée aux autres capteurs de la constellation d'instruments micro-ondes. Dans cette étude sur la matrice \mathbf{B} il n'a pas été possible de tester une expérience de type 4SSP comme pour le travail sur l'inversion Bayésienne, faute de temps. Néanmoins, il serait intéressant de documenter l'impact d'une particule très diffusante dans ce contexte. Afin d'affiner la compréhension des scores de prévision, il serait de même intéressant de réaliser un travail supplémentaire sur les covariances au sein des matrices \mathbf{B} pour en connaître leurs impacts. Dans les années à venir, les travaux concernant

l'amélioration de la dispersion de l'ensemble d'assimilation AEARP vont se poursuivre au CNRM ; suite à ce travail, la prise en compte des incertitudes associées aux opérateurs d'observation est une piste qui sera considérée.

Conclusions et perspectives générales

Deux méthodes considérant les incertitudes liées à la spécification des SSPs des hydrométéores solides et plus particulièrement celles de la neige ont été développées dans le cadre de l'assimilation de données micro-ondes spatiales pour la prévision numérique du temps. Les résultats obtenus sont très encourageants pour les études réalisées et démontrent l'intérêt de prendre en compte ces incertitudes dans le processus d'assimilation. Des perspectives communes aux différentes méthodes sont à explorer. Tout d'abord, il serait intéressant d'étudier ces deux méthodes à l'aide de la base de données de Eriksson et al. 2018[41] plus complète que celle de Liu2008 et qui est implémentée au sein de RTTOV-SCATT à partir de sa version 12. Des choix différents de SSPs sont aussi à tester, il serait par exemple intéressant d'augmenter leur nombre dans les processus d'inversion ce qui permettrait d'exploiter une base de donnée de SSPs complète. Une autre perspective serait d'implémenter ces méthodes pour les autres hydrométéores solides. Ensuite, ce travail de thèse s'est focalisé sur une seule source d'incertitudes dans le transfert radiatif en ciel nuageux, cependant le travail réalisé est suffisamment général pour être étendu à d'autres sources d'incertitudes pour les hydrométéores solides mais aussi liquide telles que le choix des distributions en taille d'hydrométéores ou bien le choix de représentation pour la distribution sous-maille des nuages qui sont toutes deux des sources d'incertitudes intéressantes à propager au sein de l'assimilation de données. Les méthodes élaborées au sein de ce manuscrit ont de plus été étudiées à l'aide d'un modèle à aire limité, AROME et d'un modèle global, ARPEGE afin qu'elles puissent être adaptées par la suite à différents modèles. Une perspective possible serait de réaliser des expériences d'assimilation avec le modèle AROME. La deuxième méthode reposant sur l'assimilation d'ensemble pourrait être, de manière analogue à l'AEARP, appliquée à l'assimilation d'ensemble AEARO (Assimilation d'Ensemble AROME).

Finalement, les travaux réalisés au sein de cette thèse font partie des recherches menées dans le cadre du projet IMOTEP (Infrarouge, Micro-Ondes et Transfert radiatif ensembliste pour la prévision des Extrêmes de Précipitations) soutenu par le CNES et dont le but est de parvenir à une assimilation en synergie des instruments infrarouges et micro-ondes. Les longueurs d'onde infrarouges étant davantage sensibles aux gouttelettes et aux cristaux de glace et les micro-ondes aux précipitations, la complémentarité de ces deux domaines spectraux permettrait à terme d'approfondir les connaissances sur les incertitudes du transfert radiatif tout en améliorant l'assimilation des observations. Ce projet est décomposé en trois axes de recherches : (i) la création d'un opérateur ensembliste prenant en compte les incertitudes du transfert radiatif dans l'opérateur d'observation, (ii) un travail pour parvenir à une assimilation en synergie entre les micro-ondes et les infrarouges et (iii) un travail de validation sur les prévisions des nuages et des précipitations.

Le premier axe a été traité au sein de ce manuscrit où différentes méthodes traitant des ensembles de SSPs ont été élaborées. Le deuxième axe est actuellement étudié au sein du CNRM dans le cadre de la thèse d'Ethel Villeneuve visant à préparer l'assimilation des futurs instruments à bord des satellites EPS-SG et MTG-I (Meteosat Third Generation - Imager). Une perspective future est ici de combiner les méthodes élaborées au sein de ce manuscrit au travail sur la synergie infrarouges/micro-ondes. Elles seront affinées à l'aide des observations des instruments infrarouges et de même étendues à cette gamme de fréquence.

De plus et afin de vérifier la pertinence des résultats produits au sein des différentes études portant sur les propriétés microphysiques dans l'axe (i) et (ii), les observations provenant des instruments micro-ondes actifs seront parallèlement exploitées au sein de l'axe (iii). La discrétisation verticale très fine des instruments radars sont en effet un bon outil de validation de résultats. Des recherches préliminaires ont été entreprises premièrement afin d'optimiser les simulations des réflectivités radars. Ce travail est en ce moment réalisé dans le cadre d'un projet EUMETSAT au CNRM par Rohit Mangla. Ainsi, une autre perspective sera de valider les résultats des travaux réalisés au sein de cette thèse à l'aide des instruments actifs. Cela contribuera à terme à développer une méthode d'assimilation des instruments actifs/passifs micro-ondes/infrarouges au sein d'un seul et même outil.

Bibliographie

- [1] Geer AJ, Baordo F. Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmospheric Measurement Techniques* 2014;7(6) :1839–1860. <https://amt.copernicus.org/articles/7/1839/2014/>.
- [2] Geer A, Ahlgrimm M, Bechtold P, Bonavita M, Bormann N, English S, et al. Assimilating observations sensitive to cloud and precipitation. *European Centre for Medium-Range Weather Forecasts*; 2017.
- [3] Haddad Z, Steward J, Tseng HC, Vukicevic T, Chen SH, Hristova-Veleva S. A data assimilation technique to account for the nonlinear dependence of scattering microwave observations of precipitation. *Journal of Geophysical Research : Atmospheres* 2015;120(11) :5548–5563.
- [4] Guerbette J, Mahfouf JF, Plu M. Towards the assimilation of all-sky microwave radiances from the SAPHIR humidity sounder in a limited area NWP model over tropical regions. *Tellus A : Dynamic Meteorology and Oceanography* 2016;68(1) :28620.
- [5] Mangla R, Indu J, Chambon P, Mahfouf JF. First steps towards an all-sky assimilation framework for tropical cyclone event over Bay of Bengal region : Evaluation and assessment of GMI radiances. *Atmospheric Research* 2021;257 :105564.
- [6] Baran AJ, Labonnote LC. A self-consistent scattering model for cirrus. I : The solar region. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2007;133(629) :1899–1912.
- [7] Kulie MS, Bennartz R, Greenwald TJ, Chen Y, Weng F. Uncertainties in microwave properties of frozen precipitation : Implications for remote sensing and data assimilation. *Journal of the Atmospheric Sciences* 2010;67(11) :3471–3487.
- [8] Ringerud S, Kulie MS, Randel DL, Skofronick-Jackson GM, Kummerow CD. Effects of ice particle representation on passive microwave precipitation retrieval in a Bayesian scheme. *IEEE Transactions on Geoscience and Remote Sensing* 2019;57(6) :3619–3632.
- [9] Faure G, Chambon P, Brousseau P. Operational Implementation of the AROME Model in the Tropics : Multiscale Validation of Rainfall Forecasts. *Weather and Forecasting* 2020;35(2) :691–710.
- [10] Bouyssel F, Berre L, Bénichou H, Chambon P, Girardot N, Guidard V, et al. The 2020 Global Operational NWP Data Assimilation System at Météo-France. In : *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)* Springer; 2022.p. 645–664.
- [11] Liu G. A Database of Microwave Single-Scattering Properties for Nonspherical Ice Particles. *Bulletin of the American Meteorological Society* 2008;89(10) :1563 – 1570. https://journals.ametsoc.org/view/journals/bams/89/10/2008bams2486_1.xml.
- [12] Saunders R, Hocking J, Turner E, Rayer P, Rundle D, Brunel P, et al. An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development* 2018;11(7) :2717–2737.
- [13] Berre L, Varella H, Desroziers G. Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quarterly Journal of the Royal Meteorological Society* 2015;141(692) :2803–2812.

-
- [14] Berre L. Simulation and diagnosis of observation, model and background error contributions in data assimilation cycling. *Quarterly Journal of the Royal Meteorological Society* 2019;145(719) :597–608.
- [15] Krishnamurti T, Xue J, Bedi H, Ingles K, Oosterhof D. Physical initialization for numerical weather prediction over the tropics. *Tellus B* 1991;43(4) :53–81.
- [16] Marécal V, Mahfouf JF. Variational Retrieval of Temperature and Humidity Profiles from TRMM Precipitation Data. *Monthly Weather Review* 2000;128(11) :3853–3866. [https://doi.org/10.1175/1520-0493\(2001\)129<3853:VROTAH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<3853:VROTAH>2.0.CO;2).
- [17] Marécal V, Mahfouf Jf, Bauer P. Comparison of TMI rainfall estimates and their impact on 4D-Var assimilation. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2002;128(586) :2737–2758.
- [18] Marécal V, Mahfouf JF. Experiments on 4D-Var assimilation of rainfall data using an incremental formulation. *Quarterly Journal of the Royal Meteorological Society* 2003;129(594) :3137–3160. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.02.120>.
- [19] Caumont O, Ducrocq V, Wattrelot É, Jaubert G, Pradier-Vabre S. 1D+ 3DVar assimilation of radar reflectivity data : A proof of concept. *Tellus A : Dynamic Meteorology and Oceanography* 2010;62(2) :173–187.
- [20] Wattrelot E, Caumont O, Mahfouf JF. Operational implementation of the 1D+ 3D-Var assimilation method of radar reflectivity data in the AROME model. *Monthly Weather Review* 2014;142(5) :1852–1873.
- [21] Borderies M, Caumont O, Delanoë J, Ducrocq V, Fourrié N, Marquet P. Impact of airborne cloud radar reflectivity data assimilation on kilometre-scale numerical weather prediction analyses and forecasts of heavy precipitation events. *Natural Hazards and Earth System Sciences* 2019;19(4) :907–926.
- [22] Ikuta Y, Okamoto K, Kubota T. One-dimensional maximum-likelihood estimation for spaceborne precipitation radar data assimilation. *Quarterly Journal of the Royal Meteorological Society* 2021;147(735) :858–875.
- [23] Errico RM, Bauer P, Mahfouf JF. Issues Regarding the Assimilation of Cloud and Precipitation Data. *Journal of the Atmospheric Sciences* 2007;64(11) :3785–3798. <https://doi.org/10.1175/2006JAS2044.1>.
- [24] Geer AJ, Bauer P, Lopez P. Lessons learnt from the operational 1D+ 4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. *Quarterly Journal of the Royal Meteorological Society* 2008;134(635) :1513–1525.
- [25] Bauer P, Geer AJ, Lopez P, Salmond D. Direct 4D-Var assimilation of all-sky radiances. Part I : Implementation. *Quarterly Journal of the Royal Meteorological Society* 2010;136(652) :1868–1885.
- [26] Geer AJ, Bauer P. Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation. ECMWF Reading, UK ; 2010.
- [27] Brandes EA, Zhang G, Vivekanandan J. Experiments in rainfall estimation with a polarimetric radar in a subtropical environment. *Journal of Applied Meteorology* 2002;41(6) :674–685.
- [28] Augros C, Caumont O, Ducrocq V, Gaussiat N, Tabary P. Comparisons between S-, C- and X-band polarimetric radar observations and convective-scale simulations of the HyMeX first special observing period. *Quarterly Journal of the Royal Meteorological Society* 2016;142 :347–362.
- [29] Ekelund R, Eriksson P, Kahnert M. Microwave single-scattering properties of non-spheroidal raindrops. *Atmospheric Measurement Techniques* 2020;13(12) :6933–6944.
- [30] Waterman PC. Symmetry, unitarity, and geometry in electromagnetic scattering. *Physical review D* 1971;3(4) :825.
- [31] Bailey MP, Hallett J. A Comprehensive Habit Diagram for Atmospheric Ice Crystals : Confirmation from the Laboratory, AIRS II, and Other Field Studies. *Journal of the Atmospheric Sciences* 2009;66(9) :2888 – 2899. <https://journals.ametsoc.org/view/journals/atsc/66/9/2009jas2883.1.xml>.

- [32] Hong G. Parameterization of scattering and absorption properties of nonspherical ice crystals at microwave frequencies. *Journal of Geophysical Research : Atmospheres* 2007;112(D11).
- [33] Petty GW, Huang W. Microwave backscatter and extinction by soft ice spheres and complex snow aggregates. *Journal of the Atmospheric Sciences* 2010;67(3) :769–787.
- [34] Tyynelä J, Nousiainen T, Göke S, Muinonen K. Modeling C-band single scattering properties of hydrometeors using discrete-dipole approximation and T-matrix method. *Journal of Quantitative Spectroscopy and Radiative Transfer* 2009;110(14-16) :1654–1664.
- [35] Ori D, Maestri T, Rizzi R, Cimini D, Montopoli M, Marzano FS. Scattering properties of modeled complex snowflakes and mixed-phase particles at microwave and millimeter frequencies. *Journal of Geophysical Research : Atmospheres* 2014;119(16) :9931–9947. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD021616>.
- [36] Leinonen J, Moisseev D. What do triple-frequency radar signatures reveal about aggregate snowflakes? *Journal of Geophysical Research : Atmospheres* 2015;120(1) :229–239. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD022072>.
- [37] Leinonen J, Szyrmer W. Radar signatures of snowflake riming : A modeling study. *Earth and Space Science* 2015;2(8) :346–358. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015EA000102>.
- [38] Kuo KS, Olson WS, Johnson BT, Grecu M, Tian L, Clune TL, et al. The Microwave Radiative Properties of Falling Snow Derived from Nonspherical Ice Particle Models. Part I : An Extensive Database of Simulated Pristine Crystals and Aggregate Particles, and Their Scattering Properties. *Journal of Applied Meteorology and Climatology* 2016;55(3) :691 – 708. <https://journals.ametsoc.org/view/journals/apme/55/3/jamc-d-15-0130.1.xml>.
- [39] Lu Y, Jiang Z, Aydin K, Verlinde J, Clothiaux EE, Botta G. A polarimetric scattering database for non-spherical ice particles at microwave wavelengths. *Atmospheric Measurement Techniques* 2016;9(10) :5119–5134. <https://amt.copernicus.org/articles/9/5119/2016/>.
- [40] Ding J, Bi L, Yang P, Kattawar GW, Weng F, Liu Q, et al. Single-scattering properties of ice particles in the microwave regime : Temperature effect on the ice refractive index with implications in remote sensing. *Journal of Quantitative Spectroscopy and Radiative Transfer* 2017;190 :26–37. <https://www.sciencedirect.com/science/article/pii/S0022407316306367>.
- [41] Eriksson P, Ekelund R, Mendrok J, Brath M, Lemke O, Buehler SA. A general database of hydrometeor single scattering properties at microwave and sub-millimetre wavelengths. *Earth System Science Data* 2018;10(3) :1301–1326.
- [42] Kneifel S, Neto JD, Ori D, Moisseev D, Tyynelä J, Adams IS, et al. Summer snowfall workshop : Scattering properties of realistic frozen hydrometeors from simulations and observations, as well as defining a new standard for scattering databases. *Bulletin of the American Meteorological Society* 2018;99(3) :ES55–ES58.
- [43] Eyre J. A fast radiative transfer model for satellite sounding systems. *ECMWF Tech Memo* 176 1991 ;.
- [44] Duruisseau F, Chambon P, Wattrelot E, Barreyat M, Mahfouf JF. Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Quarterly Journal of the Royal Meteorological Society* 2019;145(719) :620–641.
- [45] Courtier P, Freydier C, Geleyn JF, Rabier F, Rochas M. The Arpege project at Meteo France. In : *Seminar on Numerical Methods in Atmospheric Models*, 9-13 September 1991, vol. II ECMWF, Shinfield Park, Reading : ECMWF ; 1991. p. 193–232. <https://www.ecmwf.int/node/8798>.
- [46] Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres : RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research : Atmospheres* 1997;102(D14) :16663–16682.
- [47] Fouquart Y, Bonnel B. Computations of solar heating of the earth’s atmosphere- A new parameterization. *Beitraege zur Physik der Atmosphaere* 1980;53 :35–62.

-
- [48] Masson V, Le Moigne P, Martin E, Faroux S, Alias A, Alkama R, et al. The SURFEXv7. 2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development* 2013;6(4) :929–960.
- [49] Cuxart J, Bougeault P, Redelsperger JL. A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society* 2000;126(562) :1–30.
- [50] Bougeault P, Lacarrere P. Parameterization of orography-induced turbulence in a mesobeta-scale model. *Monthly weather review* 1989;117(8) :1872–1890.
- [51] Tiedtke M. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather review* 1989;117(8) :1779–1800.
- [52] Bechtold P, Köhler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, et al. Advances in simulating atmospheric variability with the ECMWF model : From synoptic to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2008 ;134(634) :1337–1351.
- [53] Bechtold P, Semane N, Lopez P, Chaboureau JP, Beljaars A, Bormann N. Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of the Atmospheric Sciences* 2014;71(2) :734–753.
- [54] Bechtold P, Bazile E, Guichard F, Mascart P, Richard E. A mass-flux convection scheme for regional and global models. *Quarterly Journal of the Royal Meteorological Society* 2001 ;127(573) :869–886.
- [55] Catry B, Geleyn JF, Bouysse F, Cedilnik J, Brožková R, Derková M, et al. A new sub-grid scale lift formulation in a mountain drag parameterisation scheme. *Meteorologische Zeitschrift* 2008 ;p. 193–208.
- [56] Lopez P. Implementation and validation of a new prognostic large-scale cloud and precipitation scheme for climate and data-assimilation purposes. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2002 ;128(579) :229–257.
- [57] Bouteloup Y, Seity Y, Bazile E. Description of the sedimentation scheme used operationally in all met eo-france nwp models. *Tellus A : Dynamic Meteorology and Oceanography* 2011 ;63(2) :300–311.
- [58] Ritchie H, Temperton C, Simmons A, Hortal M, Davies T, Dent D, et al. Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Monthly Weather Review* 1995 ;123(2) :489–514.
- [59] Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, et al. The AROME-France convective-scale operational model. *Monthly Weather Review* 2011 ;139(3) :976–991.
- [60] Pergaud J, Masson V, Malardel S, Couvreux F. A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-layer meteorology* 2009 ;132(1) :83–106.
- [61] Pinty JP, Jabouille P. A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model : simulations of a squall line and of orographic precipitations. In : *Conf. on Cloud Physics Amer. Meteor. Soc Everett, WA ; 1998. p. 217–220.*
- [62] Riette S. Development of “Physical Parametrizations with PYthon”(PPPY, version 1.1) and its usage to reduce the time-step dependency in a microphysical scheme. *Geoscientific Model Development* 2020 ;13(2) :443–460.
- [63] Caniaux G. Paramétrisation de la phase glace dans un modèle non hydrostatique de nuage : application à une ligne de grains tropicale. PhD thesis ; 1993.
- [64] Kessler E. On the distribution and continuity of water substance in atmospheric circulations. In : *On the distribution and continuity of water substance in atmospheric circulations Springer ; 1969.p. 1–84.*
- [65] Locatelli JD, Hobbs PV. Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research* 1974 ;79(15) :2185–2197.

- [66] Petty GW, Huang W. The modified gamma size distribution applied to inhomogeneous and nonspherical particles : Key relationships and conversions. *Journal of the Atmospheric Sciences* 2011;68(7) :1460–1473.
- [67] Eriksson P, Jamali M, Mendrok J, Buehler SA. On the microwave optical properties of randomly oriented ice hydrometeors. *Atmospheric Measurement Techniques* 2015;8(5) :1913–1933.
- [68] Lafore JP, Stein J, Asencio N, Bougeault P, Ducrocq V, Duron J, et al. The Meso-NH atmospheric simulation system. Part I : Adiabatic formulation and control simulations. In : *Annales geophysicae*, vol. 16 Copernicus GmbH; 1998. p. 90–109.
- [69] Chambon P, Audouin O, Birman C, Fourrie N, Loo C, Mahfouf JF, et al. Global observing system experiments within the Météo-France 4D-Var data assimilation system. *Monthly Weather Review* 2022 ;.
- [70] Tabary P. The New French Operational Radar Rainfall Product. Part I : Methodology. *Weather and Forecasting* 2007;22(3) :393 – 408. https://journals.ametsoc.org/view/journals/wefo/22/3/waf1004_1.xml.
- [71] Karbou F, Gérard É, Rabier F. Microwave land emissivity and skin temperature for amsu-A and-B assimilation over land. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2006;132(620) :2333–2355.
- [72] Kidd C, Huffman G, Maggioni V, Chambon P, Oki R. The Global Satellite Precipitation Constellation : Current Status and Future Requirements. *Bulletin of the American Meteorological Society* 2021 Oct;102(10) :E1844–E1861.
- [73] Langland RH, Baker NL. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A : Dynamic Meteorology and Oceanography* 2004;56(3) :189–201.
- [74] Candy B, Cotton EJ James and. Recent results of observation data denial experiments. *Weather Science Technical Report* 2021 ;641.
- [75] Ingleby B. Data assimilation diagnostics : Assessing the impact of observations on the forecast ; 2022.
- [76] Gandin LS. Objective analysis of meteorological field. *Gidrometeorologicheskoe Izdate'stvo* 1963;286.
- [77] Le Dimet FX, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A : Dynamic Meteorology and Oceanography* 1986;38(2) :97–110.
- [78] Parrish DF, Derber JC. The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review* 1992;120(8) :1747–1763.
- [79] Courtier P, Andersson E, Heckley W, Vasiljevic D, Hamrud M, Hollingsworth A, et al. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I : Formulation. *Quarterly Journal of the Royal Meteorological Society* 1998;124(550) :1783–1807. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712455002>.
- [80] Simmons A. The control of gravity waves in data assimilation 27 April 1999 1999 ;.
- [81] Talagrand O, Courtier P. Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Theory. *Quarterly Journal of the Royal Meteorological Society* 1987;113(478) :1311–1328.
- [82] Courtier P, Talagrand O. Variational assimilation of meteorological observations with the adjoint vorticity equation. II : Numerical results. *Quarterly Journal of the Royal Meteorological Society* 1987;113(478) :1329–1347.
- [83] Courtier P, Thépaut JN, Hollingsworth A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 1994;120(519) :1367–1387.

-
- [84] Courtier P. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society* 1997;123(544) :2449–2461.
- [85] Rabier F, Järvinen H, Klinker E, Mahfouf JF, Simmons A. The ECMWF operational implementation of four-dimensional variational assimilation. I : Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society* 2000;126(564) :1143–1170. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712656415>.
- [86] Mahfouf JF, Rabier F. The ECMWF operational implementation of four-dimensional variational assimilation. II : Experimental results with improved physics. *Quarterly Journal of the Royal Meteorological Society* 2000;126(564) :1171–1190.
- [87] Klinker E, Rabier F, Kelly G, Mahfouf JF. The ECMWF operational implementation of four-dimensional variational assimilation. III : Experimental results and diagnostics with operational configuration. *Quarterly Journal of the Royal Meteorological Society* 2000;126(564) :1191–1215.
- [88] Pereira MB, Berre L. The use of an ensemble approach to study the background error covariances in a global NWP model. *Monthly weather review* 2006;134(9) :2466–2489.
- [89] Hollingsworth A, Lönnberg P. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I : The wind field. *Tellus A* 1986;38(2) :111–136.
- [90] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research : Oceans* 1994;99(C5) :10143–10162.
- [91] Fisher M. Background error covariance modelling. In : *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean* Shinfield Park, Reading ; 2003. p. 45–63.
- [92] Derber J, Bouttier F. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A : Dynamic Meteorology and Oceanography* 1999;51(2) :195–221.
- [93] Descamps L, Labadie C, Joly A, Bazile E, Arbogast P, Cébron P. PEARP, the Météo-France short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society* 2015;141(690) :1671–1685.
- [94] Ulaby FT, Moore RK, Fung AK. Microwave remote sensing active and passive. *Radar remote sensing and surface scattering and emission theory* 1981 ;p. 848–902.
- [95] Matsapey N, Faucheu J, Flury M, Delafosse D. Design of a gonio-spectro-photometer for optical characterization of gonio-apparent materials. *Measurement Science and Technology* 2013;24(6) :065901.
- [96] Haar SKTV. *Satellite Meteorology*. Academic Press ; 12th September 1995.
- [97] Petty GW. *A first course in atmospheric radiation*. Sundog Pub ; 2006.
- [98] Petty G. 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* 1994;54(1) :79–99.
- [99] Liu Q, Simmer C, Ruprecht E. Three-dimensional radiative transfer effects of clouds in the microwave spectral range. *Journal of Geophysical Research : Atmospheres* 1996;101(D2) :4289–4298.
- [100] Szczap F, Alkasem A, Mioche G, Shcherbakov V, Cornet C, Delanoë J, et al. McRALI : a Monte Carlo high-spectral-resolution lidar and Doppler radar simulator for three-dimensional cloudy atmosphere remote sensing. *Atmospheric Measurement Techniques* 2021;14(1) :199–221. <https://amt.copernicus.org/articles/14/199/2021/>.
- [101] Joseph JH, Wiscombe W, Weinman J. The delta-Eddington approximation for radiative flux transfer. *Journal of Atmospheric Sciences* 1976;33(12) :2452–2459.
- [102] Kummerow C. On the accuracy of the Eddington approximation for radiative transfer in the microwave frequencies. *Journal of Geophysical Research : Atmospheres* 1993;98(D2) :2757–2765.

- [103] Bauer P, Moreau E, Chevallier F, O'keeffe U. Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2006 ;132(617) :1259–1281.
- [104] Liu G. A fast and accurate model for microwave radiance calculations. *Journal of the Meteorological Society of Japan Ser II* 1998 ;76(2) :335–343.
- [105] Liu Q, Weng F. A microwave polarimetric two-stream radiative transfer model. *Journal of the atmospheric sciences* 2002 ;59(15) :2396–2402.
- [106] Weinman J, Guetter PJ. Determination of rainfall distributions from microwave radiation measured by the Nimbus 6 ESMR. *Journal of Applied Meteorology and Climatology* 1977 ;16(4) :437–442.
- [107] Wendisch M, von Hoyningen-Huene W. High speed version of the method of 'successive order of scattering' and its application to remote sensing. *Beitraege zur Physik der Atmosphaere (ISSN 0005-8173)* 1991 ;64 :83–91.
- [108] Morcrette JJ, Fouquart Y. The overlapping of cloud layers in shortwave radiation parameterizations. *Journal of the atmospheric sciences* 1986 ;43(4) :321–328.
- [109] Manabe S, Strickler RF. Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of the Atmospheric Sciences* 1964 ;21(4) :361–385.
- [110] Geleyn J, et al. An economical analytical method for the computation of the interaction between scattering and line absorption of radiation 1979 ;
- [111] Hogan RJ, Illingworth AJ. Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society* 2000 ;126(569) :2903–2909. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712656914>.
- [112] Geer AJ, Bauer P, O'Dell CW. A revised cloud overlap scheme for fast microwave radiative transfer in rain and cloud. *Journal of applied meteorology and climatology* 2009 ;48(11) :2257–2270.
- [113] Liebe HJ, Manabe T, Hufford GA. Millimeter-wave attenuation and delay rates due to fog/cloud conditions. *IEEE transactions on antennas and propagation* 1989 ;37(12) :1617–1612.
- [114] Chevallier F, Di Michele S, McNally AP. Diverse profile datasets from the ECMWF 91-level short-range forecasts. *European Centre for Medium-Range Weather Forecasts* ; 2006.
- [115] Liu Q, Weng F, English SJ. An improved fast microwave water emissivity model. *IEEE Transactions on Geoscience and Remote Sensing* 2010 ;49(4) :1238–1250.
- [116] Bormann N, Geer A, English S. Evaluation of the microwave ocean surface emissivity model FASTEM-5 in the IFS. *ECMWF Reading, UK* ; 2012.
- [117] Kazumori M, English SJ. Use of the ocean surface wind direction signal in microwave radiance assimilation. *Quarterly Journal of the Royal Meteorological Society* 2015 ;141(689) :1354–1375.
- [118] Aires F, Prigent C, Bernardo F, Jiménez C, Saunders R, Brunel P. A Tool to Estimate Land-Surface Emissivities at Microwave frequencies (TELSEM) for use in numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society* 2011 ;137(656) :690–699.
- [119] Karbou F, Gérard E, Rabier F. Global 4DVAR assimilation and forecast experiments using AMSU observations over land. Part I : Impacts of various land surface emissivity parameterizations. *Weather and Forecasting* 2010 ;25(1) :5–19.
- [120] Shahabadi MB, Aparicio JM, Garand L. Impact of slant-path radiative transfer in the simulation and assimilation of satellite radiances in Environment Canada's weather forecast system. *Monthly Weather Review* 2018 ;146(12) :4357–4372.
- [121] Battaglia A, Ajewole M, Simmer C. Evaluation of radar multiple-scattering effects from a GPM perspective. Part I : Model description and validation. *Journal of applied meteorology and climatology* 2006 ;45(12) :1634–1647.

- [122] Battaglia A, Tanelli S, Mroz K, Tridon F. Multiple scattering in observations of the GPM dual-frequency precipitation radar : Evidence and impact on retrievals. *Journal of Geophysical Research : Atmospheres* 2015 ;120(9) :4090–4101.
- [123] Marshall JS, Palmer WMK. THE DISTRIBUTION OF RAINDROPS WITH SIZE. *Journal of Atmospheric Sciences* 1948 ;5(4) :165 – 166. https://journals.ametsoc.org/view/journals/atasc/5/4/1520-0469_1948_005_0165_tdorws_2_0_co_2.xml.
- [124] Abel S, Boutle I. An improved representation of the raindrop size distribution for single-moment microphysics schemes. *Quarterly Journal of the Royal Meteorological Society* 2012 ;138(669) :2151–2162.
- [125] McFarquhar GM, Heymsfield AJ. Parameterization of tropical cirrus ice crystal size distributions and implications for radiative transfer : Results from CEPEX. *Journal of the atmospheric sciences* 1997 ;54(17) :2187–2200.
- [126] Field P, Hogan R, Brown P, Illingworth A, Choullarton T, Cotton R. Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2005 ;131(609) :1997–2017.
- [127] Field PR, Heymsfield AJ, Bansemmer A. Snow size distribution parameterization for midlatitude and tropical ice clouds. *Journal of the Atmospheric Sciences* 2007 ;64(12) :4346–4365.
- [128] Heymsfield AJ, Schmitt C, Bansemmer A. Ice cloud particle size distributions and pressure-dependent terminal velocities from in situ observations at temperatures from 0 to- 86 C. *Journal of the Atmospheric Sciences* 2013 ;70(12) :4123–4154.
- [129] Eriksson P, Ekström M, Rydberg B, Murtagh DP. First Odin sub-mm retrievals in the tropical upper troposphere : ice cloud properties. *Atmospheric Chemistry and Physics* 2007 ;7(2) :471–483.
- [130] Mangla R, Borderies M, Chambon P, Geer A. Sensitivity studies of space-borne precipitation radar simulations within the RTTOV-SCATT radiative transfer model 2023 ;.
- [131] Bauer P, Microwave radiative transfer modeling in clouds and precipitation - Part I; 2004. <https://www.ecmwf.int/node/16944>.
- [132] Mie G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen der physik* 1908 ;330(3) :377–445.
- [133] Rayleigh L. On the diffraction of light by spheres of small relative index. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character* 1914 ;90(617) :219–225.
- [134] Light scattering by small particles. By H. C. Van de Hulst. New York (John Wiley and Sons), London (Chapman and Hall), 1957. Pp. xiii, 470 ; 103 Figs. ; 46 Tables. 96s. *Quarterly Journal of the Royal Meteorological Society* 1957 ;84(360) :198–199. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49708436025>.
- [135] Lawson RP, Stewart RE, Angus LJ. Observations and numerical simulations of the origin and development of very large snowflakes. *Journal of the atmospheric sciences* 1998 ;55(21) :3209–3229.
- [136] Lawson RP, Baker BA, Schmitt CG, Jensen T. An overview of microphysical properties of Arctic clouds observed in May and July 1998 during FIRE ACE. *Journal of Geophysical Research : Atmospheres* 2001 ;106(D14) :14989–15014.
- [137] Heymsfield AJ, Lewis S, Bansemmer A, Jaquinta J, Miloshevich LM, Kajikawa M, et al. A general approach for deriving the properties of cirrus and stratiform ice cloud particles. *Journal of the atmospheric sciences* 2002 ;59(1) :3–29.
- [138] Garrett T, Fallgatter C, Shkurko K, Howlett D. Fall speed measurement and high-resolution multi-angle photography of hydrometeors in free fall. *Atmospheric Measurement Techniques* 2012 ;5(11) :2625–2633.

- [139] Newman AJ, Kucera PA, Bliven LF. Presenting the snowflake video imager (SVI). *Journal of Atmospheric and Oceanic technology* 2009 ;26(2) :167–179.
- [140] Voshchinnikov N, Farafonov V. Optical properties of spheroidal particles. *Astrophysics and space science* 1993 ;204(1) :19–86.
- [141] OGUCHI T. Scattering properties of oblate raindrops and cross polarization of radio waves due to rain- Calculations at 19. 3 and 34. 8 GHz. *Radio Research Laboratories, Journal* 1973 ;20(102) :79–118.
- [142] Asano S, Yamamoto G. Light scattering by a spheroidal particle. *Applied optics* 1975 ;14(1) :29–49.
- [143] Xu YL. Electromagnetic scattering by an aggregate of spheres. *Applied optics* 1995 ;34(21) :4573–4588.
- [144] Yee K. Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media. *IEEE Transactions on Antennas and Propagation* 1966 ;14(3) :302–307.
- [145] Yang P, Ding J, Panetta RL, Liou KN, Kattawar GW, Mishchenko M. On the convergence of numerical computations for both exact and approximate solutions for electromagnetic scattering by nonspherical dielectric particles. *Electromagnetic waves (Cambridge, Mass)* 2019 ;164 :27.
- [146] Johnson BR. Invariant imbedding T matrix approach to electromagnetic scattering. *Appl Opt* 1988 Dec ;27(23) :4861–4873. <http://opg.optica.org/ao/abstract.cfm?URI=ao-27-23-4861>.
- [147] Mackowski DW. Discrete dipole moment method for calculation of the T matrix for nonspherical particles. *J Opt Soc Am A* 2002 May ;19(5) :881–893. <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-19-5-881>.
- [148] Purcell EM, Pennypacker CR. Scattering and absorption of light by nonspherical dielectric grains. *The Astrophysical Journal* 1973 ;186 :705–714.
- [149] Draine BT, Flatau PJ. Discrete-Dipole Approximation For Scattering Calculations. *J Opt Soc Am A* 1994 Apr ;11(4) :1491–1499. <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-11-4-1491>.
- [150] Draine BT, Flatau PJ. User guide for the discrete dipole approximation code DDSCAT 7.3. arXiv preprint arXiv :13056497 2013 ;.
- [151] Yurkin MA, Hoekstra AG. The discrete-dipole-approximation code ADDA : Capabilities and known limitations. *Journal of Quantitative Spectroscopy and Radiative Transfer* 2011 ;112(13) :2234–2247. <https://www.sciencedirect.com/science/article/pii/S0022407311000562>, polarimetric Detection, Characterization, and Remote Sensing.
- [152] Gans R. Strahlungsdiagramme ultramikroskopischer Teilchen. *Annalen der Physik* 1925 ;381(1) :29–38.
- [153] Debye P. Zerstreung von röntgenstrahlen. *Annalen der Physik* 1915 ;351(6) :809–823.
- [154] Born M. Quantenmechanik der stoßvorgänge. *Zeitschrift für physik* 1926 ;38(11) :803–827.
- [155] Hogan RJ, Westbrook CD. Equation for the microwave backscatter cross section of aggregate snowflakes using the self-similar Rayleigh–Gans approximation. *Journal of the Atmospheric Sciences* 2014 ;71(9) :3292–3301.
- [156] Hogan RJ, Honeyager R, Tyynelä J, Kneifel S. Calculating the millimetre-wave scattering phase function of snowflakes using the self-similar Rayleigh–Gans Approximation. *Quarterly Journal of the Royal Meteorological Society* 2017 ;143(703) :834–844.
- [157] Le Dimet FX, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A : Dynamic Meteorology and Oceanography* 1986 ;38(2) :97–110.
- [158] Kuo HL. Further studies of the parameterization of the influence of cumulus convection on large-scale flow. *Journal of Atmospheric Sciences* 1974 ;31(5) :1232–1240.

-
- [159] Jones C, Macpherson B. A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Applications* 1997;4(3) :269–277.
- [160] Županski D, Mesinger F. Four-dimensional variational assimilation of precipitation data. *Monthly Weather Review* 1995;123(4) :1112–1127.
- [161] Zou X, Kuo Y. Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Monthly Weather Review* 1996;124(12) :2859–2882.
- [162] Tsuyuki T. Variational data assimilation in the Tropics using precipitation data. Part II : 3D model. *Monthly weather review* 1996;124(11) :2545–2561.
- [163] Tsuyuki T. Variational data assimilation in the Tropics using precipitation data. Part III : Assimilation of SSM/I precipitation rates. *Monthly weather review* 1997;125(7) :1447–1464.
- [164] Gerard E, Saunders RW. Four-dimensional variational assimilation of special sensor microwave/imager total column water vapour in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society* 1999;125(560) :3077–3101. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712556014>.
- [165] Lopez P, Moreau E. A convection scheme for data assimilation : Description and initial tests. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2005;131(606) :409–436.
- [166] Tompkins AM, Janisková M. A cloud scheme for data assimilation : Description and initial tests. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2004;130(602) :2495–2517.
- [167] Moreau E, Bauer P, Chevallier F. Variational retrieval of rain profiles from spaceborne passive microwave radiance observations. *Journal of Geophysical Research : Atmospheres* 2003;108(D16). <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD003315>.
- [168] Hólm E, Andersson E, Beljaars A, Lopez P, Mahfouf JF, Simmons A, et al. Assimilation and modelling of the hydrological cycle : ECMWF’s status and plans. *Technical Memoranda* 2002;383.
- [169] Bauer P, Moreau E, Chevallier F, O’keeffe U. Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2006;132(617) :1259–1281.
- [170] Hou AY, Ledvina DV, da Silva AM, Zhang SQ, Joiner J, Atlas RM, et al. Assimilation of SSM/I-Derived Surface Rainfall and Total Precipitable Water for Improving the GEOS Analysis for Climate Studies. *Monthly Weather Review* 2000;128(3) :509 – 537.
- [171] Hou AY, Zhang SQ, da Silva AM, Olson WS. Improving Assimilated Global Datasets Using TMI Rainfall and Columnar Moisture Observations. *Journal of Climate* 2000;13(23).
- [172] Hou AY, Zhang SQ. Assimilation of Precipitation Information Using Column Model Physics as a Weak Constraint. *Journal of the Atmospheric Sciences* 2007;64(11) :3865 – 3878. <https://journals.ametsoc.org/view/journals/atasc/64/11/2006jas2028.1.xml>.
- [173] Derber JC. A variational continuous assimilation technique. *Monthly weather review* 1989;117(11) :2437–2446.
- [174] Bloom S, Takacs L, Da Silva A, Ledvina D. Data assimilation using incremental analysis updates. *Monthly Weather Review* 1996;124(6) :1256–1271.
- [175] Hou AY, Zhang SQ, da Silva AM, Olson WS, Kummerow CD, Simpson J. Improving Global Analysis and Short-Range Forecast Using Rainfall and Moisture Observations Derived from TRMM and SSM/I Passive Microwave Sensors. *Bulletin of the American Meteorological Society* 2001;82(4) :659 – 680. https://journals.ametsoc.org/view/journals/bams/82/4/1520-0477_2001_082_0659_igaasf_2_3_co_2.xml.
- [176] Geer AJ, Bauer P. Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society* 2011;137(661) :2024–2037.

- [177] Zupanski D, Zhang SQ, Zupanski M, Hou AY, Cheung SH. A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations. *Journal of Hydrometeorology* 2011 ;12(1) :118–134.
- [178] Zhang SQ, Zupanski M, Hou AY, Lin X, Cheung SH. Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Monthly Weather Review* 2013;141(2) :754–772.
- [179] Chambon P, Zhang SQ, Hou AY, Zupanski M, Cheung S. Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 2014;140(681) :1219–1235.
- [180] Zupanski M. Maximum likelihood ensemble filter : Theoretical aspects. *Monthly Weather Review* 2005;133(6) :1710–1726.
- [181] Zupanski M, Navon IM, Zupanski D. The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2008;134(633) :1039–1050.
- [182] Janisková M, Lopez P. Linearized physics for data assimilation at ECMWF. In : *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)* Springer ; 2013.p. 251–286.
- [183] Kummerow C, Hong Y, Olson W, Yang S, Adler R, McCollum J, et al. The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology* 2001 ;40(11) :1801–1820.
- [184] Randel DL, Kummerow CD, Ringerud S. The Goddard Profiling (GPROF) precipitation retrieval algorithm. In : *Satellite precipitation measurement* Springer ; 2020.p. 141–152.
- [185] Geer AJ. Physical characteristics of frozen hydrometeors inferred with parameter estimation. *Atmospheric Measurement Techniques* 2021 ;14(8) :5369–5395. <https://amt.copernicus.org/articles/14/5369/2021/>.
- [186] Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, et al. The global precipitation measurement mission. *Bulletin of the American Meteorological Society* 2014 ;95(5) :701–722.
- [187] Bormann N, Saarinen S, Kelly G, Thépaut JN. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review* 2003 ;131(4) :706–718.
- [188] Desroziers G, Berre L, Chapnik B, Poli P. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society : A journal of the atmospheric sciences, applied meteorology and physical oceanography* 2005 ;131(613) :3385–3396.
- [189] Geer AJ. Correlated observation error models for assimilating all-sky infrared radiances. *Atmospheric Measurement Techniques* 2019 ;12(7) :3629–3657.
- [190] Coopmann O, Guidard V, Fourrié N, Josse B, Marécal V. Update of Infrared Atmospheric Sounding Interferometer (IASI) channel selection with correlated observation errors for numerical weather prediction (NWP). *Atmospheric Measurement Techniques* 2020 ;13(5) :2659–2680.
- [191] Brogniez H, Clain G, Roca R. Validation of upper-tropospheric humidity from SAPHIR on board Megha-Tropiques using tropical soundings. *Journal of Applied Meteorology and Climatology* 2015 ;54(4) :896–908.
- [192] Chambon P, Meunier LF, Guillaume F, Piriou JM, Roca R, Mahfouf JF. Investigating the impact of the water-vapour sounding observations from SAPHIR on board Megha-Tropiques for the ARPEGE global model. *Quarterly Journal of the Royal Meteorological Society* 2015 ;141(690) :1769–1779.
- [193] Ikuta Y, Honda Y. Development of 1D+ 4DVAR data assimilation of radar reflectivity in JNoVA. *CAS/JSC WGNE Res Activ Atmos Oceanic Modell* 2011 ;41 :01–09.

-
- [194] Borderies M, Caumont O, Augros C, Bresson É, Delanoë J, Ducrocq V, et al. Simulation of W-band radar reflectivity for model validation and data assimilation. *Quarterly Journal of the Royal Meteorological Society* 2018;144(711) :391–403.
- [195] Gauthier P, Du P, Heilliette S, Garand L. Convergence issues in the estimation of interchannel correlated observation errors in infrared radiance data. *Monthly Weather Review* 2018;146(10) :3227–3239.
- [196] Kummerow C. GPM GMI (GPROF) Radiometer Precipitation Profiling L2A 1.5 hours 13 km V05. Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC) 2016;96(D4) :7423–7435.
- [197] Grody NC. Classification of snow cover and precipitation using the Special Sensor Microwave Imager. *Journal of Geophysical Research : Atmospheres* 1991;96(D4) :7423–7435.
- [198] Wilheit T, Kummerow CD, Ferraro R. NASDA Rainfall algorithms for AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing* 2003;41(2) :204–214.
- [199] Kubota T, Shige S, Hashizume H, Aonashi K, Takahashi N, Seto S, et al. Global precipitation map using satellite-borne microwave radiometers by the GSMaP project : Production and validation. *IEEE Transactions on Geoscience and Remote Sensing* 2007;45(7) :2259–2275.
- [200] Kubota T, Aonashi K, Ushio T, Shige S, Takayabu YN, Kachi M, et al. Global Satellite Mapping of Precipitation (GSMaP) products in the GPM era. In : *Satellite precipitation measurement* Springer ; 2020.p. 355–373.
- [201] Roberts NM, Lean HW. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review* 2008;136 :78–97.
- [202] Duncan DI, Bormann N, Hólm EV. On the addition of microwave sounders and numerical weather prediction skill. *Quarterly Journal of the Royal Meteorological Society* 2021;147(740) :3703–3718. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4149>.
- [203] Petty GW, Katsaros KB. Precipitation observed over the South China Sea by the Nimbus-7 scanning multichannel microwave radiometer during winter MONEX. *Journal of Applied Meteorology and Climatology* 1990;29(4) :273–287.
- [204] Berre L, Ștefaănescu SE, Pereira MB. The representation of the analysis effect in three error simulation techniques. *Tellus A : Dynamic Meteorology and Oceanography* 2006;58(2) :196–209.
- [205] Desroziers G, Berre L, Chabot V, Chapnik B. A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review* 2009;137(10) :3420–3436.
- [206] Daley R. *Atmospheric Data Analysis*; 1991.
- [207] Raynaud L, Berre L, Desroziers G. Accounting for model error in the Météo-France ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 2012;138(662) :249–262.
- [208] Rabier F, McNally A, Andersson E, Courtier P, Undén P, Eyre J, et al. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II : Structure functions. *Quarterly Journal of the Royal Meteorological Society* 1998;124(550) :1809–1829.
- [209] Lien GY, Miyoshi T, Kalnay E. Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecast system. *Monthly Weather Review* 2016;144(2) :643–661.
- [210] Geer AJ, Lonitz K, Weston P, Kazumori M, Okamoto K, Zhu Y, et al. All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society* 2018;144(713) :1191–1217. <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3202>.

Annexe

Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model

Fabrice Duruisseau | Philippe Chambon  | Eric Wattrelot | Marylis Barreyat |

Jean-François Mahfouf

CNRM UMR 3589, Météo-France and CNRS,
France**Correspondence**Philippe Chambon, CNRM UMR 3589,
Météo-France & CNRS 42 avenue Gaspard
Coriolis 31057 Toulouse, France.
Email: philippe.chambon@meteo.fr

The Megha-Tropiques satellite was launched in 2011 with a microwave sounder called SAPHIR onboard. This instrument probes the atmosphere with six channels around the 183.31 GHz water vapour absorption band. Its observations are sensitive to water vapour as well as to hydrometeors. This instrument was proven to be useful for data assimilation by different numerical weather prediction centres, in particular for clear-sky assimilation. At Météo-France, SAPHIR observations have been routinely assimilated in clear sky since 2015 in the ARPEGE global model. The present article introduces a framework to complement this clear-sky assimilation route by a new cloudy and rainy assimilation route for satellite microwave brightness temperatures. This framework is based on several steps including a Bayesian inversion of the SAPHIR brightness temperatures into relative humidity retrievals, which are then assimilated within the ARPEGE global model. This study presents the methodology of assimilation, including the development of two error models, one for the Bayesian inversion, and one for the observation errors of relative humidity retrievals within the ARPEGE 4D-Var data assimilation system. The forecast scores obtained with this methodology over a three-month period indicate a positive impact of SAPHIR cloudy and rainy observations within the ARPEGE system, in particular on tropical temperature and wind forecasts for which the improvements range from 0.5 to 1.7% on standard deviations with respect to the ECMWF analysis and up to a +60 h lead time.

KEYWORDS

clouds, data assimilation, Megha-Tropiques, microwave, NWP, radiative transfer, satellite, Tropics

1 | INTRODUCTION

The Megha-Tropiques satellite mission is an Indo-French mission led by the Indian Space Research Organization (ISRO) and Centre National d'Etudes Spatiales (CNES). It was launched in 2011 with a payload of three instruments dedicated to the study of the water and energy cycles in the Tropics. A summary of the outcomes of the mission 3 years after launch is summarized in Roca *et al.* (2015). One of the three instruments is a six-channel microwave radiometer called

Sondeur Atmosphérique du Profil d'Humidité Intertropicale par Radiométrie (SAPHIR) which probes the atmosphere along the 183.31 GHz water vapour absorption band. This instrument has been well calibrated and validated (Clain *et al.*, 2015) and is characterized by very stable biases and random errors over the past 7 years. From a numerical weather prediction (NWP) perspective, the observations from SAPHIR have proven to be useful and are now assimilated operationally by several weather services over the globe (e.g. Météo-France see Chambon *et al.*, 2015; UK Met Office see Doherty *et al.*,

2018; the Japanese Meteorological Agency see Kazumori, 2016; the Korean Meteorological Administration see Lee *et al.*, 2018; the European Centre for Medium-range Weather Forecasts (ECMWF) see Chambon and Geer, 2017; the National Center for Environmental Prediction (NCEP) and the Fleet Numerical Meteorology and Oceanography Center (FNMOC) see Jones *et al.*, 2017). At Météo-France, SAPHIR clear-sky observations have been routinely used since April 2015 within the four-dimensional variation (4D-Var) data assimilation system of the global model Action des Recherche Petite Echelle et Grande Echelle (ARPEGE) which provides global forecasts up to 4 days ahead (see Chambon *et al.* (2015) for details about the configuration of the clear-sky assimilation and the resulting impacts on ARPEGE forecasts). The subject of the present article is to describe the assimilation method of SAPHIR radiances in cloudy and rainy conditions which is currently under study at Météo-France to complement the clear-sky assimilation scheme.

Since the seminal paper of Errico *et al.* (2007), the assimilation of satellite observations in cloudy and rainy conditions is becoming increasingly mature in different operational weather services over the globe (e.g. Okamoto *et al.*, 2014; Zhu *et al.*, 2016; Geer *et al.*, 2017; Migliorini *et al.*, 2018; see Geer *et al.* (2018) for an extensive review of the status of these developments in several NWP centres) as well as in different research institutes (e.g. Zhang *et al.*, 2013, 2016, 2017; Haddad *et al.*, 2015; Aonashi *et al.*, 2016). In particular, ECMWF pioneered this field of research and paved the way for other centres to develop such capability. A first assimilation method, so-called “1D-Var + 4D-Var” was first implemented at ECMWF (Marécal and Mahfouf, 2002; Bauer *et al.*, 2006a; 2006b; Geer *et al.*, 2008) and further replaced by a direct assimilation method, so-called “all-sky” framework which directly assimilates both clear and cloudy brightness temperatures within the ECMWF 4D-Var system through a single framework (Bauer *et al.*, 2010; Geer *et al.*, 2010; 2017).

At Météo-France, while the spaceborne microwave observations are only assimilated in clear sky operationally, a dedicated method was developed to assimilate ground radar reflectivities in cloudy and rainy conditions (Caumont *et al.*, 2010) within the Applications of Research to Operations at Mesoscale (AROME) convective-scale model running over western Europe (Seity *et al.*, 2011; Brousseau *et al.*, 2016). This method, so-called “1D-Bayesian + 3D-Var”, has been successfully used operationally since 2010 (Wattrelot *et al.*, 2014). It is an alternative to both “1D-Var + 4D-Var” and direct assimilation methodologies. This framework has the specificity of not requiring a linearized convection scheme compared to a direct assimilation method; a linearized scheme that the ARPEGE data assimilation system is not equipped with. A similar “1D-Bayesian + 4D-Var” was implemented at the Japanese Meteorological Agency (Ikuta and Honda, 2011) for operationally assimilating observations from the Dual Frequency Precipitation Radar on board the Global Precipitation Measurement mission (Hou *et al.*, 2014). However,

TABLE 1 List of channels of SAPHIR on board the Megha-Tropiques satellite

Channel number	Central frequency (GHz)	Channel bandwidth (MHz)
1	183.31 ± 0.2	200 MHz
2	183.31 ± 1.1	350 MHz
3	183.31 ± 2.8	500 MHz
4	183.31 ± 4.2	700 MHz
5	183.31 ± 6.8	1,200 MHz
6	183.31 ± 11.0	2000 MHz

this kind of methodology has not been experimented yet for passive microwave observations into a global data assimilation system. Following the proof of concept of 1D-Bayesian inversion shown in Guerbet *et al.* (2016), the present article focuses on studying the use of such a “1D-Bayesian +4D-Var” framework for assimilating SAPHIR brightness temperatures in cloudy and rainy conditions in the global model ARPEGE.

The article is organized as follows: section 2 describes the SAPHIR instrument as well as the NWP environment used, section 3 introduces the proposed methodology of assimilation including (a) the microwave observations simulation, (b) the Bayesian inversion of brightness temperature, and (c) the errors specification. Section 4 presents the experimental set-up and the most important results, then section 5 summarizes the outcomes of the study and provides a number of perspectives to this work.

2 | DATA AND NWP SYSTEM

2.1 | The SAPHIR/Megha-Tropiques observing system

As mentioned in the introduction, the SAPHIR sounder has six channels around the 183.31 GHz water vapour absorption band (see Table 1 for the list of frequencies and the channels’ bandwidths). Channel 1 is the highest peaking channel while channel 6 is the lowest peaking channel and also the most sensitive to hydrometeors. At Météo-France, all six channels are assimilated in clear-sky conditions over oceans, and over land (note that over high terrain, the lowest peaking channels are discarded).

This cross-track sounder is on board the Megha-Tropiques satellite which is rotating around the Earth on a low-inclination orbit of 20° with respect to the Equator and at 866 km altitude. The combination of this tropical orbit and the instrument characteristics lead to a swath of 1,726 km for SAPHIR and a daily revisit varying from 3 to 6 overpasses per day for a given location in the latitude band 20°S to 20°N (Capderou, 2009).

2.2 | Overview of the NWP system

The ARPEGE prediction system, operational at Météo-France since 1992 (Courtier *et al.*, 1991), is a global

TABLE 2 Observation types assimilated within the ARPEGE 4D-Var system

Measurement	Observation type
Surface	Surface stations, ships and buoys
	Terrestrial GNSS data
	Wind profilers
Altitude	Radiosonde measurements
	Aircraft measurements
Satellite	Atmospheric motion vectors from Meteosat, GOES, Himawari
	Scatterometer winds from ASCAT
	Microwave sounding data from AMSU-A, MHS, SAPHIR, MWHS2, ATMS, SSMI/S, GMI
	Hyperspectral InfraRed data from IASI (×2), CrIS
	Clear-sky radiances from Meteosat, GOES, Himawari
	GNSS radio occultation data

spectral model which is characterized by a stretched and tilted grid (Courtier and Geleyn, 1988). The current (2018) operational suite has a spectral triangular truncation on linear grid at wave number 1,198 with a stretching factor of 2.2. Its vertical grid is made of 105 levels from the surface to 0.1 hPa and including 20 levels in the boundary layer. This setting leads to a horizontal resolution varying from 7.5 km over Europe to 36.7 km at the antipodes.

This forecast model is initialised from analyses generated by a 4D-Var incremental data assimilation system with a 6 h time window (Courtier *et al.*, 1994) which uses background errors of the day from an ensemble data assimilation system (Pereira and Berre, 2006). The control variables of this system are vorticity, divergence, temperature, surface pressure and specific humidity. Observations from a wide variety of sensors are assimilated routinely within this system as displayed in Table 2. Therefore, the experiments presented in this article consist of assimilating cloudy and rainy radiances from the SAPHIR sounder, on the top of an already quite dense observing system (~30 million assimilated observations per day).

The main difference between the NWP system used in this study and the operational one at Météo-France lies in the deep moist convection parametrization. While a mass-flux scheme derived from Bougeault (1985) is used in operations, moist convection is described by the “Prognostic Condensates Microphysics and Transport (PCMT) scheme” which has the specificity to consider prognostic convective hydrometeors. In particular, this scheme leads to a better representation of the diurnal cycle of convection in the Tropics as well as better tropical rainfall forecasts with respect to the current ARPEGE operational scheme (not shown).

3 | METHOD

3.1 | General framework for SAPHIR data assimilation

Within the ARPEGE framework (version cy42op2), the RTTOV V11.1 package (<https://www.nwpsaf.eu>) is the radiative transfer code used for simulating the six channels of

the SAPHIR sounder; these simulations are performed without taking into account scattering effects from hydrometeors. Therefore, cloudy scenes are filtered from the assimilation with a simple quality control detailed in Chambon *et al.* (2015). If this method does filter a fraction of observed clouds from the clear-sky assimilation route, one of its drawbacks is that cloudy scenes present in the first guess remain and are still assimilated with a clear-sky observation operator. When complemented by an assimilation of cloudy observations, it was considered as one of the weaknesses of the “1D-Var + 4D-Var” approach implemented at ECMWF before being replaced by the all-sky framework (Geer *et al.*, 2008).

The framework used in this study, which partly corrects this asymmetric sampling, is described below and shown in Figure 1:

- As a first step, it does split observations into two categories, the first one treated by the clear-sky assimilation route, and the second one treated by a new “1D-Bayesian + 4D-Var” route. To achieve this first step, the Radiative Transfer for TIROS Operational Vertical sounder (RTTOV-SCATT) module (Bauer *et al.*, 2006c; Geer and Baordo, 2014) is used to provide simulated brightness temperatures taking into account scattering effects to the screening procedure of observations. With these new simulations, a symmetric quality control is then applied to split the observations into two categories; only the meteorological scenes, for which both the observations and the first guess are considered as clear, go through the clear-sky assimilation route. The other cases of meteorological scenes for which either the observation, or the first guess, or both of them are cloudy, are then treated through the “1D-Bayesian + 4D-Var” route.
- As a second step, a 1D-Bayesian scheme retrieves relative humidity estimates from the observed brightness temperatures (TBs) and from *a priori* simulated TBs coming from the first guess in the vicinity of the observations.

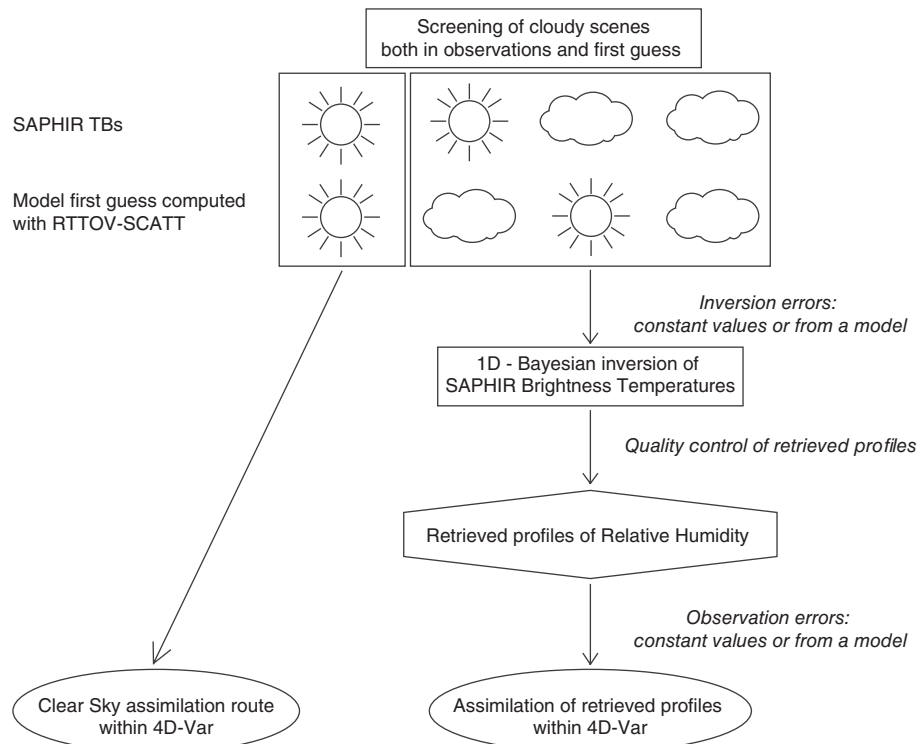


FIGURE 1 Flowchart of the 1D-Bayesian + 4D-Var framework applied to microwave SAPHIR TBs

- As a third and final step, the retrieved relative humidity estimates are assimilated within the 4D-Var with specific observation errors.

Section 3.2 describes the settings selected for the RTTOV-SCATT radiative transfer calculations as well as the cloud screening procedure, then section 3.3 explains the principle of the Bayesian inversion of brightness temperatures into relative humidity estimates. Finally, section 3.4 explains the specification of the observation errors for the assimilation of relative humidity retrievals.

3.2 | Cloudy and rainy microwave observations simulations

As previously explained, this assimilation framework requires the simulation of brightness temperatures taking into account the contribution of hydrometeors, in particular of frozen hydrometeors in addition to the contribution from other water species and the surface. The recent implementation of advanced single-scattering properties for solid hydrometeors within the RTTOV-SCATT radiative transfer code (Geer and Baordo, 2014) leads to more realistic simulations at high microwave frequencies (above 50 GHz) and in particular at the ones of the SAPHIR instrument (Guerbette *et al.*, 2016; Chambon and Geer, 2017). This observation operator is therefore used in the present article.

A detailed sensitivity study was performed by Geer and Baordo (2014) for the simulation of observations ranging between 19 and 183 GHz to finely tune both the particle size distribution and the single-scattering radiative

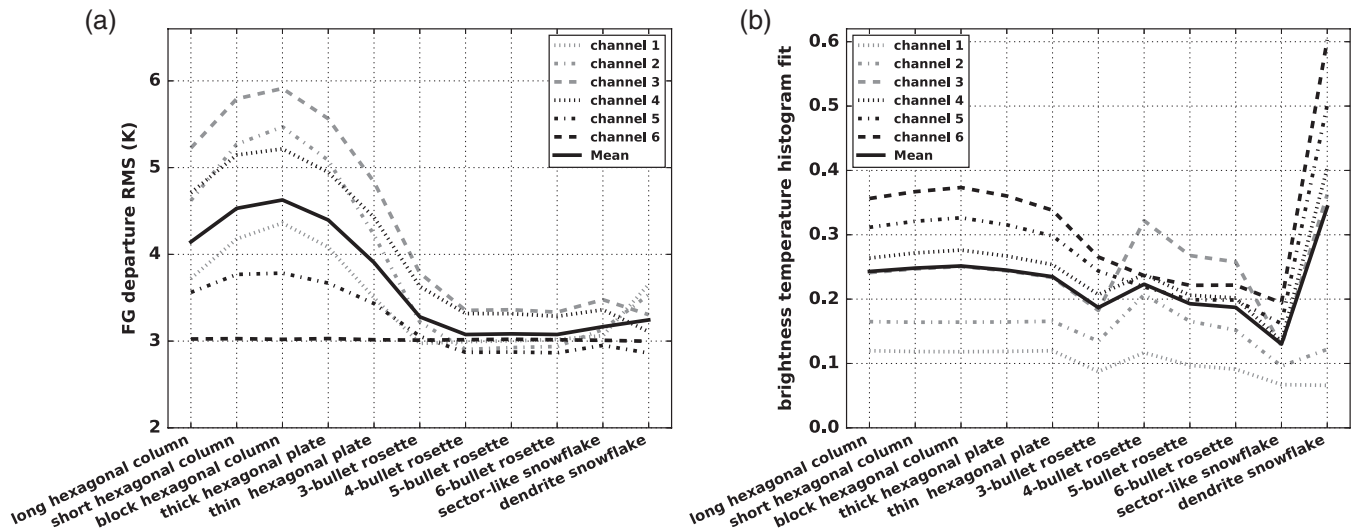
properties of hydrometeors. In particular, a fine search was conducted to derive an optimal “particle shape” from the Liu (2008) database. As a first approach for the ARPEGE model, a sensitivity of simulated brightness temperatures to the 11 particle shapes of the Liu (2008) database for the single-scattering properties of snowfall was conducted with RTTOV-SCATT. A sensitivity study to the particle size distribution was not performed for the ARPEGE model in this work but may require further attention in the future. Table 3 summarizes these 11 configurations of RTTOV-SCATT which have been tested.

Over a one-month period (January 2017), the radiative transfer simulations have been compared to SAPHIR observations over oceanic surfaces, using two different metrics. The first metric corresponds to the root-mean-square error (RMS) of first-guess departures (observed TBs minus first-guess TBs), computed for meteorological scenes which are characterized by similar scattering conditions both in the observation and the first guess (Chambon *et al.*, 2014). This categorization prevents contamination of the distributions of first-guess departures with situations where the observations are clear and the first guess cloudy or vice versa. The second metric is defined in Geer and Baordo (2014) and directly compares the brightness temperatures histograms by giving a fair weight to the less populated cold brightness-temperature bins and hence uniformly compares the histograms through the full range of brightness temperatures. Note that the smaller this histogram fit metric is, the better the observed and simulated histograms match. Both comparisons are shown in Figure 2.

Figure 2a indicates that the best simulations are obtained with the four following particle shapes: 4-bullet rosette,

TABLE 3 RTTOV-SCATT configurations tested for simulating SAPHIR with the ARPEGE model

	Particle size distribution	Single emission/absorption/scattering radiative properties
Cloud liquid water	Modified gamma	Mie sphere
Cloud ice water	Modified gamma	Mie sphere
Rain	Marshall–Palmer	Mie sphere
Snow	Field <i>et al.</i> , 2007, Tropical regime	Discrete Dipole Approximation (DDA). 11 different shapes from Liu (2008)

**FIGURE 2** Comparison of RTTOV-SCATT simulations with ARPEGE atmospheric profiles and SAPHIR observations over oceans for a time period of one month (January 2017) considering several shapes for snowfall radiative properties: (a) RMS error of first-guess departures on categorized scenes (Chambon *et al.*, 2014). (b) Geer and Baordo (2014) brightness temperatures histogram fit metric; for the six SAPHIR channels and averaged (solid curve)

5-bullet rosette, 6-bullet rosette and sector snowflake, for the six SAPHIR channels. The results obtained by using the histogram fit metric in Figure 2b are consistent with the latter results. The bullet rosettes together with the sector snowflake shape provide the simulations which compare the best to the observations.

The sector snowflake shape leading to a slightly better histogram fit than the other four potential candidates, it was then selected for the rest of the study. Note that this result is somewhat different from the one found by Guerbet *et al.* (2016) for the limited-area model Aire Limitée Adaptation Dynamique développement InterNational (ALADIN) Réunion (Fischer *et al.*, 2005) which led to selecting a block column for simulating SAPHIR data in cyclonic conditions. Nevertheless, the choice of sector snowflake is consistent with the one performed at ECMWF, for a global model and for a larger range of frequencies, giving confidence in the latter results. The bullet rosettes would still be interesting candidates for the assimilation of SAPHIR data. Therefore, the information content of these multiple particle shape simulations will be exploited as explained further in section 3.3.

As shown in Figure 1, the RTTOV-SCATT simulations are used to split the population of observations into two groups, based on a simple cloud screening (Chambon *et al.*, 2015) that has been modified for improving the symmetry in the sampling of meteorological scenes. The TBs satisfying the

following criteria are routed toward the clear-sky assimilation route and the other observations toward the Bayesian framework:

$$\text{Clear Sky Obs} : |TB_{\text{OBS}}^6 - TB_{\text{CLEARSKY RTTOV}}^6| < 5 \text{ K}$$

and

$$\text{Clear Sky Model} : |TB_{\text{RTTOV SCATT}}^6 - TB_{\text{CLEARSKY RTTOV}}^6| < 5 \text{ K}$$

where TB_{OBS}^6 corresponds to the observed brightness temperature of the 183.31 ± 11 GHz channel, $TB_{\text{CLEARSKY RTTOV}}^6$ and $TB_{\text{RTTOV SCATT}}^6$ correspond to the simulated brightness temperatures without and with taking into account the effects of hydrometeors respectively.

The RTTOV-SCATT simulations of SAPHIR channels are then considered within a Bayesian inversion described in the next section.

3.3 | Bayesian inversion of SAPHIR brightness temperatures

3.3.1 | Basic principles of the inversion

The 1D-Bayesian inversion of SAPHIR TBs undertaken in this study rely on the algorithm described in Caumont *et al.* (2010) and Wattrelot *et al.* (2014). Its objective is to retrieve from the set of observations a physical quantity that can be easily assimilated in the ARPEGE 4D-Var. This algorithm

was inspired by the Goddard Profiling (GPROF) algorithm used for rainfall estimation from passive microwave observations (Kummerow *et al.*, 1996). The main specific feature of the 1D-Bayesian inversion developed by Caumont *et al.* (2010) is that instead of using a large off-line database of atmospheric profiles, it exploits a database, limited in size, but dynamically changing in space and time. These specificities, explained further below, lead to a case-dependent database adapted to the current meteorological conditions.

In this work, the database of *a priori* information for the inversion of every SAPHIR observation is composed of a limited number of atmospheric profiles, sampled from the model first guess in the vicinity of a given observation. Considering a very large neighbourhood does increase the chances of successful retrieval, but the risk is to mix together too different meteorological situations. In the experiments which are presented in section 4, this neighbourhood is defined as an area of diameter 400 km, centred on each new observation.

For each new SAPHIR observation (\mathbf{y}_o), a number of first-guess profiles (x_i) are sampled within the neighbouring area. In the experiments which are presented in section 4, this sampling is made on the model grid, which is regularly sampled at 48 locations (7 by 7 locations minus the location of the observation). Not considering the first guess at the location of the observation also reduces the correlation between the retrieved profiles and the first guess, which was one of the drawbacks of the “1D-Var + 4D-Var” method (Geer *et al.*, 2008).

A RTTOV-SCATT simulation is performed for each sampled profile of this area, $H(x_i)$, and a weight, W_i , is then attributed to each of them following the formula:

$$W_i = \exp \left\{ -\frac{1}{2} [(\mathbf{y}_o - H(x_i))^T \mathbf{R}^{-1} (\mathbf{y}_o - H(x_i))] \right\}$$

The values of these weights depend both on how close the RTTOV-SCATT simulations $H(x_i)$ are from the observation vector \mathbf{y}_o and on the prescribed error covariance matrix \mathbf{R} . This matrix represents the radiometric errors of \mathbf{y}_o as well the errors from the observation operator. From these weights, retrieved atmospheric profiles (humidity, temperature, hydrometeors) are computed as a weighted average of the neighbours following the formula below:

$$x^{\text{retr}} = \frac{\sum_i W_i x_i}{\sum_i W_i}$$

This inversion method relies on several assumptions: (a) the database made of neighbour profiles is rather limited and therefore will only be efficient if a cloud similar to the observed one is simulated by the first guess within its vicinity, (b) the forward radiative transfer simulation on the weighted average profile must not be too different from the weighted average of the forward radiative transfer simulations on each of the neighbouring profiles.

Regarding the first assumption, if no cloud is present in the first guess in the vicinity of an observed cloud, the

1D-Bayesian inversion does not provide any result and no retrieval is then assimilated within the 4D-Var.

Regarding the second assumption, it deals with the non-linearity of the observation operator; the framework set up allows us to compute both the forward radiative transfer simulation on the weighted average and the weighted average of the forward radiative transfer simulations. Quality control is then performed to reject the observation in case of too large differences.

3.3.2 | Observation errors within the Bayesian inversion

Another important aspect of the 1D-Bayesian inversion is the prescription of the covariance matrix \mathbf{R} which shall describe the radiative transfer modelling errors together with the measurement errors, that will affect the retrieval accuracy. For ground radar reflectivities, two solutions have been experimented within Caumont *et al.* (2010): (a) the specification of a small error which leads to either an accurate Bayesian inversion or no inversion at all, and (b) the specification of a larger error which increases the number of successful inversions but smooths the results with more distributed weights. The assimilation experiments results shown in Caumont *et al.* (2010) and Wattrelot *et al.* (2014) indicate that the first solution was preferable for ground radar reflectivities. For SAPHIR TBs, Guerbet *et al.* (2016) experimented with similar solutions on a cyclonic case and selected a 1 K error for each channel of the SAPHIR sounder; but in their study, the inversion database used was made of all profiles over the geographical domain of the ALADIN Réunion limited-area model.

In this study, two options have been examined. The first one consists of a diagonal \mathbf{R} matrix with a 1 K value for each of the six SAPHIR channels. The second option is based on an empirical model described below, aiming at taking into account errors of radiative transfer simulations at 183.31 GHz which are known to be large in strong scattering situations (e.g. Kulie *et al.*, 2010; Bennartz and Greenwald, 2011). Therefore, the general idea of this error model is to specify small errors in low-scattering conditions (warm observed TBs) and to inflate them in the strongest-scattering conditions (cold observed TBs).

Following Haddad *et al.* (2015) who defined an observation operator taking into account multiple microphysical assumptions in the forward simulations, the error model defined in this study is based on the RTTOV-SCATT simulations presented in section 3.2. Indeed, a group of four different particle shapes has been identified as good candidates for the observation operator settings. The same data have been used in order to investigate whether these four configurations can lead to a significant spread between forward simulations.

For each of the six SAPHIR channels, four brightness temperatures $\text{TB}_{\text{option 1}}$, $\text{TB}_{\text{option 2}}$, $\text{TB}_{\text{option 3}}$ and $\text{TB}_{\text{option 4}}$ are simulated for each atmospheric profile with the four best options of microphysical radiative properties as discussed above. Then the warmest and the coldest values are identified

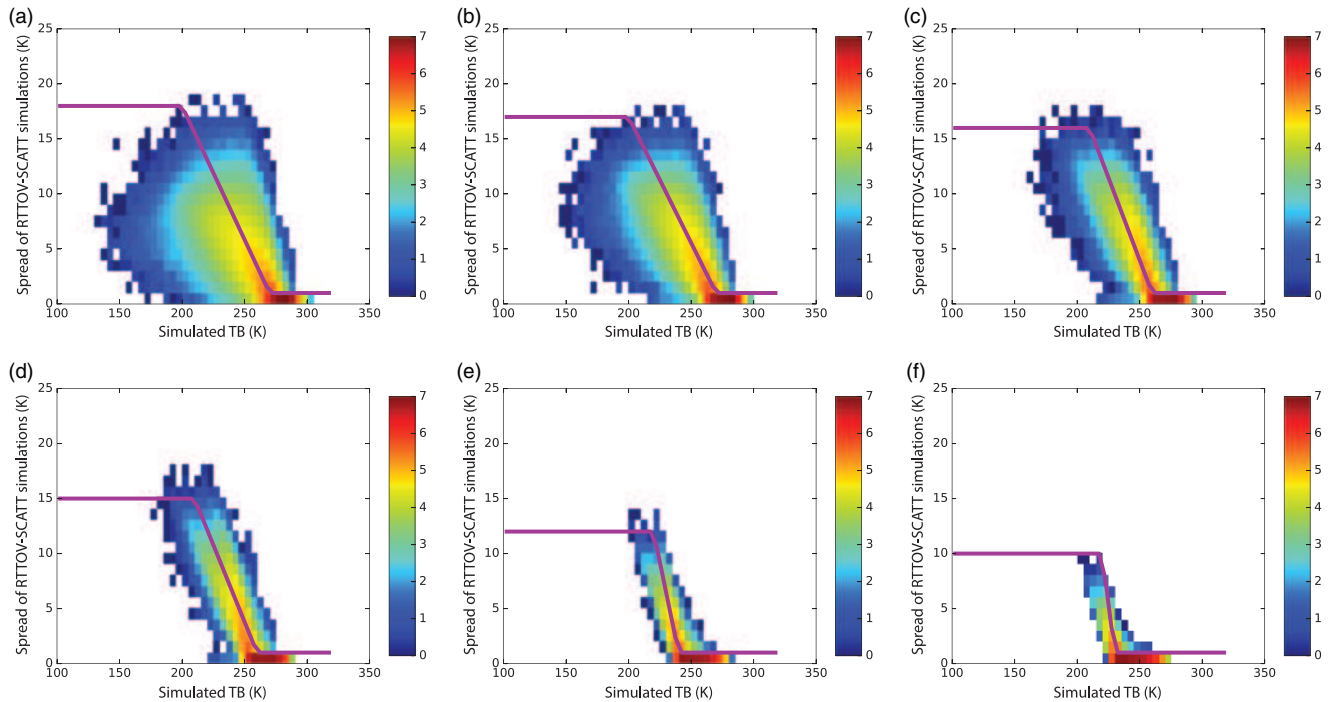


FIGURE 3 Scatter plots of estimated radiative forward simulation uncertainties for each of the six SAPHIR channels: (a) channel 6, (b) channel 5, (c) channel 4, (d) channel 3, (e) channel 2, (f) channel 1. The spread of a group of four simulations with the bullet rosettes and the sector snowflake particle shapes is plotted against each of the four forward simulations. The number of samples, corresponding to one month of SAPHIR observations in January 2017, is represented in colour with a log10 scale. The purple straight lines correspond to the empirical models fitting the modes of the distributions [Colour figure can be viewed at wileyonlinelibrary.com].

TABLE 4 Inversion error model parameters for each of the SAPHIR channels

Channel	TB threshold from which inversion errors grow (K)	TB threshold from which inversion errors saturate to the maximum value (K)	Minimum inversion error (K)	Maximum inversion error (K)
1	230	220	1	10
2	240	220	1	12
3	260	210	1	15
4	260	210	1	16
5	270	200	1	17
6	270	200	1	18

across the four TBs and a spread value is defined as the difference between the two. Figure 3 shows scatter diagrams of this spread as function of TB for each of the six SAPHIR channels. Within these scatter diagrams, each pair of (TB, spread) is represented, meaning that for a given atmospheric profile, four pairs are counted: (TB_{option 1}, spread), (TB_{option 2}, spread), (TB_{option 3}, spread) and (TB_{option 4}, spread).

In Figure 3a, one can see that SAPHIR channel 6 simulations are characterized by a small spread for warm TBs which is expected because they are not impacted by snow scattering. Then for cold scenes, this spread ranges from a few Kelvins up to 18 K. This illustrates that the observation operator is either quite linear (small spread values) or quite nonlinear (large spread values) with respect to the particle shape, depending on the atmospheric profile and/or the meteorological situation. In Figure 3f dedicated to SAPHIR channel 1, one can note that the spread tends to be smaller, and increases up to 10 K.

A simple empirical model was defined based on this dataset as an attempt to take into account these dependencies within the Bayesian inversion: it is defined as a piecewise linear model, fitting the modes of the distributions. These models (one per channel) are over-plotted in Figure 3 and their numerical values are indicated in Table 4.

The impact of the two configurations for the error covariance matrix \mathbf{R} of the inversion, either constant and equal to 1 K for all channels, or empirically modelled and variable, are illustrated in Figure 4 for a case-study along the Thailand coastline:

- The comparison of the observations and the ARPEGE first guess (Figure 4a,b) along the Thailand coastline indicates a misplacement of the clouds in the model as well as not enough scattering (TBs too warm). The latter point could arise from either the model profiles or the radiative transfer settings not fully appropriate for this convective system.

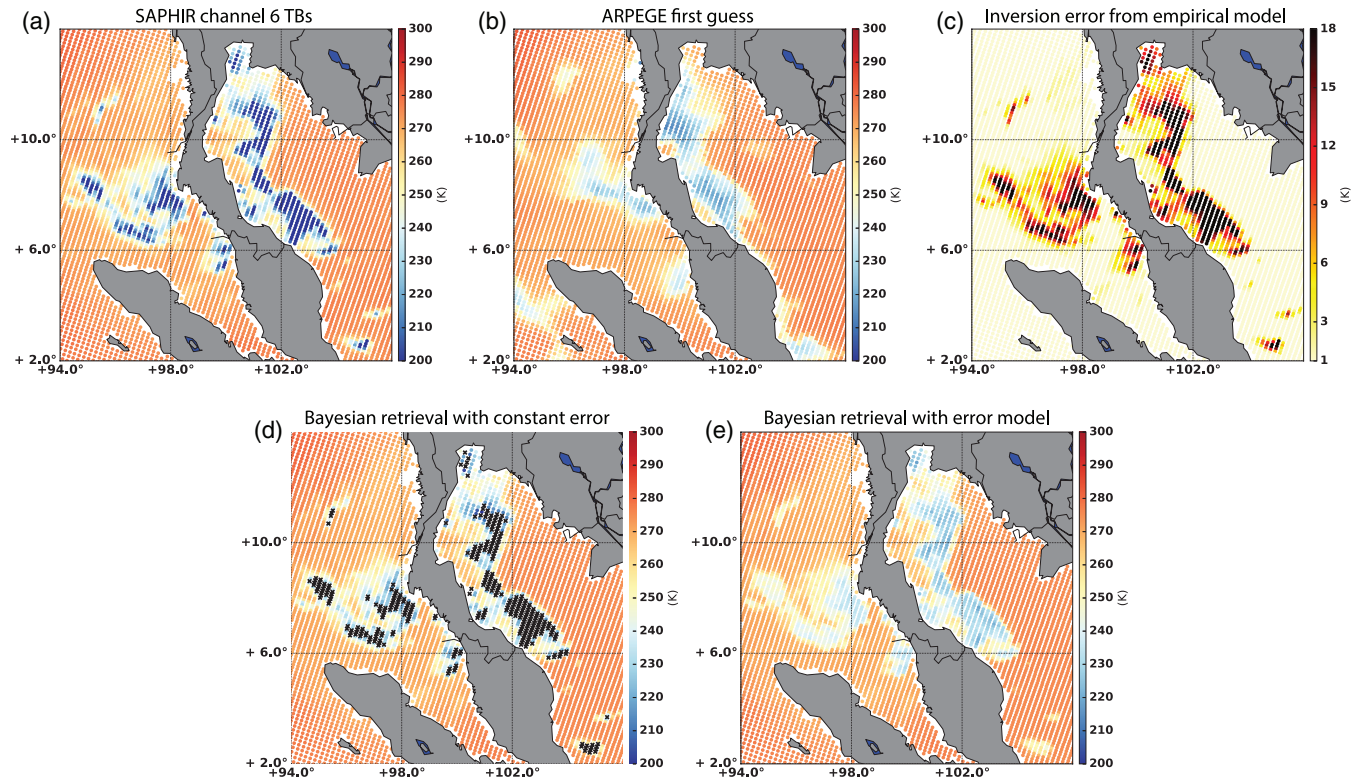


FIGURE 4 (a) Observations from SAPHIR channel 6 on 8 January 2017 at 0000 UTC; (b) the corresponding model guess at this same time and location; (c) inversion error prescribed by the inversion error model; (d) retrieval in the brightness temperature space, considering an inversion error of 1 K; (e) retrieval in the brightness temperature space, considering the inversion error model adapted to the observed scenes. Small black crosses represent locations where the Bayesian inversion is unsuccessful [Colour figure can be viewed at wileyonlinelibrary.com].

- Figure 4d,e display the results of the 1D-Bayesian inversion, for the two specifications of \mathbf{R} . With a constant value of 1 K, the inversion fails in the cores of the deep convective systems (black symbols). In such a situation, the weights W_i are too small, because the observed TBs are too far from any simulation in the first guess. When using the empirical error model, no inversion fails, and the result of the inversion depicts patterns similar to the observed ones, with a warm bias of a magnitude consistent with the prescribed large errors in strong-scattering scenes (Figure 4c).

Another possible option to increase the number of successful retrievals without inflating the inversion errors is to enlarge the database by for instance increasing the neighbourhood size. Doubling this size was experimented (not shown) and only marginally increased this number. It may be the case that enlarging the database more would lead to a different result. Nonetheless, using local meteorological situations as inversion database is one of the strengths of the method. Therefore, it is the two different options of inversion errors for the 1D-Bayesian inversion, both with a “small” neighbourhood, which have been tested in this study. With these two different settings, this framework will help testing what configurations provide the largest impact on forecast scores: more retrievals of a worse accuracy or less retrievals of a better accuracy.

3.4 | Assimilation of the 1D-Bayesian retrievals and observation errors specifications

3.4.1 | Selection of retrievals to be assimilated

From the Bayesian framework presented in the previous section, vertical profiles of temperature, specific humidity and hydrometeors are retrieved at all model levels. Therefore, (a) one or several atmospheric variables shall be selected for assimilation within the 4D-Var, (b) the retrieved profiles shall be sampled on the vertical, and (c) observation errors shall be prescribed for them.

Regarding the first aspect, relative humidity (RH) has been chosen as the variable to be assimilated following the work by Caumont *et al.* (2010) and Wattrelot *et al.* (2014) from radar reflectivity Bayesian inversion, in order to bring in information on the degree of saturation of the atmosphere to the 4D-Var analysis. This choice is as well supported by the known dependence of TBs at 183.31 GHz on RH (e.g. in clear sky: Buehler and John, 2005; Brogniez *et al.*, 2015). Whether assimilating alternative data such as specific humidity and temperature instead of RH will be considered in a future study.

Regarding the vertical sampling of the retrieved profiles, since each SAPHIR channel is sensitive to a rather bulky atmospheric layer (around 200 hPa of thickness), the number of vertical levels is set equal to the number of channels selected in the inversion. As a first experimental trial, only the four SAPHIR channels the most sensitive to hydrometeors

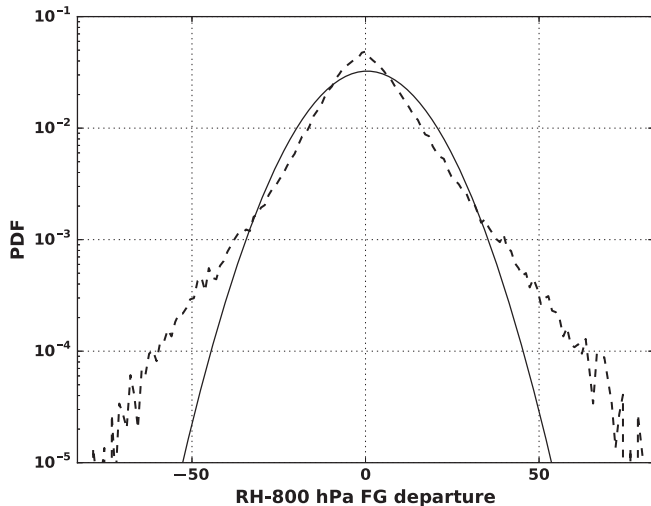


FIGURE 5 Retrieved relative humidity first-guess departures distribution at 800 hPa over January 2017 (dashed thick line). Corresponding Gaussian distribution (thin full line)

have been selected in the Bayesian inversion (channels 3 to 6). Following a side study based on RTTOV-SCATT Jacobians for specific humidity and detailed in the appendix, the four pressure levels 500, 600, 700, 800 hPa have been selected. It is likely that this step will need further refinements in the future.

3.4.2 | Observation errors for the assimilation of retrievals

First-guess departures of RH have been analysed to prescribe the errors for the retrievals at each pressure level. Figure 5 shows an example of the first-guess departure distribution for RH at 800 hPa (note that in the rest of the article, first-guess departure refers to retrieved value – first-guess value). The shape of this distribution, in particular its kurtosis, indicates that it is characterized by a similar non-Gaussian behaviour to all-sky TBs first-guess departures (e.g. Geer and Bauer, 2011). This can also be illustrated in Figure 6 where there is a strong link between the dipoles of positive and negative first-guess departures of SAPHIR channel 6 TBs (Figure 6b) and the negative and positive departures of Total Column Water Vapour (TCWV) along the Thailand coastline (Figure 6d). The blue/red patterns in first-guess departures are indicators of cloud misplacements between the observations and the model first guess. In particular, the RH first-guess departure at 800 hPa (Figure 6c) is characterized by areas which would aim to either dry (in red) or moisten (in blue) the atmosphere.

The behaviours of first-guess departure statistics of TBs and RH retrievals are compared in Figure 7. The root-mean-square (RMS) of SAPHIR channel 5 first-guess departures, categorized as a function of a scattering index, are displayed in Figure 7a. This scattering index, defined for SAPHIR in Chambon and Geer (2017), is the difference $TB_{\text{CLEARSKY RTTOV}}^6 - TB_{\text{OBS}}^6$ for the observation (horizontal axes in Figure 7a,c) and $TB_{\text{CLEARSKY RTTOV}}^6 - TB_{\text{RTTOV SCATT}}^6$ for the model first guess (vertical axes in Figure 7a,c).

This categorization of meteorological scenes highlights the small RMS of first-guess departure when the model first guess and the observation are in similar scattering conditions (diagonal of the plot), and large RMS values when the clouds are mis-located in the first guess. This latter case corresponds to the largest number of samples as indicated in Figure 7c. A similar approach of categorization is attempted for the RH retrievals but the predictor for categorizing the scenes is total column water vapour as shown in Figure 7b,d. In these figures, $TCWV_{\text{obs}}$ refers to the TCWV of the retrieved profile, and $TCWV_{\text{model}}$ to the TCWV of the first guess.

The main difference between the distributions shown is how the samples are distributed within the two different spaces. The maximum number of samples is located along the diagonal for RH while it is not the case for TBs. Indeed, high TCWV is a required but not sufficient condition for triggering convection and the water vapour mass is better constrained and forecasted by the model than clouds and precipitation as TCWV is a larger-scale meteorological field.

The main similarity between TBs and RH innovation statistics (Figure 7a,b) is that the RMS value is smaller when meteorological conditions are similar (diagonal of the plot) and larger when there are mis-locations in the first guess. These similarities on the RMS increase as a function of the magnitude of the humidity mis-location, indicating that observation errors for relative humidity retrievals could be potentially modelled to improve the Gaussianity of first-guess departure distribution. As already shown for the direct assimilation of all-sky radiances (Geer and Bauer, 2011; Minamide and Zhang, 2017; Okamoto, 2017), this could enhance the assimilation impact.

A first attempt toward this direction is presented in Figure 8a,b. It consists of modelling the gradient of increasing observation error visible in Figure 7b with a simple predictor based on TCWV. A difference between the TCWV from the Bayesian retrieval, $TCWV_{\text{obs}}$, and the TCWV from the first guess, $TCWV_{\text{model}}$, is taken as predictor to categorize the standard deviations of relative humidity first-guess departures. One example for 800 hPa is given in Figure 8a where there is an increase of the standard deviation with large differences of $(TCWV_{\text{obs}} - TCWV_{\text{model}})$.

Following the empirical approach of Geer and Bauer (2011), a simple piecewise linear model was defined to describe these variations. This linear model prescribes observation errors smaller than the standard deviation of the full departure distribution for small values of $(TCWV_{\text{obs}} - TCWV_{\text{model}})$. Then the linear model increases up to two saturation values, different for positive and negative differences of $(TCWV_{\text{obs}} - TCWV_{\text{model}})$ which is consistent with the asymmetry mentioned above. The numerical values of this linear model can be found in Table 5 for the four selected pressure levels, together with the standard deviations of each first-guess departure distribution used for the constant-error case.

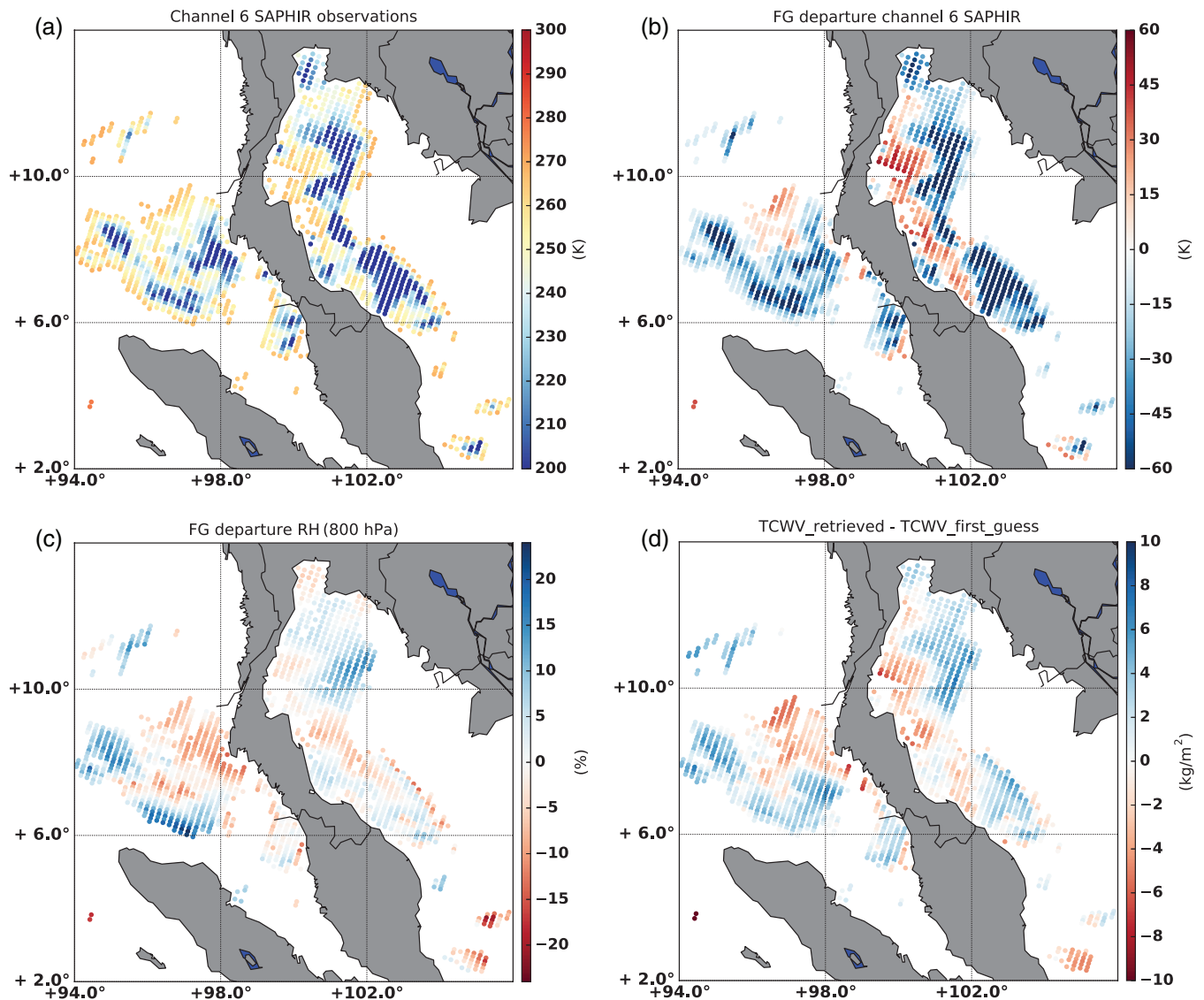


FIGURE 6 (a) Observations from SAPHIR channel 6 on 8 January 2017 at 0000 UTC which have been selected for the Bayesian framework; (b) first-guess departure in brightness temperature space; (c) retrieved RH first-guess departure at 800 hPa; (d) retrieved TCWV first-guess departure [Colour figure can be viewed at wileyonlinelibrary.com].

Normalized first-guess departure distributions are displayed in Figure 8b at 800 hPa and the skewness and kurtosis of each distribution for the four levels can be found in Table 6. The distribution, normalized with the varying error, is closer to a Gaussian curve than the one normalized with a constant error, with a kurtosis of 1.07 instead of 2.89, as well as a skewness of -0.03 instead of 0.16. For other levels, these two parameters are similarly improved as well with the variable observation error model (see Table 6).

Within the developed Bayesian framework, two options can then be used as observation errors for assimilating the relative humidity retrievals:

- a constant error, taken as the standard deviation of the first-guess departure distributions.
- a varying error, taken as the piecewise linear model function of $(TCWV_{obs} - TCWV_{model})$ which prescribes larger errors in the case of mis-located humid/dry areas.

4 | ASSIMILATION EXPERIMENTS

The methodology described in section 3 was experimented with the ARPEGE 4D-Var framework (cycle cy42op2) over a three-month period covering January to March 2017. Four assimilation cycles with 6 h windows are performed every day at 0000 UTC, 0600 UTC, 1200 UTC and 1800 UTC. Along with this cycling using 6 h ARPEGE forecasts, forecasts with a larger lead time of +102 h are performed every day at 0000 UTC. Section 4.1 describes the different configurations of experiments which have been conducted and section 4.2 describes the results in terms of analyses and forecast skill scores.

4.1 | Experimental set-up

Four different options of the Bayesian framework have been considered for the assimilation of SAPHIR data within clouds

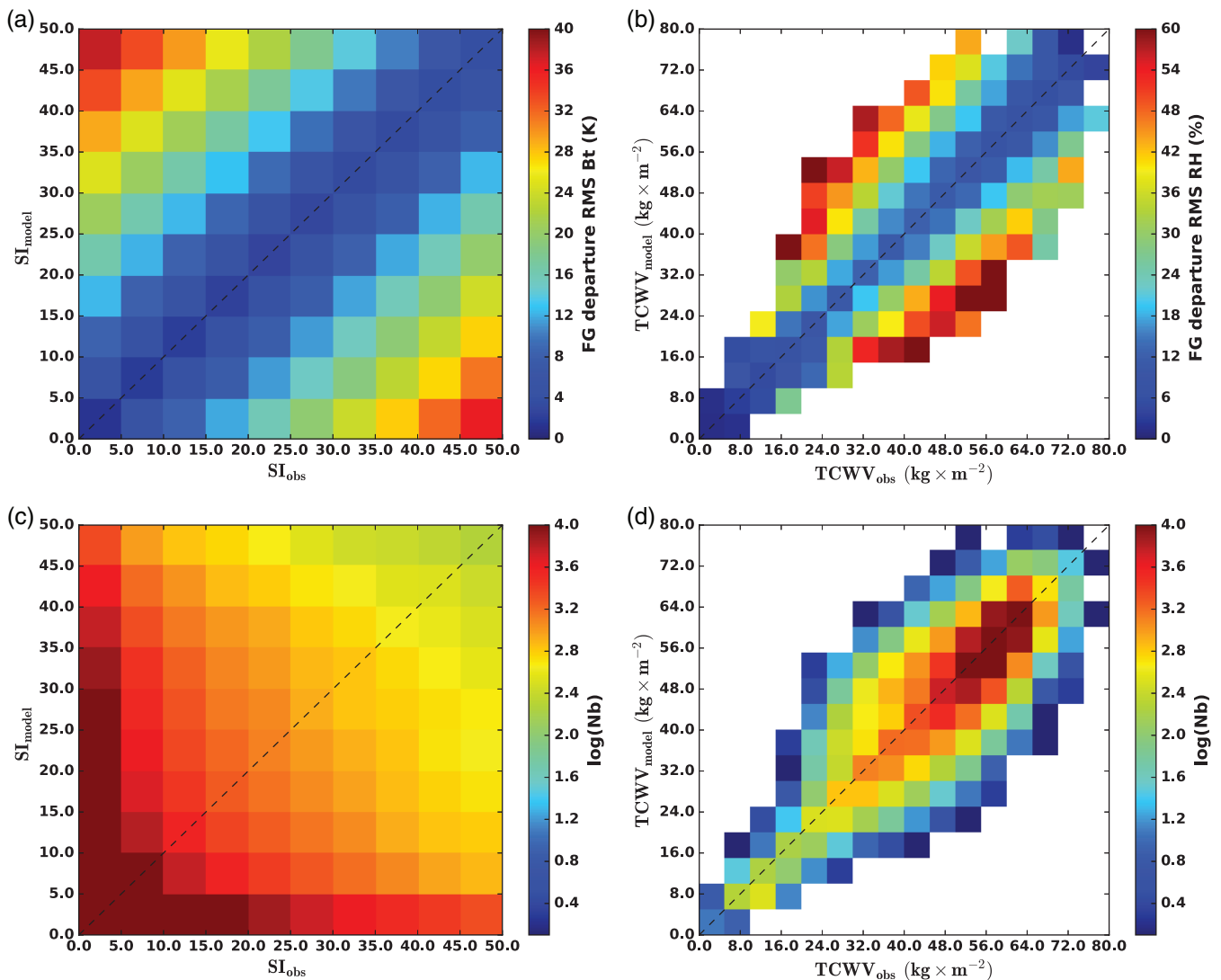


FIGURE 7 (a) SAPHIR channel-5 TBs first-guess departure RMS as function of a scattering index computed for the model guess and for the observation over a time period of 1 month; (b) relative humidity first-guess departure standard deviation at 800 hPa as function of TCWV for the model guess and the TCWV of the observations covering a time period of 1 month; (c) sample size repartition within (a); (d) sample size repartition within (b) [Colour figure can be viewed at wileyonlinelibrary.com].

and precipitation. Two options do modify the observation errors of the Bayesian inversion by using either a constant value or the error model of section 3.3.2 based on multiple microphysical assumptions (see Table 4 for the model parameters used). These options are respectively called INV-cst and INV-mod. Then two options modify the observation errors used within the 4D-Var for the assimilation of the retrieved RH at four pressure levels by using either a constant value or the error model of section 3.4.2 based on a TCWV predictor (see Table 5 for the model parameters used). These options are respectively called ASSIM-cst and ASSIM-mod.

Considering these different possibilities, five assimilation experiments of SAPHIR data have been defined: a reference run in which the full observing system detailed in Table 2 is assimilated, including SAPHIR data as in operations, and the four assimilation experiments detailed in Table 7 and called respectively: INV-cst_ASSIM-cst, INV-mod_ASSIM-cst, INV-cst_ASSIM-mod and INV-mod_ASSIM-mod. It should

be noted that, in these four experiments, the Bayesian framework is activated over oceans only as a first step.

For the horizontal thinning of observations in these four experiments, the current thinning of SAPHIR data within ARPEGE is set to 130 km as in Chambon *et al.* (2015). Therefore, the same thinning is applied to RH retrievals in cloudy areas but could be enhanced in the future as experimented with at the Global Modelling and Assimilation Office for all-sky radiances.

In the following section, the four experiments are compared to the reference run, in terms of difference of standard deviation on tropical wind forecasts, temperature forecasts and RH forecasts. The 6 h ARPEGE forecasts are validated against observations: geostationary derived winds called SATOB for wind forecasts, clear-sky AMSU-A brightness temperatures from NOAA-15, NOAA-18, NOAA-19, Aqua, MetOp-A, MetOp-B satellites for temperature forecasts, and clear-sky MHS brightness temperatures from NOAA-18, NOAA-19,

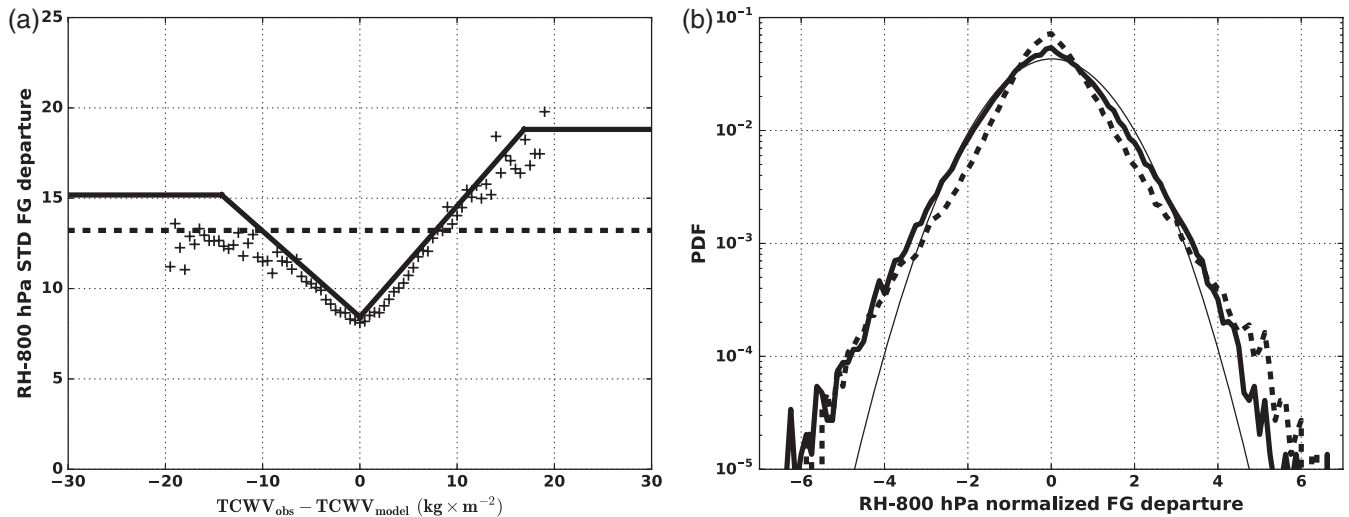


FIGURE 8 (a) Standard deviation of first-guess departures of relative humidity in % at 800 hPa, categorized as function of $TCWV_{obs} - TCWV_{model}$ (crosses), empirical model (thick line), constant observation error (thick dashed line). (b) Distributions of first-guess departures of relative humidity at 800 hPa, normalized with the empirical model (thick full line), normalized with a constant observation error (thick dashed line), compared to a Gaussian distribution (thin full line)

TABLE 5 Constant error values used for the assimilation of relative humidity retrievals (second column) and observation error model parameters for each of the four selected pressure levels

Pressure level (hPa)	Constant errors (%)	Error models' parameters				
		Error max for negative TCWV difference (%)	Error max for positive TCWV difference (%)	Minimum error for null TCWV difference (%)	Negative TCWV difference from which the error saturate	Positive TCWV difference from which the error saturate
800	14	15.18	18.81	8.43	-14.21	16.90
700	17	17.93	18.98	10.85	-13.30	14.20
600	19	24.61	24.70	12.75	-16.0	12.80
500	22	26.94	29.69	15.23	-14.20	14.20

TABLE 6 Standard deviation, skewness and kurtosis of the normalized relative humidity first-guess departure distributions over January 2017

Pressure level (hPa)	Constant observation error		Varying observation error	
	Skewness of normalized distribution	Kurtosis of normalized distribution	Skewness of normalized distribution	Kurtosis of normalized distribution
800	0.16	2.89	-0.03	1.07
700	0.09	1.91	0.02	0.76
600	0.25	1.93	0.12	1.32
500	0.03	1.64	0.07	1.21

MetOp-A, MetOp-B satellites for humidity forecasts. The longer +102 h forecasts at 0000 UTC are then validated in terms of difference of standard deviation with respect to the ECMWF analyses.

4.2 | Assimilation experiment results

Regarding the number of SAPHIR data assimilated in the five experiments, Table 7 indicates that the modification of the clear-sky route, performed over both land and oceans, leads to a decrease of assimilated SAPHIR brightness temperatures by roughly -2% in all four experiments with respect to the

reference run. However, the retrieved RH bring a significant additional amount of observations of almost +16% of data for all the four experiments, even though they are assimilated over oceans only. Hence, the total amount of assimilated SAPHIR data or SAPHIR-derived data increases by more than +13% with respect to the SAPHIR data count in the reference.

As can be seen in Figure 9, the assimilation of cloudy and rainy SAPHIR observations with the Bayesian framework does improve ARPEGE wind forecasts in the four experiments INV-cst_ASSIM-cst, INV-mod_ASSIM-cst, INV-cst_ASSIM-mod and INV-mod_ASSIM-mod.

TABLE 7 Assimilation experiments characteristics and data counts in the ARPEGE 4D-Var system over a three-month period (January–March 2017)

	Reference	INV-cst ASSIM-cst	INV-mod ASSIM-cst	INV-cst ASSIM-mod	INV-mod ASSIM-mod
Assimilated observing system	See Table 2	Reference + RH at 4 levels from SAPHIR Bayesian inversion over oceans	Reference + RH at 4 levels from SAPHIR Bayesian inversion over oceans	Reference + RH at 4 levels from SAPHIR Bayesian inversion over oceans	Reference + RH at 4 levels from SAPHIR Bayesian inversion over oceans
Inversion error	/	1 K	Model	1 K	Model
Observation errors for retrieved RH	/	Constants (see Table 5)	Constants (see Table 5)	Model (see Table 5)	Model (see Table 5)
Clear-sky route: SAPHIR data count	15,791,911	15,453,419– 2.143%	15,453,239– 2.144%	15,452,954– 2.146%	15,456,554– 2.123%
Bayesian route: retrieved RH data count	/	2,486,056+ 15.743%	2,491,923+ 15.780%	2,471,085+ 15.648%	2,476,718+ 15.683%

These improvements can be detected with SATOB observations for the 6 h forecasts. They roughly represent a 0.5% improvement of the standard deviation of innovation statistics and are statistically significant at the 95% confidence level for the four experiments and the levels 150, 200, 250 and 1,000 hPa. The results for the intermediate vertical levels vary from one experiment to the other but also correspond to levels for which the numbers of SATOB data are the smallest. These positive impacts propagate within the ARPEGE forecasts up to +48 h as can be seen in Figure 9 (right column) for the scores with respect to the ECMWF analyses. The latter improvements range between 1.8 and 0.5% of improvement of wind forecast scores.

Similar conclusions can be drawn with the temperature forecast scores shown in Figure 10. An improvement of temperature forecasts is detected for all experiments with the AMSU-A instrument, with respect to channels 7, 8 and 9 and of roughly 0.3%. A small degradation can also be seen on channels 5 and 6 by 0.1%. This latter degradation is also visible on scores with respect to the ECMWF analyses but do not last more than 6 h. A stronger and positive impact, ranging between 0.5 and 1.2%, is also visible between 700 and 150 hPa and propagates up to the +72 h lead time.

A different behaviour is found for the scores on humidity forecasts. Indeed, Figure 11 shows that the first-guess departure standard deviations of the MHS instrument are improved for the three 183.31 GHz channels, by roughly 0.5%, but the propagation in time of the positive impact in the forecast is rather different than on temperature and winds. A positive signal is found with respect to the ECMWF analyses and lasts up to +18 h/+36 h only, depending upon pressure level. Hence the positive impact of SAPHIR cloudy and rainy observations lasts much less on humidity than on the other two atmospheric variables and may be related to a poorer predictability of the water vapour environment of tropical convection (e.g. Mahfouf and Bilodeau, 2007). A degradation of roughly 0.2% on RH forecasts scores can also be found

below 800 hPa and is confined between 850 and 900 hPa. These results support the conclusion that cloudy and rainy assimilation does improve the quality ARPEGE forecasts, whatever the options used within the cloudy assimilation framework.

Comparing the four tested configurations of the cloudy framework, the three experiments using error models (INV-mod_ASSIM-cst, INV-cst_ASSIM-mod, INV-mod_ASSIM-mod) have been compared to the simplest cloudy assimilation framework (INV-cst_ASSIM-cst) using the same statistical testing method as for Figures 9, 10 and 11 (not shown). It was found that using an inversion error model (INV-mod_ASSIM-cst), which allows assimilating observations with deep convective cores associated to cold SAPHIR brightness temperatures, slightly decreases the positive impacts of the simplest configuration INV-cst_ASSIM-cst: it is a significant degradation in the case of wind forecasts at 500 hPa and temperature forecasts above 400 hPa. Using the observation error model (INV-cst_ASSIM-mod), which inflates errors of RH retrievals within the 4D-Var in the case of large cloud misplacements, is quite neutral with respect to the INV-cst_ASSIM-cst. Then, using the combination of the two models (INV-mod_ASSIM-mod) inherits the small degradations found for the INV-mod_ASSIM-cst with respect to the INV-cst_ASSIM-cst experiment.

Overall, assimilating cloudy and rainy SAPHIR observations with the Bayesian methodology described in section 3 was found to be beneficial to ARPEGE forecasts. Scores over the whole Tropics, with respect to observations and the ECMWF analysis are positive for wind and temperature forecasts up to +48 to +60 h, and on humidity forecasts up to +18 h only. The four tested configurations all lead to positive results, but the simplest cloudy assimilation framework with no error model seems to bring most of the positive signal. Several small degradations are also present in all the four experiments, lasting roughly the first 12 h of the forecasts close to the surface for temperature, and between 900 and 800 hPa for humidity; they will require

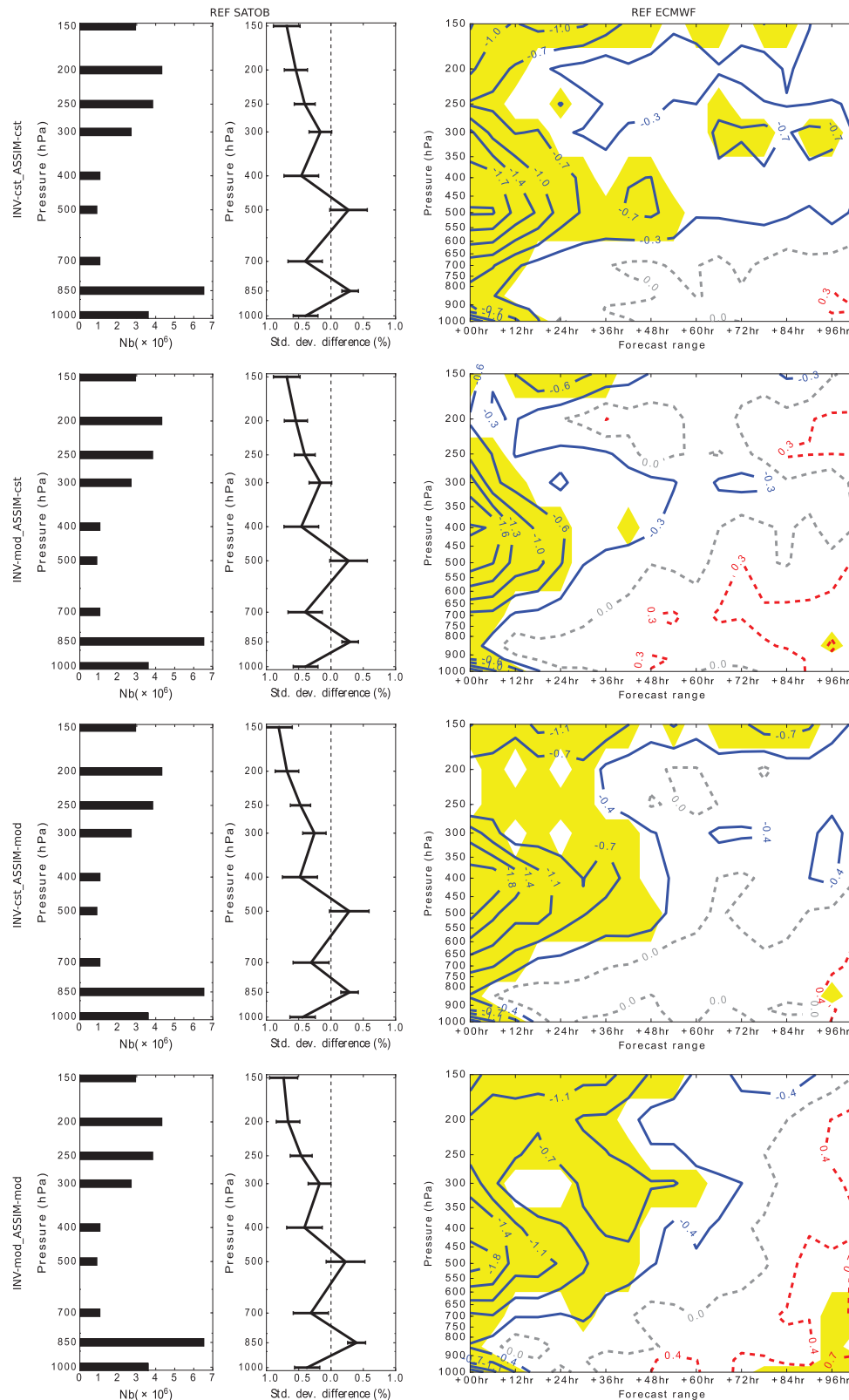


FIGURE 9 Relative differences (%) of standard deviation error for wind forecasts between the reference and the four other assimilation experiments. Negative differences indicate a beneficial impact of the SAPHIR observations in clouds and precipitation. SATOB first-guess departure standard deviation relative difference in the Tropics (middle column) with the corresponding data count (left column) and relative difference of standard deviation of tropical wind forecasts with respect to the ECMWF analysis, between the reference and the four other assimilation experiments (right column). The error bars drawn on the middle column figures correspond to confidence intervals at 95%. Yellow filled areas in the right column figures correspond to impacts which are statistically significant at the 95% confidence level. The first row corresponds to scores for the INV-cst_ASSIM-cst experiment, the second row to scores for the INV-mod_ASSIM-cst experiment, the third row to scores for the INV-cst_ASSIM-mod, and the fourth row for the INV-mod_ASSIM-mod experiment. All these scores are computed over the whole tropical belt between 30°S and 30°N, from January 2017 to March 2017 [Colour figure can be viewed at wileyonlinelibrary.com].

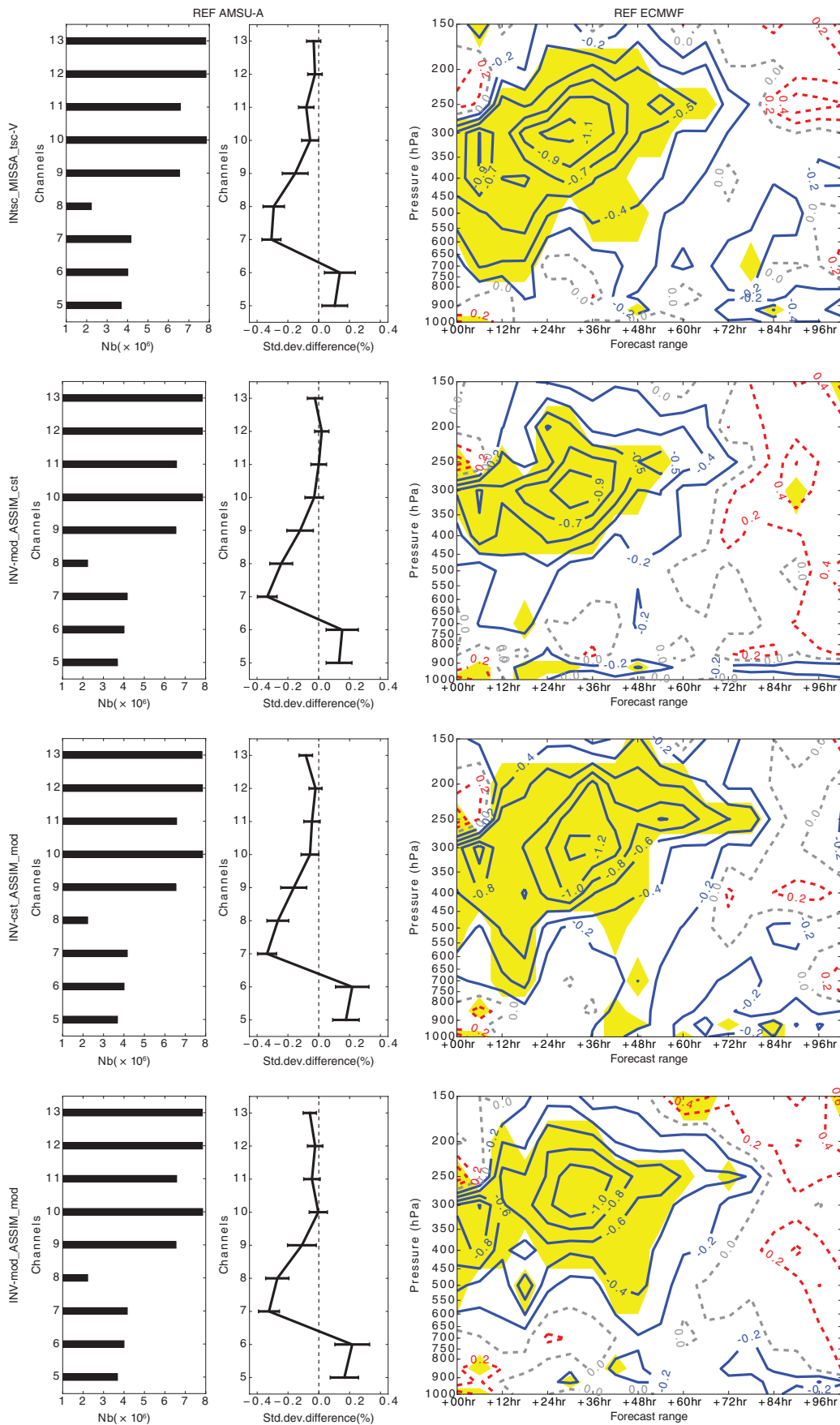


FIGURE 10 Same as Figure 9 for temperature forecasts. Observations used in the middle column correspond to AMSU-A observations NOAA-15, NOAA-18, NOAA-19, AQUA, MetOp-A, MetOp-B satellites [Colour figure can be viewed at wileyonlinelibrary.com].

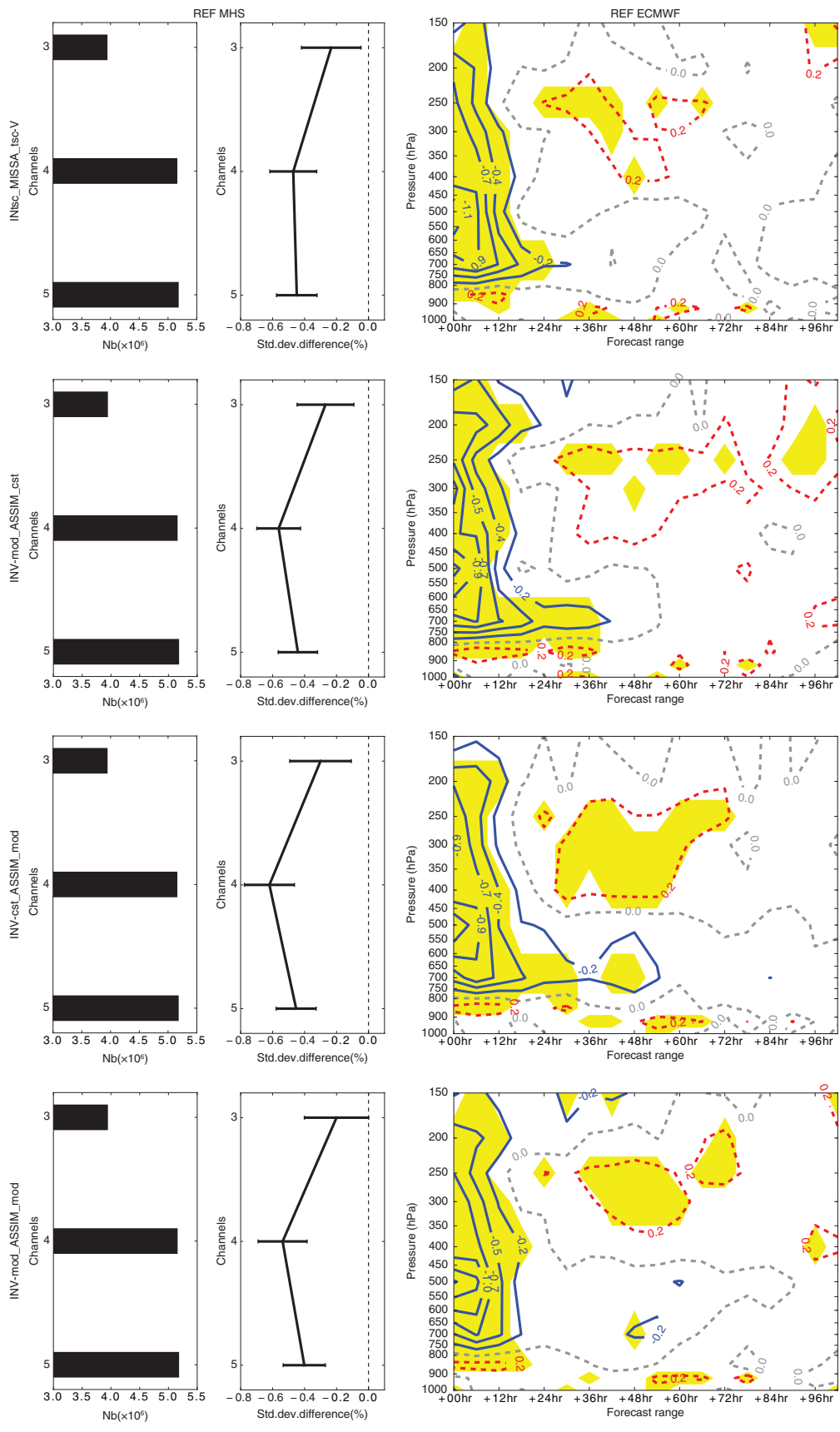


FIGURE 11 Same as Figure 9 for relative humidity forecasts. Observations used in the middle column correspond to MHS brightness temperatures from NOAA-18, NOAA-19, MetOp-A, MetOp-B satellites [Colour figure can be viewed at wileyonlinelibrary.com].

some attention for future developments described in the conclusion.

5 | CONCLUSION AND PERSPECTIVES

An original “1D-Bayesian + 4D-Var” framework has been set up to assimilate cloudy and rainy microwave brightness temperature from the SAPHIR sounder within the global numerical weather prediction model ARPEGE. This implementation has required a modification of the clear-sky assimilation route: SAPHIR TBs are only selected when both the observations and the first guess are considered as clear. The rest of the observations are routed toward a Bayesian inversion which transforms SAPHIR TBs into RH retrievals which are then assimilated within the 4D-Var ARPEGE data assimilation system.

Two error models have been developed and tested within assimilation experiments using a full observing system. The first one can be used within the Bayesian inversion and tentatively estimates the uncertainty of radiative transfer simulations in the presence of hydrometeors. This model is based on multiple particle shape RTTOV-SCATT simulations and was defined in a conservative way to prescribe large errors for the strongest scattering meteorological scenes. The second error model can be used within the 4D-Var to prescribe observation errors of the relative humidity retrievals and inflate errors when humidity structures are significantly misplaced by the model.

Four assimilation experiments using the Bayesian framework have been conducted over a three-month period and are all characterized by better wind, temperature, and relative humidity forecasts than the reference experiment. These positive impacts range between 0.2 and 1% of decrease of standard deviation with respect to either observations (SATOB, AMSU-A, MHS); with respect to the ECMWF analysis, they range between 0.5 and 1.7% of decrease of standard deviation. Onto winds and temperature forecasts, the positive impacts last up to +48 h/+60 h; but on relative humidity, they last only up to +18 h. These results are very encouraging for pursuing the developments of this Bayesian methodology.

The forecast degradations found in these experiments also provide interesting hints for the future developments of this framework:

- The selection of vertical levels for RH retrievals is currently rather crude with four fixed levels. This has somehow simplified the implementation of the framework and also the development of the observation error model, but this aspect needs to be improved. In particular, it was shown in the appendix that the selected 800 hPa level is quite low in the troposphere when considering cold SAPHIR TBs. The forecast degradations noticed close to the surface for temperature and humidity may be related

to this last level. Several possibilities could be considered: either discarding the lowest level when SAPHIR brightness temperatures are too cold following appendix results or computing online SAPHIR channels Jacobians onto the retrieved atmospheric profiles. This latter option would dynamically adapt the vertical levels to either a clear-sky observation, or an observation in a strong scattering situation.

- The slight decrease of positive impacts when activating the inversion error model could lead to several conclusions. Indeed, this model allows the assimilation of observations within deep convection cores but the linearized versions of the forecast model in the ARPEGE 4D-Var does not include a moist convection scheme. The small degradations could indicate that these developments are missing in the ARPEGE linearized physics of the ARPEGE 4D-Var to take benefit from the cold tail of SAPHIR observations. On the other hand, these degradations could also indicate that the inversion error model is overestimating the radiative transfer uncertainties and therefore smooths the weights of each inversion too much which slightly degrades its results. One possibility to keep using multiple microphysics assumptions within the Bayesian inversion would be to use small inversion errors (e.g. 1 K as in the INV-cst_ASSIM-cst experiment), but with multiple microphysics radiative transfer simulations on the neighbouring profiles. This option would take into account directly the nonlinearities of the observation operator by adapting the suitable radiative properties to meteorological situations, but this development is likely to require larger computational resources.

At last, this framework experimented with SAPHIR observations could be easily adapted to assimilate retrievals over land surfaces as well as to other microwave humidity sounders than SAPHIR which is also the next step forward of these developments.

ACKNOWLEDGMENTS

The authors are grateful to Sara Q. Zhang, Ziad Hadad and Alan Geer for fruitful discussions which helped developing this framework. Ghislain Faure is acknowledged for providing the AROME forecasts used in the appendix of the article. M. Michel Dejus (CNES/MT Project manager) and Pierre Tabary (CNES/MT Program manager) are also acknowledged. This work is supported by CNES under the Megha-Tropiques programme. The authors thank the two anonymous reviewers for their useful comments which improved the manuscript.

ORCID

Philippe Chambon  <https://orcid.org/0000-0002-3531-8890>

REFERENCES

- Aonashi, K., Okamoto, K., Tashima, T., Kubota, T. and Ito, K. (2016) Sampling error damping method for a cloud-resolving model using a dual-scale neighboring ensemble approach. *Monthly Weather Review*, 144, 4751–4770. <https://doi.org/10.1175/MWR-D-15-0410.1>.
- Bauer, P., Lopez, P., Benedetti, A., Salmund, D. and Moreau, E. (2006a) Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132, 2277–2306. <https://doi.org/10.1256/qj.05.189>.
- Bauer, P., Lopez, P., Salmund, D., Benedetti, A., Saarinen, S. and Bonazzola, M. (2006b) Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. II: 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132, 2307–2332. <https://doi.org/10.1256/qj.06.07>.
- Bauer, P., Moreau, E., Chevallier, F. and O’Keeffe, U. (2006c) Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society*, 132, 1259–1281. <https://doi.org/10.1256/qj.05.153>.
- Bauer, P., Geer, A.J., Lopez, P. and Salmund, D. (2010) Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quarterly Journal of the Royal Meteorological Society*, 136, 1868–1885. <https://doi.org/10.1002/qj.659>.
- Bennartz, R. and Greenwald, T. (2011) Current problems in scattering radiative transfer modelling for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 137, 1952–1962. <https://doi.org/10.1002/qj.953>.
- Bougeault, P. (1985) A simple parameterization of the large-scale effects of cumulus convection. *Monthly Weather Review*, 113, 2108–2121. [https://doi.org/10.1175/1520-0493\(1985\)113<2108:ASPOTL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<2108:ASPOTL>2.0.CO;2).
- Brognez, H., Clain, G. and Roca, R. (2015) Validation of upper-tropospheric humidity from SAPHIR on board Megha-Tropiques using tropical soundings. *Journal of Applied Meteorology and Climatology*, 54, 896–908. <https://doi.org/10.1175/JAMC-D-14-0096.1>.
- Brognez, H., Fallourd, R., Mallet, C., Sivira, R. and Dufour, C. (2016) Estimating confidence intervals around relative humidity profiles from satellite observations: application to the SAPHIR sounder. *Journal of Atmospheric and Oceanic Technology*, 33, 1005–1022. <https://doi.org/10.1175/JTECH-D-15-0237.1>.
- Brousseau, P., Seity, Y., Ricard, D. and Léger, J. (2016) Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142, 2231–2243. <https://doi.org/10.1002/qj.2822>.
- Buehler, S.A. and John, V.O. (2005) A simple method to relate microwave radiances to upper tropospheric humidity. *Journal of Geophysical Research*, 110, D02110. <https://doi.org/10.1029/2004JD005111>.
- Capderou, M. (2009) *Sampling: comparison with other meteorological satellites*. Megha-Tropiques Technical Memorandum No. 1, pp 51. Available at: <http://meghatropiques.ipsl.polytechnique.fr/download/megha-tropiques-technical-memorandum-n1/> [Accessed 28 January, 2019].
- Caumont, O., Ducrocq, V., Wattrelot, É., Jaubert, G. and Pradier-Vabre, S. (2010) 1D+3DVar assimilation of radar reflectivity data: a proof of concept. *Tellus A*, 62, 173–187. <https://doi.org/10.1111/j.1600-0870.2009.00430.x>.
- Chambon, P. and Geer, A.J. (2017) *All-sky assimilation of Megha-Tropiques/SAPHIR radiances in the ECMWF numerical weather prediction system*. Reading, UK: European Centre for Medium-Range Weather Forecasts, Technical Memorandum 802.
- Chambon, P., Zhang, S.Q., Hou, A.Y., Zupanski, M. and Cheung, S. (2014) Assessing the impact of pre-GPM microwave precipitation observations in the Goddard WRF ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 140, 1219–1235. <https://doi.org/10.1002/qj.2215>.
- Chambon, P., Meunier, L.F., Guillaume, F., Piriou, J.M., Roca, R. and Mahfouf, J. (2015) Investigating the impact of the water-vapour sounding observations from SAPHIR on board Megha-Tropiques for the ARPEGE global model. *Quarterly Journal of the Royal Meteorological Society*, 141, 1769–1779. <https://doi.org/10.1002/qj.2478>.
- Clain, G., Brogniez, H., Payne, V.H., John, V.O. and Luo, M. (2015) An assessment of SAPHIR calibration using quality tropical soundings. *Journal of Atmospheric and Oceanic Technology*, 32, 61–78. <https://doi.org/10.1175/JTECH-D-14-00054.1>.
- Courtier, P. and Geleyn, J. (1988) A global numerical weather prediction model with variable resolution: application to the shallow-water equations. *Quarterly Journal of the Royal Meteorological Society*, 114, 1321–1346. <https://doi.org/10.1002/qj.49711448309>.
- Courtier, P., Freydrer, C., Geleyn, J.F., Rabier, F. and Rochas, M. (1991) The Arpege project at Météo-France. In: *ECMWF Workshop, Annual Seminar on Numerical Methods in Atmospheric Models*, 9–13 September 1991. Reading: European Centre for Medium-Range Weather Forecasts.
- Courtier, P., Thépaut, J. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387. <https://doi.org/10.1002/qj.49712051912>.
- Doherty, A., Indira Rani, S., Newman, S. and Bell, W. (2018) Benefits of assimilating SAPHIR observations on analysis and forecasts of tropical fields in the Met Office global model. *Quarterly Journal of the Royal Meteorological Society*, 144(S1), 405–418. <https://doi.org/10.1002/qj.3258>.
- Errico, R.M., Bauer, P. and Mahfouf, J. (2007) Issues regarding the assimilation of cloud and precipitation data. *Journal of the Atmospheric Sciences*, 64, 3785–3798. <https://doi.org/10.1175/2006JAS2044.1>.
- Fischer, C., Montmerle, T., Berre, L., Auger, L. and Ștefanescu, S.E. (2005) An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, 131, 3477–3492. <https://doi.org/10.1256/qj.05.115>.
- Geer, A.J. and Baordo, F. (2014) Improved scattering radiative transfer for frozen hydrometeors at microwave frequencies. *Atmospheric Measurement Techniques*, 7, 1839–1860. <https://doi.org/10.5194/amt-7-1839-2014>.
- Geer, A.J. and Bauer, P. (2011) Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 137, 2024–2037. <https://doi.org/10.1002/qj.830>.
- Geer, A.J., Bauer, P. and Lopez, P. (2008) Lessons learnt from the operational 1D + 4D-Var assimilation of rain- and cloud-affected SSM/I observations at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, 134, 1513–1525. <https://doi.org/10.1002/qj.304>.
- Geer, A.J., Bauer, P. and Lopez, P. (2010) Direct 4D-Var assimilation of all-sky radiances. Part II: Assessment. *Quarterly Journal of the Royal Meteorological Society*, 136, 1886–1905. <https://doi.org/10.1002/qj.681>.
- Geer, A.J., Baordo, F., Bormann, N., Chambon, P., English, S.J., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. and Lupu, C. (2017) The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, 143, 3189–3206. <https://doi.org/10.1002/qj.3172>.
- Geer, A.J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., Liu, E.H., Collard, A., Bell, W., Migliorini, S., Chambon, P., Fourrié, N., Kim, M.-J., Köpken-Watts, C. and Schraff, C. (2018) All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, 144(713), 1191–1217. <https://doi.org/10.1002/qj.3202>.
- Guerbette, J., Mahfouf, J.F. and Plu, M. (2016) Towards the assimilation of all-sky microwave radiances from the SAPHIR humidity sounder in a limited area NWP model over tropical regions. *Tellus A*, 68(1), 28620. <https://doi.org/10.3402/tellusa.v68.28620>.
- Haddad, Z.S., Steward, J.L., Tseng, H.-C., Vukicevic, T., Chen, S.-H. and Hristova-Veleva, S. (2015) A data assimilation technique to account for the nonlinear dependence of scattering microwave observations of precipitation. *Journal of Geophysical Research – Atmospheres*, 120, 5548–5563. <https://doi.org/10.1002/2015JD023107>.
- 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, 701–722. <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Ikuta, Y. and Honda, Y. (2011) Development of 1D+4DVAR data assimilation of radar reflectivity in JNoVA. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, 41, 01.09–01.10.
- Jones, E.E., Garrett, K. and Boukabara, S. (2017) Assimilation of Megha-Tropiques SAPHIR observations in the NOAA Global Model. *Monthly Weather Review*, 145, 3725–3744. <https://doi.org/10.1175/MWR-D-16-0148.1>.
- Kazumori, M. (2016) Assimilation of Megha-Tropiques/SAPHIR radiance data in JMA’s global NWP system. *CAS/JSC WGNE Research Activities in Atmospheric and Oceanic Modelling*, 46, 01.17–01.18.
- Kulie, M.S., Bennartz, R., Greenwald, T.J., Chen, Y. and Weng, F. (2010) Uncertainties in microwave properties of frozen precipitation: implications for remote sensing and data assimilation. *Journal of the Atmospheric Sciences*, 67, 3471–3487. <https://doi.org/10.1175/2010JAS3520.1>.

- Kummerow, C., Olson, W.S. and Giglio, L. (1996) A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 1213–1232. <https://doi.org/10.1109/36.536538>.
- Lee, S., Chun, H.W. and Song, H.J. (2018) Impact of SAPHIR data assimilation in the KIAPS global numerical weather prediction system. *Atmosphere*, 28(2), 141–151. <https://doi.org/10.14191/Atmos.2018.28.2.141>.
- Liu, G. (2008) A database of microwave single-scattering properties for non-spherical ice particles. *Bulletin of the American Meteorological Society*, 89, 1563–1570. <https://doi.org/10.1175/2008BAMS2486.1>.
- Mahfouf, J. and Bilodeau, B. (2007) Adjoint sensitivity of surface precipitation to initial conditions. *Monthly Weather Review*, 135, 2879–2896. <https://doi.org/10.1175/MWR3439.1>.
- Marécal, V. and Mahfouf, J. (2002) Four-dimensional variational assimilation of total column water vapor in rainy areas. *Monthly Weather Review*, 130, 43–58. [https://doi.org/10.1175/1520-0493\(2002\)130<0043:FDVAOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0043:FDVAOT>2.0.CO;2).
- Migliorini, S., Lorenc, A.C. and Bell, W. (2018) A moisture-incrementing operator for the assimilation of humidity- and cloud-sensitive observations: formulation and preliminary results. *Quarterly Journal of the Royal Meteorological Society*, 144, 443–457. <https://doi.org/10.1002/qj.3216>.
- Minamide, M. and Zhang, F. (2017) Adaptive observation error inflation for assimilating all-sky satellite radiance. *Monthly Weather Review*, 145, 1063–1081. <https://doi.org/10.1175/MWR-D-16-0257.1>.
- Okamoto, K. (2017) Evaluation of IR radiance simulation for all-sky assimilation of Himawari-8/AHI in a mesoscale NWP system. *Quarterly Journal of the Royal Meteorological Society*, 143, 1517–1527. <https://doi.org/10.1002/qj.3022>.
- Okamoto, K., McNally, A.P. and Bell, W. (2014) Progress towards the assimilation of all-sky infrared radiances: an evaluation of cloud effects. *Quarterly Journal of the Royal Meteorological Society*, 140, 1603–1614. <https://doi.org/10.1002/qj.2242>.
- Pereira, M.B. and Berre, L. (2006) The use of an ensemble approach to study the background error covariances in a global NWP model. *Monthly Weather Review*, 134, 2466–2489. <https://doi.org/10.1175/MWR3189.1>.
- Roca, R., Brogniez, H., Chambon, P., Chomette, O., Cloché, S., Gosset, M.E., Mahfouf, J.F., Raberanto, P. and Viltard, N. (2015) The Megha-Tropiques mission: a review after three years in orbit. *Frontiers in Earth Science*, 3, 17. <https://doi.org/10.3389/feart.2015.00017>.
- Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C. and Masson, V. (2011) The AROME-France convective-scale operational model. *Monthly Weather Review*, 139, 976–991. <https://doi.org/10.1175/2010MWR3425.1>.
- Wattrelot, E., Caumont, O. and Mahfouf, J.F. (2014) Operational implementation of the 1D+3D-Var assimilation method of radar reflectivity data in the AROME model. *Monthly Weather Review*, 142, 1852–1873. <https://doi.org/10.1175/MWR-D-13-00230.1>.
- Zhang, S.Q., Zupanski, M., Hou, A.Y., Lin, X. and Cheung, S.H. (2013) Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Monthly Weather Review*, 141, 754–772. <https://doi.org/10.1175/MWR-D-12-00055.1>.
- Zhang, F., Minamide, M. and Clothiaux, E.E. (2016) Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, 43, 2954–2963. <https://doi.org/10.1002/2016GL068468>.
- Zhang, S.Q., Matsui, T., Cheung, S., Zupanski, M. and Peters-Lidard, C. (2017) Impact of assimilated precipitation-sensitive radiances on the NU-WRF simulation of the West African Monsoon. *Monthly Weather Review*, 145, 3881–3900. <https://doi.org/10.1175/MWR-D-16-0389.1>.
- Zhu, Y., Liu, E., Mahajan, R., Thomas, C., Groff, D., Van Delst, P., Collard, A., Kleist, D., Treadon, R. and Derber, J.C. (2016) All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Monthly Weather Review*, 144, 4709–4735. <https://doi.org/10.1175/MWR-D-15-0445.1>.

How to cite this article: Duruisseau F, Chambon P, Wattrelot E, Barreyat M, Mahfouf J-F. Assimilating cloudy and rainy microwave observations from SAPHIR on board Megha Tropiques within the ARPEGE global model. *Q J R Meteorol Soc* 2019;145:620–641. <https://doi.org/10.1002/qj.3456>

APPENDIX A: INFORMATION CONTENT ANALYSIS OF SAPHIR CLOUDY AND RAINY OBSERVATIONS THROUGH RTTOV-SCATT JACOBIANS COMPUTATION

The current implementation of the Bayesian framework described in this article requires us to prescribe pressure levels, used to vertically sub-sample the retrieved atmospheric profiles after the Bayesian inversion. These samples are then assimilated within the ARPEGE 4D-Var data assimilation system. In order to prescribe a first set of pressure levels, a side study was performed to estimate the atmospheric levels which are characterized by a large information content on humidity in the SAPHIR channels, both in clear sky and in strong scattering situations.

An analysis of specific humidity Jacobians derived from the RTTOV-SCATT radiative transfer model simulations was performed on a large number of atmospheric profiles, provided by a NWP regional model called AROME (Seity *et al.*, 2011; Brousseau *et al.*, 2016). This model is characterized by a non-hydrostatic dynamical core and provides more realistic and fine-scale structures of humidity within cloudy and precipitating areas than the global model ARPEGE. Therefore, this model is more likely to represent the variability of SAPHIR observations and is used instead of ARPEGE for this side study. AROME runs operationally with a 3D-Var data assimilation system over western Europe at 1.3 km resolution (Brousseau *et al.*, 2016), but it has also been run since 2016 over five geographical domains in the Tropics to cover French overseas territories. These tropical versions of AROME do not have a data assimilation system yet and are initialized and coupled at the lateral boundaries with the ECMWF model instead of the ARPEGE model for AROME over western Europe. The AROME model used here covers the southwest Indian Ocean (32°E–67°E; 26°S–7°S) at a 2.5 km horizontal resolution with 90 vertical levels and provides forecasts every 6 h up to a +42 h lead time.

For this study, the AROME forecasts initialized at 0000 UTC and 1200 UTC over the January to February 2017 period are used at a single lead time of +36 h. This choice prevents the selected profiles from being contaminated by the spin-up from their initial conditions. These forecasts are sampled every 50 km over the full geographical domain to consider independent samples. For each profile, the Jacobians of SAPHIR channels for humidity are computed with RTTOV-SCATT, normalized with 10%

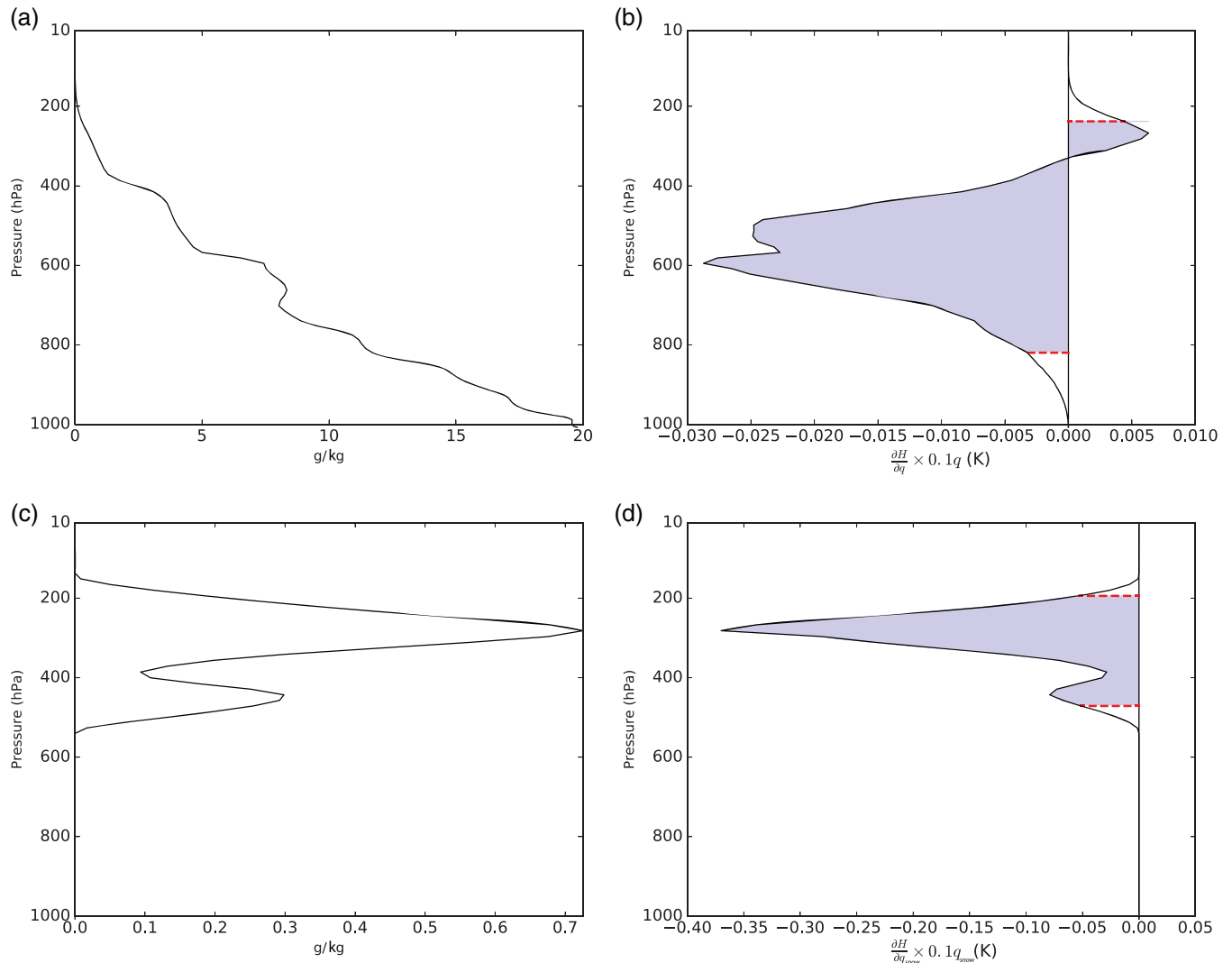


FIGURE A1 (a) Specific humidity profile, (b) corresponding Jacobians for SAPHIR channel 6, normalized by 10% of the specific humidity profile, (c) snow-specific content profile, (d) corresponding Jacobians for SAPHIR channel 6, normalized by 10% of the snow profile. Atmospheric profile comes from the 36 h forecast of the AROME model initialized on 7 January 2017 at 37.5°E and 20°S. The shaded areas in (b) and (d) show the bounds containing 95% of the integrals of normalized Jacobians [Colour figure can be viewed at wileyonlinelibrary.com].

of the used atmospheric profiles and then integrated over the vertical pressure levels. Then lower and upper bounds, respectively called $P_{\text{Lower Bound}}$ and $P_{\text{Upper Bound}}$ are defined as follows so that the integral between the two bounds are equal to 95% of the total integral of the normalized Jacobians:

$$\int_{P_{\text{Surface}}}^{P_{\text{Lower Bound}}} \frac{\partial H}{\partial q} \times 0.1 \times q \times dP = \frac{(1 - \alpha)}{2} \times \int_{P_{\text{Surface}}}^{P_{\text{Model top}}} \frac{\partial H}{\partial q} \times 0.1 \times q \times dP,$$

$$\int_{P_{\text{Upper Bound}}}^{P_{\text{Model top}}} \frac{\partial H}{\partial q} \times 0.1 \times q \times dP = \frac{(1 - \alpha)}{2} \times \int_{P_{\text{Surface}}}^{P_{\text{Model top}}} \frac{\partial H}{\partial q} \times 0.1 \times q \times dP,$$

where α is equal to 0.95, H is the observation operator and q is specific humidity or snow content.

An example of the results of these calculations which provide an estimate of the layers to which SAPHIR channels are sensitive is given in Figure A1. For this particular case, the forward RTTOV-SCATT simulation estimates the brightness temperature for SAPHIR channel 6 to 241 K. The upper and lower bounds of the corresponding Jacobian for humidity (respectively for snow), defined here as bounding 95% of the sensitivity, are 237 hPa (resp. 194 hPa) and 819 hPa (resp. 471 hPa).

These lower and upper bounds, characterizing the levels “from” and “up to” which SAPHIR observations have some sensitivities, have been computed on a large number of atmospheric profiles. The distributions of lower and upper bound pressures have then been categorized by bins of 20 K of SAPHIR channel 6 forward simulations, in order to examine their variations as function of a proxy for scattering intensity.

Figure A2b indicates that a significant number of $P_{\text{Lower Bound}}$ values falls in between 850 and 650 hPa, as

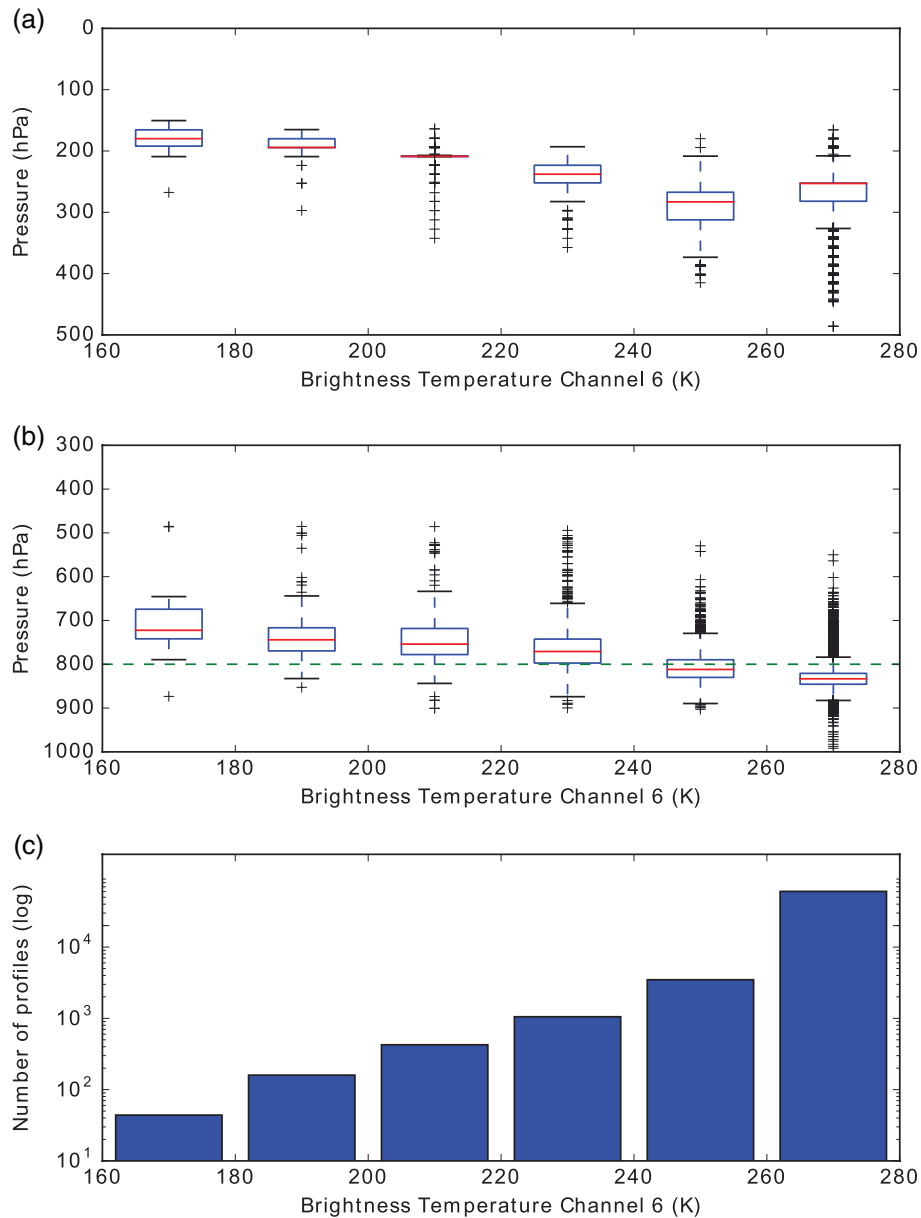


FIGURE A2 Whisker plots of (a) $P_{Upper Bound}$, (b) $P_{Lower Bound}$, for normalized Jacobians of specific humidity for SAPHIR channel 6, and (c) number of atmospheric profiles considered, categorized by bins of 20 K of the simulated brightness temperature for SAPHIR channel 6 with the AROME model (36 h forecasts in January–February 2017 over the Indian Ocean). The ends of the whiskers indicate 1.5 times the interquartile range [Colour figure can be viewed at wileyonlinelibrary.com].

indeed the 25 to 75% quartiles of each of the categories on the horizontal axis are included within these values. The black crosses along the whisker plots also indicate that each of the distributions are characterized by a tail toward higher altitudes. As expected, the colder the forward simulations for SAPHIR channel 6 are, the higher in the atmosphere both bounds are also located. In particular, the 800 hPa value seems a kind of compromise between the clear-sky/low-scattering scenes for which this value lies well within the quartiles, and the stronger scattering scenes for which it lies only in the lower whisker. Regarding the upper bounds, most of the $P_{Upper Bound}$ values range between 150 and 300 hPa.

A sensitivity of these results was performed to the α value for the definition of $P_{Lower Bound}$ and $P_{Upper Bound}$. Two

other values have been tested, 0.99 and 0.90 (not shown) and as expected when increasing the α value, the 800 hPa value is either well contained in the quartiles for all categories of meteorological scenes, or only in the lower whisker when forward simulations are colder than 260 K. These results are therefore sensitive to this prescribed parameter. Nonetheless, they give an indication of what pressure levels can be used for the Bayesian inversion and up to which extent they are adapted to SAPHIR channels sensitivity.

As a first simple implementation for the Bayesian inversion, four fixed pressure levels have been prescribed to retrieve relative humidity (section 3.4.1). Based on these results, an empirical selection was hence performed as 800, 700, 600 and 500 hPa, having in mind that constraining humidity in

the lower layers is likely to be more important for clouds and precipitations than in the upper troposphere. Nonetheless, the results presented here indicate that the lowest level 800 hPa has a limited validity which one could set to brightness temperatures greater than 220 K with the 95% information content calculations. While consistent with the SAPHIR

weighting functions of channel 3 to 6 in clear sky (e.g. Brogniez *et al.*, 2016), these pressure levels may be too low in the atmosphere for the strongest scattering scenes. These results should hence be kept in mind for interpreting the scores of the assimilation experiments presented in this article.

Table des sigles et acronymes

ABI	Advanced Baseline Imager
ADDA	Amsterdam Discrete Dipole Approximation
AEARO	Assimilation d'Ensemble AROME
AEARP	Assimilation d'Ensemble ARPEGE
AHI	Advanced Himawari Imager
AMSR-E	Advanced Microwave Sensing Radiometer for EOS (Earth Observing System)
ARAMIS	Application Radar À la Météorologie Infra-Synoptique
AROME	Applications de la Recherche à l'Opérationnel à Mésos-Échelle
ARPEGE	Action de Recherche Petite Échelle Grande Échelle
ARTS	Atmospheric Radiative Transfer Simulator
ASL	Atmospheric Science Letter
ATMS	Advanced Technology Microwave Sounder
AVHRR	Advanced Very High Resolution Radiometer
BLUE	Best Linear Unbiased Estimator
CEPMMT	Centre Européen de Prévisions Météorologiques à Moyen Terme
CNRM	Centre National de Recherches Météorologiques
COSMO	COnsortium for Small-scale MOdelling
CrIS	Crosstrack Infrared Sounder
DDA	Discrete Dipole Approximation
DDMM	Discrete Dipole Moment Method
DDSCAT	Discrete Dipole SCATtering
DMSP	Defense Meteorological Satellite Program
DPR	Dual Frequency Radar
EBCM	Extended Boundaries Condition Method
ECMWF	European Centre for Medium-range Weather Forecasts
EMV	Estimateur de Maximum de Vraisemblance
EPS-SG	Eumetsat Polar System-Second Generation
ETR	Équation du Transfert Radiatif
EUMETSAT	EUropean organization for the Exploitation of METeorological SATellites
FASTEM	FAST microwave ocean Emissivity Model
FDTD	Finite-Difference Time-Domain
FSO_i	Forecast Sensitivity to Observation impacts
FSS	Forecast Skill Score
GEOS-DAS	Goddard Earth Observing System - Data Assimilation System
GMI	GPM Microwave Imager
GNSS-RO	Global Navigation Satellite System (GNSS)-based radio occultation (RO)
GPM	Global Precipitation Measurement

GPROF	Goddard PROFiling algorithm
IASI	Infrared Atmospheric Sounding Interferometer
IAU	Incremental Analysis Update
ICI	Ice Cloud Imager
IFS	Integrated Forecast System
IGOM	Improved Geometric-Optics Method
II-TM	Invariant Imbedding T-matrix Method
JMA	Japan Meteorological Agency
MetOp-SG	Meteorological Operationnal-Second Generation
MHS	Microwave Humidity Sounder
MODIS	MODerate-resolution Imaging Spectroradiometer
MSG	Metop Second Generation
MTG-I	Meteosat Third Generation - Imager
MWI	MicroWave Imager
MWS	MicroWave Sounder
NASA	National Aeronautics and Space Administration
NCEP	National Center of Environmental Prediction
NFM	Null Field Method
NMC	National Meteorological Center
OI	Optimal Interpolation
OMM	Office Météorologique Mondial
OSCAR	Observing Systems Capability Analysis and Review tool
PNT	Prévision Numérique du Temps
PSD	Particle Size Distribution
PMM	Point Matching Method
RASTA	Radar Airbone System Tool for Atmosphere
RGa	Rayleigh-Gans Approximation
RRTM	Rapid and accurate Radiative Transfer Model
RTTOV	Radiative Transfer for TOVS (TIROS Operational Vertical Sounder)
RTTOV-SCATT	RTTOV SCATTering
SAPHIR	Sondeur Atmosphérique du profil d'Humidité Intertropicale par Radiométrie
SEVIRI	Spinning Enhanced Visible & InfraRed Imager
SOS	Successive Order of Scattering
SSM/I	Special Sensor Microwave/Imager
SSM/I-S	Special Sensor Microwave Imager/Sounder
SSP	Single Scattering Properties
SSRGA	Self Similar Rayleigh-Gans Approximation
SURFEX	SURFace EXternalisée
SVM	Seperation Variable Method
TELSEM	Tool to Estimate Land Surface Emissivity at Microwave frequencies
TIROS-1	Television Infrared Observation Satellite 1
TKE	Turbulent Kinetic Energy
TL/AD	Tangent Linéaire/ADjoint
TMI	TRMM (Tropical Rainfall Measuring Mission) Microwave Imager
UTC	Coordinated Universal Time
VCA	Variationnal Continuous Assimilation
WRF	Weather Research and Forecasting Model
WRF-EDAS	WRF-Based Ensemble Data Assimilation System
3D-VAR	Assimilation de données variationnelle tridimensionnelle
4D-VAR	Assimilation de données variationnelle quadridimensionnelle

Table des figures

1.1	Résolution horizontale de la grille basculée/étirée de la version du modèle ARPEGE opérationnelle depuis 2019.	9
1.2	Processus microphysiques considérés dans le schéma ICE3. Les processus froids sont représentés en bleus et les processus chauds en rouge.	13
1.3	Carte des domaines des modèles AROME-OM (boîtes oranges) et climatologie entre 1980 et 2018 des trajectoires de cyclones (traits verts)	14
1.4	Densité du réseau d'observations (a) en surface, (b) mesurées par radiosondages, (c) par les avions, (d) par les imageurs et sondeurs micro-ondes, (e) par les sondeurs infrarouges hyperspectraux, (f) par les vecteurs de mouvement atmosphériques déduits des imageurs infrarouges, (g) de vents mesurés par les diffusiomètres et (h) et d'angles de courbures mesurés par radio occultation : Global Navigation Satellite System (GNSS)-based radio occultation (RO)	15
1.5	Exemple de ballon-sonde, lancé automatiquement par Météo-France. Les différents instruments mesurant les composantes de l'atmosphère sont situés dans la nacelle à l'extrémité inférieure du dispositif.	16
1.6	Proportions de nombres d'observations disponibles et utilisées par type d'observations pour l'analyse pour les modèles (a) ARPEGE et (b) AROME pour le mois de juin 2022. . . .	18
1.7	Chronologie des satellites et capteurs micro-ondes fournissant des données prêtes à être analysées. Les barres et les flèches indiquent respectivement les observations mesurées à partir d'une fauchée et au nadir. Les barres bleues représentent les radiomètres d'imagerie/à balayage conique, les barres vertes représentent les radiomètres de sondage/cross-track et les barres rouges indiquent les capteurs actifs (radars). Les triangles indiquent les capteurs qui fournissent actuellement des données (au 2021-02-09) et peuvent continuer à le faire, ainsi que les missions futures. (Source des données : basée sur la base de données OSCAR (Observing Systems Capability Analysis and Review Tool) de l'Office météorologique mondial (OMM) et EUMETSAT (EUropean organisation for the exploitation of METeorological SATellites))	19
1.8	Scores FSOi des observations assimilées par le modèle ARPEGE pour les années 2020 et 2021	20
1.9	Version de l'interpolation optimale (OI) consistant à utiliser, pour tous les points d'une boîte d'analyse (en rouge), toutes les observations situées dans une boîte plus grande (en bleue) de sorte que la plupart des observations sélectionnées dans deux boites voisines soient identiques. Ici les observations sont représentées par des points noirs.	22
1.10	Schéma de la méthode du 4D-Var : la prévision commence à partir de l'ébauche X_b et suit la trajectoire en rouge. La prévision corrigée en bleu part de l'analyse X_a qui a été modifiée pour ajuster la nouvelle trajectoire en fonction des observations disponibles dans la fenêtre d'assimilation.	25

2.1	Le spectre électromagnétique	30
2.2	Représentation schématique des différents niveaux d'énergie d'une molécule	31
2.3	Pourcentage de transmission à travers l'atmosphère terrestre le long de la verticale en conditions de ciel clair pour un rayonnement micro-ondes dont la fréquence varie entre 1 et 300 GHz	33
2.4	Représentation de l'angle solide $d\Omega$, avec θ l'angle zénithal et ϕ l'angle azimutal du rayonnement incident.	34
2.5	Illustration du changement de variable entre ds et dz	37
2.6	Schéma illustrant le passage de l'espace du modèle à l'espace des observations à l'aide des propriétés microphysiques et macrophysiques adaptées aux luminances micro-ondes.	38
2.7	Exemples de différentes méthodes pour calculer la fraction nuageuse totale effective.	41
2.8	Exemples de distributions en goutte pour des contenus de pluie allant de 1×10^{-3} à $1 \times 10^{-1} \text{ kg/m}^3$. Les lignes pleines représentent des exemples de distributions de Marshall et Palmer[123] et les lignes pointillées de distributions de Abel et Boutle[124].	44
2.9	Graphique représentant les différents régimes de diffusion en fonction de la longueur d'onde du rayonnement et de la taille des particules considérées provenant d'une présentation de Dubuisson (2018)	46
2.10	Différents hydrométéores solides en fonction de la sursaturation par rapport à la glace et de la température.	48
2.11	Base de données Liu2008 comprenant 11 particules solides contrastées (Liu 2008)	51
3.1	Exemple d'inversion Bayésienne sur les côtes Thaïlandaises. (a) Observations du canal 183.31 ± 11.0 du sondeur SAPHIR le 8 Juin 2017 à 00:00 UTC; (b) estimation du modèle correspondant à la même heure et au même endroit; (c) erreur d'observation du modèle empirique; (d) Résultats de l'inversion en considérant une erreur d'observation de 1 K; (e) Résultats de l'inversion en considérant le modèle d'erreur d'inversion adapté aux scènes observées. Les petites croix noires représentent les endroits où l'inversion Bayésienne a échoué.	59
3.2	Illustration des effets de l'assimilation d'une température de brillance du sondeur MHS à la fréquence 183 GHz, représentée par une croix noire, durant une fenêtre d'assimilation (21:00 UTC jusqu'à 09:00 UTC). La première colonne montre les effets à 21:00 UTC, la deuxième à 00:00 UTC et la troisième à 09:00 UTC. La première ligne montre les contenus en taux neige en kg/m^2 , la deuxième les incréments de pression au niveau moyen de la mer en hPa et la troisième les incréments de contenus en neige.	60
4.1	Nombre cumulé de bases de données de SSPs pour le spectre micro-ondes publiées entre 1995 et 2016. La couleur indique quelles classes de particules sont incluses dans les ensembles de données : les monocristaux sont en bleu, les agrégats en jaune, les particules fondues en rouge et les particules givrées en vert.	64
4.2	Ensemble de formes de particules solides utilisé dans l'étude de Baran et Labonnote (2007)[6].	65
4.3	Moyenne des innovations du modèle IFS et des observations du radiomètre SSMI/S pour le mois de Juin 2012 à la fréquence 183.31 ± 7 pour (1) les sphères de Mie, (2) la SSP <i>Sector snowflake</i> et (3) la SSP <i>Rosette 3-bullets</i> de la base de Liu2008[11]. Les moyennes sont calculées sur une grille $5^\circ \times 5^\circ$	67

4.4	Satellite GPM Core Observatory avec à son bord le radar DPR et le radiomètre micro-ondes GMI. Les fauchées observées par le radar DPR sont représentées respectivement en vert et en rose pour la bande Ku et la bande Ka et la fauchée observée par GMI en bleue.	68
4.5	Distributions des températures de brillance sur océan simulées à la fréquence 183.31 ± 7 GHz avec les particules de la base Liu2008[11] et le modèle ARPEGE sur une période de 10 jours en avril 2021 pour le domaine géographique entre -60° et 60° de latitude. Les distributions des observations du radiomètre GMI sont tracées en noires.	70
4.6	Distributions des températures de brillance sur océan simulées à la fréquence 183.31 ± 7 GHz avec les particules de la base Liu2008[11] et le modèle AROME-Antilles sur une période de 2 mois en septembre et octobre 2017. Les distributions des observations du radiomètre GMI sont tracées en noires.	71
4.7	Ensemble des domaines sur lesquels les calculs de score ont été réalisés au sein de ce manuscrit. Le domaine NORD20, TROPICS et SUD20 sont ici respectivement délimité par des latitudes minimales de 20°N , 20°S et 90°S et maximales de 90°N , 20°N et 20°S .	73
6.1	a,b,c - Diagrammes en violon des distributions des observations nuageuses (définies Section 7.1) en fonction des précipitations de surface GPROF pour chaque SSP de l'expérience 3SSP. e,f,g,h - Idem pour les 4 SSPs de l'expérience 4SSP. d, i - Histogrammes du nombre d'observations nuageuses par intervalle de précipitations de surface respectivement pour les expériences 3SSP et 4SSP.	128
6.2	a,b,c - Diagrammes en violon des distributions des observations nuageuses (définies Section 7.1) en fonction du contenu intégré en glace GPROF pour chaque SSP de l'expérience 3SSP. e,f,g,h - Idem pour les 4 SSPs de l'expérience 4SSP. d, i - Histogrammes du nombre d'observations nuageuses par intervalle de contenu intégré en glace respectivement pour les expériences 3SSP et 4SSP.	129
6.3	Différences relatives de scores de prévisions entre trois expériences, pour le vent, le domaine géographique NORD20 et pour 4 niveaux pressions : 700 hPa, 850 hPa, 925 hPa, 1000 hPa. Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du CEPMMT, pour les échéances d'ARPEGE jusqu'à +102h sur une période de 2 mois. L'expérience de référence, notée 1SSP, assimile les données de GMI en ciel nuageux avec une seule SSP pour simuler les effets de la présence de neige dans l'atmosphère. Les deux autres expériences, nommées 3SSP et 4SSP, assimilent les données de GMI en ciel nuageux avec respectivement 3 ou 4 SSPs pour la neige.	132
6.4	Exemple de calcul des fractions $O_{(n)(i,j)}$ ou $M_{(n)(i,j)}$ pour un point (i,j) représenté par le carré vert. Pour un voisinage n donné, représenté par un carré rouge ici, on calcule la fraction de profils dépassant le seuil q de pluie. Le diagramme en secteur à droite représente cette fraction.	133
6.5	a - Scores FSS des expériences GMI-BAYRAD 1SSP (en rouge) et 3SSP (en noir) sur les Tropiques. b - Différences relatives des scores FSS entre les expériences GMI-BAYRAD 1SSP et 3SSP. c - Idem à (a) pour les expériences GMI-BAYRAD 1SSP (en rouge) et 3SSP (en noir). d - Idem à (c) pour les expériences GMI-BAYRAD 1SSP et 3SSP. Tous les résultats ont été calculés pour des échéances comprises entre 24 et 96 h et des seuils de pluie de 3 (ligne pleine) et 40 mm/hr (ligne pointillée).	135
7.1	Schéma explicatif de l'expérience EDA-MULTI. Pour chaque cycle d'assimilation et chaque membre de l'ensemble, une SSP est tirée aléatoirement et spécifiée à l'opérateur d'observation. Son modèle d'erreur associé est également exploité. Les SSPs pouvant être choisies aléatoirement sont celles des particules suivantes : <i>Thin plate</i> , <i>Rosette 3-bullet</i> et <i>Sector snowflake</i> de la base de données de Liu2008[11].	139

7.2 (a) Écart-types des innovations calculés à partir de prévisions 6 h du modèle ARPEGE et d'observations du radiomètre GMI à la fréquence 183.31 ± 7 GHz en fonction du prédicteur nuageux symétrique et pour la SSP *Thin plate* (en pointillé bleu). (b) Même représentation mais pour la SSP *Rosette 3-bullet*. (c) Même représentation mais pour la SSP *Sector snowflake*. Le nombre de profils de l'ébauche par intervalle du prédicteur nuageux est représenté par l'histogramme en gris. Les courbes rouges représentent les modèles d'erreurs définis pour chaque particule. Ils ont été choisis de manière à coïncider avec les écart-types des innovations et avec celui défini dans Geer et Bauer (2010[26]). 141

7.3 (a) Différences des écart-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI par intervalles du rapport de dispersion β de l'expérience EDA-NOGMI. (b) Histogramme des profils par intervalle du rapport de dispersion β pour l'expérience EDA-NOGMI. (c) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-NOGMI au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 146

7.4 (a) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-CLEARSKY et EDA-NOGMI au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-NOGMI au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 147

7.5 (a) Différences des écart-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY par intervalles du rapport de dispersion β de l'expérience EDA-CLEARSKY. (b) Histogramme des profils par intervalle du rapport de dispersion β pour l'expérience EDA-CLEARSKY. (c) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY au niveau 80 (vers 810 hPa) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-CLEARSKY au niveau 80 (vers 810 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 149

7.6 (a) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-ALLSKY et EDA-CLEARSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-CLEARSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 150

7.7 (a) Différences des écart-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY par intervalles du rapport de dispersion β de l'expérience EDA-ALLSKY. (b) Histogramme des profils par intervalles du rapport de dispersion β pour l'expérience EDA-ALLSKY. (c) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. (d) Carte du rapport de dispersion β de l'expérience EDA-ALLSKY au niveau 80 (vers 810 hPa sur océans) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 151

7.8 (a) Carte des différences des écart-types en humidité spécifique entre les expériences EDA-MULTI et EDA-ALLSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. (b) Carte du rapport de dispersion β de l'expérience EDA-ALLSKY au niveau 100 (vers 990 hPa) du modèle ARPEGE. Les résultats sont moyennés sur une période de 3 mois allant de août à novembre 2021. 152

7.9 Différences relatives de scores de prévisions entre trois expériences, pour l'humidité spécifique, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du Centre Européen, pour les échéances d'ARPEGE jusqu'à +96h pour une période de 2 mois. L'expérience de référence est ici l'expérience 4D-NOGMI qui n'assimile ici aucune données de GMI et est couplée aux matrices **B** de l'expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices **B** des expériences EDA-ALLSKY et EDA-MULTI. 155

7.10 Différences relatives de scores de prévisions entre trois expériences, pour la température, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du Centre Européen, pour les échéances d'ARPEGE jusqu'à +96h pour une période de 2 mois. L'expérience de référence est ici l'expérience 4D-NOGMI qui n'assimile ici aucune données de GMI et est couplée aux matrices **B** de l'expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices **B** des expériences EDA-ALLSKY et EDA-MULTI. 156

7.11 Différences relatives de scores de prévisions entre trois expériences, pour le vent, le domaine géographique NORD20 (a, d), TROPIQUES (b, e), SUD20 (c, f) et pour 2 niveaux pressions : 400 hPa (a, b, c), 925 hPa (d, e, f). Le score de prévision est un écart-type entre les prévisions ARPEGE et l'analyse du Centre Européen, pour les échéances d'ARPEGE jusqu'à +96h pour une période de 2 mois. L'expérience de référence est ici l'expérience 4D-NOGMI qui n'assimile ici aucune donnée de GMI et est couplée aux matrices **B** de l'expérience EDA-NOGMI. Les deux autres expériences, nommées 4D-ALLSKY et 4D-MULTI, assimilent les données de GMI en ciel nuageux et sont couplées respectivement aux matrices **B** des expériences EDA-ALLSKY et EDA-MULTI. 157

7.12 a - Scores FSS des expériences 4D-ALLSKY (en rouge) et 4D-MULTI (en noir) sur les Tropiques. b - Différences relatives des scores FSS entre les expériences 4D-ALLSKY et 4D-MULTI. Tous les résultats ont été calculés pour des échéances comprises entre 24 et 96 h et des seuils de pluie de 3 (ligne pleine) et 40 mm/hr (ligne pointillée). Les barres d'erreurs représentent les différences significatives à 95% 158

Liste des tableaux

1.1	Caractéristiques du modèle ARPEGE	10
1.2	Caractéristiques du modèle AROME-France	11
1.3	Caractéristiques liées aux hydrométéores définis dans le modèle ICE3. Ici p_w représente la densité de l'eau.	14
1.4	Les différentes résolutions du 4D-Var opérationnel à Météo-France	26
2.1	Niveaux d'énergie d'une molécule selon différentes caractéristiques. ΔE représente la quantité d'énergie pour passer d'une énergie E_1 à une énergie de niveau supérieure E_2 en électronvolt (eV, $1 \text{ eV} = 1.602176634 \times 10^{-19} \text{ J}$)	31
2.2	Régime de diffusion selon plusieurs hydrométéores à la fréquence 30 GHz.	47
2.3	Tableau récapitulatif des caractéristiques d'un ensemble de bases de données de SSPs dans le domaine des micro-ondes pour divers hydrométéores solides	52
4.1	Liste des caractéristiques des canaux du radiomètre GMI. IFOV (the Instantaneous Field of View) désigne les champs de visée instantanés et NE Δ T (the Noise Equivalent Delta Temperature) la température équivalente au bruit.	69
4.2	Configuration de RTTOV-SCATT choisie pour les études présentées dans le manuscrit. Grâce à l'équilibre thermodynamique local, l'absorption est ici égale à l'émission.	70
6.1	Observations assimilées au sein du 4D-Var de l'étude	130
6.2	Nombre total d'observations du radiomètre GMI assimilées pour les différentes expériences sur la période d'étude de deux mois. Les observations en ciel clair sont assimilées directement au sein du 4D-Var déterministe tandis que les observations en ciel nuageux sont assimilées à l'aide de la méthode Bayésienne.	131
7.1	Les différentes résolutions de l'AEARP	138
7.2	Pressions atmosphériques sur océans associées aux niveaux verticaux du modèle ARPEGE	140
7.3	Les différentes résolutions du 4D-Var utilisé dans cette étude	153