



HAL
open science

Téledétection satellite du cycle de l'eau - De l'optimisation du traitement des observations à l'intégration des produits restitués

Victor Pellet

► **To cite this version:**

Victor Pellet. Téledétection satellite du cycle de l'eau - De l'optimisation du traitement des observations à l'intégration des produits restitués. Sciences de l'environnement. UPMC - Paris 6 Sorbonne Universités, 2019. Français. NNT: . tel-02868582v1

HAL Id: tel-02868582

<https://theses.hal.science/tel-02868582v1>

Submitted on 8 Nov 2019 (v1), last revised 15 Jun 2020 (v2)

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 DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité : Terre, environnement

École doctorale n° 129 : Sciences de l'environnement d'île-de-France

réalisée au

LERMA, Observatoire de Paris, PSL

sous la direction de Filipe AIRES

présentée par

Victor PELLET

pour obtenir le grade de :

DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Sujet de la thèse :

**Téledétection satellite du cycle de l'eau -
De l'optimisation du traitement des observations à
l'intégration des produits restitués**

soutenue le 9 novembre 2018

devant le jury composé de :

M ^{me} Véronique Ducrocq	CNRM, Toulouse	Rapporteuse
M. Stephen English	ECMWF, Reading	Rapporteur
M. Filipe Aires	LERMA, Paris	Directeur de thèse
M ^{me} Cathy Clerbaux	LATMOS, Paris	Examinatrice
M. Philippe Drobinski	LMD, Palaiseau	Président du jury
M. Fabrice Papa	LEGOS, Toulouse	Examineur

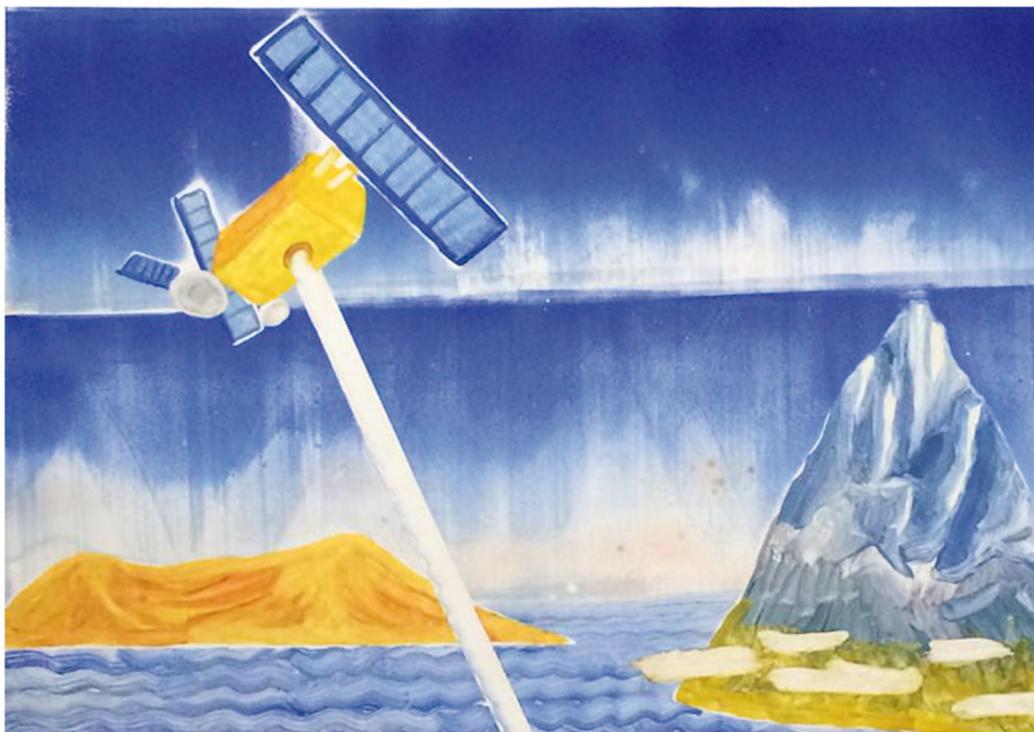


Figure .1. *Maximilien Pellet, l'observation du cycle de l'eau depuis les satellites. Monotype, 2017*

Résumé

L'observation satellite offre l'opportunité unique d'une couverture globale et d'un suivi temporel continu en météorologie et climatologie. On rencontre toutefois de nombreuses difficultés dans leur exploitation, ce qui requiert l'utilisation de techniques d'optimisation dans les diverses étapes de leur traitement. On traitera dans cette thèse deux questions : (1) au niveau du traitement des données satellite, en amont de la chaîne de restitution, et (2) en aval, pour l'intégration des multiples produits satellites disponibles. On propose ainsi une nouvelle approche de compression/sélection de canaux (les "Bottlenecks Channels") pour réduire la dimension des données hyperspectrales (dans l'infrarouge avec l'instrument IASI et dans les microondes pour un nouveau concept de sondeur). On développe également un nouveau schéma d'intégration de multiples produits satellites pour l'étude du cycle de l'eau (surface, océan et atmosphère). Cette approche d'intégration permet de rendre plus cohérentes les composantes du cycle de l'eau. Elle est présentée ici dans le contexte du cycle de l'eau méditerranéen.

Mots clés : Télédétection, cycle de l'eau, observations satellites, optimisation, réduction de dimension, intégration

Abstract

Satellite observation offers the unique opportunity for global coverage and continuous temporal monitoring in meteorology and climatology. However, there are many difficulties in their exploitation, which requires the use of optimization techniques at the different steps of their processing. This thesis will deal with two questions : (1) the processing of satellite data before the restitution, and (2) the integration of the multiple satellite products available. A new approach for channel compression/selection (Bottlenecks Channels) is proposed to reduce the dimension of hyperspectral data (in the infrared with the IASI instrument and in microwaves for a new sounder concept). A new scheme is also developed to integrate multiple satellite products for the study of the water cycle (surface, ocean and atmosphere). This integration approach makes the components of the water cycle more coherent to each other. It is presented here in the context of the Mediterranean water cycle.

Keywords : Remote sensing, satellite observations, optimization, dimension reduction, integration, water cycle

Remerciements

Moi qui voulais devenir astronaute ou faire capitaine d'industrie je suis sur le point de finir ma thèse en science de l'Environnement. Si la formation ingénieur me permit de me spécialiser en traitement du signal et analyse de données, le champ des futurs possibles était encore trop vaste et moi trop éparpillé pour choisir. Mon orientation professionnelle je la dois à une personne et à un café, dans un bar de la Gare de Lyon à Paris. Par la suite, cette personne a continué à jouer le premier rôle dans le début de ma carrière scientifique et a été chronologiquement mon maître de stage, un collègue, et très vite mon directeur de thèse. Depuis plus de quatre ans, Filipe Aires me fait confiance, m'encourage, m'accompagne et me guide. Cette thèse n'aurait pas pu voir le jour sans son aide totale et son encadrement précieux et amical. Jusqu'au derniers instants de la rédaction, ce manuscrit manifeste l'investissement de mon directeur de thèse dans mon projet doctoral et je l'en remercie sincèrement.

Je remercie également Catherine Prigent pour ses conseils, sa bienveillance et son suivi important et assuré, bien que bénévole dans mon projet doctoral. Catherine, que je considère comme ma co-directeur.e de thèse, a toujours su trouver du temps pour moi, et les paroles rassurantes qui me manquaient. Mme Ducrocq et M. English ont acceptés d'être les rapporteurs de cette thèse et je les en remercie. Je tiens également à remercier Annarita Mariotti et Wolfgang Ludwig pour leur implication dans mon comité de thèse, leur vision extérieurs et leurs conseils et les partenaires du projet WACMOS-MED : Gabriel Jordà, Diego Fernandez, Wouter Dorigo, Jan Polcher, et Lucca Brocca pour leurs discussions fructueuses tout au long de mon projet doctoral.

Merci à mes voisin de bureau et d'étage avec qui j'ai passé quatres riches années au LERMA : Binh, Xaoni, Dié, Carlos, Martina. Merci en particulier à Lise, Fabienne et Samuel pour les relectures laborieuses de mes chapitres de thèse et merci à Jordane pour l'initiation à la lunette astronomique et sa disponibilité. Je n'oublie pas les équipes administratives du LERMA (ancienne et renouvelé et Woihiba toujours là) ainsi que le pôle informatique qui m'ont permis de vivre sereinement toutes missions à l'étranger et tout bug informatique. Leurs présence, moteur d'ambiance et de convivialité du LERMA-8e étage, ont assuré un cadre de travail agréable au cours de ces quatre dernières années.

Sur un plan plus personnel, je voudrais remercier Azar Khalatbari et Philippe Lognonne, mes parrain et marraine de sciences, qui, depuis toujours, accompagnent mon épanouissement dans ce domaine avec leurs discussions, conseils et encouragements. Merci aussi à Bernard Pechberty pour son soutien tout au long de ce trajet.

Je remercie également Fifou, mon ami de toujours. Sa personnalité multiple est un

bien que je chérie. Pendant ces trois ans, mais aussi loin que remonte ma mémoire, son amitié demeure un repaire et un appui.

Un tendre merci à Daphné qui est avec moi à chaque instant dans ce projet de recherche mais surtout ailleurs, et bien au-delà. Merci de m'accompagner et de m'offrir un horizon heureux. Je suis enchanté de t'avoir rencontré et de vivre avec toi.

Au terme de cet exercice, je voudrais joindre la dédicace de ma thèse à la mémoire de mon grand-père Jean et au reste de ma famille qui m'a accompagné de près ou de loin, mais toujours, dans mes projets. Merci à mes grand-mères Tatine et Myvonne, à mon grand-père Claude et ma belle-grand-mère Jeannie, pour leur gentillesse à toute épreuve ; Merci à ma mère Martine et mon père Éric, pour leurs enseignements et leur amour ; et merci à Mon grand frère Romain, sa femme Sonia, à ma grande sœur Sandra, son mari Bastien, et à mon frère jumeau Maximilien pour la formidable famille qu'ils me sont. Merci.

Mon manuscrit de thèse a été corrigé une dernière fois avant impression, grâce à l'aide de Véronique Ducrocq, mon père et ma sœur ; et je leurs suis reconnaissant pour cet effort.

Publications

Aires, F., V. Pellet, C. Prigent, and J.-L. Moncet (2015) *Dimension reduction of satellite observations for remote sensing, Part I : A comparison of compression, channel selection, and bottleneck channel approaches*, Quarterly Journal of the Royal Meteorological Society.

Pellet, V., F. Aires, and C. Prigent (2015) *Dimension reduction of satellite observations for remote sensing, Part II : A comparison of compression, channel selection, and bottleneck channel approaches*, Quarterly Journal of the Royal Meteorological Society.

Pellet, V. and F. Aires (2018) *"Bottleneck Channels" algorithm for satellite data dimension reduction : A case study for IASI* IEEE T. Geosci. Remote Sens.

Pellet, V., F. Aires, A. Mariotti, and D. Fernandez (2018a) *Analyzing the Mediterranean water cycle via satellite data integration*, Pure Appl. Geophys.

Pellet, V., F. Aires, S. Munier, G. Jordá, D. F. Prieto, A. Dorigo, J. Polcher, and L. Brocca (2018b) *Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle - Application to the Mediterranean region*, Hydrology and Earth System Sciences, *in review*

Table des matières

I	Introduction générale	1
1	L'observation de la terre par les satellites	1
2	Les défis	2
2.1	Conception d'un produit à partir des observations satellites	2
2.2	Le volume des observations satellites	4
2.3	L'hétérogénéité des produits	5
3	Un exemple d'outil intégrateur : l'assimilation	6
4	Le travail de thèse	8
4.1	La réduction de dimension	8
4.2	L'intégration de produits satellites pour l'étude du cycle de l'eau	9
4.3	Organisation de la thèse	10
II	Le sondage satellite et la réduction de dimension	11
1	Le transfert radiatif	11
1.1	Émission du rayonnement	11
1.2	Réflexion et diffusion	12
1.3	Absorption et énergie interne	13
1.4	Transmittance de l'atmosphère et fonction de poids	14
1.5	Équation du transfert radiatif	15
2	Les sondeurs atmosphériques	16
2.1	Infrarouge	17
2.2	Micro-ondes	19
2.3	Le sondage satellite hyperspectral	19
3	Méthodologie de réduction de la dimension des données hyperspectrales	21
3.1	Compression de données : ACP et variantes	22
3.2	La sélection de canaux	26
3.3	Les "Bottleneck Channels" (BC)	34
4	Réduction de dimension pour un instrument hyperspectral micro-onde	38
4.1	Matériels	38
4.2	Résultats en termes de compression	42
4.3	Résultats en termes de sélection de canaux	46
4.4	Analyse du contenu en information	48
5	Réduction de dimension pour l'instrument IASI dans l'infrarouge	50
5.1	Matériels	51
5.2	Évaluation avec des données simulées	57
5.3	Évaluation avec des données réelles	59
6	Conclusion du chapitre	62

III Bases de données pour l'étude du cycle de l'eau	67
1 Les observations satellites	68
1.1 Précipitations	70
1.2 Evapotranspiration	74
1.3 Contenu en eau dans le sol par gravimétrie	80
1.4 La vapeur d'eau atmosphérique intégrée	83
1.5 La hauteur d'eau par altimétrie	84
2 Les données <i>in situ</i> et de modélisation	85
2.1 Le ruissellement des rivières	85
2.2 Mesure de l'écoulement net au détroit de Gibraltar	86
2.3 Pluviomètres	87
2.4 Sites micro-météorologiques	87
3 La ré-analyse ERA-I de l'ECMWF	88
3.1 Divergence de l'humidité verticalement intégrée dans ERA-I	89
3.2 La vapeur d'eau atmosphérique verticalement intégrée	90
IV L'observation satellite du cycle de l'eau méditerranéen	93
1 La mer Méditerranée et sa région	93
1.1 Le climat méditerranéen	94
1.2 Situation socio-économique et importance de l'eau dans la région	95
1.3 Principaux projets de recherche sur le cycle de l'eau en Méditerranée	97
2 Outils d'analyses	99
2.1 Définition des sous-bassins	99
2.2 Le bilan hydrique	100
3 Une multitude d'observations satellites	101
3.1 Une grande dispersion	101
3.2 Deux méthodes simples d'utilisation des données satellites	104
4 Le cycle de l'eau - premier aperçu	109
4.1 La saisonnalité des composantes du cycle	109
4.2 Le cycle de l'eau	115
4.3 Constat sur le cycle de l'eau en Méditerranée vu par les satellites	117
V Optimisation de données par fermeture du cycle de l'eau	119
1 État de l'art de l'intégration sous contrainte des produits satellites pour l'étude du cycle de l'eau	119
1.1 L'estimateur optimal	120
1.2 Le filtre de Kalman	126
1.3 L'assimilation via le 3D-VAR	130
1.4 Discussion sur les méthodes d'intégration dans le contexte de l'étude du cycle de l'eau par les satellites	132
2 Nouveaux développements théoriques et conditions expérimentales	135
2.1 Hypothèses sur les incertitudes des observations	135
2.2 Fermeture globale du cycle de l'eau : terrestre, océanique et atmosphérique	137
2.3 Fermeture inter-dépendante des sous-bassins	141
2.4 Contraintes hydrologiques à plusieurs échelles temporelles	142

3	Extrapolation spatiale et temporelle	145
3.1	Extrapolation spatiale mois par mois, suivant la distance géographique : INT	145
3.2	Vers une extrapolation temporelle via une calibration (CAL)	147
4	Résultats	149
4.1	Validation des nouvelle solutions intégrées	149
4.2	Les estimations annuelles	155
4.3	Étude des tendances des variables hydrologiques	160
4.4	Estimation des incertitudes des produits satellites	164
5	Conclusion du chapitre	165
VI	Conclusions et perspectives	169
1	Retour sur les objectifs de cette thèse	169
2	La réduction de dimension	170
3	L'intégration sous contrainte	171
A	Équivalence dans le calcul des contaminations dans la restitution des profils atmosphériques	179
B	Autre méthode d'extrapolation spatiale de la contrainte hydrologique avec régression statistique et indice pixelaire	181
	Bibliographie	201

Chapitre I

Introduction générale

1 L'observation de la terre par les satellites

L'observation de la terre par les satellites a considérablement modifié les études sur la météorologie, l'environnement et le climat.

Des réseaux hydrométéorologiques *in situ* se sont rapidement développés dans le monde entier après la Seconde Guerre mondiale, afin d'améliorer les prévisions météorologiques dont l'importance stratégique fut démontrée pendant la guerre. Les collaborations internationales se sont dans un premier temps, limitées à la collecte et à la centralisation de ces données *in situ*. L'Organisation des Nations Unies (ONU) a, par exemple, créé l'Organisation Météorologique Mondiale (OMM) en 1950 pour coordonner les services d'observations hydrologiques à l'échelle mondiale, pour améliorer la qualité de l'observation et pour promouvoir une plus grande collaboration dans les domaines de la météorologie et de l'hydrologie.

L'ère de l'observation satellite commence avec le satellite américain TIROS-1 (1960) qui prend en cliché la Terre pour la première fois depuis l'espace et met en évidence la nature globale et spatialement continue des composantes météorologiques comme les nuages. Mais il faudra attendre les années 1970 et le lancement des premiers satellites géostationnaires de la National Ocean Atmosphere Agency (NOAA) GOES-1 (Geostationary Operational Environmental Satellite-1) pour obtenir des données quantitatives sur la température ou l'humidité atmosphérique, à partir de mesures radiométriques (voir Table I.1).

Les satellites météorologiques ont pour mission de recueillir les données utilisées pour la prévision du temps et l'étude du climat de la Terre. Il s'agit des seules observations permettant de fournir une couverture spatiale globale avec une répétitivité temporelle élevée. Cette propriété est essentielle au-dessus des zones inhabitées qui sont généralement dépourvues de réseaux d'observations *in situ* provenant des stations de surface, radiosondages, avions, bouées... Les zones les moins surveillées *in situ* sont les océans, les zones désertiques, les régions polaires, et, dans une moindre mesure, les pays du Sud, bien moins équipés en réseaux d'observations. Le suivi par satellites se révèle être un outil intégrateur. Pour chaque nouvelle génération de satellite, les capteurs utilisés sont plus performants et capables d'effectuer des mesures de plus en plus complexes donnant lieu à des estimations de plus en plus fiables.

À partir de 1970, des produits satellites globaux sont obtenus à partir des premiers instruments satellites. La communauté du climat commence alors à élaborer des programmes de recherche à l'échelle du globe. Le programme mondial de recherches

sur le climat, lancé en 1980 conjointement par l'OMM et le Conseil International pour la Science, marque les débuts des grands programmes de recherches centrés autour des observations satellites : définition des besoins et des futures missions à partir de collaborations d'agences spatiales nationales.

L'observation satellite est basée : (1) sur une constellation de satellites défilants afin que l'ensemble de la surface terrestre soit observée avec une bonne résolution spatiale (2) et sur des satellites géostationnaires pour une meilleure résolution temporelle. La Table I.1 liste quelques satellites/instruments utilisés pour l'étude du cycle de l'eau. Les instruments d'observations satellites mesurent le rayonnement émis par l'atmosphère et la surface de la Terre (du visible aux micro-ondes en passant par l'infrarouge) afin de restituer quantitativement l'état des variables géophysiques. L'ensemble des instruments d'observation du cycle de l'eau est largement représenté par les sondeurs et les imageurs. L'imageur fournit une observation 2D, sans information sur la structure verticale de la scène. Les sondeurs peuvent, quant à eux, fournir de l'information sur la structure verticale de l'atmosphère (3D) grâce à des observations dans de multiples canaux.

Aujourd'hui, ces produits issus de l'observation satellite sont devenus indispensables pour la prévision du temps et pour l'étude du climat ; ils tendent à remplacer les réseaux *in situ* coûteux et compliqués à entretenir. Lorsque l'UNESCO a lancé la décennie hydrologique internationale en 1965 avec la création du Global Runoff Data Centre (GRDC) pour rassembler les données mondiales sur le débit des rivières, le centre collectait des relevés mensuels de débit dans plus de 3 500 stations et jusqu'à 8 000 stations à son apogée. Depuis, la collecte de données a périclité en raison non seulement de volontés politiques de non-partage d'informations sensibles, comme le stock d'eau, mais aussi du coût d'entretien des réseaux hydrologiques. Par exemple, la mission Surface Water Ocean Topography (SWOT, Table I.1) a pour objectif de mesurer les hauteurs d'eau des rivières à l'échelle du globe afin d'estimer le débit de ces rivières. En ce sens, l'instrument satellite vient palier l'effondrement du GRDC.

2 Les défis

L'avènement de l'observation satellite qu'ont connu ces vingt dernières années soulève des difficultés particulières auxquelles la communauté scientifique doit faire face.

2.1 Conception d'un produit à partir des observations satellites

Les "produits" satellites nécessitent des méthodes d'inversion (i.e. restitution) sophistiquées pour estimer l'état des variables géophysiques à partir de l'observation satellite (i.e. la mesure du rayonnement). La Figure I.1 schématise la genèse de produits satellites dérivés à partir d'observations satellites. Outre les mesures satellites, les données en entrée des algorithmes de restitution peuvent également provenir de modèles numériques de prévisions météorologiques ou d'observations *in situ*. Des modèles physiques (comme les modèles de transfert radiatif inverse) ou bien des modèles statistiques sont utilisés afin de générer un produit satellite représentant une variable géophysique. Ainsi les produits satellites sont plus ou moins directement issus de l'ob-

Plateforme satellite	Début de mission	Fin de mission	Instrument pour le cycle de l'eau	Objets d'intérêts
NOAA-TIROS-1 (Television Infrared Observational Satellite)	avril 1960	juin 1960	deux caméras (un grand angle)	Couverture nuageuse, météo
NOAA-GOES-1 to 15 (Geostationary Operational Environmental Satellite)	Nf1 : octobre 1975 Nf15 : mars 2010	GOES-1 to 11 terminé	Imageur, IR,VIS et sondeurs	température humidité cyclones
ESA-ERS1 (European Remote Sensing Satellite)	juillet 1991	mars 2000	Radar (altimétrie) Sondeurs micro-onde SAR Radiomètre micro-onde	température (surface, atmosphère) topographie vent et humidité
NASA/JAXA-TRMM (Tropical Rainfall Measuring Mission)	novembre 1997	avril 2015	Radar (précipitation) Imageur micro-onde TMI Scanner infrarouge VIRS capteur Radiatif (CERES)	précipitation (3D) nuages flux radiatifs
NASA-Terra NASA-Aqua	février 2000 mai 2002	opérationnel opérationnel	MODIS (Moderate Resolution Imaging Spectrometer)	nuages et propriétés, humidité, température de surface albédo, végétation, neige évapotranspiration
NASA-GRACE (Gravity Recovery and Climate Experiment)	mars 2002	opérationnel	train de deux satellites identiques	gravité terrestre, variation du contenu en eau humidité du sol
ESA/EUMETSAT-MetopA	mars 2006	opérationnel	sondeur IR IASI sondeur micro-ondes AMSU/MHS ASCAT	température humidité, ozone vent
NOAA-POES 1 to 19 (Polar Orbiting Environmental Satellite)	N-19 : février 2009	opérationnel	sondeur micro-onde AMSU-A et MHS Radiomètre AVHRR	Variables atmosphériques, nuages, humidité température
NOAA-Soumi-NPP (N-POES Preparatory project)	octobre 2011	opérationnel	sondeur micro-onde ATMS sondeur IR CrIS imageur et radiomètre VIIR	température, humidité, climat, océan
NASA/JAXA-GPM (Global Precipitation Mission)	février 2014	opérationnel	imageur micro-onde GPM Dual- précipitation radar DPR	précipitation
ESA-Sentinel-1A -1B	avril 2014 avril 2016	opérationnel opérationnel	Radar C-band	glace de mer, courant, vent, inondation tremblement de terre
ESA-Sentinel-2A	juin 2015	opérationnel opérationnel	imageur multi-spectral	végétation, humidité des sols, zones inondées région côtière
NASA/CNES-Swot	2020	- -	Ka-band radar radar interféromètre	hauteur d'eau sur mer et sur lac, rivière.

TABLE I.1 – Quelques exemples de satellites/instruments utilisés pour l'étude du cycle de l'eau.

servation satellite et incorporent plus ou moins de données auxiliaires non satellitaires.

Les erreurs dans la restitution de produits satellites peuvent être causées, par exemple, par (1) le bruit instrumental au niveau de l'instrument, (2) des erreurs d'éta-lonnage, qui ont une incidence sur la précision des mesures, (3) des déficiences inhérentes au processus de restitution/ inversion (4) des erreurs dans les données auxiliaires utilisées. La correction ou la réduction de ces quatre sources d'erreurs suscitées font l'objet de travaux de recherche à part entière. Connaissant la nature des observations satellites et les limites actuelles de la restitution des produits satellites, peut-on développer des outils statistiques qui permettent d'optimiser leur exploitation en l'état actuel ? Dans cette thèse, j'aborderai des problèmes particuliers : le volume des observations, en amont, avant la restitution et l'hétérogénéité des produits satellites dérivés

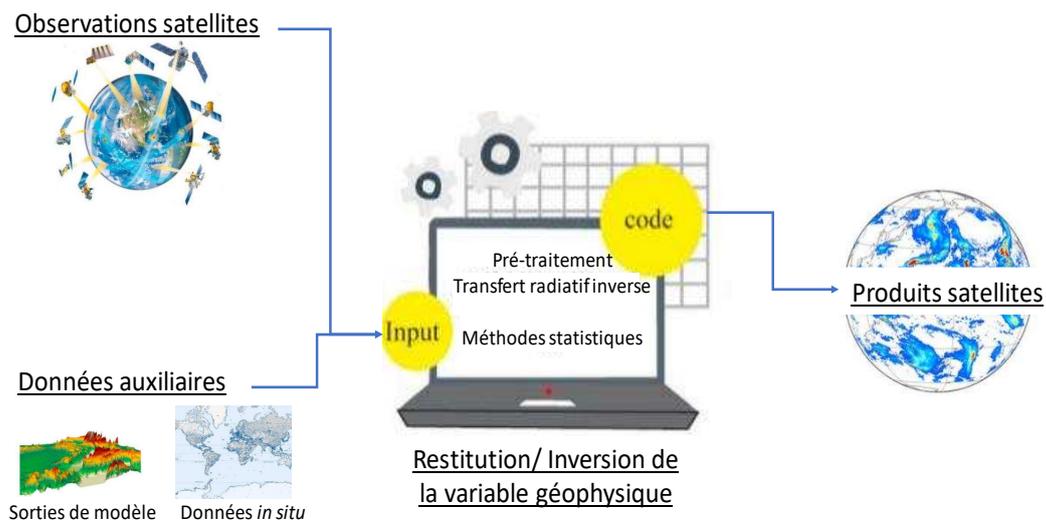


Figure I.1. Schéma de génération d'un produit satellite à partir de la mesure satellite. Modifié à partir de Sun et al. (2018)

en aval.

2.2 Le volume des observations satellites

Le volume des observations satellites ne cesse d'augmenter, non seulement parce que de plus en plus de missions satellites sont lancées, mais aussi parce que les instruments imageurs et sondeurs ont une résolution temporelle, spatiale et/ou spectrale de plus en plus grande, ce qui représente une quantité d'observations de plus en plus importante. Par exemple, l'imageur SEVIRI des satellites Géostationnaire Météosat Second Generation (en orbite depuis 2002) présente un volume d'observations considérablement augmenté par rapport aux observations des capteurs Météosat de la première génération (1977-1997). SEVIRI transmet des images tous les quarts d'heure avec douze canaux spectraux à la résolution horizontale maximale de 1 km . Les capteurs de la première génération ne disposaient que de trois canaux spectraux et transmettaient les images toutes les trente minutes avec une résolution horizontale maximale de 2.5 km . Un autre exemple d'amélioration technique est l'apparition de sondeurs infrarouges hyperspectraux (à très haute résolution spectrale) : ainsi l'instrument Infrared Atmospheric Sounding Interferometer (IASI), lancé en 2006, comporte 8461 canaux (Hilton et al., 2012) et IASI-NG (Crevoisier et al., 2014) qui sera à bord de MetOp-SG dès 2019 possèdera 16922 canaux.

Les performances de ces instruments, à la fois en résolution temporelle, résolution spectrale et en résolution spatiale vont permettre aux météorologistes de décrire plus précisément les processus atmosphériques (Collard et al., 2010) mais cela s'accompagne de difficultés pratiques. Le stockage et la transmission des données peuvent devenir difficiles pour les instruments actuels et futurs. De plus, les algorithmes de restitution et d'assimilation peuvent difficilement traiter une telle quantité d'information.

Dès lors, des approches pertinentes de réduction de l'information sont nécessaires afin de développer des méthodes d'inversion efficaces et rapides, tout en bénéficiant de toute l'information collecté depuis le satellite.

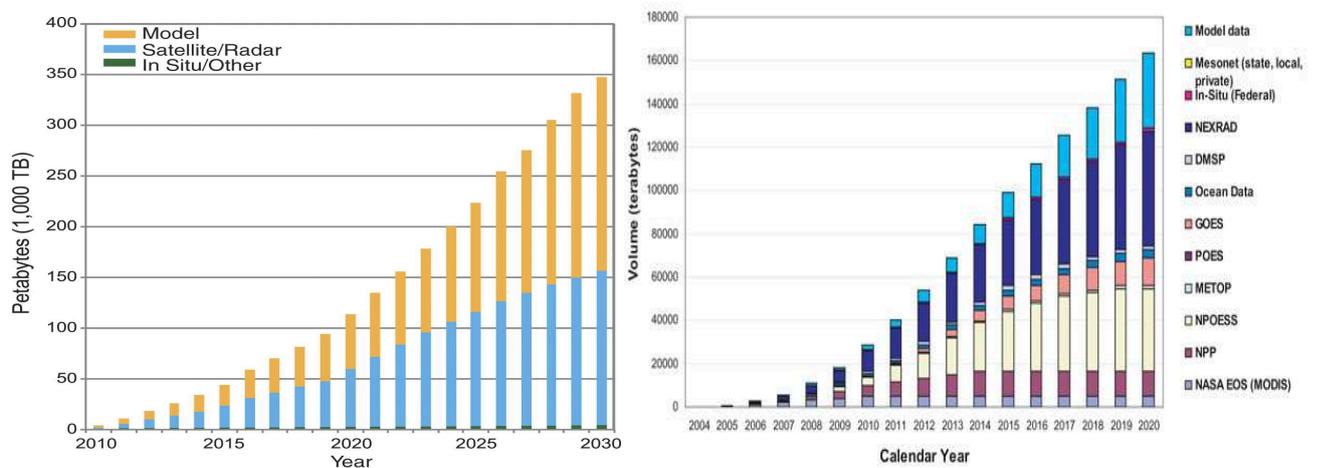


Figure I.2. *Tendance de la croissance mondiale du volume de données climatiques 2010-2030 (Overpeck et al., 2011) (à gauche). Le détail du volume de données climatiques avec, pour la partie observations satellites, le volume en fonction de la plateforme satellite sur 2006-2020 (document de la NOAA) (à droite).*

La Figure I.2 à gauche montre l'augmentation prévue des données climatiques mondiales : les sorties de modèles, les données satellites et les données *in situ*. Avec des valeurs en *Petabytes* = $10^3 \cdot \text{Terabytes}$, le volume total de données satellites sera multiplié par 3, d'ici 2030, par rapport à aujourd'hui (2018), générant des problèmes d'exploitation de plus en plus importants. La Figure I.2 à droite détaille le volume des données sorties des modèles (Model data), celui des données *in situ* (Mesonet, Federal et NEXRAD) ainsi que la part des différentes plateformes satellites dans le volume total actuel d'observations satellites. La plus grande partie des observations est issue des satellites américains de la NOAA (NPOESS, voir Table I.1).

2.3 L'hétérogénéité des produits

Outre le problème du volume "brut" d'observations satellites à gérer de façon pertinente, l'observation satellite est marquée par son nombre croissant de produits géophysiques restitués.

Le processus, depuis la mesure instrumentale par télédétection à bord des instruments satellites jusqu'à l'estimation de variables géophysiques complexes (voir la Figure I.1), incorpore beaucoup de choix plus ou moins subjectifs, pris par les différents acteurs de la chaîne de traitement, et qui diffèrent souvent d'un centre de production à un autre. Il existe plusieurs sources d'incertitudes et de différences entre les produits comme par exemple :

- le nombre et le choix de canaux de l'instrument, utilisés dans la restitution ;
- le transfert radiatif utilisé (en particulier pour les cas complexes comme les nuages ou les précipitations) ;
- la mise en cohérence (spatiale et temporelle) des observations de différents instruments sondeurs, imageurs ou radars ;

- les différents algorithmes de restitution ;
- la diversité des schémas d'assimilation ;
- l'utilisation ou non de données *in situ* pour la calibration.

In fine différentes estimations existent pour une même variable géophysique et ces estimations ne concordent pas toujours entre elles. La différence entre ces estimations atteint parfois 100% de la valeur donnée, ce qui est problématique pour les applications éventuelles.

Il faut noter que la chaîne de traitement est heureusement perfectible au niveau de chaque maillon : depuis le dé-bruitage instrumental jusqu'à la calibration *a posteriori* en utilisant des données *in situ*, en passant par l'amélioration de l'algorithme de restitution.

En l'absence de produits de référence à l'échelle globale, la multitude des produits satellites devient un avantage pour, par exemple, la correction des biais. Il est alors nécessaire d'intégrer les diverses sources d'information, à savoir les produits satellites, afin de s'appuyer au mieux sur la synergie des différentes estimations satellites.

3 Un exemple d'outil intégrateur : l'assimilation

L'assimilation de données est un exemple d'intégration de diverses observations et /ou produits satellites. Pionnière dans les approches multi-instruments et dans les approches à multiples variables géophysiques, l'assimilation est un cadre d'analyse qui permet de résoudre les difficultés mentionnées grâce à des outils physico-statistiques : le pré-traitement des observations par satellite, leurs calibrations, l'inversion et la fusion des observations et /ou des produits satellites avec les sorties de modèles et les données *in situ*.

Dans les centres de la Prévision Numérique du Temps (PNT) comme l'European Center for Medium-range Weather Forecast (ECMWF), les observations de plusieurs instruments dont des hyperspectraux comme IASI et Atmospheric Infrared Sounder (AIRS) ainsi que les mesures de sondeurs micro-ondes, comme l'Advanced Microwave Sounding Unit (AMSU), sont fusionnées afin de contraindre le modèle de prévision météorologique (English et al., 2013). Ici des techniques de réduction du volume des observations hyperspectrales sont nécessaires car les schémas d'assimilation et la puissance de calcul actuelle ne permettent pas d'assimiler tout le spectre mesuré par les instruments hyperspectraux. Collard (2007) a, par exemple, sélectionné 300 canaux sur les 8461 que compte l'instrument IASI.

L'assimilation consiste à faire confiance à la fois à un modèle physique descriptif et aux observations satellites, qu'elles soient instrumentales¹ ou bien déjà dérivées en produits satellites, afin d'estimer l'état le plus probable d'une variable géophysique, état désigné sous le terme "analyse"². La Figure 1.3 montre très schématiquement le principe d'assimilation dans lequel on a un état initial, ébauche, issu du modèle physique, et des observations satellites. À l'issue de l'assimilation, l'état du modèle est mis à jour (c'est-à-dire contraint) par l'information issue des observations satellites et cette "analyse" est alors utilisée comme nouveau point de départ pour incrémenter le modèle et établir une "prévision".

1. c'est-à-dire les spectres mesurés par les instruments satellites.

2. Typiquement les variables atmosphériques dans les modèles de prévision du temps.

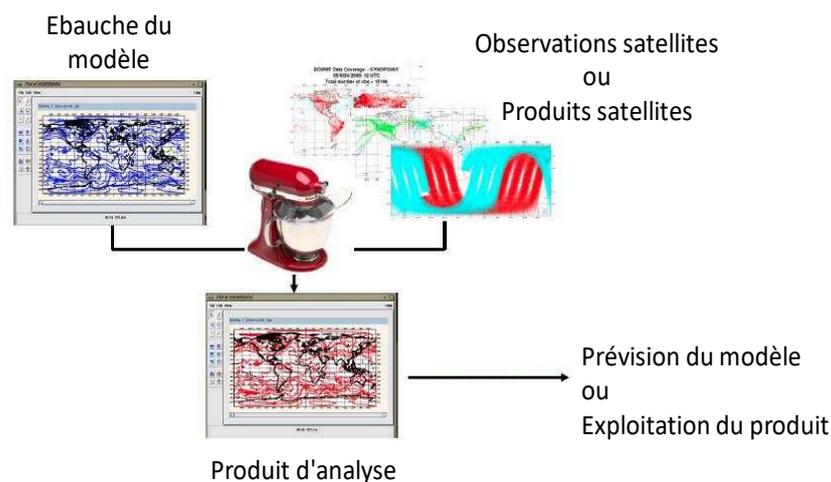


Figure I.3. Schéma illustrant l'assimilation d'observations satellites au sein d'un modèle.

L'approche multi-instruments utilisée en assimilation nécessite des techniques de pré-traitement des observations. Ces techniques de dé-bruitage, de correction de biais et de fusion de données sont couramment utilisées dans l'assimilation. L'utilisation d'un modèle physique permet, entre autres, d'obtenir un produit à une résolution temporelle et spatiale plus fine que les observations par satellites (en faisant confiance à la prévision du modèle entre deux observations ou bien pour la dynamique à l'intérieur du pixel satellite). La cohérence et les résolutions temporelle et spatiale des produits issus de l'analyse présentent des avantages certains qui expliquent que la plupart des études menées sur le climat, et en particulier sur le cycle de l'eau, utilisent davantage ces sorties d'analyse que directement les produits satellites (Mariotti et al., 2002 ; Trenberth et al., 2011).

L'assimilation s'est d'abord développée pour des observations satellites, dans les centres de PNT. Celle-ci nécessite des modèles de transferts radiatifs fiables afin que l'assimilation des observations satellites soit effective. Les incertitudes liées au transfert radiatif pour les variables de surface ont conduit à étendre l'assimilation aux produits satellites déjà restitués. Ainsi, de nombreux schémas d'assimilation ont été développés avec des produits satellites. On parle alors d'approche multiproduits.

En plus d'utiliser l'information provenant des produits satellites pour contraindre les sorties du modèle, on peut implémenter une "contrainte" que l'analyse, en sortie d'assimilation, doit respecter. Par exemple, Pan et Wood (2006) assimilent les produits satellites des différentes composantes du cycle de l'eau continentale (précipitation, évapotranspiration) afin de contraindre les sorties du modèle hydrologique Variable Infiltration Capacity (Liang et al., 1994, VIC) à respecter la fermeture du cycle de l'eau.

Les outils utilisés dans l'assimilation (e.g. fusion, calibration, débruitage, intégration sous contrainte) peuvent également être utilisés pour le traitement des données satellites (sans assimilation dans un modèle). Ainsi, des études récentes tendent à ob-

tenir une base de données à variables géophysiques multiples en intégrant plusieurs produits.

Actuellement, les produits satellites sont moins utilisés que des produits plus intégrés comme les produits de la ré-analyse de l'ECMWF (Dee et al., 2011). L'assimilation répond aux défis que pose l'observation satellite et permet d'obtenir des produits à haute résolution. Mais les produits de ré-analyse incorporent des informations résultant de l'utilisation des modèles descriptifs. Afin d'obtenir des produits au plus proche de l'observation satellite (pour la calibration ou la validation des sorties de modèle par exemple), il faut développer des techniques proches de l'assimilation mais sans l'utilisation de modèles physiques. Les outils statistiques comme les estimateurs optimaux et l'intégration sous contrainte sont prometteurs. Dans cette thèse, seront examinées les deux modalités d'amélioration du traitement des données satellites : la réduction de dimension, en amont, et l'intégration de produits, en aval.

4 Le travail de thèse

4.1 La réduction de dimension

L'un des facteurs de l'augmentation du volume d'observations par satellite, est l'observation hyperspectrale. Un certain nombre d'instruments satellites se sont éloignés des imageurs et des sondeurs traditionnels au profit d'instruments dits hyperspectraux qui possèdent des milliers de canaux comme IASI (Hilton et al., 2012) et IASI-NG (Crevoisier et al., 2014). Alors qu'un instrument traditionnel moyennait le rayonnement reçu sur une certaine bande spectrale, les instruments hyperspectraux, avec leurs milliers de canaux, échantillonnent de façon très fine le rayonnement et permettent une résolution spectrale beaucoup plus grande. Par rapport aux sondeurs traditionnels, les sondeurs hyperspectraux peuvent mieux résoudre la structure verticale des variables atmosphériques comme la température ou l'humidité en discrétisant plus finement la colonne atmosphérique.

La Figure I.4 compare la capacité de deux sondeurs infrarouges à échantillonner verticalement l'atmosphère. Les pics des courbes indiquent l'altitude à laquelle les canaux de l'instrument permettent une mesure. Plus la courbe présente un pic étroit, plus la mesure à cette altitude est précise. Les observations hyperspectrales permettent de collecter d'importantes informations sur l'état de l'atmosphère et les propriétés de surface : profil de température et d'humidité atmosphérique, contenu en gaz mineurs, température de surface de la mer. Toutes ces mesures nourrissent et améliorent les prévisions numériques.

Toutefois, l'augmentation du nombre d'observations satellites génère des difficultés aussi bien théoriques que pratiques, et en particulier pour les applications en quasi-temps réel. L'utilisation directe des données originales (spectre complet d'observations d'un instrument hyperspectral) ne peut satisfaire les besoins des centres de prévision numérique du temps et des méthodes de réduction de dimension pertinentes (on parle de dimension spectrale pour caractériser le nombre de mesures aux différentes longueurs d'onde dans une observation hyperspectrale) doivent être mises en place.

Des méthodes classiques de réduction de dimension existent en opérationnel : la compression de données et la sélection de canaux. J'ai développé une approche innovante de réduction de dimension des spectres à la croisée de ces deux approches gé-

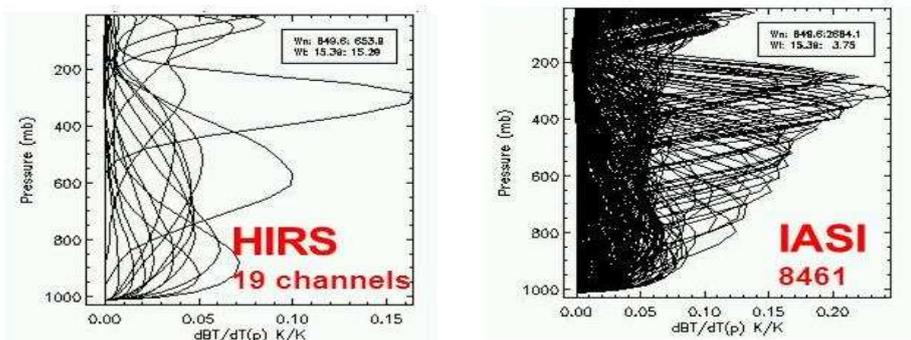


Figure I.4. Capacité d'échantillonnage verticale de l'atmosphère pour un sondeur traditionnel infrarouge à 19 canaux, le High-resolution Infrared Radiation Sounder (HIRS/4) (à droite) et pour le sondeur hyperspectral infrarouge IASI et ses 8461 canaux (à gauche).

nérales : les "Bottleneck Channels". Dans cette approche, une compression de l'information est obtenue mais avec une contrainte imposée aux composantes : celle d'être au plus proche des canaux réels.

L'étude de la réduction de dimension présentée dans cette thèse est issue d'un travail de master. J'ai poursuivi ce travail en parallèle de mon travail de thèse sur l'optimisation des données satellites pour l'étude du cycle de l'eau.

4.2 L'intégration de produits satellites pour l'étude du cycle de l'eau

S'inspirant du cadre de l'assimilation, il est possible de développer des méthodes d'intégration de nombreux produits satellites afin de les rendre plus cohérents entre eux et de réduire leurs biais et leurs incertitudes. Ainsi Sheffield et al. (2009) et Munier et Aires (2017) estiment les incertitudes de plusieurs produits satellites pour toutes les composantes du cycle de l'eau continentale. Cette approche à multiples variables, estimées par plusieurs produits dérivés, est de plus en plus importante dans la communauté des utilisateurs de produits satellites, notamment dans l'étude du climat et du cycle de l'eau.

Nous avons choisi de travailler sur le cycle de l'eau en Méditerranée, le pourtour méditerranéen étant un "hot-spot"³ du changement climatique. Une analyse pertinente du cycle de l'eau est capitale pour la région méditerranéenne où les répercussions socio-économiques de l'appauvrissement en eau sont cruciales à l'échelle internationale comme le montre l'actuelle crise des migrants "climatiques"⁴ et sa gestion par l'Europe.

Dans le contexte d'optimisation de produits satellites, la fusion de multiples produits

3. zone très sensible au changement climatique.

4. La Banque mondiale estime que 140 millions de personnes vont migrer d'ici 30 ans pour fuir les effets du réchauffement climatique et en particulier les pénuries d'eau.

satellites pour une même composante du cycle de l'eau réduit les incertitudes d'estimation pour cette variable et une contrainte hydrologique est utilisée afin de rendre cohérentes les diverses estimations des composantes du cycle de l'eau. Dans son rapport spécial, le Groupe d'experts Intergouvernemental sur l'Evolution du Climat (GIEC) sur "la gestion des risques de catastrophes et de phénomènes extrêmes pour les besoins de l'adaptation au changement climatique" (2012), définit le cycle de l'eau comme "Le cycle dans lequel l'eau s'évapore des océans et de la surface terrestre, est transportée par la circulation atmosphérique sous forme de vapeur d'eau, se condense pour former des nuages, précipite sous forme de pluie ou de neige, est interceptée par les arbres et la végétation, assure le ruissellement à la surface terrestre, s'infiltré dans les sols, recharge les eaux souterraines et/ou ruisselle dans les cours d'eau, s'écoule dans les océans et, finalement s'évapore à nouveau des océans ou de la surface terrestre". Il est alors important que ce cycle soit fermé. La prise en compte de l'équation du bilan hydrique comme contrainte d'intégration permet d'obtenir une meilleure estimation du bilan hydrique de la région méditerranéenne.

4.3 Organisation de la thèse

Le résultat de ma recherche est organisé en six chapitres.

Le Chapitre 2 présente des techniques de compression et de sélection de canaux pour la restitution de variables atmosphériques par observations hyperspectrales. Nous présenterons notre méthode de réduction de dimension les "Bottleneck Channels" qui sera évaluée dans le cadre d'un concept d'instrument hyperspectral micro-ondes et de l'instrument IASI.

Le Chapitre 3 présente les divers produits satellites disponibles pour l'étude du cycle de l'eau en Méditerranée. Une attention particulière est portée sur leurs algorithmes de restitution et les observations satellites qu'ils utilisent.

Le Chapitre 4 met en évidence les capacités, différences et limites des produits satellites pour l'étude du cycle de l'eau en Méditerranée. Un premier aperçu du cycle de l'eau est obtenu.

Le Chapitre 5 s'attache à présenter des méthodes d'intégration de produits satellites, avec une contrainte de fermeture du bilan hydrique. De nouvelles estimations sont données pour les différentes composantes du cycle, une analyse des incertitudes est obtenue pour les différents produits satellites dérivés.

Dans le Chapitre 6, les conclusions et perspectives de ce travail de thèse sont détaillées.

Chapitre II

Le sondage satellite et la réduction de dimension

Où l'on présente le transfert radiatif, principe physique sous-jacent à la télédétection satellite, et son utilisation pour le sondage dans l'infrarouge et les micro-ondes. Et où l'on se rend compte de la nécessité de réduire la dimension des observations hyperspectrales de façon pertinente pour la restitution des variables géophysiques. Les techniques de réduction de dimension classiques sont comparées à une nouvelle approche : les "Bottleneck Channels" dans le contexte d'un concept d'instrument micro-onde ainsi que celui du sondeur IASI. Ce chapitre s'appuie sur (Aires et al., 2015; Pellet et al., 2015; Pellet et Aires, 2018)

1 Le transfert radiatif

Le transfert radiatif décrit le trajet d'un rayonnement à travers l'atmosphère, depuis la surface de la terre vers l'instrument satellite. La luminance énergétique spectrale (ou radiance) est la grandeur physique mesurée par l'instrument satellite (puissance par unité de surface, par unité d'angle solide, visée de l'instrument, et pour une longueur d'onde donnée en $W.m^{-2}.sr^{-1}.m^{-1}$). Lorsque ce rayonnement traverse le milieu atmosphérique, il subit plusieurs processus physiques : l'absorption (α), la réflexion (ρ) ou la transmission (τ), dépendantes de la longueur d'onde λ et liées par la conservation de l'énergie ($\rho_\lambda + \alpha_\lambda + \tau_\lambda = 1$).

Dans cette section, j'expliquerai l'origine de la radiance à partir de la loi de Planck ainsi que les processus physiques intervenant dans le transfert radiatif. Finalement, je donnerai une équation du transfert radiatif, sous certaines hypothèses simplificatrices utilisées en télédétection opérationnelle. Cette section est inspirée des manuscrits de thèse de Paul (2013) et Vincensini (2013).

1.1 Émission du rayonnement

1.1.a La théorie du corps noir

Un corps noir est un corps parfait qui absorbe tout rayonnement incident et le ré-émet et ce dans toutes les directions et quelle que soit la longueur d'onde. La théorie de Planck énonce que pour un corps noir en équilibre thermodynamique et à une température donnée, la distribution spectrale du rayonnement émis ne dépend que de la

température du corps et de la longueur d'onde. Ainsi pour une longueur d'onde λ , la fonction de Planck est donnée par :

$$B_{\lambda}(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda kT}} - 1}, \quad (\text{II.1})$$

où T est la température (en Kelvin), h la constante de Planck ($h = 6.6237 \cdot 10^{34} \text{ J.s.}$), k la constante de Boltzmann ($k = 1.381 \cdot 10^{23} \text{ J.K}^{-1}$) et c la vitesse de la lumière dans le vide ($c=299\,792\,458 \text{ m.s}^{-1}$). En d'autres termes, à une température donnée, la luminance est déterminée pour chaque longueur d'onde. La fonction de Planck étant inversible, la luminance émise à une certaine longueur d'onde, ne dépend que de la température du corps noir. Pour les températures des surfaces terrestres, le maximum d'émission, selon la théorie du corps noir, est situé dans l'infrarouge vers $10 \mu\text{m}$, et le rayonnement émis dans les longueurs d'ondes du visible est à peu près nul.

1.1.b Le "corps gris "

Un corps naturel est considéré comme un corps gris. À la différence du corps noir, un corps gris n'absorbe qu'une partie du rayonnement incident, le reste est réfléchi et/ou transmis. Aussi, à température égale, un corps gris n'émet pas autant qu'un corps noir. L'expression de la luminance spectrale d'un corps gris à la température T est :

$$I_{\lambda}(\theta, \phi, T) = \epsilon_{\lambda}(\theta, \phi, T) \cdot B_{\lambda}(T), \quad (\text{II.2})$$

où $\epsilon_{\lambda}(\theta, \phi, T)$ est l'émissivité directionnelle spectrale (à la longueur d'onde λ) dans la direction (θ, ϕ) et $B_{\lambda}(T)$ est la luminance spectrale d'un corps noir de température T .

1.1.c L'émissivité de la surface continentale

L'émissivité directionnelle spectrale de la surface terrestre est donc le rapport entre la luminance émise par la surface réelle et la luminance émise par un corps noir à la même température. Cette émissivité dépend généralement de la direction d'émission, de la température de la surface, de la longueur d'onde, ainsi que d'autres propriétés de la surface comme l'indice de réfraction et la composition chimique. L'émissivité varie entre 0 et 1 (l'émissivité d'un corps noir est égale à 1). La loi de Kirchhoff permet d'établir l'égalité entre l'émission et l'absorption à l'équilibre thermodynamique local. Pour les surfaces continentales, considérées à l'équilibre thermodynamique, on a $\epsilon_{\lambda} = 1 - \rho_{\lambda}$ où ρ_{λ} est la réflectance spectrale à la longueur d'onde λ soit le rapport entre l'énergie réfléchi et l'énergie incidente en fonction de la longueur d'onde. Elle est comprise entre 0 (corps noir) et 1 (tout est réfléchi).

1.2 Réflexion et diffusion

En fonction de la nature de la surface réfléchissante et de la longueur d'onde, deux types de réflexion sont observés :

Lien entre température de brillance, luminance et température réelle

Un corps à une température physique T émet donc un rayonnement moins énergétique que celui du corps noir à la même température physique. On lui associe alors une température de brillance TB définie comme étant la température d'un corps noir qui rayonnerait la même luminance (dans la direction (θ, ϕ)) que ce corps naturel à la température physique T :

$$TB = B_{\lambda}^{-1}[\epsilon(T) \cdot B_{\lambda}(T)] \quad (\text{II.3})$$

où B_{λ}^{-1} est l'inverse de la fonction de Planck. La notion de température de brillance (TB) sera utilisée par la suite pour caractériser le rayonnement reçu au niveau de l'instrument satellite lors de la télédétection. L'inversibilité de la fonction de Planck assure la bijection entre radiance et température de brillance à une longueur d'onde donnée.

- *La réflexion directionnelle* se produit lorsque les irrégularités géométriques de la surface sont très petites par rapport à la longueur d'onde du rayonnement incident. La surface, plane à l'échelle de la longueur d'onde, réfléchit le rayonnement selon la loi de Descartes (angle de réflexion = angle d'incidence).
- *La réflexion diffuse* s'observe quand la surface est rugueuse à l'échelle de la longueur d'onde. L'énergie réfléchie est diffusée dans toutes les directions.

La diffusion caractérise le rayonnement incident dévié et dispersé dans diverses directions par des particules. Deux régimes sont considérés :

- *La diffusion de Rayleigh* lorsque la longueur caractéristique de la particule diffusante est inférieure à la longueur d'onde du rayonnement incident. La diffusion de Rayleigh est une diffusion considérée comme isotrope (les particules émettent dans toutes les directions de façon égale) et est inversement proportionnelle à la longueur d'onde à la puissance quatre ($1/\lambda^4$). Ainsi, plus la longueur d'onde est petite, plus le pouvoir diffusant des molécules est fort.
- *La diffusion de Mie* lorsque la longueur caractéristique de la particule diffusante est du même ordre de grandeur que la longueur d'onde du rayonnement incident. L'intensité du rayonnement diffusé varie alors de façon inversement proportionnelle à la longueur d'onde ($1/\lambda$) et de façon anisotrope.

1.3 Absorption et énergie interne

Tout corps peut absorber une partie du rayonnement incident qu'il reçoit. Le rapport entre l'énergie absorbée et celle reçue à la longueur d'onde λ est appelé coefficient d'absorption α_{λ} , compris entre 0 et 1. L'énergie absorbée augmente l'énergie interne du corps.

À l'échelle moléculaire, l'énergie interne est définie par le niveau d'excitation des électrons autour du noyau. Chaque molécule peut se trouver dans différents états électroniques, correspondant à des énergies internes plus ou moins hautes. Ces états électroniques sont subdivisés en plusieurs états de vibration et de rotation. L'énergie absorbée peut donc entraîner des processus de rotation et de vibration (faible changement d'énergie), d'excitation voire de dissociation ou d'ionisation (grand changement

Équation de Beer-Lambert

Soit un coefficient d'extinction du milieu pour la longueur d'onde λ , $K_\lambda = k_\lambda + \sigma_\lambda$, avec k_λ le coefficient d'absorption et σ_λ le coefficient de diffusion (incluant toutes les différentes diffusions). La variation dI_λ de la luminance incidente I_λ (i.e., intensité spectrale) lors de la traversée une couche atmosphérique infinitésimale de longueur dl suit la loi de Beer-Lambert :

$$dI_\lambda = -K_\lambda \cdot I_\lambda \cdot dl \quad (\text{II.4})$$

En intégrant l'équation précédente sur l'épaisseur de l'atmosphère entre la surface (altitude 0) et l'altitude z avec un angle de visée θ , la luminance $I_\lambda(z, \theta)$ vaut :

$$I_\lambda(z, \theta) = I_\lambda(0) \cdot e^{-\frac{\int_0^z K_\lambda(z) dz}{\cos\theta}} \quad (\text{II.5})$$

d'énergie). Le rayonnement dans le micro-onde contient une faible énergie (l'énergie E d'un photon est proportionnelle à sa fréquence de propagation f : $E = h \cdot f$) et ne peut qu'engendrer des rotations de la molécule. Dans l'infrarouge, l'énergie transportée est plus importante et peut donc engendrer des vibrations en plus des rotations. Les rayonnements dans le domaine visible et ultra-violet transportent encore plus d'énergie et peuvent engendrer des changements de l'état électronique des molécules, voire des dissociations ou des ionisations.

Afin de retrouver un état stable (non excité), les molécules de l'atmosphère perdent de l'énergie en émettant un rayonnement. L'émission aura lieu dans les mêmes bandes spectrales que l'absorption mais pas forcément dans la même direction que le rayonnement incident (diffusion).

1.4 Transmittance de l'atmosphère et fonction de poids

Tout corps peut transmettre une partie du rayonnement incident qu'il reçoit. Le rapport entre l'énergie transmise et l'énergie incidente est appelé coefficient de transmission (ou transmittance) τ_λ , compris entre 0 et 1. Dans l'hypothèse d'une atmosphère plane et parallèle, absorbante et diffusante, la transmittance $\tau_\lambda(z, \theta)$ à une longueur d'onde λ d'un rayonnement quittant l'atmosphère vers l'espace avec un angle θ et à l'altitude z est définie *via* l'intégration de l'équation de Beer-Lambert entre la surface et l'altitude z .

On définit alors le facteur de transmittance atmosphérique à la longueur d'onde λ comme :

$$\tau_\lambda(z, \theta) = e^{-\frac{\int_0^z K_\lambda(z) dz}{\cos\theta}}. \quad (\text{II.6})$$

1.4.a Fonction de poids

La fonction de poids à la longueur d'onde λ et à l'altitude z est la dérivée de la transmission atmosphérique, elle s'écrit donc :

$$\frac{d\tau}{dz}(z) = \frac{K_\lambda(z)}{\cos\theta} \cdot e^{-\frac{\int_0^z K_\lambda dz}{\cos\theta}}. \quad (\text{II.7})$$

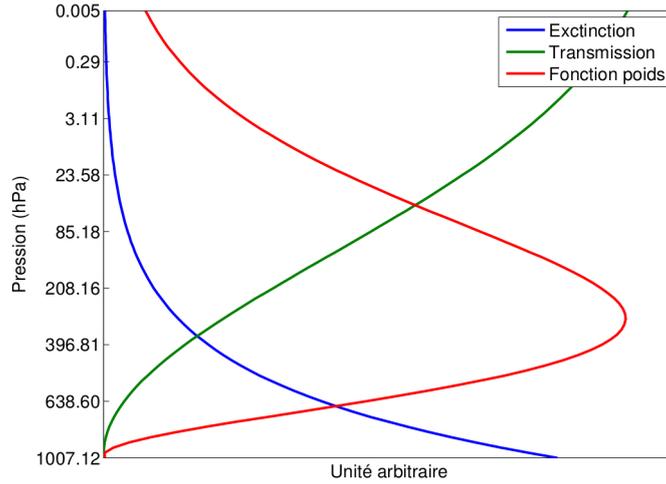


Figure II.1. Sur le graphique, la variation du coefficient d'extinction atmosphérique est représentée en bleu. La transmission atmosphérique est en vert, et la fonction de poids résultante est représentée en rouge. Ces courbes sont données à titre d'exemple, elles ne représentent pas une longueur d'onde particulière (Paul, 2013).

Le coefficient d'extinction K_λ augmente avec le nombre de molécules par volume d'air et donc décroît avec l'altitude (atmosphère de moins en moins dense). Inversement, la transmission atmosphérique croît de 0, à la surface, jusqu'à 1, au sommet de l'atmosphère (pour une visée au nadir). Ainsi la fonction de poids croît jusqu'à une certaine altitude qui correspond à l'altitude du maximum de sensibilité, puis elle décroît jusqu'au sommet de l'atmosphère. La Figure II.1 représente un exemple schématique de fonction de poids suivant les variations du coefficient d'extinction atmosphérique. L'absorption atmosphérique (et donc les pics des fonctions de poids) varie en fonction de la longueur d'onde considérée. En analysant le rayonnement à plusieurs longueurs d'onde, il est possible de restituer des informations sur la température, l'humidité et la concentration en gaz tout le long du profil vertical atmosphérique.

1.5 Équation du transfert radiatif

Dans le cas d'un instrument observant la Terre avec un angle θ par rapport au nadir, l'équation du transfert radiatif est résolue depuis le sol jusqu'au satellite en supposant une atmosphère plane et parallèle afin de simplifier le problème. La Figure II.2 illustre les différentes composantes du rayonnement depuis la surface vers le satellite. Dans cette configuration, le capteur reçoit un rayonnement $I_\lambda^{sat}(\theta)$ dû à la fois : (1) à l'émission naturelle de la surface terrestre $I_\lambda^{surf}(\theta)$ en partie atténuée par l'atmosphère, (2) au rayonnement ascendant émis par l'ensemble des couches atmosphériques $I_\lambda^{atm,\uparrow}(\theta)$ et enfin (3) au rayonnement descendant de l'atmosphère $I_\lambda^{atm,\downarrow}(\theta)$ (excluant la contribution du soleil) réfléchi par la surface puis atténué par l'atmosphère :

$$I_\lambda^{sat}(\theta) = \tau_\lambda(z_{sat}, \theta) I_\lambda^{surf}(\theta) + I_\lambda^{atm,\uparrow}(\theta) + (1 - \epsilon_\lambda^{surf}) \tau_\lambda(z_{sat}, \theta) I_\lambda^{atm,\downarrow}(\theta), \quad (\text{II.8})$$

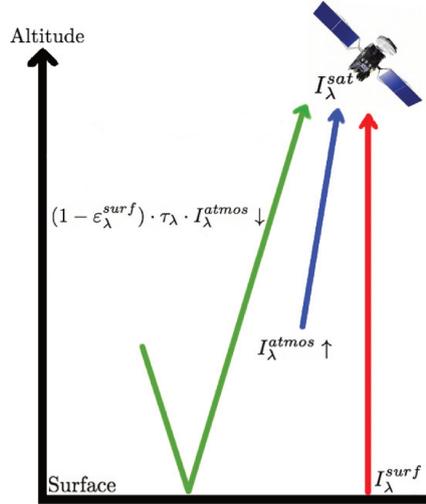


Figure II.2. Schéma des différents rayonnements reçus par le satellite au sommet de l'atmosphère : l'émission de la surface, le rayonnement montant émis par l'atmosphère et le rayonnement descendant émis par l'atmosphère puis réfléchi par la surface (Paul, 2013).

où la surface de la terre est un corps gris à la température T^{surf} et d'émissivité $\epsilon_\lambda^{surf}(\theta)$ et de réflectance $(1 - \epsilon_\lambda^{surf})$, avec :

$$I_\lambda^{surf}(\theta) = \epsilon_\lambda^{surf}(\theta) B_\lambda(T^{surf}). \quad (\text{II.9})$$

Chaque couche de l'atmosphère de largeur infinitésimale dz' et de transmission $d\tau$ est approximée par un corps noir à l'équilibre thermodynamique, on a alors :

$$I_\lambda^{atm,\uparrow}(\theta) = \int_0^{z_{sat}} B_\lambda(T(z')) \frac{d\tau}{dz'} dz', \quad (\text{II.10})$$

$$I_\lambda^{atm,\downarrow}(\theta) = \int_{z_{sat}}^0 B_\lambda(T(z')) \frac{d\tau}{dz'} dz', \quad (\text{II.11})$$

Le principe de télédétection passive est basé sur cette équation du transfert radiatif. À chaque longueur d'onde, la luminance (radiance ou température de brillance) émise est fonction de la composition atmosphérique le long du profil vertical atmosphérique. En inférant plusieurs mesures de luminance à différentes longueurs d'ondes, il est possible de restituer le profil vertical des concentrations en gaz et en déduire l'état de l'atmosphère ainsi que des caractéristiques de surface comme la température de surface. Les instruments sondeurs et imageurs sont basés sur ce principe pour quantifier l'état les variables géophysiques atmosphériques et de surface.

2 Les sondeurs atmosphériques

La Figure II.3 présente l'absorption atmosphérique dans les différentes gammes de rayonnement électromagnétique (depuis les rayons gamma jusqu'aux ondes radio, en passant par le visible, l'infrarouge mesuré par le sondeur IASI et le micro-onde mesuré

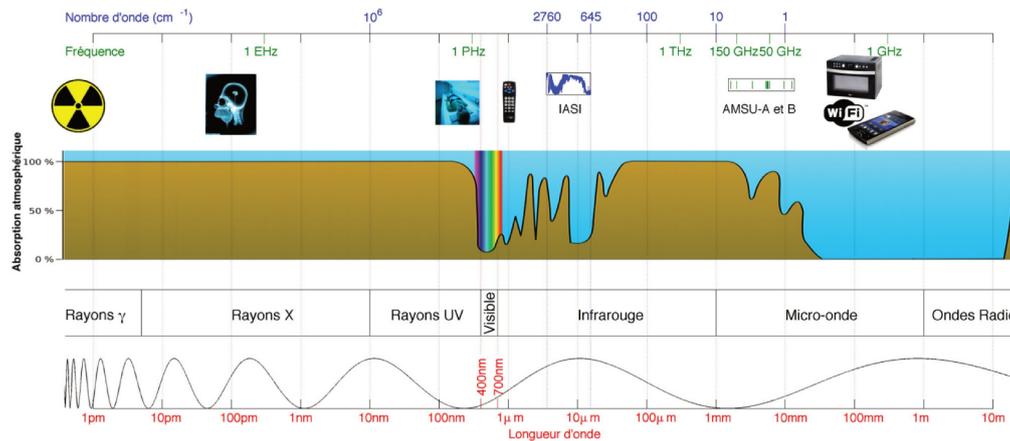


Figure II.3. Représentation du spectre électromagnétique et correspondance entre les différentes unités. Les utilisations les plus communes des différentes gammes de longueur d'onde sont représentées. Les instruments IASI (Infrared Atmospheric Sounding Interferometer) et AMSU (Advanced Microwave Sounding Unit) sont également représentés (Paul, 2013).

par le sondeur AMSU). Le rayonnement est absorbé par l'atmosphère pour une grande partie du spectre électromagnétique et le sondage satellite y est impossible à ces fréquences (aucun rayonnement ne serait transmis depuis l'atmosphère jusqu'au satellite). Les instruments de sondage atmosphérique IASI et AMSU mesurent le rayonnement à des longueurs d'onde pour lesquelles l'atmosphère est peu absorbante voire transparente.

2.1 Infrarouge

L'absorption dans l'infrarouge est majoritairement due aux changements d'états vibrationnel et rotationnel des différents absorbeurs atmosphériques. Les principaux absorbeurs sont la vapeur d'eau (H_2O), le dioxyde de carbone (CO_2) et l'ozone (O_3). De nombreux autres gaz, présents à l'état de trace dans l'atmosphère, y sont également actifs : le méthane (CH_4), le monoxyde de carbone (CO) ou le protoxyde d'azote (N_2O). La Figure II.4 montre la transmittance relative de l'atmosphère (en %) dans l'infrarouge et le micro-onde ainsi que les raies d'absorption des éléments actifs. Le dioxygène (O_2) et diazote (N_2), majoritaires dans l'atmosphère, sont presque inactifs dans l'infrarouge. Cela permet d'avoir accès à la mesure de l'absorption par les composants mineurs sus-cités. La plupart de ces gaz se trouvent dans la troposphère (partie basse de l'atmosphère). La mesure du rayonnement dans les bandes d'absorption de chaque constituant atmosphérique, permet de déterminer leur concentration et donc la composition chimique de l'atmosphère. Le sondage dans l'infrarouge permet la mesure de la température atmosphérique (*via* l'absorption de CO_2 , gaz dont la distribution verticale est connue et stable) et de la vapeur d'eau (*via* l'absorption de H_2O). La Table II.5 recense les différents absorbeurs atmosphériques. Cependant, l'absorption du rayonnement par les hydrométéores rend les nuages imperméables (opaques) aux rayonnements infrarouges. Il n'est pas possible de mesurer, depuis le satellite, le rayonnement infrarouge sous un nuage. Dans la suite, les expériences menées sur la réduction de

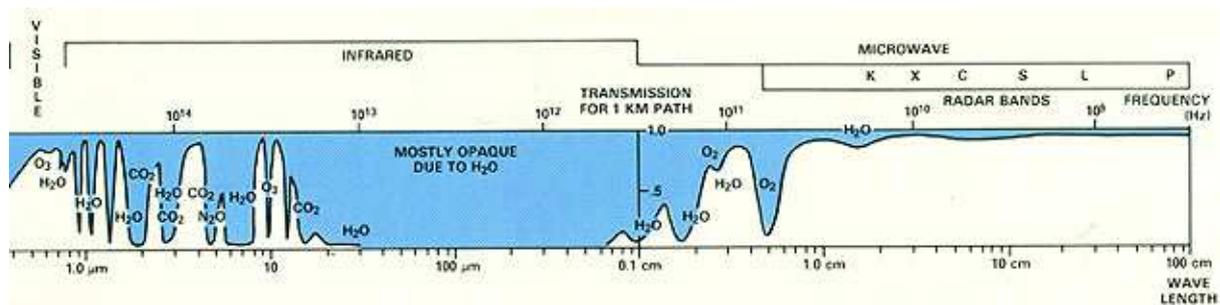


Figure II.4. La transmittance relative de l'atmosphère pour un rayonnement à différentes longueurs d'onde. Les zones blanches désignent les fenêtres atmosphériques, dans lesquelles le rayonnement n'interagit pas beaucoup avec les molécules d'air et n'est donc pas absorbé. (<http://geoinfo.amu.edu.pl>)

dimension des observations hyperspectrales seront limitées au cas simple en ciel clair.

On caractérise communément les raies d'absorption par leur nombre d'ondes en infrarouge. En spectroscopie, le nombre d'ondes ν est le nombre d'oscillations de l'onde par unité de longueur ($\nu = 1/\lambda$, où λ est la longueur d'onde). L'unité du nombre d'ondes est donc le m^{-1} . Les sondeurs infrarouges hyperspectraux mesurent précisément le rayonnement émis par l'atmosphère et la surface, à haute résolution spectrale. Disposant d'un grand nombre de canaux, ils échantillonnent l'atmosphère avec une fine résolution verticale, en présentant des fonctions de poids aux pics très étroits. On peut citer les sondeurs IASI sur Metop, AIRS sur Aqua, CrIS (Cross-track Infrared Sounder) sur Suomi NPP (National Polar-orbiting Partnership) et le futur instrument IASI-NG sur MetOpSG. La Table II.1 résume les caractéristiques de ces quatre sondeurs infrarouge hyperspectraux. L'instrument IASI sera détaillé à la section 5 et servira de cas concret à une étude sur les techniques de réduction de dimension des observations hyperspectrales dans l'infrarouge.

	AIRS	IASI	CrIS	IASI-NG
Plateforme Satellite	Aqua	MetOp-A MetOp-B	Suomi NPP	MetOp-SG
Lancement	2002	2006 2012	2011	2020
Résolution spatiale en km (nadir)	13.5	12	14	12
Techniques	Grating	FTS	FTS	FTS
Nombre de canaux	2211	8461	2378	16922
Couverture spectrale	17 bandes spectrales discontinues	LWIR, MWIR & SWIR (continu)	LWIR, MWIR & SWIR (discontinu)	LWIR, MWIR & SWIR (continu)
Résolution spectrale (m^{-1})	0.4-2.1 (selon la région)	0.25	0.625	0.125
Reference	Chachine et al. (2006)	Hilton et al. (2012)	Han et al. (2013)	Crevoisier et al. (2014)

TABLE II.1 – Caractéristiques instrumentales des sondeurs hyperspectraux infrarouge AIRS, IASI, CrIS, IASI-NG. FTS : "Fourier Transforms Sounder". LWIR : "Long Wave InfraRed", MWIR : "Middle Wave InfraRed", SWIR : "Short Wave InfraRed".

2.2 Micro-ondes

L'absorption dans le micro-onde est dominée par l'eau (liquide ou gaz) et le dioxygène (voir Figure II.4). On trouve la plus forte raie rotationnelle de l'eau à 556,9 GHz et sa raie d'absorption à 138,3 GHz. La mesure dans la bande d'absorption du dioxygène (O_2) permet d'avoir accès indirectement à des mesures de températures. Le dioxygène présente, en particulier, une bande d'absorption à 57 GHz et une autre raie d'absorption à 118,75 GHz. La Table II.4 résume les raies d'absorption H_2O et O_2 majoritairement sondées par les instruments dans le micro-onde. De nombreuses autres molécules présentent des raies d'absorption dans les micro-ondes (l'ozone O_3 , le dioxyde d'azote NO_2 , le monoxyde de chlore ClO , le dioxyde de soufre SO_2 et le protoxyde d'azote N_2O). Cependant, la faible concentration de ces constituants dans l'atmosphère et leur faible capacité absorbante, les rend difficilement détectables dans le micro-onde. On caractérise communément les raies d'absorption par leurs fréquences (GHz) dans le micro-onde.

Le sondage dans le micro-onde permet de traverser, au moins en partie, les nuages, car les hydrométéores diffusant à ces fréquences rendent le nuage non totalement opaque. Ainsi, des informations peuvent être obtenues sous les nuages. Plus important, la mesure d'absorption dans le nuage permet d'estimer des précipitations ou de mesurer la quantité totale d'eau dans la colonne atmosphérique sondée ce qui est très utile pour obtenir une information sur le profil de précipitation (voir chapitre II). L'ensemble des sondeurs micro-ondes échantillonnent en particulier la raie d'absorption de l' O_2 à 57 GHz pour restituer la température atmosphérique (gaz dont la distribution verticale est connue et stable) et celle de H_2O à 183 GHz pour restituer la vapeur d'eau. On peut citer l'instrument SSMIS (Special Sensor Microwave Imager/Sounder) à bord du DMSP (Defense Meteorological Satellite Program), le couple AMSU-A et MHS (Microwave humidity sounder) à bord de MetOp-A et MetOp-B, l'instrument ATMS (Advanced Technology Microwave Sounder) à bord du NPP (National Polar-orbiting Partnership) et le futur instrument MWS (Micro-Wave Sounder) qui sera à bord de la plateforme MetOp-SG. La Table II.2 résume les caractéristiques de ces quatre sondeurs micro-ondes.

De nos jours, les observations satellites dans le micro-onde sont les plus utilisées par les centres PNT. Ceci s'explique en partie par leur(s) faible(s) sensibilité(s) aux nuages et aux aérosols et donc la possibilité de les utiliser en permanence. cela s'explique également par leurs faibles sensibilités aux constituants mineurs de l'atmosphère (peu connus et donc difficilement quantifiables) qui pourraient corrompre la restitution de la température et/ou de l'humidité (la Section 3.2.b de ce chapitre traitera de ces contaminations).

2.3 Le sondage satellite hyperspectral

La Figure II.5 illustre le principe du sondage satellite pour la température dans l'infrarouge (Phulpin et al., 2003). Les raies les plus intenses du CO_2 se situent vers 675 cm^{-1} . La luminance mesurée à ce nombre d'ondes (température de brillance = 233 K) permet d'accéder à la température de la couche atmosphérique dont la fonction poids est centrée à 100 hPa (à droite), soit 225 K. De même, la luminance mesurée à 775 cm^{-1} (fenêtre atmosphérique) donne la température de la couche la plus basse, située vers 970 hPa, soit 285 K. Ainsi le sondage à différentes longueurs d'ondes,

	SSMIS	AMSU/MHS	ATMS	MWS
Plateforme Satellite	DMSP (F16)	MetOp-A MetOp-B	Suomi NPP	MetOp-SG
Lancement	2003	2006 2012	2011	2020
Fauchée (km)	1700	2340	2500	2300
Scan	Conique	Cross-Track	Cross-Track	Cross-Track
Résolution spatiale au nadir (km) T/q	37.5/12	48/- -/16	32/16	20/17
Nombre de canaux	24	20	22	24
Couverture spectrale (GHz) min/max	37/183.31	23.8/191.31	23.8/183.31	23.8/229

TABLE II.2 – Caractéristiques instrumentales des sondeurs micro-ondes SSMIS, AMSU/MHS, ATMS, MWS. La quatrième ligne indique la résolution spatiale au nadir pour les canaux de la raie d'absorption de l' O_2 à 57 GHz (T) et de H_2O à 183 GHz (q).

autour de la raie d'absorption permet de connaître la température atmosphérique à différentes altitudes. Plus la raie est échantillonnée finement avec des canaux présentant des fonctions de poids étroites, plus grande est la résolution verticale du profil de température ainsi sondé.

La Figure II.6 compare les fonctions de poids liées aux canaux du sondeur micro-onde AMSU (à droite) et à ceux du sondeur hyperspectral IASI (à gauche). Avec leurs milliers de canaux, les sondeurs hyperspectraux ont une multitude de fonctions de poids étroites, ce qui signifie que l'atmosphère est sondée sur un plus grand nombre

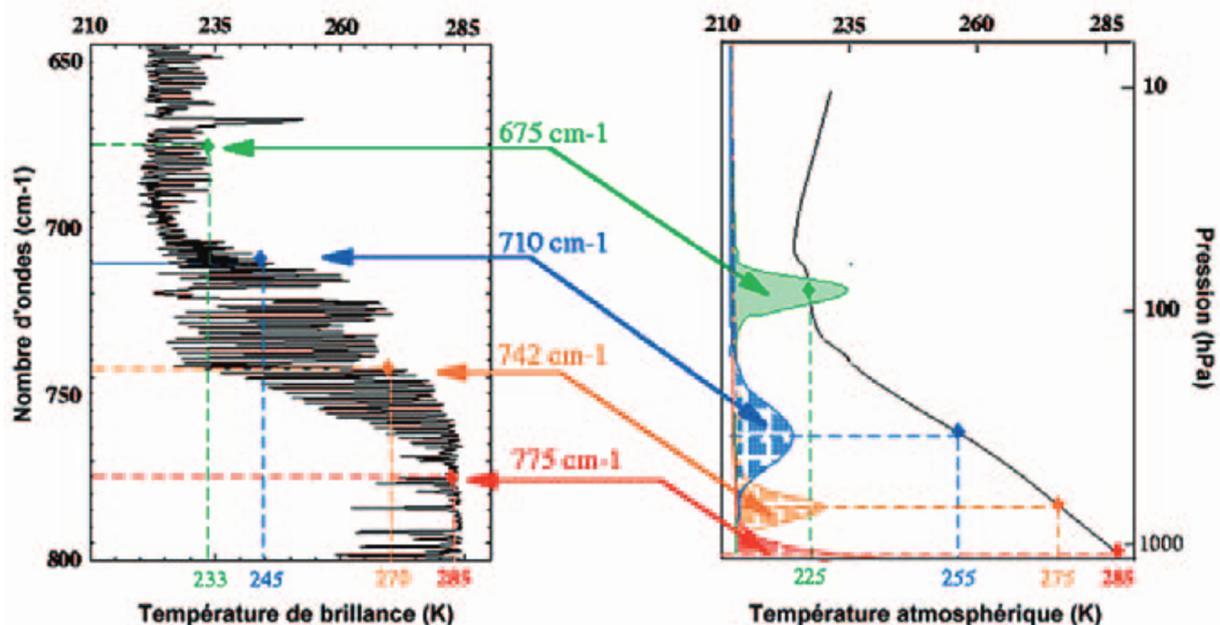


Figure II.5. Principe du sondage satellite. À gauche, spectre d'émission atmosphérique dans l'infrarouge entre 650 et 800 cm^{-1} . Ce spectre donne la température de brillance en fonction du nombre d'ondes (l'inverse de la longueur d'onde). À droite, profil de la température atmosphérique avec les fonctions de poids de quatre canaux (Phulpin et al., 2003).

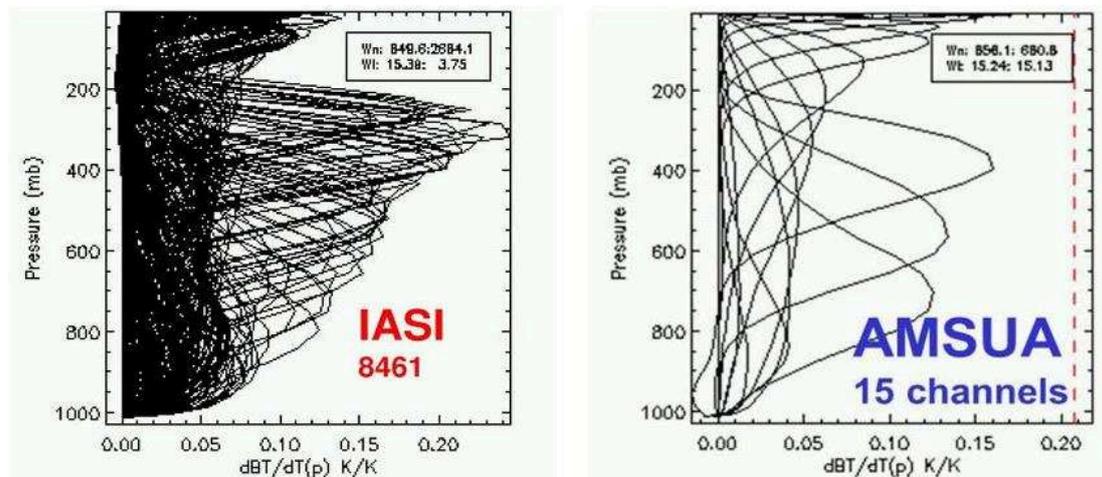


Figure II.6. Comparaison des fonctions de poids liées aux canaux du sondeur micro-onde AMSU (à droite) et à ceux du sondeur hyperspectral infrarouge IASI (à gauche).

de couches plus minces et permet une plus grande résolution verticale. En outre, la résolution spectrale des instruments hyperspectraux permet le sondage de gaz mineurs et d'aérosols dont les bandes d'absorption sont très fines. Alors que le signal de ces éléments serait écrasé dans le cas d'un sondage traditionnel, l'observation hyperspectrale discrimine leurs impacts dans le transfert radiatif. Si les sondeurs hyperspectraux opérationnels dans l'infrarouge ont déjà été présentés, il n'y a pas encore d'instrument hyperspectral opérationnel dans les micro-onde. Dans la Section II.4 de ce manuscrit, on s'intéressera au concept d'un instrument hyperspectral micro-ondes étudié pour l'European Space Agency (ESA).

Des stratégies de traitement des données encore plus sophistiquées et des modèles de transfert radiatif plus détaillés sont nécessaires pour convertir les informations hyperspectrales supplémentaires en produits dérivés utiles. Mais avant tout, l'augmentation du volume d'observations satellites génère des difficultés de stockage et de transmission des données et pour des raisons d'efficacité informatique, les algorithmes d'inversion du transfert radiatif ne sont pas capables de traiter autant d'informations et des méthodes de réduction de dimension pertinentes doivent être mises en place.

3 Méthodologie de réduction de la dimension des données hyperspectrales

Le but de la réduction de dimension est de fournir à la méthode d'inversion, qui cherche à inférer l'état de l'atmosphère à partir des spectres observés, l'information la plus complète et la plus pertinente à partir des observations originales, mais avec une dimension moindre. Deux stratégies ont été élaborées par la communauté des PNT

pour répondre à ce défi.

(1) Une façon de réduire la dimension originale est la compression qui utilise l'ensemble des observations et extrait ses caractéristiques les plus pertinentes. L'analyse en Composantes Principales (ACP ou PCA en anglais) classique et ses dérivés reste la technique de réduction de dimension la plus utilisée grâce à sa vitesse, sa simplicité et ses performances (Aires et al., 2002c ; Atkinson et al., 2010). Malgré leur manque de signification physique, les composantes ACP sont utilisées dans des schémas d'assimilation (Collard et McNally, 2009) et des codes de transfert radiatif dédiés ont été développés. Pour la restitution¹, environ 50 composantes sont généralement suffisantes dans le cas d'un instrument hyperspectral infrarouge comme IASI. En outre, les techniques de compression permettent de réduire le bruit des observations lié à l'instrument (bruit thermique au niveau du capteur) et à l'incertitude du transfert radiatif (hypothèses simplificatrices, incertitude sur la composition atmosphérique).

(2) Une autre façon de réduire la dimension est de ne sélectionner qu'une partie de l'observation (quelques canaux) comme entrées pour les algorithmes de restitution. La sélection de canaux est particulièrement utile lorsque l'on se concentre sur les gaz mineurs. On obtient, certes, des canaux réels, plus intuitifs à manipuler, mais cette approche induit aussi des pertes d'information et de redondance, et limite la capacité à débruiter les observations originales. Typiquement, les algorithmes d'inversion pour AIRS, CrIS, ou IASI (Collard, 2007) fonctionnent avec 200-300 canaux sélectionnés pour la température et l'humidité.

Dans la suite de ce chapitre, diverses techniques de réduction de dimension sont présentées et comparées. Outre les plus populaires, nous décrivons une méthode innovante, au croisement de l'ACP et de la sélection de canaux, qui permet d'extraire les caractéristiques de tout le spectre tout en conservant la nature et la caractéristique des canaux physiques : la représentation "Bottleneck Channels" (BC).

Les notations utilisées dans cette section sont données dans la Table II.3.

3.1 Compression de données : ACP et variantes

Dans ce qui suit x représente les profils atmosphériques de température T ou d'humidité q . Le vecteur d'observation est bruité par une erreur d'observation (à la fois liée à l'instrument mais aussi au modèle de Transfert radiatif, RTM en anglais) :

$$\begin{aligned}\tilde{y} &= y + \varepsilon, \\ &= \text{RTM}(x) + \varepsilon_{RTM} + \varepsilon_{inst},\end{aligned}\tag{II.12}$$

où ε_{RTM} est l'erreur de transfert radiatif et ε_{inst} l'erreur de l'instrument.

Dans cette section, les techniques de compression sont utilisées pour représenter le spectre des températures de brillance (TB) observé y de dimension m (nombre total de canaux) avec un nombre limité $n \leq m$ de "variables" (ou composantes). On veut réduire la dimension ($n \ll m$) tout en préservant au mieux l'information contenue dans y . L'objectif ici est de compresser le vecteur y indépendamment de son utilisation pour la restitution d'une variable géophysique x (voir Table II.7).

1. L'inversion du transfert radiatif par des méthodes physiques ou statistiques

y	Température de Brillance (TB) vecteur, dimension m
m	Nombre total de canaux dans y
\tilde{y}	Vecteur d'observation bruité $\tilde{y} = y + \varepsilon$, avec ε le bruit instrumental
\sim	Symbole pour une variable bruitée
$\hat{\cdot}$	Symbole pour l'estimation d'une variable
x	Variable géophysique à restituer. Ici vecteur de dimension $2 * 91$ pour les profils de T et q
T	Profil de température, dimension 91 (niveaux atmosphériques)
q	Profil de l'humidité spécifique, dimension 91 (niveaux atmosphériques)
H	Jacobien (matrice) $\frac{\partial y}{\partial x}$, dimension $m \times (2 * 91)$
H_T	Jacobien en T seulement $\frac{\partial y}{\partial T}$, dimension $m \times (91)$ (idem pour H_q)
\mathcal{B}	Base de donnée de e couples (x, y) variable géophysique et vecteur d'observation associé
η	Bottleneck Channels (BC)
e	Nombre d'échantillons dans \mathcal{B}
n	Nombre de composantes, BCs ou canaux sélectionnés
h	Vecteur de PCA ou NAPCA, dimension n
F	Matrice de filtre PCA ou NAPCA (voir texte)
s	Vecteur de sélection de canaux, dimension n
R	Matrice de covariance du bruit instrumental (251×251)
B	Matrice de covariance d'erreur (<i>a priori</i>)
A	Matrice de covariance d'erreur de restitution, estimée théoriquement contenu en information
f	Modèle d'inversion tel que $f(\tilde{y}) = x'$ avec $\tilde{y} \simeq RTM(x')$ (voir texte)
P	Matrice de projection de \tilde{y} vers les BCs $\tilde{\eta}$
Q	Matrice de projection des BCs $\tilde{\eta}$ vers y

TABLE II.3 – Notations pour la réduction de dimension des données hyperspectrales.

3.1.a Analyse en Composantes Principales (ACP ou PCA en anglais)

La PCA est une technique de compression largement utilisée pour compresser et réduire le bruit dans le vecteur d'observation \tilde{y} . Il s'agit d'une approche à la fois géométrique et statistique. En pratique, la compression est la recherche d'un sous-espace (dans l'espace décrit par toutes les observations aux différentes longueurs d'onde) sur lequel la projection de l'ensemble des observations est optimale. Dans la PCA, la recherche d'un sous-espace optimal de compression se fait d'abord par un changement de base (géométrie). En considérant l'espace vectoriel défini par l'ensemble des observations (nuage de points) et du produit scalaire, il est possible de trouver une Base OrthoNormée (BON) qui décrit cet espace. Le caractère orthogonal de la BON assure que chaque "direction" de variabilité dans la BON est indépendante des autres. La diagonalisation de la matrice de covariance du nuage de points (symétrique, positive et définie positive) permet d'obtenir cette BON comme étant la famille des vecteurs propres de la matrice de covariance (Théorème Spectral). Dès lors que cette BON existe, on peut trouver un sous-espace décrit par un nombre plus restreint des vecteurs de la BON pour lequel la variabilité (statistique) est optimale. Pour cela, on hiérarchise les vecteurs de la BON : plus un vecteur de la base est une direction de grande variabilité dans le nuage de point, mieux il décrit le nuage (en termes statistiques) plus son "numéro d'ordonnement" sera élevé.

La Figure II.7 schématise le changement de base opéré par PCA. Dans cet exemple simple, le nuage de points (observations) y est défini par deux variables (en abscisse et en ordonnée). Les directions données en vert sont les deux vecteurs de la BON décrivant l'espace des observations (ici un plan). La PCA conduit à décrire le nuage

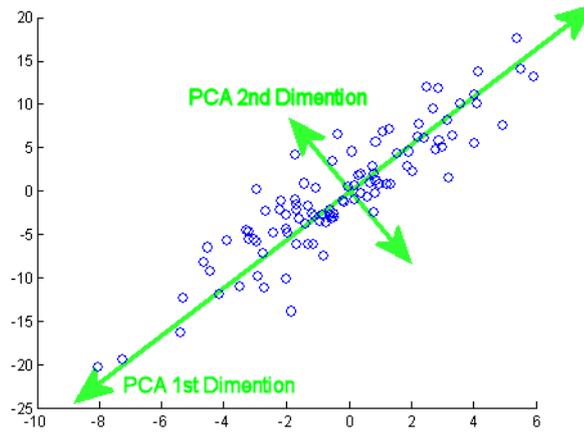


Figure II.7. Exemple de représentation graphique d'un nuage de points de dimension 2. Elle peut être décrite par l'abscisse et l'ordonnée d'origine ou bien les nouveaux axes verts. Finalement, la projection sur l'unique axe PCA-1st Dimension semble suffisante pour rendre compte de la variabilité du nuage.

de points dans les nouvelles coordonnées définies par les vecteurs en vert. Dans cette nouvelle projection, la variation sur le vecteur PCA-2nd Dimension ne semble être que du bruit. On peut dès lors caractériser tout le nuage de points seulement par sa projection sur PCA-1st Dimension. L'information est compressée.

La BON est d'abord calculée sur un ensemble de données synthétiques et sans bruit y de sorte que les composantes sont extraites de y sans la contamination du bruit. On utilise alors le changement de coordonnées (grâce à la matrice de projection) sur des observations réelles avec le bruit instrumental \tilde{y} .

Soit $\mathcal{D} = \{y^i ; i = 1, \dots, e\}$ une base de données de vecteurs de e observations y de dimension m . Soit C_y la matrice de covariance ($m \times m$) de la base de données \mathcal{D} . Le théorème spectrale permet de diagonaliser C_y comme : $C_y = V \cdot L \cdot V^t$ avec V la matrice dont les m colonnes sont égales aux vecteurs propres de C_y et ordonnées selon leur valeurs propres décroissantes (V est de dimension $m \times m$) et L la matrice diagonale avec les m valeurs propres ordonnées de façon décroissantes.

Le filtre $F = L^{-1/2} \cdot V^t$ (matrice $m \times m$) est utilisé pour projeter les vecteurs d'observations y sur la base orthonormale composée par les colonnes de F . Les vecteurs de la BON sont par la suite appelés PCs ("Principal Components" en anglais) ou composantes.

$$\begin{aligned}
 h &= F \cdot y = F_{1\star} \cdot y_1 + \dots + F_{m\star} \cdot y_m, \\
 y &= F^{-1} \cdot h \\
 &= F^t \cdot h = h_1 \cdot F_{\star 1} + \dots + h_m \cdot F_{\star m}.
 \end{aligned}
 \tag{II.13}$$

La matrice h représente les projections du nuage sur la BON. Les vecteurs $\{F_{i\star}; i = 1, \dots, m\}$, i.e., les lignes de F , sont appelés les filtres et les vecteurs propres normalisés $\{F_{\star i}; i = 1, \dots, m\}$ sont appelés fonctions de base de la PCA (composantes). Puisque ces vecteurs propres sont une base orthogonale pour représenter le vecteur y , les composantes de y dans cette nouvelle base ne sont pas corrélées.

La PCA est la technique de compression la plus utilisée en raison de son efficacité. Cependant, elle possède des limites bien connues en statistique comme pour la télédétection :

- La qualité de la base de données utilisée pour obtenir les PCs est très importante. La PCA est optimale près du centre du nuage de données \mathcal{D} (en pratique, les observations sont centrées avant d'être projetées). La PCA est optimale lorsque l'ensemble de données à une distribution Gaussienne, mais elle est moins adaptée aux structures plus complexes : distributions bimodales ou distributions asymétriques. Les PCs reproduiront les corrélations présentes dans l'ensemble de données mais il n'est pas possible de trouver des signaux dans les PCs qui ne soient pas présents dans \mathcal{D} . La PCA hérite seulement des corrélations linéaires de l'ensemble des observations.
- La PCA agit sur l'ensemble du spectre et tend à filtrer (ou à déléguer dans les PCs d'ordre supérieur moins fiables) les signaux fins dûs à des constituants atmosphériques mineurs. De sorte que l'information subtile (fines raies d'absorption dans le spectre) est négligée par rapport aux variabilités dominantes. Ceci peut être résolu en construisant une base de données dédiée spécifiquement aux fines absorptions.
- Un problème particulièrement important est celui du "mélange", plusieurs variabilités d'origines physiques différentes, sont mélangées dans les premiers PCs afin de maximiser la variance expliquée. La PCA a tendance à agir sur l'ensemble du vecteur d'observation \tilde{y} , ce qui implique un mélange des différentes composantes physiques. Cela signifie que les PCs pourraient ne pas être interprétables physiquement. En outre, les PCs mélangent des signaux verticaux, ce qui rend difficile la sélection dynamique, pour par exemple ne retenir que les informations relatives à l'atmosphère claire au-dessus du sommet du nuage. C'est la raison pour laquelle certains centres PNT préfèrent utiliser seulement des radiances spectrales directement plutôt que leur combinaison linéaire en PCs.

Diverses alternatives ont été étudiées pour limiter le problème de mélange. La PCA sur une bande limitée de l'observation a été testée de manière empirique. Cette localisation des filtres facilite l'interprétation physique des PCs extraits et devrait également améliorer l'inversion.

3.1.b ACP-Ajustée avec le bruit ("Noise-Adjusted PCA", NAPCA en anglais)

La PCA ne prend pas en compte les aspects d'inversion ni explicitement les sensibilités des canaux aux variables géophysiques, mais des informations auxiliaires peuvent être ajoutées dans le cadre général de la PCA. Le "Noise-Adjusted PCA" (NAPCA) est une extension de la PCA qui prend explicitement en compte le bruit ε et sa matrice de covariance C_ε dans le vecteur d'observation \tilde{y} (Blackwell, 2005). Les vecteurs de la BON ainsi obtenus optimisent leur rapport signal/bruit. Afin de pondérer les PCs par leur signal/bruit, on utilise le filtre $m \times m$ de "blanchiment" P tel que $P \cdot C_\varepsilon \cdot P = I_{m \times m}$ (qui correspond à une normalisation des variances du bruit dans le nuage de point). P est d'abord obtenu par $P = Q \cdot \Delta_\varepsilon^{-\frac{1}{2}}$, où Q est la matrice avec les vecteurs propres de C_ε , et Δ_ε la matrice de covariance du bruit diagonalisé ($Q^t \cdot C_\varepsilon \cdot Q = \Delta_\varepsilon$). Ensuite, une PCA est

appliquée aux variables $y' = P \cdot y$ en diagonalisant la matrice $Cov(y') = G^t \Delta G$. Le filtre global NAPCA combine les deux filtres : $F = P \cdot G$. L'Eq. (II.13) peut alors être utilisées avec cette nouvelle matrice de projection.

3.1.c Radiances reconstruites

On a vu que la PCA utilise seulement les premiers PCs pour représenter au mieux (et de façon compressée) l'information contenue dans les observations hyperspectrales. Le bruit d'observation est censé être présent uniquement selon les directions des PCs d'ordre supérieur. Ainsi, une observation "dé-bruitée" peut être reconstruite à partir de ces premiers PCs en utilisant la compression/décompression des Eqs. (II.13) (où h ne contiendrait plus qu'un certain nombre de colonnes). Ces "radiances reconstruites" peuvent être considérées comme des radiances pour lesquelles le bruit de l'instrument a été réduit et subir une sélection de canaux (Collard, 2007). Cependant, les radiances reconstruites peuvent présenter des corrélations inter-canaux plus élevées que dans la réalité en raison de la projection/ dé-projection linéaire sur le sous-espace optimal dans la PCA. Cela peut devenir un problème, en particulier pour la détection des gaz mineurs qui présentent des signatures (i.e. variabilité) spectrales subtiles.

3.2 La sélection de canaux

Il a été mentionné plus haut que la compression est souvent indépendante de l'algorithme de restitution. Cela signifie que le critère d'optimalité est général et basé uniquement sur spécificités des TBs . Par exemple, la corrélation, l'information mutuelle et la redondance des canaux peuvent être utilisées pour supprimer les canaux moins informatifs, et ce, sans aucun lien avec l'utilisation ultérieure qui en sera faite lors de la restitution. Cette approche est légitime, mais elle peut être sous-optimale si l'utilisateur cherche à restituer une variable particulière.

Une autre façon de réduire la dimension des observations hyperspectrales est de sélectionner puis d'utiliser uniquement une partie de l'observation : on se concentre alors sur les variables à restituer x (par exemple T et q) afin de sélectionner certains canaux. L'hypothèse principale est que :

- les spectres observés contiennent de nombreuses caractéristiques redondantes qui ne fournissent plus d'informations supplémentaires après que certains canaux aient été sélectionnés.
- ou certaines des caractéristiques ne sont pas pertinentes pour la restitution considérée.

Cette approche est particulièrement pertinente lorsqu'on s'intéresse à certains canaux spécifiques, comme dans le cas de la restitution de gaz mineurs. La sélection de canaux a l'avantage de conserver les sens physiques des canaux qu'elle manipule, ce qui facilite la restitution. Cependant, cette approche induit généralement une perte d'information et la redondance des canaux n'est pas pleinement exploitée pour réduire les bruits d'observations (incertitudes du modèle de transfert radiatif et bruit instrumental).

3.2.a Principes généraux

Dans cette partie, nous allons nous concentrer sur la sélection de canaux pour la restitution de T et q . Les notations utilisées sont rappelées dans la Table II.3. Afin

de sélectionner certains canaux $s = (s_1, \dots, s_n)$, il est nécessaire d'avoir une mesure de qualité (critère de sélection) ainsi qu'une stratégie de recherche pertinente.

3.2.a.1 Critère de sélection

Erreur d'inversion empirique et théorique - La mesure de qualité d'une sélection de canaux s est fortement liée à la capacité de restituer de façon fiable la variable géophysique recherchée, par inversion du transfert radiatif.

Lorsqu'un algorithme de restitution f est assez rapide, la matrice de covariance d'erreurs de restitution A peut aussi être estimée empiriquement, simplement en estimant la matrice de covariance des différences entre la variable recherchée x , connu par une autre source d'information (par exemple une mesure *in situ* par radiosonde) et celle restituée par la mesure de \tilde{y} .

$$A_{\text{emp}} = \text{cov}(x - f(\tilde{y})). \quad (\text{II.14})$$

L'inversion f utilise une information *a priori* x_o , des informations auxiliaires, et les observations satellites \tilde{y} . f peut être linéaire ou non-linéaire. L'avantage de cette estimation empirique est qu'aucune hypothèse probabiliste (Gaussienne ou autre) n'est faite sur les erreurs, ni de linéarité sur le transfert radiatif. Cependant, cette méthode peut être couteuse en temps car elle nécessite un nombre suffisant d'échantillons pour obtenir une estimation fiable de A_{emp} .

La matrice de covariance d'erreur d'inversion théorique A_{th} a été définie par Rodgers (2000) et peut être dérivée comme métrique pour la sélection de canaux. Cette métrique mesure l'apport informationnel d'une observation (ici liée à un ou plusieurs canaux) pour la restitution d'une variable géophysique particulière. Pour cela, certaines notions doivent être introduites. Comme les fonctions de poids représentent une sensibilité d'un canal, d'une certaine longueur d'onde, à l'absorption atmosphérique (voir la Section II.8 de ce chapitre), les Jacobiens représentent les sensibilités des canaux aux variables géophysiques comme la température ou l'humidité. Le Jacobien est la matrice H qui caractérise ces sensibilités et est constituée des dérivations premières des TB par rapport aux différentes variables géophysiques (en l'occurrence T et q). Le bruit instrumental (en K) est l'incertitude sur la mesure de TB par un canal dans une situation donnée. On note R la matrice de covariance des bruits instrumentaux de tous les canaux considérés. On parle d'ébauche (ou "first guess") lors de l'estimation première d'une variable géophysique avant la mesure instrumentale des TB correspondantes. Cette ébauche est une information *a priori*, caractérisée par sa matrice de covariance d'erreur B . La mesure de qualité pour la sélection doit inclure une information sur la capacité des canaux sélectionnés à restituer fidèlement une variable géophysique lors de l'inversion. Dans le cas d'observations hyperspectrales, les algorithmes de sélection de canaux sont basés sur (1) les Jacobiens des canaux ainsi que (2) leur bruit instrumental. Dans la pratique, les sélections utilisent les Jacobiens simulés par un RTM pour mesurer le contenu informationnel des canaux de l'instrument par rapport à une variable géophysique. On peut alors définir la matrice de covariance d'erreur d'inversion théorique A_{th} comme un compromis entre l'incertitude sur la variable x avant l'observation (qui est mesurée par B) et l'incertitude de la restitution par

les observations sachant le bruit instrumental R et le Jacobien simulé H :

$$A_{\text{th}} = (B^{-1} + H^T R^{-1} H)^{-1}. \quad (\text{II.15})$$

Cette estimation des erreurs de restitution est basée sur certaines hypothèses : l'inversion utilise des informations *a priori* (i.e. l'état de l'atmosphère connu ou supposé avant mesure), le transfert radiatif est linéarisé autour de cet *a priori* et les erreurs d'observation ainsi que celles de l'*a priori* sont des variables aléatoires Gaussiennes qui peuvent être caractérisées par les matrices de covariance R et B . Cette matrice A_{th} est très importante pour mesurer la qualité d'une sélection de canaux particulière. Le calcul de la covariance d'erreur de restitution *a posteriori* par apport d'une information d'observation (Rodgers, 2000) sera explicité et détaillé dans le contexte de l'estimateur Bayésien au chapitre V.

Minimiser les erreurs relatives de A - La première stratégie pour trouver une technique efficace de réduction de dimension est de "minimiser" la matrice de covariance des erreurs A . Par exemple, la trace ou le déterminant de A sont des critères efficaces qui peuvent être minimisés pour améliorer la qualité de la restitution. Il est également possible de minimiser directement la racine carrée des éléments diagonaux de A ($\sqrt{\text{diag}(A)}$) puisque ces éléments représentent l'écart-type des erreurs de restitution pour chaque variable à restituer dans x^2 .

L'inconvénient de la minimisation absolue des erreurs est que les erreurs dans x sont directement comparées dans le critère de qualité alors qu'elles peuvent être de natures très différentes. C'est un problème majeur lorsque les unités sont différentes (par exemple K pour la température et kg/kg pour l'humidité) ou lorsque les plages de variabilité sont très différentes (par exemple pour l'humidité dans la partie inférieure ou supérieure de l'atmosphère). L'utilisation d'un critère d'erreur absolue focalise la sélection et l'extraction des canaux vers les variables ayant des valeurs plus élevées et négligera celles ayant des valeurs plus basses.

Pour éviter de tels problèmes, une stratégie consiste à comparer les matrices A à d'autres matrices de covariance (à partir d'une autre sélection ou de l'information *a priori* B). Par exemple, lorsque deux sélections de canaux, s^1 et s^2 , doivent être comparées, s^1 est meilleure que s^2 quand $\text{Trace}\left(\frac{A(s^1)}{A(s^2)}\right) < 1$. Dans le cadre de cette recherche, nous définissons l'*amélioration* par rapport à l'*a priori* B par :

$$\text{Imp}(i) = \left(\frac{\sqrt{B(i,i)}}{\sqrt{A(i,i)}} - 1 \right) \times 100, \quad (\text{II.16})$$

pour $i = 1, \dots, m$, l'index des variables à restituer dans x . L'*amélioration* $\text{Imp}(i)=10\%$ signifie que l'incertitude associée à l'information *a priori* a diminué de 10% après la restitution de x_i par l'observation satellite. La restitution peut être pire que l'information *a priori*. Dans ce cas, l'*amélioration* peut être négative.

Pour obtenir un critère sur l'ensemble des variables à restituer, il est possible de faire la moyenne de l'*amélioration* :

$$C_{\text{Mean}}(s) = \text{Mean}(\text{Imp}(i); i = 1, \dots, m), \quad (\text{II.17})$$

2. C'est ce qui est fait lorsqu'on utilise le critère des moindres carrés, par exemple, pour les régressions statistiques.

ou de prendre l'amélioration maximale :

$$C_{\text{Max}}(s) = \text{Max}(\text{Imp}(i); i = 1, \dots, m), \quad (\text{II.18})$$

où C représente le critère de qualité. Les deux approches dans les Eqs. (II.17) et (II.18) ont leurs propres avantages et inconvénients et peuvent être choisies pour des applications particulières.

L'utilisation de l'information *a priori* comme référence pour la qualité des sélections de canaux n'est pas sans implication. Si cette information *a priori* n'est pas réaliste, elle peut fausser le critère de qualité utilisé pour la sélection. Une alternative peut être de remplacer le terme $\sqrt{\text{diag}(B(i, i))}$ dans l'Eq. (II.16) par l'écart-type des x_i ("Standard Deviation" en anglais) : $\text{STD}(x_i)$ dans la base de données \mathcal{D} . De cette façon, la comparaison est faite par rapport à la variabilité de x dans \mathcal{D} , et non par rapport à des hypothèses sur la qualité de l'*a priori* caractérisée par sa matrice de covariance B .

S'il est difficile de comparer les erreurs absolues, la comparaison d'erreurs relatives a aussi ses complexités. Par exemple, les erreurs *a priori* dans les couches atmosphériques supérieures peuvent être élevées et la sélection de canaux en utilisant une comparaison relative à l'*a priori* pourrait alors se concentrer sur l'amélioration de ces couches, même si aucune humidité n'est réellement présente. Avec les mesures de qualité basées sur une amélioration, il est donc nécessaire de se limiter aux couches atmosphériques pour lesquelles on dispose d'une information réelle lors de la sélection de canaux.

Chaque critère de qualité pour les sélections de canaux a ses avantages et ses inconvénients, et le choix d'une stratégie doit être fait en connaissance de causes et en fonction de l'application.

3.2.a.2 Stratégies de recherche

Comme tous les algorithmes d'optimisation, il faut une procédure de recherche pour explorer l'espace des solutions s afin de trouver la solution optimale \hat{s} . La procédure de recherche est plus ou moins exhaustive, déterministe ou stochastique, en fonction de la complexité du critère de qualité à optimiser et de la taille de l'espace à explorer. Il est possible de distinguer : (1) les "méthodes locales" qui utilisent un principe d'optimisation pour construire la solution de façon itérative, un canal à la fois ; (2) les "méthodes globales" qui visent à construire simultanément l'ensemble de la solution optimale. Les méthodes globales sont optimales. Par exemple, lors de la recherche du meilleur couple de canaux, les méthodes locales identifieront d'abord le meilleur canal, puis chercheront le deuxième meilleur canal supplémentaire. Les méthodes globales tenteront de trouver directement le meilleur couple de canaux. Le meilleur couple devrait être meilleur que le premier meilleur canal avec le deuxième meilleur canal.

Recherche exhaustive - Toutes les combinaisons de canaux possibles sont examinées et testées. C'est la seule technique qui assure que la solution optimale \hat{s} peut être trouvée. Cependant, les méthodes globales et exhaustives sont coûteuses en temps de calcul. Elle sont souvent trop longues pour être utilisées dans la pratique, et impossible pour les problèmes du monde réel avec des solutions s dans des espaces de dimension supérieure à 10 (le nombre de combinaisons possibles de n canaux parmi m ($n \leq m$) est $C_m^n = \frac{m!}{n!(m-n)!}$, par exemple $C_{10}^5 = 252$).

Recherche heuristique - Les recherches heuristiques explorent, au hasard, l'espace des candidats. Elles sont particulièrement intéressantes lorsque cet espace de recherche a une grande dimension. Les recherches heuristiques ne sont qu'en partie aléatoires, on peut utiliser les probabilités pour explorer plus précisément les parties intéressantes de l'espace des solutions. Ce processus est plus efficace pour trouver des solutions quasi-globales.

"Backward elimination" - Tous les canaux sont d'abord introduits dans la sélection. Ensuite à chaque étape de la procédure d'optimisation, un canal est supprimé en fonction de sa redondance avec les autres canaux de la sélection. Comme le processus commence avec tous les canaux, ce processus peut initialement prendre en compte toutes les interactions possibles entre les canaux (ce qui n'est pas le cas pour la "recherche locale"). L'inconvénient est que cette méthode peut être aussi très coûteuse en temps de calcul.

"Forward selection" - Il s'agit là aussi d'une stratégie de recherche incrémentale. À chaque itération, un nouveau canal est ajouté à la sélection de l'étape précédente, en identifiant le meilleur candidat relatif à la sélection déjà existante. Ce processus peut commencer par une sélection prédéfinie qui doit être améliorée, ou bien il peut commencer par une sélection vide. Le processus d'itération est arrêté lorsque le nombre de canaux désiré est atteint ou lorsqu'aucune amélioration n'est remarquée lors de l'ajout d'un nouveau canal. Cette technique est rapide, mais elle peut être piégée dans un minimum local en raison de sa nature itérative. Cette approche peut aussi passer à côté de combinaisons très informatives de canaux qui sont peu pertinents individuellement.

3.2.b Considérations sur les problèmes à plusieurs variables

3.2.b.1 Restitution de plusieurs variables

Dans le cas de la restitution de plusieurs variables (typiquement T et q), deux approches peuvent être envisagées. Tout d'abord, il est possible d'obtenir deux sélections de canaux indépendamment pour T et q . Par exemple, on a l'intention de trouver les 50 meilleurs canaux pour T d'abord et ensuite les 50 meilleurs canaux pour q . La restitution de la température et de l'humidité peut, par la suite, être effectuée indépendamment et les résultats de la sélection sont plus faciles à interpréter. Il peut arriver toutefois que les canaux contiennent à la fois des informations sur la température et l'humidité, c'est le cas des canaux sondant la troposphère (Collard, 2007). Ainsi des canaux d'humidité sont souvent sélectionnés pour restituer la température. Contraindre la sélection de canaux sur les bandes CO_2 pour T et H_2O pour q est un moyen simple d'éviter ce problème. Si on se ramène à l'Eq. (II.15), cela signifie que la matrice jacobienne H ne contient que la sensibilité par rapport à T ou q et que la matrice B est limitée aux covariances de T ou q (les corrélations croisées sont donc ignorées avec cette approche). C'est l'approche utilisée dans les centres PNT. Cependant, cette approche a aussi ses limites. En particulier, la solution finale est une sélection de canaux pour T et une sélection pour q indépendantes. En tant que problème d'optimisation, il est toujours préférable de construire l'ensemble de la solution simultanément plutôt

que pièce par pièce. Comme on vient de le voir, certains canaux peuvent être simultanément sensibles à T et q et des informations redondantes peuvent être présentes dans les deux sélections indépendantes de 50 canaux pour la température et pour l'humidité. Il est alors possible de bénéficier de ces canaux en dissociant leurs informations en utilisant la synergie de "un-mixing" (Aires et al., 2011b).

Une autre approche consiste à rechercher les 100 meilleures canaux pour la restitution simultanée de T et q . S'il est bien construit, le critère de qualité relatif (amélioration par rapport à *a priori* ou à l'écart type de la variable à restituer) devrait équilibrer le choix des observations pour restituer T et q . Cependant, il est plus difficile dans cette approche de contrôler comment la sélection de canaux se concentre sur l'une ou l'autre des variables. Par exemple, si le potentiel d'amélioration est plus élevé pour q que pour T , la sélection peut se focaliser sur les canaux d'humidité.

3.2.b.2 Variable externe et contamination

Un autre aspect à considérer est celui de l'impact des paramètres géophysiques (surface ou atmosphère) non restitué dans x . On parle alors de variables "externes" au schéma d'inversion. Par exemple, le profil de température affecte certains des canaux micro-ondes qui sont considérés pour la restitution de l'humidité. Même si nous ne sommes pas intéressés par la restitution de la température, les profils d'humidité peuvent en être affectés. Toute simulation du transfert radiatif pour des canaux sensibles à l'humidité, se base sur une hypothèse de profil de température. Toute erreur sur ce profil induit des erreurs dans les simulations et donc dans la restitution de l'humidité qui en résulte.

Dans la plupart des analyses de contenu en information et des sélections de canaux, les variables externes ne sont pas prises en compte. Cela peut conduire à surestimer artificiellement la qualité des restitutions des observations satellites. Collard (2007) a tenté de tenir compte des effets des gaz mineurs non inclus dans la simulation du transfert radiatif en augmentant artificiellement les erreurs d'observation pour les canaux ayant une sensibilité pour ces espèces. Une approche plus complète a été adoptée par Ventress et Dudhia (2014), qui ont utilisé la variabilité climatologique des espèces constitutives de l'atmosphère pour modéliser leurs "bruits" sur les TBs pendant le processus de sélection de canaux. Ils ont appliqué cette technique non seulement aux gaz mineurs, mais aussi à l'humidité.

Ce paragraphe décrit comment prendre en compte la contamination par des variables externes dans l'analyse du contenu de l'information et la sélection des canaux à venir.

Soit V la variable à restituer et v la variable externe. Il y a deux façons possibles de prendre en compte l'effet de contamination.

L'Eq. (II.15) peut être utilisée avec un Jacobien et des *a priori* qui incluent toutes les variables :

$$A_{V+v}^{-1} = B_{V+v}^{-1} + H_{V+v} \cdot R^{-1} \cdot H_{V+v}^t \quad (\text{II.19})$$

Cette approche peut être très coûteuse en pratique, surtout si plusieurs variables sont prises en compte en même temps (T , q , O_3 , nuages, *etc.*) parce que les inversions devront être réalisées sur des matrices beaucoup plus grandes et il faut estimer toutes ces quantités.

Il est possible de simplifier le problème en définissant explicitement les effets de la contamination. Mesurons par exemple l'impact de l'incertitude de l'humidité sur les canaux lors d'une sélection pour restituer T . La matrice de covariance d'erreur de v est donnée par B_v et l'impact de cette incertitude sur les canaux est une distribution Gaussienne avec pour matrice de covariance $H_v^t \cdot B_v \cdot H_v$. Ces incertitudes peuvent simplement être ajoutées aux erreurs d'observations ayant pour covariance R . Par conséquent, l'estimation du contenu de l'information pour V n'est pas :

$$A_V^{-1} = B_V^{-1} + H_V^t \cdot R^{-1} \cdot H_V, \quad (\text{II.20})$$

comme dans l'Eq. (II.15), mais avec contamination :

$$A_V^{-1} = B_V^{-1} + H_V^t \cdot (R + H_v \cdot B_v \cdot H_v^t)^{-1} \cdot H_V. \quad (\text{II.21})$$

Il est montré dans l'Annexe A de ce document que l'Eq. (II.19) est équivalente à l'Eq. (II.21) si B_{V+v} ne tient pas compte des corrélations entre V et v .

3.2.c Sélection par minimisation d'un critère de qualité

Dans cette partie, trois critères de sélection de canaux, couramment utilisés pour une sélection "forward", sont décrits. Un exemple d'implémentation opérationnelle de tels algorithmes y est aussi détaillé.

3.2.c.1 "Minimization of Relative Error" (MRE)

Avec le critère de MRE, l'accent est mis sur la simplicité. Puisque le but de la sélection des canaux est de réduire le nombre de canaux pour les restitutions tout en minimisant les erreurs de restitution, le critère d'optimalité est simplement basé sur les erreurs de restitution estimées. Plus spécifiquement, le critère de qualité est donné par la racine carrée des éléments diagonaux de la matrice théorique A (voir Section 3.2.a, l'Eq. II.15).

$$Crit(s|s_{ref}) = \frac{1}{N} \sum_{i=1}^N \frac{\sqrt{A(i,i)} - \sqrt{A_{ref}(i,i)}}{\sqrt{A_{ref}(i,i)}}, \quad (\text{II.22})$$

où p est le nombre de variables dans x , A et A_{ref} sont les matrices de covariance d'erreur de la sélection de canaux s et celle de la référence s_{ref} . Ce critère est une estimation de la réduction moyenne de la covariance d'erreur de restitution (sur toute les couches atmosphériques) relative à la sélection de canal s par rapport à une sélection de référence s_{ref} .

Le choix de la sélection de canaux de référence est très flexible : (1) il peut être une sélection de canaux intermédiaire dans le processus d'optimisation, (2) il peut être lié aux canaux d'un instrument actuel ou futur que nous voulons optimiser, (3) dans le cas où il n'y a pas de sélection préexistante, la matrice de covariance d'erreur $A_{ref} = B$ ne décrit que les statistiques d'incertitude *a priori* sur les profils avant l'observation satellite. Cette matrice est, dans ce cas, calculée grâce aux climatologies atmosphériques.

Entropie au sens de Shannon

Claude Shannon (1916-2001) a introduit le concept d'entropie en théorie de l'information. L'entropie d'une variable statistique correspond à la quantité d'information qu'il faut connaître pour pouvoir la déterminer de façon certaine. Une variable déterministe aura une entropie nulle, car il suffit de connaître la valeur que prend la variable pour la déterminer. Une variable aléatoire aura une entropie supérieure car connaître sa moyenne ne suffit pas à la déterminer complètement.

Soit X une variable aléatoire discrète pouvant prendre les valeurs (x_1, x_2, \dots, x_n) avec respectivement une probabilité (p_1, p_2, \dots, p_n) . L'entropie de la variable X au sens de Shannon est :

$$E(X) = \sum_{i=1}^n p_i \cdot \log\left(\frac{1}{p_i}\right) \quad (\text{II.23})$$

Cette entropie est maximale si chaque état x_i a la même probabilité $p_i = \frac{1}{n}$.

L'entropie est donc une mesure dans l'espace des probabilités, on peut alors comparer ou additionner les entropies de deux variables différentes.

3.2.c.2 "Degree of Freedom of the Signal" (DFS) et "Entropy Reduction" (ER)

Dans ces approches de sélection de canaux, une méthode itérative proposée par Rodgers (1996a) est utilisée pour choisir les canaux en fonction de leur contenu en information. Le contenu en information de chaque canal est déterminé par sa réduction d'entropie (ER) ou ses degrés de liberté du signal (DFS en anglais). Ces deux mesures sont issues de la théorie de l'information.

L'entropie est une fonction de l'ensemble des probabilités des états possibles et est maximale si tous les états ont une probabilité égale. La réduction d'entropie est alors une mesure de détermination au sens statistique (variance expliquée versus variance totale), plus l'entropie de l'état d'une variable géophysique est réduite par l'utilisation d'un canal, plus celui-ci est pertinent pour connaître l'état le plus probable. Plus l'entropie est réduite, plus l'état est connu. La réduction d'entropie expliquée de B vers A est calculée par :

$$\text{ER} = \frac{1}{2} \log_2 \left(\frac{B}{A} \right) \quad (\text{II.24})$$

avec les matrices A et B issues de l'Eq. (II.15).

Le DFS mesure le nombre d'informations indépendantes par rapport au bruit dans un signal (Rabier et al., 2002). La DFS est une mesure de détermination comme l'ER. Si l'observation d'un canal permet de connaître toutes les directions de variation indépendantes de l'état d'une variable géophysique, alors la DFS est nulle. Si l'ajout d'un canal n'ajoute pas d'information, la DFS n'est pas modifiée. La DFS peut être calculée mathématiquement comme :

$$\text{DFS} = \text{Tr}(I - AB^{-1}) \quad (\text{II.25})$$

Une méthode itérative est utilisée pour sélectionner les canaux les plus informatifs, l'un après l'autre. Afin d'accélérer les calculs, un algorithme a été développé en sup-

posant que les erreurs d'observation entre canaux ne sont pas corrélées (Rodgers, 1996a). À chaque étape, le canal dont l'augmentation en information est le plus important par rapport à celui des canaux précédents est sélectionné, compte tenu du contenu informationnel des canaux précédemment sélectionnés.

Les Eqs. (II.24) et (II.25) nécessitent une mise à jour de la matrice de covariance d'erreur A de l'Eq. (II.15) à chaque itération de la sélection qui commence sans canaux sélectionnés et choisit séquentiellement le canal avec l'augmentation du contenu en information le plus élevé. Tout d'abord, le Jacobien H (qui est constant pendant le processus itératif) est normalisé par la matrice de covariance d'erreur R :

$$H' = R^{-1/2} H \quad (\text{II.26})$$

D'après Rodgers (1996a), la matrice de covariance d'erreur mise à jour à chaque étape d'itération i peut être calculée à partir de l'étape précédente ($i - 1$) comme :

$$A_i = A_{i-1} - \frac{(A_{i-1} h')(A_{i-1} h')^T}{1 + (A_{i-1} h')^T h'} \quad (\text{II.27})$$

où h' est le vecteur colonne égal à la ligne H' pour le canal considéré.

Ensuite, le changement de DFS ou de ER entre deux itérations peut être écrit comme :

$$\delta ER = \frac{1}{2} \log_2(1 + h'^T A_{i-1} h') \quad (\text{II.28})$$

$$\delta DFS = \frac{h'^T A_{i-1} h'}{1 + h'^T A_{i-1} h'}. \quad (\text{II.29})$$

La procédure itérative est initialisée avec $A_0 = B$, la matrice de covariance d'erreur *a priori* représentant l'incertitude sur les profils avant l'observation.

Les DFS et ER peuvent être calculés indépendamment pour toutes les situations atmosphériques de l'ensemble de données. La méthode la plus simple pour évaluer une sélection de canal sur tous les profils atmosphériques est basée sur le rang moyen de sélection de chaque canal calculé sur tous les profils. D'autres méthodes ont été envisagées avec le DFS ou l'ER. Lipton (2003) a proposé une méthode basée sur le calcul d'un DFS normalisé qui est additionné pour chaque canal sur tous les profils. Cette méthode a été utilisée par Collard (2007) pour la sélection de canaux opérationnelle dans l'assimilation des modèles des centre PNT.

3.3 Les "Bottleneck Channels" (BC)

Dans cette nouvelle approche de réduction de dimension des observations hyperspectrales, l'objectif est de trouver s , c'est-à-dire un nombre limité n de canaux ou "pseudo-canaux" représentatifs de l'ensemble du spectre complet y (dimension m) tel que $n \leq m$. Cette stratégie s'inspire de la méthode Optimal Spectral Sampling (OSS) utilisée pour accélérer les calculs de transfert radiatif (Moncet et al., 2008, 2010) : des radiances monochromatiques calculées à des fréquences sélectionnées (nodes) sont utilisées pour modéliser un spectre complet. Cette approche peut réduire d'un facteur 10 ou plus le temps de calcul d'un modèle de transfert radiatif.

Dans (Aires et al., 2015 ; Pellet et al., 2015), la stratégie des Bottleneck Channels (BC) a été introduite puis utilisée non pas pour trouver des fréquences monochromatiques mais plutôt des canaux instrumentaux représentatifs du reste du spectre,

appelés BCs. Á notre connaissance, cette approche n'avait pas encore été envisagée pour ce type d'application.

3.3.a BCs "stricts"

Pour obtenir une sélection de BCs s , une régression linéaire est calculée de telle sorte que le vecteur d'observation complet y (sans bruit) puisse être reconstruit à partir des canaux sélectionnés (bruité) :

$$\begin{aligned}\bar{y} &= Q \cdot \tilde{\eta}_s, \text{ avec} \\ \tilde{\eta}_s &= \tilde{y}_s,\end{aligned}\tag{II.30}$$

où les BCs $\tilde{\eta}_s$ sont une sélection s du spectre bruité \tilde{y} . La matrice Q peut être obtenue en minimisant simplement les erreurs des moindres carrés : $\sum_{e \in \mathcal{D}} \|y^e - \bar{y}^e\|^2$.

Nous adoptons ici une procédure itérative pour trouver cette sélection de canaux mais n'importe laquelle des stratégies de recherche vues dans la section 3.2.a.2 pourrait être utilisée à la place. La sélection de canal est d'abord initiée par le vecteur vide : $s^0 = \emptyset$. Á chaque itération i ($1 \leq i \leq m$), tous les canaux j encore disponibles sont ajoutés individuellement à la sélection précédente afin de créer une sélection candidate : $s_j^i = \{s^{i-1}; j\}$. Pour chaque s_j^i , une régression linéaire Q est obtenue à partir de tous les échantillons \mathcal{D} , le canal j' qui minimise l'erreur :

$$\begin{aligned}\sum_{e \in \mathcal{D}} \|y^e - Q \cdot \tilde{y}_{s_j^i}^e\|^2 \text{ ou ,} \\ \text{Max}_{e \in \mathcal{D}} (y^e - Q \cdot \tilde{y}_{s_j^i}^e),\end{aligned}\tag{II.31}$$

est sélectionné $s^i = (s^{i-1}; j')$. D'autres critères d'optimalité que ceux de Eq. (II.31) pourraient être utilisés à la place, en fonction des objectifs. On utilise dans ce travail de recherche la distance maximal (en haut dans l'Eq. (II.31)). Á la fin de ce processus, une sélection de canaux s^m est obtenue avec ($n \leq m$) et où les m canaux sont hiérarchisés par ordre d'importance. Les BCs $\tilde{\eta}_s$ sont définies comme les observations bruitées de la sélection résultante s (Figure II.8).

Il faut noter que les BCs ne tiennent pas compte des paramètres géophysiques qui doivent être restitués par la suite ; l'objectif de la sélection est ici limité à la compression du vecteur observé \tilde{y} et à une reconstruction du vecteur complet y . Il s'agit d'une sélection de canaux puisque le but est d'obtenir m BCs parmi les canaux d'observation. La recherche des BC optimales prend en compte le bruit instrumental puisque les observations réelles \tilde{y} sont utilisées comme candidats dans le processus : les canaux avec un bruit plus élevé seront moins enclins à être choisis que les canaux avec un bruit plus faible. Cependant, comme toute technique de sélection de canal, la redondance du spectre complet n'est pas pleinement exploitée. Cela signifie que les capacités de dé-bruitage de cette approche sont théoriquement moins efficaces que la PCA et la NAPCA.

Des modèles statistiques plus complexes pourraient être utilisés à la place de la régression linéaire Q . Par exemple, un réseau neuronal serait plus adéquat pour bénéficier des interactions complexes et potentiellement non linéaires entre les canaux afin de mieux représenter les liens physiques dans le vecteur y . Cependant, les modèles plus sophistiqués sont plus exigeants en termes de calculs et, dans la pratique, il serait difficile de les utiliser dans un algorithme BC-strict.

3.3.b BCs issus d'une projection linéaire

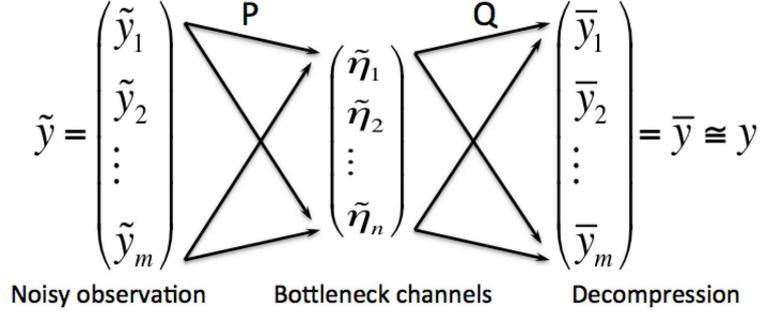


Figure II.8. Représentation schématique des BCs stricts (a) et projetés (b). De gauche à droite, les observations bruitées \tilde{y} , les BCs $\tilde{\eta} \approx y_s$, et le spectre reconstruit \bar{y} . Les matrices P et Q sont les deux régressions linéaires entre ces trois vecteurs.

Un autre algorithme plus sophistiqué basé sur les BCs est possible. Il bénéficie cette fois de l'utilisation complète du spectre observé, par opposition à une sélection de sous-ensembles de canaux. Dans cette approche, le BC $\tilde{\eta}$ est un ensemble de "pseudo-canaux" résultant de l'ensemble du vecteur observé et imitant une sélection de canaux $s : \tilde{\eta}_s = P \cdot \tilde{y}$, où P est une régression linéaire entre \tilde{y} et y_s (voir Figure II.8). De cette façon, $\tilde{\eta}_s$ est presque une sélection de canal, mais elle résulte d'une combinaison linéaire du vecteur d'observation complet, comme la PCA. Les BCs projetés permettent de mieux exploiter la redondance dans \tilde{y} et de mieux dé-bruiter les observations que les BCs stricts.

La Figure II.9 représente la corrélation entre les observations \tilde{y} (dans l'axe horizontal) et la sélection des $m = 100$ premiers BCs dans le contexte d'un sondeur avec 251 canaux dans micro-onde. Chaque BC est corrélé à de nombreux canaux. Même si chaque BC essaie de reproduire un canal particulier, il est en fait lié à de nombreux canaux car les $m = 251$ canaux d'origine sont fortement corrélés puisque sensibles aux mêmes variables atmosphériques. Ces corrélations sont utilisées de façon à ce que les BCs $\tilde{\eta}_s$ soient plus proches des observations sans bruit y_s que les observations bruitées réelles \tilde{y}_s .

À chaque itération j pour trouver un nouveau BC, pour chaque sélection de candidats $s_j^i = s^{i-1}$; j , une régression linéaire est d'abord obtenue pour relier $\tilde{y}_{s_j^i}$ et \tilde{y} . Ensuite, une régression linéaire est effectuée pour estimer la matrice Q reliant les BCs au spectre dé-bruité y (de la même manière que les BC-stricts de la section (3.3.a)). Le canal optimal j' est choisi pour minimiser, par exemple, l'erreur Maximum de restitution :

$$\text{Max}_{e \in \mathcal{D}} (y^e - Q \cdot P \cdot \tilde{y}^e) \quad (\text{II.32})$$

ou le critère des moindres carrés. La première régression linéaire $\tilde{\eta}_s = P \cdot \tilde{y}$ peut être faite une fois pour toutes avant les itérations pour trouver la meilleure sélection de BCs.

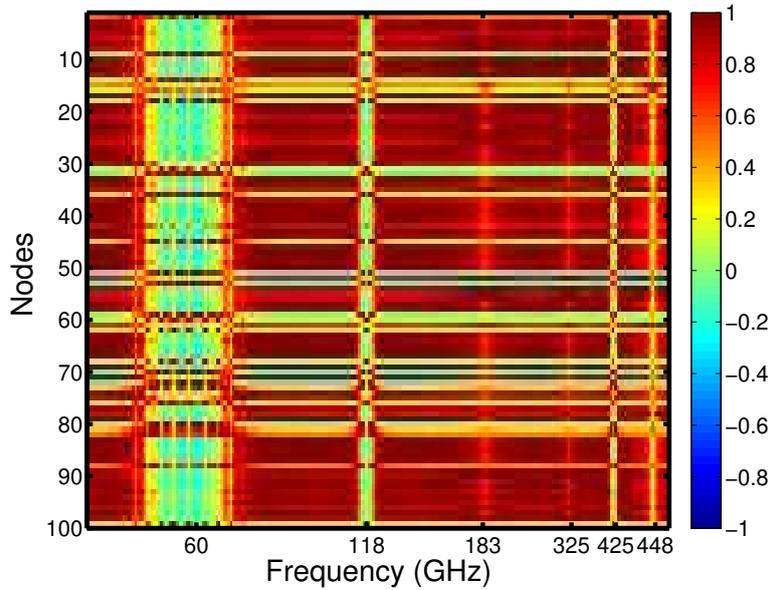


Figure II.9. Exemple de corrélation entre les BC-projetés (version linéaire) sur l'axe des y et les canaux de température de Brillance (TB) (axe des x) pour un instrument hyperspectral (251 canaux) micro-onde.

À la fin de ce processus, une matrice P de dimension $n \times m$ est obtenue afin de projeter le vecteur observé \tilde{y} dans l'espace BCs, et une matrice Q de dimension $m \times n$ pour projeter les BCs résultants dans l'espace d'observation (voir Figure II.8). Ces deux matrices sont équivalentes aux matrices de projection de la PCA, de sorte que cette approche "BC-projeté" est très similaire à la PCA, sauf que les BC sont ici contraints de représenter une sélection de canaux, ce qui leur donne une signification physique. En outre, dans la PCA, P et Q sont inverses l'un par rapport à l'autre, ce qui n'est pas le cas pour les BCs où une contrainte explicite de suppression du bruit (de \tilde{y} vers y dans la Figure II.8) est utilisée. Les BC-projetés sont une technique de compression qui effectue une extraction de caractéristiques sur l'ensemble du vecteur \tilde{y} . C'est aussi une sélection de canaux puisque les BCs $\tilde{\eta}_s$ tentent de reproduire une sélection de canaux s . La version BC-projeté est donc une approche intermédiaire entre la méthode de compression et la méthode de sélection des canaux.

Une caractéristique intéressante des BC (stricts et projetés) est qu'il est très facile de contraindre la sélection de canal s pour inclure certains canaux choisis *a priori* par l'utilisateur. Par exemple, l'utilisateur peut imposer à la sélection certains canaux (*a priori*), connus pour être sensibles aux gaz mineurs. L'utilisation de la distance du Max (voir Section précédente et l'Eq. (II.31)) devrait imposer, à l'étape de sélection de canaux, une reconstruction aussi uniforme que possible de tous les canaux m , y compris ceux contenant une variabilité très subtile comme les gaz traces. Cependant, les sélectionner *a priori* peut être un moyen plus direct de contraindre la sélection. Pour les BC-projetés, cela résulterait en une matrice P avec certaines de ses lignes incluant seulement un "1" dans la colonne correspondant au canal désiré *a priori*.

3.3.c BCs issus d'une projection non-linéaire

Chaque projection linéaire de la Figure II.8 peut être remplacée par un modèle non linéaire. Puisque la deuxième régression (des BCs vers le vecteur d'observation reconstruit) doit être faite plusieurs fois dans la recherche de la meilleure sélection de canaux s , il peut devenir extrêmement coûteux d'utiliser une régression non linéaire. Cependant, comme indiqué précédemment, la première régression (c'est-à-dire la matrice P) doit être faite une fois pour toutes afin qu'elle puisse être définitivement remplacée par une régression non linéaire comme, par exemple, un réseau neuronal : NN_1 . Une fois que NN_1 et la sélection de canal optimale s sont obtenus, il est également possible de remplacer la régression linéaire Q par un modèle non linéaire NN_2 .

Au terme de cette section, les différentes méthodes de réductions de dimension ont été présentées, les avantages et les inconvénients des méthodes classiques de compression ou de sélection de canaux ont été détaillés. Leurs utilisations varient en fonction de l'objectif fixé préalablement mais toutes ces méthodes possèdent une grande flexibilité. Une nouvelle méthode de réduction de dimension a été introduite à la croisée des méthodes classiques : les BCs. Ces méthodes seront comparées dans le contexte d'un concept d'instrument hyperspectral micro-onde (Section 4) et de l'instrument IASI (Section 5).

4 Réduction de dimension pour un instrument hyperspectral micro-onde

Pour illustrer et tester les diverses approches présentées à la Section 3, des expériences sont réalisées pour un concept d'instrument hyperspectral dans les micro-ondes. Les résultats qui suivent sont issus de Pellet et al. (2015). L'instrument hyperspectral possède des canaux dans les raies d'absorption O_2 et H_2O entre 1 et 500 GHz, pour la restitution des profils de température atmosphérique (T) et d'humidité (q). L'étude est faite dans des conditions de ciel clair et au nadir au-dessus de l'océan. L'accent est mis sur la restitution combinée de la température et l'humidité. Dans cette section, nous analysons la possibilité de réduire la taille de ces observations en les comprimant, en effectuant une sélection de canal ou en utilisant une représentation avec les BCs.

4.1 Matériels

Un ensemble de données sur les profils atmosphériques est nécessaire pour étudier les techniques de réduction de dimension dans diverses conditions atmosphériques. L'ensemble de données doit décrire au mieux la variabilité statistique et les cas plus extrêmes (Paul et Aires, 2014) présents dans la nature. Un ensemble de données atmosphériques et des informations *a priori* de l'ECMWF ont été choisis ici. Les expériences se limitent à des cas de ciel clair au-dessus de l'océan, au nadir. Un code de transfert radiatif est sélectionné (Atmospheric Radiative Transfer Simulator, ARTS) et des simulations sont effectuées sur la base de données des profils.

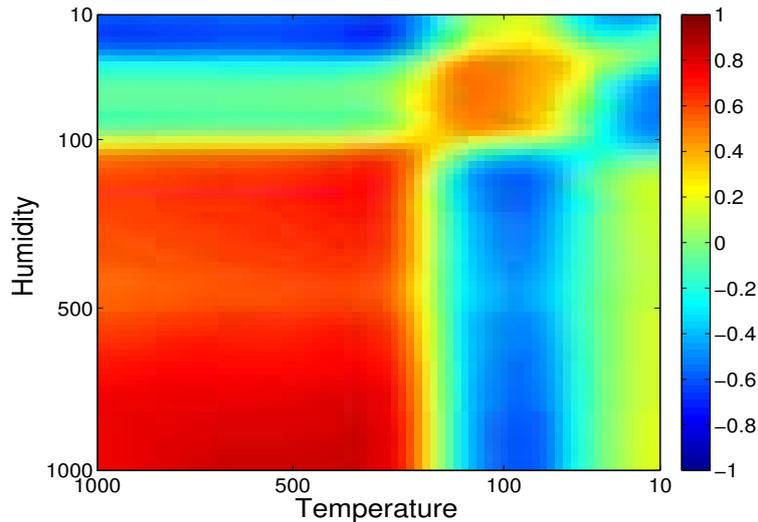


Figure II.10. *Corrélation croisée entre le profil de température et celui de l'humidité le long de l'atmosphère. Les axes vertical et horizontal représentent la pression atmosphérique des niveaux verticaux (hPa).*

4.1.a Base de données des profils atmosphériques

Les profils atmosphériques ont été obtenus à partir du système de prévision de l'ECMWF, avec une résolution horizontale d'environ 25 km sur 91 niveaux de pression. Trois bases de données en ciel clair de 5000 profils sont disponibles, une pour l'échantillonnage de la température, une pour l'humidité et une autre pour l'ozone. Les données n'ont été sélectionnées que lorsque les matrices de covariance d'erreur d'*a priori* B étaient disponibles. Cela a réduit le nombre de profils disponibles de 15 000 à 1989 profils. Chaque situation atmosphérique comprend des propriétés de surface telles que la température et la pression. Seules les situations océaniques sont conservées ici, et l'émissivité de l'océan est fixée à 0.6. Ces profils représentent une large gamme de profils de température et d'humidité avec une grande proportion d'atmosphères sèches et froides.

La Figure II.10 montre la corrélation entre les profils de température et d'humidité et illustre le lien étroit entre ces deux variables. Le profil d'humidité dans la tropopause est fortement couplé aux températures troposphériques, et ce jusqu'à la surface. Ainsi les canaux sensibles à la température de surface peuvent aider à restituer l'humidité dans la tropopause, mais pour cela, des informations sur la température sont nécessaires. Cela implique également qu'une procédure de sélection de canaux pour la restitution simultanée de T et q peut bénéficier de ces liens à condition que les effets de contamination soient connus et qu'une stratégie de dé-mélange soit mise en place afin de séparer les deux informations (T et q).

4.1.b Matrice de covariance d'erreur *a priori*

Les méthodes de restitution utilisent souvent une information *a priori* sur l'état de l'atmosphère en plus des observations satellites. Cette information doit être prise en compte dans les techniques de sélection des canaux. Pour chaque situation atmosphérique de 1989, la matrice de covariance *a priori* B représente les covariances

Spectral band	Type	Nb chan.	Res. (MHz)	Inst. Noise (K)	RT Noise (K)
Around 60 GHz	O ₂ - Temperature	94	100	0.40	0.32
Around 118 GHz	O ₂ - Temperature	51	200	0.42	0.92
Around 183 GHz	H ₂ O - Humidity	51	400	0.40	1.07
Around 325 GHz	H ₂ O - Humidity	21	1000	0.73	1.30
Around 425 GHz	O ₂ - Temperature	17	1000	0.93	1.47
Around 448 GHz	H ₂ O - Humidity	17	1000	0.93	1.47

TABLE II.4 – Description des six bandes d’absorption associées aux 251 canaux simulés. Pour chaque bande, la résolution spectrale est indiquée avec le nombre de canaux et le bruit instrumental choisi.

d’erreur de l’information *a priori* pour la température et la vapeur d’eau. Afin d’être au plus proche des conditions opérationnelles des centres PNT, les matrices de covariance d’erreur B ont été dérivées du système d’assimilation des données d’ensemble de l’ECMWF. Elles résultent de la combinaison entre des écarts-types d’erreur *a priori* et d’une matrice de corrélation des erreurs. (1) Les écarts-types des erreurs *a priori* sont ceux utilisés à l’ECMWF et correspondent aux dispersions de l’ensemble des prévisions. Ils varient temporellement et dans l’espace, en raison de la variabilité géographique. Ces dispersions sont de bonnes approximations et facilitent l’assimilation des observations satellites dans le modèle de prévision. (2) Les matrices de corrélation verticale correspondent à des moyennes climatologiques calculées sur deux saisons. Leurs principales variations résultent des contrastes géographiques et de l’orographie, les variations saisonnières et diurnes étant faibles par rapport aux variations géographiques. Ces matrices sont séparées pour chaque variable et sont disponibles pour la température et l’humidité. Pour chacune des 1989 situations atmosphériques en ciel clair et au-dessus de l’océan, la matrice de covariance d’erreur *a priori* B $(2 \times 91) \times (2 \times 91)$ représente les erreurs de T et q .

4.1.c Le modèle de transfert radiatif

Le code de transfert radiatif ARTS est utilisé pour simuler les températures de brillance observables dans les micro-ondes par un instrument hyperspectrale et ce, pour chaque situation atmosphérique. Les calculs de transfert radiatif sont effectués pour la visée d’un instrument au nadir. Les bandes de fréquences ont été choisies pour analyser la gamme de 1 à 500 GHz. Ils comprennent des canaux situés dans les lignes d’absorption (60, 118, 183, 325 et 420 GHz) pour permettre la restitution des profils verticaux de la température et de l’humidité atmosphérique (Aires et al., 2015). Les caractéristiques des bandes choisies dans cette étude sont présentées dans la Table II.4. La résolution spectrale choisie est comparable aux résolutions des instruments prévus dans le cadre des futures missions satellites météorologiques comme MeTOp-SG.

ARTS détermine les Jacobiens pendant le processus de simulation. Les unités pour les Jacobiens sont $\frac{\partial T_b}{T_{leveli}} = \frac{K}{K}$ par rapport au profil de température et $\frac{\partial T_b}{\partial v_{mr}} = \frac{K}{1}$ en ce qui concerne les changements dans les profils de gaz mineurs, de H₂O et de O₃. La Figure II.11 représente les Jacobiens pour la température et l’humidité, moyennés sur toutes les atmosphères considérées. Ces Jacobiens varient selon le type d’atmosphère, mais la Figure II.11 illustre le lien moyen entre les variables géophysiques à

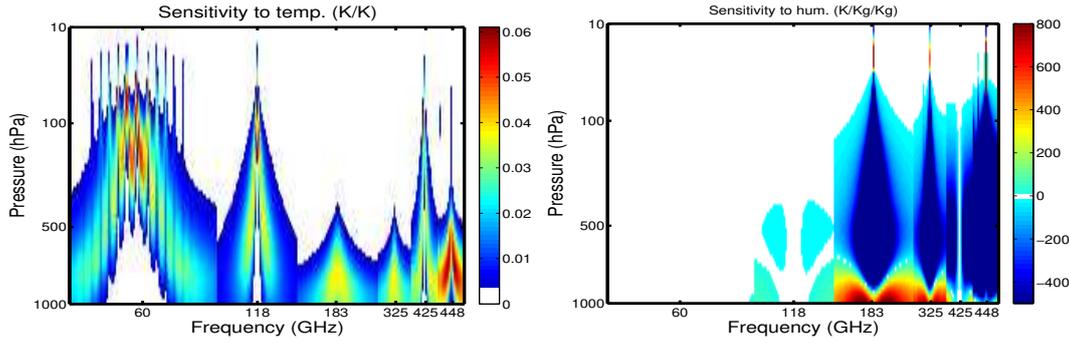


Figure II.11. *Jacobiens en Température (gauche) et humidité (droite) pour les 251 canaux choisis dans les bandes d'absorption à 60, 118, 183, 325, 425, and 448 GHz. L'axe horizontal n'est pas continu en fréquence.*

restituer (T et q) et les radiances observables. On peut voir qu'un même canal peut être sensible aux deux variables (par exemple à 183 GHz) et donc qu'une sélection de canaux doit en tenir compte pour dé-mélanger les deux signaux dans sa mesure. Les Jacobiens pour l'humidité sont sensibles dans la troposphère, principalement pour les canaux situés dans la raie d'absorption de H_2O . Plus le canal est éloigné du centre de la raie, plus il sonde à basse altitude. Les Jacobiens pour la température sont élevés pour la plupart des canaux de sondage. On s'attend à ce que tous les canaux soient affectés par des signatures spectrales qui dépendent de la température. Certains canaux dans la bande des 60 GHz sont sensibles à la température à haute altitude. La bande d'humidité de 448 GHz est plutôt sensible à la température dans la basse atmosphère, beaucoup plus que la bande d'humidité de 183 GHz. Cela aura probablement un impact sur la sélection des canaux pour T et q .

4.1.d Les bruits d'observation

Dans la plupart des méthodes de réduction de dimension des données, le bruit de l'instrument et les incertitudes du transfert radiatif doivent être pris en compte.

Les niveaux de bruit des instruments choisis dans cette étude reposent sur des spécifications récentes en matière de bruit des récepteurs. Les détails peuvent être trouvés dans (Aires et al., 2015). La Table II.4 fournit les hypothèses de bruit pour cette étude.

Les erreurs de transfert radiatif sont complexes à estimer. Dans cette section, nous présentons un modèle simplifié avec des erreurs basées sur une hypothèse Gaussienne mais réalistes. Le bruit du transfert radiatif en ciel clair comprend les contributions de l'atmosphère et de la surface. La principale source de bruit du transfert radiatif vient de la modélisation de surface. Ce bruit est proportionnel à l'erreur d'émissivité, δ_{emis} , estimée à 0.01. L'erreur est alors :

$$\frac{dT_b}{de} = T_s \tau^2. \quad (II.33)$$

Les spécifications d'erreur du transfert radiatif sont données dans la Table II.4. La corrélation d'erreur pour des canaux proches pour une même raie d'absorption doit

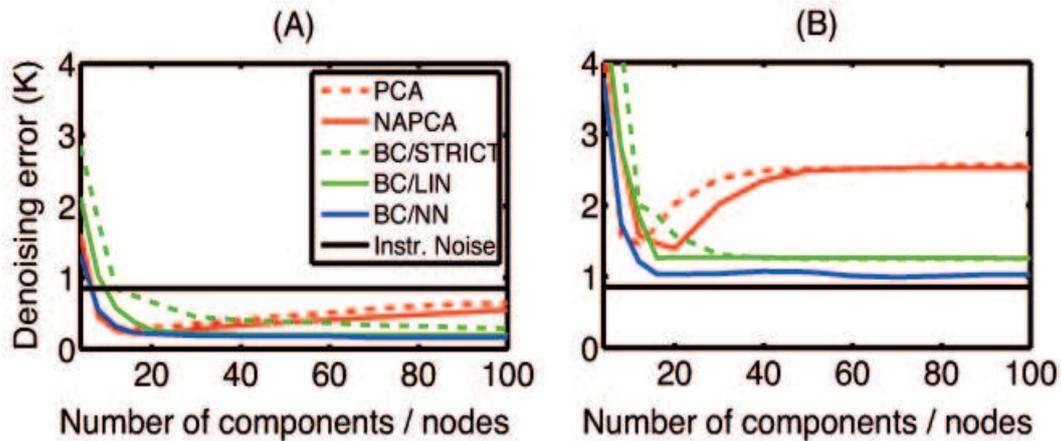


Figure II.12. Évolution de l'erreur de dé-bruitage (i.e., $y - \bar{y}$), avec un nombre croissant de composantes (pour la PCA et NAPCA) ou de BCs stricts ou projetés (linéaires ou non). (A) représente l'erreur de dé-bruitage moyen sur tout le spectre d'observation, et (B) représente l'erreur de dé-bruitage maximal atteint sur les plus subtiles canaux. Le bruit instrumental moyen comprenant l'erreur du modèle direct est représenté par sa déviation standard à des fins de comparaison (en noir).

être prise en compte. La matrice de corrélation est construite en utilisant deux Gaussiennes tronquées pour modéliser la dépendance de corrélation avec les fréquences. Pour un canal particulier, la plage de corrélation considérée est de 1 GHz de part et d'autre du canal. À partir des spécifications de bruit pour chaque canal et de la matrice de corrélation, il est alors possible de calculer la matrice de covariance d'erreur résultante.

4.2 Résultats en termes de compression

4.2.a Comparaison en compression

Les cinq techniques de compression sont testées ici : deux techniques classiques la PCA et la NAPCA, et trois configurations différentes de la nouvelle approche par représentation BCs : BC-stricts, BC-Projetés/Lin et BC-Projetés/non-Lin. Afin de mesurer la capacité de la méthode de compression à réduire le bruit dans les observations (bruit d'instrument et incertitudes du transfert radiatif), les statistiques sont effectuées sur $(y - \bar{y})$ où y est le vecteur d'observation sans bruit et \bar{y} est la compression puis la décompression des observations bruitées \bar{y} . La déviation standard de $(y - \bar{y})$ est appelée erreur de dé-bruitage ; plus elle est faible, plus la technique de compression est capable de réduire le bruit d'observation dans \bar{y} . Dans la Figure II.12-(A), les erreurs de dé-bruitage sont représentées pour un nombre croissant de composantes ou BC, pour les méthodes PCA, NAPCA et BC.

La NAPCA surpasse toujours l'approche PCA puisqu'elle utilise explicitement l'information du bruit d'observation dans \bar{y} . Par la suite, les résultats ne se concentreront plus que sur l'approche NAPCA au lieu de la PCA classique. Pour les méthodes PCA et NAPCA, les erreurs de débruitage diminuent avec l'augmentation du nombre de composantes jusqu'à ce qu'un optimum soit atteint à environ $n=16$. En effet, la va-

riabilité liée au bruit d'observation n'est représentée que par des vecteurs de la BON d'ordre élevée (voir section sur l'ACP et ses variantes). Chaque vecteur essaie de représenter les résidus des projections selon les vecteurs précédents, de sorte que leur contenu informationnel est vraiment négligeable à partir d'un certain rang. La plupart du temps, les composantes d'ordre supérieur ne font que coder des variabilités de haute fréquence dans le vecteur d'observation et qui sont très susceptibles de n'être que du bruit. Le nombre optimal de composantes n'est défini que pour l'ensemble du vecteur d'observation : c'est le meilleur compromis de dé-bruitage lorsqu'on considère les $m = 251$ canaux. Il est possible que certains canaux représentant des caractéristiques spectralement rares nécessitent plus de vecteurs pour être bien représentés par la PCA, et que d'autres canaux soient sur-représentés (ce qui signifie que la PCA codera même leur bruit instrumental ce qui est inefficace).

L'approche BC-strict est moins performante que la NAPCA lorsqu'un nombre limité de BCs est utilisé (Figure II.12-(A)). Avec 10 BCs, seuls 10 canaux sont utilisés alors que pour la compression, 10 composantes pour le PCA ou NAPCA utilisent le vecteur d'observation complet \tilde{y} . Cependant, l'approche BC a une erreur de dé-bruitage décroissante monotone qui converge vers l'optimum des approches PCA et NAPCA. Avec environ 90 BC, l'erreur de dé-bruitage est équivalente à l'utilisation des 16 composantes NAPCA.

Les BC-projetés ont de bien meilleurs résultats : la version linéaire converge vers une erreur identique à NAPCA avec seulement 30 BCs, et la version non linéaire converge avec seulement 20 BCs. Le BC-projeté/non-Lin montre une amélioration significative par rapport à la version BC-projetés/Lin. Elle est comparable à la méthode NAPCA en termes de taux optimal de compression, et les erreurs de décompression n'augmentent pas avec le nombre croissant de BCs.

Le choix de l'une ou l'autre approche devrait être basé sur la nécessité d'une représentation des observations satellites en composantes indépendantes (avec le NAPCA) ou d'une représentation similaire aux canaux (BC). La plupart des centres PNT préfèrent l'utilisation de canaux purs (afin de mieux découpler les variables et de faciliter le processus d'assimilation).

Toutes les méthodes (PCA, NAPCA et BC) prennent en compte un critère d'erreur moyen sur l'ensemble du spectre d'observation. La Figure II.12-(B) présente l'erreur maximale de dé-bruitage sur le vecteur d'observation, en fonction du nombre de composantes ou de BC. $Max(y - \tilde{y})$ converge vers $max(y - \tilde{y})$ qui est égal à 2,5 K (l'instrument et les bruits RT pour le canal 448 GHz, voir Table II.4). Cette erreur maximale peut aider à décider combien de BCs/PCs doivent être utilisés. Même si elle n'est pas optimisée pour la réduction de l'erreur maximale, la méthode BC-projeté a une diminution monotone de l'erreur. Au contraire, les méthodes basées sur la PCA ont une augmentation très importante de l'erreur maximale lorsque le nombre de composantes dépasse environ 20 composantes.

4.2.b Analyse spectrale des erreurs de compression

La Figure II.13 représente la variation spectrale des erreurs de dé-bruitage ($y - \tilde{y}$), pour les six bandes spectrales (à gauche), et autour de la bande 60 GHz (au milieu), et autour de la bande 183 GHz (à droite). Les résultats sont représentés pour les méthodes PCA, NAPCA, BC-Strict, BC-Projeté/Lin et BC-Projeté/non-Lin pour un nombre différents de composantes ou BCs. Comme pour le diagnostic moyen présenté pré-

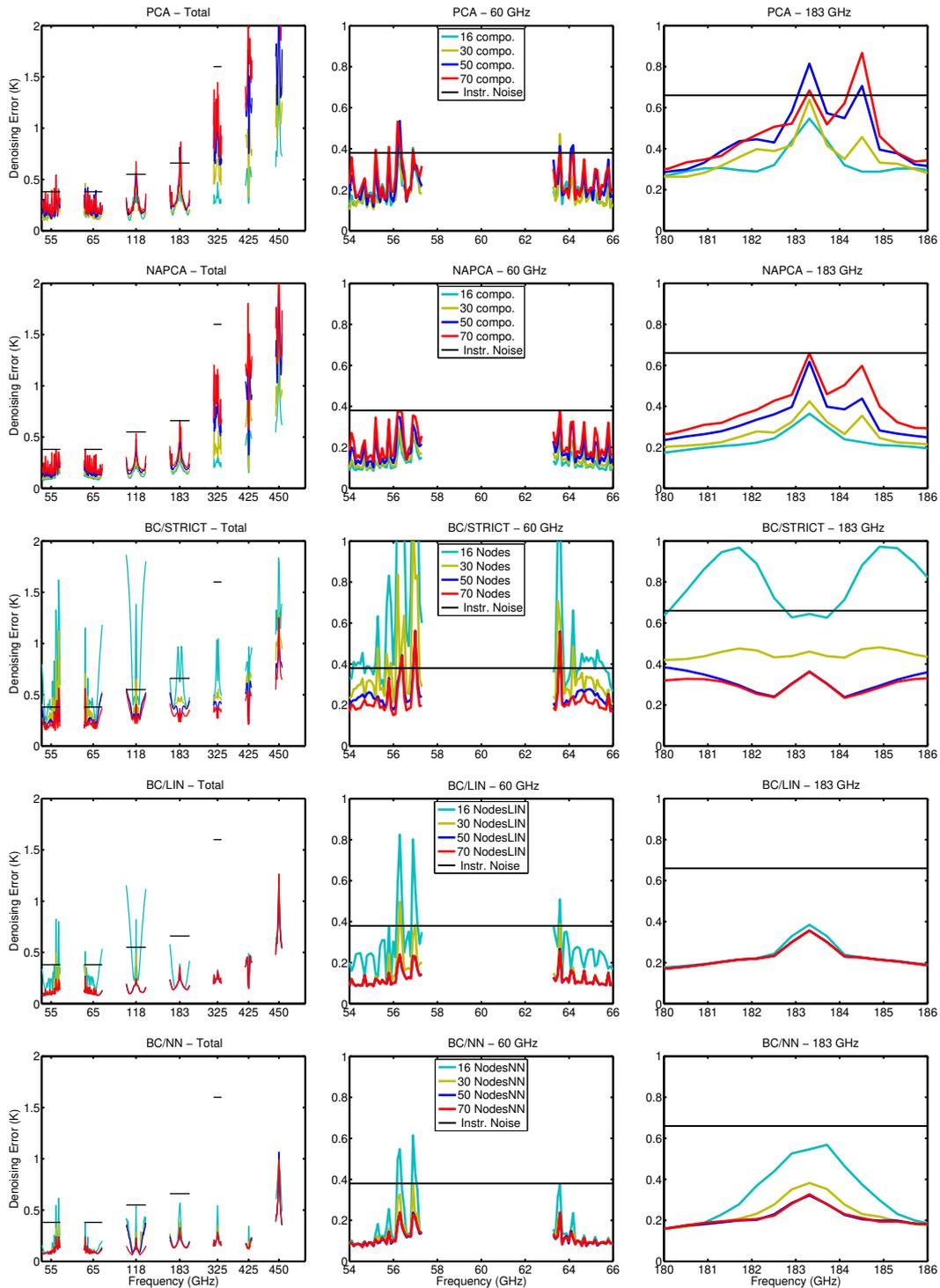


Figure II.13. Erreur de dé-bruitage ($y - \bar{y}$), pour les six bandes spectrales (à gauche), et autour de la bande 60 GHz (au milieu), et autour de la bande 183 GHz (à droite). Les différentes méthodes de compression sont présentées, de haut en bas : PCA, NAPCA, BC-stricts, BC-Projetés/Lin et BC-Projetés/non-Lin pour un nombre différent de composantes ou BCs. Le bruit instrumental avec l'erreur du modèle est représenté en noir à des fin de comparaison.

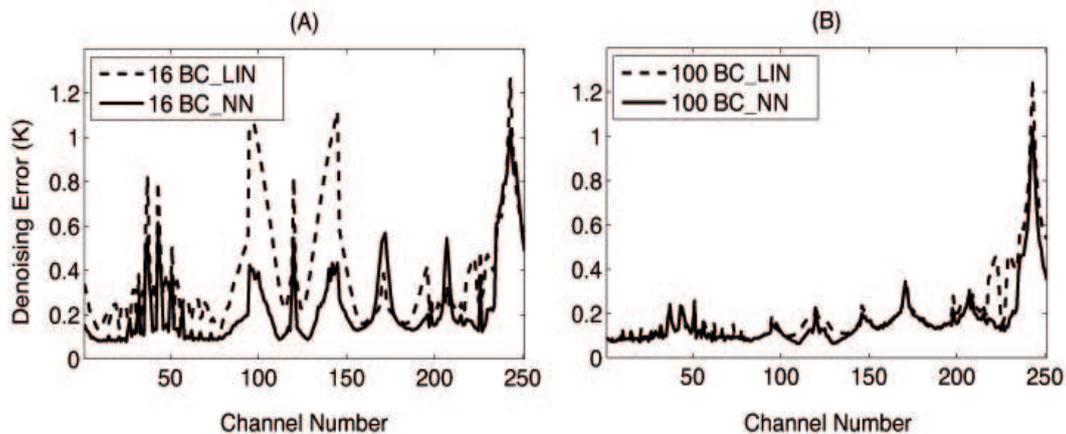


Figure II.14. Erreur de dé-bruitage pour un instrument micro-onde $STD(y - \bar{y})$ où $\bar{y} = Q \cdot P \cdot \tilde{y}$, est une régression linéaire, ou bien $\bar{y} = NN_2(NN_1(\tilde{y}))$, une régression utilisant un réseau de neurones; avec 16 (A) ou 100 (B) BCs utilisés.

cédemment, pour les approches liées à la PCA, une fois que le nombre optimal de composantes est atteint, l'erreur de dé-bruitage tend à augmenter. Pour les approches BC, l'erreur de dé-bruitage diminue et atteint un plateau. Pour toutes les méthodes, il est possible de sélectionner un nombre de composantes pour lesquels l'erreur de dé-bruitage est inférieure au bruit de l'instrument pour presque tous les canaux. Notez que contrairement à NAPCA, certains canaux reconstruits par PCA avec de nombreux PCs ont finalement un bruit supérieur à leur bruit instrumental; cette représentation non optimale est probablement liée au manque d'informations sur le bruit dans la PCA classique.

La Figure II.14 représente l'écart-type des erreurs entre les observations sans bruit y et le résultat \bar{y} de la régression linéaire $\bar{y} = Q \cdot P \cdot \tilde{y}$ ou la régression du réseau de neurone $\bar{y} = NN_2(NN_1(\tilde{y}))$. La régression NN a toujours une erreur inférieure, de sorte qu'il est avantageux d'utiliser une régression non linéaire puisque le coût de calcul supplémentaire est négligeable. Ces erreurs sont inférieures au bruit instrumental et cette diminution de l'erreur par rapport au bruit instrumental ne résulte que de la redondance d'information dans les observations. L'impact du modèle non linéaire est plus important lorsque le nombre de BCs est limité (c'est-à-dire 20 BC). Avec 100 BCs, l'avantage d'un NN est considérablement réduit. Seuls certains canaux particulièrement difficiles à représenter bénéficient de la non-linéarité (ce qui pourrait servir d'indice de la complexité de certains canaux). Par conséquent, le choix d'utiliser ou non des BC-projetés/non-Lin devrait être basé sur la complexité du vecteur d'observation à comprimer et sur le taux de compression requis pour l'application particulière. Les inconvénients sont le coût de calcul mais aussi le fait que les Jacobiens des projections non linéaire de canaux ne peuvent être obtenus directement via les projection NN, ce qui ne facilite pas leur utilisation future dans une analyse de contenu d'information.

Puisque les BC-Stricts sont moins efficaces que les BC-Projetés, les premiers seront privilégiés dans ce qui suit. En outre, les BC-Projetés/non-Lin sont trop coûteux en temps de calcul dans la pratique, de sorte que l'analyse du contenu en information se concentrera sur les BC-Projetés/Lin.

4.3 Résultats en termes de sélection de canaux

4.3.a Comparaison avec les méthodes de sélection de canaux

Ici, les méthodes de sélection des canaux sont comparées. Notre étude comprend la méthode BC-Projetés/Lin, la sélection de canaux par MRE et par ER. Ces méthodes de sélection de canaux suivent des objectifs différents : pour le MRE et l'ER, l'objectif est de sélectionner des canaux qui optimisent la restitution d'une variable donnée alors que pour la méthode BC-Projetés/Lin, comme pour les approches basées sur la PCA, l'objectif est de sélectionner les canaux qui peuvent reproduire au mieux l'ensemble des observations. Pour les méthodes de sélection des canaux qui sont orientées vers la restitution, des restitutions indépendantes et combinées des deux variables T et q sont également testées.

Les fréquences sélectionnées sont présentées dans la Figure II.15 de haut en bas, pour les méthodes BC-Projetés/Lin, MRE et ER. Les résultats sont affichés lorsqu'on sélectionne 50 canaux dans les bandes O_2 pour restituer T (à gauche), 50 canaux dans les bandes H_2O pour restituer q (au milieu), ou 100 canaux dans les bandes O_2 et H_2O pour la restitution simultanée de T et q (à droite). Puisque les BC-Projetés/Lin ne tiennent pas compte des variables à restituer, il n'y a pas de différence entre les approches indépendantes ou combinées de T et q . Comme la PCA ou la NAPCA, la méthode BC-Projetés/Lin tend à représenter au mieux l'ensemble des observations. Elle sélectionne les canaux dans toutes les bandes (60, 118, 183, 325, 420 et 448 GHz), quelle que soit la variable à estimer.

Les configurations indépendantes pur- T , pur- q et combinés $T + q$ diffèrent pour les méthodes de sélection de canal MRE et ER. Pour T , les canaux sont quasi exclusivement choisis dans la bande 60 GHz O_2 . Pour la restitution de l'humidité q , les canaux sélectionnés sont situés principalement dans la bande 183 GHz mais la bande 448 GHz est également échantillonnée, même si la contamination par température est plus élevée (Figure II.11) pour cette bande, et le bruit plus important. Cependant, certains canaux dans la bande 118 GHz sont sélectionnés dans la restitution combinée $T + q$, probablement en raison de la sensibilité accrue à l'humidité par rapport à la bande 60 GHz, et de la contribution possible de la bande 183 GHz à la séparation des deux contributions dans T et q . Les premiers canaux sélectionnés par la méthode ER sont tous situés dans la bande 183 GHz H_2O .

4.3.b Comparaison des Jacobiens

Les Jacobiens sont représentés dans la Figure II.16 pour les composantes NAPCA, les BC-Projetés/Lin, et les canaux sélectionnés via MRE et ER. Pour les méthodes de pure sélection de canal (MRE et ER), les Jacobiens des canaux résultant des deux types de sélections (indépendante et combinée pour T et q) y sont montrés. Les Jacobiens des composantes de NAPCA n'ont pas de signification physique réelle, ils peuvent osciller entre les valeurs positives et négatives et ce avec des ordres de grandeur non liés à la physique d'un canal. La première composante du NAPCA s'efforcera de transmettre un maximum d'information liée à la fois à la température et à l'humidité, et ce à différents niveaux atmosphériques.

Les pures méthodes de sélection de canaux (MRE et ER) permettent d'obtenir les Jacobiens de canaux réels. Leur répartition le long de la verticale atmosphérique montre bien comment les canaux échantillonnent l'atmosphère pour T et q . Lorsqu'elle

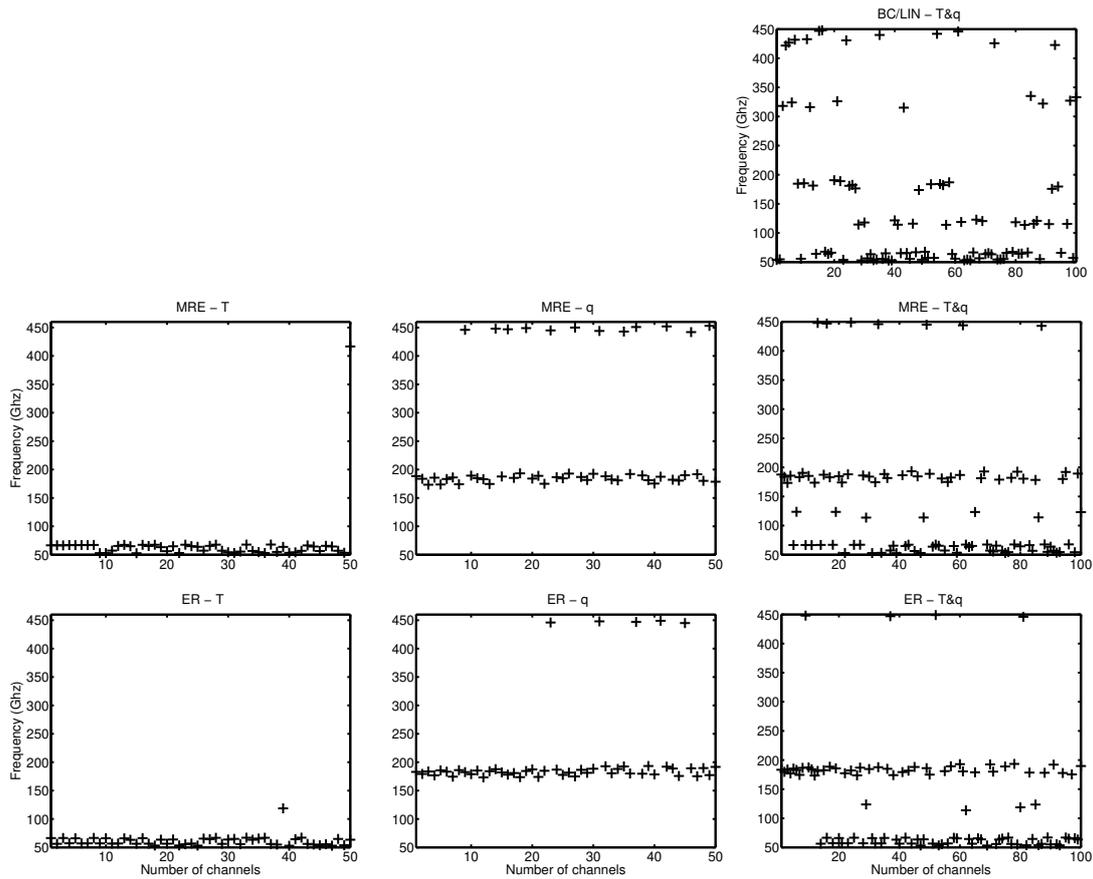


Figure II.15. De haut en bas : fréquence de canaux sélectionnés par BC-Projeté/Lin, MRE et ER. De gauche à droite : sélection de 50 canaux dans la bande O_2 pour la restitution de T uniquement (à gauche), sélection de 50 canaux dans la bande H_2O pour la restitution de q uniquement (au milieu), 100 canaux dans les bande O_2 et H_2O pour la restitution combinée de $T + q$ (à droite).

est effectuée sur T seulement, la bande des 60 GHz apparaît en premier, mais ER semble décrire plutôt les niveaux plus élevés par rapport à MRE. Lorsqu'elles sont effectuées avec la restitution combiné $T + q$, les deux méthodes semblent se concentrer un peu plus sur l'humidité, et moins sur la température, parce que les potentiels d'améliorations de restitution sont plus élevés pour l'humidité.

Pour la sélection combinée $T + q$, le critère de qualité MRE ne permet pas de sélectionner des canaux avec des sensibilités en haut de l'atmosphère car les améliorations dans la haute atmosphère ne contribuent pas tant à l'amélioration du critère de sélection MRE. Les sensibilités par rapport q des différentes méthodes MRE et ER sont plutôt proches, et elles ne diffèrent pas sur leurs configurations pur- q et $T + q$.

Les Jacobiens des BC-Projeté/Lin sont calculés à l'aide de la matrice de projection P (définissant les BC-Projetés/Lin, voir section précédente) à partir des jacobiens du spectre complet. Les Jacobiens des BCs sont donc une combinaison linéaire des Jacobiens originaux. Les Jacobiens des BC-Projeté/Lin ont un comportement très proche des canaux issus des méthodes de sélection de canal confirmant que les BCs tentent de reproduire des canaux réels (mais avec un bruit réduit). Ils échantillonnent bien les températures verticales et ont une bonne description du profil d'humidité. Comme

pour la technique de compression NAPCA, il n'y a pas de configuration pur- T ou pur- q puisque les BCs ont l'intention de représenter au mieux le spectre complet observé, sans aucune considération sur la variable à restituer. Les BCs tendent à reproduire un canal donné et leur signification physique est imposée.

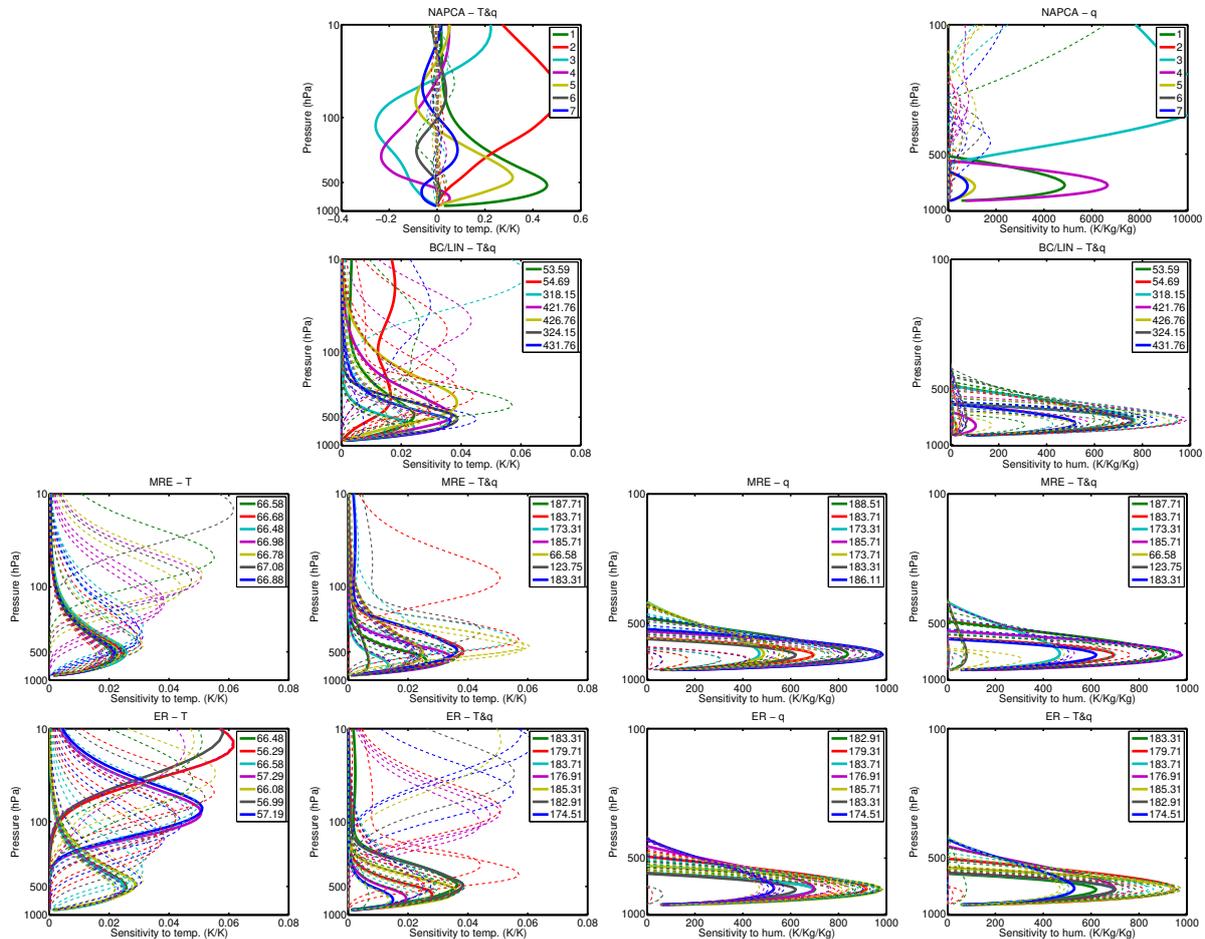


Figure II.16. Pour les PC/BCs/canaux obtenus : Jacobien en température T (K/K) (les deux colonnes de gauche) et Jacobien en humidité q (K/Kg/Kg) (les deux colonnes de droite). De haut en bas : NAPCA, BC-Projeté/Lin, MRE et ER. Pour MRE et ER, les Jacobiens T sont affichés pour la sélection pure- T ou $T + q$, et les Jacobiens q sont affichés pour la configuration pur- q ou $T + q$. Chaque fois, les sept premiers PC/BCs/canaux sont indiqués en gras et leurs fréquences sont indiquées, les autres PC/BCs/canaux d'ordre supérieur sont représentés par une ligne en pointillés.

4.4 Analyse du contenu en information

La Figure II.17 représente l'amélioration de la restitution des variables T et q avec les canaux sélectionnés par rapport à l'information *a priori*, définie dans l'Eq. II.16. Deux configurations de canaux sont illustrées : l'une avec $2 \cdot 10$ canaux sélectionnés indépendamment (10 pour T et 10 pour q), l'autre avec 20 canaux sélectionnés simultanément pour T et q . Les améliorations sont plus élevées pour l'humidité que pour

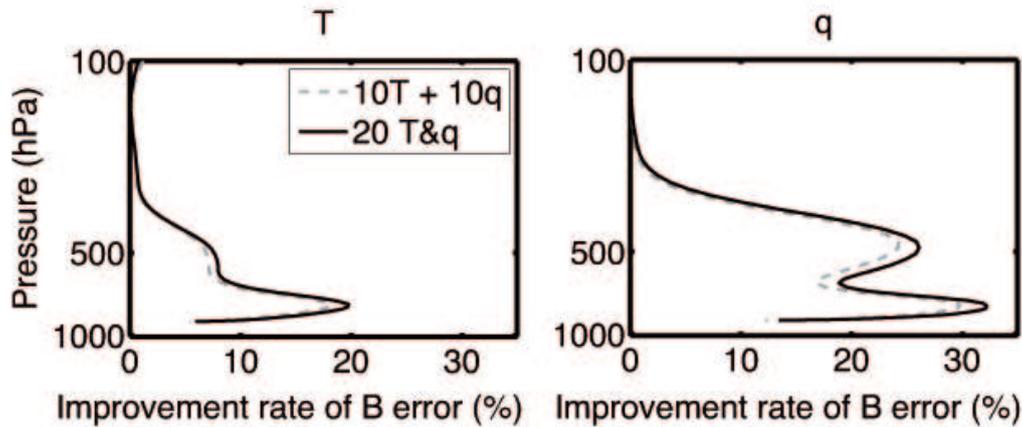


Figure II.17. Taux d'amélioration (en %) de l'erreur de restitution par rapport à l'*a priori* pour la température (en haut) et l'humidité (en bas) lors de l'utilisation de 20 canaux sélectionnés dans le micro-onde : 10 canaux sélectionnés pour T plus 10 canaux pour q (ligne grise en pointillés) ou 20 canaux simultanément pour T et q (ligne noire continue).

la température car l'erreur de l'*a priori* sur la température est relativement faible pour la plupart des situations atmosphériques, ce qui réduit les possibilités d'amélioration. La Figure II.17 montre que les améliorations sont toujours plus importantes lorsqu'on utilise une technique de sélection de canaux simultanément sur T et q . Cela confirme que la sélection des canaux sur toutes les variables (T et q ici) est plus optimale que de le faire indépendamment pour chaque variable. Cependant, dans cet exemple, les différences sont limitées : environ 3% pour T et q . Plus le nombre de canaux est faible, plus la différence est élevée et plus le nombre de canaux est élevé, plus la différence est faible. Dans cette application dans les micro-ondes, de nombreux canaux sont sélectionnés et il n'est pas problématique de sélectionner indépendamment les canaux T et q . Si le problème était moins linéaire et si le rapport entre le nombre de canaux sélectionnés et le nombre total de canaux était réduit, la stratégie simultanée devrait être utilisée à la place. Dans la suite, seules les sélections de canaux indépendantes pour T et q seront comparées aux méthodes de compression.

La Figure II.18 représente des améliorations d'erreur de restitution par rapport à l'*a priori* vu dans l'Eq. II.16, en comparant les quatre méthodologies, ER, MRE, BC-Projeté/Lin et NAPCA, lorsque 20 et 100 composantes/BCs/canaux sont sélectionnés pour la restitution de la température et de l'humidité. Lorsque seulement 20 composantes/BCs/canaux sont sélectionnées, les techniques de compression (NAPCA et BC-Projeté/Lin) sont nettement plus performantes que les sélections de canaux (ER et MRE). Les BCs sont presque aussi bons que les PCs de la NAPCA, même si l'erreur de compression NAPCA est, par construction, optimale. Lorsque suffisamment de canaux sont sélectionnés (p. ex. 100), les différences entre méthodes deviennent négligeables. Cependant, 100 canaux représentent plus d'un tiers du nombre total de canaux dans cette application, donc ce n'est pas idéal. Ces résultats montrent clairement que dans un problème où le bruit d'observation est significatif, les stratégies de compression présentent un avantage certain. Les BCs offrent cet avantage avec la possibilité supplémentaire de représenter le spectre d'observation avec des pseudo-

températures de brillance.

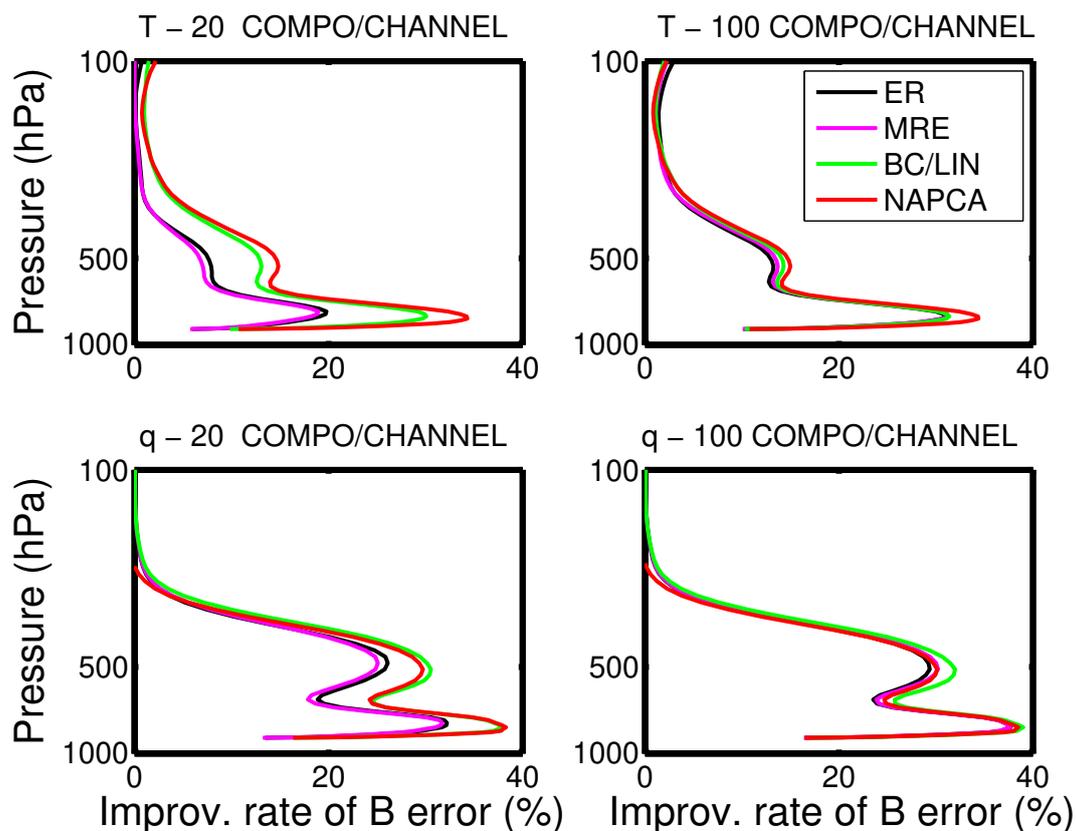


Figure II.18. Taux d'amélioration (en %) de l'erreur de restitution par rapport à l'a priori pour température (en haut) et l'humidité (en bas) lors de l'utilisation de 20 (à gauche) ou 100 (à droite) PCs/BCs/canaux, pour les méthodes : ER, MRE, BC/LIN et NAPCA. La sélection a été effectuée simultanément pour T et q .

5 Réduction de dimension pour l'instrument IASI dans l'infrarouge

Des expériences similaires à celles menées précédemment sur le concept d'instrument hyperspectral dans les micro-ondes sont ici présentées dans le contexte opérationnel du spectromètre infrarouge IASI et ses 8461 canaux. La dimension du spectre de IASI, bien supérieure aux 251 canaux de l'instrument micro-onde, les spécifications réelles du bruit instrumental et l'existence de sélections de canaux opérationnelles pour l'instrument (Collard, 2007) nous ont permis de tester et de comparer l'approche BC avec les autres méthodes de réduction de dimension dans un cas de données hyperspectrales et concrètes. En outre, les configurations locales et contraintes de BCs ont été testées. Les résultats qui suivent sont issus de (Pellet et Aires, 2018).

Outre la restitution des profils de température atmosphérique T et d'humidité q , on s'intéresse aussi au contenu en ozone O_3 . Les conditions sont toujours en ciel clair et au nadir mais au-dessus de l'océan comme au-dessus de la surface terrestre. Après

avoir présenté l'instrument IASI et le matériel nécessaire à l'étude, les différentes méthodes de réduction de dimension sont d'abord testées sur des mesures de températures de brillance simulées pour IASI puis une évaluation avec des données réelles pour l'inversion sera testée avec les différentes méthodes de réduction de dimension.

5.1 Matériels

5.1.a L'instrument IASI

Le sondeur IASI est un spectromètre infrarouge à transformée de Fourier, basé sur un interféromètre Michelson (Cayla et al., 1995). Il fait partie des instruments à bord des plateformes satellites Metop-A (lancée en 2006) et Metop-B (2012). L'objectif principal de l'IASI est de soutenir les centres PNT (Collard, 2007). Son spectre s'étend de 645 à 2760 cm^{-1} avec une résolution spectrale de 0.25 cm^{-1} . Le spectre de l'IASI est divisé en trois bandes principales : B1 de 645 à 1210 cm^{-1} , B2 de 1210 à 2000 cm^{-1} , et B3 de 2000 à 2760 cm^{-1} . La Table II.5 indique les variables géophysiques relatives à la mesure de l'IASI. L'IASI permet de restituer les profils de température T et de vapeur d'eau q dans la troposphère et la basse stratosphère (Aires et al., 2002c), ainsi que de mesurer les concentrations d'ozone O_3 , de monoxyde de carbone CO , de méthane CH_4 et d'autres composés (Clerbaux et al., 2009).

Wavenumber (cm^{-1})	Geophysical variable
650-770	temperature (T)
770-980	cloud and surface properties
1000-1070	ozone (O_3)
1080-1150	cloud and surface properties
1210-1650	humidity (q), T , N_2O , CH_4 , SO_2
2100-2150	total CO
2150-2250	T and total N_2O
2350-2420	T
2420-2700	cloud and surface properties
2700-2760	total CH_4

TABLE II.5 – Résumé des variables géophysiques relatives à la mesure de l'IASI

La Figure II.19 (en haut) montre, pour une situation atmosphérique donnée, les TBs qui sont mesurées par l'instrument IASI. À un nombre d'onde donné, la TB résulte de l'émission et de l'absorption des différentes composantes atmosphériques le long de l'atmosphère à ce nombre d'onde. Plus la TB est chaude, moins le signal a été absorbé par son trajet dans l'atmosphère (pas de bande d'absorption des composantes atmosphériques à ce nombre d'onde) et plus les canaux de IASI sondent profondément dans l'atmosphère. Par exemple, les TBs les plus chaudes (280 K entre 770-980, 1080-1150 et 2420-2700 cm^{-1}) correspondent aux canaux de la fenêtre, caractérisant les propriétés de surface. L'échantillonnage spectral d'une bande d'absorption (CO_2 ou H_2O par exemple, voir Table II.5) fournit des informations sur ces concentrations de gaz.

Le bruit de l'instrument IASI a été modélisé par le Centre National d'Études spatiales (CNES) à l'aide d'une distribution de bruit Gaussien centrée, caractérisée par son écart-type $std(\nu, TB)$ fonction du nombre d'onde ν et de la température de brillance

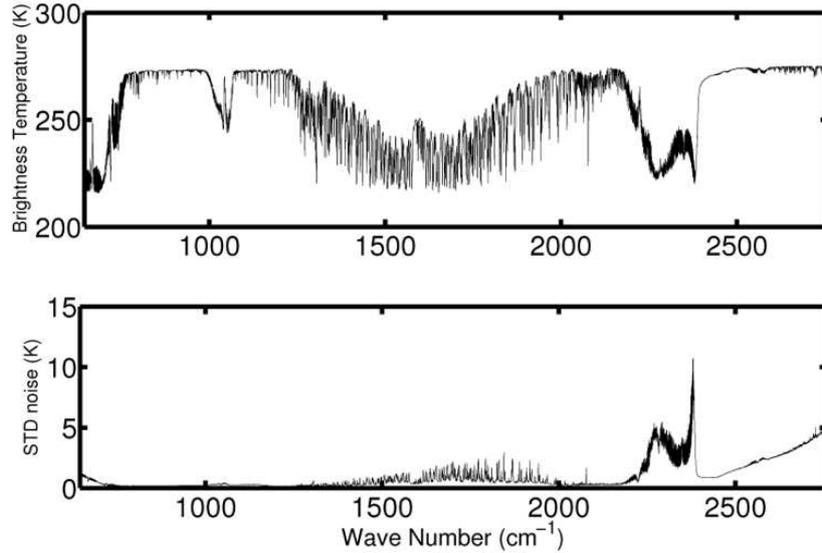


Figure II.19. Spectre en température de brillance de IASI simulé (en haut) avec le bruit instrumental correspondant (en bas).

mesuré TB :

$$std(\nu, TB) = \frac{\frac{\partial B(TB=280, \nu)}{\partial TB}}{\frac{\partial B(TB=T, \nu)}{\partial TB}} \cdot std_{280}, \quad (\text{II.34})$$

où $B(TB, \nu)$ est la fonction de Planck, et std_{280} est l'écart type du bruit instrumental à 280 K (NEDT), qui caractérise l'instrument IASI. La Figure II.19 (en bas) représente l'écart-type du bruit instrumental pour une situation atmosphérique donnée. Le niveau élevé de bruit dans la bande B3 explique que cette bande n'est généralement pas utilisée pour la restitution. Nous n'utiliserons pas non plus B3 dans cette étude.

5.1.b Base de données TIGR des profils atmosphériques

La base de données Thermodynamic Initial Guess Retrieval (TIGR) est développée au Laboratoire de Météorologie Dynamique (Chedin et al., 1985 ; Chevallier et al., 1998). Il comprend plus de 2300 profils de radiosondes du monde entier. La Figure II.20 montre la distribution spatiale de TIGR sur le globe. Les profils sont choisis afin de représenter la variabilité maximale des conditions atmosphériques de température et d'humidité. Cela signifie que le TIGR comprend non seulement des profils atmosphériques standards, mais aussi des situations extrêmes pour couvrir un large éventail de conditions atmosphériques, des tropiques aux régions polaires. Chaque situation est caractérisée de la surface au sommet de l'atmosphère par des profils T , q et O_3 , dans une grille verticale de 98 niveaux de pression.

5.1.c Le transfert radiatif via le Optimal Spectral Sampling (OSS)

Le code de transfert radiatif OSS est développé par le centre Atmospheric and Environment Research (Moncet et al., 2010 ; Lipton et al., 2009). OSS est utilisé ici pour calculer les températures de brillance de IASI pour chaque situation atmosphérique, de 645 cm^{-1} à 2760 cm^{-1} et avec une résolution spectrale de 0.25 cm^{-1} , dans des conditions de ciel clair.

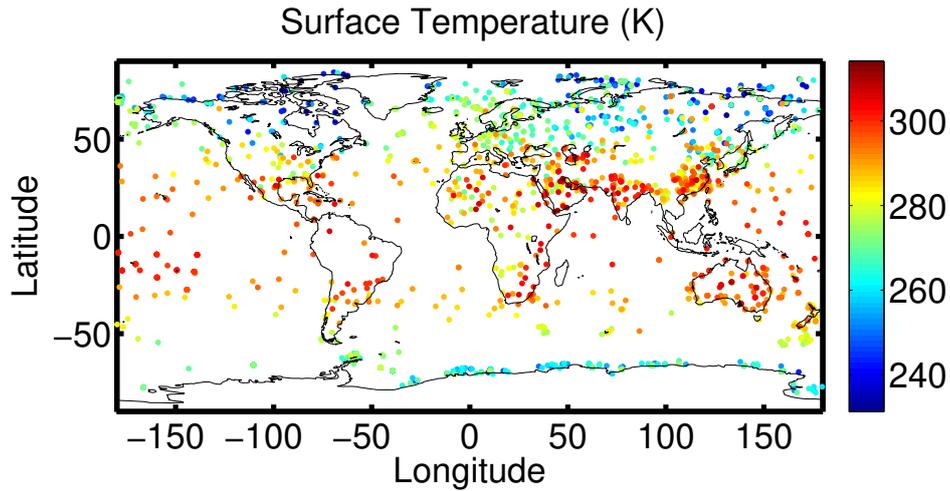


Figure II.20. Distribution spatiale des situations dans la base de données TIGR, la couleur du pixel représente la température de surface (T_s) en K.

Le code OSS a également été utilisé ici pour estimer les Jacobiens du transfert radiatif dans l'infrarouge. Pour cette étude, seuls les Jacobiens en température de surface, température atmosphérique, humidité, ozone sont pris en compte. Les unités pour les Jacobiens par rapport à la température à un niveau particulier sont $\frac{\partial T_b}{\partial T} = \frac{K}{K}$ et $\frac{\partial T_b}{\partial q} = \frac{K}{Kg/Kg}$ pour q et O_3 . Les deux dernières variables sont mesurées en termes de contenu absolu (kilogramme par kilogramme d'air). La Figure II.21 représente les Jacobiens de l'IASI pour T , q et O_3 pour une situation donnée (comme dans la Figure II.19). Tous les canaux IASI semblent être sensibles à T à différents niveaux de l'atmosphère : autour de la surface pour les canaux de 770 à 980 cm^{-1} , dans l'atmosphère moyenne de 1210 à 1650 cm^{-1} , et en haut de l'atmosphère de 650 à 770 cm^{-1} et de 2350 à 2420 cm^{-1} . Les canaux les plus sensibles à T sont aussi les plus sensibles à q . Ceci est dû à la dépendance de l'émission de la vapeur d'eau à la température et ce à différents niveaux atmosphériques, produisant ainsi une corrélation élevée entre la température et la vapeur d'eau. Enfin, moins de quatre cents canaux (de 1000 à 1100 cm^{-1}) sont sensibles à la concentration de O_3 dans la haute atmosphère où O_3 est présent (voir Table II.5).

5.1.d Matrice de covariance d'erreurs *a priori*

Dans les expériences suivantes, nous avons construit la matrice de covariance des erreurs *a priori* pour faciliter l'interprétation des résultats. La matrice de covariance B est définie comme : $Cov(x, z) = \sigma(x) \cdot \sigma(z)$ où x et z représentent deux variables géophysiques à restituer, σ l'écart-type. Il est donc nécessaire de décider d'une matrice de corrélation et d'un écart-type pour les variables T , q et O_3 :

- On suppose communément que la corrélation des erreurs d'humidité (respectivement d'ozone) entre les couches verticales est proche de la corrélation associée à l'humidité même (resp. ozone). La matrice de covariance d'erreurs *a priori* B pour q et O_3 basée sur : (1) les matrices de corrélation entre l'humidité et l'ozone, et (2) une variance d'erreur de 20% pour q (resp. 30% pour O_3) sur chaque couche.
- La température est fortement corrélée entre les couches basses de l'atmo-

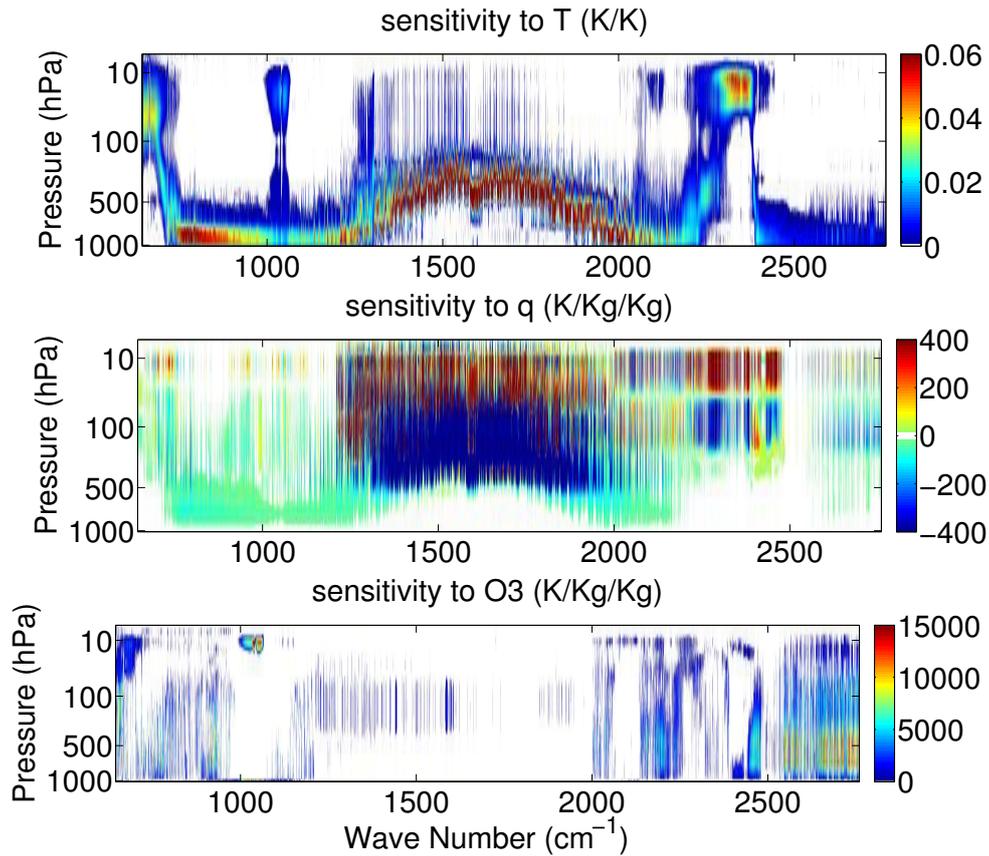


Figure II.21. De haut en bas : Jacobiens des canaux de l'instrument IASI en température (T), humidité (q), et ozone (O_3), donnés pour une situation particulière dans TIGR.

sphère, de sorte que l'utilisation de la matrice de corrélation de T comme matrice de corrélation des erreurs *a priori* signifierait qu'un seul élément d'information serait nécessaire pour restituer la partie basse du profil T , et qu'un nombre limité de canaux serait sélectionné dans la tropopause. Afin d'éviter cette difficulté, nous avons décidé d'utiliser une erreur indépendante entre les couches atmosphériques. De cette façon, la sélection des canaux cherchera à obtenir des canaux plus uniformément répartis le long de la verticale. Une variance d'erreur de 1 K est généralement utilisée et est choisie ici.

5.1.e Le point sur les nouveaux développement des méthodes

Pour cette étude, des implémentations particulières de ER/DFS et BCs ont été testées.

BCs localisés - Il a été mentionné précédemment que les BCs sont une projection linéaire de l'ensemble du spectre observé : $\eta_s = P \cdot \tilde{y}$. Cela signifie qu'une fréquence sélectionnée s_i peut résulter d'une combinaison linéaire de canaux très dispersés le long du spectre. Pour l'instrument IASI par exemple, un canal sélectionné pour T dans la plage 660 cm^{-1} peut être estimé à partir de mesures dans cette bande, mais aussi à partir de canaux dans la région 2300 cm^{-1} . Il est possible de contraindre les BCs s_i à ne résulter que des fréquences incluses dans leur environnement. Cette version

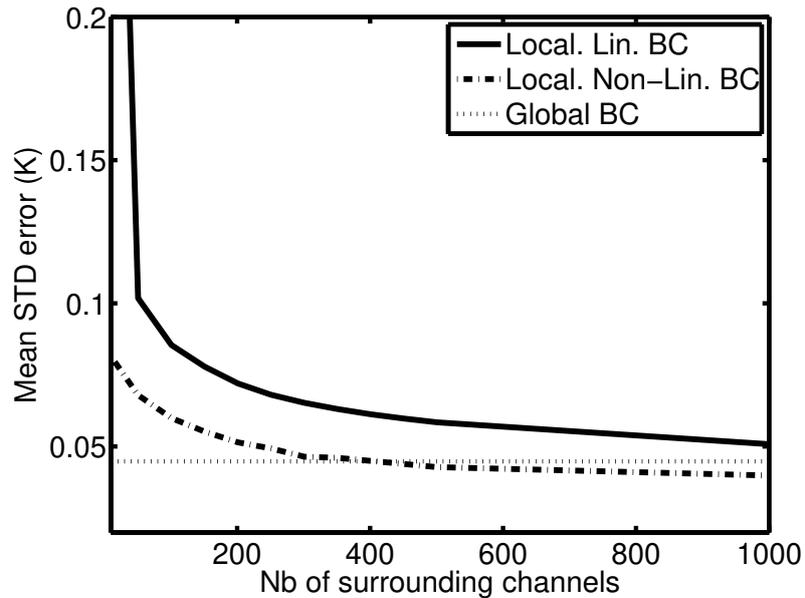


Figure II.22. Évolution des erreurs de projection, $E[(\eta_s - TB)^2]$. En utilisant 100 BC, avec un nombre croissant de canaux voisins pour les approches BC-Localisé/linéaire et BC-Localisé/non-linéaire, comparées à l'erreur de projection des BC-projetés standards.

de BCs est appelée ici les BCs localisés. Il convient de noter que les BCs localisés présentent un risque de contenir un contenu d'information réduit. Par exemple, certaines informations relatives aux gaz à l'état de traces dont la bande d'absorption est étroite (c'est-à-dire accessible par quelques canaux seulement) ne seraient pas représentées dans les BCs si aucun pseudo-canal n'est suffisamment proche de la bande d'absorption de ces gaz mineurs. Cependant, les canaux localisés peuvent avoir un avantage pratique. La caractéristique de localité évite qu'une contamination sur une région spectrale particulière n'ait un impact sur l'ensemble des BCs en limitant spectralement la construction des BCs aux canaux environnant. Par exemple, la contamination des nuages peut être gérée : seuls les BCs situés parmi les canaux des fenêtres seraient touchés par les nuages, mais les autres ne le seraient pas et le dernier serait toujours utile dans le schéma de restitution. En plus de leurs contraintes de fréquence supplémentaires, les BCs localisés sont plus rapides à estimer parce que le nombre d'entrées du modèle est limité dans la régression P .

Le composé le plus subtil sur lequel nous nous concentrons est l'ozone. L'ozone présente une bande d'absorption autour de 1000-1150 cm^{-1} , ce qui signifie que IASI utilise environ 450 canaux pour échantillonner la bande d'absorption d'ozone (avec une résolution spectrale de 0,25 cm^{-1}). Ici, nous avons choisi de considérer 1000 canaux comme le nombre limite de voisins dans la version localisée. Au-delà de 1000 voisins, un BC lié à la bande d'ozone serait construit à partir de canaux non liés à l'ozone et les considérations locales seraient alors perdues.

La Figure II.22 représente les erreurs de reconstruction des BCs du spectre IASI pour diverses configurations : les approches BC-Projeté (générique), BC-Localisé/linéaire, BC-Localisé/non-linéaire avec un nombre croissant de voisins pris en compte dans les versions localisées. On peut voir que la méthode BC-Localisé/linéaire présente une erreur de reconstruction plus élevée que la méthode BC-global, même en utilisant

1000 canaux voisins. Les BCs-Localisés/linéaire, même s'ils peuvent avoir un avantage en considération physique, introduisent une limite sur la capacité de compression de la méthode BC. Les BCs-Localisés/non-linéaires présentent une erreur de reconstruction plus faible que la méthode générique BC-projeté et devraient permettre d'obtenir un meilleur taux de compression. Cependant, cette méthode est très coûteuse en temps et n'a pas été testée plus en détails.

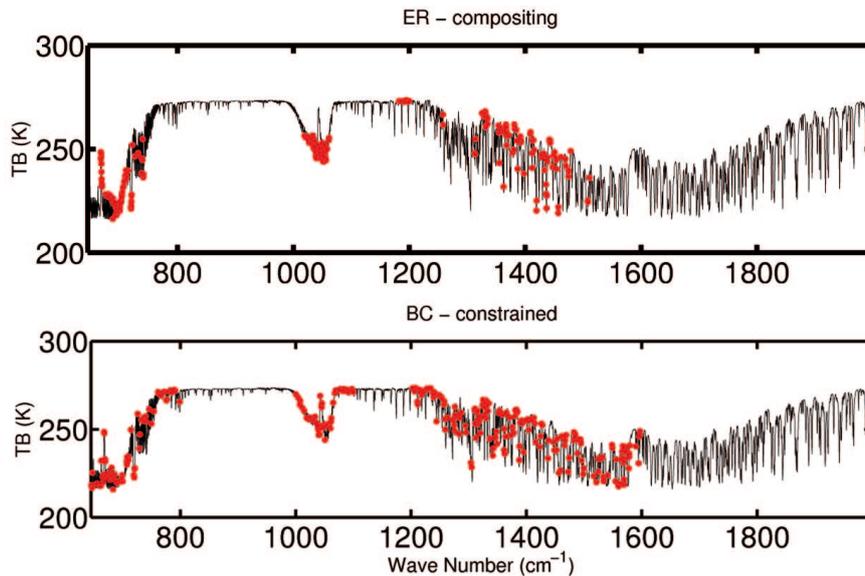


Figure II.23. De haut en bas : ER-selection et BC-constraints.

ER/DFS-compositing - Les fréquences de canaux sélectionnés par ER sont représentées dans la Figure II.23 (en haut). Dans une sélection sans contrainte (pour laquelle n'importe quel canal de la bande B1 et B2 peut être sélectionné), certains canaux sélectionnés pour T peuvent être localisés dans la bande d'absorption H_2O . Ce comportement résulte du fait que la mesure de l'absorption de H_2O dépend de la température. Afin d'obtenir une sélection de température "pure", il est possible de contraindre la sélection ER en B1 seulement pour T et seulement en B2 pour q afin d'éviter ce problème de mélange. Un autre problème est celui de l'impact de la température de surface (T_s) : (1) la corrélation est élevée entre T_s et le profil T dans la tropopause et (2) le bruit instrumental est faible pour les canaux fenêtres. Par conséquent, la sélection pour la restitution de T va être biaisée vers les canaux fenêtres. Afin d'éviter cette difficulté, deux sélections indépendantes peuvent être mises en place : respectivement 20 canaux pour T_s et 100 autres canaux pour le profil T . La sélection pour l' O_3 ne souffre pas de problème de mélange. Les sélections contraintes sont utilisées dans le contexte de l'instrument IASI et sont représentées dans la Figure II.23. La sélection résultante est une composition ("compositing") des quatre sélections indépendantes pour T_s , T , q et O_3 .

BC-Projeté/Lin/contraint - La stratégie de contrainte dans ER-compositing peut être utilisée avec les BCs. Dans le cas de la technique BC-Projeté/Lin/contraint, l'espace du canal candidat est réduit à des bandes d'absorption particulières pour se concentrer sur la restitution de certaines variables géophysiques. Ceci peut également être

utilisé dans le cadre du NAPCA en calculant les PCs sur une bande d'absorption particulière. Les BC-Projetés/Lin/contraints utilisent les mêmes critères liés à la reconstruction (minimisant l'erreur quadratique moyenne de reconstruction du spectre entier dans Eq. (II.32)) que les BCs-Projetés/Lin génériques, mais l'espace des candidats pour sélectionner η_s est réduit et centré autour de bandes d'absorption spécifiques pour les diverses variables atmosphériques d'intérêt. Par la suite seule, la version BC-Projeté/Lin/contraint sera comparée dans l'étude.

5.2 Évaluation avec des données simulées

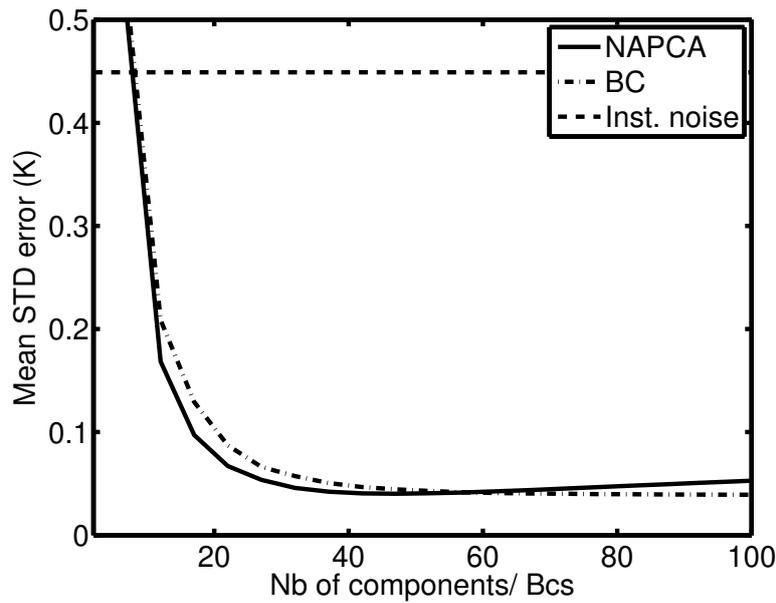


Figure II.24. Erreur de débruitage ($y - \hat{y}$) pour un nombre croissant de PCs ou BCs. Le bruit instrumental moyen est représenté par une ligne noire en pointillés.

Comme pour l'instrument dans le micro-onde, les BCs-Projetés/Lin/contraints sont comparés à la NAPCA pour leur capacité à dé-bruiter le spectre observé. Dans la Figure II.24, les erreurs de dé-bruitage ($y - \hat{y}$) sont représentées pour un nombre croissant de PCs ou de BCs. Pour NAPCA, l'erreur de dé-bruitage diminue avec l'augmentation des composants (PC) jusqu'à ce qu'un optimum soit atteint (environ 47) et que l'erreur commence à croître. De manière asymptotique, l'erreur de dé-bruitage converge vers environ 0,45 K, le bruit moyen de l'instrument (c'est-à-dire la reconstruction parfaite des spectres bruités). L'approche par BC est légèrement moins efficace que celle de la NAPCA. La contrainte supplémentaire, présente dans la construction des BC-Projetés/Lin/contraints, réduit légèrement leur capacité de compression. Cependant, l'approche de la BC a une erreur de dé-bruitage constamment décroissante qui converge vers l'optimum NAPCA. Avec environ 100 BCs, l'erreur de dé-bruitage équivaut à l'utilisation optimale de 47 PCs sur l'ensemble du spectre. La différence entre les erreurs de dé-bruitage spectral avec NAPCA (47 composantes) et BC (100 BCs) est inférieure à $5 \cdot 10^{-3}$ K le long du spectre.

Dans ce paragraphe, on compare les BCs-projetés/lin/contraints à la sélection de canaux par ER-compositing. Les BCs-projetés/lin/contraints sont représentés dans la

Figure II.23. Par définition, les BCs-projetés/lin/contraints sont plus localisés que les BCs-projetés/lin génériques et donc plus proches de la sélection de canaux ER.

Les canaux sélectionnés par ER sont présentés dans la Figure II.23. Il s'agit de la combinaison *a posteriori* des trois sélections indépendantes pur- T pur- q et pur- O_3 . Par définition, les sélections se situent autour des bandes d'absorption des variables à restituer : bande CO_2 pour la température, bande H_2O pour l'humidité et bande O_3 pour l'ozone. Pour T , de nombreux canaux sont choisis dans la région $670-710\text{ cm}^{-1}$ sensible à la haute troposphère et à la basse stratosphère. Ceci est en partie une conséquence de la difficulté de restituer la température dans ces couches et des niveaux de bruit instrumental relativement plus élevés pour ces canaux par rapport aux canaux sensibles à T dans des niveaux atmosphériques inférieurs (voir Figure II.19). La sélection pur- q est située dans la bande d'absorption H_2O centrée à environ 1600 cm^{-1} mais les canaux sont plus rassemblés dans la région $1200-1600\text{ cm}^{-1}$ qui a des niveaux de bruit d'instrumentaux plus bas. Le contenu de l'information sur l'humidité est plus riche verticalement que celui de l'ozone et nécessite une grande distribution le long de la bande d'absorption pour échantillonner du bas vers le haut de l'atmosphère. La sélection d'ozone est bien située dans la bande d'absorption O_3 . Étant donné qu'il y a surtout un contenu d'information intégré dans l'ozone, il n'est pas nécessaire d'avoir un échantillonnage étendu de canaux.

Les Jacobiens en température (en K/K), humidité et ozone (en $K/Kg/Kg$) sont représentés dans la Figure II.25 (panel de gauche) pour les BCs et canaux ER résultants. Comme en micro-onde, les Jacobiens des BCs sont très proches de ceux des canaux réels, c'est l'un de leurs principaux avantages. Les sélections ER et BC échantillonnent T le long de l'atmosphère. Pour la sélection T , les canaux CO_2 et H_2O ont des sensibilités différentes. Les canaux H_2O sont sensibles à la température jusqu'à 100 hPa avec une sensibilité plus élevée que les canaux CO_2 pour le même niveau de pression. On peut noter que la méthode BC-projeté/lin/contraint sélectionnent plus de canaux H_2O que CO_2 . L'allure des Jacobiens en humidité et en ozone est fortement dépendante de l'unité des variables. En effet, $\frac{K}{Kg/Kg}$ amplifiera artificiellement la sensibilité lorsque la concentration de H_2O et O_3 est faible. Puisque le contenu d'information sur l'ozone est très localisé autour de 10 hPa, les jacobiens de l'ozone sont représentés de 100 à 0,1 hPa seulement. L'échantillonnage de l'humidité et de l'ozone est moins distribué le long de la verticale en raison du pauvre contenu en information dans les hautes atmosphères pour ces variables.

Le contenu de l'information est mesuré à l'aide de la métrique présentée par (Rodgers, 1996a) : la matrice de covariance des erreurs de restitution est estimée à l'aide de Eq. (II.15). La Figure II.25 (panel de droite) représente des améliorations d'erreur de restitution par rapport à l'*a priori* Eq (II.16), en comparant les trois méthodologies NAPCA, ER et BC-projeté/lin/contraint pour la restitution de la température, l'humidité et l'ozone. Lorsque seulement 50 éléments d'information (composants ou canaux) sont utilisés, encore une fois NAPCA surpasse nettement la sélection du canal ER. Il y a à peine une amélioration entre 50 et 300 PC. Pour le ER, les 50 premiers canaux sélectionnés regroupent les 17 premiers canaux pour T , 17 premiers canaux pour q , et 16 premiers canaux pour O_3 . Les résultats des BC-projeté/lin/contraint sont presque aussi bons que ceux de la NAPCA. Avec seulement 50 canaux, la sélection ER atteint presque le maximum d'informations en humidité. Ceci peut s'expliquer par le fait que certains canaux CO_2 entre 700 et 800 cm^{-1} sont également sensibles à l'humidité. Il

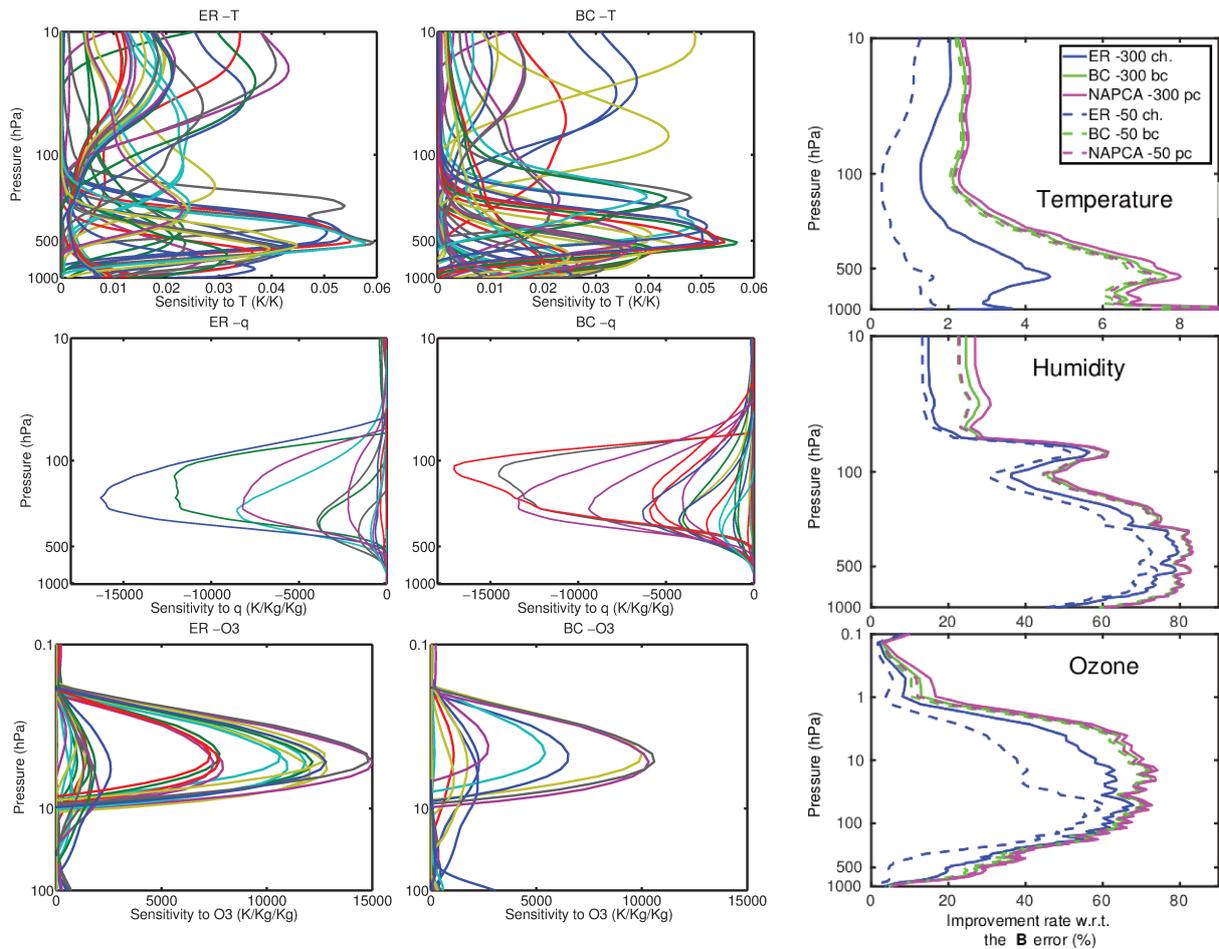


Figure II.25. Jacobien des canaux (colonne à gauche) ou des BCs (colonne au milieu) et amélioration de la restitution, par rapport à l'a priori pour les différentes approches (colonne de droite), en température (en haut), humidité (au milieu) et ozone (en bas).

en va de même pour la température : plus de 17 CO_2 sont utilisés puisque les canaux q sont également sensibles à T , mais ce n'est pas suffisant car l'erreur T a priori est assez faible (1 K) et des informations supplémentaires sont nécessaires pour l'améliorer. Notez que pour 300 canaux, la méthode ER obtient un bon contenu d'information de l'observation principale.

5.3 Évaluation avec des données réelles

Les méthodes précédentes sont testées sur des observations réelles. Pour ce faire, nous nous sommes appuyés sur l'ensemble de données LMD_ICO (IASI Composite Observatory) généré au Laboratoire de Météorologie Dynamique. Cet ensemble de données se compose de températures de brillance réelles mesurées par IASI, mises en coïncidence avec des mesures radiosondes pour des situations atmosphériques recouvrant le monde entier. Le jeu de données commence en juillet 2007 (lancement de MetOpA) et est mis à jour mensuellement. L'information sur les radiosondes provient de la base de données Analyzed RadioSoundings Archive (ARSA) fournie par

le groupe d'Analyse des Rayonnements Atmosphériques³ du Laboratoire de Météorologie Dynamique. ARSA est basée sur les observations de stations réparties dans le monde entier. La validation des données ARSA a été effectuée à l'aide de nombreuses comparaisons entre les températures de brillance IASI simulées et observées. Les observations de radiosondes fournissent, sur 43 niveaux atmosphériques, les profils de la température, l'humidité spécifique et l'ozone. La base de données ARSA actuelle (environ 6 millions d'éléments) commence en janvier 1979, et s'étend, sur une base mensuelle.

L'ensemble de données LMD_ICO utilisé dans l'étude comprend également le flag de couverture nuageuse, la masse d'air et les caractéristiques de visée de l'instrument IASI pour chaque pixel. La coïncidence est effectuée en utilisant une différence de temps maximale de 30 min et une distance spatiale de moins de 100 km. L'étude est menée sur une année de données (2014) avec un nombre total de coïncidences égal à 49 109. Deux critères ont été utilisés pour filtrer les données : une couverture nuageuse nulle, et la géométrie du nadir a été sélectionnée en utilisant un angle d'incidence compris entre -10° et $+10^\circ$. Les collocations spatio-temporelles entre les mesures de IASI et les mesures de radiosondage introduisent des incertitudes dans la base de données. Le même schéma d'inversion via un réseau de neurones est testé pour les quatre types d'entrées :

- la sélection de canaux ER composée de pur- T_s , pur- T , pur- q , pur- O_3 . Les canaux étant hiérarchisés dans les sélections, on peut tester l'augmentation du nombre de canaux et son impact dans l'inversion (30 ou 270 dans le test) ;
- NAPCA avec 30 ou 270 PCs ;
- l'approche BC-projeté/lin/contraint avec 30 ou 270 BCs
- la sélection de 270 canaux sur IASI utilisée en opérationnel (notée COL pour Collard (2007) dans la suite).

Pour les méthodes de compression NAPCA et BC-projeté/lin/contraint, les PCs et BCs sont obtenus grâce aux matrices de projection issues des données simulées. La base de données LMD_ICO (TB/radiosondes) est utilisée pour entraîner et tester un modèle de restitution statistique basé sur un réseau de neurone (NN) avec 30 ou 270 entrées. La couche cachée est choisie avec 45 neurones ; ce nombre a été testé pour être assez grand pour les deux configurations d'entrée. Le nombre de sorties est de 43 correspondant aux 43 niveaux atmosphériques de température ou d'humidité.

L'entraînement du NN s'effectue à l'aide d'un algorithme d'optimisation : la rétro-propagation de la descente de gradient.

La Figure II.26 représente l'écart-type de l'erreur d'inversion NN de la température (à gauche) et de l'humidité (à droite). Des statistiques sont fournies pour la compression NAPCA, la technique BC, et les sélections de canaux ER et COL, lorsqu'on utilise 30 (en pointillés) et 270 (continu) des éléments d'information. Notez que la restitution basée sur NN n'utilise aucune information *a priori* sur l'état de l'atmosphère (souvent inférieure à 1 K pour la température). De plus, les statistiques montrés dans la Figure II.26 incluent les incertitudes dues aux erreurs de collocation spatio-temporelle. Cela conduit à des performances de restitution plus faibles que ce qui pourrait être obtenu théoriquement.

Pour la restitution de T (Figure II.26-Gauche), les NAPCA et BC ont des statistiques similaires en comparant les configurations 30 et 270 entrées : les deux sont des

3. <http://ara.abct.lmd.polytechnique.fr/index.php?page=arsa>

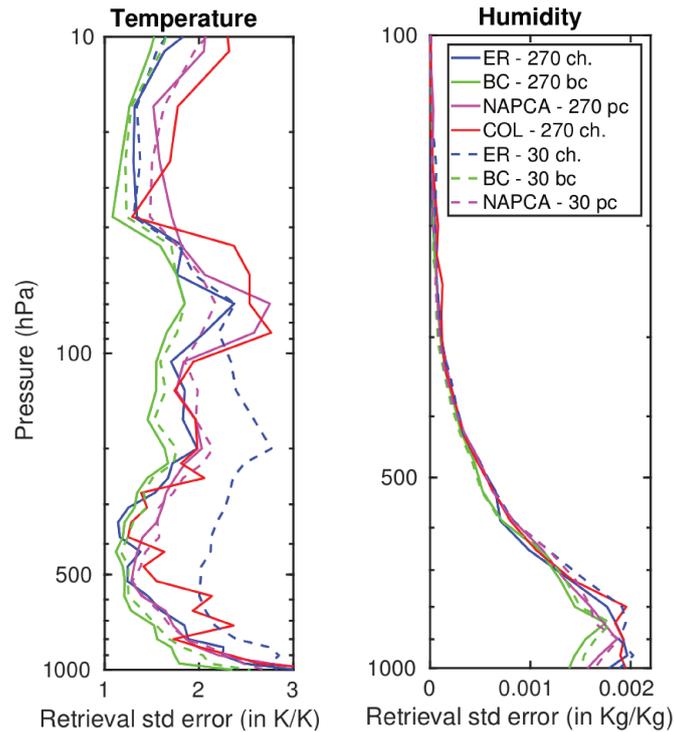


Figure II.26. Déviation standard (STD) de l'erreur de restitution NN de la température (à gauche) et de l'humidité (à droite), pour le NAPCA, le ER, la sélection Col-lard (2007) et BC-projeté/lin/contraints, avec 30 (lignes pointillées) et 270 composants/canaux (lignes continues).

techniques de compression, de sorte que les informations de nombreux canaux sont déjà composées dans les 30 premières informations. Dans certaines couches (supérieures à 100 hPa) pour la température, l'utilisation de plus de composants peut même être préjudiciable. Pour les techniques de sélection de canaux (ER, COL), l'augmentation du nombre de canaux a un impact important en raison de l'exploitation de la redondance des canaux pour réduire l'impact du bruit d'observation. Les performances des deux sélections ER et du COL sont similaires, mais ER surpasse COL pour les couches atmosphériques supérieures. L'approche BC semble être meilleure que les autres approches lorsqu'elle utilise 30 ou 270 entrées, pour presque tous les niveaux atmosphériques.

La Figure II.26 (à droite) représente une statistique similaire pour la restitution de l'humidité q . Encore une fois, les techniques de compression telles que NAPCA ou BC ont de plus faibles statistiques d'erreur puisqu'ils exploitent mieux la redondance des canaux. Les BCs surpassent la NAPCA pour la restitution de q sur toutes les couches atmosphériques. La Figure II.26 confirme que les BCs sont un bon compromis entre la compression, qui utilise la redondance de canal, et la sélection de canal pour préserver le canal.

Une signification physique de la réduction des données facilite l'inversion, non seulement théorique sur des données simulées (voir le contenu en information) mais aussi une inversion opérationnelle (inversion statistique avec NN) avec des données réelles (TB réelles et profils atmosphériques réels à partir de radiosondes).

La Table II.6 représente les statistiques : moyenne, écart-type et l'erreur quadratique moyenne ("Root Mean Square Error" RMS en anglais) pour T et q , lorsqu'on fait

		ER	BC	NAPCA	COL
T (in K)	Mean	0.1	-0.005	-0.2	-0.1
	STD	1.7	1.5	1.8	1.9
	RMS	1.8	1.5	1.9	2.0
q (10^{-4} Kg/Kg)	Mean	0.3	-0.1	-0.2	-0.6
	STD	4.8	4.2	4.7	5.0
	RMS	4.9	4.2	4.6	5.0

TABLE II.6 – Statistiques de restitution(s) de la température T et l'humidité q , pour les ER, BCs-constraints, NAPCA et COL (Collard, 2007), lorsqu'on utilise 270 composants/canaux. La moyenne, l'écart-type et l'erreur quadratique moyenne ("Root Mean Square Error" RMS, en anglais) sont fournis.

la moyenne de la surface à 10 hPa T et de la surface à 500 hPa pour q . Des statistiques sont fournies pour la sélection ER, la technique BC, la compression NAPCA et la sélection du canal Collard, lorsqu'on utilise 270 éléments d'information. Notez que dans cette restitution NN, aucune information n'est fournie *a priori*, contrairement aux méthodes itératives ou d'assimilation, donc les scores sont inférieurs à ce qui pourrait être obtenu autrement. Mais cette approche nous permet de comparer de manière robuste l'information fournie par chaque méthode de réduction des données pour la restitution, indépendamment de toute source d'information externe. Les statistiques de la Table II.6 confirment d'abord que toutes les extractions NN sont non biaisées, c'est généralement le cas pour les extractions statistiques. De plus, les BCs obtiennent de meilleures statistiques de restitution que les autres techniques de réduction.

6 Conclusion du chapitre

Dans ce chapitre, nous avons introduit les notions de sondeurs hyperspectraux dans le contexte du sondage météorologique par télédétection. Les algorithmes d'inversion qui permettent l'estimation de variables atmosphériques telles que le profil de température ou d'humidité ne peuvent pas gérer l'ensemble des observations disponibles. Dès lors des techniques de réduction de dimension doivent être développées. Après avoir introduit deux méthodes classiques de réduction de dimension, nous avons présenté une nouvelle approche de représentation, les "Bottleneck Channels", et illustré les avantages des techniques classiques de compression et de sélection des canaux avec ceux de la nouvelle approche, d'abord pour un concept d'instrument en micro-onde puis pour l'instrument IASI. La Table II.7 synthétise les propriétés, avantages et limites des différentes méthodes présentées dans ce chapitre.

Plusieurs points sont à retenir. (1) Le NAPCA est plus efficace que la PCA sans coût de calcul supplémentaire réel. (2) Il a été démontré que la sélection du canal prenant en compte deux variables à restituer de façon combinée n'est pas plus performante que les sélections indépendantes de multiples variables lorsque le nombre de canaux est limité. (3) À dimension égale, les approches de compression sont plus efficaces : elles utilisent, dès les premières pièces d'information (PC ou BC), l'ensemble du vecteur d'observation et exploitent la redondance pour réduire le bruit d'observation. (4) Dans ce contexte, les BCs semblent être un bon compromis entre la compression et la sélection du canal. Les performances des BCs sont proches des PCs mais la si-

gnification physique des BCs, représentant des pseudo-canaux, peut faciliter leur(s) utilisation(s) dans des schémas d'inversion ou d'assimilation.

Problème de mélange - Les composantes PCA mélangent diverses informations. Dans la Figure II.27, la corrélation entre les 50 premières composantes issues du spectre d'IASI et les profils T , q et O_3 est représentée. On peut voir que toutes les composantes ont une information sur les constituants T , q et O_3 . Par conséquent, il est impossible d'utiliser les composantes NAPCA standards pour un seul composant atmosphérique, ce qui peut être utile par exemple dans les schémas de d'inversion indépendante où la température est d'abord récupérée, puis l'humidité et l'ozone. Au contraire, il est conceptuellement plus facile avec les BCs de sélectionner des canaux qui sont liés à un seul constituant atmosphérique. Par exemple, les BCs peuvent se concentrer sur la restitution de l'ozone. Comme mentionné précédemment, les BCs peuvent être limités aux canaux d'échantillonnage seulement dans la bande d'absorption $1000-1100\text{ cm}^{-1} O_3$. En outre, les statistiques de dé-bruitage ($y - \bar{y}$) des BCs peuvent être aussi bons que ceux de la NAPCA.

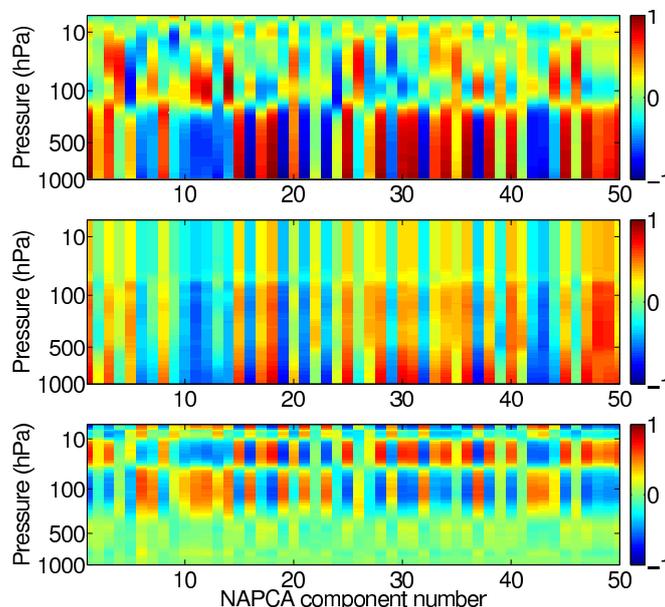


Figure II.27. Corrélation entre les 50 premières composantes NAPCA issues du spectre d'IASI et les profils T (en haut), q (au milieu) et O_3 (en bas).

Contamination - Nous avons établi que la NAPCA mélange plusieurs constituants en un seul composant. Lorsqu'on utilise ces composants dans une inversion, il est difficile de ne pas tenir compte de tous ces constituants. Les restitutions individuelles et indépendantes devraient être évitées. Nous définissons les "effets de contamination" comme les difficultés survenant lorsque les variations d'autres variables (non prises en compte/ non restituées) impactent la restitution de la variable considérée. Par rapport aux sélections de canaux qui peuvent se concentrer sur des domaines spectraux spécifiques, cette difficulté est apparue lors de l'utilisation de composantes NAPCA pour l'inversion. Par exemple, si la présence de nuages affecte une région spectrale, ils auront un impact sur la plupart des composantes de NAPCA, sauf lorsque ces composantes ont été calculées dans des régions spectrales sans nuages. Les BCs-projetés/lin peuvent offrir une solution plus facile et pratique. Ils sont plus localisés spectralement que les composantes de NAPCA parce qu'ils essaient de reproduire

des canaux physiques. Par conséquent, certains BCs seront contaminés par la présence de nuages, mais d'autres ne le seront pas. En comparant les TB simulés à l'aide d'un modèle de transfert radiatif et les TB des observations, il est possible de détecter les canaux qui seront contaminés par les nuages (procédure établie pour détecter les nuages dans les centres opérationnels PNT) et de les supprimer dans le schéma d'inversion ou d'assimilation. En revanche, avec NAPCA, si une composante est fortement contaminée, il est difficile de la supprimer sans supprimer toutes les composantes d'ordre inférieur : puisque chaque composante est calculée sur les anomalies laissées par les composantes antérieures, l'erreur se propagera en cascade sur les composantes suivantes.

Constituants mineurs - Les bandes d'absorption étroites pour les gaz à l'état de traces peuvent être négligeables en termes de variabilité statistique par rapport aux constituants majeurs tels que CO_2 pour la température ou H_2O . Par conséquent, l'information sur les gaz à l'état de traces peut être perdue ou mélangée et disséminée dans des composantes d'ordre supérieur de la NAPCA. Avec la stratégie BC, même les canaux isolés peuvent être choisis par le processus de sélection car le critère de qualité BC est de réduire uniformément les erreurs dans tous les canaux du spectre. Si cela n'est pas suffisant, il est facile de choisir *a priori* des canaux particuliers lors de la phase de sélection des canaux de l'algorithme BC. Cette flexibilité du BC permet d'introduire automatiquement des informations pour restituer les constituants majeurs de l'atmosphère, et de sélectionner manuellement des canaux dédiés aux gaz à l'état de traces.

Calibration - Lors de l'utilisation d'observations satellites dans l'assimilation ou dans un algorithme de restitution, les simulations de transfert radiatif nécessaires peuvent avoir des biais systématiques par rapport aux observations réelles. Une calibration doit être mise en place afin de corriger les biais. Avec l'utilisation de composantes NAPCA, la calibration implique toutes les TB utilisées pour calculer les composantes et ne peut se faire dans l'espace des composantes. Avec les BCs, l'étalonnage peut se faire seulement sur un petit nombre de canaux, seul un nombre limité d'informations est alors nécessaire et la calibration est plus facile à réaliser.

Accélération de l'inversion ou de l'assimilation - Des simulations de transfert radiatif doivent être effectuées pour la plupart des techniques d'inversion ou d'assimilation. Pour les approches NAPCA et BC, la matrice de projection peut être obtenue une fois pour toutes et utilisée directement pour l'assimilation ou l'inversion. Avec les composantes NAPCA, ces simulations doivent être effectuées sur l'ensemble du domaine spectral utilisé pour calculer les composantes. Un transfert radiatif dédié travaillant dans l'espace des composantes peut être développé. Toutefois, cette contrainte n'est pas nécessaire lors de l'utilisation des BCs, puisque les simulations de transfert radiatif ne peuvent être effectuées que dans le nombre limité de canaux sélectionnés. Ils peuvent être simulés, individuellement, par un code de transfert radiatif pour des fréquences particulières (comme le modèle OSS). Des simulations de RT monochromatiques sont effectuées sur des "noeuds" (les "pseudo-canaux") qui sont ensuite utilisés pour reconstruire l'ensemble du spectre. Ceci peut être appliqué dans le cadre des BCs pour accélérer le calcul et faciliter le schéma d'inversion ou d'assimilation.

Ces méthodes de réduction de dimension des observations hyperspectrales sont incontournables pour gérer le volume de données de ces observations de plus en plus important. Obtenir une représentation pertinente de l'ensemble de l'observation mais

	Nom	Description			Propriétés			
		Domaine d'application	Utilisation du bruit	orienté restitution	taux de compression	De-bruitage	signification physique	Localisation spectrale
Compression	PCA	Global	Non	Non	Bon	Oui	Non	Non
	NAPCA	Global	Oui	Non	Bon	Oui	Non	Non
	BC-projeté	Global	Oui	Non	Bon	Oui	Oui	Oui
Sélection de canaux	BC-strict	qq canaux	Oui	Non	Moyen	Non	Oui	Oui
	MRE	qq canaux	Oui	Oui	Moyen	Non	Oui	Oui
	DFS	qq canaux	Oui	Oui	Moyen	Non	Oui	Oui
	ER	qq canaux	Oui	Oui	Moyen	Non	Oui	Oui

TABLE II.7 – Comparaison des propriétés des techniques de compression, BC et de sélection de canaux.

de façon compressée est la première étape dans d'optimisation de l'exploitation des observations satellites. Une mauvaise réduction générera des erreurs qui se propageront jusqu'aux produits dérivés satellites et seront alors source d'incertitude dans l'estimation satellite de variables géophysiques complexes comme la précipitation. La genèse des produits dérivés satellites pour l'estimation des composantes du cycle de l'eau est l'objet du chapitre suivant. Nous nous interrogerons sur l'origine instrumentale et les algorithmes de restitution des multiples produits estimant les différentes composantes du cycle de l'eau.

Chapitre III

Bases de données pour l'étude du cycle de l'eau

Où l'on introduit les produits dérivés satellites que l'on utilisera par suite pour décrire le cycle de l'eau en Méditerranée en insistant sur les instruments sondeurs ou imageurs, infra-rouge ou micro-onde sur lesquelles ils reposent (voir la Table III.1). Et où l'on s'aperçoit que, outre les observations satellites, les produits dérivés incorporent aussi des informations annexes au sein de leurs algorithmes de restitution. Ces produits satellites sont couramment utilisés pour l'étude du cycle de l'eau.

Termes du bilan en eau

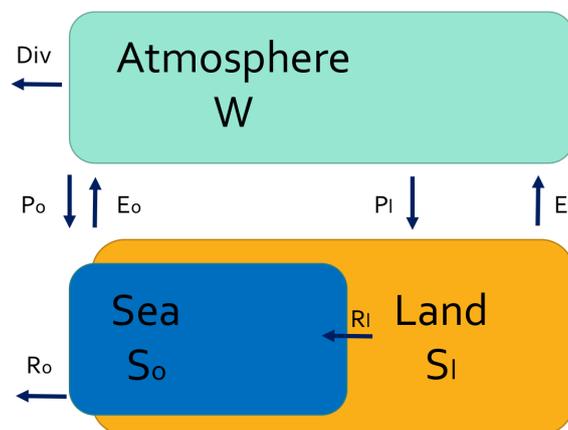


Figure III.1. Diagramme schématique du cycle de l'eau méditerranéen.

Les principales composantes du cycle de l'eau en Méditerranée et leurs interactions à travers le bilan hydrique sont représentées dans la Figure III.1. Les principaux échanges d'eau entre la surface et l'atmosphère sont l'évapo(transpi)ration et les précipitations. Le cycle de l'eau continental (qui a lieu au niveau de la surface continentale) est lié au budget océanique par le biais du ruissellement (égalité dans les échanges d'eau douce). Les termes de stockage dans la terre (S_I), la mer (S_o) et l'atmosphère (W) apparaissent également dans le diagramme.

D'après cette représentation schématique, le cycle de l'eau peut être décrit à l'aide des trois équations du bilan hydrique variant dans le temps :

$$\begin{aligned}
\frac{\delta S_l}{\delta t} &= P_l - E_l - R_l \text{ (Contiental)} \\
\frac{\delta S_o}{\delta t} &= P_o - E_o + R_l^* - R_o \text{ (Oc anique)} \\
\frac{\delta W_{l/o}}{\delta t} &= E_{l/o} - P_{l/o} - \text{div}(\overrightarrow{Q_{l/o}}), \text{ (Atmosph rique)}
\end{aligned}
\tag{III.1}$$

o  l'indice l est pour la partie continentale et o oc anique, P est la pr cipitation ; E le taux d' vapo(transpi)ration ; S le contenu en eau ; R_l ou R_l^* l' coulement de la terre vers l'oc an (avec $Area_{Med.Sea} \cdot R_l^* = Area_{Land} \cdot R_l$) ; R_o l' coulement net   travers le d troit de Gibraltar ; W le contenu total en vapeur d'eau atmosph rique ; et $\text{div}(\overrightarrow{Q_{l/o}})$ la divergence du flux d'humidit  atmosph rique int gr e verticalement (par la suite, cette composante sera plus simplement nomm  divergence de l'humidit  atmosph rique).

1 Les observations satellites

Dans cette section, les instruments satellites sondeur, imageur ou radar sont d sign s par le couple : "Satellite//Capteur" (e.g. METOP-A//IASI-A). Il faut diff rencier la *mesure* faite par un instrument (e.g. radiance) des *produits d riv s* satellites (comme les pr cipitations) qui n cessitent des algorithmes d'inversion(s) (mod les, hypoth ses) ainsi que, souvent, des sources d'informations annexes. De nombreux algorithmes ont  t  construits par les diff rents centres m t orologiques. Cependant, la fiabilit  des produits reste relativement variable d'un produit   l'autre. Dans la suite de cette section, je pr senterai les diff rents produits satellites en d crivant les mesures instrumentales utilis es par les algorithmes correspondants.

Dataset	Couverture temp.	Résolution spat. (°)	Résolution Temp.	Description	Producteur	Référence
Précipitation						
GPCP	1979-2015	2.5	daily	multiple satellites et jauges	U. of Maryland	Adler et al. (2003)
CMORPH	1998-2015	0.25	30min	micron-ondes et infrarouge	NOAA	Joyce et al. (2004)
TMPA	1998-2015	0.25	3h	multiple satellites et jauges	NASA	Huffman et al. (2007)
PERSIANN	2000-2013	0.25	3h	micron-ondes et infrarouge	CHRS	Ashouri et al. (2015)
ERA-I Precipitation	1980-2015	0.25	12h	ré-analyse	ECMWF	Dee et al. (2011)
EOBS Precipitation	1950-2006	0.25	1j	<i>in situ</i> griddé	project ENSEMBLES	Haylock et al. (2008)
FLUXnet precipitation	2002-2010	-	1mois	<i>in situ</i>	FLUXnet	Falge et al. (2017)
Évapotranspiration						
GLEAM	1980-2012	0.25	1j	observations satellites, jauges and ré-analyse	U. of Amsterdam and U. of Ghent	Martens et al. (2016)
MOD16	2000-2012	0.25	8j	observations satellites	NTSG	Mu et al. (2011)
NTSG	1983-2012	0.25	1m	observations satellites et ré-analyse	NSTG	Zhang et al. (2010)
ERA-I evapotranspiration	1980-2015	0.25	12h	ré-analyse	ECMWF	Dee et al. (2011)
FLUXnet evapotranspiration	2002-2010	-	1m	<i>in situ</i>	FLUXnet	Falge et al. (2017)
Évaporation						
Oaflux	1985-2015	1	1j	observations satellites et ré-analyse	WHOI	Sun et al. (2003)
Seaflux	1998-2015	0.25	3h	observations satellite, ré-analyse et <i>in situ</i>	GEWEX	Curry et al. (2004)
ERA-I Evaporation	1980-2015	0.25	6h	ré-analyse	ECMWF	Dee et al. (2011)
Contenu en eau						
CSR	2002-2012	0.25	1mois	GRACE	CSR	Bettadpur (2012)
GFZ	2002-2012	0.25	1mois	GRACE	GFZ	Dahle et al. (2013)
GRGS (land only)	2002-2012	0.25	1moi	GRACE	CNES	Biancale et al. (2005)
JPL	2002-2012	0.25	1mois	GRACE	JPL	Watkins et Yuan (2014)
Eau précipitable						
Globalvapor	1996-2015	0.5	1j	observations satellites	DWD, GEWEX	Schneider et al. (2013)
ERA-I Water vapor	1979-2015	0.25	6h	ré-analyse	ECMWF	Dee et al. (2011)
Ruissellement						
CEFREM	1980-2009	< 0.25	1an	<i>in situ</i>	Cefrem	Ludwig et al. (2009)
ORCHIDEE	1980-2009	0.5	1mois	modèle	LMD	Polcher et al. (1998)
Divergence d'humidité						
ERA-I Moisture divergence	1979-2015	0.25	6h	ré-analyse	ECMWF	Dee et al. (2011)
Écoulement à Gibraltar						
IMEDEA- netflow	2004-2010	-	1mois	<i>in situ</i> & modèle	IMEDEA	Jordà et al. (2016)

TABLE III.1 – Aperçu des différents produits utilisés dans cette étude. Leur période de couverture temporelle commune, sur laquelle le bilan du cycle de l'eau est estimé, est 2004-2009.

des nuages et donc de la quantité et de la taille des hydrométéores qui s'y trouvent : aérosols, gouttelettes d'eau et cristaux de glace. Une relation entre les nuages épais et la précipitation est établie. De même, les estimations des précipitations par IR sont basées sur les mesures de température de brillance au sommet des nuages. Un certain nombre d'algorithmes pour le visible ou l'infrarouge ont été développés, citons l'algorithme de Geostationary Operational Environmental Satellites (GOES) Precipitation Index (GPI) (Arkin et Meisner, 1987) et l'algorithme de précipitations multispectral GOES (Ba et Gruber, 2001). Cependant la relation entre la température du sommet des nuages et les précipitations est indirecte : la température au sommet d'un nuage ne représente pas forcément une faible pluie plus bas dans la colonne d'atmosphère (Sapiano et al., 2009) et au contraire, tous les nuages convectifs ne forment pas des précipitations. Néanmoins, les observations dans ce domaine de longueur d'onde ont l'avantage d'avoir une bonne résolution spatiale ainsi une fréquence d'échantillonnage élevée car les imageurs sont typiquement installés sur des satellites géostationnaires (GEO).

1.1.b Méthodes issues des micro-ondes passifs

Le sondage passif dans les micro-ondes est une méthode plus directe de mesure des précipitations car les particules d'eau absorbent le rayonnement à ces fréquences. Le domaine des micro-ondes correspondant à la gamme de fréquence comprises entre 1 GHz et 300 GHz (e.g. longueurs d'ondes entre 30 cm et 1 mm , voir Figure II.3). Dans le domaine des micro-ondes, les interactions entre le rayonnement et les nuages sont marquées par l'absorption et la diffusion des hydrométéores au sein du nuage. En présence de précipitations, l'absorption par les gouttelettes d'eau est importante aux basses fréquences [19 et 36,5 GHz], alors qu'on observe la diffusion par les cristaux de glace aux fréquences élevées (86 GHz). Contrairement aux imageurs dans le visible ou l'infrarouge, les sondeurs micro-ondes peuvent sonder les nuages et distinguer différents types de pluie. Bien que les méthodes basées sur les micro-ondes permettent de faire des estimations plus réalistes des précipitations instantanées que par IR, les sondeurs sont généralement à bord de satellites défilant, ce qui conduit à un échantillonnage temporel relativement faible (environ 2 fois par jour) par rapport aux instruments IR basés sur les GEO (30 minutes ou moins). De plus les sondeurs ont une faible résolution spatiale (45 km au nadir pour l'instrument AMSU-A).

1.1.c Produits satellites dérivés de multi-sources

Les estimations satellites les plus précises proviennent d'une combinaison d'imageurs infrarouge sur des satellites GEO à haute fréquence d'échantillonnage et des sondeurs micro-ondes sur des satellites en orbite basse (LEO) à échantillonnage temporel moins fréquent (Huffman et al., 2007).

Cinq produits satellites multi-sources, utilisant cinq algorithmes différents, ont été sélectionnés dans notre étude. Deux d'entre eux sont des algorithmes présentant une calibration à l'aide de données de jauges *in situ* au sol : TMPA-3B42.V7 présenté dans (Huffman et al., 2007) et le Global Precipitation Climatology Project (GPCP, v2) présenté par Adler et al. (2003). Trois sont des produits "purement" satellites : Joyce et al.

(2004) ont développé la technique de morphing CPC de la NOAA (CMORPH, v1), Ashouri et al. (2015) ont développé l'estimation des précipitations à l'aide d'un réseau de neurones artificiels (PERSIANN, v1) et Turk et al. (2010) ont conçu le produit du Naval Research Laboratory (NRL). Dans cette étude, nous utilisons une combinaison d'ensemble de produits de précipitations calibrées et non-calibrées. Ce choix est motivé par la volonté de préserver les attributs spatiaux des données satellites originales pour les zones géographiques où la densité de jauges est limitée. La Figure III.3 illustre les estimations de précipitation des quatre produits satellites pour janvier 2004. On peut noter les différences de couverture spatiale, TMPA étant limité à 50°N ; les différences de résolution, GPCP ayant une résolution de 2.5°x2.5° alors que les trois autres produits sont à 0.25°x0.25° ; et surtout une différence de structures spatiales dans l'estimation entre les produits calibrés *in situ* TMPA et GPCP et les deux autres.

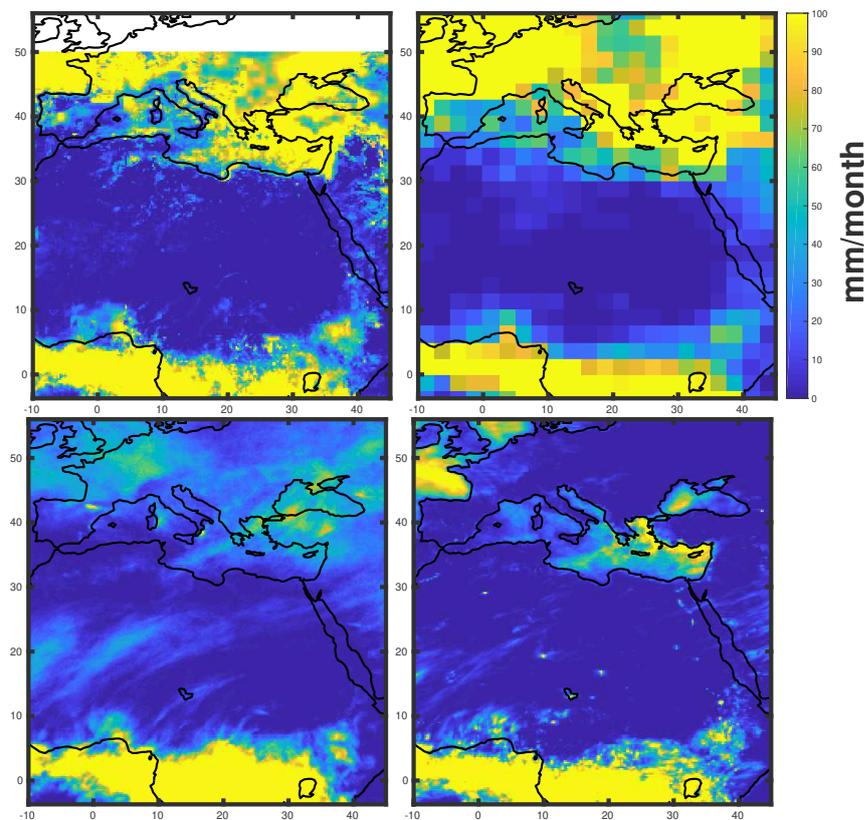


Figure III.3. Estimation de précipitation pour le mois de janvier 2004 des produits TMPA 3B42-V7 (haut, gauche) et GPCP-2 (haut, droite) PERSIANN (bas, gauche) et CMORPH (bas, droite). On peut remarquer la très grande disparité de ces estimations, issues d'algorithmes différents, même à l'échelle du mois. Cela fera l'objet d'étude et de commentaires au Chapitre IV.

1.1.c.1 TMPA 3B42.V7

L'algorithme TRMM (Tropical Rainfall Measuring Mission (Simpson et al., 1996)) Multi-satellite Precipitation Analysis (TMPA-3B42) est développé par la NASA.

Deux principaux produits de précipitation découlent de l'algorithme TMPA : (1) la version en quasi temps réel (3B42-RT) et (2) le produit de recherche (3B42-V7). Le

produit TMPA-3B42.V7 a été conçu pour intégrer les estimations de différents satellites avec les jauges au sol, pour une meilleure estimation issue de la mesure de l'imageur micro-onde TRMM/TMI. Huffman et al. (2007) décrivent la construction de TRMM-3B42. Dans un premier temps, l'algorithme Goddard profiling algorithm (GPROF) (Kummerow et al., 2001) est utilisé pour l'estimation des précipitations intégrant les données micro-onde des satellites LEO TRMM//TMI, AQUA//AMSR-E, DMSP//SSM-I et AMSU-B. Toutes les données micro-ondes ont été inter-calibrées avec des paramètres mensuels du TRMM//TMI Combined Instrument (TCI), qui intègrent également des informations provenant du radar micro-onde(s) de précipitation TRMM/PR. Ces données d'estimations micro-ondes calibrées sont fusionnées avec les données infrarouges issues des satellites GOES- W, GOES-E, GMS, METEOSAT-5, METEOSAT-7 et NOAA-12. Enfin, à la différence du produit TRMM-3B42.RT, les mesures des données au sol du Global Precipitation Climatology Centre (GPCC) sont incluses dans la calibration du produit TRMM-3B42.v7 (Huffman et al., 2007). Les produits en sortie de l'algorithme TRMM-3B42 ont une résolution temporelle de 3 heures avec des valeurs de taux de précipitations en mm.h^{-1} . La zone géographique couverte s'étend de la latitude 50° N à 50° S pour une grille de résolution spatiale de $0,25^\circ \times 0,25^\circ$. Ce produit est disponible depuis le 1^{er} janvier 1998 et jusqu'à nos jours.

1.1.c.2 GPCP-2

Le produit de Global Precipitation Climatology Project (GPCP-2, (Adler et al., 2003)) est le dataset, avec TMPA, le plus largement reconnu et utilisé par la communauté scientifique. La version 2 a été lancée en 2002. GPCP-2 incorpore les estimations de précipitations dérivées des imageurs et sondeurs micro-ondes DMSP//SSM-I et METop-B//AMSU-B dans la calibration de l'algorithme GPI de Arkin et Meisner (1987). Cet algorithme détermine les précipitations dérivées des mesures infrarouges issues des sondeurs Aqua//AIRS, TIROS//TOV et NOAA-15,-16,18. Ensuite, ces données satellites sont calibrées sur les mesures de jauges du Global Precipitation Climatology Centre (GPCC). GPCP-2 est distribué à une résolution 2.5° depuis 1980 et jusqu'à nos jours sur l'ensemble du globe. GPCP comme TMPA permettent une estimation efficace des pluie temporellement étendue. La calibration par les jauge de précipitation permettent une meilleur estimation des pluie fortes et de très courte durée. Alors que TMPA est limité au latitude $\pm 50^\circ$, GPCP possèdent une couverture plus vaste.

1.1.c.3 PERSIANN

L'algorithme Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network (PERSIANN) au Center for Hydrometeorology and Remote Sensing a été développé par Hsu et al. (1997). Il est basé sur une procédure de classification/approximation utilisant un réseau de neurones. Les données en entrée sont les températures de brillance des sondeurs infrarouges des instruments GOES-8, GOES-9/10, GOES-12, GMS-5, METEOSAT-6 et -7 fournies par la NOAA. L'information issue de l'imageur micro-ondes TRMM//TMI y est utilisée dans la phase d'apprentissage du réseau de neurones. Les taux de précipitation mesurés toutes les 30 minutes sont agrégés en précipitations cumulées sur 3 heures. Les données de sortie ont une résolution spatiale de $0,25^\circ \times 0,25^\circ$. La zone géographique couverte par l'algorithme s'étend de la latitude 60° N à 60° S et le produit est disponible depuis le 2 mars 2000

jusqu'à nos jours.

1.1.c.4 CMORPH

L'algorithme Climate Prediction Center MORPHING technique (CMORPH) a été développée à la NOAA (Joyce et al., 2004). Le produit satellite CMORPH utilise des estimations de précipitations obtenues à partir d'observations micro-ondes des instruments DMSP//SSM/I, la NOAA-POES//AMSU-B, AQUA//AMSR-E et TRMM//TMI. Les données infrarouges issues des satellites GOES 8, GOES 10, Meteosat-5, Meteosat-7 et GEO Meteorological Satellite-5 (GMS-5) sont utilisées pour propager les caractéristiques des précipitations entre deux mesures micro-ondes. Une interpolation linéaire pondérée dans le temps est utilisée pour modifier la forme et l'intensité des précipitations. Le produit CMORPH est disponible depuis le 3 décembre 2002 jusqu'à nos jours, à différentes résolutions spatiales et temporelles, avec une couverture de 60° N à 60° S

1.1.c.5 NRL

La technique NRL pour estimer la précipitation est générée par le Naval Research Laboratory (NRL). La technique dite "mixte" (blend) NRL produit des estimations de précipitations toutes les trois heures à une résolution de 0.25° entre 60°S et 60°N. L'algorithme NRL est basé sur une collocation temporelle et spatiale, et ce en temps réel, de température de brillance des imageurs dans le visible ou l'infrarouge présents sur les satellites GOES-11, GOES-12, Meteosat-7, Meteosat-9 (MSG-2), et GMS-6 (MTSAT-1R) avec la précipitation mesurée par les sondeurs micro-ondes NOAA//AMSU-B et NOAA//MHS METOP-B//MHS, DMSP//SSMIS et EOS//ASMR-E, des imageurs micro-onde DMSP//SSM/I et TRMM//TMI et du radar micro-onde TRMM//PR. Cette collocation est alors utilisée pour construire une table de correspondance (Look Up Table). Cette table est ensuite utilisée pour restituer le taux de précipitation à chaque nouvelle mesure infrarouge, et est mise à jour après le passage d'un instrument micro-onde.

1.2 Evapotranspiration

Bien que le terme "satellite" soit communément utilisé pour les produits en évaporation ou évapotranspiration décrits ci-dessous, aucun de ces deux processus n'est "observable" directement depuis les satellites. Des modèles d'interface sol-air et des informations annexes sont donc toujours nécessaires à l'estimation de ces variables.

1.2.a Évaporation sur mer

L'estimation de l'évaporation sur mer se fait à partir de paramétrisations Bulk des flux de chaleur latente et sensible décrites par Liu et al. (1979). Le taux d'évaporation peut ainsi être calculé à partir de la paramétrisation de ces deux flux qui nécessite comme entrées les variables observables : la température atmosphérique, la vitesse du vent près de la surface, la température de surface et l'humidité spécifique de l'air à la surface.

Deux produits d'évaporation utilisant la même paramétrisation Bulk mais avec différentes données en entrée ont été choisis : le Objectively Analyzed air-sea Fluxes for Global Oceans (Sun et al., 2003) ; et le SeaFlux du Global Energy and Water Cycle Exchanges (GEWEX) Project (Curry et al., 2004). La Figure III.4 montre l'estimation

Paramétrisation Bulk

Les flux de chaleur latente (LHF) et sensible (SHF) permettent le calcul du taux d'évaporation ($Evap$) (Liu et al., 1979) :

$$\begin{aligned}LHF &= \rho \cdot L_e \cdot c_e \cdot U \cdot (q_s - q_a), \\SHF &= \rho \cdot c_p \cdot c_h \cdot U (T_a - \theta), \\Evap &= \frac{LHF}{\rho \cdot L_e} \\&= c_e \cdot U \cdot (q_s - q_a),\end{aligned}\tag{III.2}$$

où ρ est la densité de l'air, L_e la chaleur latente de vaporisation (i.e. enthalpie), U est la vitesse du vent près de la surface à une altitude z , q_a est l'humidité spécifique de l'atmosphère près de la surface, q_s est l'humidité spécifique à saturation au niveau de la surface de la mer, à la température (T_s) et c_e et c_p sont des coefficients d'échange en régime turbulent et sont déterminés entre autres par la stabilité atmosphérique, les différences de température air-mer et la vitesse du vent (Fairall et al., 2003). On peut alors exprimer q_s en fonction de la température de surface T_s et l'humidité de saturation d'eau pure q_{sat} :

$$q_s = 0.98 \cdot q_{sat}(T_s),\tag{III.3}$$

où un facteur multiplicateur de 0,98 est utilisé pour la réduction de l'humidité spécifique à saturation causée par une salinité typique de 34 psu. D'autre part, θ inclue une correction γ pour le taux de réduction adiabatique :

$$\theta = T_a + \gamma \cdot z,\tag{III.4}$$

où T_a est la température à l'altitude z .

Il existe actuellement plusieurs formes d'algorithmes bulk (Brunke et al., 2002). Les différences entre algorithmes résident dans les hypothèses sur les conditions de vent léger et de stratification stable de l'atmosphère pour le traitement de l'état de la mer (houle, effets directionnels) dans le paramétrage des coefficients d'échange c_e et c_h .

Dans l'Eq. (III.2), U , T_s , T_a et q_a sont indépendants. Une fois que l'information de ces quatre variables observables est donnée, les termes de flux et d'évaporation peuvent être paramétrés.

de ces deux produits pour janvier 2004. On peut noter la différence de résolution entre OAFlux à 1° et SeaFlux à 0.5°. Par ailleurs les deux estimations semblent proches en terme de structure spatiale d'évaporation, avec un gradient Ouest-Est présent dans les deux produits. Je détaille par la suite ces deux produits.

1.2.a.1 OAFLUX

OAflux est un produit de flux de surface global et journalier à la résolution spatiale

de 1° et sur la période 1981-2015. La paramétrisation sous-jacente à ce produit utilise, pour les quatre variables d'entrée, non seulement des données d'observations satellites mais aussi des champs de sortie de la ré-analyse² de l'ECMWF : ERA-40 (Uppala et al., 2005) et ré-analyses de National Centers for Environmental Prediction (NCEP) (Kalnay et al., 1996a). Nous détaillons ci-après les sources satellites des variables d'entrée :

- Il y a quatre sources satellites sur la vitesse du vent U : le radar QuikSCAT//SeaWinds et trois instruments micro-ondes l'imageur DMSP//SSM-I (F16 et F17) et les sondeurs Aqua//AMSR-E et WindSat.
- La température de surface T_s est le produit d'analyse de la NOAA optimum interpolation produite par Reynolds et al. (2007). L'analyse a deux produits : l'un utilise les données infrarouges du sondeur à très haute résolution AVHRR et l'autre combine l'infrarouge AVHRR avec les données micro-ondes AMSR-E. Les deux produits utilisent également les données *in situ* des navires et des bouées de mesures et incluent un ajustement des biais satellites par rapport aux données *in situ*.
- L'humidité de l'air proche de la surface de la mer (q_a) est difficilement observable depuis les satellites. En revanche, l'eau précipitable³ au-dessus des océans peut plus facilement être estimée par inversion du transfert radiatif. Chou et al. (1995) utilisent l'information sur l'eau précipitable pour déduire l'humidité près de la surface. Sa méthode consiste à décomposer le signal de l'eau précipitable provenant de SSM/I au dessus des océans en utilisant la méthode d'analyse en composantes principales (voir chapitre I section 2.1) afin d'extraire a composante du signal liée à l'humidité spécifique. L'humidité de surface résultante a été validé en utilisant des radiosondes.
- La température de l'air proche de la surface de la mer (T_a) vient des produits de sortie de la ré-analyse.

Les données satellites/ré-analyse sont combinées linéairement afin d'obtenir le meilleur estimateur, pour chaque variable considérée, au sens des moindres carrés.

1.2.a.2 SeaFlux

SeaFlux est un produit de flux de surface global à la résolution spatiale de 1°, à la résolution temporelle de 3h et sur la période 1998-2007. Comme OAFlux, ce produit utilise, pour les quatre variables d'entrée, non seulement des données d'observation satellites mais aussi des champs de sortie de la ré-analyse. :

- Les principales données d'entrée du champ de vitesse du vent de surface sont fournies par l'analyse effectuée sur Cross-Calibrated Multi-Platform (CCMP) Ocean Surface Wind Components data (Atlas et al., 2011). Cette analyse utilise les données satellites issues des radiomètres ADEOS//NSCAT, TRMM//TMI, ADEOS-II//SeaWinds et du radar QuikSCAT//SeaWinds. L'analyse est produite toutes les six heures en utilisant les données de ré-analyse de l'ECMWF pour combler les jours ou heures manquant(e)s.
- Le produit de Température de surface de l'eau (T_s) est le même que celui utilisé dans OAflux (Reynolds et al., 2007) mais celui-ci est interpolé en utilisant

2. la ré-analyse fait l'objet d'une section de ce même Chapitre III.3

3. L'eau précipitable est la vapeur d'eau atmosphérique contenue dans toute la colonne verticale jusqu'au sommet de l'atmosphère.

l'information du rayonnement solaire journalier ainsi que le taux de précipitation moyen et la vitesse de vent moyen afin d'estimer la variabilité diurne (Clayson et al., 2013) nécessaire à SeaFlux. Le rayonnement solaire est le produit Surface Radiation Budget (SRB) de GEWEX et les précipitations proviennent du Global Precipitation Climatology Project (GPCP) (Huffman et al., 2001).

- Comme pour OAflux, l'estimation de l'humidité près de la surface (q_a) à partir de l'observation satellite dans les micro-ondes est issue de la forte corrélation entre l'humidité de surface et la vapeur d'eau totale intégrée (Liu et al., 1979). Cependant, cette corrélation, très forte à l'échelle du mois, s'estompe sur des échelles de temps plus courtes. Des algorithmes de type réseaux de neurones sont alors nécessaires pour estimer l'humidité près de la surface à la résolution de 3h (Roberts et al., 2010).
- La température de l'air proche de la surface de la mer (T_a) est aussi dérivée des observations micro-ondes. Roberts et al. (2010) montrent que l'utilisation de T_s dans l'*a priori* de restitution de la T_a contribue de façon significative à l'amélioration de l'estimation de T_a .

SeaFlux est très proche de OAflux (voir Figure III.4). Mais à la différence de la pondération linéaire entre sortie de ré-analyse et produits satellites dans OAflux, la procédure d'interpolation utilise les tendances physiques et dynamiques de la ré-analyse pour calibrer les produits satellites. De plus, les données de ré-analyse dans SeaFlux proviennent de Modern Era Retrospective-analysis for Research and Applications (MERRA) (Bosilovich et al., 2008).

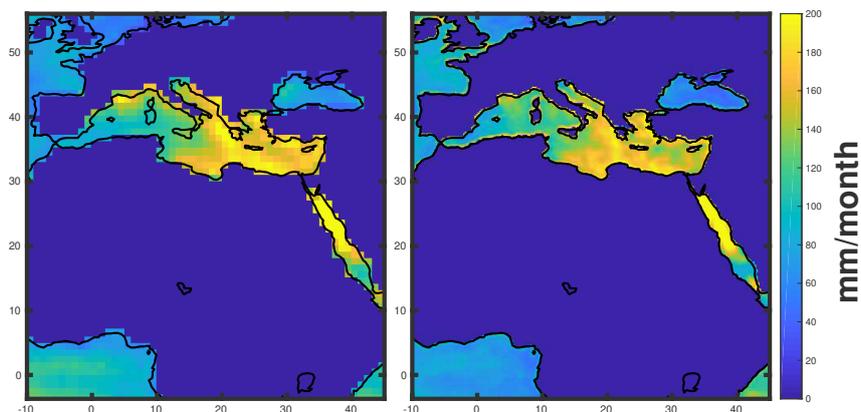


Figure III.4. Estimation du taux d'évaporation pour le mois de janvier 2004 des produits OAflux (à gauche) et SEAflux (à droite). À la différence des produits satellites de précipitation, les produits d'évaporation, incorporent des sortie de ré-analyse et une paramétrisation qui permet une certaine cohérence entre les deux estimations.

1.2.b Évapotranspiration continentale

L'évapotranspiration (ET) est un phénomène physique qui implique le sol, les plantes et l'air dans les conditions météorologiques environnantes. L'évapotranspiration n'est pas facile à mesurer mais comme pour l'évaporation sur mer, certaines paramétrisations ont été développées afin d'estimer l'évapotranspiration à partir des données de

surface et d'atmosphère. L'équation de Penman-Monteith (équation PM) est considérée comme le modèle standard et recommandée par la Food and Agriculture Organization of the United Nations (Allen et al., 1998). L'équation ne donne pas directement l'évapotranspiration "réelle" mais une évapotranspiration de "potentiel". Selon Allen et al. (1998), l'évapotranspiration des différentes cultures est liée à un taux d'évapotranspiration "potentiel" modulé par des coefficients propres à la culture et au sol, comme le facteur de stress.

Paramétrisation PM

Le taux d'évapotranspiration "potentiel" (ET_o) est donné par :

$$ET_o = 0.408 \cdot \delta \cdot (Rn - G) + \gamma \cdot \left(\frac{900}{T_{mean} + 273} \right) \cdot U_2 \cdot (es - e),$$

$$\delta = \lambda \cdot \frac{es}{Rv \cdot T_{mean}^2}, \quad (III.5)$$

$$\gamma = cp \cdot \frac{P}{\epsilon \cdot \lambda},$$

$$es = 2.53 \cdot 10^8 \cdot \exp\left(\frac{-5420}{T_{mean} + 273}\right),$$

$$e = UR \cdot es, \quad (III.6)$$

où ET_o est l'évapotranspiration de référence ($mm.jour^{-1}$); δ est la pente de la courbe de pression de vapeur (constante, $kPa.^{\circ}C^{-1}$); Rn est le rayonnement net à la surface de la culture ($MJ.m^{-2}.jour^{-1}$); G est le flux de chaleur dans le sol par conduction ($MJ.m^{-2}.jour^{-1}$); T_{mean} est la température moyenne quotidienne de l'air à 2m de hauteur ($^{\circ}C$); U_2 est la vitesse du vent à 2m de hauteur ($m.s^{-1}$); es est la pression de vapeur saturante (kPa); e est la pression de vapeur réelle (kPa); UR est l'humidité relative; $(es - e)$ est le déficit de pression de vapeur saturée (kPa); γ est la constante psychométrique ($kPa.^{\circ}C^{-1}$); λ est la chaleur latente de vaporisation de l'eau ($2450 kJ.m^{-2}.jour^{-1}$); cp est la chaleur spécifique à pression constante ($1.01 kJ.kg^{-1}.K^{-1}$); P est la pression atmosphérique (kPa); ϵ est le rapport poids moléculaire de la vapeur d'eau/air sec, qui est égal à 0.622. Cette paramétrisation nécessite en outre les valeurs des variables météorologiques suivantes : températures maximale et minimale de l'air, rayonnement solaire global et pression de vapeur réelle (Allen et al., 1998). Ces dernières peuvent être estimées à partir de la durée réelle de l'ensoleillement et de l'humidité relative, respectivement. Le premier terme de l'équation (III.5) est appelé le terme "rayonnement", tandis que le second terme, appelé "aérodynamique", dépend directement du déficit en vapeur (Allen et al., 1998).

Le rayonnement net est exprimé sous la forme :

$$Rn = (1 - A) \cdot R_s - [(0.1 + 0.9 \cdot n) \cdot (0.34 - 0.14 \cdot e^{0.5} \cdot \sigma \cdot (T_{mean} + 273)^4)] \quad (III.7)$$

où R_s est le rayonnement global à la surface ($MJ.m^{-2}.jour^{-1}$); A est l'albédo ou le coefficient de réflexion de la canopée; n est la durée réelle de l'ensoleillement (heures); σ est la constante de Stefan-Boltzmann, qui est égale à $5.67 \cdot 10^8 W.m^{-2}.K^{-4}$.

En raison de l'indisponibilité fréquente des variables météorologiques nécessaires aux calculs des variables et malgré le contexte théorique plus solide du modèle PM, de nombreux modèles de cultures utilisent des approches empiriques pour calculer l'évapotranspiration de référence. Priestley et Taylor (1972) (PT) ont proposé une équation empirique pour calculer l'évapotranspiration de référence.

Paramétrisation PT

Équation empirique de l'évapotranspiration de référence ET_o

$$ET_o = \alpha \cdot W \cdot (Rn - G), \quad (\text{III.8})$$

où W est l'indice qui dépend de la constante psychométrique γ ($kPa \cdot K^{-1}$) et α une constante de calibration.

Trois produits satellites ont été choisis pour l'évapotranspiration continentale. Ils diffèrent sur la paramétrisation de l'évapotranspiration et sur les données satellites en entrée. La Figure III.5 montre l'estimation de ces trois produits pour janvier 2004. Deux d'entre eux ne donnent pas d'estimation évapotranspiration pour la zone désertique du Sahara. Les structures spatiales de ces produits diffèrent particulièrement pour la région équatoriale influencée par le phénomène de mousson. Je détaille par la suite ces trois produits.

1.2.b.1 GLEAM-V3.a

Le Global Land Evaporation Amsterdam Model (GLEAM) (Martens et al., 2016 ; Miralles et al., 2011) est un ensemble d'algorithmes dédiés à l'estimation de l'évaporation journalière terrestre à partir de données satellites et de la ré-analyse à la résolution de 0.25° . Il utilise les données satellites à la fois comme forçage (i.e. en entrées) et en assimile d'autres (voir partie sur la ré-analyse). On détaille ci-après l'utilisation des observations satellites :

$$ET = ET_o \cdot S + E_i \quad (\text{III.9})$$

où S est le paramètre de stress dépendant du type de culture et E_i l'interception (évaporation sur la canopée humide).

GLEAM dérive séparément les différentes composantes de l'évaporation terrestre : (1) la transpiration, (2) l'évaporation du sol et des eaux libres, (3) l'interception de la canopée et la sublimation (Martens et al., 2016 ; Miralles et al., 2011). Le flux de vapeur est calculé pour chacune de ces composantes séparément, puis agrégé en fonction de chaque type de couverture terrestre. Premièrement, l'équation de Priestley et Taylor (1972) est utilisée pour calculer l'évaporation potentielle de la culture en fonction de la température de l'air (mesurée par l'instrument Aqua//AIRS) et le rayonnement net (mesuré par l'instrument Aqua//CERES). Puis l'évaporation potentielle est convertie en transpiration (ou évaporation du sol) à l'aide du facteur de stress S . Ce dernier est lié à l'épaisseur optique de la végétation mesurée dans les micro-ondes (SMOS), à la présence de neige (produit GLOBSNOW (Luus et al., 2013) issu de mesures en micro-ondes : SMMR, SSM-I et AMSR-E) et à un modèle d'infiltration forcé par les précipitations TMPA3B42-v7 et MSWEP (Beck et al., 2017) et pour lequel les observations

d'humidité du sol (SMOS) sont assimilées. Outre l'utilisation de ces données satellites, GLEAM-V3.a incorpore pour toutes les variables sus-citées des données issues de la ré-analyse (ERA-Interim) afin d'obtenir la longue couverture temporelle 1980-2016.

1.2.b.2 MOD16

Les spectro-radiomètres Terra//MODIS et Aqua//MODIS (MODerate Resolution Imaging Spectroradiometer) fournissent des informations sur la végétation et l'énergie à la surface. L'algorithme issu de MODIS Global Evapotranspiration Project (Mu et al., 2011, MOD16) estime l'évapotranspiration intégrée sur 8 jours à la surface de la Terre avec une résolution de 0.25° par la paramétrisation de PM. Cette paramétrisation, comme pour l'algorithme GLEAM, rend compte de l'interface surface-atmosphère dans les conditions climatiques particulières et liées aux caractéristiques du sol. Il inclut un certain nombre de modules comme l'évaporation du sol nu, l'interception de la canopée et la transpiration des arbres. Cet algorithme repose sur plusieurs entrées, la plupart sont des données météorologiques pour les variables atmosphériques et les produits satellites sont utilisés pour les caractéristiques de la surface. Ces produits sont issus de MODIS : couverture terrestre (land cover), albédo, fraction du rayonnement absorbé, indice de surface foliaire (LAI).

1.2.b.3 NTSG

L'algorithme de Numerical Terradynamic Simulation Group product (Zhang et al., 2010, NTSG) quantifie l'évaporation d'eau de surface à l'aide d'une approche PT et quantifie la transpiration de la canopée et l'évaporation du sol à l'aide d'une approche PM, la conductance de la canopée étant déterminée à partir de l'indice de végétation à différence normalisée (NDVI).

Dans cette étude, les données satellites d'entrées de l'algorithme comprennent l'indice NDVI, la couverture terrestre et les termes de rayonnement, tandis que les autres données quotidiennes de météorologie de surface ont été obtenues à partir de la ré-analyse (NCEP-NCAR).

Le produit NDVI est issu de deux produits : le Global Inventory Modeling and Mapping Studies (GIMMS) NDVI (basé sur AVHRR) à la résolution de 8 km et bimensuel (Tucker et al., 2005) et MODIS-IGBP à 1km.

Les termes de rayonnement solaire journalier net proviennent du produit de Surface Radiation Budget de GEWEX. Les données météorologiques quotidiennes (i.e. températures et pression de vapeur d'eau dans l'air) sont issues de la ré-analyse du National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) (Kalnay et al., 1996b).

1.3 Contenu en eau dans le sol par gravimétrie

Le contenu en eau dans le sol (Terrestrial Water storage, TWS) est l'eau stockée dans et sur les surfaces terrestres, comme l'humidité du sol, les lacs, les zones humides (marais), la glace, l'eau et la glace souterraine, et la neige. Les changements mensuels de ces masses d'eau stockées impactent le champ de gravité terrestre. Dès

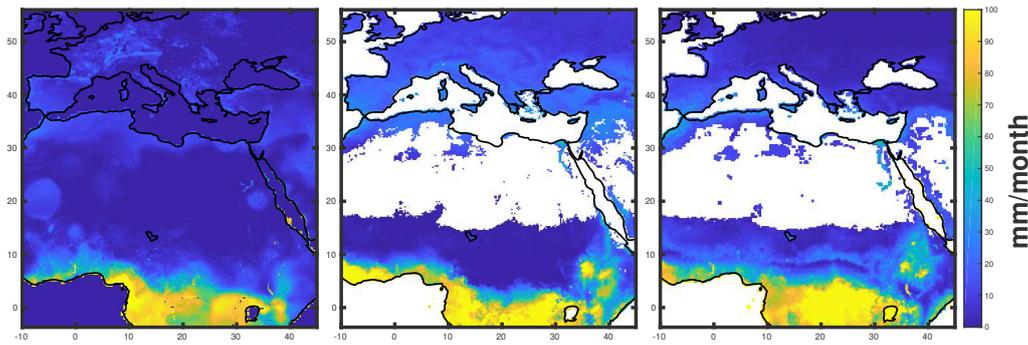


Figure III.5. Estimation du taux d'évapotranspiration pour le mois de janvier 2004 des produits GLEAM (à gauche), MOD16 (au centre) et NTSG (à droite). Deux d'entre eux ne donnent pas d'estimation évapotranspiration pour la zone désertique du Sahara. Outre ce constat pour cette zone désertique, on observe de forte différences entre le produit GLEAM et MOD16 au dessus des reliefs Alpains. La paramétrisation des différents produits au dessus des régions complexes (désert ou montagne) est problématiques ce qui met en évidence les limites de l'estimation satellite de l'évapotranspiration.

lors, il est possible de restituer cette information de stock ou plus précisément les variations grâce à la mesure satellite du champ de gravitation. Cela est possible depuis le lancement de l'instrument GRACE (Gravity Recovery and Climate Experiment) en 2002. GRACE est constitué de deux satellites qui se suivent à une distance de 220 km sur une même orbite, à 550 kilomètres d'altitude. Ils sont munis d'un instrument qui mesure les variations de la distance qui les sépare, avec une précision de dix micromètres. La Figure III.6 illustre schématiquement le principe de mesure par gravimétrie. Lorsque le satellite de tête arrive au-dessus d'une zone où la gravité croît, sa vitesse augmente et donc aussi sa distance au satellite qui le suit. De cette augmentation de distance (accélération) on déduit la variation correspondante du champ de gravité. Les satellites GRACE effectuent une couverture globale de la Terre en un mois. Outre les changements dans la gravité causés par la redistribution de masse dite "solide" (tremblement de terre, ou rebond post-glaciaire, également appelé ajustement isostatique), la plupart des variations mensuelles de la gravité sont en fait causées par des variations de stockage d'eau dans les réservoirs hydrologiques (e.g. pompage ou infiltration) et par le déplacement des masses de glace (e.g. fonte de glacier), d'eau océanique (e.g. courants marins profonds), atmosphérique (vapeur d'eau) et terrestre (e.g. crues, fonte de neige).

Les produits sur le contenu en eau sont tous dérivés de la mission GRACE : le produit du Jet Propulsion Laboratory (JPL, Watkins et Yuan, 2014)) ; celui du Centre for Space Research (CSR, Bettadpur, 2012), et du German Research Centre for Geoscience (GFZ, Dahle et al., 2013)) ; le produit issu du Groupe de Recherche de Géodésie Spatiale (GRGS, Biancale et al., 2005)). La mesure de GRACE étant mensuelle, tous ces produits donnent la variation du contenu en eau à l'échelle du mois et à la résolution spatiale de 1°. La Figure III.7 montre l'estimation du changement de contenu en eau entre décembre 2003 et janvier 2004 pour les quatre produits. Alors que trois produits sont très semblables, le quatrième, GRGS, diffère dans sa structure spatiale. En outre, il ne semble pas estimer de changement en masse d'eau pour le bassin du

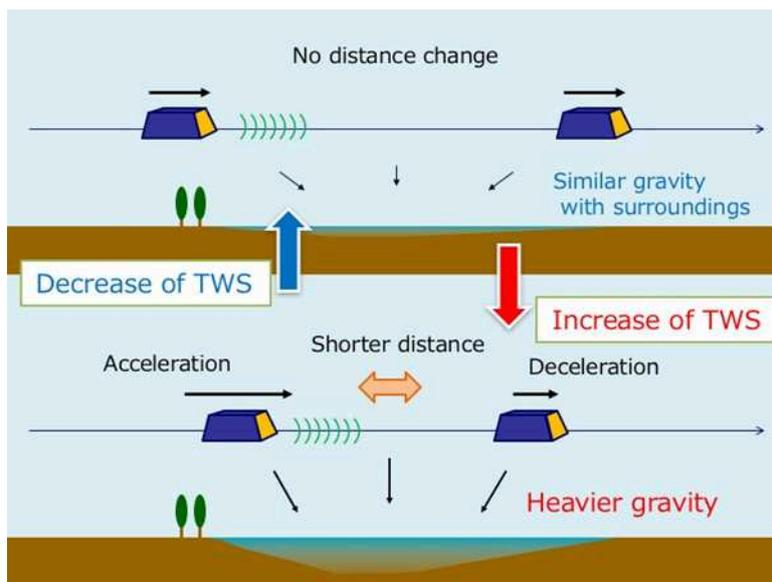


Figure III.6. Schéma explicatif de la mesure des variations de masse d'eau sur terre depuis les deux instruments GRACE (Japan Agency for Marine-Science and Technology, JAMSTEC; Nagoya University).

Nil.

1.3.a JPL, CSR et GFZ

Les trois centres CSR, GFZ et JPL font partie du système GRACE Ground System et génèrent les produits de contenu en eau en épaisseur équivalente d'eau à la surface (en *cm*). L'obtention des variations mensuelles du champ de gravité à partir des observations GRACE nécessite une inversion complexe des observations de la distance relative entre les deux engins spatiaux, en combinaison avec une détermination précise de l'orbite par GPS et diverses corrections pour les accélérations des satellites non liées aux changements de gravité. Les variations du contenu en eau s'obtiennent par décomposition du signal de champ de gravité à l'aide d'harmoniques sphériques (famille de fonctions formant une base dans l'espace des fonctions de carré intégrable sur la sphère qui permet la décomposition du signal). En outre, la masse de l'atmosphère est éliminée pendant le post-traitement à l'aide des champs de pression atmosphérique de l'ECMWF. Les produits CSR, GFZ et JPL ne contiennent donc pas de variabilité de masse atmosphérique. Les produits issus de GRACE Ground System traite indépendamment les surfaces continentales et les océans. Par souci de clarté, la Figure III.7 ne montre que les produit continentaux de GRACE Ground System. Un modèle océanique est utilisé pour éliminer les vents forts et les mouvements océaniques dus à la pression ou à la marée. Les champs de gravité résultant ne reflètent donc pas non plus la variabilité de l'océan mais seulement un apport local en eau. La différence entre les produits JPL, CSR, GFZ étant seulement liée à des choix de paramétrisation, leurs produits sont très proches et il est recommandé par GRACE Ground System de les utiliser en les fusionnant.

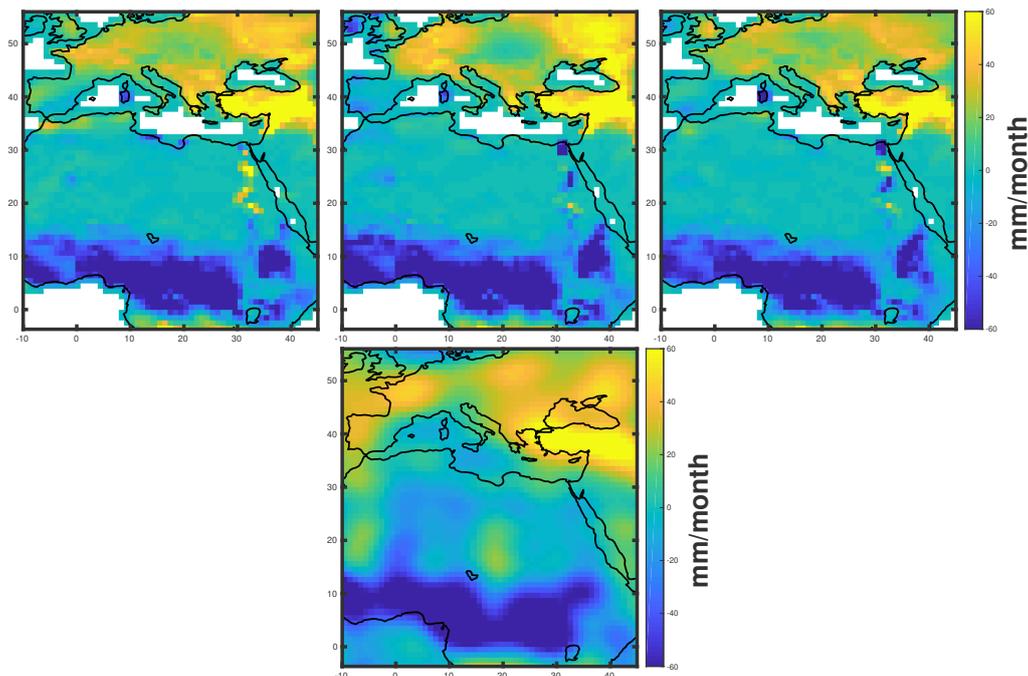


Figure III.7. Estimation des changements de masse d'eau entre décembre 2003 et janvier 2004 des produits "land-only" JPL (en haut, à gauche), CSR (en haut, au centre), GFZ (en haut, à droite) et du produit GRGS (en bas). Alors que trois produits sont très semblables, le quatrième, GRGS, diffère dans sa structure spatiale.

1.3.b GRGS

Le produit du GRGS propose lui aussi une estimation du contenu en eau à partir des mesures de GRACE selon une décomposition en harmoniques sphériques. Dans le CNES-GRGS, les variations gravitationnelles telles que les marées terrestres, les marées océaniques (FES2012, LEGOS), les champs de pression atmosphérique (ECMWF, ERA-Interim) et océanique (TUGO, LEGOS) ont été prises en compte dans la chaîne de traitement. Toutes ces corrections sont basées sur des sorties de la ré-analyse ou bien sur des modèles atmosphériques.

1.4 La vapeur d'eau atmosphérique intégrée

Seul un produit satellite est utilisé pour l'estimation de la vapeur d'eau intégrée verticalement (un autre produit issu de la ré-analyse sera présenté dans la section III.3).

Le produit satellite GlobVapour (Schneider et al., 2013) est dit "multi-sources", il combine des mesures issues de deux instruments selon le type de surface : DMSP//SSM-I (F13, F14) pour les surfaces océaniques et ENVISAT//MERIS pour les surfaces continentales. L'inversion du spectre SSM-I pour l'estimation de la vapeur d'eau verticalement intégrée se fait à partir des mesures de température de brillance via un schéma 1D-Var utilisant comme *a priori* des profils de la ré-analyse ERA-Interim et des données climatologiques qui sont colocalisés (spatialement et temporellement) avec les observations satellites. Pour obtenir le produit de vapeur d'eau MERIS, un réseau de neurones est d'abord appliqué pour filtrer les cas nuageux. La restitution en vapeur d'eau verticalement intégrée est ensuite basée sur une méthode d'optimisation itéra-

tive.

La vapeur d'eau de l'atmosphère est intégrée verticalement dans une colonne allant de la surface à 200 hPa et est exprimée en kg/m^2 . GlobVapour est un produit mensuel, à la résolution de 0.5° , sur la période 2003-2008. La Figure III.8 montre l'estimation du changement en vapeur d'eau atmosphérique entre décembre 2003 et janvier 2004. Pour ces mois, on observe une corruption du signal sur de nombreux pixels, particulièrement au-dessus de la région équatoriale impactée par la mousson. Cela est dû à la présence de nuages convectifs qui rend difficile la restitution de ENVISAT//MERIS.

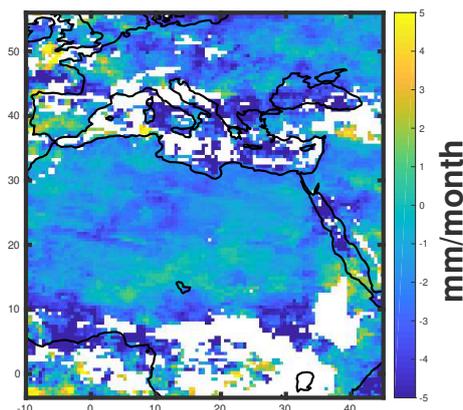


Figure III.8. Estimation des changements de masse d'eau atmosphérique entre décembre 2003 et janvier 2004 du produit GlobVapour. On observe pour ce mois particulier, une estimation très bruitée.

1.5 La hauteur d'eau par altimétrie

La hauteur d'eau par altimétrie sera utilisée pour la validation du contenu en eau sur mer. Nous utiliserons alors une estimation indépendante de la variation du contenu en eau de la Méditerranée. Les variations du contenu en eau peuvent être estimées (Fenoglio-Marc et al., 2006 ; Jordà et Gomis, 2006) comme étant le niveau total de la mer moins les variations thermiques (c'est-à-dire les variations du niveau de la mer dues à l'expansion/contraction thermique). Le niveau total de la mer est obtenu à partir du produit altimétrique Ssalto//Duacs produit et distribué par le Service de surveillance marine et environnementale de Copernicus (CMEMS⁴). Les variations thermiques du niveau de la mer sont estimées en utilisant une réanalyse régionale océanique MEDRYS (Hamon et al., 2016 ; Bahurel et al., 2012) et deux produits mondiaux qui incluent la Méditerranée (le Met Office Hadley Centre EN-v4 (Good et al., 2013)) et le produit ISHII (Ishii et al., 2003).

4. <http://www.marine.copernicus.eu>

2 Les données *in situ* et de modélisation

2.1 Le ruissellement des rivières

Aucun produit satellite n'est encore⁵ dédié à l'estimation du débit des rivières. De plus, peu de données publiques ou privées concernant les rivières sont encore partagées avec le Global Runoff Data Centre (GRDC) qui recueille des données sur les débits à l'échelle mondiale. Par conséquent, deux sources indirectes d'information sur les débits et ruissellements seront nécessaires. Dans cette étude, le terme de ruissellement est utilisé pour caractériser tout écoulement vers la mer. Si le point côtier est une embouchure de rivière, on parle d'écoulement fluvial, sinon d'un ruissellement de surface.

L'ensemble de données CEFREM-V2 de l'écoulement annuel côtier dans la mer Méditerranée (Ludwig et al., 2009) est basé sur des observations *in situ* et quelques estimations indirectes utilisant la formule de Pike (1964). Cet ensemble de données a une résolution spatiale élevée ($< 0.25^\circ$) et une grande étendue temporelle (1980-2009). Malheureusement, les valeurs fournies ne sont que des valeurs annuelles. Comme nous nous intéressons au bilan hydrique à l'échelle mensuelle, la dynamique mensuelle de ces écoulements doit être fournie par une autre source.

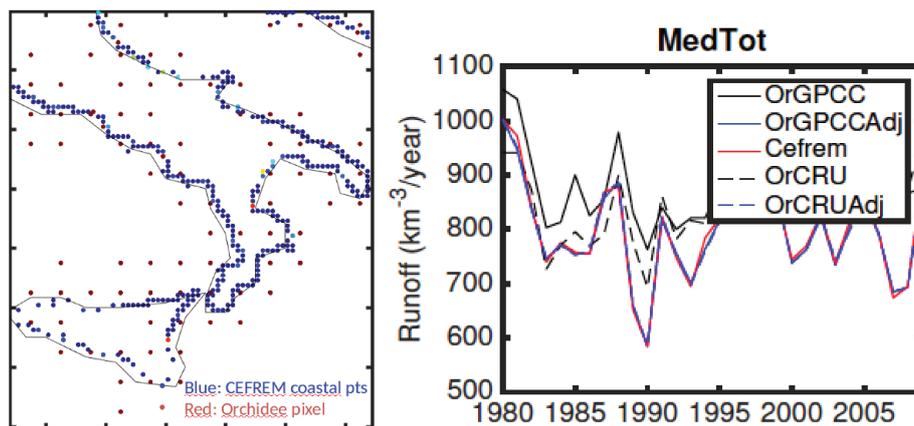


Figure III.9. Visualisation des côtes dans les produits CEFREM-V2 et ORCHIDEE au niveau de l'Italie (gauche). Débit d'eau annuel entrant dans la Méditerranée entre 1980 et 2009 pour les deux produits CEFREM-V2 et ORCHIDEE (droite).

Organising Carbon and Hydrology In Dynamic EcosystEms (ORCHIDEE) est le schéma de surface du modèle de climat de l'Institut Pierre Simon Laplace (IPSL) développé au Laboratoire de Météorologie Dynamique (Polcher et al., 1998). Il résout explicitement les bilans d'eau et d'énergie des surfaces continentales, ainsi que la phénologie (floraison, feuillaison) et le cycle du carbone de la biosphère terrestre. Il est choisi ici pour décrire la dynamique mensuelle des écoulements fluviaux et ruissellements. Deux produits sont disponibles à partir de son schéma de routage, ils sont issus de deux forçages de précipitations différents : GPPC et le produit du Climatic Research UNit (CRU).

5. La mission SWOT prévue en 2021 a pour objectif, entre autres, de mesurer le débit des rivières de largeur supérieure à 100 m.

La couverture temporelle des données du modèle ORCHIDEE est identique à CEFREM (1980-2009) mais sa résolution spatiale est plus grossière, à 0.5° . Nous avons projeté les dynamiques mensuelles modélisées par ORCHIDEE sur la grille CEFREM en conservant les valeurs annuelles de CEFREM. La Figure III.9, à gauche, montre les résolutions spatiales des deux produits au dessus de l'Italie du Sud et la Sicile. Le produit final a alors la résolution spatiale et la valeur cumulative annuelle du CEFREM, mais avec la dynamique mensuelle du modèle ORCHIDEE. La Figure III.9, à droite, compare les débits annuels simulés par ORCHIDEE avec les valeurs CEFREM. L'apport annuel moyen d'eau douce en Méditerranée (sans la mer Noire) est $401 \text{ km}^3/\text{an}$ ($375 \text{ km}^3/\text{an}$ dans ORCHIDEE) pour la période 1980-2009.

2.2 Mesure de l'écoulement net au détroit de Gibraltar

La seule estimation pluriannuelle (2004-2016) du débit net (entrée-sortie) de Gibraltar basée sur les observations est celle présentée dans (Jordà et al., 2016). Le produit est basé sur la reconstruction du champ de vitesse d'écoulement sur une section verticale du détroit en utilisant une technique d'interpolation optimale des mesures *in situ*. L'interpolation utilise des modèles de covariances spatiales déduits de simulations numériques à haute résolution spatiale. Nous utilisons une reconstruction mensuelle du transport net où les effets de la pression atmosphérique ont été supprimés. Ceci est fait par souci de cohérence avec les estimations GRACE du stockage de l'eau dans la mer. La technique de reconstruction utilisée pour générer cette estimation s'est avérée efficace pour simuler la variabilité, mais les incertitudes sur la moyenne sont importantes. Par conséquent, dans ce travail, nous substituons la valeur moyenne 2004-2016 de l'estimation de Jordà et al. (2016) par l'estimation proposée dans (Jordà et al., 2017) issue d'un consensus entre experts lors d'une évaluation du transport à Gibraltar. La Figure III.10 montre l'estimation de l'écoulement à Gibraltar avant et après correction de biais (i.e. la moyenne).

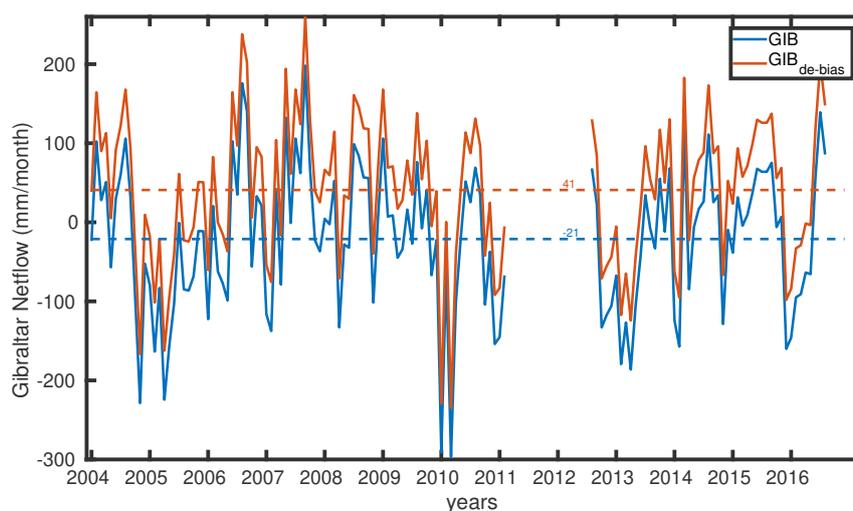


Figure III.10. Écoulement net à Gibraltar (en mm.mois^{-1}) avant (bleu) et après (rouge) l'ajustement de biais (Jordà et al., 2017). L'aire totale de la Méditerranée incluant la mer Noire est de $3.5 \times 10^6 \text{ km}^2$.

2.3 Pluviomètres

Les données de pluviométrie par jauges seront utilisées pour la validation. Le produit d'observation Climat Assessment & dataset EOBS a été développé à partir du projet EU-FP6 ENSEMBLES (Haylock et al., 2008). Il s'agit d'un ensemble de données régionales (2316 jauges sur l'Europe) interpolé pour produire une base de données à la résolution spatiale de 0.25° et couvrant la période 1950-2007. EOBS a été bien documentée et validée avec des données *in situ*. La Figure III.11 montre la distribution spatiale des jauges utilisées dans le produit EOBS.

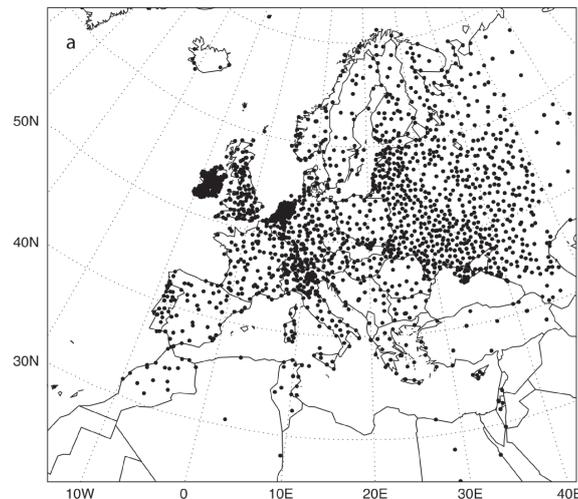


Figure III.11. *Distribution spatiale des jauges utilisées dans le produit EOBS (Haylock et al., 2008). On remarque la répartition hétérogène des jauges de précipitation. L'ensemble des données *in situ* pour la région nord-africaine est relativement épars ce qui limite la résolution du produits EOBS pour cette région. En outre aucune estimation n'est obtenue pour la mer Méditerranée.*

2.4 Sites micro-météorologiques

Les données d'évapotranspiration sur les sites micro-météorologiques seront utilisées pour la validation de la précipitation et de l'évapotranspiration. FLUXNET est un réseau mondial de sites de tours micro-météorologiques qui utilisent des méthodes de covariance tourbillonnaire (ou flux des turbulences) pour mesurer les échanges de dioxyde de carbone, de vapeur d'eau et d'énergie entre les écosystèmes terrestres et l'atmosphère (Falge et al., 2017). Plus de 500 stations à travers le monde fonctionnent en continu. L'objectif de FLUXNET est de fournir des informations pour valider les produits de télédétection pour l'évaporation. La Figure III.12 montre la distribution spatiale des sites micro-météorologiques dans la base de données FLUXNET.

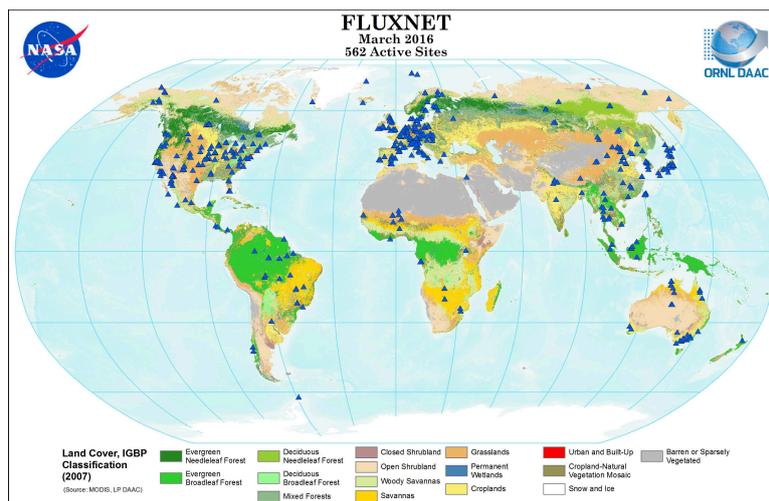


Figure III.12. Distribution spatiale des sites présents dans la base de données FLUXNET. Les différentes zones de Köppen sont précisées sur la carte (<http://fluxnet.fluxdata.org/sites/>).

3 La ré-analyse ERA-I de l'ECMWF

L'idée de l'analyse est de fusionner des observations satellites irrégulières dans le temps avec des modèles continus et ce, pour de nombreux processus physiques, afin de générer une estimation (i.e. une "analyse") dynamique de l'état du système (atmosphère et/ou surface), avec une homogénéité spatiale et une continuité temporelle. Un système d'analyse comprend un modèle de prévision de base et une routine d'assimilation de données. Les observations assimilées dans le système d'analyse influencent/corrigent toutes les prévisions suivantes du système. La Figure III.13 illustre l'amélioration de la prévision lors de l'assimilation de différents types d'observations satellites à deux dates différentes. Il a été vu au Chapitre 2 que les observations des sondeurs micro-ondes sont les plus assimilées par les centres de PNT. Grâce à leur capacité à traverser les nuages ce sont aussi les observations qui ont permis la plus grande amélioration de prévisions (voir Figure III.13).

La ré-analyse a pour but d'améliorer la représentation du climat sur toute une période considérée (1980-2012 dans ERA-I). Alors que l'analyse a pour but de fournir des *a priori* (i.e. une première approximation) pour la prédiction future, la ré-analyse conserve ces analyses et assimile d'autres données historiques non disponibles lors de l'analyse initiale en temps réel. Le modèle de base et les routines d'assimilation sont conservés sur toute la période. On peut citer : le système de réanalyse NCEP/NCAR (NCEP1 et NCEP2), ECMWF réanalyse (ERA-40 et ERA-I) et le système MERRA (Modern-Era Retrospective Analysis for Research and Application system).

ERA-Interim (ERA-I) applique l'assimilation de données variationnelles quadridimensionnelles (4D-Var) qui ajoute à l'assimilation spatiale de données (3D-Var) la dimension du temps : toutes les observations antérieures sont re-utilisées à chaque nouveau cycle d'assimilation. Les produits ERA-I sont mis à jour une fois par mois, avec un délai de deux mois afin de contrôler la qualité des produits.

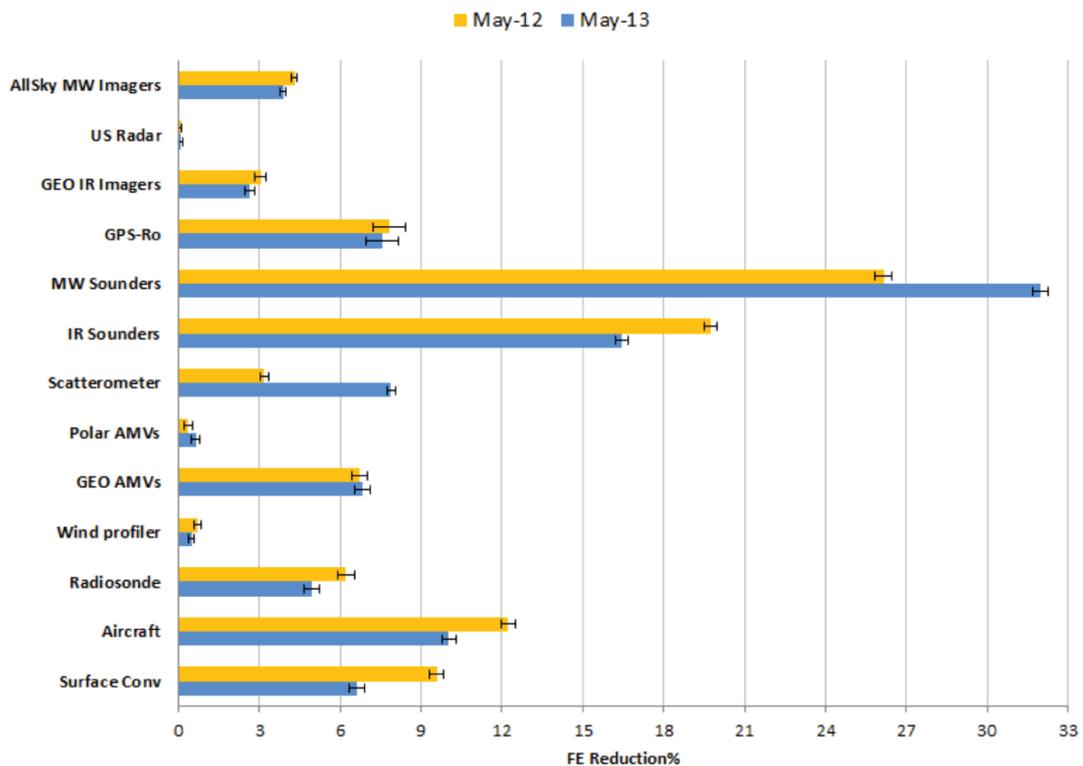


Figure III.13. Réduction de l'erreur de prévision ("Forecast Error", FE) en pourcentage par assimilation de différents types d'observations satellites dans ERA-I à deux dates (English et al., 2013).

3.1 Divergence de l'humidité verticalement intégrée dans ERA-I

Nous utilisons le produit mensuel de la réanalyse ERA-I (Dee et al., 2011) pour représenter la divergence de l'humidité verticalement intégrée (le modèle atmosphérique de ERA-I possède une résolution verticale de 60 couches atmosphériques sur 80 km). Le produit mensuel est obtenu par intégration des moyennes journalières, elles-mêmes obtenues comme moyenne des quatre analyses quotidiennes (à 00h, 06h, 12h et 18h). La résolution temporelle de ce produit est donc de 6 heures. La Figure III.14 montre la ré-analyse de la divergence d'humidité atmosphérique intégrée le long de la colonne d'atmosphère pour le mois de janvier 2004. On distingue des zones convergentes (valeurs négatives) comme au-dessus des Alpes et des zones divergentes (valeurs positives) comme au-dessus de la mer Méditerranée. Parmi les différentes réanalyses, ERA-I a été adopté pour notre étude en raison de résultats démontrant les avantages de la représentation de la variabilité à long terme du vent (Stopa et Cheung, 2014) qui joue un rôle clé dans la représentation de la divergence de l'humidité. Néanmoins Seager et Henderson (2013) ont montré la limitation de la ré-analyse qui ne représente pas les événements de divergence d'humidité plus courts que l'échelle temporelle de 6 heures. Cette limitation devra être prise en compte lors de la fermeture du cycle de l'eau.

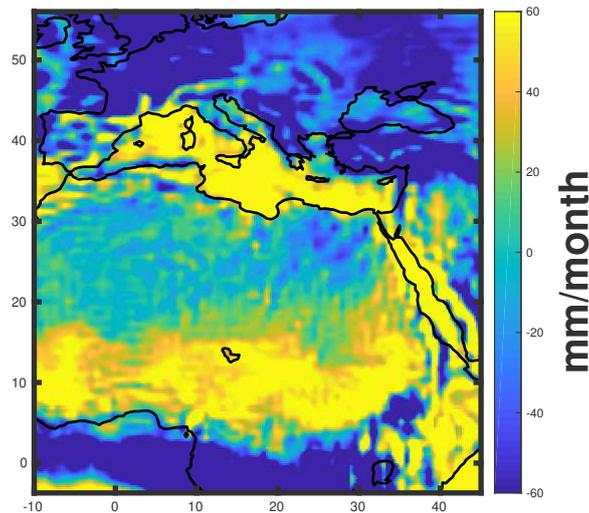


Figure III.14. Estimation de la divergence d'humidité atmosphérique verticalement intégrée pour le mois de janvier 2004 dans le produit de ré-analyse ERA-I.

3.2 La vapeur d'eau atmosphérique verticalement intégrée

Pour la représentation de la vapeur d'eau atmosphérique, verticalement intégrée, outre le produit satellite GobVapour, le produit de ré-analyse de 6 heures de ERA-I (Dee et al., 2011) est lui aussi considéré. La ré-analyse ERA-I a été choisie ici car l'eau précipitable, bien qu'elle soit basée sur un modèle, est largement contrainte par l'observation par satellite en particulier via l'assimilation des températures de brillance des sondeurs. La Figure III.15 montre la ré-analyse du changement de contenu en vapeur d'eau atmosphérique entre décembre 2003 et janvier 2004.

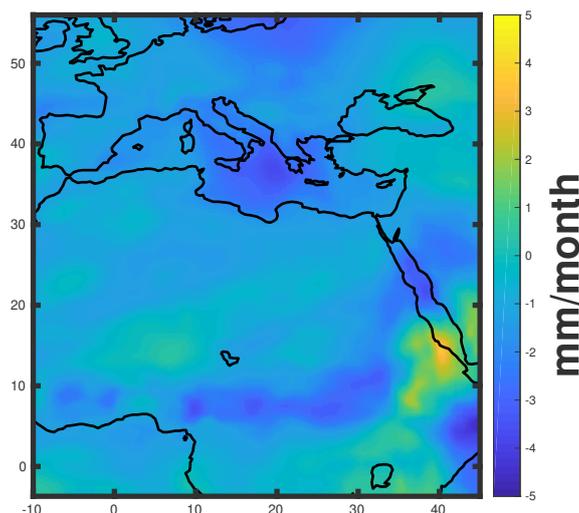


Figure III.15. Estimation des changements de masse d'eau atmosphérique entre décembre 2003 et janvier 2004 dans le produit de ré-analyse ERA-I. Contrairement au produit satellite bruité (voir Figure III.8), l'utilisation d'un modèle permet ici d'obtenir une estimation plus continue spatialement.

Conclusion du chapitre

Ce chapitre a décrit l'ensemble des produits utilisés, par la suite, dans l'étude du cycle de l'eau méditerranéen. Afin de les intégrer, les produits ont été projetés sur une grille de résolution spatiale commune de 0.25° et ré-échantillonnés temporellement à l'échelle du mois. Les différents produits satellites sont dérivés par des algorithmes plus ou moins complexes : si la vapeur d'eau atmosphérique peut être directement restituée par inversion du transfert radiatif, la restitution de la précipitation nécessite des modèles statistiques et une synergie entre instrument de mesure tandis que l'évapotranspiration est paramétrée à partir de produits eux-mêmes restitués (température de surface et vitesse du vent) et seule la ré-analyse permet de décrire la divergence d'humidité atmosphérique. Les produits satellites choisis dans cette thèse sont communément utilisés et représentent bien la variabilité qui existe entre produits pour une même variable géophysique.

Chapitre IV

L'observation satellite du cycle de l'eau méditerranéen

Où une meilleure compréhension du cycle hydrologique méditerranéen dans son intégralité et une quantification de ses processus entre les composantes atmosphérique, continentale et océanique sont obtenues en utilisant principalement les produits satellites. Et où la meilleure façon d'utiliser la multiplicité des produits satellites y est étudiée afin de minimiser les erreurs faites sur la fermeture du cycle de l'eau. Ce chapitre s'appuie sur (Pellet et al., 2018).

1 La mer Méditerranée et sa région



Figure IV.1. La mer méditerranée et sa région (www.earth.google.fr)

La superficie de la mer Méditerranée, à l'exclusion de la mer Noire, est d'environ $2.5 \cdot 10^6 \text{ km}^2$; la profondeur moyenne est de 1500 m avec une valeur maximale de 5150 m dans la mer Ionienne (au sud de la mer Adriatique, entre l'Italie et la Grèce). Elle est entourée de 21 pays d'Afrique, d'Asie et d'Europe. Les pays riverains de la Méditerranée appartiennent au bassin de façon très variée : certains s'y trouvent entièrement (Albanie, Grèce, Italie...), d'autres assez partiellement (Espagne, France, Turquie et tous les pays du Sud et du Proche-Orient). Réciproquement, le bassin méditerranéen est partagé très inégalement entre les pays riverains ; l'Italie, l'Égypte, la Turquie et l'Espagne ont les plus grandes parts du bassin.

"Le Nil n'est méditerranéen à son terme que par un accroc de l'histoire géologique" (Margat, 2004).

La plupart des études sur la mer Méditerranée considèrent son bassin hydrographique sans le bassin du Nil (Mariotti et al., 2002 ; Philandras et al., 2011). Ce bassin présente des particularités hydrologiques comme le barrage d'Assouan (construit en 1960) ou son climat tropical et équatorial à sa source près du lac Victoria. Il est souvent considéré comme un bassin clos et autophage ne faisant pas partie du bassin hydrographique méditerranéen car son débit à l'embouchure de la mer Méditerranée est négligeable. Ludwig et al. (2009) et Margat (2004) ont considéré que la plus grande partie de l'eau du Nil (68 % de l'apport du Nil Blanc et de ses affluents) est perdue par évaporation dans les lacs, les deltas intérieurs, les marécages et par transpiration. Antérieurement à toute influence humaine, le Nil ne devait fournir en Méditerranée que $73 \text{ km}^3 \cdot \text{an}^{-1}$ d'eau douce. Sous l'effet des irrigations millénaires en Égypte, le débit du Nil est tombé à $40 \text{ km}^3 \cdot \text{an}^{-1}$ au milieu du XX^e siècle, avant les aménagements modernes (barrage d'Assouan). Néanmoins l'incertitude reste élevée sur la quantité d'eau charriée dans ce bassin ainsi que son débit à l'embouchure et nous avons donc décidé ici de garder le sous-bassin du Nil dans notre étude et de vérifier cette hypothèse. Le bassin hydrographique du Nil est de plus de 3 millions km^2 .

La mer Méditerranée est un bassin semi-fermé, relié à l'océan Atlantique par le détroit de Gibraltar (14.5 km de large et moins de 300 m de profondeur). La région méditerranéenne a une morphologie complexe (Figure IV.1). Les montagnes entourant la mer Méditerranée influent sur la circulation atmosphérique de la région. Le sommet du Mont Blanc, culminant la région méditerranéenne à une hauteur de 4 800 m , se trouve dans les Alpes. L'orographie de la région permet la présence de glaciers permanents et une couverture de neige étendue en hiver.

Îles, golfes, péninsules et bassins régionaux décrivent une distribution terre-mer complexe avec des conséquences importantes sur la circulation océanique et atmosphérique : une grande variabilité spatiale. La topographie océanique est également compliquée avec des bassins profonds reliés par des détroits beaucoup moins profonds. La circulation en mer Méditerranée est caractérisée par la formation d'eau dense et des gyres locaux.

La mer Méditerranée et la mer Noire s'étendent sur une surface totale de 3.0 million km^2 et leurs bassins hydrographiques cumulés représentent plus de 5 millions km^2 du pourtour méditerranéen. Les plus grands fleuves méditerranéens sont le Nil, le Rhône, le Pô, l'Èbre et le Tibre. Le Danube et le Dniepr sont les principaux contributeurs en eau douce de la mer Noire.

1.1 Le climat méditerranéen

Le "climat méditerranéen" est l'un des climats dans la classification qualitative des différents types de climat sur Terre (par exemple Köppen (1936)). Il a été utilisé pour définir le climat d'autres régions en plus de la région méditerranéenne elle-même¹. Le concept de climat "méditerranéen" est caractérisé par des hivers doux et humides et

1. La limite de la culture de l'olivier correspondait historiquement à la zone d'extension de ce climat.

des étés chauds et secs. Il est localisé sur la côte ouest des continents entre environ 30° et 40° de latitude (Californie, centre du Chili, région du Cap en Afrique du Sud, et Ouest de l'Australie).

La mer Méditerranée est située dans une zone de transition, où les latitudes moyennes et la variabilité tropicale sont concurrentes. La partie nord de la région méditerranéenne présente un climat océanique, tandis que la partie sud est caractérisée par un climat subtropical. Il y a une transition continue Nord/Sud entre le climat océanique, le climat méditerranéen puis le climat désertique. La position intermédiaire des régions au climat méditerranéen entre les dépressions tempérées et les anticyclones subtropicaux fait que la saison froide est humide et variable et que la saison chaude est sèche et peu variable. Plus on va vers le Nord, plus la sécheresse estivale est courte et moins marquée ; plus on s'approche de l'équateur, plus la sécheresse estivale dure. Le climat méditerranéen est exposé à la mousson sud-asiatique en été et à l'anticyclone sibérien en hiver. La partie sud de la région est principalement sous l'influence de la branche descendante de la cellule de Hadley, tandis que la partie nord est davantage liée à la variabilité des latitudes moyennes, caractérisée par le NAO (North Atlantic Oscillation).

La climatologie pluviométrique de la région méditerranéenne est caractérisée par des étés secs associés à de très longues périodes de sécheresse, suivis de précipitations automnales et hivernales très intenses. Il en résulte une grande variabilité saisonnière dans la recharge des aquifères, le débit des rivières, le contenu en eau du sol et les caractéristiques de la végétation. En outre, l'interaction avec l'atmosphère n'est pas encore bien connue.

Les régions de climat méditerranéen sont réputées pour leurs vents locaux catabatiques (la tramontane en Languedoc-Roussillon et le mistral en Provence). Ces vents descendant la montagne sont froids et asséchant et soufflent du continent européen vers la mer. Le sirocco et le khamsin, chauds et secs, soufflent de l'Afrique du Nord vers la mer. Au contraire, la circulation atmosphérique océanique, amène la chaleur et l'humidité depuis la mer vers les pays du pourtour méditerranéen.

1.2 Situation socio-économique et importance de l'eau dans la région

L'amplification des problèmes liés à l'eau dans la région méditerranéenne a amené le Plan Bleu à publier un essai de synthèse et de prospective sur l'eau en Méditerranée. À cette fin, Margat (2004) a rendu un rapport à l'Organisation des Nations Unies sur la situation de l'eau en Méditerranée.

La région méditerranéenne ne dispose que de 3% des ressources en eau du monde, alors qu'elle rassemble 7,3% de l'humanité. Près de 60% de la population mondiale pauvre en eau, c'est-à-dire disposant de moins de 1000 m^3 de ressource moyenne annuelle par habitant, se concentre dans les seuls pays méditerranéens du Sud et du Proche-Orient.

La Figure IV.2 montre les ressources en eau naturelles renouvelables et ressources exploitables, par habitant, dans les pays méditerranéens (en référence aux populations de l'année 2000). Les ressources en eau rapportées aux populations sont un indicateur d'abondance ou de rareté de l'eau. Pour comparer les pays au moyen de cet indicateur, des seuils sont couramment utilisés :

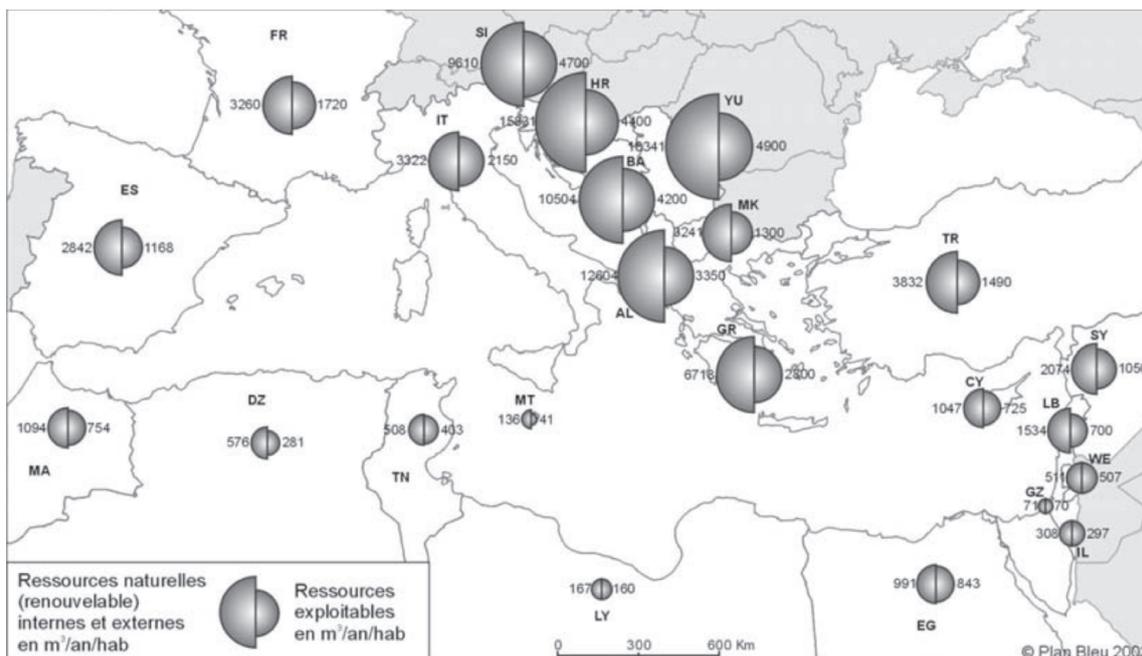


Figure IV.2. Ressources en eau naturelles renouvelables et ressources exploitables, par habitant, dans les pays méditerranéens (pour l'année 2000, Margat (2004))

- des ressources en eau moyennes par habitant comprises entre $1000 \text{ m}^3 \cdot \text{an}^{-1}$ et $500 \text{ m}^3 \cdot \text{an}^{-1}$ sont caractéristiques d'une "tension" ou pauvreté en eau ("Water stress" en anglais)
- des ressources en eau moyennes par habitant inférieures à $500 \text{ m}^3 \cdot \text{an}^{-1}$ sont caractéristiques d'une pénurie en eau ("Scarcity" en anglais)

Les ressources naturelles par habitant des pays varient entre moins de $100 \text{ m}^3 \cdot \text{an}^{-1}$ (Gaza) et plus de $10\,000 \text{ m}^3 \cdot \text{an}^{-1}$ (Balkans). Les ressources exploitables par habitant varient dans une moindre mesure mais encore très largement (de $<100 \text{ m}^3 \cdot \text{an}^{-1}$ à $>2000 \text{ m}^3 \cdot \text{an}^{-1}$, en excluant les Balkans). Suivant les valeurs-seuils admises, deux pays ou territoires se trouvaient dès l'an 2000 en situation de pauvreté en eau (Chypre, Égypte) et sept en situation de pénurie (Algérie, Cisjordanie, Gaza, Israël, Libye, Malte, Tunisie). Ces neuf pays représentent 153 millions d'habitants. Ainsi, sur 433 millions de méditerranéens, en 2000, plus d'un tiers vivent dans des pays au-dessous du seuil de pauvreté en eau, dont plus de 50 millions dans les pays en situation de pénurie en eau.

Globalement les pays méditerranéens utilisent près du quart de leurs ressources renouvelables naturelles en eau, et presque la moitié de leurs ressources exploitables réelles, avec une forte dissymétrie Nord/Sud. L'agriculture irriguée reste de beaucoup le principal utilisateur (63% pour l'ensemble du bassin méditerranéen, 80% au Sud et au Proche-Orient).

1.3 Principaux projets de recherche sur le cycle de l'eau en Méditerranée

Alors que le rapport du Plan Bleu de l'ONU (Margat, 2004) tire la sonnette d'alarme sur la disponibilité et le partage des ressources en eau en Méditerranée, la région méditerranéenne est identifiée comme l'un des principaux "hot spot"² du changement climatique par le Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC) : sa sensibilité aux changements climatiques est élevée et son évolution reste très incertaine (Bernstein et al., 2014). La région méditerranéenne est dominée par un réchauffement significatif, une forte diminution des précipitations moyennes annuelles et une augmentation de leur variabilité pendant la saison chaude et sèche. La région connaît de nombreuses interactions et rétroactions aux niveaux océanique, atmosphérique et hydrologique, tout en faisant face à des activités anthropiques élevées.

Plusieurs programmes ou réseaux internationaux de recherche sont dédiés à la compréhension du climat et des interactions atmosphère, océanographie, hydrologie et biologie à l'échelle mondiale. Le bassin méditerranéen fait naturellement l'objet d'une composante régionale forte pour chacun de ces projets internationaux.

Parallèlement au rapport du Plan Bleu de l'ONU, le World Hydrological Cycle Observing System (WHYCOS) est lancé en 1993 par l'OMM avec l'appui de la Banque Mondiale. Le projet vise à améliorer la connaissance des ressources en eau pour le développement durable. WHYCOS a d'abord été développé en tant qu'organisation mondiale, avec pour objectif l'établissement d'environ 1000 stations sur les principaux fleuves du monde afin de mesurer des variables clés comme le débit des rivières, les sédiments, la chimie de l'eau et des variables météorologiques. Sa mise en œuvre a évolué en une implantation de composantes régionales plus autonomes et indépendantes dans le temps. La première composante régionale du projet WHYCOS sur la Méditerranée, MED-HYCOS, est lancée en 1995. Ce programme météorologique avait pour principal objectif le partage en temps réel des données hydrologiques entre pays du bassin méditerranéen ainsi que le renforcement des capacités techniques et institutionnelles de leurs services hydrologiques (MED-HYCOS, 2001).

Le réseau scientifique MedCLIVAR (Mediterranean Climate Variability), créé en 2003 au sein du projet CLIVAR, est axé sur les conséquences de la variabilité de l'océan et de ses processus dans les régions froides et leur sensibilité au changement climatique. CLIVAR, lancé en 1995, traite des processus au-dessus des océans, souvent appelés processus climatiques "lents". MEDCLIVAR est la première initiative de recherche internationale cherchant à développer une vision multidisciplinaire de l'évolution du climat méditerranéen à travers des études qui intègrent les composantes atmosphériques, marines et terrestres du climat à des échelles de temps allant de la paléo-reconstruction aux scénarios climatiques futurs (Lionello et al., 2006).

Le méta-programme décennal (2010-2020) et international MISTRALS (Mediterranean Integrated Studies at Regional And Local Scales) recoupe plusieurs programmes de recherche autour de questions scientifiques liant sciences de l'atmosphère, océanographie, hydrologie et biologie. L'un de ces programmes, HyMeX (Hydrological cycle in Mediterranean Experiment), a pour objectif de mieux caractériser le cycle de l'eau dans le bassin méditerranéen à des échelles spatio-temporelles s'étendant de l'évènement météorologique à la caractérisation de l'évolution climatique multi-décennale

2. un hot-spot identifie une région dont le climat est particulièrement sensible aux changements globaux

(Drobinski et al., 2013). HyMeX est une composante régionale du programme GEWEX³. Ce programme a essayé une multitude de projets régionaux et noyauté encore la recherche sur le cycle de l'eau aujourd'hui. La Figure IV.3 illustre les différents groupes de travail répartis au sein d'HyMeX.

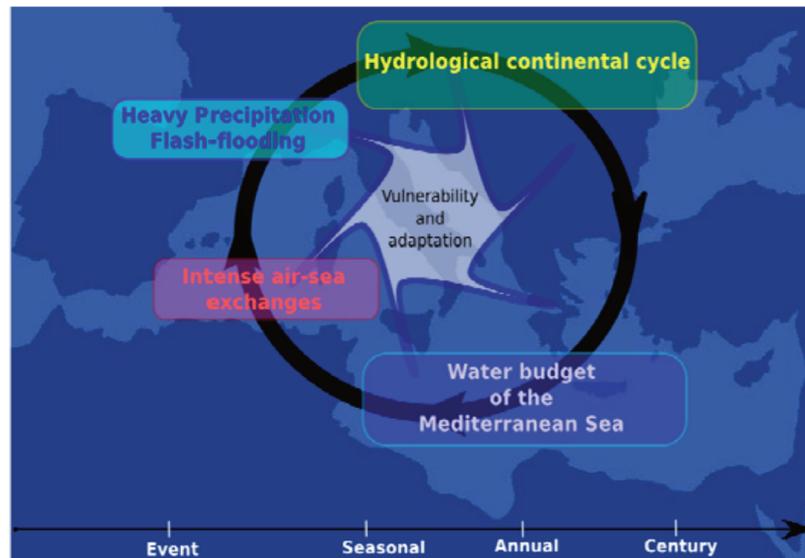


Figure IV.3. Organisation des différents groupes de travail (WG pour Working Group) autour du projet HyMeX.

La mission ESA, WACMOS-MED est une régionalisation du programme Water Cycle Observation Multi-mission Strategy (WACMOS) en soutien au programme HyMeX. WACMOS a été lancé par l'ESA en 2009, en collaboration avec GEWEX pour développer de nouvelles méthodes et de nouveaux produits satellites destinés à l'étude du cycle de l'eau pour évaluer ses variations et ses tendances. L'initiative a notamment pour but d'aider les scientifiques des pays membres de l'ESA à développer et valider et maximiser l'utilisation des produits satellites de l'ESA. En outre, le projet permet d'aborder le couplage entre les données terrestres et les données atmosphériques du cycle de l'eau et d'évaluer ses influences sur le climat. WACMOS-MED a pour but de contribuer aux objectifs scientifiques de HyMeX en exploitant les données satellites afin d'améliorer notre compréhension du cycle de l'eau dans la région méditerranéenne grâce à une approche intégrée impliquant conjointement les composantes terrestres, océaniques et atmosphériques. Ce travail de thèse a été effectué dans le cadre du projet WACMOS-MED (Pellet et al., 2018,b).

Plusieurs études ont déjà porté sur le bassin méditerranéen pour suivre l'évolution du cycle de l'eau. Sanchez-Gomez et al. (2011) ont décrit le cycle de l'eau de la mer Méditerranée à l'aide d'un ensemble de modèles de climat régionaux ("Regional Climate Model", RCM en anglais), Mariotti et al. (2002) ont exploré la relation entre le cycle de l'eau et le NAO puis la variation à long terme des précipitations et de l'évapotranspiration (Mariotti, 2010 ; Senay et al., 2009). Certaines études ont porté spécifiquement

3. Lancée par l'OMM en 1990, GEWEX a joué un rôle central dans l'élargissement des activités de recherche sur le cycle de l'eau à différentes échelles temporelles et ses liens avec le cycle énergétique. Un des objectifs était de déterminer le cycle hydrologique de la Terre et les flux d'énergie à l'aide de d'observations satellites.

sur la variabilité de l'eau douce et des nutriments (Ludwig et al., 2009 ; Struglia et al., 2004), et d'autres sur les impacts du changement climatique (Mariotti et al., 2008). Cependant, la plupart de ces études utilisent des simulations de modèles ou des données de ré-analyse, voir des bases de données de mesures *in situ* mais très peu d'observations satellites.

La région méditerranéenne est marquée par une sévère dissymétrie en termes de système d'observations et de suivis des variables géophysiques. Si les pays européens ont pu mettre en place très tôt des réseaux de mesure au sol à grande échelle (nationale ou internationale), l'Afrique du nord en est totalement dépourvue. Ainsi pour ces pays du Sud, l'observation satellite demeure l'unique opportunité d'un suivi en temps réel des composantes de son climat local.

2 Outils d'analyses

Afin d'analyser le cycle de l'eau méditerranéen, plusieurs outils sont nécessaires, aussi bien la définition des sous bassins de la Méditerranée que la définition mathématique du bilan hydrique. Cette section introduit le matériel de ce travail de recherche.

2.1 Définition des sous-bassins

La région sur laquelle cette étude se concentre est représentée dans la Figure IV.4. Il s'agit du bassin versant de la mer Méditerranée et de la mer Noire (il comprend tous les bassins des fleuves qui se déversent dans les deux mers). Comme aucune hypothèse n'est faite pour l'écoulement au Bosphore, et ce en raison du manque de données, dans cette étude, la mer Méditerranée se réfèrera par la suite à la mer Méditerranée et à la mer Noire (une seule couleur dans la Figure IV.4). La région méditerranéenne est principalement située entre les latitudes moyennes de l'Europe et le désert du Sahara, mais le sous-bassin du Nil (LY-EG, défini ci-dessous) s'étend jusqu'à l'équateur au niveau du lac Victoria.

L'aire du bassin versant a été calculée à partir de chaque pixel côtier, à l'aide d'un modèle hydrologique (Wu et al., 2011) à une résolution spatiale de 0.25 °.

La subdivision du bassin méditerranéen en sous-bassins versants facilite l'analyse des climats locaux et des caractéristiques hydrologiques spécifiques. Ludwig et al. (2009) ; Cruzado (1985) ont divisé la région méditerranéenne en onze sous-bassins à la fois océaniques (divisant la mer Méditerranée) et continentaux, et ce en se basant sur la présence de plusieurs types de nutriments dans les régions côtières. Dans cette étude, nous avons décidé de les agréger en six sous-bassins continentaux sur la base de frontières naturelles et des caractéristiques hydrologiques. Les massifs montagneux de l'Atlas, des Pyrénées et des Alpes et des considérations climatiques ont été utilisés pour rassembler les zones de drainage. Cette agrégation a l'avantage de réduire en moyenne, à l'échelle du sous-bassin, les erreurs dans le contenu en eau estimé par GRACE (qui a une résolution grossière de 300 km²), et dans la divergence de l'humidité de ERA-I où les caractéristiques locales sont moins bien représentées, (Mueller et al., 2011). La mer Méditerranée et les six sous-bassins terrestres utilisés dans cette étude sont définis comme suit :

- le Maghreb occidental basé principalement sur le débit d'eau douce des montagnes de l'Atlas (**MA-DZ-TN**) ;

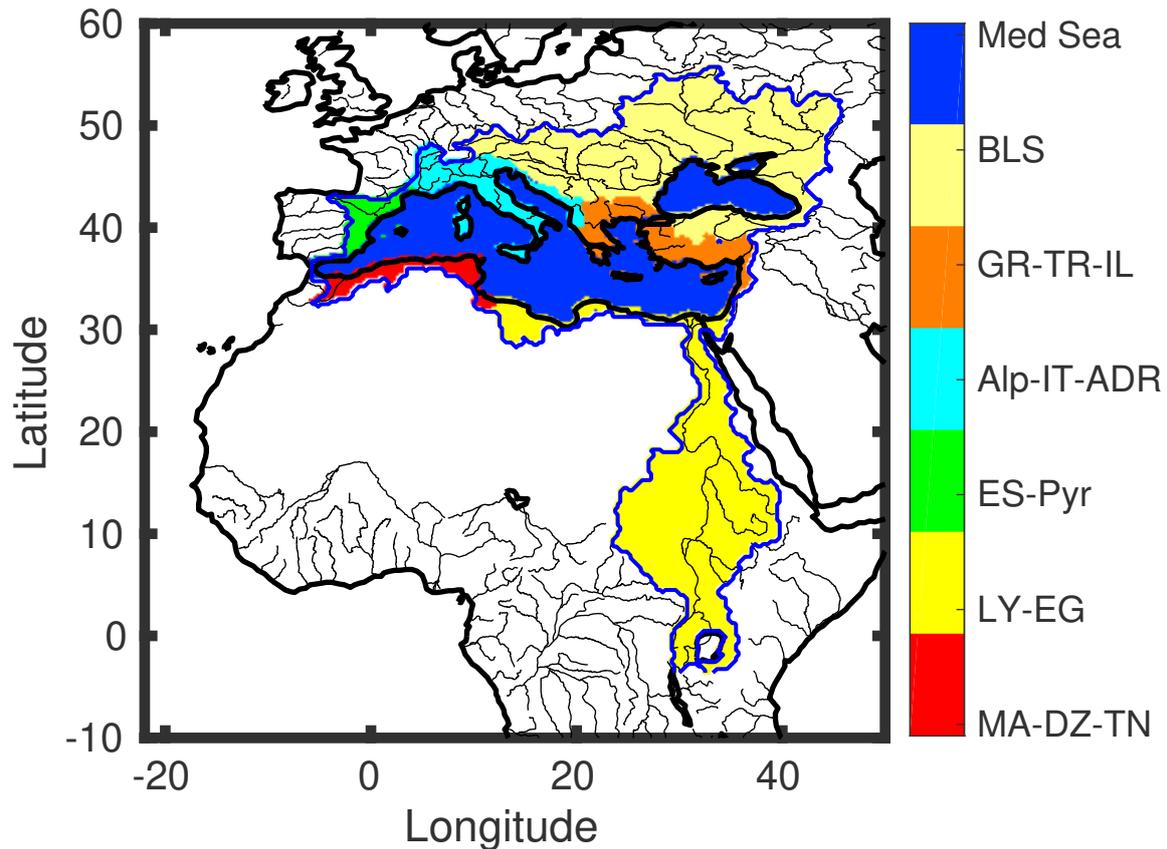


Figure IV.4. Sous-bassins de l'aire de drainage méditerranéenne considérés dans cette étude. Les sous-bassins ont été définis à l'aide d'un modèle hydrologique (Wu et al., 2011), et les rivières proviennent de HydroShed (<http://www.hydrosheds.org/>). Voir le texte pour la définition des sous-bassins.

- le bassin du Nil et la côte libyenne caractérisant le climat saharien et subsaharien (LY-EG) ;
- la côte espagnole et les Pyrénées (ES-Pyr) ;
- la côte française, l'Italie et la mer Adriatique, bassins forcés par les Alpes et les Balkans (Alp-IT-ADR) ;
- la partie orientale de la Méditerranée, la Grèce, la Turquie et Israël (GR-TR-IL) ;
- l'ensemble du bassin versant de la mer Noire, Bulgarie, Géorgie, Roumanie, Russie, Turquie, Ukraine, Slovaquie, Hongrie, Autriche, Slovénie, Bosnie, Serbie (BLS).

La définition hydrologique de ces sous-bassins assure l'indépendance hydrique des sous-bassins entre eux. Cette indépendance permettra de considérer séparément le cycle de l'eau et le bilan hydrique d'un bassin à l'autre.

2.2 Le bilan hydrique

Si les équations du cycle de l'eau présenté au chapitre précédent (Eq. (III.1)) ne sont pas satisfaites, il reste des résidus. Les résidus du bilan hydrique sont définis comme suit :

$$\begin{aligned}
Res1_l &= \hat{P}_l - \hat{E}_l - \hat{R}_l - \frac{\delta \hat{S}_l}{\delta t}, \\
Res1_o &= \hat{P}_o - \hat{E}_o + \hat{R}_l^* - \hat{R}_o - \frac{\delta \hat{S}_o}{\delta t}, \\
Res2_{l/o} &= \hat{E}_{l/o} - \hat{P}_{l/o} - \text{div}(\hat{Q}_{l/o}) - \frac{\delta \hat{W}_{l/o}}{\delta t},
\end{aligned} \tag{IV.1}$$

où le symbole $\hat{}$ souligne une estimation (satellite, *in situ*,...) des composantes de l'eau. *Res1* est le résidu du budget hydrique de surface ;

Res2 est le résidu du budget hydrique dans l'atmosphère. Les résidus du budget hydrique (*Res1*, *Res2*) devraient être nuls si le cycle de l'eau est "fermé". Par conséquent, le calcul des résidus est une mesure d'erreur.

Les trois relations dans l'Eq. (IV.1) sont en $mm.mois^{-1}$ (normalisée par la surface) pour des raisons de commodité et afin de comparer directement les résultats sur des sous-bassins de tailles diverses.

Tout au long de cette thèse, les capacités des diverses méthodologies seront mesurées par la RMS des résidus du bilan hydrique. L'analyse sera effectuée à l'échelle temporelle mensuelle et à l'échelle du bassin méditerranéen (c'est-à-dire l'ensemble de l'aire de drainage de la Méditerranée et de la mer Noire) et du sous-bassin (c'est-à-dire l'aire de drainage d'une région côtière particulière). La meilleure façon d'utiliser la multiplicité des ensembles de données y est étudiée afin de minimiser les erreurs sur le cycle de l'eau (section 3). Je décrirai alors le cycle de l'eau en terme de climatologies saisonnières en comparaison avec la ré-analyse de ERA-I (section 4). Enfin, la fermeture du bilan hydrique sera évaluée (section 4.2).

3 Une multitude d'observations satellites

Tous les produits satellites introduits au chapitre 3 sont utilisés ici pour rendre compte du cycle de l'eau sur la région méditerranéenne. Le nombre important de produits disponibles pour chacune des composantes du cycle doit être pris en compte pour obtenir une description la plus fiable possible du cycle.

3.1 Une grande dispersion

Dans la Figure IV.5, les climatologies saisonnières des composantes du cycle de l'eau (en ligne) sont représentées pour tous les sous-bassins méditerranéens (en colonne). Les climatologies (saisonnières) sont définies comme une saison "moyenne" sur la période commune à tous les produits satellites. La figure compare l'accord et les divergences dans les moyennes saisonnières des ensembles de données en termes de biais systématiques des produits satellites.

Pour les précipitations (première ligne), les résultats montrent un écart important entre les produits, sur tous les sous-bassins, à l'exception du LY-EG où le phénomène de mousson est similaire dans les différents produits satellites. Les produits calibrés sur des données *in situ* (GPCP et TMPA) ont une climatologie saisonnière similaire, et proche de celle de la ré-analyse (ERA-I) qui elle aussi utilise des données *in situ*. Ce n'est pas le cas des produits CMORPH, NRL et PERSIANN qui ont une saisonnalité

différente. Les différences sont plus importantes en hiver pour chaque sous-bassin ; et les climatologies saisonnières peuvent même être opposées pour des sous-bassins particuliers tels que MA-DZ-TN et Alp-IT-ADR. D'un point de vue climatologique, seuls les produits calibrés et la ré-analyse ont une saison de précipitations typiquement méditerranéenne (c'est-à-dire des précipitations plus élevées en hiver qu'en été) pour les sous-bassins côtiers et la mer Méditerranée. Les erreurs des produits satellites non-calibrés sont particulièrement importantes pour la mer Méditerranée et pour le sous-bassin MA-DZ-TN en raison de la contamination côtière lors de la restitution de la pluie sur de nombreux pixels. Pour le sous-bassin Alp-IT-ADR, les erreurs peuvent s'expliquer par des problèmes d'estimation de pluies neigeuses (Levizzani et al., 2011). Guo et al. (2015) soulignent la limitation de l'estimation des précipitations par satellite sur les régions montagneuses. La saison des précipitations sur le LY-EG est opposée à la saison sur les autres sous-bassins. Ceci est dû à la mousson à la source de ce sous-bassin dans la région tropicale près du lac Victoria.

La cohérence est bonne pour les trois produits satellites d'évapotranspiration sur terre et les deux produits d'évaporation sur mer et ce pour tous les sous-bassins. Ils sont également proches de la ré-analyse. Leurs climatologies saisonnières (deuxième ligne de la Figure IV.5) sont cohérentes avec le climat méditerranéen (c'est-à-dire une évapotranspiration plus élevée en été sur terre et une évaporation plus élevée en hiver sur la mer). La variabilité de l'évapo(transpi)ration est cohérente avec les précipitations : moins d'évapo(transpi)ration et de précipitations en Afrique du Nord que dans les Alpes. Au-dessus de la mer, OAflux est plus proche de la ré-analyse que SeaFlux car OAflux utilisent les données de la réanalyse ERA en entrée alors que SeaFlux utilise la réanalyse MERRA.

Le ruissellement (troisième ligne de la Figure IV.5) est exprimé en $mm.mois^{-1}$ (normalisé par l'air du sous-bassin) à des fins de comparaison. Comme expliqué dans le Chapitre 3, le ruissellement ici est la climatologie d'ORCHIDEE normalisée par la valeur annuelle de CEFREM. Les deux différents forçages ORCHIDEE donnent des estimations similaires pour les ruissellements. Le sous-bassin Alp-IT-ADR, défini par le ruissellement alpin, est la plus grande source d'eau douce de la Méditerranée par unité de surface (ramenée à l'aire du bassin versant). L'apport en eau par ruissellement est plus grand en hiver, ce sous-bassin hydrologique étant influencé par la fonte des neiges alpines. Les sous-bassins nord-africains ne représentent, au contraire, qu'une petite partie de l'apport fluvial à la mer, avec un ruissellement quasi nul en été. Ludwig et al. (2009) ont estimé un débit quasi nul à l'embouchure du Nil depuis la construction du barrage d'Assouan en 1961. Contrairement aux sous-bassins hydrologiques des Alpes ou des Balkans, le sous-bassin BLS ne présente pas une réduction du ruissellement en été ; cela vient de fait que les rivières de plus haute latitude, sous des climats continentaux marqués par les hivers longs, se déversent dans la mer Noire avec une plus grande constance. On peut noter une similitude particulièrement en été entre les climatologies de ruissellement du sous-bassin BLS et le ruissellement total dans Med-Sea. À cette période de l'année, le ruissellement méditerranéen est faible et seule la mer Noire connaît un apport important en eau.

La quatrième ligne de la Figure IV.5 montre la climatologie des changements mensuels du contenu en eau. On s'intéresse ici au changement mensuel et non au contenu lui-même (voir l'Eq. (III.1)). Cela permet en outre de s'émanciper des hypothèses sur la masse d'eau moyenne sur Terre, qui de toute manière n'est pas mesurée par GRACE⁴.

4. GRACE ne mesure que les anomalies par rapport au champ gravitationnel moyen et ne peut avoir accès

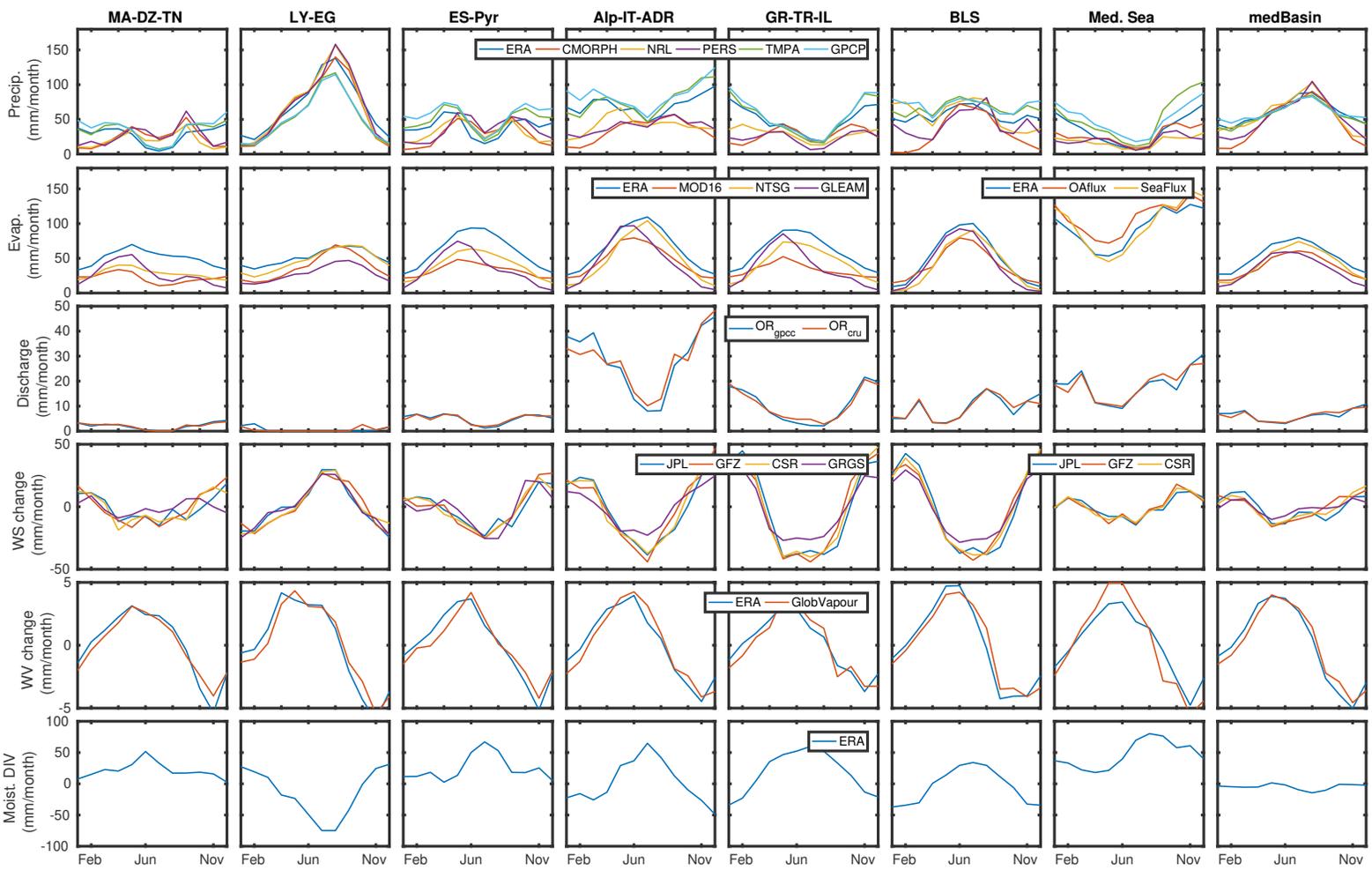


Figure IV.5. Climatologie saisonnière des composantes du cycle de l'eau (en lignes) sur les différents sous-bassins (en colonnes) dans la région méditerranéenne, à partir des observations satellitaires de la ré-analyse.

Les quatre estimations sur les terres et les trois sur la mer sont semblables pour

à cette valeur moyenne

chaque saison. Le produit GRGS se distingue davantage des autres, notamment sur le sous-bassin Alp-IT-ADR. Ceci est dû à un post-traitement différent des trois autres qui ont utilisé le même modèle hydrologique GLDAS pour l'étalonnage (Longuevergne et al., 2010). Les valeurs négatives représentent une perte de masse d'eau et les valeurs positives un gain par rapport au mois précédent. À l'échelle du sous-bassin, la similitude entre les climatologies de précipitation et celles des changements mensuels du contenu en eau illustre la recharge des sols par précipitation.

Les deux produits décrivant le changement mensuel de vapeur d'eau dans l'atmosphère sont représentés dans la cinquième ligne de la Figure IV.5. Il n'est pas surprenant que le produit Globvapor soit proche de la ré-analyse (ERA-I) sur chaque sous-bassin, car la vapeur d'eau est fortement contrainte par les observations atmosphériques dans la réanalyse ERA-I. En outre, le produit Globvapor utilise un *a priori* de profil atmosphérique issue de la réanalyse ERA-I pour la restitution de la vapeur d'eau au dessus de la mer Méditerranée ce qui explique leur faible différence pour la partie océanique.

Enfin, la climatologie de la divergence d'humidité verticalement intégrée est représentée dans la dernière ligne de la Figure IV.5. Cette composante n'est disponible ici que par ERA-I. En général, il existe une relation négative entre les précipitations et la divergence de l'humidité, en particulier pour les sous-bassins avec des précipitations montagneuses (par exemple MA-DZ-TN à travers l'Atlas, ES-Pyr à travers les Pyrénées et Alp-IT-ADR à travers les Alpes). Lorsque la divergence augmente, la zone devient plus évacuative et les précipitations diminuent. Le sous-bassin LY-EG montre une relation négative encore plus forte entre les précipitations et la divergence d'humidité en raison de la forte accumulation d'humidité lors de la mousson.

3.2 Deux méthodes simples d'utilisation des données satellites

Dans cette section, deux approches exploitant la multiplicité des produits satellites pour examiner le cycle de l'eau sont comparées : (1) une "sélection optimale" où la "meilleure" combinaison de produits est sélectionnée ; et (2) une approche par "fusion de données" qui combine tous les produits disponibles pour chaque composante du cycle de l'eau. Ces deux approches seront testées afin d'évaluer leur capacité à réduire les résidus du bilan hydrique.

3.2.a Sélection optimale (OS)

Une recherche systématique de la meilleure combinaison est effectuée : pour toutes les combinaisons possibles, les résidus du bilan hydrique sont calculés pour chaque sous-bassin en suivant l'Eq. (IV.1). La combinaison de produit avec les résidus du bilan hydrique le plus faible (sur l'ensemble des sous-bassins) est sélectionnée.

La sélection optimale est donnée par la précipitation **TMPA** ; associée à l'évapotranspiration **GLEAM** ; l'évaporation **OAFflux** ; le contenu en eau **GRGS** sur terre et **JPL** sur mer ; le ruissellement issue d'ORCHIDEE forcés par GPCC ; et la vapeur d'eau atmosphérique **Globvapor**. Un seul produit est disponible pour la divergence d'humidité. Il est à noter qu'en raison des différences importantes entre les produits de précipitations (voir la Figure IV.5), la sélection optimale dépend fortement du choix du produit

de précipitation.

3.2.b Fusion de données par pondération (Simple Weighting, SW)

Les produits satellites, et plus généralement toute estimation d'une variable par observation, présentent deux types d'incertitudes : (1) des erreurs systématiques liées, par exemple, à l'étalonnage absolu du capteur. Ceux-ci peuvent être un biais ou un facteur multiplicatif. L'utilisation de plusieurs observations avec différentes erreurs systématiques peut aider à réduire ce type d'incertitude. (2) Des erreurs aléatoires liées aux incertitudes de l'algorithme de restitution ou à des informations auxiliaires manquantes ou inexactes (par exemple la présence de nuage dans le cadre d'une restitution par observation dans l'infrarouge, voir chapitre 2). Ceux-ci sont souvent caractérisés par un écart type utilisant une hypothèse gaussienne. D'un point de vue statistique, l'utilisation d'une somme pondérée de plusieurs estimations réduit les erreurs aléatoires de l'estimation si aucune erreur de biais n'est présente dans les diverses estimations. Par exemple, si on considère n observations indépendantes caractérisées par un même écart type σ , l'estimateur lié à la moyenne des observations disponibles aura un écart type de $\frac{\sigma}{\sqrt{n}}$.

Par conséquent, la méthodologie SW, présentée par exemple dans (Aires, 2014), est d'abord basée sur une correction des biais saisonniers (pour réduire les biais systématiques à l'échelle de la saison) et est ensuite suivie d'une moyenne pondérée des estimations corrigées (pour réduire les erreurs aléatoires).

3.2.b.1 Correction de biais saisonniers

Dans cette section, on s'intéresse au bilan hydrique à partir des climatologies décrites dans la Figure IV.5. Cette analyse préliminaire indique si les estimations sont à peu près cohérentes les unes aux autres, à l'échelle de la climatologie, et si elles permettent en outre la fermeture du cycle de l'eau à cette échelle.

Afin de corriger le biais systématique des produits satellites, on vient contraindre la climatologie des différents produits décrivant chaque composante du cycle de l'eau, à avoir la même climatologie de référence sur la période commune à tous les produits d'une même composante. La climatologie de référence pour chacune des composantes du cycle de l'eau est choisie, par défaut, comme la moyenne des climatologies des différents produits décrivant la composante. On choisit ici d'assigner à chaque produit le même "poids" dans l'estimation de la référence, tant il est difficile d'obtenir une évaluation fiable des incertitudes des différents produits (par manque de données *in situ*). Pour l'ensemble des composantes du cycle de l'eau hormis la précipitation, les climatologies des différents produits se ressemblent suffisamment pour que la référence par défaut soit pertinente.

Pour les précipitations, il ne serait pas pertinent d'utiliser la moyenne de climatologies si différentes. Afin de définir une climatologie de référence pour la précipitation, on cherche la "meilleure" climatologie, au sens du bilan hydrique, parmi celles des cinq produits disponibles. La sélection est basée sur la cohérence entre la climatologie des produits de précipitations et les climatologies de référence par défaut des autres composantes du cycle de l'eau. Le bilan hydrique est estimé suivant l'Eq. (III.1) sur chaque sous-bassin à l'échelle mensuelle et ce pour chaque produit de précipitations. La RMS

des résidus du bilan hydrique est présentée dans la Table IV.1. Plus la RMS est petite, meilleure est la climatologie de précipitations au sens de la fermeture du cycle de l'eau.

	sous-bassins						LAND	SEA
	MA-DZ-TN	LY-EG	ES-Pyr	Alp-IT-ADR	GR-TR-IL	BLS		
RMS ($mm.mois^{-1}$) des résidus du bilan hydrique continental (ou océanique) à l'échelle de la climatologie								
GPCP	12.6	15.4	17.2	15.7	14.7	22.5	13.2	28.5
CMORPH	19.3	28.8	22.0	50.1	35.9	31.8	19.5	49.1
NRL	19.3	34.5	17.2	40.2	28.2	18.1	17.4	56.5
PERSIANN	18.7	33.4	17.1	42.3	35.2	21.4	17.0	57.1
TMPA	10.3	16.0	12.1	17.3	14.2	18.8	10.6	26.5
ECMWF	10.2	29.8	12.8	13.5	11.9	10.4	14.1	36.0
RMS ($mm.mois^{-1}$) des résidus du bilan hydrique atmosphérique à l'échelle de la climatologie								
GPCP	32.8	13.6	46.3	42.5	37.3	27.3	19.0	8.5
CMORPH	28.0	14.3	27.8	27.2	25.9	30.0	12.6	35.6
NRL	22.9	17.6	25.8	16.8	15.7	22.4	15.9	43.9
PERSIANN	30.3	18.5	33.3	20.3	17.4	19.8	13.7	44.6
TMPA	28.9	12.4	39.9	35.9	34.7	24.0	14.5	14.8
ECMWF	24.4	18.6	29.9	27.5	29.8	14.6	22.5	17.7

TABLE IV.1 – RMS des résidus du bilan hydrique continental ou océanique (partie supérieure) et atmosphérique (partie inférieure) sur les sous-bassins méditerranéens en utilisant la climatologie saisonnière de six produits de précipitations et la climatologie de référence pour les autres composantes du cycle de l'eau. Voir le texte pour la notion de climatologie de référence. Les résidus ont été calculés à l'échelle mensuelle. Les meilleurs résultats en terme de fermeture du cycle de l'eau sont indiqués en gras.

La Table IV.1 montre qu'aucune des combinaisons de produits satellites ne permet la fermeture du cycle de l'eau à l'échelle de la climatologie. Le résidu oscille entre 13 à 20 $mm.mois^{-1}$ sur la terre, et jusqu'à 57 $mm.mois^{-1}$ pour l'océan. Ce résidu résulte de :

- certaines hypothèses comme négliger les échanges d'eau souterraines dans l'Eq. (III.1) ;
- les incertitudes des produits satellites (Rodell et al., 2015) ;
- les erreurs associées à des données non satellites utilisées dans notre analyse (la dynamique du débit des rivières à partir des sorties du modèle ORCHIDEE ou la ré-analyse de la divergence de l'humidité de l'ECMWF (Trenberth et al., 2011, 2007)).

Le bassin océanique (mer Méditerranée + mer Noire) est le sous-bassin le plus déficitaire en eau d'après son bilan hydrique, et ce en raison de la forte incertitude qui existe pour l'écoulement net au détroit de Gibraltar (Jordà et al., 2016). Dans l'ensemble, ces résultats montrent les limites des produits satellites pour l'étude du cycle de l'eau.

La climatologie de TMPA a le résidu de bilan hydrique continental le plus faible en terme de RMS, ce qui signifie qu'il est le plus cohérent avec les climatologies de référence des autres composantes du cycle de l'eau. TMPA permet une meilleure fermeture que la ré-analyse lorsqu'on considère l'ensemble du bassin méditerranéen. Pour le bilan atmosphérique, le produit CMORPH réduit mieux les résidus à l'échelle globale (c'est-à-dire en considérant tous les sous-bassins en même temps). Le NRL semble être meilleur sur la plupart des sous-bassins pour le budget atmosphérique, mais ce

Une évaluation des produits de précipitation

Corrélation et RMS de la différence entre les produits satellites de précipitations et le produit *in situ* griddé EOBS (voir la description de ce produit au chapitre 3).

	Climatic sub-basin					
	MA-DZ-TN	ES-Pyr	Alp-IT-ADR	GR-TR-IL	BLS	Europe
Correlation						
GPCP	0.87	0.97	0.95	0.92	0.86	0.91
CMORPH	-0.07	0.38	0.12	0.16	0.53	0.03
NRL	-0.12	0.51	0.28	0.68	0.47	0.27
PERSIANN	-0.04	0.18	0.30	0.50	0.45	0.20
TMPA	0.81	0.96	0.92	0.92	0.91	0.89
ECMWF	0.85	0.94	0.91	0.90	0.94	0.89
RMSD						
GPCP	13.45	20.74	26.74	25.54	23.18	18.76
CMORPH	30.51	25.18	38.65	29.32	30.76	28.24
NRL	30.35	21.01	30.55	17.18	28.32	18.79
PERSIANN	31.77	28.93	32.76	26.21	31.39	24.75
TMPA	14.35	14.53	21.42	23.47	18.35	13.19
ECMWF	12.54	8.29	14.83	19.10	8.13	6.93

La Figure IV.5 souligne la nécessité d'évaluer les précipitations. Logiquement, les deux produits satellites calibrés sur des mesures de jauges (GPCP et TMPA) sont fortement corrélés avec les observations *in situ*; et ce plus que la ré-analyse. Dans les régions de plaine comme le sous-bassin BLS, les produits de précipitation concordent raisonnablement avec les EOBS, avec une corrélation de 0,45 pour le PERS à 0,91 pour TMPA. Cependant, sur les régions montagneuses (Alpes ou Atlas), les corrélations vont de -0,12 pour le NRL à 0,95 pour le GPCP. En moyenne sur l'Europe, la corrélation est proche de zéro (0,03) pour CMORPH et 0,91 pour GPCP. Cette première évaluation souligne l'écart entre les produits calibrés avec les jauges, et ceux qui ne le sont pas comme CMORPH. Cela souligne une limitation des produits "purement" satellites pour leur utilisation fiable dans l'étude du climat et la météorologie.

produit est éparé sur la période 2004-2009. Du fait des bons résultats de TMPA sur la plupart des sous-bassins et étant donné que le cycle de l'eau est principalement déterminé par le bilan hydrique continental, TMPA est choisi ici comme climatologie de référence pour les précipitations. Il n'est pas surprenant que la meilleure solution choisie ici soit un des produits calibrés sur des données de jauges. En outre, l'encadré sur l'évaluation des produits de précipitation suggère la relative bonne performance de ce produit par rapport aux données *in situ*

3.2.b.2 La moyenne pondérée

La méthodologie SW utilise la multiplicité des estimations de chaque composante du cycle de l'eau pour réduire les erreurs aléatoires. Considérons les p observations de précipitations P_i associées à des erreurs ϵ_i qui suivent une distribution gaussienne $N(0, \sigma_i)$, où σ_i est l'écart-type de l'estimation P_i . L'estimation des précipitations issue de SW est donnée par la moyenne pondérée :

$$P_{SW} = \frac{1}{p-1} \sum_{i=1}^p \frac{\sum_{k \neq i} (\sigma_k)^2}{\sum_k (\sigma_k)^2} P_i. \quad (IV.2)$$

Cette équation est valide lorsqu'il n'y a pas d'erreur de biais dans les P_i (d'où notre correction préliminaire du biais saisonnier) et est optimale lorsque les erreurs ϵ_i sont statistiquement indépendantes les unes des autres. Cette expression est valable pour les autres composantes du cycle de l'eau.

Si les p erreurs sont indépendantes les unes des autres, la variance de l'estimation P_{SW} est donnée par :

$$var(P_{SW}) = \frac{1}{(p-1)^2} \sum_{i=1}^p \left(\frac{\sum_{k \neq i} (\sigma_k)^2}{\sum_k (\sigma_k)^2} \right)^2 \sigma_i^2. \quad (IV.3)$$

Il s'agit d'une information importante qui donne l'incertitude de l'estimation de l'Eq. (IV.2). Dans le cas simple où $\sigma_i = \sigma$ pour $i = 1, \dots, p$, la variance devient :

$$var(P_{SW}) = \frac{\sigma^2}{p}. \quad (IV.4)$$

Plus le nombre d'observations p est élevé, plus l'incertitude de l'estimateur SW de l'Eq. (IV.2) est faible.

Ce cas simple montre comment les erreurs P_{SW} peuvent être réduites de manière significative en utilisant la multiplicité des observations P_i .

Dans cette thèse, on estime chaque composante du cycle de l'eau, à l'aide de la méthodologie SW.

3.2.c Comparaison des méthodes OS et SW

Les résidus des bilans hydriques continentaux, océaniques et atmosphériques de la région méditerranéenne sont calculés à l'échelle du mois et du sous-bassin. Des statistiques sont présentés dans la Table IV.2. Outre les deux méthodes, la table montre aussi les résultats obtenus à partir d'une combinaison de produits de ré-analyse pour toutes les composantes du cycle de l'eau exceptés le ruissellement et le contenu en eau.

En moyenne, sur l'ensemble du bassin méditerranéen, la méthodologie SW montre une RMS mensuelle des résidus du bilans hydrique inférieure à celle issue de la sélection optimale, non seulement sur les continents mais aussi sur l'océan et l'atmosphère. SW utilise la multiplicité des produits satellites disponibles pour réduire les erreurs aléatoires d'estimation de chaque composante qui impactent le bilans hydrique (voir l'Eq. (IV.2)). La valeur efficace très élevée des résidus du bilan océanique est due à l'erreur sur le débit net de Gibraltar. En effet, Jordà et al. (2016) souligne qu'une forte anomalie (décembre 2004-juin 2005) a un impact sur la reconstruction de l'écoulement net ce qui augmente les résidus du bilan hydrique sur la même période.

Les produits de ré-analyses tels que ERA-I sont connus pour ne pas fermer le cycle de l'eau. Ceci est dû à l'absence de contrainte de conservation de l'eau pendant le processus d'assimilation lors de la genèse des produits ERA-I mais aussi aux erreurs de certaines sources de données extérieures à la ré-analyse (le ruissellement et le changement de contenu en eau). Les résultats de l'ERA-I sont meilleurs que ceux de SW pour l'atmosphère, mais ils sont pires pour les bilans hydriques continentaux

et océaniques. Hirschi et al. (2006) estiment le contenu en eau dans le sol sur certains bassins fluviaux méditerranéens (Danube, Wisla, Odra, Elbe, Elbe, Weser, Rhin, Rhône, Pô, Seine, Loire, Garonne, Ebre) sur la base des bilans hydriques atmosphériques et continentaux. La RMSD entre leurs estimations et les données de GRACE oscille entre 3 à 39 $mm.mois^{-1}$ sur les différents bassins considérés ce qui est similaire à nos résultats (de 16 à 25 $mm.mois^{-1}$ sur terre).

		MA-DZ-TN	LY-EG	ES-Pyr	Alp-IT-ADR	GR-TR-IL	BLS	LAND	SEA
Statistique sur les résidus du bilan hydrique <i>continental ou océanique</i> ($mm.mois^{-1}$)									
ERA-I	mean	-23.8	16.8	-25.7	-21.1	-20.0	-1.7	2.1	-52.5
	std	24.8	19.7	27.9	23.1	23.3	18.0	13.6	69.6
	rms	34.3	25.9	37.8	31.2	30.6	18.0	13.6	86.7
OS	mean	3.2	23.6	10.0	3.5	6.2	17.4	16.8	-5.3
	std	25.1	14.4	25.7	28.5	25.3	18.5	10.6	75.7
	rms	25.1	27.6	27.5	28.5	25.8	25.4	19.8	75.2
SW	mean	4.8	15.0	8.1	1.9	4.8	18.2	13.8	-10.6
	std	17.7	13.9	15.6	21.5	17.0	17.4	9.4	74.1
	rms	18.2	20.5	17.5	21.5	17.6	25.1	16.6	74.3
Statistique sur les résidus du bilan hydrique <i>atmosphérique</i> ($mm.mois^{-1}$)									
ERA-I	mean	0.9	-2.3	-3.2	-9.4	-7.42	-2.5	-3.2	12.6
	std	15.4	7.6	18.0	10.0	9.5	7.7	5.5	5.8
	rms	15.3	8.0	18.1	13.7	12.0	8.0	6.2	13.8
OS	mean	-31.6	0.1	-40.3	-32.1	-37.7	-21.9	-14.0	9.8
	std	17.3	18.7	16.6	19.9	12.3	16.3	5.8	22.9
	rms	36.0	18.6	43.5	37.7	39.7	27.3	15.1	24.7
SW	mean	-27.8	-0.7	-36.8	-32.5	-32.5	-22.6	-14.9	5.3
	std	15.5	14.3	17.5	20.5	14.5	13.9	7.3	14.8
	rms	31.8	14.2	40.7	38.3	35.6	26.5	16.6	15.7

TABLE IV.2 – Statistiques sur les résidus mensuels des bilans hydriques continentaux, océaniques et atmosphériques, calculées sur les sous-bassins méditerranéens pour la période 2004-2009, et pour les deux méthodes : OS et SW. Les résidus des bilans hydriques pour la ré-analyse de ERA-I sont également montrés.

4 Le cycle de l'eau - premier aperçu

Dans cette section, le cycle de l'eau du bassin méditerranéen est décrit à l'aide des estimations de chaque composante du cycle via la méthode de fusion de donnée SW. Cette approche a été choisie, au terme de la comparaison préliminaire avec OS, car elle permet de mieux réduire les résidus des bilans hydriques.

4.1 La saisonnalité des composantes du cycle

Les moyennes mensuelles des composantes du cycle de l'eau sont représentées dans les Figures IV.6 et IV.8 (en lignes), sur la période 1980-2012, à l'exception du contenu en eau (GRACE n'est disponible que depuis 2002) et de l'estimation de l'écoulement (disponible entre 1980 et 2009). Ils sont représentés pour janvier, avril, juillet et octobre (en colonnes). Pour les deux termes de flux principaux (précipitations et

évapo(transpi)ration), la climatologie de ERA-I est également représentée à des fins de comparaison.

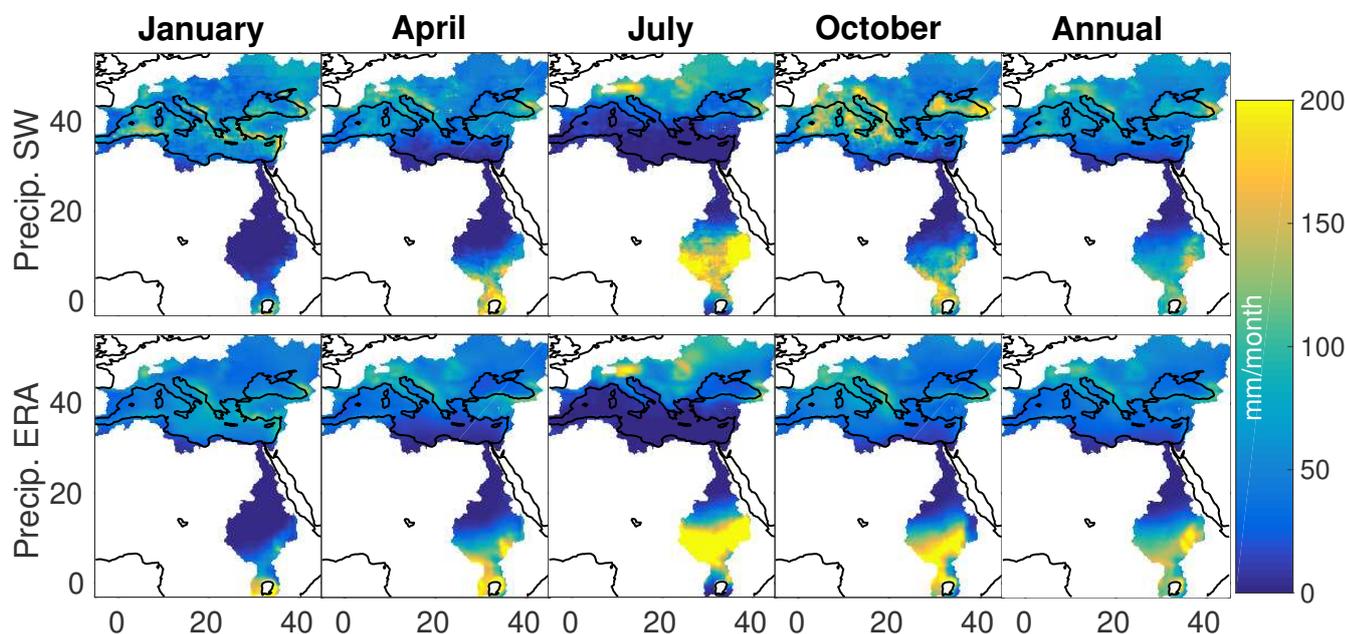


Figure IV.6. Moyenne mensuelle pour la précipitation via SW (première rangée) et ERA-I (deuxième rangée). De gauche à droite, pour : Janvier, avril, juillet et octobre. La dernière colonne représente la moyenne annuelle.

Précipitation - La comparaison des moyennes de précipitations mensuelles et annuelles est présentée dans la Figure IV.6. Les estimations de SW sont cohérentes avec ERA-I. À l'échelle annuelle, les précipitations sont principalement caractérisées par un gradient latitudinal (le gradient méridien) depuis la côte africaine sèche (moins de 22 mm.mois^{-1} en moyenne sur l'année) jusqu'aux terres plus humides des latitudes moyennes, en particulier au-dessus des montagnes (Alpes, Balkans et Caucase) où les précipitations peuvent atteindre 160 mm.mois^{-1} en moyenne sur l'année. Les précipitations les plus élevées se trouvent dans la zone équatoriale au-dessus des montagnes éthiopiennes (200 mm.mois^{-1}).

À l'échelle mensuelle, le pourtour de la Méditerranée se caractérise par des hivers humides et des étés secs. Les précipitations ne se produisent en Afrique du Nord que pendant la saison des pluies (octobre à mars), suivant un climat méditerranéen standard. Durant cette saison, l'ouest est pluvieux plus tôt que l'est en raison de la dépression longitudinale de l'Atlantique. Au printemps et en été, les précipitations ne se produisent qu'au-dessus des montagnes et dans les hautes latitudes en Europe centrale (climat continental). C'est à cette saison que le gradient méridien à travers la région méditerranéenne est le plus marqué. La saison des pluies est opposée dans la région tropicale qui est sujette à la mousson, où l'on peut observer un hiver sec (les précipitations sont alors situées uniquement dans la région équatoriale) et un été humide (plus de 200 mm.mois^{-1}) entre 0 et 15°N . La partie supérieure du bassin du Nil, lorsque le fleuve traverse le Sahara, est marquée par une absence totale de précipitation tout au long de l'année. À n'importe quelle échelle temporelle, les précipitations sont plus abondantes en moyenne dans le produit de ré-analyse ERA-I dans les zones tropicales et dans le sud-ouest des Alpes. La plus grande différence entre les estima-

tions concerne la mer Noire à l'automne, où les précipitations sont bien plus faibles dans le produit ERA-I.

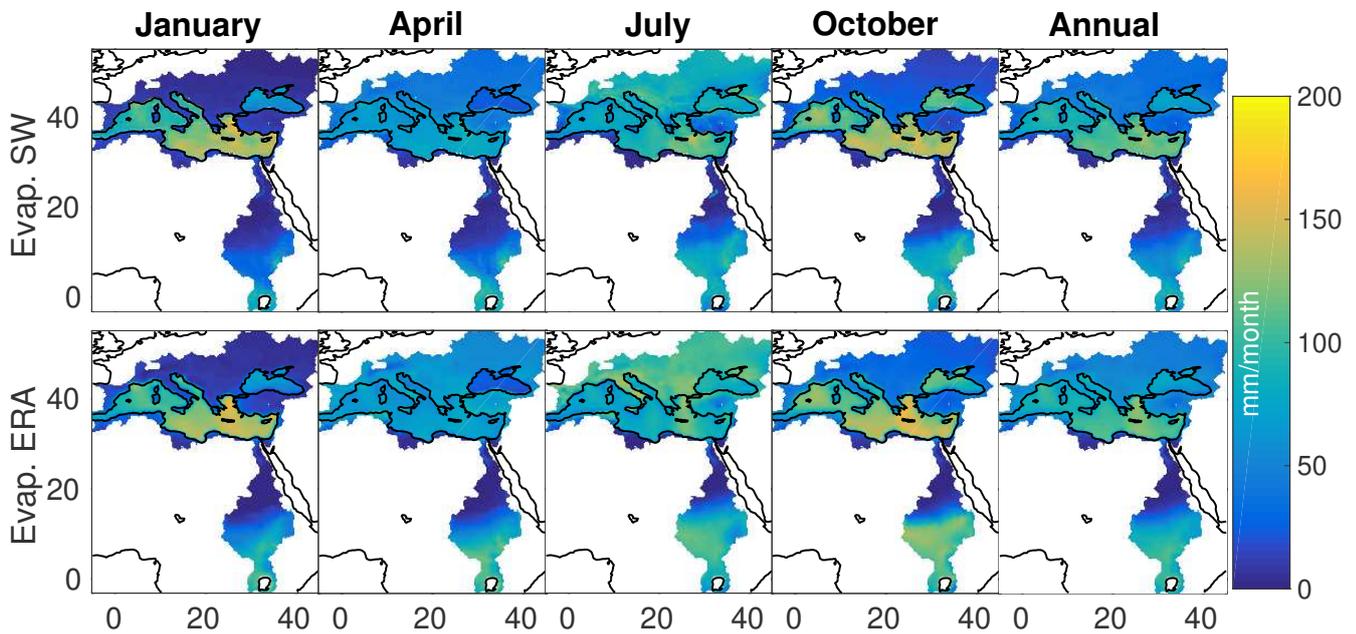


Figure IV.7. Figure IV.6 mais pour l'évapotranspiration

Evapo(transpiration) - Le panneau inférieur de la Figure IV.7 compare la moyenne mensuelle et annuelle du taux d'évapo(transpi)ration dans SW et ERA-I. Les deux estimations sont en accord, avec des valeurs légèrement plus élevées pour ERA-I. La similarité est due au fait que les algorithmes de restitution de l'évapo(transpi)ration utilisent des informations auxiliaires (par exemple la température de surface) et des modèles d'interface surface-atmosphère (Martens et al., 2016) similaires à ceux qui sont utilisés dans la ré-analyse (voir Chapitre 3).

Les deux processus physiques derrière l'évapotranspiration sur terre et l'évaporation sur mer sont très différents. Ce qui explique le grand contraste entre ces deux variables. À l'échelle de l'année, la mer Méditerranée présente en moyenne un taux d'évaporation bien plus élevé (environ 120 mm.mois^{-1} en moyenne sur l'année) que l'évapotranspiration sur terre (environ 60 mm.mois^{-1} sur l'année pour l'Europe des latitudes moyennes et 90 mm.mois^{-1} pour les zones équatoriales). Les valeurs les plus élevées sont observées dans la partie orientale de la mer.

À l'échelle mensuelle, les caractéristiques spatiales de l'évapotranspiration continentale suivent celles des précipitations dans la région tropicale : là où le rayonnement de surface est élevé et où le manque d'eau est donc le principal facteur limitant de l'évapotranspiration (en termes de quantité d'eau évacuée dans l'atmosphère). L'évapotranspiration estivale est quasi nulle à partir de la côte africaine et en descendant jusqu'à l'Équateur. En Europe, l'évapotranspiration estivale augmente au nord du bassin, vers l'Europe centrale et orientale. Dans les deux estimations présentées dans la Figure IV.6, avril est le mois qui présente le taux d'évapotranspiration le plus faible. Juillet est le seul mois où l'évapotranspiration peut dépasser l'évaporation sur mer.

Sur mer, l'évaporation méditerranéenne est plus intense en hiver, principalement en raison des vents plus forts et de l'air plus sec dans l'atmosphère. L'évaporation hivernale présente un gradient longitudinal très marqué. Ce gradient souligne la différence

des deux bassins dans la circulation thermohaline de la mer Méditerranée. L'évaporation très active dans le bassin oriental densifie les eaux de surface provenant du bassin occidental qui plongent pour constituer des eaux plus profondes avant de ressortir par le détroit de Gibraltar. L'évaporation estivale est inférieure à celle de l'hiver sur la majeure partie du bassin.

L'évaporation à la surface la mer Méditerranée représente une grande source de chaleur pour toute la région. La mer Méditerranée est considérée comme le réservoir de chaleur de toute la région méditerranéenne (Lionello et al., 2006).

Ruissellement - Les moyennes mensuelles et annuelles du ruissellement sont présentées dans la Figure IV.8 (première ligne). Le ruissellement total est ici ramené à l'échelle du sous-bassin, à des fins de représentation. Comme les caractéristiques spatiales du ruissellement sont difficilement observables, l'interprétation de la climatologie du ruissellement dans le bassin, se résume aux commentaires de la Figure IV.5.

L'eau douce continentale provient principalement d'Europe, de 15 mm.moiss^{-1} en moyenne dans le sous-bassin Est (GR-TR-IL) à 30 mm.moiss^{-1} dans le sous-bassin Alp-IT-ADR. Ce sous-bassin fournit la majeure partie de l'eau douce continentale à la mer Méditerranée. Presque aucun ruissellement ne provient de la côte africaine. À l'échelle mensuelle, le ruissellement est plus élevé pendant la saison des pluies (d'octobre à février) en raison du forçage des précipitations (qui comprend la neige et la fonte des neiges en plus de la pluie).

Variations du contenu en eau - Le produit SW (Figure IV.8, deuxième ligne) est basé sur deux estimations indépendantes sur terre et sur mer. Afin d'éviter la contamination terre-mer, le traitement des données satellites (filtrage côtier) par le "GRACE Ground System" se limite à quelques "pixels" sur la Méditerranée et la mer Noire, du fait de la faible résolution de l'instrument. La Figure IV.8 montre la variation mensuelle du contenu en eau. Les valeurs négatives représentent une perte en eau par les sols ou sur mer et les valeurs positives un gain par rapport au mois précédent. À l'échelle annuelle, les faibles valeurs de changement du contenu en eau montrent que l'hypothèse de variation nulle annuelle est raisonnable pour la région méditerranéenne entre 2002 et 2012.

Cependant, les changements de contenu en eau ne sont pas négligeables pour le bilan hydrique à l'échelle mensuelle et ne peuvent donc pas être négligés dans notre analyse mensuelle.

Pendant l'hiver et au début du printemps, après de fortes précipitations, les sols continentaux européens sont rechargés et stockent une grande quantité d'eau (jusqu'à $100 \text{ mm.moiss}^{-1}$ de plus que le mois précédent). Au contraire, dans la zone tropicale, une perte d'eau peut être observée pour le même mois (moins 50 à $100 \text{ mm.moiss}^{-1}$) illustrant l'assèchement de la région après la mousson. Au pic de la mousson en été, la zone subsaharienne se recharge (de plus de $100 \text{ mm.moiss}^{-1}$) et particulièrement les hauteurs du plateau éthiopien. Pendant cette période estivale, le continent européen libère son eau par le prélèvement naturel de plantes et le pompage pour l'agriculture. L'évapotranspiration augmente ainsi que la fonte des neiges. Tous ces processus réduisent le stock d'eau contenu dans le sol. Pour la mer, les changements de niveau d'eau sont de moindre importance mais nous pouvons encore détecter des variations saisonnières avec des gains en eau plus élevés en hiver et automne qu'en été. Ces différences peuvent être reliées à la précipitation plus grande sur mer en hiver et un

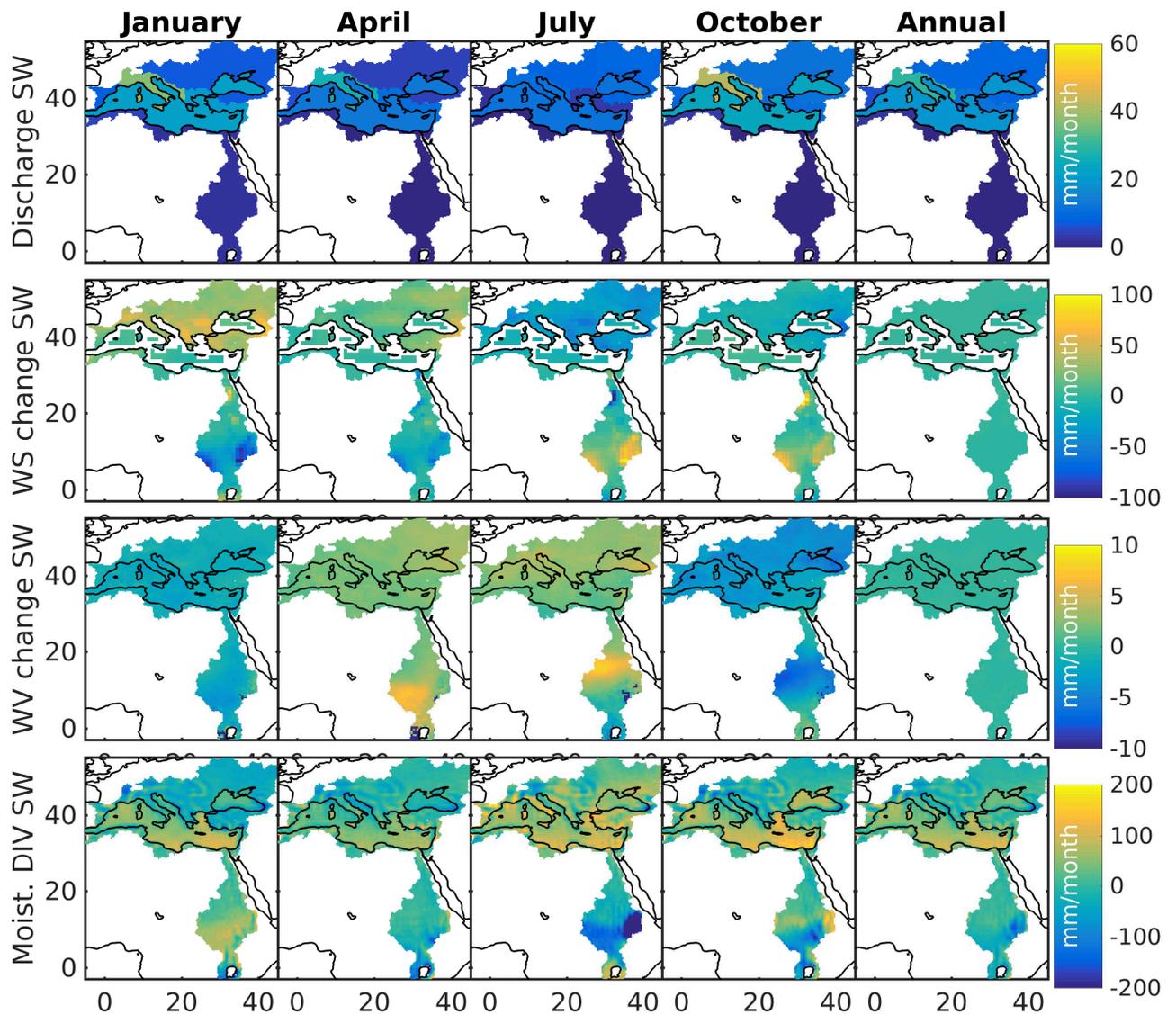


Figure IV.8. De haut en bas, les moyennes mensuelles pour : le ruissellement, les variations du contenu en eau sur terre et sur mer, les variations de la vapeur d'eau atmosphérique et la divergence de l'humidité estimés via SW. De gauche à droite, pour : Janvier, avril, juillet et octobre. La dernière colonne représente la moyenne annuelle.

apport en eau continental plus important à cette même saison.

Il est également intéressant d'observer la rétention d'eau par le barrage d'Assouan sur le Nil près de son embouchure suivie d'une libération d'eau en juillet, certainement pour l'irrigation.

Variation de vapeur d'eau atmosphérique - La variation de la vapeur d'eau atmosphérique est représentée dans la troisième ligne de la Figure IV.8. Les valeurs né-

gatives représentent une perte de vapeur d'eau atmosphérique et les valeurs positives un gain par rapport au mois précédent. Le changement de la vapeur d'eau atmosphérique a été négligé dans la plupart des études précédentes. À l'échelle annuelle, c'est raisonnable compte tenu des faibles valeurs observées. Cependant, il n'est pas optimal de négliger les changements à l'échelle mensuelle avec des écarts jusqu'à $\pm 10 \text{ mm.mois}^{-1}$.

La vapeur d'eau augmente sur l'ensemble de la région de janvier à juillet (un gain mensuel d'environ 7 mm.mois^{-1} sur l'Europe et 10 mm.mois^{-1} sur le sud de LY-EG en été), suivi d'une baisse en octobre. La configuration spatiale est différente au nord ou au sud de la Méditerranée : dans le nord, le changement en vapeur d'eau suit les caractéristiques d'évapotranspiration continentale, alors que dans la zone tropicale soumise à la mousson au sud, les caractéristiques spatiales de la vapeur d'eau atmosphérique sont en opposition de phase par rapport aux patterns de précipitation.

Divergence d'humidité - La dernière ligne de la Figure IV.8 montre la divergence d'humidité verticalement intégrée. Les valeurs positives correspondent à un déplacement d'humidité hors du pixel (évacuation) tandis que les valeurs négatives correspondent à une circulation vers le pixel (accumulation). La circulation atmosphérique est fortement influencée par la topographie complexe de la terre entourant la Méditerranée qui joue un rôle crucial dans la direction du flux d'air. En moyenne, la mer Méditerranée évacue l'humidité de la mer vers la terre, tout au long de l'année (environ 100 mm.mois^{-1} de valeurs positives à l'échelle annuelle). Au contraire, l'accumulation d'humidité peut être observée au-dessus des montagnes (Alpes, Caucase, montagnes éthiopiennes). La mer Méditerranée est véritablement le réservoir d'humidité de toute la région méditerranéenne.

À l'échelle mensuelle, la divergence d'humidité saisonnière sur le sous-bassin LY-EG est influencée par la Zone de Convergence InterTropicale (ZCIT). Cette zone définit une zone de forte accumulation d'humidité - une zone proche de l'équateur en hiver mais qui se déplace vers 5°N en été. La ZCIT est antérieure aux précipitations de mousson de l'Afrique de l'ouest. À l'exception des montagnes qui sont toujours des zones de convergence de l'humidité, l'Europe est principalement une région d'accumulation en hiver et une région d'évacuation en été. Cette variabilité peut être due à plusieurs variables de couplage terre-atmosphère telles que la température à la surface de la terre et l'humidité du sol (Seneviratne et al., 2006) ou comme la température à la surface de la mer (Li, 2006).

Écoulement à Gibraltar - La Figure IV.9 donne la climatologie saisonnière de l'écoulement net à Gibraltar. L'écoulement montre une nette saisonnalité, l'eau s'écoulant principalement de la mer Méditerranée vers l'Atlantique en hiver (débit net nul ou négatif). Ce mécanisme est entraîné par la différence de pression qui pousse les eaux profondes lourdes, froides et salées de la Méditerranée vers l'ouest dans l'Atlantique. En été, l'augmentation des flux de chaleur a un impact sur la densité de l'eau de surface, puis sur le débit entrant (alors que la densité du débit sortant est supposée constante). Cela augmente la différence de densité et renforce l'échange d'eau de l'Atlantique vers la Méditerranée (Soto-Navarro et al., 2010).

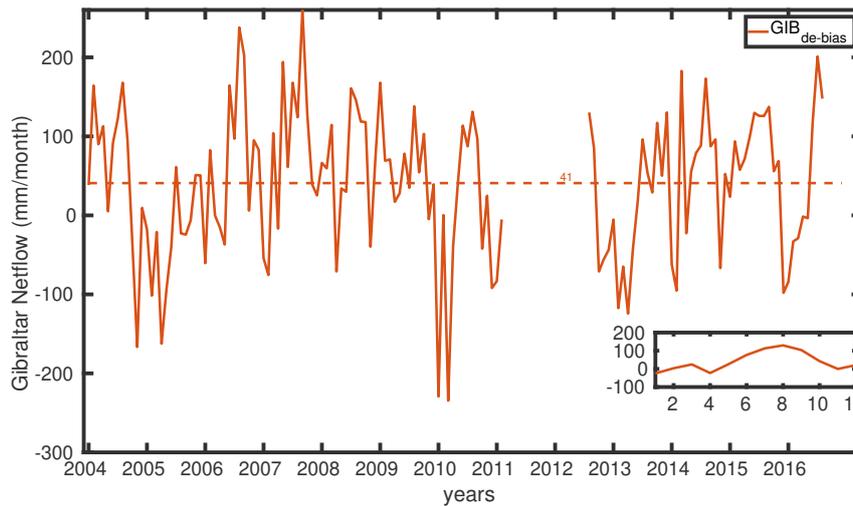


Figure IV.9. Écoulement net à Gibraltar après l'ajustement de biais (Jordà et al., 2017) entre 2004 et 2016 et sa climatologie saisonnière en $mm.mois^{-1}$. L'aire totale de la Méditerranée avec la mer Noire est de $3.5 \cdot 10^6 \text{ km}^2$.

4.2 Le cycle de l'eau

La Figure IV.10 montre le bilan hydrique annuel vu par les produits satellites à l'échelle des sous-bassins. Ne sont représentés que les composantes de flux car à l'échelle de l'année, les variations des composantes de stock en eau (dans le sol, sur mer et dans l'atmosphère) sont négligeables. À l'échelle du bassin méditerranéen total, les résidus du bilan hydrique continental et atmosphérique mettent en évidence une incohérence entre les produits de précipitation et ceux d'évapotranspiration, la partie continentale étant excédentaire et l'atmosphère déficitaire d'une même quantité d'eau (environ 14 mm.mois^{-1}). Les moyennes des autres flux du cycle de l'eau sont bien plus faibles. L'opposition de signe entre les résidus des bilans hydriques de surface (positif sur les sous-bassins continentaux et négatif pour la partie océanique) et ceux de l'atmosphère (négatif au-dessus de la terre et positif au-dessus des mers) à l'échelle des sous bassins est un bon indicateur d'incohérence entre les deux flux majeurs que sont les précipitations et l'évapotranspiration. Néanmoins, déficit et excédent ne s'équilibrent qu'à l'échelle du bassin totale. Alors que l'excédent est quasi nul pour le bilan hydrique du sous-bassin Alp-IT-ADR, le déficit en eau dans l'atmosphère atteint la moitié des précipitations annuelles ($35.1 \text{ mm.mois}^{-1}$) pour ce sous-bassins. Au contraire, le sous-bassin LY-EG présente un bilan atmosphérique plutôt équilibré alors que son excédent en eau à la surface est très grand (15 mm.mois^{-1}). Globalement les petits sous-bassins côtiers présentent un large déficit dans l'atmosphère ce qui pourrait être dû à l'estimation de la divergence d'humidité sur les côtes ou bien des erreurs de contamination terre/mer dans les précipitations. La variabilité spatiale des résidus (excédents/déficits) des bilans hydriques d'un sous-bassin à l'autre souligne l'intérêt de suivre le cycle de l'eau à cette échelle du sous-bassin. Si à grande échelle les deux flux majeurs, précipitation et évaporation, pourraient expliquer seuls les erreurs, à une échelle plus fine, les autres composantes du cycle de l'eau et leurs incertitudes respectives doivent être prises en compte.

Dans l'ensemble, la saisonnalité du cycle de l'eau via SW concorde bien avec la

littérature ayant analysé la région méditerranéenne. Par exemple : Rodell et al. (2015) ont trouvé les flux annuels moyens suivants : $P=48 \text{ mm.mois}^{-1}$, $E=118 \text{ mm.mois}^{-1}$ et $DIV=73.7 \text{ mm.mois}^{-1}$ au-dessus de la mer Méditerranée ; et $P=74.8$, $E=79.6$ $DIV=43 \text{ mm.mois}^{-1}$ au-dessus de la mer Noire (alors que notre étude considère les deux mers à la fois). Rodell et al. (2015) ont trouvé des résidus inférieurs pour le bilan hydrique continental avec $Res_l= 3.1 \text{ mm.mois}^{-1}$ pour l'ensemble du continent eurasien, et supérieurs pour le bilan océanique ($Res_o= 5.9 \text{ mm.mois}^{-1}$).

En utilisant la ré-analyse et des données GRACE, Trenberth et Fasullo (2013) ont trouvé un budget terrestre résiduel $Res_l= 4.6 \text{ mm.mois}^{-1}$ pour le continent eurasien.

Au-delà de ces similitudes, cette étude montre la saisonnalité de la durée de stockage dans le sol et dans l'atmosphère. Ces quantités sont souvent négligées alors que leur caractère saisonnier contribue à la compréhension globale des échanges d'eau. De plus, négliger le terme de stockage pour le sous-bassin BLS (jusqu'à $\pm 50 \text{ mm.mois}^{-1}$) conduirait à une erreur de même amplitude, équivalente à la moitié des précipitations pour ce sous-bassin.

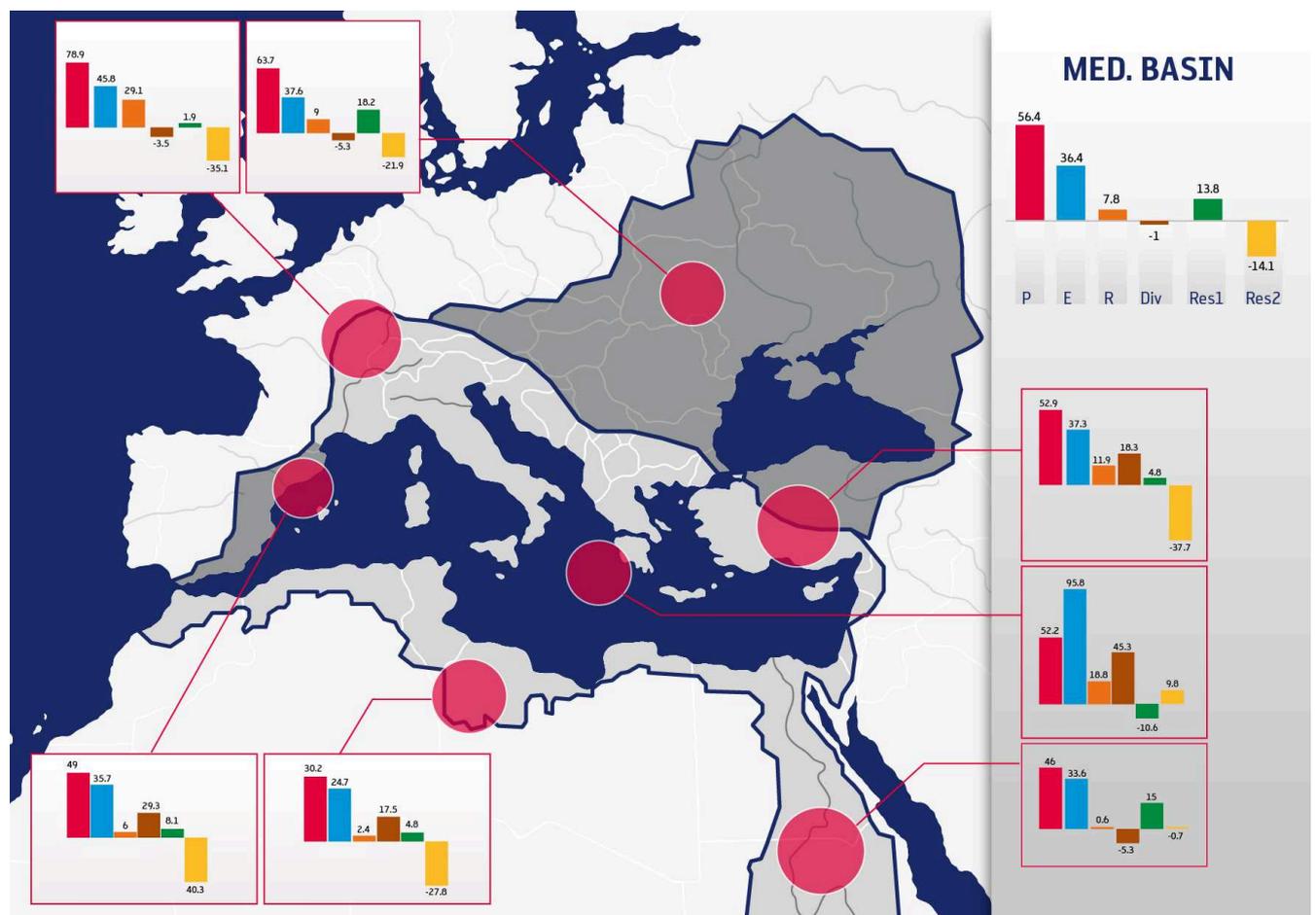


Figure IV.10. Cycle de l'eau annuel moyen vu par les produits satellites pour les six sous-bassins, la partie océanique et l'ensemble du bassin méditerranéen : précipitations (rouge), évapotranspiration (bleu), ruissellement (rose), divergence (orange), résidus du bilan hydrique au niveau de la surface (vert) et dans l'atmosphère (jaune). Les moyennes annuelles sont exprimées en mm.mois^{-1} .

4.3 Constat sur le cycle de l'eau en Méditerranée vu par les satellites

Ce chapitre est centré sur le suivi du cycle de l'eau dans la région méditerranéenne, à l'échelle du bassin et du sous-bassin. Contrairement à la plupart des études précédentes : (1) l'analyse utilise principalement des observations satellites ; (2) tient compte de la totalité de la zone de drainage méditerranéenne et le domaine d'étude est étendu à la mer Noire à l'Est et jusqu'au lac Victoria au Sud ; et (3) le cycle de l'eau a été analysé non seulement sur le continent mais aussi sur la mer et dans l'atmosphère. Par rapport à d'autres études antérieures sur le cycle de l'eau et faites à l'échelle du globe (Rodell et al., 2015 ; Trenberth et Fasullo, 2013), notre analyse est faite à l'échelle du mois et du sous-bassin.

L'observation du cycle de l'eau méditerranéen par satellites est un défi en raison de la topographie complexe de la région (zones côtières et montagnes) et des processus complexes (système karstique pour l'hydrologie continentale et échange surface/atmosphère au-dessus de la mer). Le nombre de produits satellites pour l'observation de la Terre augmente. La plupart des composantes du cycle de l'eau peuvent maintenant être suivies depuis l'espace ; et pour chaque composante, plusieurs produits sont disponibles.

Cependant, ces estimations sont corrompues par des biais systématiques et des incertitudes dans les mesures et/ou algorithmes de restitution. Par exemple Adler et al. (2003) considère que les valeurs climatologiques moyennes des précipitations au-dessus de l'océan sont encore sujettes à discussion, en particulier aux latitudes moyennes et élevées. L'utilisation de produits satellites doit alors être optimisée pour réduire ces erreurs. De plus, les produits satellites ne sont pas cohérents entre eux (ni pour une même composante ni entre les composantes du cycle de l'eau), ce qui limite leur utilisation pour ce type d'étude.

Deux méthodologies ont été testées ici pour traiter cette multiplicité de données : la "sélection optimale" et la "pondération simple" (SW). Leur comparaison montre que la procédure de fusion SW fournit une meilleure estimation pour les composantes du cycle de l'eau et son suivi :

- SW gère mieux les incertitudes, les erreurs systématiques et aléatoires ;
- il permet une réduction des résidus du bilan hydrique ;
- les composantes du cycle de l'eau obtenues sont plus cohérentes entre elles ;
- une comparaison effectuée avec la ré-analyse ERA-I, la référence pour l'étude du cycle de l'eau sur la région méditerranéenne, montre que la saisonnalité des composantes est similaire pour les deux estimations.

De nombreuses perspectives intéressantes quant au suivi du cycle de l'eau ont été ouvertes par l'approche SW. Cependant, la fermeture du bilan hydrique n'est toujours pas satisfaisante, ce qui souligne la nécessité de mettre en place des techniques de fusion de données plus sophistiquées telles que l'assimilation (Pan et Wood, 2006) ou l'intégration (Aires, 2014 ; Munier et al., 2014 ; Rodell et al., 2015). Le chapitre suivant est dédié à la présentation d'une méthodologie d'optimisation des produits satellites basée sur une contrainte d'intégration hydrologique : la fermeture du cycle de l'eau.

Chapitre V

Optimisation de données par fermeture du cycle de l'eau

Où l'on montre que l'utilisation de la contrainte de fermeture du cycle de l'eau dans l'estimateur Bayésien permet d'obtenir des estimations plus cohérentes des composantes du cycle. Le cadre théorique et de nouveaux développements de l'intégration de la contrainte hydrologique sont introduits. Et où une description globale et cohérente du cycle de l'eau méditerranéen est obtenue avec les produits satellites intégrés. Ce chapitre s'appuie sur (Pellet et al., 2018b).

1 État de l'art de l'intégration sous contrainte des produits satellites pour l'étude du cycle de l'eau

Dans le chapitre précédent, issu de (Pellet et al., 2018), les données satellites sont utilisées pour suivre le cycle de l'eau dans la région méditerranéenne, mais il a été démontré que la fermeture du bilan hydrique n'est pas obtenue avec les produits satellites.

Je rappelle ci-après l'expression mathématique du cycle de l'eau. Les principaux échanges d'eau entre l'atmosphère et la surface sont l'évapo(transpi)ration et les précipitations. Le budget terrestre est lié au budget océanique par le biais de l'écoulement fluvial (égalité dans les échanges d'eau douce). Les termes de stockage sur le continent (S_l), dans la mer (S_o) et l'atmosphère (W) varient alors au cours du temps :

$$\begin{aligned}\frac{\delta S_l}{\delta t} &= P_l - E_l - R_l \quad (\text{Terrestrial}) \\ \frac{\delta S_o}{\delta t} &= P_o - E_o + R_l^* - R_o \quad (\text{Oceanic}) \\ \frac{\delta W_{l/o}}{\delta t} &= E_{l/o} - P_{l/o} - \text{div}(\overrightarrow{Q_{l/o}}), \quad (\text{Atmospheric})\end{aligned}\tag{V.1}$$

où l'indice l est pour la partie continentale (land) et o pour l'océan. P est la précipitation ; E le taux d'évapo(transpi)ration ; S le stockage en eau ; R_l ou R_l^* le ruissellement en eau douce de la terre vers l'océan (avec $Area_{Med.Sea} \cdot R_l^* = Area_{Land} \cdot R_l$) ; R_o

l'écoulement net à travers le détroit de Gibraltar ; W le contenu total en vapeur d'eau atmosphérique ; et $div(\overrightarrow{Q_{I/O}})$ la divergence d'humidité.

Les trois relations dans l'Eq. (V.1) sont en $mm.mois^{-1}$. Elles ont été normalisées par la surface pour des raisons de commodité et afin de comparer directement les résultats sur des sous-bassins de tailles diverses.

On rappelle que l'utilisation des produits satellites pour le suivi du cycle de l'eau reste un défi pour plusieurs raisons : (1) les incertitudes inhérentes aux observations satellites (erreurs systématiques et aléatoires), (2) la multiplicité des produits pour la même variable géophysique, (3) l'incohérence entre les différents produits satellites (pour la même variable ou bien entre les différentes composantes du cycle de l'eau). La combinaison optimale de produits pour étudier le cycle de l'eau a récemment fait l'objet de beaucoup d'attention dans la littérature. Les caractéristiques de certaines méthodes d'intégration sont synthétisées dans la Table V.1 pour une meilleure comparaison des approches considérées dans ce qui suit.

Dans cette section, différentes méthodologies d'intégration sont présentées dans le contexte de l'utilisation de contraintes hydrologiques. Ces méthodes d'intégration sont décrites ici dans le cas simple d'erreur gaussienne et d'opérateur linéaire. Mais il faut noter que leurs implementations necessite souvent des developpement complexe pour gerer les cas d'operateur non linéaire et des erreur de différente nature. ceci est vrai en particulier dans le cas du filtre de Kalman (Section 1.2) et le 3D-VAR (Section 1.3) toute deux developpées en opérationelle. Les notations utilisées dans cette section sont cohérentes mais différeront de la Section 2. Outre les papiers référents aux diverses techniques de contraintes hydrologiques, cette section est également basée sur la présentation de Bonan (2014).

1.1 L'estimateur optimal

Dans cette section, je rappelle d'abord le cadre théorique de l'estimateur optimal à partir de l'inversion Bayésienne et introduit les notations. Je présente ensuite trois exemples d'estimateur optimal avec une ou plusieurs contrainte(s), dans le contexte de la fermeture du cycle de l'eau, tous basés sur la théorie Bayésienne.

1.1.a L'inversion Bayésienne

L'inversion Bayésienne consiste à utiliser une observation indirecte y afin d'inférer une certaine information sur l'état d'un système x , connaissant le lien H de x à y .

Considérons le vecteur d'état x comme une variable aléatoire. Le vecteur observation y est lié à x par la relation suivante :

$$y = H(x) + \epsilon. \quad (V.2)$$

Le vecteur d'état, noté x , contient l'état du système étudié. L'inversion Bayésienne a généralement pour but d'estimer au mieux ce vecteur. Typiquement, dans le cas du transfert radiatif, x représente les variables atmosphériques à quantifier. Généralement, nous disposons d'une estimation du vecteur d'état appelée *a priori* et notée x_b (b pour "background"). Cette "ébauche" est issue de la connaissances du système avant

l'observation (climatologie, processus d'assimilation antérieur, prévision du modèle,...). Cette connaissance se traduit par la densité de probabilité ("probability density fonction", pdf en anglais) $P(x)$, caractérisé par sa matrice de covariance d'erreur B .

Le vecteur d'observation - noté y , auquel on a accès par la mesure, est une observation indirecte de l'état du système x . La mesure est impactée par un bruit ϵ , variable aléatoire indépendante de x souvent considérée comme gaussienne et de matrice de covariance d'erreur Q . $P(y)$ est la pdf *a priori* de la mesure y . Dans le cas du transfert radiatif, y est la mesure satellite du rayonnement passif émis par l'atmosphère et ϵ est le bruit instrumental au niveau du capteur.

L'opérateur d'observation, noté H , est un opérateur déterministe non nécessairement linéaire qui relie chaque point de l'espace des états x à un point dans l'espace des mesures y .

Le théorème de Bayes donne un formalisme pour inverser statistiquement l'opérateur H et calculer une densité de probabilité *a posteriori* par mise à jour de la pdf *a priori* $P(x)$ avec la pdf de mesure $P(y)$:

- Soit $P(x)$ la pdf *a priori* de x . $P(x)dx$ est alors la probabilité, avant la mesure, que x appartienne au volume $(x, x + dx)$; ce qui traduit quantitativement notre connaissance de x avant que la mesure ne soit faite.
- $P(y)$ est la pdf *a priori* de la mesure y avant que celle-ci soit effectuée.
- $P(x, y)$ est la probabilité jointe *a priori*. $P(x, y)dxdy$ est la probabilité que x soit dans $(x, x + dx)$ et y dans $(y, y + dy)$.
- $P(y/x)$ est la probabilité conditionnelle de y sachant x , $P(y/x)dy$ représente l'information sur y si l'état était x . C'est l'équivalent probabiliste du transfert radiatif dans le cas où y est l'observation satellite.
- $P(x/y)$ est la probabilité conditionnelle de x sachant y .

On a alors :

$$\begin{aligned} P(x) &= \int P(x, y)dy, \\ P(y/x) &= \frac{P(x, y)}{P(x)}, \\ P(x/y) &= \frac{P(x, y)}{P(y)}. \end{aligned} \tag{V.3}$$

En éliminant $P(x, y)$ dans la 3^e équation, on obtient le théorème de Bayes pour la relation entre les deux pdfs conditionnelles :

$$P(x/y) = \frac{P(y/x)P(x)}{P(y)}.$$

$P(x/y)$ est alors la pdf *a posteriori* de x une fois que la mesure est effectuée. Cette information est utilisée ensuite pour mettre à jour la pdf *a priori* $P(x)$ de x .

On peut alors trouver x_a (a pour analyse) tels que la probabilité $P(x = x_a/y)$ soit maximale. C'est l'estimateur de Bayésien. Le cas où l'opérateur H est linéaire et où toutes les variables (x, ϵ, y) sont gaussiennes constitue un problème inverse simple où l'estimateur optimal peut être explicité en prenant le logarithme des pdfs.

$$\begin{aligned}
-2 \cdot \ln[P(x)] &= (x - x_b)^t B^{-1} (x - x_b) + c1, \\
-2 \cdot \ln[P(y/x)] &= -2 \cdot \ln[\epsilon] = (y - Hx)^t Q^{-1} (y - Hx) + c2,
\end{aligned}
\tag{V.4}$$

où $c1$ et $c2$ sont des constantes. En appliquant le théorème de Bayes on peut exprimer $P(x/y)$ par :

$$-2 \cdot \ln[P(x/y)] = (y - Hx)^t Q^{-1} (y - Hx) + (x - x_b)^t B^{-1} (x - x_b) + c3,
\tag{V.5}$$

avec $c3$ une constante. Cette forme est quadratique en x , avec B et R réelles, positives et définies positives, et peut donc être factorisée en :

$$-2 \cdot \ln[P(x/y)] = (x - x_a)^t S_a^{-1} (x - x_a) + c4.
\tag{V.6}$$

On a alors la pdf *a posteriori* qui est aussi gaussienne avec comme moyenne x_a et comme covariance d'erreur S_a . En posant $y - Hx = y - Hx_b - H(x - x_b)$ dans l'Eq. (V.5), on peut égaliser terme à terme les Eq. (V.5) et (V.6) et on obtient :

$$\begin{aligned}
S_a^{-1} &= H^t Q^{-1} H + B^{-1}, \\
x_a &= x_b + S_a H^t Q^{-1} (y - Hx_b).
\end{aligned}
\tag{V.7}$$

L'estimateur optimal est alors la valeur moyenne x_a de la pdf *a posteriori* $P(x/y)$. Les termes constants $c1$, $c2$ et $c4$ sont tels que la pdf $P(x/y)$ soit normale.

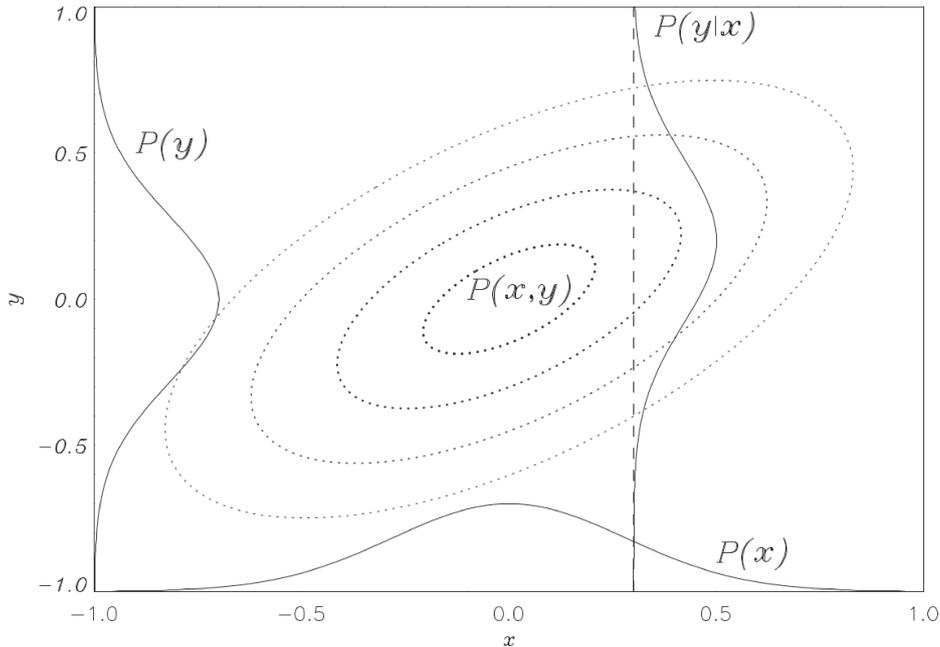


Figure V.1. Illustration du théorème de Bayes en deux dimensions. (Rodgers, 2000)

Le théorème de Bayes peut être généralisé au cas où il existe plusieurs observations indépendantes y_i d'un même système x . Pour $i = 1 \dots l$ on a :

$$y_i = H_i \cdot x + \epsilon_i, \quad (\text{V.8})$$

où chaque ϵ_i possède une matrice de covariance Q_i indépendantes des autres. Le théorème de Bayes donne :

$$\begin{aligned} P(x/y_1, y_2, \dots, y_l) &= \frac{P(y_1, y_2, \dots, y_l/x)P(x)}{P(y_1, y_2, \dots, y_l)} \\ &= \frac{P(y_1/x)P(y_2/x) \cdots P(y_l/x)P(x)}{P(y_1)P(y_2) \cdots P(y_l)} \\ &= P(x) \cdot \frac{\prod P(y_i/x)}{P(y_i)}. \end{aligned} \quad (\text{V.9})$$

L'Eq. (V.5) devient alors :

$$-2 \cdot \ln[P(x/y)] = (y - H_1 x)^t Q_1^{-1} (y - H_1 x) + (y - H_2 x)^t Q_2^{-1} (y - H_2 x) + (x - x_b)^t B^{-1} (x - x_b) + c, \quad (\text{V.10})$$

et La solution :

$$\begin{aligned} S_a^{-1} &= +B^{-1} + H_1^t Q_1^{-1} H_1 + H_2^t Q_2^{-1} H_2, \\ x_a &= x_b + S_a [H_2^t Q_2^{-1} (y_2 - H_2 x_b) - H_1^t Q_1^{-1} (y_1 - H_1 x_b)]. \end{aligned} \quad (\text{V.11})$$

Bien que l'estimateur Bayésien ait été introduit ici dans le cadre du transfert radiatif pour lequel, l'espace des variables d'état x diffère de l'espace des observations y . On peut élargir le problème d'inverse Bayésien au cas particulier où l'observation y est le résultat d'une contrainte H sur le vecteur d'état x . Dans ce cas l'opérateur H est une contrainte considérée comme une "observation" virtuelle dans le contexte de l'inversion Bayésienne (Rodgers, 2000).

1.1.b Un exemple d'estimateur optimal avec une seule contrainte

Dans le cadre du projet NEWS¹ qui vise à mieux caractériser le cycle de l'eau à l'aide des produits satellites, Rodell et al. (2015) ainsi que L'Ecuyer et al. (2015) mettent en place une contrainte hydrologique et énergétique afin de fermer simultanément les cycles de l'eau et de l'énergie, à l'échelle du globe, aussi bien sur la partie continentale que sur l'océan et dans l'atmosphère. Les deux cycles sont contraints simultanément grâce à l'équivalence entre l'évapotranspiration et la chaleur latente. Afin de prendre en compte ces contraintes, ils utilisent un estimateur linéaire où les erreurs sont considérées comme gaussiennes. Dans la suite je décrirai les méthodes d'intégration appliquées seulement à la contrainte hydrologique continentale.

Dans (Rodell et al., 2015), les composantes du cycle de l'eau se distinguent en :

1. La NASA a lancé en 2003 un programme d'étude sur les cycles de l'énergie et de l'eau (NASA Energy and Water cycle Study, NEWS). Ce programme vise à bénéficier des progrès obtenus dans l'observation satellite. En particulier, NEWS documente les cycles de l'eau et d'énergie globaux en utilisant les produits satellites les plus récents. Le programme s'intéresse aussi aux changements du cycle de l'eau et de ses composantes en réponse au changement climatique. NEWS a ainsi évalué les effets du changement climatique sur la fonte des neiges sous les hautes latitudes et a également permis l'estimation de l'évolution de la répartition de l'eau au sein des composantes continentales du cycle.

- termes de flux x (précipitation, évaporation, évapotranspiration, divergence de l'humidité atmosphérique, ruissellement); pour lequel on dispose d'un *a priori* (i.e. première estimation) x_b donné par les produits satellites avec une incertitude caractérisée par la matrice B de covariance d'erreurs *a priori*;
- termes de stock (contenu du sol en eau, contenu des océans en eau, contenu atmosphérique en vapeur d'eau), vue comme une observation y bruitée des flux;
- $H = (1, -1, -1)^t$ représente l'équilibre entre les termes de flux et ceux de stock pour l'équation du bilan hydrique continental (voir la section 1.4, pour une discussion sur les hypothèses de Rodell et al. (2015)) et ϵ l'erreur dans les stocks qui déséquilibrent le bilan hydrique (ϵ est supposé gaussien) caractérisée par sa matrice de covariance d'erreur R .

On peut écrire l'équation du cycle de l'eau comme un problème inverse dans le cadre du Théorème de Bayes :

$$y = H \cdot x + \epsilon. \quad (V.12)$$

En considérant les observations de flux comme *a priori* sur le vecteur d'état des flux et les produits satellites des stocks comme une observation bruitée des flux, avec l'opérateur d'observation $H = (1, -1, -1)^t$, Rodell et al. (2015) trouvent une estimation des flux x_a pour laquelle la contrainte de fermeture est respectée suivant l'Eq. (V.7).

1.1.c Un exemple d'estimateur optimal avec plusieurs contraintes

Dans le cadre du projet ESA WATCHFUL sur la fermeture du cycle de l'eau, Aires (2014) a développé plusieurs méthodologies (voir la Table V.1) pour intégrer de nombreux produits satellites avec une contrainte de fermeture du bilan hydrique. Ces méthodes consistent à partitionner les résidus du bilan, c'est-à-dire le déséquilibre de l'Eq. (III.1), entre les composantes en fonction de leur incertitude.

Soit $x = (P, E, \Delta S, R)^t$ la variable d'état des quatre composantes du cycle de l'eau continental. La fermeture du bilan hydrique peut être considérée comme l'observation de l'équilibre et s'écrit :

$$y_1 = 0 = H_1 \cdot x + \epsilon_1, \quad (V.13)$$

avec $H_1 = (1, -1, -1, -1)$. On a alors ϵ_1 qui représente la non fermeture du cycle de l'eau (voir l'Eq. (V.1), première ligne) et peut être caractérisée par sa matrice de covariance Q_1 .

Cette approche peut être étendue à plusieurs produits satellites disponibles pour chacune des composantes du cycle de l'eau. Soit le vecteur rassemblant tous les produits satellites pour chacune des composantes :

$$y_2 = (P_1, \dots, P_p, E_1, \dots, E_q, R_1, \dots, R_r, \Delta S_1, \dots, \Delta S_s)^t. \quad (V.14)$$

En considérant les différents produits satellites et leur incertitude, on a l'équation d'observation suivante :

$$y_2 = H_2 \cdot x + \epsilon_2, \quad (\text{V.15})$$

x est toujours le vecteur d'état des quatre composantes du cycle de l'eau et y_2 représente les produits satellites estimant chacune de ces composantes. Les produits satellites possèdent une incertitude ϵ_2 à cause du bruit instrumental mais aussi à cause des différents algorithmes d'inversion et de leurs hypothèses. Cette incertitude est caractérisée par la matrice de covariance de bruit Q_2 . H_2 est l'opérateur d'observation et a la structure suivante²

$$H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Le but de cette approche est d'exploiter simultanément toutes les informations disponibles et de trouver le meilleur compromis pour estimer x avec les deux contraintes et un *a priori*.

On peut alors considérer le problème dans le cadre de l'estimateur optimal avec plusieurs observations/contraintes indépendantes. On dispose :

- d'un *a priori* x_b donné par la climatologie avec une incertitude caractérisée par la matrice B de covariance d'erreurs *a priori*.
- d'une contrainte hydrologique y_1 bruitée de covariance d'erreur Q_1 représentant la contrainte hydrologique
- d'une observation y_2 caractérisée par sa matrice de covariance d'erreur Q_2 .

D'après le théorème de Bayes étendu au cas gaussien linéaire avec plusieurs observations indépendantes (Eq. (V.10)), la solution à ce problème existe où $y_1 = 0$ dans Eq. (V.11) :

$$\begin{aligned} S_a^{-1} &= +B^{-1} + H_1^t Q_1^{-1} H_1 + H_2^t Q_2^{-1} H_2, \\ x_a &= x_b + S_a [H_2^t Q_2^{-1} (y_2 - H_2 x_b) - H_1^t Q_1^{-1} (H_1 x_b)]. \end{aligned} \quad (\text{V.16})$$

On observe alors que l'inverse de la covariance d'erreur S_a de l'estimateur optimal x_a est l'information de Fisher des pdf des informations disponibles dans l'Eq. V.16. Le

2. S'il y avait une pondération particulière des produits satellites, alors ce ne serait pas des 1 dans la matrice H_2 .

premier terme à droite est lié à l'information *a priori* fournie par l'ébauche x_b . Puisque l'objectif de ce travail est d'obtenir une estimation des composantes du cycle de l'eau indépendante des modèles, cette ébauche doit provenir de la climatologie et non de la ré-analyse. Si pour l'ébauche on considérait la prévision d'un modèle, comme par exemple dans le cadre d'un filtre de Kalman, l'estimateur optimal serait alors identique à l'assimilation avec une contrainte d'égalité dans la mise à jour du vecteur d'état du système. Cette approche présentée par Pan et Wood (2006) est détaillée dans la Section V.1.2. Le deuxième terme est lié aux observations satellites rassemblées dans y_2 . Le troisième terme représente la contrainte de fermeture du bilan hydrique avec une matrice de covariance d'erreur Q_1 . L'estimateur optimal peut être utilisé sans l'information *a priori* x_b , mais, bien sûr, cela nuirait à la qualité de la solution.

1.1.d Vers une méthode plus simple : Filtrage post-traitement

Aires (2014) montre que la solution optimale trouvée précédemment est très proche d'un simple Filtrage post-traitement (PF) introduit par Pan et Wood (2006) et appliqué directement sur un premier estimateur lié aux données satellites : le produit issu de la méthode SW, vu dans la partie précédente.

Soit x_{SW} l'estimation obtenue du vecteur d'état x par pondération des produits satellites disponibles (voir Chapitre 4, Section 3.2.b). On a une information sur x_{SW} par sa matrice de covariance B liée aux incertitudes des produits dérivés satellites. On peut alors appliquer la contrainte de fermeture $0 = H_1 \cdot x$ déterministe (avec H_1 défini dans l'Eq. (V.13)) et obtenir l'estimation du vecteur d'état avec un bilan hydrique équilibré :

$$\begin{aligned} x_{PF} &= x_{SW} + K(0 - H_1 \cdot x_{SW}) \\ &= K_{PF} \cdot x_{SW}, \end{aligned} \tag{V.17}$$

où $K = B H_1^t (H_1 B H_1^t)^{-1}$ et $K_{PF} = [Id - K H_1]$ avec Id la matrice identité.

L'approche PF a été utilisée dans cette thèse pour mettre en place la contrainte hydrologique mais plusieurs améliorations ont été apportées et celles-ci seront détaillées dans la section V.2.

1.2 Le filtre de Kalman

Le filtre de Kalman est une technique de fusion d'information entre des produits satellites et un modèle descriptif. Comme pour l'estimateur Bayésien (Section V.1.1), les notations et cadre général du filtre de Kalman sont introduits, puis un exemple pratique d'utilisation du filtre est détaillé dans le contexte de la fermeture du cycle de l'eau.

1.2.a Principe

Dans le cadre du filtre de Kalman, le système dynamique est décrit par un vecteur d'états x , et son évolution temporelle entre chaque instant k , est décrite par l'équation suivante :

$$x_{k+1} = F_k(x_k) + M_k(u_{k+1}) + w_k. \tag{V.18}$$

Cette équation représente la dynamique (transition) du système entre les instants k et $k+1$, forcé par l'entrée u_k . Les paramétrisations du modèle de surface sont contenues dans $M_k(\cdot)$ et $F_k(\cdot)$. En raison d'hypothèses simplificatrices du modèle (i.e. les paramètres) ou d'hétérogénéités spatiales et d'incertitudes dans le forçage, la prédiction de l'état du système x à l'instant $k+1$ par $M_k(\cdot)$ et $F_k(\cdot)$ reste incertain. Cette incertitude est représentée par un bruit additif w_k .

Les sorties du système sont décrites par l'équation suivante :

$$z_k = H_k(x_k) + v_k, \quad (\text{V.19})$$

où $H_k(\cdot)$ décrit la relation entre l'état x_k et les sorties observables. Ces sorties sont mesurées avec une certaine erreur v_k telle que z_k représente la mesure bruitée des sorties du système. Le but de l'assimilation de données est d'améliorer nos connaissances sur x_k grâce à l'utilisation des observations de z_k .

Afin de simplifier le raisonnement, nous décrivons d'abord la résolution du filtre de Kalman dans le cas discret, linéaire et où les incertitudes du système et des observations sont des bruits additifs gaussiens $w_k \sim \mathcal{N}(0, B_k)$ et $v_k \sim \mathcal{N}(0, Q_k)$, indépendants les uns des autres, à tout instant k . Si les Eq. (V.18) et (V.19) sont linéaires on obtient :

$$\begin{aligned} x_{k+1} &= F_k \cdot x_k + M_k \cdot u_{k+1} + w_k, \\ z_k &= H_k \cdot x_k + v_k. \end{aligned} \quad (\text{V.20})$$

L'estimation optimale de l'état du système à l'instant k , $\hat{x}_{k|k}$, au sens du minimum de covariance d'erreur $P_{k|k}$, connaissant la mesure z_k , est issue d'une mise à jour de l'estimateur prédit par évolution du système depuis l'état précédent $\hat{x}_{k|k-1}$ après observation de z_k . Avant observation, on a :

$$\begin{aligned} \hat{x}_{k|k-1} &= F_k \cdot \hat{x}_{k-1|k-1} + M_k \cdot u_k, \\ P_{k|k-1} &= F_k P_{k-1} F_k^t + B_k, \end{aligned} \quad (\text{V.21})$$

où $\hat{x}_{k|k-1}$ est souvent appelé "forecast" et noté x_k^f .

L'observation z_k permet une innovation \tilde{y}_k de l'estimateur précédent avec une matrice de covariance d'erreur S_k :

$$\begin{aligned} \tilde{y}_k &= z_k - H_k \cdot \hat{x}_{k|k-1}, \\ S_k &= H_k \cdot P_{k|k-1} \cdot H_k^t + Q_k. \end{aligned} \quad (\text{V.22})$$

On met alors à jour l'estimateur de l'état du système $\hat{x}_{k|k}$ et sa matrice de covariance d'erreur $P_{k|k}$ en utilisant l'observation disponible z_k :

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k \cdot \tilde{y}_k \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k \cdot (z_k - H_k \cdot \hat{x}_{k|k-1}), \\ P_{k|k} &= (I - K_k \cdot H_k) \cdot P_{k|k-1}, \\ K_k &= P_{k|k-1} H_k^t \cdot S_k^{-1} \\ K_k &= P_{k|k-1} H_k^t \cdot (H_k \cdot P_{k|k-1} \cdot H_k^t + Q_k)^{-1}, \end{aligned} \quad (\text{V.23})$$

où K_k est le gain de Kalman optimal. Et $\hat{x}_{k|k}$ est souvent appelé "analyse" et noté x_k^a .

Les Éq. (V.21), (V.22) et (V.23) forment l'algorithme du filtre de Kalman. Celui-ci est illustré dans la Figure V.2. Dans le cas du filtre de Kalman linéaire, l'"analyse" est alors identique à l'estimateur optimal avec comme ébauche *a priori* l'évolution du système depuis la dernière assimilation d'observation.

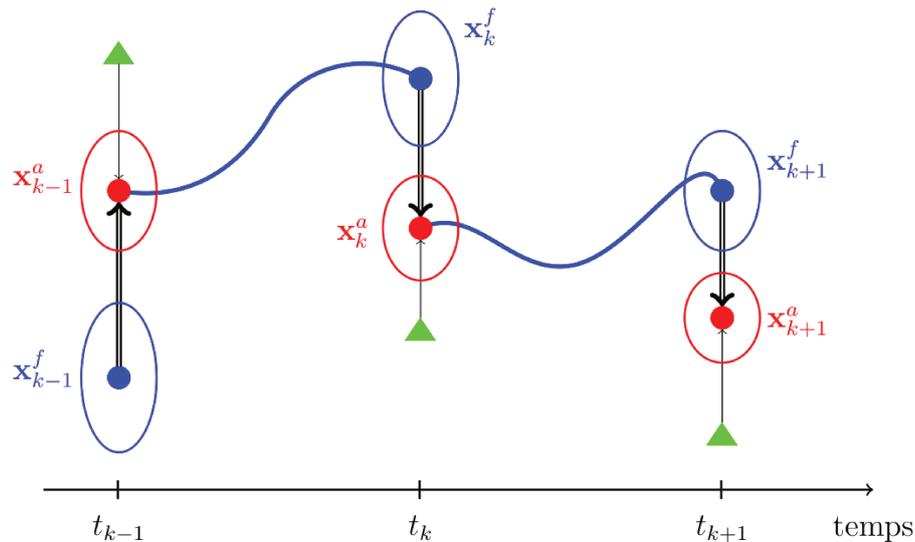


Figure V.2. Fonctionnement séquentiel du filtre de Kalman. A un instant donné k , nous disposons de x_k^f (point bleu) une prévision faite grâce au modèle à partir de l'état analysé au temps précédent, (équivalent à $\hat{x}_{k|k-1}$ dans le texte) de sa matrice de covariance d'erreur P_k^f (symbolisée par l'ellipse bleue) et d'une observation y_k (triangle vert). La prévision est corrigée à l'aide de l'observation pour produire l'analyse x_k^a (point rouge, équivalent à $\hat{x}_{k|k}$ dans le texte) avec pour matrice de covariance d'erreur P_k^a (symbolisée par l'ellipse rouge). (Bonan, 2014)

Le filtre de Kalman peut être généralisé aux modèles et opérateurs d'observations non-linéaires. Pour cela, il suffit de considérer l'Eq. (V.18) et de linéariser les opérateurs $B_k(\cdot)$ et $F_k(\cdot)$ autour de x_k . Noter que dans ce cas, l'estimateur via le gain de Kalman n'est plus optimal. Cette version généralisée du filtre de Kalman est appelée filtre de Kalman étendu (EKF pour "Extended Kalman Filter"). Pour les modèles non linéaires et pour lesquels les erreurs de mesures sont non stationnaires et non gaussiennes, l'approche classique de EKF peut présenter d'importantes difficultés pour le calcul du gain K . Comme alternative à la résolution explicite du gain de Kalman, on peut chercher à estimer et mettre à jour l'état du système, sa covariance d'erreur et donc le gain de Kalman par une méthode de Monte Carlo où l'estimation est approchée par un ensemble N_e d'états représentatifs du système. Cette approche développée par Evensen (1994) est appelée filtre de Kalman d'ensemble (EnKF pour "Ensemble Kalman Filter").

1.2.b Ajout de la contrainte de fermeture du bilan hydrique dans le filtre de Kalman

Après avoir obtenu des estimations par filtre de Kalman non contraint, l'approche pour contraindre l'état du système hydrologique est de construire une deuxième équation d'observation comme contrainte d'égalité (bilan hydrique) avec une observation considérée sans bruit, qui permet alors une nouvelle mise à jour. Pan et Wood (2006) ont présenté en premier lieu un travail dans lequel ils visaient à fermer le bilan hydrique à l'aide des produits satellites. Dans cette étude, les composantes du cycle de l'eau continental ont été assimilées dans le modèle de surface VIC (Liang et al., 1994) en utilisant un filtre de Kalman et une contrainte de fermeture sur l'état du système (voir la Table V.1). En supposant ces nouvelles observations parfaites pendant le calcul du gain de Kalman et la mise à jour de Kalman, on peut obtenir des singularités dans la matrice de covariance des erreurs d'observation, mais ce problème peut être dépassé par l'utilisation de la pseudo-inverse de Moore-Penrose (Pan et Wood, 2006). Yilmaz et al. (2011) ont résolu ce problème en considérant un "bruit d'observation" dans cette seconde équation qui correspond à une incertitude sur la contrainte d'égalité du bilan hydrique.

1.2.b.1 Contrainte stricte implémentée dans VIC

Le modèle de surface continentale VIC (Liang et al., 1994) est un modèle numérique décrivant les échanges d'eau et d'énergie entre la surface continentale et l'atmosphère. On a alors le vecteur d'état $x_k = [s_{1,k}, s_{2,k}, s_{2,k}]^t$, représentant l'humidité volumétrique dans trois couches du sol, superficielle, intermédiaire et profonde. VIC est un modèle de surface, forcé par les précipitations $u_k = p_k$ et ses sorties observables sont $H(x_k) = [s_{1,k}, s_{2,k}, e_k, q_k]^t$ avec $e_k = e_k(s_{1,k}, s_{2,k}, s_{2,k}, p_k)$ l'évapotranspiration et $q_k = q_k(s_{1,k}, s_{2,k}, s_{2,k}, p_k)$ le ruissellement.

VIC est un modèle hydrologique unidimensionnel, ce qui signifie que le modèle traite la terre (sol et végétation) comme une colonne 1D et que les flux latéraux d'humidité sont ignorés. Ainsi, le bilan hydrique défini pour cette colonne 1D peut s'écrire à l'aide de p, e, q et le changement dans le stockage total de l'eau dans le sol ($s_1 + s_2 + s_3$) entre les instants $k-1$ et k :

$$\begin{aligned} p_k - e_k - q_k - [(s_{1,k}, s_{2,k}, s_{2,k}) - (s_{1,k-1}, s_{2,k-1}, s_{2,k-1})] &= 0 \\ G_k(x_k, u_k) &= g_k \\ G_k(x_k^*) &= g_k. \end{aligned} \quad (V.24)$$

Où x_k^* est le vecteur d'état "étendu" (qui prend le terme de forçage en plus du vecteur d'état), $G_k(\cdot)$ est le filtre de contrainte avec $g_k = 0$ à tout instant k pour un bilan hydrique équilibré et ce avec une matrice de covariance d'erreur R_{2k} nulle. Dans la suite nous présenterons la contrainte utilisée par Pan et Wood (2006) dans le cadre du EKF où les opérateurs ont été linéarisés autour de la prédiction x_k même si l'implémentation telle qu'elle est décrite dans (Pan et Wood, 2006) est basée sur un filtre EnKF. Pour le système représenté dans l'Eq.(V.24), le gain de Kalman K_{2k} est, suivant l'Eq. (V.23) :

$$K_{2k} = P_{k|k}^* G_k^t \cdot (G_k \cdot P_{k|k}^* \cdot G_k^t)^{-1}, \quad (V.25)$$

où $P_{k|k}^*$ est la matrice de covariance d'erreur de l'état estimé par le filtre de Kalman non contraint (étendu au terme de forçage). On peut alors mettre à jour l'estimateur de l'état du système (suivant l'Eq. (V.23)), maintenant contraint par son bilan hydrique nul :

$$\begin{aligned}\hat{x}_{k+c|k}^* &= \hat{x}_{k|k}^* - K_{2k} \cdot G_k \cdot \hat{x}_{k|k}^* \\ \hat{x}_{k+c|k}^* &= \hat{x}_{k|k}^* - P_{k|k}^* G_k^t \cdot (G_k \cdot P_{k|k}^* \cdot G_k^t)^{-1} \cdot \hat{x}_{k|k}^*,\end{aligned}\tag{V.26}$$

où $k+c$ indique l'instant où l'on prend en compte la contrainte hydrique.

1.2.b.2 Contrainte relaxée

Appliquer une contrainte forte lors de l'assimilation (i.e., imposer un résidu nul $g_k=0$ avec une covariance d'erreur nulle $R_2=0$) permet de préserver la quantité totale d'eau en termes de contenu en eau. Toutefois, le problème que pose une contrainte forte est d'équilibrer le bilan hydrique quelles que soient les erreurs individuelles des produits intégrés. Il semble inapproprié d'imposer la fermeture d'un bilan hydrique imparfaitement observé.

Dans (Yilmaz et al., 2011), il est démontré que le filtre de Kalman contraint (Pan et Wood, 2006) peut conduire à des estimations d'état très irréalistes. Si ces erreurs sont suffisamment importantes, la contrainte budgétaire entraînera la valeur de certaines variables d'état au-delà de leur plage de variation naturelle. Yilmaz et al. (2011) ont étendu l'approche de Pan et Wood (2006) en relaxant la contrainte de fermeture du bilan hydrique. Ici, une contrainte relaxée est imposée, ce qui est dû à l'incertitude dans le bilan hydrique même. Yilmaz et al. (2011) ont introduit une contrainte de fermeture caractérisée par une incertitude ϕ , l'Eq. (V.26) devient alors :

$$\hat{x}_{k+c|k}^* = \hat{x}_{k|k}^* - P_{k|k}^* G_k^t \cdot (\phi + G_k \cdot P_{k|k}^* \cdot G_k^t)^{-1} \cdot \hat{x}_{k|k}^*.\tag{V.27}$$

On remarque que plus l'incertitude ϕ est faible plus le filtre se rapproche du filtre de Pan et Wood (2006).

1.3 L'assimilation via le 3D-VAR

Introduite dès 1955 par Sasaki (1955), cette méthode est utilisée de façon opérationnelle conjointement avec les filtres de Kalman par les services de prévision météorologique. Les méthodes variationnelles (3D-VAR et 4D-VAR) ne sont pas principalement basées sur la théorie de l'estimation statistique, comme l'estimateur optimal, mais sur la théorie de l'optimisation. Néanmoins dans le cas linéaire et gaussien, les deux méthodes aboutissent au même estimateur.

1.3.a Principe

L'objectif de l'assimilation 3D-Var est de trouver une estimation x_a du vecteur d'état x grâce à l'observation y indirecte de x : $y = H(x) + \epsilon$. On cherche x_a minimisant la fonction coût suivante :

$$J(x) = J^b(x) + J^o(x), \quad (\text{V.28})$$

avec $J^b(x)$ représentant le coût de l'écart de l'ébauche x_b

$$J^b(x) = \frac{1}{2}(x - x_b)^t B^{-1}(x - x_b), \quad (\text{V.29})$$

et $J^o(F)$ représentant le coût de l'écart des observations y

$$J^o(x) = \frac{1}{2}(y - H(x))^t R^{-1}(y - H(x)). \quad (\text{V.30})$$

Minimiser J représente donc un compromis (au sens des moindres carrés) entre être proche de l'ébauche et être proches des observations. Cette propriété utilise une "distance" proportionnelle à l'inverse de leurs incertitudes respectives comme l'information de Fisher pour le cas gaussien.

Le 3D-Var contourne efficacement le problème lié aux opérateurs d'observation non linéaires H et aux erreurs non gaussiennes ϵ en minimisant numériquement la fonction coût J . Les méthodes, type "descente du gradient", permettent la minimisation de cette fonction de cout.

Le schéma d'assimilation peut être utilisé avec un modèle dynamique comme le filtre de Kalman. La Figure V.3 représente le cycle d'assimilation du 3D-VAR utilisé dans le contexte d'une mise à jour de modèle dynamique.

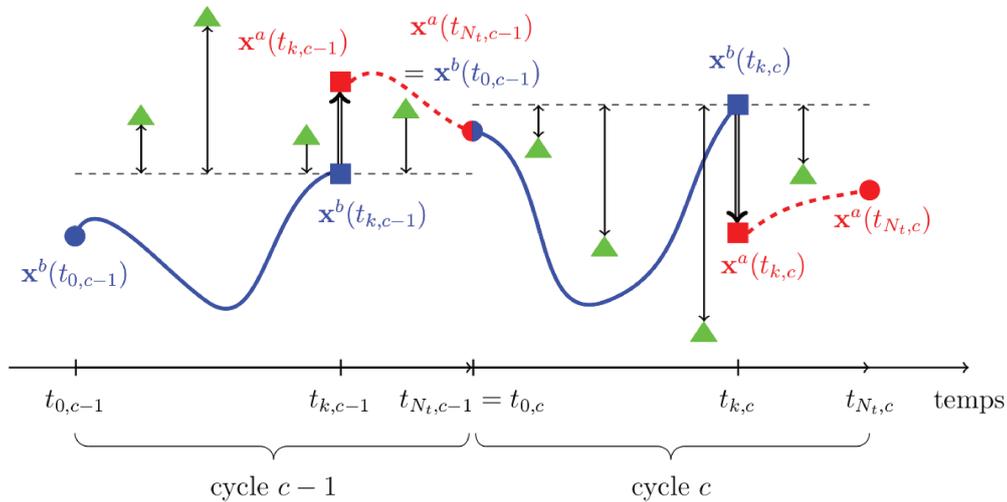


Figure V.3. Cycle du 3D-Var. A chaque cycle c , le modèle d'évolution propage l'ébauche du système x_b jusqu'à l'instant $t_{k,c}$ pour produire $x_b(t_{k,c})$ (courbe bleue). L'assimilation par 3D-VAR consiste ensuite à utiliser les observations (triangles verts) pour produire une analyse $x_a(t_{k,c})$. Pour cela, il faut calculer la distance entre les observations (triangles verts) et les sorties du modèle $H_k(x)$. Après la minimisation, $x_a(t_{k,c})$ peut être utilisé comme condition initiale pour le cycle suivant. (Bonan, 2014).

Les méthodes 3D-VAR n'ont, à ma connaissance, jamais été utilisées avec une contrainte hydrologique, comme on a pu le faire avec l'estimateur optimal (Section V.1.1). Elles pourraient pourtant être utilisées de la même façon, en utilisant l'équation de fermeture comme une équation d'observation virtuelle.

1.3.b Le 4D-VAR

Alors que le schéma 3D-VAR ne prend pas en compte la dynamique des observations (elles sont toutes comparées à l'état du système lors de l'assimilation), dans le cadre du 4D-VAR (Dimet et Talagrand, 1986), c'est la trajectoire optimale sur une fenêtre de temps donnée qui est recherchée. Cette variante du 3D-Var qui ajoute la dynamique du système à l'assimilation est celle utilisée en opérationnel dans les systèmes de l'ECMWF et de Météo-France.

Dans le 4D-VAR, la fonction coût J à minimiser devient alors :

$$J(x_o) = J^b(x_o) + J^o(x_o), \quad (\text{V.31})$$

avec $J^b(x_o)$ représentant l'écart à l'ébauche à l'instant initial de la fenêtre de temps considéré dans le 4D-VAR :

$$J^b(x_o) = \frac{1}{2}(x_o - x_b)^t B^{-1}(x - x_b), \quad (\text{V.32})$$

et $J^o(F)$ représentant l'écart aux observations prenant en compte la dynamique du modèle et la temporalité des observations :

$$J^o(x_o) = \frac{1}{2} \sum_{k=0}^N (y_k - H_k(x_k))^t R_k^{-1}(y_k - H(x_k)), \quad (\text{V.33})$$

$J^o(x_o)$ fait intervenir l'état du système à différents instants : $x_{k+1} = M_k(x_k)$ et un instant initial connu x_o . Lors de l'assimilation 4D-VAR, ce sont les distances des observations y_k aux sorties du modèle $H_k(x_k)$ aux différents instants qui sont pris en compte.

Les schémas 3D-VAR et 4D-VAR pourraient alors contraindre les différents produits dans le cycle de l'eau sans hypothèse d'erreur gaussienne comme c'est le cas pour l'approche par filtre de Kalman d'ensemble (Section V.1.2.b).

1.4 Discussion sur les méthodes d'intégration dans le contexte de l'étude du cycle de l'eau par les satellites

Les caractéristiques des méthodes présentées par Pan et Wood (2006), Rodell et al. (2015) et Aires (2014) sont synthétisées dans la Table V.1.

Si (Pan et Wood, 2006) est le premier papier de l'équipe de Princeton sur l'implémentation d'une contrainte hydrologique dans le modèle de surface VIC, plusieurs articles ont été publiés sur cette approche, d'abord sur l'assimilation au niveau du bassin

	Princeton	NASA NEWS	ESA
Méthodes d'intégration	Simple Weighting + Kalman contraint	Estimateur optimal	Simple Weighting + Post-Filtering
References	Pan et Wood (2006) Sahoo et al. (2011); Pan et al. (2012)	Rodell et al. (2015) L'Ecuyer et al. (2015)	Aires (2014); Munier et al. (2014) Munier et Aires (2017)
Stratégie	Assimilation dans VIC	Optimisation des flux	Optimisation
Sources	Observations + modèle	Observations + modèle	Observations
Bilan	Continental seulement	Continental, océanique & atmosphérique	Continental, océanique & atmosphérique
Échelle spatiale	bassin ³	continent	pixel et bassin
Multiplicité des produits	oui fusion de données	pour <i>E</i> seulement	oui fusion de données
Référence pour les incertitudes	jauges & produit moyen	produit moyen	prescrit (littérature)
Multi-échelle spat.	non	oui : continents et un océan	oui : bassin et sous-bassins
Multi-échelle temp.	non : mensuel	non : annuel + interpolation ⁴	oui : mensuel & annuel
Vecteur d'état	$X_T = [P_l E_l R_l \Delta S_l]^t$	$F = [P E R Div]^t$ $Res = [\Delta S \Delta W]^t$	$X_{lo} = [P_l E_l R_l \Delta S_l \Delta W_l Div_l]^t$ sur contient $X_o = [P_o E_o \Delta S_o Gib]^t$ pour la mer $X_{lo} = [X_l X_o]$ pour les deux
Covariance d'erreurs	B_T est la matrice de covariance de X_T	S_{Res} and S_F matrices de covariance	B_{lo} est la matrice de covariance de X_{lo}
Math.	$G_T = [1, -1, -1, -1]$	A : Matrice du Bilan ⁵	$G_l = \begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}$ $G_o = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & -1 \\ -1 & 1 & 0 & -1 & -1 & 0 \end{bmatrix}$ $L_{lo} = \begin{bmatrix} 0 & 0 & \frac{A_{land}}{A_{sea}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ $G_{lo} = \begin{pmatrix} G_l & 0 \\ L_{lo} & G_o \end{pmatrix}$
Équation de fermeture	$G_T \cdot X_T = 0$	$Res = A \cdot F$	$G_{lo} \cdot X_{lo} = r$, $r \sim \mathcal{N}(0, \Sigma)$ with $\sqrt{\Sigma} = 2 \text{ mm/month}$
Type de contrainte	contrainte forte	contrainte forte + interpolation	contrainte relaxée
Solution	$X_{Tc} = X_T + K_T \cdot (0 - G_T X_T)$ avec $K_T = B_T G_T \cdot (G_T B_T G_T^t)^{-1}$	$F_c = F + Q^{-1} J^t S_{Res}^{-1} (Res - AF)$ J le Jacobien de Res p/r à F et $Q = (J^t S_{Res}^{-1} J + S_F^{-1})^{-1}$	$X_{loc} = (I - K_{PF} G_{lo} \Sigma^{-1} G_{lo}^t) \cdot X_{lo}$ $K_{PF} = (B_{lo}^{-1} + G_{lo} \Sigma^{-1} G_{lo}^t)^{-1}$

TABLE V.1 – Les trois principales initiatives en matière de contrainte de fermeture du cycle de l'eau ("Water Cycle" WC en anglais) et leurs différences. Dans la troisième colonne, les caractères en gras indiquent les nouvelles caractéristiques de la méthodologie présentée dans ce travail de thèse. Les indices sont : *l* pour la terre, *o* pour l'océan, les deux incluent l'atmosphère au-dessus de leurs surfaces. Toutes les notations sont résumées dans la Table V.2.

versant d'une rivière (Pan et Wood, 2006 ; Pan et al., 2012 ; Sahoo et al., 2011) à l'aide de plusieurs estimations satellites pour chacune des variables prenant en compte leurs incertitudes (Sheffield et al., 2009). Une assimilation au niveau du pixel, sur l'ensemble du globe a ensuite été entreprise dans VIC (Zhang et al., 2017). Les bases de données d'"analyse", résultant de ces études, ne sont pas des produits satellites pur, puisque le modèle du VIC est largement utilisé pour leur genèse. De plus, l'utilisation d'un modèle de surface limite fortement le bilan hydrique à sa composante continentale. Même relaxée, cette contrainte peut déséquilibrer le bilan hydrique dans l'atmosphère.

Dans (Rodell et al., 2015), le cycle de l'eau est fermé dans ses trois dimensions (continental, océanique et atmosphérique) et avec le cycle d'énergie (L'Ecuyer et al.,

2015). Cependant, cette fermeture est obtenue à l'échelle de 7 continents et un méga-océan, et annuellement seulement. Rodell et al. (2015) utilisent alors les résultats obtenus pour fermer le cycle de l'eau à l'échelle du mois. Ils utilisent ensuite une interpolation de Lagrange entre la somme des mois et l'année pour obtenir une contrainte "relaxée" et plus réaliste mensuellement. Ils utilisent ce même schéma d'interpolation pour obtenir une fermeture à l'échelle des océans (considérant alors 9 mers et océans dont la mer Méditerranée et la mer Noire) à partir de la fermeture du méga-océan. Rodell et al. (2015) utilisent plusieurs équations seulement valables à grandes échelles temporelles et spatiales qu'ils assouplissent ensuite à plus fine échelle par une interpolation de Lagrange. Dès lors ces estimations à plus fine échelle ne sont pas optimales. De plus, ils négligent les variations du contenu en eau dans le sol et dans l'atmosphère aussi bien annuellement que mensuellement. Certains termes d'échange entre bassins océaniques ne sont pas quantifiés. En particulier, entre la mer Noire et la mer Méditerranée (déroit du Bosphore) et entre la Méditerranée et l'Océan Atlantique (déroit de Gibraltar). Si Rodell et al. (2015) intègrent des produits satellites, leur approche n'est pas d'utiliser les nombreuses sources d'information disponibles pour chaque composante du cycle de l'eau. En effet seule l'évapotranspiration et la convergence d'humidité atmosphérique sont quantifiées par plusieurs sources (satellites et ré-analyses). Les autres variables ne sont estimées que par un unique produit, ce qui peut limiter la réduction d'erreurs aléatoires (voir le chapitre 5). En outre la précipitation n'est estimée que par GPCP et l'information du ruissellement est issue d'une fusion entre le modèle de surface VIC et les rivières observées. La fermeture est à l'échelle globale et on peut supposer que les résidus du bilan hydrique sont encore significatifs à l'échelle régionale et mensuelle. Seules les séries chronologiques des moyennes spatiales sont corrigées et aucune tentative de spatialisation n'a été entreprise.

Aires (2014) a comparé plusieurs méthodologies (Table V.1) pour intégrer différents produits satellites avec une contrainte de fermeture du bilan hydrique. Aucun modèle de surface n'a été utilisé dans ces méthodes d'intégration, ce qui rend le produit obtenu intéressant pour l'étalonnage et la validation de modèles. Parmi les méthodes d'intégration, Aires (2014) présente l'utilisation de l'estimateur optimal, proche de celui utilisé par Rodell et al. (2015). L'autre méthode proposée, appelée Filtrage post-traitement (PF), utilise la contrainte d'égalité présentée dans (Pan et Wood, 2006) mais l'applique hors du cadre d'assimilation dans le modèle VIC par filtre de Kalman puisqu'il s'agit de l'appliquer sur une "ébauche" issue directement des observations satellites (SW). La méthode est alors appelée SW+PF. Cette méthode SW+PF a été appliquée par Munier et al. (2014) au-dessus du bassin du Mississipi avec une validation du concept utilisant des données *in situ*. Puis elle a été utilisée sur plusieurs bassins des grandes rivières du globe (Munier et Aires, 2017) en utilisant des ensembles de données satellites pour P , E et ΔS et des observations *in situ* pour R . Si une attention particulière est portée sur la multiplicité des sources d'informations satellites pour les différentes composantes du cycle de l'eau, en particulier à travers l'approche de fusion SW, seule la composante continentale du cycle est fermée dans ces études.

2 Nouveaux développements théoriques et conditions expérimentales

Dans le cadre de cette thèse, nous sommes partis des développements menés par Aires (2014); Munier et al. (2014); Munier et Aires (2017) sur la mise en place d'une contrainte hydrologique, afin d'intégrer les produits satellites pour l'analyse du cycle de l'eau. On se focalise ici sur la région méditerranéenne. Plusieurs développements ont été ajoutés au cadre de base (Aires, 2014) et sont détaillés dans cette section. Ils sont indiqués en gras dans la Table V.1. Les notations sont données dans la Table V.2.

2.1 Hypothèses sur les incertitudes des observations

Il faut, tout d'abord, revenir sur les hypothèses faites quant aux incertitudes des produits satellites qui sont importante car elle fixe les plages de variation qu'entraînera la fermeture sur chacune des composantes respectivement. On parlera d'incertitudes en terme de biais et d'erreurs aléatoires, caractérisées par leur écart type. On exprimera ces incertitudes sur les variables mensuelles puisqu'on travaille à cette échelle.

Certaines études visent à caractériser ces incertitudes comme l'incertitude relative de nombreux produits en fonction de la distance au produit moyen (Pan et al., 2012; Zhang et al., 2016) ou bien comme la distance à un produit référent non satellite comme les données *in situ* ou la ré-analyse (Sahoo et al., 2011). Ces caractérisations sont généralement spécifiques aux produits et au domaine géographique de l'étude. Pour certains produits utilisés dans cette thèse, aucune caractérisation de l'incertitude ne peut être trouvée dans la littérature. Il est donc difficile de privilégier un produit par rapport à un autre pour une même composante du cycle de l'eau. Nous avons considéré la même incertitude pour tous les produits d'une variable géophysique donnée (voir Section IV.3.2).

La Table V.3 résume les incertitudes utilisées dans les différentes techniques d'intégration qui existent dans la littérature. L'incertitude est associée à un poids qui est le rapport entre l'incertitude de la variable considérée et la somme de toutes les incertitudes de l'équation du bilan hydrique (calculée comme $\sigma_i^2 / \sum_i \sigma^2$ et exprimée en pourcentage). Les incertitudes présentées dans la Table V.3 correspondent aux produits après fusion par pondération et non à un produit satellite particulier (voir la méthodologie SW Section IV.3.2). Les incertitudes considérées dans ce travail de thèse sont issues de (Munier et al., 2014) mais légèrement modifiées pour traiter le cas particulier de la région méditerranéenne. Munier et al. (2014) ont utilisé des valeurs d'incertitude de 10 mm pour chacun des quatre produits P et les trois produits E (menant à $\sigma_P=5$ et $\sigma_E=5,8$ mm pour P et E après fusion par pondération SW), 5 mm pour chacun des trois produits ΔS (menant à $\sigma_{\Delta S}=2,9$ mm pour le produit fusionné) et $\sigma_R=1$ mm pour le seul produit R .

Afin d'être plus proche de Rodell et al. (2015), d'une part, nous avons décidé de réduire l'incertitude σ_P à 4 mm. Ceci est justifié puisque la correction de biais a été faite vers le produit TMPA qui est calibré par des données de jauges (voir chapitre 5). D'autre part, nous avons augmenté l'incertitude σ_E jusqu'à 6 mm. L'incertitude du ΔS fusionné est globalement la même que dans (Munier et al., 2014) puisqu'elle est principalement due à la faible résolution spatiale des pixels de GRACE. Enfin, l'incertitude du

Symboles mathématiques	
M^t	transpose
ΔM	Différenciation
$\frac{\delta M}{\delta t}$	Dérivation
N	Distribution normale
σ	Ecart-type
RMS	Root Mean Square
RMSD	Root Mean Square of the Difference
Indices	
M_T	Continental ("Terrestrial" en anglais)
M_I	Sur terre (continental plus atmosphérique au-dessus de la surface continentale)
M_I^i	Sur le i^{eme} sous-bassin (continental plus atmosphérique)
M_o	Sur mer (océanique plus atmosphérique)
M_{Io}	Global : sur terre et sur mer
M_c	Contraint
M_{SW}	Estimation par la solution SW
M_{PF}	Estimation par la solution SW+PF
M_{INT}	Estimation par la solution INT
Composantes du cycle de l'eau	
P	Précipitation
E	Evapotranspiration
S	Contenu en eau
W	Vapeur d'eau
Div	Divergence de l'humidité verticalement intégrée
Gib	Écoulement net à Gibraltar
Bos	Écoulement net au Bosphore
Vecteur d'état du cycle de l'eau et matrice d'incertitude associée	
X_T, B_T	Continental
X_I	Sur terre (continental plus atmosphérique)
$X_I^{(i)}$	Sur le i^{eme} sous-bassin (continental plus atmosphérique)
X_o	Sur mer (océanique plus atmosphérique)
X_{Io}, B_{Io}	Global
X_{Io}^{Month}	Global pour un mois <i>month</i> particulier
r, Σ	Résidu du bilan hydrique toléré
Matrices de fermeture	
G_T	Bilan hydrique continental
G_I	Bilan hydrique sur terre (continental plus atmosphérique)
$G_I^{(i)}$	Bilan hydrique sur le i^{eme} sous-bassin (continental plus atmosphérique)
G_o	Bilan hydrique sur mer (océanique plus atmosphérique)
G_{Io}	Bilan hydrique global
L_{Io}	Égalité en ruissellement d'eau douce
A_{Iand}	Aire du bassin versant global de la Méditerranée avec la mer Noire
$A_I^{(i)}$	Aire du sous-bassin versant i^{eme}
A_{Sea}	Aire maritime (Méditerranée + mer Noire)
$L_{Io}^{(i)}$	Égalité en ruissellement d'eau douce depuis le sous-bassin i^{eme}
GA_{Io}	Bilan hydrique global sur tout les mois et l'année entière
N_{Io}	Version modifiée de G_{Io} sans l'impact de $DeltaS$
$N_I^{(i)}$	Version modifiée de $G_I^{(i)}$ sans l'impact de $DeltaS$
Filtre	
K_T	Contrainte continentale
K_{merge}	Matrice de fusion de données par SW
K_{PF}	Contrainte de fermeture globale via PF
K_{an}	Filtre d'analyse théorique

TABLE V.2 – Notations utilisées dans la section

ruissellement R a été augmentée puisque le produit est partiellement basé sur des simulations du modèle ORCHIDEE et que l'écoulement souterrain n'est pas inclus dans l'analyse (voir Section III.2.1). Pour les variables atmosphériques, nous considérons une incertitude selon la plage de variabilité de la variation en vapeur d'eau atmosphérique : 1 mm . Suite à l'étude de Seager et Henderson (2013), l'incertitude sur la divergence d'humidité de la ré-analyse a été fixée à 6 mm en raison de sa grande variabilité et de son incapacité à représenter des systèmes convectifs d'une durée inférieure à 6h (voir Section III.3.1).

	Zhang et al. (2017)	Sahoo et al. (2011)	Munier et al. (2014)	Rodell et al. (2015)	Our study
Area	Europe	Danube basin	Mississippi basin	Eurasia	Med. region
P	- 36%	- 47%	5 mm 37%	3 mm 24%	4 mm 25%
E	- 41%	- 32%	5.8 mm 49%	5 mm 65%	6 mm 55%
R	- 7%	- 3%	1 mm 1.5%	3 mm 11%	2 mm 6%
ΔS	- 14%	- 18%	2.9 mm 12.5%	- -	3 mm 14%

TABLE V.3 – Incertitudes pour les composantes continentales du cycle de l'eau, en termes d'écart-type d'erreur mensuelle et de pourcentage d'incertitude totale sur le résidu du bilan hydrique. Les pondérations associées à une variable (calculées comme le rapport entre la variabilité de la variable particulière et celle du résidu du bilan hydrique $\sigma_i^2 / \sum_i \sigma^2$) sont exprimées en pourcentage.

2.2 Fermeture globale du cycle de l'eau : terrestre, océanique et atmosphérique

Pan et Wood (2006) et Aires (2014) ne se sont intéressés qu'au bilan hydrique continental. Ceci s'explique par l'importance de cette composante du cycle de l'eau en termes d'accès aux ressources en eau disponible, son temps de séjour à la surface et les conséquences directes sur les sociétés humaines. Néanmoins, comme suggéré dans l'introduction de ce chapitre, contraindre le bilan hydrique sur sa composante continentale peut engendrer un déséquilibre dans le bilan hydrique atmosphérique. La Figure V.4 montre la différence de correction apportée sur les 6 composantes du cycle de l'eau lorsqu'on utilise une contrainte de fermeture seulement continentale et lorsqu'on utilise une contrainte de fermeture du cycle de l'eau globale (continental, océanique et atmosphérique). La dernière ligne de la Figure V.4 montre la valeur absolue des résidus du bilan hydrique global (continental+ océanique + atmosphérique). Comme attendu, la fermeture de la partie continentale seule ne permet pas de fermer le cycle de l'eau sur sa partie océanique et atmosphérique ; bien que les précipitations et l'évapotranspiration soient contraints, elles ne sont pas cohérentes avec les autres composantes atmosphériques (vapeur d'eau et divergence de l'humidité). Dans certains cas, la fermeture continentale seule peut dégrader la fermeture sur les autres parties du cycle de l'eau.

Afin d'équilibrer le bilan hydrique, non seulement sur sa composante continentale mais aussi océanique et atmosphérique, nous considérons d'abord les six composantes terrestres de l'eau $X_l^t = (P_l, E_l, R_l, \Delta S_l, \Delta W_l, Div_l)$ et les six composantes océaniques de l'eau $X_o^t = (P_o, E_o, \Delta S_o, \Delta W_o, Div_o, Gib)$ (^t voir Table V.2). On définit alors le vecteur d'état représentatif de l'ensemble du cycle de l'eau sur la Méditerranée $X_{lo}^t = [X_l, X_o]^t$. La fermeture du bilan hydrique peut être relaxée (Yilmaz et al., 2011), en utilisant une variable aléatoire gaussienne centrée r telle que $X^t \cdot G_{lo}^t = r$, avec $r \sim \mathcal{N}(0, \Sigma)$

avec :

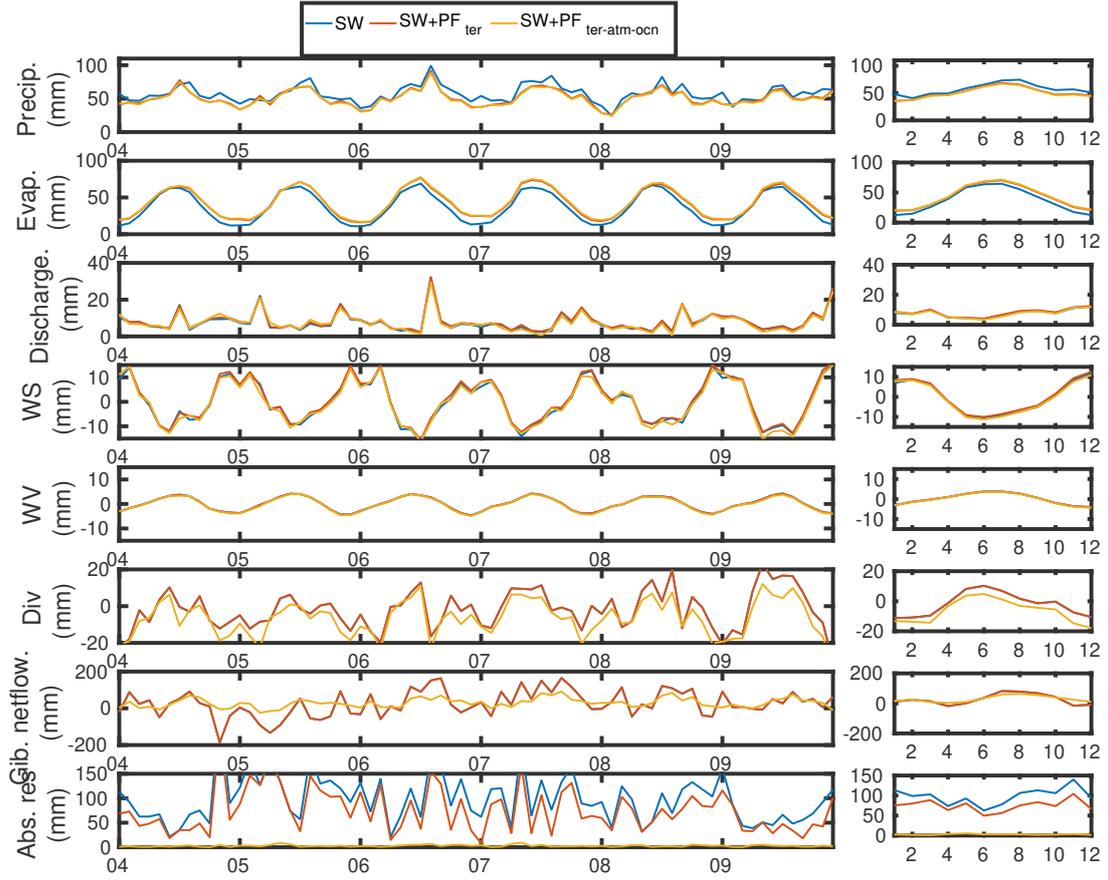


Figure V.4. Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'une des intégration (en rouge) ne considère comme contrainte seulement l'équation continental, l'autre (en jaune) contraint le cycle de l'eau dans sa globalité. La fermeture de la partie continentale seule ne permet pas de fermer le cycle de l'eau sur sa partie océanique et atmosphérique.

$$G_{lo} = \begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{A_{land}}{A_{sea}} & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & -1 & 0 \end{bmatrix}, \quad (V.34)$$

qui représente les équations du bilan hydrique dans l'Eq. (V.1).

Soit le vecteur d'observation :

$$Y_l^t = (P_1, \dots, P_p, E_1, \dots, E_q, R_1, \dots, R_r, \dots, \Delta S_1, \dots, \Delta S_s, \Delta W_1, \dots, \Delta W_v, Div_1, \dots, Div_d), \quad (V.35)$$

ce vecteur est de dimension $n_l = p + q + m + s + v + d$, il rassemble les multiples observations disponibles pour chaque composante du cycle de l'eau sur terre :

- (P_1, P_2, \dots, P_p) , les p estimations de précipitation ;
- (E_1, E_2, \dots, E_q) , les q sources d'information pour l'évapotranspiration ;

- (R_1, R_2, \dots, R_m) , les n estimations du ruissellement ;
- $(\Delta S_1, \Delta S_2, \dots, \Delta S_s)$, les s sources d'information pour le changement du contenu en eau dans le sol.
- $(\Delta W_1, \Delta W_2, \dots, \Delta W_v)$, les v estimations du changement en vapeur d'eau dans l'atmosphère.
- $(Div_1, Div_2, \dots, Div_d)$, les d estimations pour la divergence d'humidité.

De même, Y_o est le vecteur d'observation de dimension n_o , défini sur la partie océanique.

On cherche le filtre K_{an} permettant obtenir X_{an} ("an" pour "analyse"⁶) une estimation de X_{lo} basée sur les observations Y_{lo} :

$$X_{an} = K_{an} \cdot Y_{lo} \quad \text{avec} \quad Y_{lo} = [Y_l, Y_o], \quad (\text{V.36})$$

avec K_{an} une matrice $12 \times (n_l + n_o)$.

Dans l'approche SW, chaque composante du cycle de l'eau est pondérée sur la base de son incertitude *a priori* (Section IV.3.2) mais aucune contrainte de fermeture n'est imposée à la solution X_{SW} . L'approche PF a été introduite (voir (Pan et Wood, 2006) et (Aires, 2014)) pour imposer la contrainte de fermeture sur une solution obtenue préalablement. Ici, nous avons utilisé X_{SW} comme première estimation sur le vecteur d'état X_{lo} . Dans (Aires, 2014), le PF a été utilisé et testé sans modèle, comme une simple étape de post-traitement après le SW. Suite à (Yilmaz et al., 2011), on a mis en œuvre le filtre PF avec une contrainte de fermeture relaxée caractérisée par sa covariance d'incertitude $\Sigma = Id \cdot 4 \text{ mm}$.

$$X_{PF} = (I - K_{PF} \cdot G_{lo} \Sigma^{-1} G_{lo}^t) \cdot X_{SW}, \quad (\text{V.37})$$

où $K_{PF} = (B_{lo}^{-1} + G_{lo} \Sigma^{-1} G_{lo}^t)^{-1}$ et B_{lo} est la matrice de covariance d'erreur de X_{SW} . On peut ré-écrire l'équation précédente à l'aide de Y_{lo} et expliciter le filtre K_{an} de l'Eq. (V.36) :

$$\begin{aligned} X_{PF} &= (I - K_{PF} \cdot G_{lo} \Sigma^{-1} G_{lo}^t) \cdot K_{merge} \cdot Y_{lo}, \\ X_{PF} &= K_{an} \cdot Y_{lo}, \\ &= X_{an}. \end{aligned} \quad (\text{V.38})$$

L'étape PF consiste à répartir le résidu du bilan hydrique entre les douze composantes. Cette technique permet d'obtenir une fermeture satisfaisante du bilan hydrique indépendamment pour chaque bassin. On peut noter que le terme résiduel r pourrait être réduit dans l'approche SW+PF en diminuant la variance Σ dans l'Eq. (V.37). Si la relaxation est trop faible, la fermeture est assurée mais au détriment des incertitudes de nos hypothèses (comme l'eau souterraine inconnue) et d'une certaine incertitude quant à l'aire exacte du bassin de drainage utilisé pour calculer la moyenne spatiale des composantes du cycle l'eau.

Suivant Munier et al. (2014) nous avons imposé la fermeture du bilan hydrique indépendamment sur les basses et les hautes fréquence du signal des composantes du

6. Attention, ce terme, "analyse", n'est pas à prendre dans le sens de l'assimilation; aucun modèle n'est utilisé.

cycle de l'eau afin d'éviter que la fermeture n'impacte les variables de basse fréquence telles que l'évapotranspiration (principalement due à la croissance annuelle de la végétation (Allen et al., 1998)). Pour cela, nous avons d'abord décomposé le signal de chaque variable en composantes hautes et basses fréquences (en utilisant une décomposition par "Transformation de Fourier rapide", FFT en anglais) en considérant une fréquence de séparation égale à 6 mois. La fermeture est ensuite appliquée indépendamment sur les basses et les hautes fréquences. La composante haute fréquence de E n'est alors pas incluse dans l'équation de fermeture du cycle de l'eau continental. Aucune hypothèse n'est faite sur l'évaporation océanique qui peut, elle, avoir une composante de haute fréquence. La linéarité de PF et FFT assure la fermeture du bilan hydrique par recombinaison. Dans le cadre d'une fermeture multi-échelles temporelles (Section 2.4), la contrainte annuelle n'est appliquée qu'à la fermeture du budget en basse fréquence.

La Figure V.5 montre les différences de correction apportées par la fermeture du bilan hydrique continental avec ou sans décomposition en fréquence. On peut voir que la contrainte sans décomposition fréquentielle engendre des hautes fréquences dans la correction du signal de l'évapotranspiration. Après fermeture par décomposition/recombinaison, le bilan hydrique continental reste assuré.

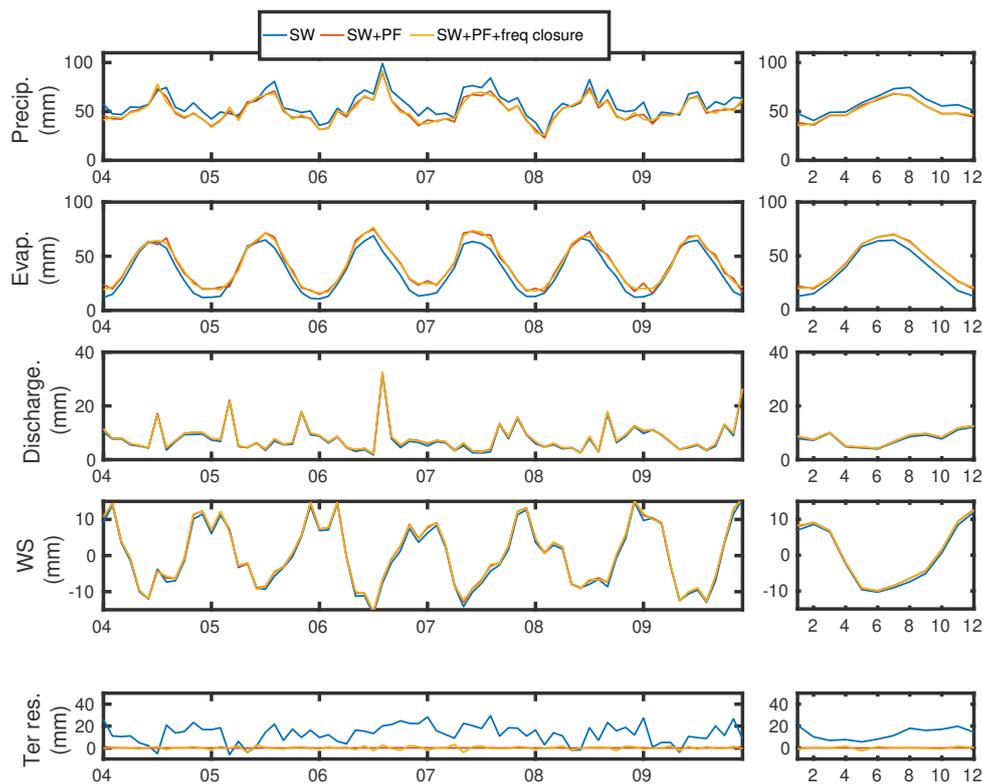


Figure V.5. Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'intégration est effectuée sans (en rouge) ou avec (en jaune) décomposition fréquentielle. La contrainte sans décomposition fréquentielle engendre des hautes fréquences dans la correction du signal de l'évapotranspiration.

2.3 Fermeture inter-dépendante des sous-bassins

Une contrainte de fermeture à l'échelle du bassin méditerranéen ne permet pas d'assurer une fermeture à l'échelle du sous-bassin. La Figure V.6 montre l'impact de la contrainte hydrique globale (à l'échelle de bassin entier) sur le résidu du bilan hydrique du sous-bassin BLS comparé à une contrainte de fermeture à l'échelle du sous-bassin, suivant la méthode SW+PF. Ainsi la courbe CCM correspond à l'estimation des résidus du bilan hydrique, après que toutes les composantes du cycle de l'eau ont été corrigées, telle que la fermeture du cycle de l'eau soit assurée à l'échelle du bassin entier. La Figure V.6 montre que la contrainte de fermeture à l'échelle du bassin entier n'est pas suffisante pour assurer la fermeture du cycle à une échelle spatiale plus fine.

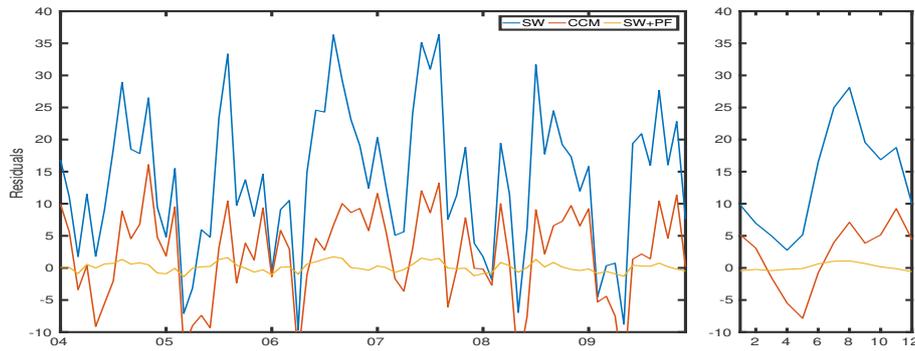


Figure V.6. Série temporelle entre 2004 et 2009 du résidu du bilan hydrique au niveau du sous-bassin BLS avec une correction issue d'une fermeture seulement à l'échelle du bassin méditerranéen total (CCM) et une fermeture à l'échelle du sous-bassin BLS (SW+PF).

Il est toutefois possible d'imposer une fermeture du bilan hydrique simultanément sur les six sous-bassins continentaux, sur l'ensemble du bassin continental, sur l'océan (Méditerranée et mer Noire) et dans l'atmosphère. Considérons le vecteur d'état du cycle de l'eau total :

$$X^t = [X_l^{(1)}, X_l^{(2)}, X_l^{(3)}, X_l^{(4)}, X_l^{(5)}, X_l^{(6)}, X_o]^t. \quad (V.39)$$

Ce vecteur inclut les 6 vecteurs d'états X_l^i de chaque sous-bassin continental i de surface $A_l^{(i)}$ et la partie océanique o . La matrice de fermeture devient alors :

$$G_{lo} = \begin{pmatrix} G_l^{(1)} & 0 & \dots & 0 & 0 \\ 0 & G_l^{(2)} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & G_l^{(6)} & 0 \\ L_{lo}^{(1)} & L_{lo}^{(2)} & \dots & L_{lo}^{(6)} & G_o \end{pmatrix}, \quad (V.40)$$

avec

$$\begin{aligned}
G_l^{(i)} &= \begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}, \\
L_{lo}^{(i)} &= \begin{bmatrix} 0 & 0 & \frac{A_l^{(i)}}{A_{sea}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \\
G_o &= \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 1 \\ -1 & 1 & 0 & -1 & -1 & 0 \end{bmatrix}.
\end{aligned} \tag{V.41}$$

La dernière ligne de G_{lo} représente la fermeture globale du cycle de l'eau, dans tous les sous-bassins et avec l'océan. La dimension des matrices de covariance B_{lo} et Σ est augmentée en fonction de la nouvelle taille du vecteur d'état X_{lo} . On considère une plus faible incertitude sur les deux dernières équations $G_{lo}(13 : 14, :)$ avec $\Sigma(13, 13) = \Sigma(14, 14) = 2 \text{ mm}$. Cette fermeture plus contrainte à l'échelle du bassin total permet une meilleure correction de la divergence d'humidité pour laquelle les flux entre sous-bassins sont négligés lors de leurs fermetures inter-dépendantes. Aucun terme croisé dans B_{lo} et Σ n'est inclus, ce qui signifie qu'il n'y a pas de dépendance d'erreurs dans la première estimation X_{SW} , ainsi que dans la fermeture entre les sous-bassins.

La Figure V.7 montre comment la fermeture inter-dépendante des sous-bassins avec le bassin total et la partie océanique assure un équilibre du bilan hydrique multi-échelle : à grande échelle (bassin total) et à l'échelle du sous-bassin (voir Figure V.9).

2.4 Contraintes hydrologiques à plusieurs échelles temporelles

Il est également possible d'imposer une contrainte sur la fermeture du cycle de l'eau simultanément à l'échelle mensuelle et annuelle. Une fermeture mensuelle assure une fermeture annuelle par sommation, bien que la relaxation ne soit pas la même à l'échelle mensuelle (moins forte) ou annuelle (plus forte).

Mais la fermeture annuelle nous permet d'imposer une fermeture en supposant qu'aucun changement de contenu en eau dans le sol n'est intervenu à l'échelle annuelle, ce qui représente une contrainte supplémentaire sur ΔS_l et permet de s'assurer qu'aucun biais n'est introduit pour cette variable pendant le processus PF. Dans ce cadre, les fermetures mensuelles deviennent inter-dépendantes dans l'année et le nouveau vecteur d'état devient :

$$X_{year}^t = [X^{Jan}, \dots, X^{Dec}]^t, \tag{V.42}$$

avec X^m le vecteur d'état total X défini dans Eq. (V.39), pour le mois m . La fermeture est appliquée indépendamment pour les quatre années de la période 2004-2009 mais les 12 mois de chaque année ont maintenant une inter-dépendance dans leur fermeture.

La matrice de fermeture GA_{lo} qui inclut la fermeture pour les douze mois de l'année et l'année complète est dérivée de la contrainte mensuelle G_{lo} et est définie comme suit :

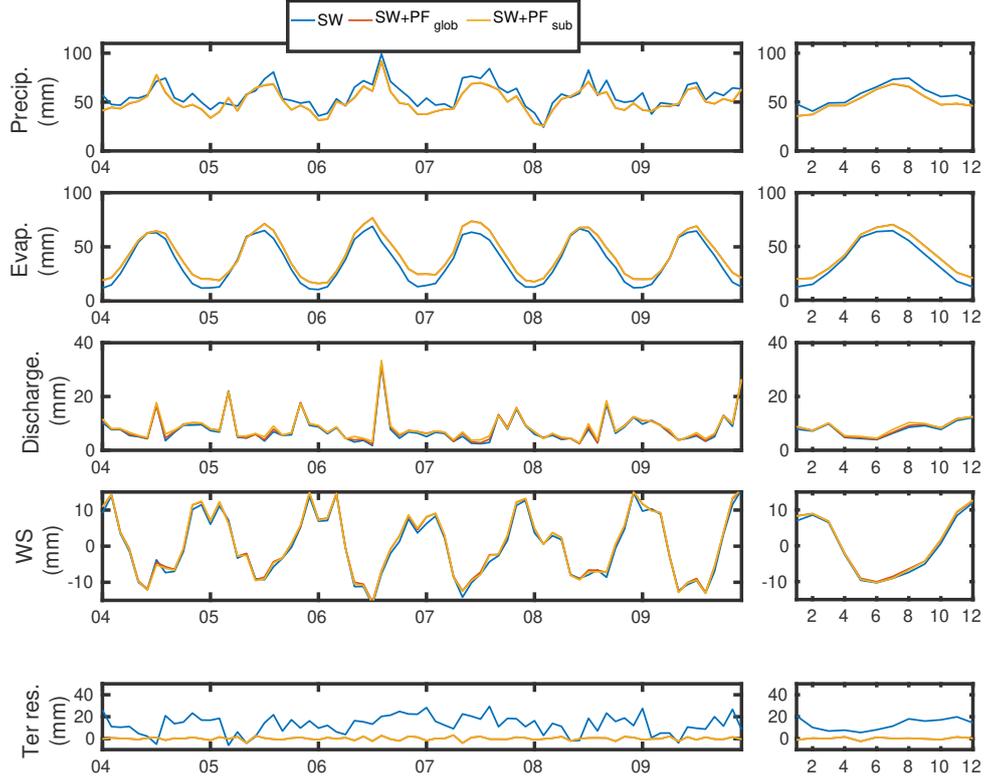


Figure V.7. Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'une des intégration (en rouge) ne considère comme contrainte seulement le cycle de l'eau à l'échelle du bassin méditerranéen entier, l'autre (en jaune) contraint le cycle de l'eau dans sa globalité et à l'échelle du sous-bassin et de façon interdépendante entre les sous-bassins. La fermeture inter-dépendante des sous-bassins avec le bassin total et la partie océanique assure un équilibre du bilan hydrique à grande échelle.

$$GA_{lo} = \begin{pmatrix} G_{lo} & 0 & \cdots & 0 \\ 0 & G_{lo} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & G_{lo} \\ N_{lo} & N_{lo} & \cdots & N_{lo} \end{pmatrix}, \quad (\text{V.43})$$

avec N_{lo} une version modifiée de la matrice G_{lo} dans laquelle $G_l^{(i)}$ est ré-écrit $N_l^{(i)}$ en négligeant le terme de contenu en eau ΔS_l :

$$N_l^{(i)} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & -1 & -1 \end{bmatrix}. \quad (\text{V.44})$$

La dernière ligne de GA_{lo} représente la fermeture du budget annuel sans changement du contenu en eau dans le sol à l'échelle annuelle. La dimension des matrices de covariance B_{lo} et Σ est de nouveau augmentée suivant la taille du vecteur d'état X_{year} . Aucun terme croisé dans B_{lo} et Σ sont inclus, ce qui signifie qu'il n'y a pas de

	Climatic sub-basins										LAND		OCEAN	
	MA-DZ-TN		ES-Pyr		Alp-IT-ADR		GR-TR-IL		BLS		surf	atm	surf	atm
	surf	atm	surf	atm	surf	atm	surf	atm	surf	atm				
ERA-I	34.3	15.3	37.8	18.1	31.2	13.7	30.6	12.0	18.0	8.0	13.6	13.8	86.7	6.2
OS	25.1	36.0	27.5	43.5	28.5	37.7	25.8	39.7	25.4	27.3	19.8	15.1	75.2	24.7
SW	18.2	31.8	17.5	40.7	21.5	38.3	17.6	35.6	25.1	26.5	16.6	16.6	74.3	15.7
SW+PF	4.46	3.04	4.38	3.99	4.42	3.07	4.46	3.21	3.64	2.82	2.78	2.28	7.18	3.13
	75%	90%	74%	90%	79%	91%	74%	90%	85%	89%	83%	91%	91%	80%

TABLE V.4 – RMS des résidus du bilan hydrique mensuel (en *mm*) sur les sous-bassins en utilisant la solution OS, SW et SW+PF, pour la période 2004-2009. Le pourcentage d'amélioration du RMS des résidus, par rapport à la solution SW, des méthodes d'intégration est également indiqué. À des fins de comparaison, les résultats obtenus avec ERA-I sont également représentés.

dépendance dans la première estimation ni dans l'erreur de fermeture entre les mois.

Les résidus du bilan hydrique de surface (terrestre ou océanique) et atmosphérique pour la région méditerranéenne sont calculés à l'échelle mensuelle, sur la période 2004-2009. Les racines carrées de la moyenne quadratique (RMS) des résidus sont résumées dans la Table V.4 pour les quatre types de solution (ERA-I, OS, SW, SW+PF). Le pourcentage d'amélioration du RMS des résidus des méthodes d'intégration par rapport à la solution SW est également indiqué à des fins de comparaison.

ERA-I représente la fermeture du bilan hydrique en utilisant le produit de ré-analyses pour toutes les variables à l'exception du contenu en eau et du ruissellement (afin de maintenir la cohérence de la comparaison). Il convient de noter qu'ERA-I n'a pas de contrainte de conservation d'eau. La sélection optimale (OS) est donnée par la combinaison de produits suivante : Précipitation TMPA ; évapotranspiration GLEAM et évaporation OAFflux ; changement de stockage d'eau GRGS sur terre et JPL sur mer ; décharge ORCHIDEE-CEFREM forcée par GPCC ; et la vapeur Globvapor pour le changement de vapeur d'eau atmosphérique (voir Section IV.3.1). Comme on peut le voir dans (Aires, 2014 ; Munier et al., 2014 ; Pellet et al., 2018), la procédure de fusion SW réduit les résidus du budget du cycle de l'eau à l'échelle du sous-bassin, en réduisant les erreurs aléatoires des produits satellites. Le produit surpasse la ré-analyse ERA-I et le produit OS. Cependant, la fermeture n'est généralement pas satisfaisante avec cette technique.

La solution SW+PF ferme le bilan hydrique sur tous les sous-bassins, à la surface comme dans l'atmosphère, avec un RMS du résidu d'environ 4 *mm*. Les résidus du bilan hydrique de surface sont considérablement réduits : de 72 % sur le sous-bassin GR-TR-IL et jusqu'à 94 % pour la mer Méditerranée. Cela montre la nécessité d'utiliser une contrainte de fermeture du bilan hydrique qui relie les six composantes du cycle de l'eau et les rend plus cohérentes entre elles.

Cette technique SW+PF ne peut traiter que des séries temporelles (la moyenne spatiale sur les sous-bassins considérés), et non des cartes (pixels) puisque le ruissellement n'est pas disponible à cette résolution. Par conséquent, afin d'obtenir un produit à composantes multiples qui ferment le bilan hydrique et qui ont des caractéristiques spatiales, d'autres techniques doivent être utilisées.

3 Extrapolation spatiale et temporelle

Dans cette section, on détaille une approche d'extrapolation spatiale qui permet d'étendre, du bassin au pixel, la correction liée à la contrainte hydrologique (SW+PF). La spatialisation mois par mois de la fermeture avec interpolation et extrapolation basée sur la distance géographique à l'extérieur des sous-bassins que nous présenterons dans la suite est la technique la plus naturelle dans le contexte du pourtour méditerranéen. Pour d'autres environnements ou pour des études en global, une autre méthode d'extrapolation spatiale est présentée dans l'Annexe B. Cette méthode issue de (Munier et Aires, 2017) est basée sur une régression statistique qui permet d'appliquer au niveau du pixel l'équivalence de la contrainte de fermeture. Grâce à un indice pixelaire, il est possible de moduler cette correction à partir de caractéristiques de surface, ce qui permet une plus grande flexibilité et une prise en compte de caractéristiques locales pour l'interpolation.

Dans un second temps, je présenterai une méthode d'extrapolation temporelle qui permet d'étendre l'impact de fermeture au-delà de la période commune pour les produits satellites (2004-2009).

3.1 Extrapolation spatiale mois par mois, suivant la distance géographique : INT

Contrairement à l'approche présentée par Munier et Aires (2017) qui tend à approximer la correction liée à la contrainte hydrologique par un apprentissage statistique (voir Annexe B), on cherche ici à spatialiser directement la solution SW+PF sur la période 2004-2009 en utilisant une méthodologie : INT basée sur une inter-/extrapolation. La méthodologie est présentée dans ses différentes étapes dans la Figure V.8 pour les précipitations et l'évaporation, et pour un mois particulier. Dans cette approche, nous ne considérons que les sous-bassins méditerranéens et leur voisinage proche, de sorte qu'une simple interpolation spatiale de la correction de fermeture peut être suffisante.

La solution SW+PF (Figure V.8, seconde ligne) permet la minimisation des résidus du bilan hydrique, pour chaque mois $m = 1, \dots, 72$ de la période 2004-2009 mais nous n'obtenons qu'une seule valeur pour chaque mois et chaque bassin.

La méthode INT nécessite un facteur d'échelle pour passer de la solution SW à la solution SW+PF à l'échelle du sous-bassin.

Nous définissons $\beta^{(i)}(m) = P_{PF}^{(i)}(m) / P_{SW}^{(i)}(m)$ (pour les précipitations ici), le rapport entre la solution SW et la solution SW+PF, pour chaque sous-bassin i et au mois m . Ce facteur est utilisé pour amener la solution SW vers la solution SW+PF à l'échelle du bassin, pour un mois particulier m , de la manière suivante :

$$P_{INT}^{(i)}(m) = \beta^{(i)}(m) \cdot P_{SW}^{(i)}(m) \quad (V.45)$$
$$(\text{=} P_{PF}^{(i)}(m)).$$

Pour le changement du contenu en eau et la divergence d'humidité, ce facteur β peut être négatif. Dans ce cas on utilise, une correction homogène $\gamma^{(i)}(m) = P_{PF}^{(i)}(m) - P_{SW}^{(i)}(m)$:

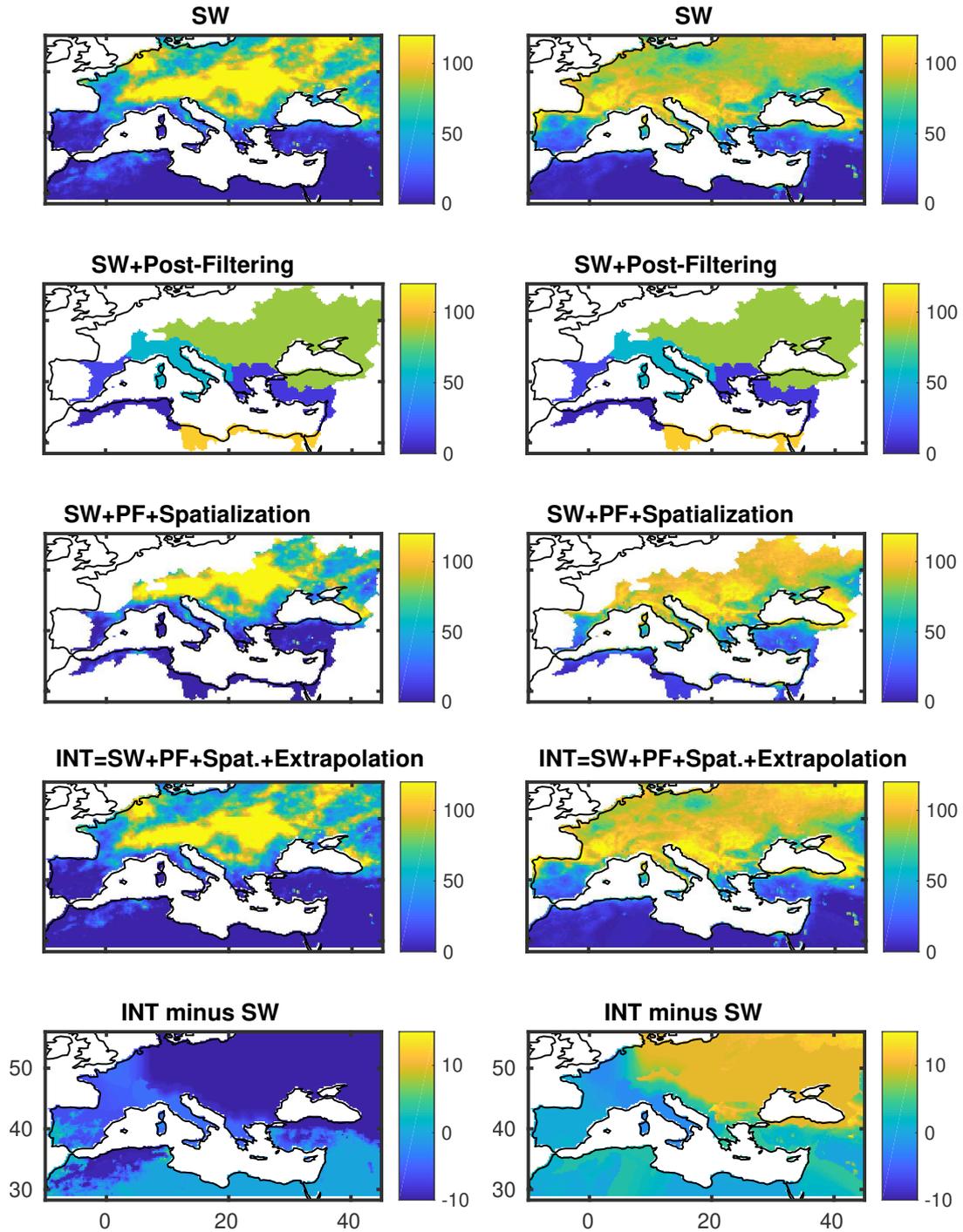


Figure V.8. Etapes de la spatialisation de la contrainte hydrologique depuis SW jusqu'à INT pour la précipitation (à gauche) et l'évapotranspiration (à droite), pour juillet 2008. Les unités sont en $mm.mois^{-1}$.

$$\begin{aligned} \Delta S_{INT}^{(i)}(m) &= \Delta S_{SW}^{(i)}(m) + \gamma^{(i)}(m) \\ &= \Delta S_{PF}^{(i)}(m). \end{aligned} \quad (V.46)$$

Le facteur β est défini à l'échelle du sous-bassin, mais s'il est interpolé spatialement, il pourrait être utilisé à l'échelle du pixel pour obtenir une solution véritablement

spatialisée. On définit alors le facteur pixelaire α de telle sorte que : pour chaque pixel j dans le sous-bassin i , pour chaque mois m : $\alpha(j, m) = \beta^{(i)}(m)$ (ou $\gamma^{(i)}(m)$). Lorsqu'elle est utilisée telle quelle, la convolution de SW par α permet la spatialisation de la fermeture des sous-bassins (Figure V.8, troisième ligne) avec :

$$\begin{aligned} \iint_{j \in A_i^{(i)}} P_{SW}(j, m) \cdot \alpha(j, m) &= \beta^{(i)}(m) \cdot P_{SW}^{(i)}(m) \\ &= P_{INT}^{(i)}(m). \end{aligned} \quad (V.47)$$

Cependant, ce produit présente non seulement une discontinuité entre les sous-bassins (où différents facteurs d'échelle β sont définis), mais aucune valeur ne peut être fournie en dehors des sous-bassins.

Pour résoudre ces deux problèmes, les facteurs pixelaires α sont interpolés/extrapolés :

- *Interpolation* - Une région de 200 km de part et d'autre de la frontière entre deux sous-bassins i_1 et i_2 est définie, et une interpolation lissante est effectuée entre les deux facteurs d'échelle $\beta^{(i_1)}(m)$ et $\beta^{(i_2)}(m)$ en fonction de la distance à la frontière. Cette interpolation des facteurs α entre deux sous-bassins peut introduire des erreurs (les résidus de fermeture peuvent légèrement augmenter) mais il sera montré que cet effet est limité (voir Table V.6) et que les équations entre parenthèses dans les Eqs. (V.45)-(V.46) sont maintenues.
- *Extrapolation* - Une extrapolation des indices α est alors effectuée pour avoir un facteur α en dehors du domaine des sous-bassins. Cette extrapolation est pondérée en fonction des distances respectives des deux sous-bassins les plus proches.

Le produit INT est la convolution entre la solution SW et la carte des facteurs pixelaires α interpolés et extrapolés qui en résulte. INT est alors une version optimisée de SW dans laquelle la fermeture du cycle de l'eau a été prise en compte à l'échelle du pixel.

La quatrième ligne de la Figure V.8 montre le produit INT résultant et sa couverture spatiale. Les problèmes de continuité entre les sous-bassins ont été résolus et l'extrapolation permet une couverture spatiale sur l'ensemble du domaine. La différence entre l'estimation SW et INT, représentée dans la dernière ligne de la Figure V.8, est alors directement liée aux facteurs α interpolés. La discontinuité entre les sous-bassins est lissée. Le nord de l'Europe hors France est principalement déterminé par le facteur d'échelle lié au sous-bassin BLS. Ceci est conforme à la classification climatique de Köppen (1936), en effet le nord et le centre de l'Europe sont soumis au même climat des latitudes moyennes.

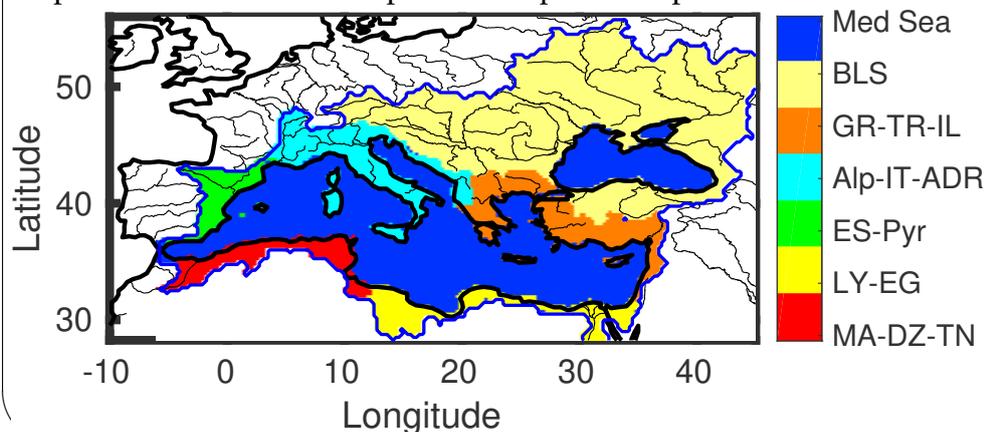
Comme la solution SW+PF n'est disponible que sur la période 2004-2009, INT n'est défini que sur cette période.

3.2 Vers une extrapolation temporelle via une calibration (CAL)

Pour obtenir la solution INT, de nombreux produits satellites ont été pris en compte : plusieurs produits pour chaque composante hydrologique (la partie SW) et avec toutes

Réduction de la région pour INT

Bien que la solution SW+PF existe pour le sous-bassin du Nil, LY-EG, les caractéristiques équatoriales et tropicales de la source de ce fleuve rendent l'extrapolation des résultats de fermeture trop délicate à ces basses latitudes. Le désert du Sahara offrant une frontière naturelle, nous avons considéré une région réduite aux pays du pourtour méditerranéen pour l'extrapolation spatiale avec INT.



les composantes du cycle de l'eau (la partie PF). Cependant, si l'un des produits est manquant, la solution INT ne peut pas être estimée.

Dans cette section, un modèle (statistique) de correction de fermeture est introduit pour corriger, indépendamment, chaque composante du cycle de l'eau sur la base des résultats de l'intégration INT. Le modèle de correction de fermeture corrige par exemple directement la précipitation, sans l'utilisation des autres composantes (E , R , ΔS , ΔWV et Div). Le modèle de correction de fermeture est représenté par des transformations linéaires simples définies par un facteur d'échelle a , un biais b et un coefficient c qui permet de traiter le cas où la variable est nulle afin de ne pas avoir des valeurs négatives.

$$P_{CAL} = a \cdot P_{SW} + b(1 - e^c \cdot P_{SW}). \quad (V.48)$$

Cette équation est valable pour toutes les composantes du cycle de l'eau. Les paramètres a et b du modèle sont calculés grâce aux régressions entre les estimations P_{SW} et P_{INT} au niveau du pixel sur la période 2004-2009, puis la relation peut être propagée sur la période entière de P_{SW} (1980-2012) pour obtenir une base de données spatiale avec longue série temporelle.

Cette calibration permet d'obtenir une solution de plus long terme à l'échelle du pixel comme la solution SW (Table V.5), mais avec des statistiques de fermeture du cycle de l'eau plus proches de la solution INT. Dans nos tests, la régression linéaire est satisfaisante pour la calibration, et il n'a pas été jugé nécessaire d'utiliser un outil de régression statistique plus complexe tel qu'un réseau de neurones.

Les diverses techniques de fusion/intégration présentées et utilisées dans cette thèse sont comparées dans la Table V.5.

	EO merging	Spatial resol.	Coverage period*	WC budget closure
OS	no	pixel	1993-2012	- -
SW	yes	pixel	1980-2012	-
SW+PF	yes	basin scale	2004-2009	++
INT	yes	pixel	2004-2009	++
CAL	yes	pixel	1980-2012	+

TABLE V.5 – Principales caractéristiques des cinq méthodes développées dans cette thèse : EO signifie "Earth Observation" (satellite d'observation de la Terre), * ne tient pas compte de la période de GRACE. La dernière colonne montre la capacité de la méthodologie à fermer le bilan hydrique. Le terme "- -" signifie une très mauvaise fermeture, "-" signifie une mauvaise fermeture, "+" signifie une bonne fermeture et "++" signifie une très bonne fermeture.

4 Résultats

Dans cette section, les solutions INT et CAL, présentées précédemment, sont d'abord évaluées en terme de fermeture du cycle de l'eau. Les données *in situ* disponibles (précipitations et évapotranspiration) sont ensuite utilisées pour valider les composantes de l'eau des solutions intégrées. La validation est effectuée à deux échelles spatiales différentes : l'échelle du sous-bassin (en utilisant une moyenne spatiale de la composante sur chaque bassin) et l'échelle du pixel. Dans un second temps, les bases de données obtenues sont utilisées afin de décrire complètement et de façon cohérente le cycle de l'eau dans la région méditerranéenne.

4.1 Validation des nouvelles solutions intégrées

4.1.a Équilibre du bilan en eau

L'impact de la contrainte hydrologique (PF) ainsi que des processus INTegration (INT) et CALibration (CAL) sur la moyenne spatiale des estimations des composantes du cycle de l'eau et des résidus du bilan hydrique sur les différents sous-bassins méditerranéens est résumé sur la Figure V.9.

Comme pour la Table V.4, les résidus du bilan hydrique de surface (terrestre ou océanique) et atmosphériques pour la région méditerranéenne sont calculés à l'échelle mensuelle, sur la période 2004-2009. Les statistiques de la RMS de ces résidus sont résumées dans la Table V.6 pour les six produits considérés (ERA-I, OS, SW, SW, SW+PF, INT et CAL). Le pourcentage d'amélioration des solutions avec contrainte (par rapport à la solution SW) est également indiqué à des fins de comparaison. Cette table a déjà été commentée dans la section 2.4 pour les méthodes ERA-I, OS, SW et SW+PF.

Le produit INT fournit des résultats de fermeture du cycle de l'eau satisfaisants (de 61% à 94%), même s'ils sont légèrement dégradés par rapport à la solution SW+PF (en raison du processus d'interpolation entre sous-bassins). Puisqu'aucune interpolation n'a été appliquée sur la mer Méditerranée, les résidus de la solution INT sont égaux à ceux de SW+PF pour la partie océanique.

	Climatic sub-basins										LAND		OCEAN	
	MA-DZ-TN		ES-Pyr		Alp-IT-ADR		GR-TR-IL		BLS		surf	atm	surf	atm
	surf	atm	surf	atm	surf	atm	surf	atm	surf	atm				
ERA-I	34.3	15.3	37.8	18.1	31.2	13.7	30.6	12.0	18.0	8.0	13.6	13.8	86.7	6.2
OS	25.1	36.0	27.5	43.5	28.5	37.7	25.8	39.7	25.4	27.3	19.8	15.1	75.2	24.7
SW	18.2	31.8	17.5	40.7	21.5	38.3	17.6	35.6	25.1	26.5	16.6	16.6	74.3	15.7
SW+PF	4.46	3.04	4.38	3.99	4.42	3.07	4.46	3.21	3.64	2.82	2.78	2.28	7.18	3.13
	75%	90%	74%	90%	79%	91%	74%	90%	85%	89%	83%	91%	91%	80%
INT	5.23	5.82	5.15	6.47	7.70	7.65	6.62	8.16	4.21	3.20	3.79	4.07	7.18	3.13
	71%	81%	70%	84%	64%	80%	62%	77%	83%	87%	77%	84%	91%	80%
CAL	13.14	14.48	13.38	17.77	20.13	20.21	14.51	16.77	18.00	13.03	12.79	11.44	24.63	12.50
	27%	54%	23%	56%	6%	47%	17%	52%	28%	50%	22%	56%	66%	17%

TABLE V.6 – RMS des résidus du bilan hydrique (en $mm.mois^{-1}$) calculés sur le sous-bassin en utilisant les solutions OS, SW, SW+PF, INT et CAL et pour la période 2004-2009. Le pourcentage d'amélioration du RMS des résidus de la solution SW aux méthodes contraintes est également indiqué. À des fins de comparaison, les résultats obtenus à l'aide de l'ensemble de données ERA-I sont également représentés.

Le produit CAL améliore moins les résidus du bilan hydrique par rapport à INT. Néanmoins, les résidus liés à ces produits sont réduits sur l'ensemble des sous-bassins.

La Figure V.9 donne la série temporelle 2004-2009 de toutes les estimations des composantes de l'eau pour les différentes méthodologies (SW, SW+PF, INT et CAL) sur les différents sous-bassins ainsi que la fonction de densité de probabilité (pdf) du résiduel. Cette figure montre l'impact de la fermeture du cycle de l'eau sur les séries chronologiques 2004-2009.

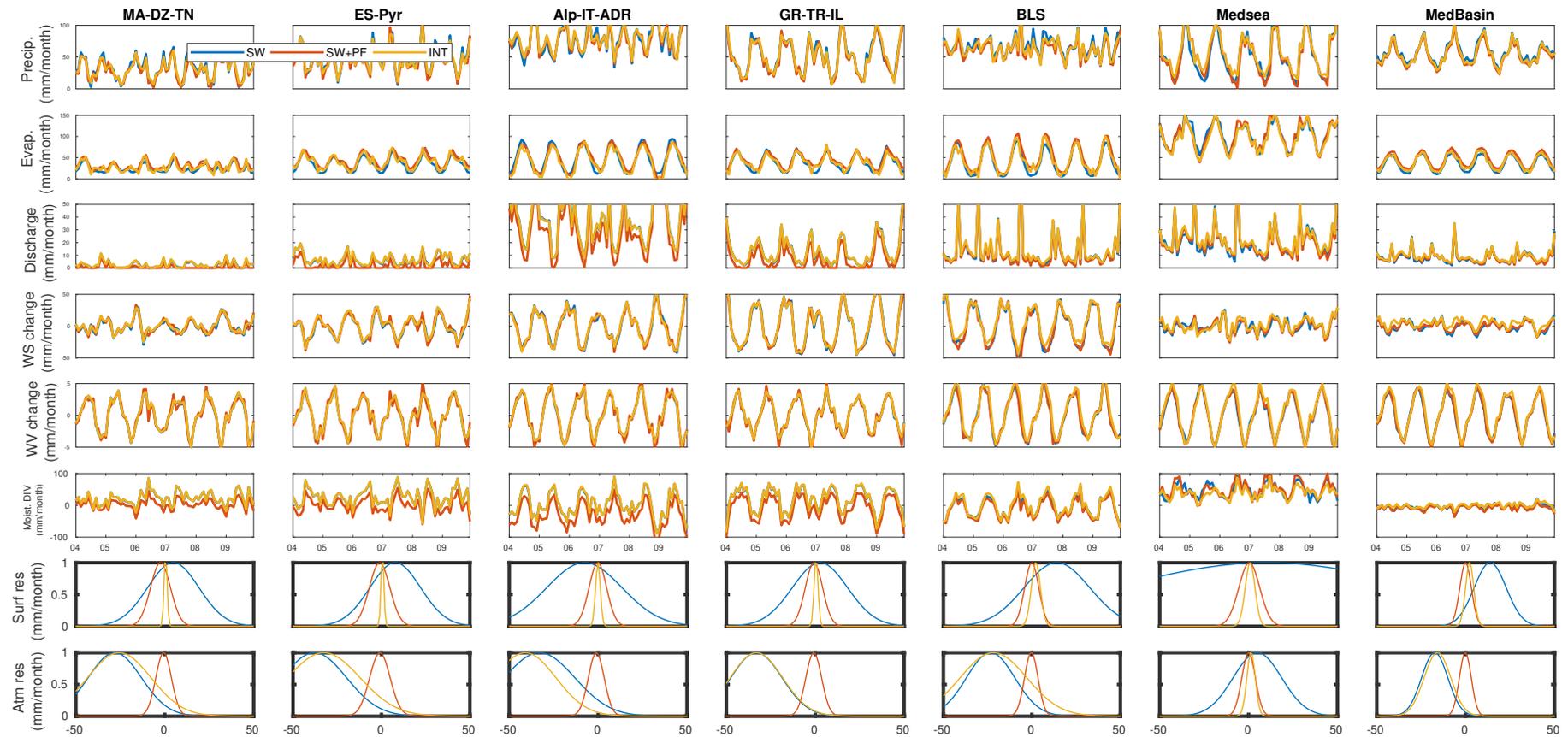


Figure V.9. Comparaison des estimations des composantes du cycle de l'eau et de la densité de probabilité des résidus du bilan hydrique (en ligne). Les estimations concernent la période 2004-2009 sur les 6 sous-bassins continentaux, la partie océanique et le bassin méditerranéen total (en colonne) issus des différentes méthodologies présentées dans l'étude : SW, SW+PF, INT et CAL.

4.1.b Évaluation à l'échelle du sous-bassin

Comme la contrainte de fermeture du cycle de l'eau a été imposée à l'échelle du sous-bassin (voir la Section V.2.4), l'évaluation du produit intégré se fait également à cette échelle. Deux métriques sont utilisées ici, la RMS des Différences (RMSD) avec les mesures *in situ* et la corrélation (CORR). Seules les solutions SW+PF, INT et CAL sont comparées avec SW.

Précipitations continentales - La validation des précipitations présente l'avantage de disposer d'un plus grand nombre de mesures (jauges) que pour toute autre composante du cycle de l'eau. Le produit de précipitation EOBS présenté au Chapitre III est utilisé ici.

La Table V.7 fournit la comparaison des solutions SW, SW+PF, INT et CAL avec le produit de précipitations griddées EOBS en termes de corrélation temporelle (aux échelles mensuelles et du sous-bassin), et de RMSD, pour chaque sous-bassin et pour l'échelle "continentale" (l'espace continental est défini dans la Figure de l'encart "Réduction de la région pour INT", Section V.3.1). Comme le produit SW+PF n'est défini qu'à l'échelle des sous-bassins, aucune statistique n'est présentée pour cette approche sur la région continentale (dernière colonne). Pour la RMSD, les résultats sont également fournis avec l'amélioration par rapport à la solution SW.

		Climatic sub-basins					Continental
		MA-DZ-TN	ES-Pyr	Alp-IT-ADR	GR-TR-IL	BLS	
Correlation	SW	0.81	0.88	0.87	0.87	0.79	0.78
	SW+PF	0.84	0.90	0.88	0.87	0.81	-
	INT	0.84	0.89	0.88	0.87	0.81	0.80
	CAL	0.81	0.88	0.87	0.87	0.79	0.79
RMSD	SW	14.01	16.69	21.78	23.04	20.56	15.68
	SW+PF	13.60	14.10	22.42	21.98	16.64	-
		2%	15%	-3%	4%	19%	-
	INT	13.59	14.35	21.88	21.83	16.84	12.93
		2%	14%	-1%	5%	18 %	17%
	CAL	14.00	14.83	22.06	21.64	17.23	13.16
		0%	11%	-2%	6%	16%	16%

TABLE V.7 – Comparaison des solutions de précipitation SW, SW+PF, INT et CAL avec l'ensemble de données EOBS, en termes de corrélation, de RMSD (en $mm.mois^{-1}$) et de pourcentage d'amélioration de la RMSD par rapport à la solution SW. Ces scores sont calculés sur chaque sous bassin continental et la région entière couverte par le produit EOBS (dernière colonne).

Sur l'ensemble des sous-bassins, la méthodologie SW+PF améliore les résultats par rapport à la méthode, sans contrainte, SW. Même si la corrélation de SW avec EOBS est déjà bonne, la contrainte de fermeture améliore jusqu'à 0.84 la corrélation sur le sous-bassin MA-DZ-TN, jusqu'à 0.81 sur le grand sous-bassin BLS, et même sur les sous-bassins complexes comme Alp-IT-ADR. SW+PF réduit également la RMSD avec EOBS (jusqu'à 20 %). Ces résultats montrent l'intérêt de la contrainte de fermeture sur les précipitations. Sans contraindre explicitement les produits de précipitations satellites vers les données *in situ*, en prenant en compte la contrainte de fermeture du cycle de leau, SW+PF améliore les statistiques des produits satellites avec des jauges.

Le produit INT présente une corrélation et des statistiques de RMSD similaires à celles de SW+PF sur les sous-bassins méditerranéens, avec une légère diminution de la corrélation avec EOBS sur le sous-bassin ES-Pyr. Sur la région continentale totale, INT améliore la corrélation par rapport à SW (0,80 pour 0,78) tout en réduisant de 17 % la RMSD. Par conséquent, le processus d'interpolation entre les sous-bassins (voir la spatialisation dans la Section 3.1) ne dégrade pas la solution dans les sous-bassins, tandis que l'extrapolation à l'extérieur des sous-bassins permet de dépasser les statistiques de SW sur l'ensemble du continent. INT présente des performances comparables à celles de SW+PF en termes de capacité de fermeture et de proximité aux mesures *in situ* avec l'avantage d'avoir une variabilité spatiale à l'échelle du pixel.

Enfin, le produit de précipitation CAL montre des résultats aussi bons que SW pour la corrélation (légèrement meilleurs pour l'ensemble de la région continentale), et une RMSD par rapport à EOBS plus faible. Le produit CAL ne ferme pas aussi bien le bilan hydrique que la solution INT, mais il a l'avantage d'être disponible sur une période plus longue (1980-2012) que INT.

Changement du niveau de la mer - Le changement de contenu en eau dans la mer Méditerranée (étroitement lié aux variations du niveau de la mer) est validé à l'aide de données altimétriques sur la période 2004-2009. Tout d'abord, les estimations de dilatation thermique (qui entraîne aussi une variation du niveau de la mer) des quatre bases de données présentées dans le Chapitre III sont fusionnées en une seule estimation. L'estimation de dilatation thermique est ensuite soustraite de l'estimation du niveau de la mer Méditerranée par l'altimètre AVISO. La variation mensuelle est alors calculée à l'aide du même filtre de dérivation que pour GRACE (voir Section III.1.3).

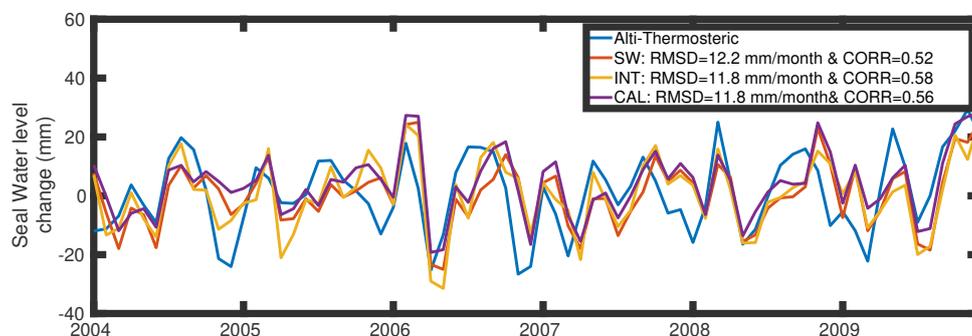


Figure V.10. Validation du contenu en eau sur mer des produits SW, INT et CAL par rapport à l'altimétrie moins le contenu thermique.

La Figure V.10 compare les variations du contenu en eau de l'estimation par altimétrie avec les estimations des différentes méthodologies présentées dans ce travail de thèse. Comme la mer Méditerranée est considérée sans la mer Noire pour la validation, il n'y a pas d'estimation SW+PF (qui regroupe la Méditerranée et la mer Noire). La légende de la Figure V.10 résume les performances des trois solutions satellites par rapport aux données altimétriques. Ici encore, l'estimation de INT surpasse la méthodologie SW tant en corrélation qu'en RMSD. CAL présente également de meilleurs résultats que SW mais la corrélation avec l'altimétrie est légèrement réduite par rapport à INT. La prise en compte de la fermeture du bilan hydrique améliore les estimations des changements de contenu en eau de la mer.

L'avantage de la méthodologie INT et de la méthodologie CAL est d'obtenir des composantes du cycle de l'eau plus cohérentes entre elles, mais aussi plus proches des données *in situ*.

4.1.c Évaluation à l'échelle du pixel

Les estimations INT et CAL sont ici évaluées à l'échelle du pixel, pour les précipitations et l'évapotranspiration. Les améliorations de INT et CAL, par rapport à SW, sont mesurées en utilisant les mesures des précipitations et de l'évapotranspiration de la base de données FLUXnet (voir la description au Chapitre III), disponible sur la région méditerranéenne, pour la période 2004-2009. Les sites européens FLUXnet sont principalement localisés en Allemagne, en France et en Italie. Pour chaque site, la différence entre la solution satellite (SW, INT et CAL) et l'estimation FLUXnet pour l'évapotranspiration et les précipitations est calculée sur la période 2004-2009.

La Figure V.11 présente les diagrammes de dispersion de la RMSD pour la précipitation. Sur la plupart des stations (82 %), les solutions INT et CAL améliorent l'estimation des précipitations par rapport à SW. L'emplacement des sites améliorés est indiqué avec le même code couleur que dans la Figure V.12.

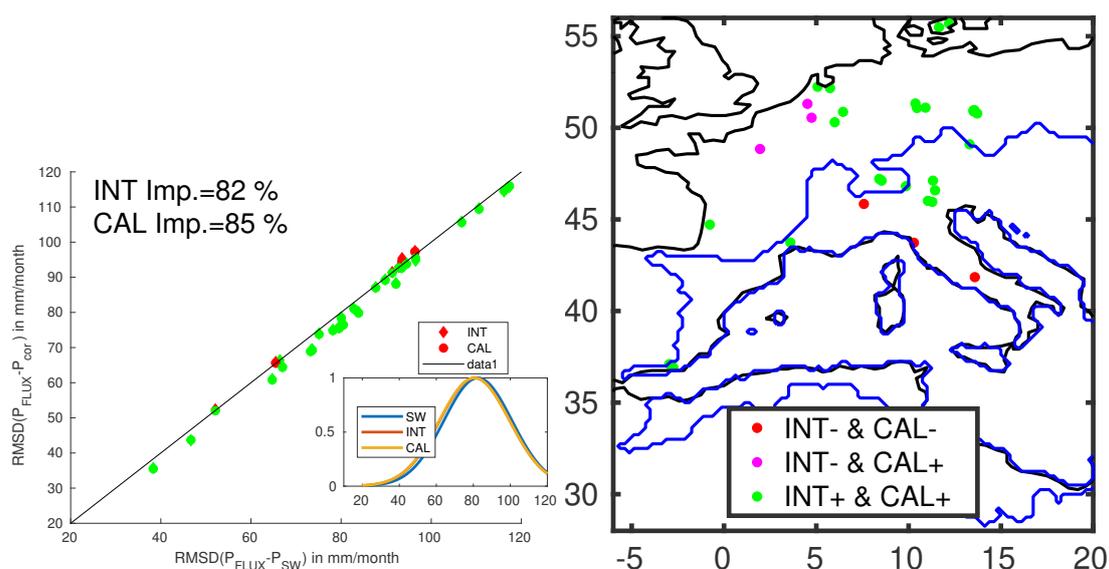


Figure V.11. Gauche : Diagramme de dispersion de la RMSD entre la station FLUXnet et les produits SW, INT et CAL, pour les précipitations. Les points sous la ligne 1:1 (verts) indiquent une amélioration et les points au-dessus de la ligne (rouges) indiquent une dégradation. La figure encapsulée montre la distribution des différences avec les estimations Fluxnet. Droite : Emplacement des stations FLUXnet utilisées pour la validation : les points verts indiquent une amélioration pour INT et CAL par rapport à SW, les points magenta indiquent une amélioration uniquement pour CAL, et les points bleus uniquement pour INT. Les points rouges sont ceux où aucune amélioration n'est observée.

La Figure V.12 montre les diagrammes de dispersion de la RMSD des produits SW, INT et CAL avec les données d'évapotranspiration FLUXnet, pour chaque station. On n'observe pas d'impact significatif de la fermeture au niveau des sites Fluxnet

pour l'évapotranspiration. INT et CAL améliorent les estimations de l'évapotranspiration pour seulement 53 % des sites. La distribution des différences (voir le diagramme encapsulé de la Figure V.12) est légèrement réduite par INT et CAL par rapport à la solution SW. L'emplacement des sites améliorés par INT et CAL est indiqué en vert, en jaune par CAL seulement, et en bleu par INT seulement. Les points rouges représentent la dégradation à la fois par INT et par CAL. Ces résultats doivent être pris avec précaution car les mesures des sites FLUXnet sont ponctuelles alors que l'estimation satellite est effectuée à la résolution 0.25° . En outre, les quelques stations peuvent ne pas être représentatives de l'ensemble du domaine spatial considéré dans cette analyse.

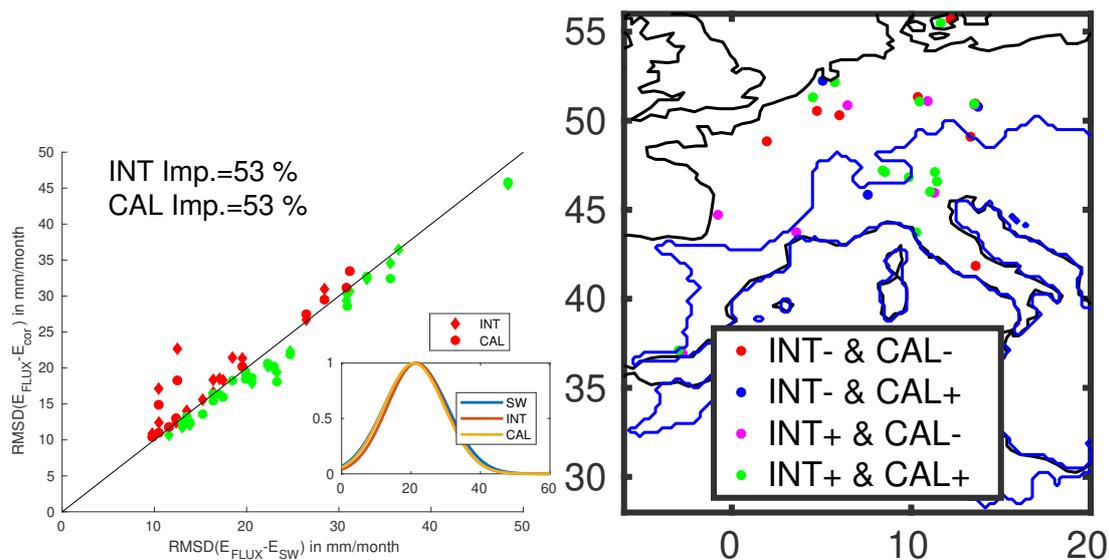


Figure V.12. Même Figure que V.11 pour l'évapotranspiration.

4.2 Les estimations annuelles

Les flux annuels du cycle de l'eau en Méditerranée et leur variabilité sur la période 2004-2009 sont représentés sur la Figure V.13. La variabilité est calculée comme l'écart-type des valeurs annuelles sur la période. Ces valeurs ont été calculées sur les sous-bassins continentaux ou océaniques :

- En considérant l'ensemble des sous-bassins européens (avec Turquie) ou Africains (sans tenir compte du bassin du Nil pour lequel seul le débit est représenté),
- En considérant indépendamment la mer Noire et la mer Méditerranée.

Les nombres en petite police sont pour SW tandis que la police large sont les estimations résultant de la méthodologie INT. Les deux valeurs pour l'estimation de l'écoulement net au détroit du Bosphore sont estimées comme le terme de déficit de l'équation du bilan hydrique ($Bos = E - P - R - Gib$), calculé indépendamment pour la Méditerranée et la mer Noire. En utilisant l'estimation INT (c.-à-d. après la fermeture des deux mers ensemble), les deux valeurs sont en meilleur accord qu'avant la fermeture. Dans ce qui suit, seules les valeurs INT sont décrites.

La Figure V.13 montre la contribution inégale en quantité d'eau douce déversée à la mer entre l'Europe ($314 \pm 57 \text{ km}^3 \text{ an}^{-1}$) et l'Afrique ($21 \pm 30 \text{ km}^3 \cdot \text{an}^{-1}$). En outre, il

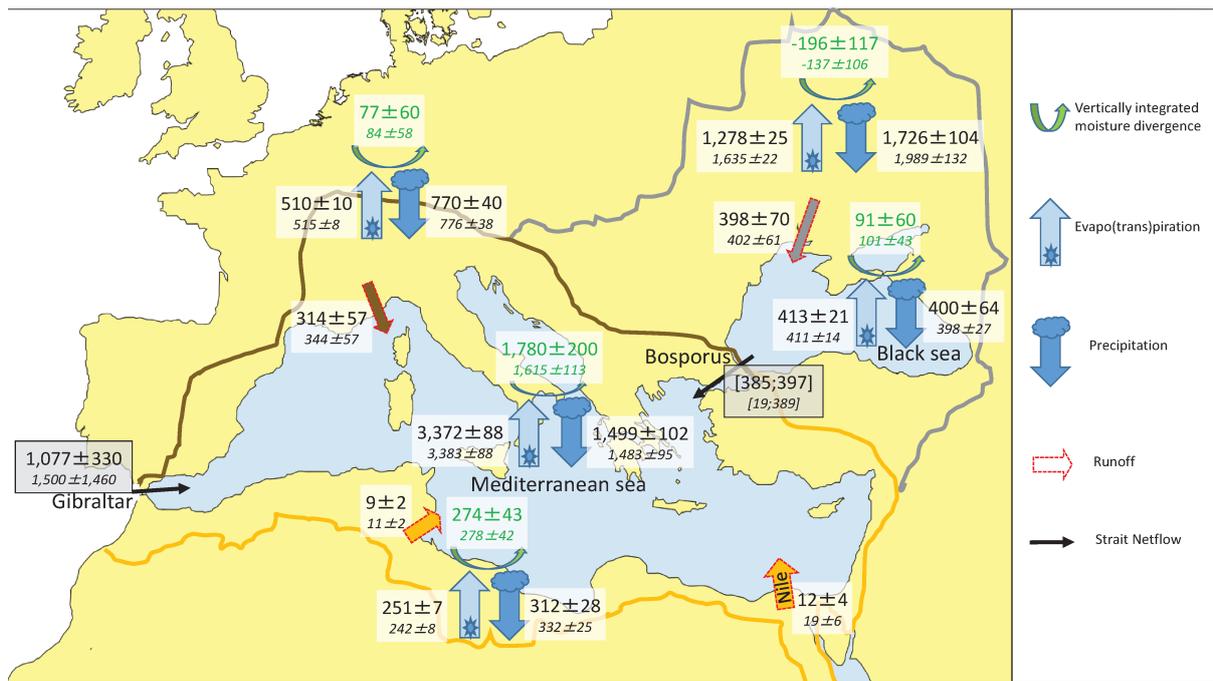


Figure V.13. Flux annuels moyens ($\text{km}^3 \cdot \text{an}^{-1}$) du cycle de l'eau en Méditerranée et incertitudes associées dans SW (petite police) et INT (grande police) au cours de la période 2004-2009.

montre le rôle de la mer Noire dans le cycle de l'eau global de la Méditerranée. La plupart des eaux douces européennes s'écoulent vers la mer noire ($398 \pm 70 \text{ km}^3 \cdot \text{an}^{-1}$; ce qui représente plus de 50 % du ruissellement européen total), où l'équilibre $E - P$ permet une contribution égale à la mer Méditerranée par l'apport du détroit du Bosphore. L'optimisation par la fermeture du cycle de l'eau diminue l'estimation du débit du Nil (de 19 ± 6 à $12 \pm 30 \text{ km}^3 \cdot \text{an}^{-1}$). Des discussions récentes sur le débit du Nil peuvent être trouvées dans (Jordà et al., 2017). Les nouvelles estimations du ruissellement comprennent, par la fermeture du cycle de l'eau continental, l'écoulement d'eau souterrain dans les aquifères méditerranéens.

Après l'optimisation de la fermeture, les précipitations annuelles, l'évapotranspiration et la divergence d'humidité sur le bassin européen contributeur de la mer Méditerranée sont estimées à : 770 ± 40 ; 510 ± 10 et $77 \pm 60 \text{ km}^3 \cdot \text{an}^{-1}$ respectivement quant à la région contributrice de la mer Noire, ces mêmes flux sont estimés à $1,726 \pm 104$; $1,278 \pm 25$ et $-196 \pm 117 \text{ km}^3 \cdot \text{an}^{-1}$. L'Europe accumule la plus grande partie de l'humidité provenant de la mer Méditerranée ($1780 \pm 200 \text{ km}^3 \cdot \text{an}^{-1}$) tandis que la mer Noire évacue une plus faible humidité vers la terre ($91 \pm 60 \text{ km}^3 \cdot \text{an}^{-1}$). Sur terre, la contribution de la partie africaine à la divergence globale de l'humidité est faible ($274 \pm 43 \text{ km}^3 \cdot \text{an}^{-1}$) principalement en raison de la présence des montagnes de l'Atlas qui sont une zone de convergence. Les deux estimations de l'écoulement net au détroit du Bosphore sont très proches, avec une différence inférieure à l'incertitude associée dans la Figure V.13. Les apports d'eau douce dans les deux détroits méditerranéens (Bosphore et Gibraltar) compensent la très importante perte d'eau par évaporation ($3,372 \pm 88 \text{ km}^3 \cdot \text{an}^{-1}$) en Méditerranée. Ce processus représente plus de deux fois les précipitations ($1,499 \pm 102 \text{ km}^3 \cdot \text{an}^{-1}$).

La Figure V.13 représente l'ensemble du cycle de l'eau dans la région méditerranéenne : le rôle de la mer Méditerranée comme réservoir d'humidité et d'énergie pour

les terres environnantes ; la faible contribution de la côte africaine en termes de ressources en eau, et le rôle de la mer Noire comme processus tampon pour l'apport d'eau douce entre l'Europe et la Méditerranée. Ce processus quasi-triangulaire met l'accent sur le lien hydrologique entre les terres environnantes et les deux mers.

References		E	P	E-P	R	Bos	Gib	Div
Sanchez-Gomez et al. (2011) 1957-2002	HIRHAM	1,377±55	425±57	952±80	116±30	116±30	720±100	-
	MEAN	1,254(±164)	442(±84)	812(±180)	124(±46)	87(±60)	540(±150)	-
	CRCM	1,208±72	606±80	602±107	73±40	110±50	420±130	-
Mariotti et al. (2002) 1979-1993	NCEP	1,113	433	680	100	75	494	659
	ERA-40	934	331	603	100	75	370	488
Jordà et al. (2017) 2005-2010	Prescribed values	-	-	900±200	200±10	100±20	850±400 600±200	-
Rodell et al. (2015) 2000-2010	orginal	1,391±157	576±76	815±157	-	0	0	866±131
	optimized	1,420±109	571±73	849±109	-	0	0	848±105
Current study 2004-2009	SW	1,300±34	573±36	726±57	144±21	2±615	575±561	620±44
	INT	1,295±33	577±40	719±60	155±15	129±60	428±124	677±77
	CAL	1,295±34	574±36	721±57	155±20	80±250	428±196	680±53

TABLE V.8 – Comparaison des flux moyens annuels et de leur variabilité associée (en $mm.an^{-1}$) pour la mer Méditerranée (sans la mer Noire) dans la littérature. La variabilité des produits est ici l'écart-type des valeurs annuelles, sauf pour l'incertitude associée à la moyenne des modèles climatiques régionaux qui est l'écart entre les modèles. La période d'analyse des différentes études est rappelée.

La Table V.8 résume la comparaison des diverses estimations des flux d'eau obtenues dans ce travail de thèse avec ce que l'on peut trouver dans la littérature. Les diverses estimations annuelles sont basées sur différentes périodes de temps et la comparaison doit être prise avec précaution car une certaine part de variabilité est probablement due au changement de régime hydrologique sur un plus long terme. Sanchez-Gomez et al. (2011) se sont concentrés sur le bilan thermique et hydrique de la mer Méditerranée à l'aide d'un ensemble de modèles climatiques régionaux (RCM) à haute résolution issue de ERA-40 : la base de données FP6-EU ENSEMBLE. Le budget atmosphérique n'a pas été pris en compte dans (Sanchez-Gomez et al., 2011) et aucune estimation de la divergence d'humidité n'a été fournie. A des fins de comparaison, nous avons décidé de choisir l'estimation moyenne de l'ensemble des RCMs et deux modèles particuliers : le modèle danois HIRHAM (Hesselbjerg Christensen et Meteorologisk Institut, 1996) et le modèle canadien CRCM (Plummer et al., 2006). Ces deux modèles ont été sélectionnés car leurs estimations P-E sont les extrêmes de l'ensemble des RCMs. Dans (Sanchez-Gomez et al., 2011), le flux net à Gibraltar a été estimé comme le terme de déficit du bilan hydrique : $Gib = E - P - R - Bos$.

Mariotti et al. (2002) ont analysé le cycle de l'eau sur la région méditerranéenne dans le contexte de la téléconnexion NAO ("North Atlantic Oscillation") sur la période 1979-1993 en utilisant deux ré-analyses : ERA-40 et NCEP-NCAR pour les précipitations, l'évaporation et la divergence de l'humidité. Ils ont utilisé les données de débit des rivières de MED-HYCOS et de GRDC en ne considérant que les rivières observées pour le ruissellement total (conduisant à l'estimation de $100 mm.an^{-1}$ pour l'apport en eau dans la méditerranée) et la valeur de $75 mm.an^{-1}$ pour le débit au détroit du Bosphore (Lacombe et Tchernia, 1972). Ils ont estimé le débit net à Gibraltar comme le déficit en eau de la Méditerranée en utilisant l'équation $Gib = Div - R - Bos$ provenant du budget océanique et atmosphérique. Le changement de contenu en eau dans la mer Méditerranée à l'échelle du mois est considéré nul. Ils ont utilisé d'anciennes ver-

sions des ré-analyses et plusieurs commentaires ont déjà été émis sur les estimations des précipitations et de l'évapotranspiration dans ces versions. Néanmoins, d'après mes connaissances, ils ont fourni la dernière tentative d'estimation du cycle de l'eau de la Méditerranée en tenant compte de la composante atmosphérique.

Jordà et al. (2017) proposent un état de l'art de la quantification des diverses estimations des composantes du cycle de l'eau. Les valeurs présentées dans la Table V.8 pour les différentes composantes sont les valeurs consensuelles au sein de la communauté scientifique. Elles sont basées sur plusieurs travaux et prennent en compte le résultat de Mariotti et al. (2002) et Sanchez-Gomez et al. (2011) par exemple. En particulier, la moyenne climatologique de l'estimation de l'écoulement net dans le détroit de Gibraltar à partir de (Jordà et al., 2016) a été ré-évaluée et corrigée dans (Jordà et al., 2017).

La Table V.8 montre aussi les résultats de Rodell et al. (2015) avant et après leur optimisation. Les fermetures du bilan hydrique sur la mer Méditerranée et la mer Noire ont été faites indépendamment (sans échange au détroit du Bosphore). Le cycle de l'eau dans la mer Méditerranée a été corrigé sans prendre en compte l'écoulement net à Gibraltar. Rodell et al. (2015) n'a fourni aucun ruissellement explicite pour la mer Méditerranée, mais seulement pour le continent eurasiatique tout entier.

Pour les quatre articles mentionnés, seule la Méditerranée sans la mer Noire est prise en compte. Aucune estimation de la méthodologie SW+PF n'est donc fournie dans la Table V.8. L'étude présentée dans ce travail de thèse est la seule à utiliser des observations directes pour le flux net à Gibraltar et calculer le débit au détroit du Bosphore comme le terme de déficit de l'équation du bilan hydrique ($Bos = E - P - R - Gib$).

Evaporation - La moyenne de l'ensemble RCM pour l'évaporation annuelle est de 1254 mm.an^{-1} avec un écart inter-modèle de 164 mm.an^{-1} . Certains RCM ont évalué l'évaporation annuelle plus élevée comme HIRHAM qui l'a estimée à $1377 \pm 55 \text{ mm.an}^{-1}$. Au contraire, Mariotti et al. (2002) ont trouvé une évaporation relativement faible avec les réanalyses (1113 et 934 mm.an^{-1} avec respectivement NCEP et ERA). Rodell et al. (2015) ont estimé une évaporation moyenne beaucoup plus élevée et une variabilité annuelle plus élevée avec une valeur annuelle moyenne de $1391 \pm 157 \text{ mm.an}^{-1}$ en utilisant seulement OAFflux et $1420 \pm 109 \text{ mm.an}^{-1}$ après optimisation. Notre solution SW non contrainte donne une valeur annuelle de $1300 \pm 34 \text{ mm.an}^{-1}$ et notre produit INT contraint donne $1295 \pm 33 \text{ mm.an}^{-1}$. L'estimation de CAL est proche de INT.

Précipitation - La moyenne de l'ensemble RCM pour les précipitations annuelles était de $442 \pm 84 \text{ mm.an}^{-1}$ ce qui est assez proche de la valeur du NCEP dans (Mariotti et al., 2002) alors que l'estimation par satellite dans (Rodell et al., 2015) et celle présentée dans cette thèse indiquent des précipitations plus élevées. La valeur présentée dans (Rodell et al., 2015) passe de 576 à 571 mm.an^{-1} après optimisation alors qu'avec la contrainte de fermeture via la méthode INT, la précipitation change de 573 à 577 mm.an^{-1} . Les produits SW, INT et CAL présentent des estimations de précipitations proches à l'échelle annuelle en raison de l'incertitude relativement faible des précipitations dans le schéma d'intégration. Certains RCMs ont calculé des précipitations encore plus importantes que celles qui ont été restituées par satellite ($606 \pm 80 \text{ mm.an}^{-1}$). Sanchez-Gomez et al. (2011) avaient déjà noté que les ensembles de données satellitaires calibrés comme le GPCP et le TMPA ont tendance à donner des

valeurs de précipitations plus élevées que ce qui a été simulé dans les RCMs. Les précipitations sur la mer sont une variable sensible et leur validation est difficile en raison de l'absence de bouées pour la mesure *in situ*. La valeur des ré-analyses de ERA-40 dans (Mariotti et al., 2002) est plus faible que celle du NCEP.

Evaporation moins Précipitation - Sanchez-Gomez et al. (2011) se sont concentrés sur le budget $E - P$ pour évaluer la cohérence physique des RCMs. Ils ont considéré qu'un modèle ayant une forte évaporation tend à avoir des précipitations élevées. Le budget $E - P$ moyen des RCM est de $812 \pm 180 \text{ mm.an}^{-1}$ et la fourchette des valeurs se situe entre 952 ± 80 (modèle HIRHAM) et $602 \pm 107 \text{ mm.an}^{-1}$ (modèle CRCM). L'écart entre les modèles reste élevé pour le budget $E - P$, soulignant les difficultés à fournir une évaluation réaliste du bilan hydrique. Rodell et al. (2015) ont trouvé un budget $E - P$ similaire à la moyenne de l'ensemble des RCMs, mais la variabilité associée est également élevée en raison de l'incertitude sur l'évaporation. Nos estimations $E - P$ sont respectivement de 726 ± 57 et $719 \pm 60 \text{ mm.an}^{-1}$ avant et après la contrainte de fermeture. Ces valeurs sont plus basses mais se situent toujours dans la plage de l'ensemble RCM. Ils sont plus proches de celles trouvées avec les ré-analyses du NCEP (Mariotti et al., 2002). Jordà et al. (2017) considèrent que le bilan $E - P$ est de $900 \pm 200 \text{ mm.an}^{-1}$ ce qui est en accord avec l'estimation du modèle CRCM. Rodell et al. (2015) ont trouvé un budget $E - P$ similaire mais avec une estimation d'évaporation beaucoup plus élevée qui semblait irréaliste. De plus, on peut remarquer que leur contrainte de fermeture tend à augmenter la valeur d'évaporation et donc le budget E-P.

Ruissellement d'eau douce - Seuls les RCMs fournissant le ruissellement ont été utilisés pour calculer la valeur annuelle de R ($124 \pm 46 \text{ mm.an}^{-1}$) dans (Sanchez-Gomez et al., 2011). Mariotti et al. (2002) ont trouvé des valeurs comparables pour le ruissellement en ne considérant que les rivières surveillées. Rodell et al. (2015) ne considèrent pas le ruissellement dans la mer Méditerranée puisque la fermeture a été effectuée à l'échelle mondiale (continent eurasién) et aucune valeur n'a été fournie pour l'apport d'eau douce à la Méditerranée. Notre estimation de ruissellement est augmentée de 144 ± 21 pour SW à $155 \pm 15 \text{ mm.an}^{-1}$ pour INT après optimisation. Cette augmentation est principalement due à la réévaluation du débit du Nil qui présente un débit plus important ($76 \text{ km}^3.\text{an}^{-1}$) après la fermeture. Toutes ces estimations de ruissellement sont inférieures à la valeur prescrite dans (Jordà et al., 2017) ($200 \pm 10 \text{ mm.an}^{-1}$).

Écoulement au détroit du Bosphore - La valeur moyenne de l'ensemble des RCMs pour l'apport d'eau douce au détroit du Bosphore est de $87 \pm 60 \text{ mm.an}^{-1}$, ce qui souligne les écarts importants entre les RCMs pour cette variable. Rodell et al. (2015) ont fermé indépendamment la Méditerranée et la mer Noire sans échange entre les deux bassins océaniques (c'est-à-dire que l'écoulement net est égal à zéro). L'approche développée dans cette thèse, donne l'écoulement comme le terme de déficit de l'équation du bilan hydrique ($Bos = E - P - R - Gib$). Le produit SW présente une valeur irréaliste pour l'écoulement au détroit du Bosphore ($2,0 \pm 615 \text{ mm.an}^{-1}$), ce qui est principalement dû à la forte incertitude associée à l'écoulement net à Gibraltar. La contrainte de fermeture améliore l'estimation de l'écoulement au détroit de Bosphore qui est égal à $129 \pm 60 \text{ mm.an}^{-1}$ pour la solution INT, après optimisation. La valeur est proche du déficit du bilan hydrique de la mer Noire (calculé après optimisation) : $132 \pm 60 \text{ mm.an}^{-1}$ (non présentée dans la Table V.8) soulignant la cohérence entre les bilans hydriques

des deux mers. La valeur est toujours supérieure à l'estimation de 75 mm.an^{-1} de Mariotti et al. (2002).

Écoulement au détroit de Gibraltar- Rodell et al. (2015) n'ont considéré aucun échange à Gibraltar lors de la fermeture du cycle de l'eau méditerranéen et n'ont fourni aucune estimation pour cette variable. Sanchez-Gomez et al. (2011) et Mariotti et al. (2002) ont évalué le flux net en fermant le cycle de l'eau sur la région méditerranéenne mais ils ont utilisé des hypothèses et des équations différentes. L'estimation de Sanchez-Gomez et al. (2011) est basée sur la fermeture seulement océanique alors qu'elle est basée sur la fermeture océanique et atmosphérique dans (Mariotti et al., 2002). La moyenne de l'ensemble des RCMs est de $540 \pm 150 \text{ mm.an}^{-1}$ dans (Sanchez-Gomez et al., 2011), tandis que Mariotti et al. (2002) ont trouvé une valeur inférieure avec les ré-analyses (493 et 370 mm.an^{-1} avec NCEP et ERA). Jordà et al. (2017) donnent deux valeurs pour le débit net à Gibraltar, l'une d'observation directe mais qui souffre de grandes incertitudes ($850 \pm 400 \text{ mm.an}^{-1}$) et l'autre comme le déficit du budget de l'eau ($600 \pm 200 \text{ mm.an}^{-1}$). Dans ce manuscrit, la valeur utilisée dans l'estimation de SW non contrainte est celle de Jordà et al. (2016) avec une correction de biais⁷ basée sur la valeur moyenne prescrite dans (Jordà et al., 2017). L'estimation de l'écoulement net après optimisation est de $428 \pm 124 \text{ mm.an}^{-1}$ ce qui montre une diminution de l'estimation par rapport à SW (et donc par rapport à (Jordà et al., 2017)) à ce que l'on peut trouver dans (Jordà et al., 2017) mais dans la fourchette du déficit du bilan hydrique des RCMs.

Divergence d'humidité - Aucune divergence d'humidité n'a été fournie par les RCMs dans (Sanchez-Gomez et al., 2011). Mariotti et al. (2002) ont trouvé une divergence d'humidité de 659 mm.an^{-1} avec NCEP et 488 mm.an^{-1} avec ERA, alors que Rodell et al. (2015) ont estimé la divergence à $848 \pm 105 \text{ mm.an}^{-1}$ après optimisation. La différence entre les estimations de Rodell et al. (2015), Mariotti et al. (2002) et ce que l'on trouve dans la présente étude est principalement due à l'écart entre les trois réanalyses : Modern-Era Retrospective Analysis for Research and Applications (MERRA) utilisé par Rodell et al. (2015), NCEP et ERA-40 dans (Mariotti et al., 2002) et ERA-I utilisé dans la présente étude. Des travaux récents sur les comparaisons de réanalyses atmosphériques ont montré les avantages de l'ERA-I. Stopa et Cheung (2014) ont mis l'accent sur les performances ERA-I dans la représentation de la variabilité à long terme du vent, critique pour la représentation de la divergence d'humidité. Brown et Kummerow (2013) ont fait remarquer que le budget $E - P$ (SeaFlux- GPCP) dérivé du satellite est bien corrélé avec la divergence d'humidité atmosphérique ERA-Interim.

4.3 Étude des tendances des variables hydrologiques

4.3.a Tendances au niveau du mois et du sous-bassin

Pour chaque sous-bassin et pour chaque mois, la tendance annuelle sur la période 1980-2012 est calculée à l'aide d'une régression linéaire sur la série temporelle des composantes de flux du cycle de l'eau. Les résultats sont présentés dans la Figure V.14. Les deux composantes majeures de flux du cycle de l'eau estimées via CAL

7. Le lecteur peut se reporter à la Figure III.9.

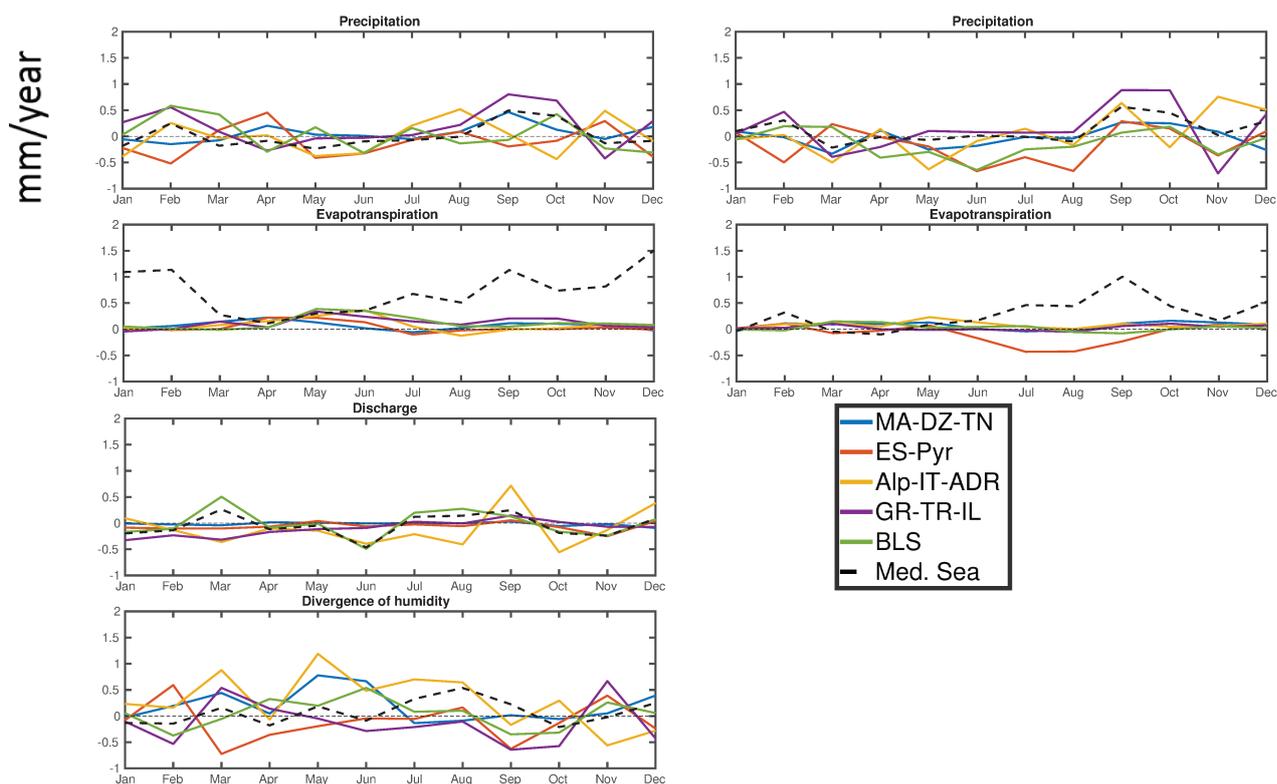


Figure V.14. *Tendance linéaire mensuelle des composantes du flux du cycle de l'eau estimée via CAL. De haut en bas : précipitations, évapo(transpi)ration, ruissellement et divergence de l'humidité. Les lignes zéro sont tracées pour faciliter la lecture. Les tendances des précipitations et d'évapo(transpi)ration de l'ERA-I sont également représentées à des fins de comparaison.*

sont comparées à ERA-I.

Précipitation - Les tendances des précipitations mensuelles dans ERA-I et CAL représentées dans la première ligne de la Figure V.14 sont en accord, quel que soit le sous-bassin considéré de septembre à avril mais différent en été. Ils sont particulièrement proches pour certains sous-bassins tels que Alp-IT-ADR ou GR-TR-IL. Les tendances pour les précipitations océaniques sur la mer Méditerranée sont très proches dans les deux produits. Les produits ERA-I et CAL montrent une précipitation qui augmente globalement (depuis 1980) en septembre et octobre plus que tout autre mois, sur tous les sous-bassins et pour la mer Méditerranée. Toutefois, les tendances pour la fin du printemps et l'été semblent différer. De plus, une réduction de $1 \text{ mm.mois}^{-1}.an^{-1}$ des précipitations peut être observée en mai dans les sous-bassins des Alpes et des Balkans au cours des 30 dernières années. Cela peut être dû à un printemps et un automne plus chaud. ERA-I montre une diminution forte des précipitations estivales pour le sous-bassin ES-Pyr.

Evapo(transpi)ration - La tendance d'évapo(transpi)ration est montrée dans la deuxième ligne de la Figure V.14. Les deux produits CAL et ERA-I concordent avec une intensification de l'évapotranspiration en mai et septembre et une réduction pendant l'été, sur tous les sous-bassins. Ces tendances sont plus marquées dans le produit satellite CAL. Les tendances pour l'évaporation sur la mer Méditerranée diffèrent seulement en

hiver entre les deux produits avec une augmentation de l'évaporation hivernale plus marquée dans le produit CAL. La tendance de cette variable est positive pour tous les mois de l'année, ce qui souligne l'intensification de la perte d'eau douce dans la mer Méditerranée par évaporation.

Ruissellement - Les tendances mensuelles du ruissellement dans la mer Méditerranée, représentées dans la troisième ligne de la Figure V.14, ne sont significatives que pour les sous-bassins Alp-IT-ADR et BLS qui sont les principaux contributeurs d'eau douce à la mer Méditerranée. L'augmentation du ruissellement total dans la Mer en septembre est due à l'augmentation des précipitations au cours du même mois dans le sous-bassin Alp-IT-ADR (déjà mentionné). En revanche la tendance positive en mars et négative en juin pour la Mer Méditerranée est liée à l'augmentation ou la diminution de l'apport d'eau douce dans le sous-bassin BLS pour ces mêmes mois.

Divergence d'humidité - La dernière ligne de la Figure V.14 montre l'intensification de l'accumulation et de l'évacuation du flux d'humidité sur les différents sous-bassins. La tendance positive dans le sous-bassin Alp-It-ADR indique une augmentation de l'évacuation sur les Alpes et les Balkans qui sont généralement des zones d'accumulation. La Mer Méditerranée montre une intensification de la divergence de l'humidité au cours de l'été.

4.3.b Tendence au niveau du pixel

Les tendances spatiales annuelles des composantes de flux du cycle de l'eau peuvent être analysées à l'échelle du pixel en utilisant une régression linéaire de la série temporelle. La signifiante est donnée par un test de Kandell-Mann avec plus de 90 % de confiance. Le ruissellement n'a pas d'étendue spatiale et l'analyse de ses tendances est donc donnée dans la quatrième ligne de la Figure V.14.

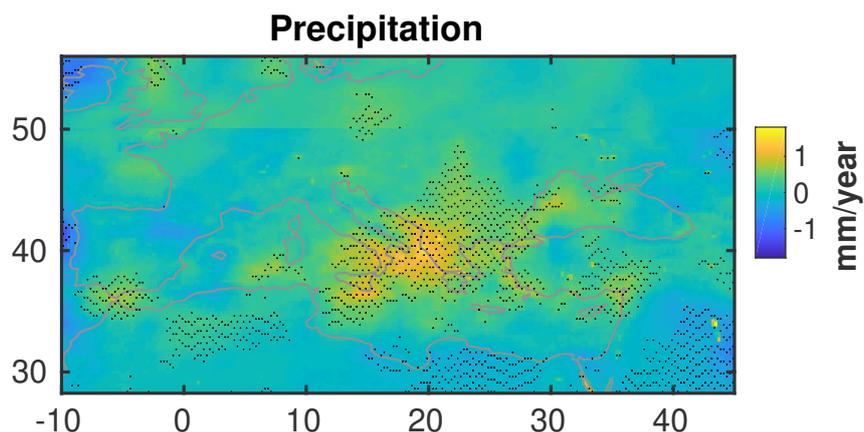


Figure V.15. Tendence annuelle des précipitations dans CAL sur la période 1980-2012. La tendance avec une signifiante statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.

Précipitation - La structure spatiale des tendances de précipitation est montrée sur la Figure V.15. Celle-ci est proche de ce que Mariotti (2010) a analysé avec seulement le produit GPCP. Le test de Mann-Kendall montre des tendances positives sur la mer Adriatique, la Grèce et l'Italie du sud qui pourraient être liées aux caractéristiques

orographiques de la région. La côte égyptienne présente une diminution de la précipitation.

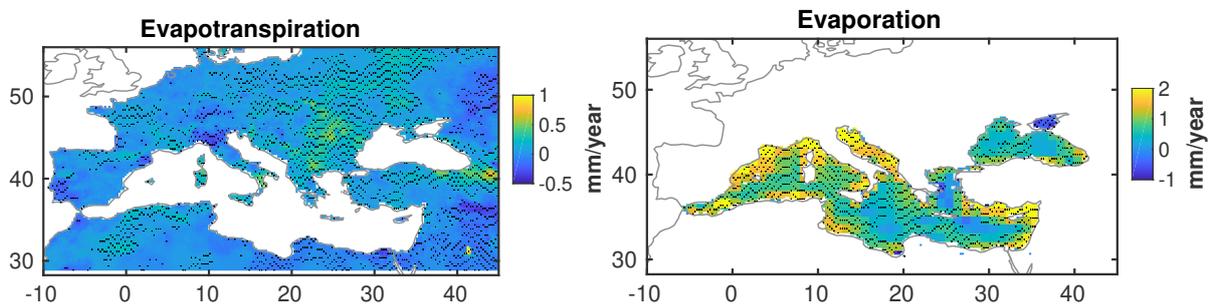


Figure V.16. Tendances annuelles de l'évaporation (à droite) et de l'évapotranspiration (à gauche) dans CAL sur la période 1980-2012. La tendance avec une signification statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.

Evapo(transpi)ration - La Figure V.16 montre la tendance annuelle de l'évapotranspiration et de l'évaporation. Le produit satellite CAL souffre d'une contamination terre/mer près de la côte (c'est-à-dire que certains pixels contiennent de la terre et de la mer), ce qui conduit à une surestimation de la tendance sur l'évapo(transpi)ration près de la côte. L'évaporation de la mer Méditerranée augmente en moyenne, en particulier dans le bassin de l'ouest.

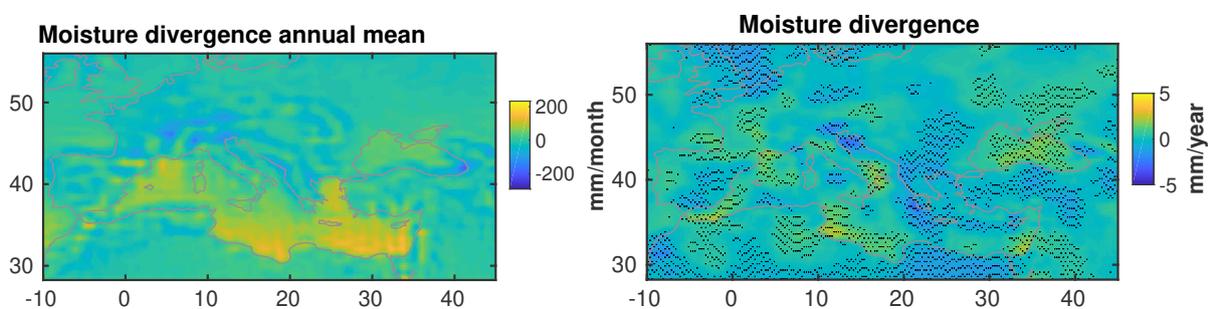


Figure V.17. Tendances annuelles de la divergence de l'humidité dans CAL sur la période 1980-2012 à droite. Comme la signification de la tendance varie selon la valeur moyenne (positive ou négative) de la divergence, cette dernière est représentée à gauche. La tendance avec une signification statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.

Divergence d'humidité - Les tendances de la divergence de l'humidité (intégrée verticalement) sont représentées dans la Figure V.17. Les tendances de la divergence d'humidité sont plus fortes principalement au-dessus des montagnes (Alpes et Caucase) pour toutes les saisons. La tendance montre une augmentation de l'évacuation de l'humidité à l'ouest de la mer Méditerranée et au nord-est de l'Europe. Globalement, l'accumulation d'humidité dans la région méditerranéenne diminue puisque la tendance de la divergence est positive. A l'échelle annuelle, le test de Mann-Kendall met l'accent sur l'accumulation accrue d'eau précipitable sur la Méditerranée orien-

taie (région d'évacuation). Les Alpes présentent aussi une plus forte augmentation de l'accumulation d'humidité.

4.4 Estimation des incertitudes des produits satellites

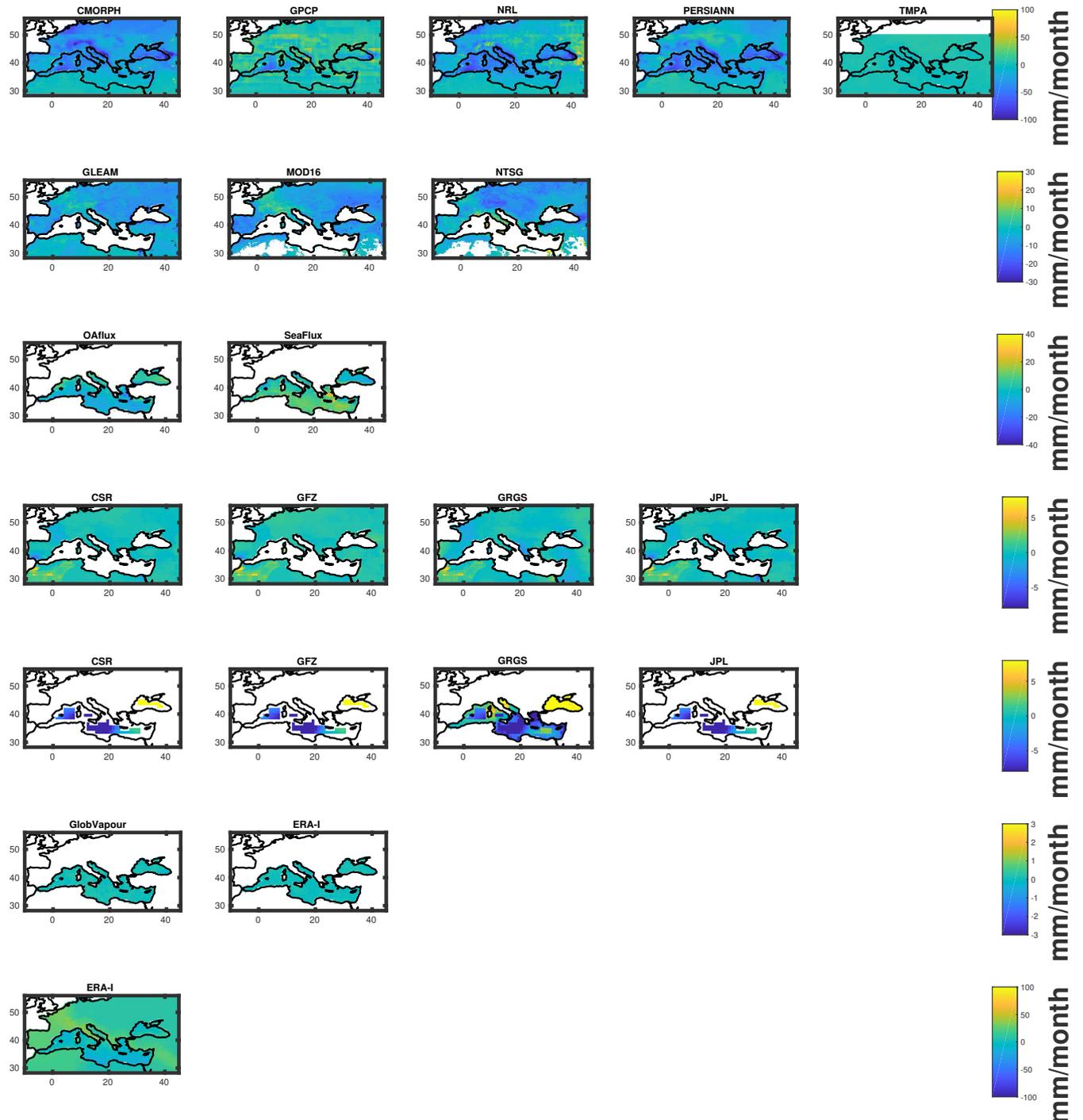


Figure V.18. Estimation de biais des produits satellites décrivant les différentes composantes du cycle de l'eau. La moyenne du produits INT est pris comme référence pour chacune des variables.

Les biais et incertitudes de chaque ensemble de données sont calculés par rapport

au produit intégré INT. Les résultats sont présentés dans les Figures V.18 et V.19. En moyenne, pour les produits de précipitation, GPCP et TMPA ont des biais positifs tandis que NRL PERSIANN et le CMORPH ont des biais négatifs. Pour l'évapotranspiration, des biais négatifs sont obtenus pour les trois datasets MOD16, GLEAM et le NTSG sur l'ensemble de l'Europe, seuls GLEAM et MOD16 montrent un biais positif au-dessus des montagnes et au nord de la France. NSTG montre un biais positif au-dessus des Alpes et un biais négatif au sud de l'Italie. Concernant l'évaporation sur mer, on observe un biais négatif pour OAFflux et un biais positif pour SeaFLUX au-dessus de la Méditerranée (présentant un gradient nord/sud) et inversement au-dessus de la mer Noire. Les produits de changement de contenu en eau sur terre montrent des biais beaucoup plus faibles. Les structures spatiales des biais sont très proches entre les produits issus de GRACE Ground System (JPL, CSR et GFZ), seul GRGS diffère. Les changements de contenu en eau sur mer ont des biais plus importants mais on retrouve la remarque générale, faite précédemment, sur la similitude des biais des produits de GRACE Ground System. Au-dessus de la mer, GRGS montre un biais en opposition aux trois autres produits. La figure ne montre aucun biais significatif pour les deux produits GobVapour et ERA-I. Le biais du produit ERA-I de la divergence, seul dataset utilisé pour représenter cette variable, est finalement la différence entre les produits SW et INT pour cette variable.

Les structures des biais et des incertitudes varient d'un produit à l'autre mais en général, les incertitudes sont élevées là où les biais sont forts. Dans l'ensemble, les biais et les incertitudes du GPCP et TMPA sont plus faibles sur terre que sur mer puisqu'ils sont corrigés à l'aide de jauges que l'on trouve principalement sur terre. Le produit GRGS montre une plus grande variabilité sur mer que les autres produits. Le produit de la divergence ne montre une variabilité que sur mer ce qui est dû au type de correction apporté dans INT pour cette variable particulière où seul le biais a été corrigé pour la partie continentale (voir Section 2.3.1 de ce chapitre).

5 Conclusion du chapitre

Ce chapitre rend compte du potentiel de certaines méthodes d'intégration à rendre cohérents plusieurs produits dérivés satellites, et ce dans une approche multi-variée. L'intégration sous contrainte réduit, *a posteriori*, les erreurs des différents produits dérivés satellites en utilisant un principe de synergie : l'observation du bilan hydrique. Les composantes du cycle de l'eau étant dépendantes les unes aux autres, l'observation de la précipitation, par exemple, apporte une information sur les autres composantes du cycle. L'utilisation de la contrainte de fermeture du cycle de l'eau dans l'estimateur linéaire de Bayes permet d'obtenir *a posteriori* des estimations cohérentes des composantes du cycle. Des outils de dé-biaisage, de fusion de données, et d'intégration sont ici développés hors de tout schéma d'assimilation. Ces méthodes permettent alors d'obtenir des estimations aux plus près des observations, sans information liée à un modèle et avec une fiabilité accrue.

Au cours de ce chapitre, diverses méthodologies d'intégration ont été présentées et une attention particulière a été accordée à la méthode INTegration. Cette approche assure pleinement les objectifs énoncés précédemment : il s'agit d'un produit spatialisé, mais dans lequel la fermeture du bilan hydrique est prise en compte. Le produit INT permet de réduire la RMS du résidu du bilan hydrique à $3,55 \text{ mm.mois}^{-1}$ sur terre

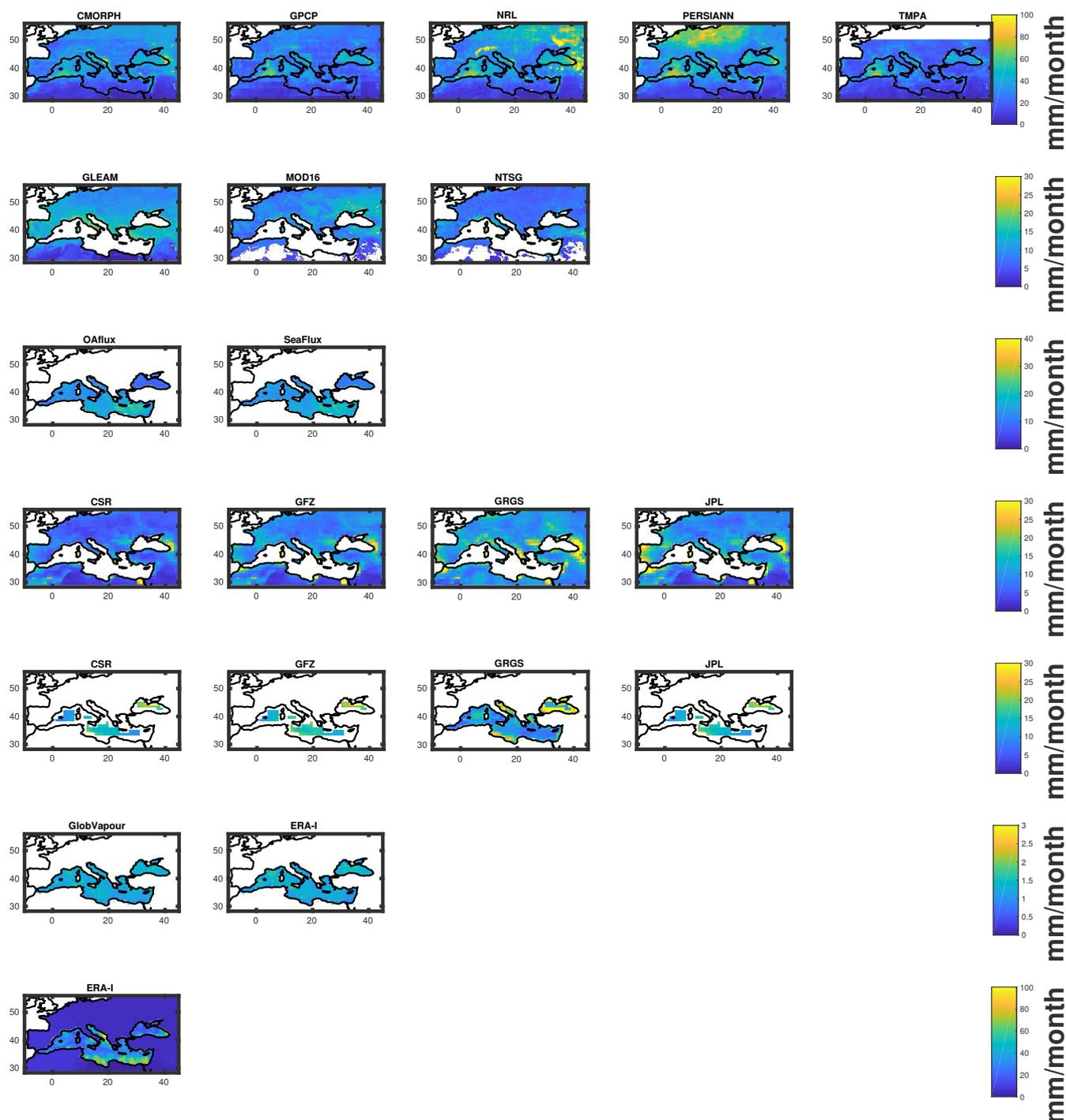


Figure V.19. Même que Figure que V.18 mais pour les erreurs aléatoires (incertitudes) des produits satellites.

et $5,27 \text{ mm.mois}^{-1}$ dans l'atmosphère. Ces réductions représentent une amélioration de respectivement 78% et 80% par rapport à la meilleure combinaison de données satellites sans contrainte. La couverture temporelle de INT est limitée à la période de couverture commune (2004-2009) aux différents produits satellites (voir Table III.1).

Le produit INT a été évalué à différentes échelles. Même si l'évaluation est une tâche difficile et que le travail présenté n'est pas exhaustif, nos résultats montrent que la prise en compte de la fermeture du cycle de l'eau dans l'estimation des différentes

composantes du cycle permet de réduire les différences avec les mesures *in situ*. À l'échelle du sous-bassin, les précipitations globales sont plus proches de l'ensemble de données *in situ* EOBS après avoir été contraintes. L'estimation du niveau de la mer est également améliorée par rapport à l'estimation par altimétrie. A l'échelle du pixel, l'estimation INT montre une meilleure concordance avec les mesures des stations *in situ* de la base de données FLUXnet2015.

Le cycle de l'eau a été analysé en termes de moyennes climatologiques sur la période 2004-2009 et comparé à la littérature existante. La méthodologie INT a amélioré les estimations des composantes de l'eau en Méditerranée. Le produit INT fournit des valeurs plus réalistes pour les flux nets d'écoulement au Bosphore et à Gibraltar. Il convient de noter que l'estimation du Bosphore est principalement déterminée par l'estimation de Gibraltar et pourrait donc être améliorée par une évaluation plus précise de l'écoulement net à Gibraltar.

Le cycle de l'eau méditerranéen a déjà été bien étudié par Mariotti et al. (2002) et Sanchez-Gomez et al. (2011) en s'appuyant sur des modèles et des ré-analyses. A l'échelle mondiale, Rodell et al. (2015) ferment indépendamment la Méditerranée et la mer Noire en utilisant l'observation par satellites tandis que Sanchez-Gomez et al. (2011) ferment le cycle de l'eau de la mer Méditerranée en estimant le flux net de Gibraltar comme étant le déficit budgétaire du bilan hydrique.

Cette étude a permis une description complète du cycle de l'eau, basée sur moins d'hypothèses. C'est aussi le premier effort pour fermer le cycle de l'eau sur toutes ses composantes : à la surface et dans l'atmosphère sur l'ensemble de la région méditerranéenne, en utilisant des observations par satellites et des mesures *in situ*. Mais l'intérêt de notre approche est de proposer une méthodologie d'intégration des produits satellites qui prend en compte la fermeture du cycle de l'eau. Alors que Sanchez-Gomez et al. (2011) et Mariotti et al. (2002) n'ont apporté aucune correction sur les produits satellites, notre approche, similaire à Rodell et al. (2015), tend à contraindre les produits satellites dans le contexte de l'estimateur Bayésien en considérant l'équilibre du bilan hydrique comme une observation "virtuelle" (Rodgers, 2000). À la différence de Pan et Wood (2006), on propose une intégration, type fusion de données plus contrainte, sans l'aide d'un modèle d'hydrologie de surface. Les outils ainsi développés permettent de rester au plus près des observations satellites et les produits intégrés pourraient être utilisés dans la calibration de modèle sans problème de dépendance.

Il y a encore des incertitudes concernant l'observation du cycle de l'eau et de ses composantes, mais la méthodologie INT est le meilleur moyen d'inclure de la cohérence entre les estimations des différentes composantes indépendantes. On a pu dès lors utiliser INT comme référence à l'estimation des biais et incertitudes des différents produits satellites originels. Ce produit à composantes multiples est prometteur pour forcer, calibrer ou contraindre des modèles régionaux avec conservation de l'eau.

Outre la capacité d'introduire la fermeture du cycle de l'eau avec l'approche INT, ce chapitre a également introduit le produit CAL qui est un étalonnage des produits satellites pour extrapoler dans le temps la contrainte de fermeture. Le produit CAL est moins efficace pour fermer le cycle de l'eau mais présente l'avantage d'avoir une couverture temporelle plus longue. Elle permet en outre d'étudier les tendances sur 1980-2012.

Chapitre VI

Conclusions et perspectives

1 Retour sur les objectifs de cette thèse

L'observation satellite permet une couverture globale et un suivi temporel cruciaux pour la météorologie et la surveillance du climat. Les satellites offrent l'opportunité unique de compléter les réseaux d'observations *in situ* locaux et permettent une démocratisation de l'accès à l'information concernant la météorologie et le climat. Si les produits satellites sont maintenant omniprésents, leurs incertitudes demeurent grandes. Fruit d'une longue chaîne de traitement via divers algorithmes et hypothèses, la variabilité inter-produits pour une région comme la Méditerranée, peut atteindre 100% de la valeur de l'estimation surtout pour des zones géographiques complexes comme les massifs montagneux. Beaucoup de scientifiques préfèrent l'utilisation de la ré-analyse avec son modèle dynamique à l'utilisation directe de produits satellites. Ce travail de thèse a pour objectif de contribuer à une meilleure exploitation des observations par satellite, dans la perspective de l'étude du cycle de l'eau en particulier.

Cette thèse s'est concentrée sur deux types d'optimisation, l'une intervenant au niveau de l'observation satellite, l'autre au niveau de l'intégration de produits dérivés. Nous avons, d'une part, proposé une manipulation plus pertinente des spectres hyperspectraux de températures de brillance mesurées en infrarouge ou micro-ondes dans un objectif d'une réduction de dimension pour les algorithmes de restitution (statistique ou physique). Ces méthodes de réduction de dimension sont incontournables pour le pré-traitement (comme pour le dé-bruitage) des observations satellites hyperspectrales. Obtenir une représentation pertinente de l'ensemble de l'observation mais de façon compressée est la première étape dans d'optimisation de l'exploitation des observations satellites hyperspectrales.

D'autre part, une approche de fusion de données puis une méthode d'intégration des différents produits satellites ont été développées pour une estimation plus fiable des variables géophysiques qu'ils représentent. Dans le cas particulier de l'étude du cycle de l'eau, l'équation du bilan hydrique peut être vue comme une contrainte permettant d'intégrer les estimations des différentes composantes du cycle de l'eau afin de les rendre plus cohérentes entre elles. Cette intégration permet la correction d'une partie des erreurs des produits satellites.

2 La réduction de dimension

Le Chapitre II a présenté un état de l'art des méthodes classiques de réduction de dimensions utilisées en opérationnel. Les techniques de compression que sont l'Analyse par Composantes Principales (ACP) et ses variantes permettent une réduction optimale au sens de la variabilité statistique. Ces techniques conservent l'information contenue dans de grandes bases de données (comme les spectres de IASI) par un nombre réduit de composantes. Ces méthodes permettent alors la restitution opérationnelle de variables atmosphériques et de surface en limitant le nombre d'entrées des algorithmes d'inversion. Malheureusement, l'usage de composantes de l'ACP souffre de l'absence de sens physique de ces composantes et des problèmes de mélange d'information qu'elles révèlent. Dans l'objectif de restituer une variable particulière, on cherche à sélectionner les canaux les plus pertinents. Les sélections de canaux sont des méthodes de réduction de données "orientées" vers la/les variable(s) géophysique(s) à restituer. Si cette approche permet la conservation du sens physique de la mesure, elle est loin d'être optimale en terme de redondance d'information pour, par exemple, le dé-bruitage. On a présenté une nouvelle approche de réduction de dimension des spectres à la croisée des deux méthodes classiques : les "Bottleneck Channels". Dans cette approche, une compression de l'information est obtenue de façon analogue à l'ACP mais avec une contrainte imposée aux composantes : celle d'être au plus proche des canaux réels.

Afin d'évaluer le potentiel de cette nouvelle approche, deux types d'évaluation ont été développées. L'une en terme de compression pure basée uniquement sur une réduction de dimension des spectres, l'autre sur la capacité à restituer de façon fiable les variables géophysiques comme les profils de température et d'humidité atmosphérique ainsi que le contenu en ozone par inversion des spectres via un réseau de neurones. Les BCs sont un bon compromis entre la compression et la sélection de canaux. Ils présentent la capacité de compresser l'information efficacement tout en possédant une représentation identique aux canaux réels. Ils permettent une meilleure gestion des problèmes de mélange et de contamination et représentent un nombre limité de pseudo-canaux faciles à assimiler.

Ainsi est-il possible d'intervenir sur l'observation satellite et d'optimiser la gestion du contenu en information des mesures physiques de l'instrument satellite (ici des sondes), soit pour le stockage et la transmission des données, soit pour l'inversion du transfert radiatif et la restitution de variables géophysiques. Cette optimisation se traduit par une meilleure restitution des variables géophysiques dans le cas d'une inversion par réseau de neurones.

Perspectives

Du fait de sa généralité, l'approche BC est un bon candidat pour les développements futurs de méthodes de compression d'information et les BCs pourront être validés en opérationnel pour la restitution ou l'assimilation réelle des variables géophysiques. L'approche BC pourra être améliorée dans le contexte des situations nuageuses, en examinant comment la contamination des nuages peut être traitée dans les BCs. Dans le contexte des instruments hyperspectraux, l'implémentation de BCs non-linéaires basée sur des algorithmes d'apprentissage profond (deep-learning) pour-

rait être plus performante en terme de compression pure. Mais d'autres architectures de réseaux peuvent être envisagées, par exemple, l'utilisation de réseaux neuronaux convolutifs qui prendraient en compte une information sur le profil de la variable géophysique à restituer afin de contraindre la compression des spectres. D'autres techniques mixtes de compression/sélection permettraient une utilisation optimale de l'information présente dans les spectres des sondeurs hyperspectraux en vue de l'inversion.

À ce jour, plus de 510 satellites ont été lancés pour l'observation de la Terre. Les instruments sont développés avec des résolutions spatiales et spectrales de plus en plus élevées : alors que l'ESA travaille sur un prototype d'instrument hyperspectral micro-onde, le sondeur IASI-NG et ses 16922 canaux sera en orbite dès 2022. La quantité de données satellites augmente d'un téraoctet par jour. En 2014, pour la seule Agence ESA, la quantité de données d'observation de la Terre dépassait 1.5 PB. Dans l'avènement du "Big Earth Data", l'observation satellite constitue une autre source de plus en plus importante de données climatiques. Le volume d'observation pour la transmission et l'exploitation des données nécessitera l'utilisation de techniques physico-statistiques de plus en plus avancées.

3 L'intégration sous contrainte

Dans le contexte d'optimisation de produits satellites dérivés, la fusion de multiples produits satellites pour une même composante du cycle de l'eau réduit les incertitudes d'estimation et permet une réduction des résidus du bilan hydrique, mais la fermeture n'est toujours pas satisfaisante d'où la nécessité de mettre en place des techniques d'intégration. Le Chapitre 4 a présenté une contrainte de fermeture du bilan hydrique comme approche d'intégration *a posteriori* de multiples produits de variables différentes.

Plusieurs types d'intégration existent. L'assimilation des produits satellites dans un modèle de surface via un filtre de Kalman ou des méthodes variationnelles avec la contrainte du bilan hydrique à satisfaire, permet d'obtenir des produits optimisés. Malheureusement, l'utilisation d'un modèle dans l'intégration des produits satellites inhibe l'opportunité de calibrer des modèles *a posteriori* avec ces produits intégrés (car dès lors ce ne sont plus des observations indépendantes du modèle). L'autre approche consiste à intégrer les produits satellites uniquement via une contrainte de fermeture dans le cadre de l'estimateur Bayésien. La prise en compte de l'équation du bilan hydrique permet de corriger les produits satellites, de les rendre à la fois plus cohérents dans l'équilibre entre les diverses composantes du cycle de l'eau (précipitation, évapotranspiration, ruissellement, contenu en eau dans le sol, contenu en vapeur d'eau atmosphérique et divergence de l'humidité atmosphérique) et plus proches des données *in situ* à l'échelle du sous-bassin comme à l'échelle du pixel.

Perspectives

La technique d'intégration pourrait être améliorée avec une meilleure caractérisation d'incertitude des divers produits satellites ainsi qu'un plus grand nombre de produits pour certaines composantes comme la vapeur d'eau atmosphérique ou bien la

divergence d'humidité. D'autre part, une estimation des échanges d'eau souterrains à travers les systèmes karstiques méditerranéens serait bénéfique à l'analyse du cycle de l'eau dans la région méditerranéenne. Dans le cas de larges bassins versants suffisamment bien surveillés avec des mesures d'écoulement des affluents et confluents, il serait possible de mettre en place une contrainte de fermeture "en cascade" d'amont en aval. Ces contraintes de fermeture dépendantes permettraient une optimisation plus fiable des données sur l'ensemble du bassin. La méthode d'intégration présentée dans cette thèse à l'échelle du bassin et sous-bassin sera implémentée sur d'autres bassins du globe comme par exemple en Asie pour le Mékong, le Brahmaputra, le Gange et le Yangtze. L'optimisation des produits satellites sur ces bassins himalayens permettra de mieux décrire et par conséquent suivre le cycle de l'eau sur cette partie très sensible du globe.

Les approches d'intégration sous contrainte des composantes du cycle de l'eau via les méthodes variationnelles pourraient être implémentées dans le cadre de l'assimilation opérationnelle de produits satellites dans les centres PNT. Dans le cas de l'assimilation de produits de surface (comme l'humidité du sol), on pourrait ajouter des contraintes de bilan hydrique à respecter comme des observations virtuelles afin d'obtenir une analyse qui prendrait en compte ces contraintes physiques.

Le développement de produits plus intégrés, comprenant, par exemple, de la ré-analyse et des observations satellites permettrait de se rapprocher des besoins des utilisateurs. Une source majeure d'amélioration serait d'augmenter la résolution spatiale et temporelle des produits satellites afin de mieux représenter les événements extrêmes. Une approche basée sur la ré-analyse de données non-satellites (stations météorologiques, données aéroportées, radiosondes) peut, par exemple, être envisagée. La descente d'échelle ("downscaling") des produits satellites par l'utilisation de la ré-analyse permettrait d'obtenir des bases de données satellites avec une meilleure résolution spatiale.

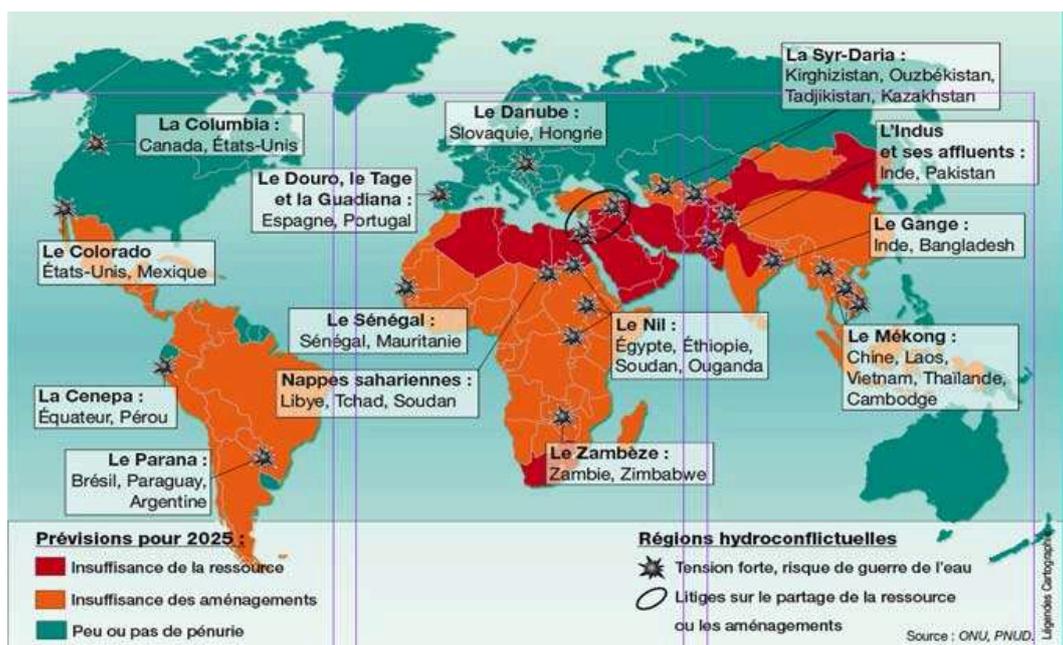


Figure VI.1. Cartographie des tensions et conflits liées aux pénuries d'eau (ONU).

Dans l'attente de la publication du prochain rapport du GIEC sur "l'impact d'un réchauffement global de 1.5°C¹", il me paraît opportun de rappeler les conclusions du rapport technique spécialement consacré à la question de l'eau publié par le GIEC en 2007. Ce dernier permet de faire un point sur les conséquences probables du réchauffement planétaire sur les ressources en eau. Les experts du GIEC estiment que le nombre d'êtres humains souffrant d'un manque d'eau pourrait atteindre le chiffre de 3.2 milliards (la population mondiale va encore augmenter de plus de 2,5 milliards de personnes dans les quarante ans à venir), si les températures devaient augmenter de plus de 4°C. De son côté, la Banque Mondiale estime que 140 millions de personnes migreront d'ici 30 ans pour fuir les effets du réchauffement climatique et en particulier les pénuries d'eau (mais aussi la baisse de la production agricole et la hausse du niveau de la mer). Les populations les plus pauvres seront les premières touchées : 86 millions de migrants climatiques potentiels en Afrique sub-saharienne, 40 millions en Asie du Sud et 17 millions en Amérique latine. La Figure VI.1 montre les projections pour 2025 en termes de disponibilité de l'eau.

Outre les migrations d'origine climatique, qui seront un enjeu pour la coopération internationale, les tensions et les conflits régionaux liés aux pénuries d'eau, pour l'approvisionnement en eau et la maîtrise des crues, seront de plus en plus exacerbés. Pas moins de 260 bassins fluviaux sont partagés entre deux ou plusieurs États et le changement climatique va accentuer encore les concurrences potentielles. La Figure VI.1 montre les régions "hydro-conflictuelles" potentielles d'ici 2025. Si ces litiges n'ont, jusqu'à présent, presque jamais débouché sur des conflits ouverts, ils soulignent la nécessité des processus de discussion et de coopération entre les acteurs internationaux et ces processus requièrent des partenariats internationaux pour le suivi du cycle de l'eau à l'échelle locale et globale. L'observation du cycle de l'eau, l'estimation des ressources disponibles et leurs projections à court, moyen et long termes sont d'une importance capitale, non seulement comme objet de recherche mais aussi comme enjeu socio-économique. En raison de ses inévitables répercussions sociales, l'évolution du cycle de l'eau demeurera un sujet d'étude majeur pour le siècle à venir.

1. Attendu pour octobre 2018.

ACRONYMES

Liste des différents acronymes utilisés dans cette étude :

- **AIRS** : Atmospheric Infrared Sounder
- **ACP** : Analyse en Composantes Principales
- **ADEOS** : Advanced Earth Observing Satellite
- **AER** : Atmospheric and Environment Research
- **AMSR-E** : Advanced Microwave Sounding Radiometer-EOS
- **AMSU-A/B** : Advanced Microwave Sounding Unit A/B
- **ARTS** : Atmospheric Radiative Transfer Simulator
- **ASCAT** : Advanced SCATterometer
- **ATMS** : Advanced Technology Microwave Sounder
- **AVHRR** : Advanced Very High Resolution Radiometer
- **BC** : Bottleneck Channels
- **BON** : Base OrthoNormée
- **CMORPH** : CPC MORPHING technique
- **CPC** : Climate Prediction Center
- **CSR** : Centre for Space Research
- **CEFREM** : Centre de Formation et de Recherche sur les Environnements Méditerranéens
- **CEPMMT** : Centre Européen pour les Prévisions Météorologiques à Moyen Terme
- **CERES** : Clouds and the Earth's Radiant Energy System
- **CNES** : Centre National d'Etudes Spatiales
- **CMEMS** : Copernicus Marine Environment Monitoring Service
- **CORR** : Correlation
- **CRU** : Climatic Research Unit
- **CrIS** : Cross-track Infrared Sounder
- **DFS** : Degree of Freedom of the Signal
- **ECMWF** : European Center for Medium-range Weather Forecast
- **EOS** : Earth Observation System
- **ESA** : European Space Agency
- **ER** : Entropy Reduction
- **ERA** : European ReAnalysis
- **FAO** : Food and Agriculture Organization (of the United Nations)
- **GEWEX** : Global Energy and Water-Cycle Experiment
- **GFZ** : GeoForschungsZentrum (German Research Centre for Geoscience)
- **GEO** : GEOstationary (satellite)
- **GOES** : Geostationary Operational Environmental Satellites
- **GPI** : GOES Precipitation Index
- **GMS** : GEO Meteorological Satellite
- **GIMMS** : Global Inventory Modeling and Mapping Studies

- **GLEAM** : Global Land Evaporation Amsterdam Model
- **GPCC** : Global Precipitation Climatology Centre
- **GPCP** : Global Precipitation Climatology Project
- **GRACE** : Gravity Recovery and Climate Experiment
- **GRDC** : Global Runoff Data Centre
- **GRGS** : Groupe de Recherche de Géodésie Spatiale
- **HIRS** : High-resolution Infrared Radiation Sounder
- **IMEDEA** : Institut MEDiterranni d'Estudis Avançats (Mediterranean Institute for Advanced Studies)
- **IASI** : Infrared Atmospheric Sounding Interferometer
- **IASI-NG** : IASI Next Generation
- **ICA** : Independent Components Analysis
- **IPSL** : Institut Pierre Simon Laplace
- **JPL** : Jet Propulsion Laboratory
- **JAMSTEC** : Japan Agency for Marine-Science and Technology
- **LEO** : Low-Earth Orbit
- **LEGOS** : Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
- **LAI** : Leaf Area Index, on utilise l'acronyme anglais pour indice de surface foliaire
- **MERIS** : MEdium-spectral Resolution, Imaging Spectrometer
- **MERRA** : Modern-Era Retrospective Analysis for Research and Application system
- **MHS** : Microwave humidity sounder
- **MOP** : MicroOnde Passif
- **MODIS** : MODerate-resolution Imaging Spectroradiometer
- **MOD16** : MODIS Global Evapotranspiration Project
- **MRE** : Minimization of Relative Error
- **MSWP** : Multi-Source Weighted-Ensemble Precipitation
- **NASA** : National Aeronautic and Space Administration
- **NAPCA** : Noise-Adjusted PCA
- **NCEP** : National Centers for Environmental Prediction
- **NCAR** : National Center for Atmospheric Research
- **NDVI** : Normalized Difference Vegetation Index
- **NOAA** : National Oceanic and Atmospheric Administration
- **NPP** : National Polar-orbiting Partnership
- **NTSG** : Numerical Terradynamic Simulation Group product
- **NSCAT** : The NASA Scatterometer
- **NWP** : Numerical Weather Prediction center
- **OAFIux** : Objectively Analyzed air-sea Fluxes
- **OMM** : Organisation Météorologique Mondiale
- **ONU** : Organisation des Nations unies
- **ORCHIDEE** : Organising Carbon and Hydrology In Dynamic Ecosystems
- **OSS** : Optimal Spectral Sampling
- **PDF** : Probability Density Function, on utilise l'acronyme anglais pour la densité de probabilité
- **PERSIANN** : Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network
- **PR** : Precipitation Radar
- **PM** : Penman-Monteith (equation)

- **PNT** : Pr evision Num erique de Temps (**NWP** en anglais)
- **PT** : Priestley et Taylor (equation)
- **PC** : Principal Component
- **RMS** : Root Mean Square, on utilise l'acronyme anglais pour la racine de l'erreur quadratique
- **RMSD** : Root Mean Square of the Difference, on utilise l'acronyme anglais pour la racine de la diff erence quadratique
- **RTM** : Radiative Transfer Model (mod ele de transfert radiatif en anglais)
- **SMOS** : Soil Moisture and Ocean Salinity
- **SSM/I** : Special Sensor Microwave/Imager
- **SSMIS** : Special Sensor Microwave Imager/Sounder
- **SRB** : Surface Radiation Budget
- **SWOT** : Surface Water Ocean Topography
- **TMI** : TRMM Microwave Image
- **TMPA** : TRMM Multi-satellite Precipitation Analysis
- **TRMM** : Tropical Rainfall Measuring Mission
- **TIROS** : Television and InfraRed Observational Satellite
- **TIGR** : Thermodynamic Initial Guess Retrieval
- **TOVS** : TIROS Operational Vertical Sounder
- **VI** : Visible
- **IR** : InfraRouge
- **WCRP** : World Climate Research Programme

Annexe A

Équivalence dans le calcul des contaminations dans la restitution des profils atmosphériques

Considérons l'effet de contamination de l'incertitude sur l'humidité q sur la restitution de la température T .

Effet de contamination- On définit :

$$R_q = H_q \cdot B_q \cdot H_q^t \quad (\text{A.1})$$

la matrice de covariance d'erreurs gaussiennes liée à la contamination de l'incertitude sur le profil d'humidité q via le transfert radiatif sur les canaux. La matrice de covariance d'erreur de restitution de T devient avec ce terme :

$$A'_T{}^{-1} = B_T{}^{-1} + H_T^t \cdot (R + R_q)^{-1} \cdot H_T \quad (\text{A.2})$$

L'identité matricielle de *Woodbury* peut alors être utilisée :

$$(R + H_q \cdot B_q \cdot H_q^t)^{-1} = R^{-1} - R^{-1} \cdot H_q \cdot (B_q^{-1} + H_q^t \cdot R^{-1} \cdot H_q)^{-1} \cdot H_q^t \cdot R^{-1} \quad (\text{A.3})$$

En combinant les Eqs. (A.2) et (A.3) :

$$A'_T{}^{-1} = B_T{}^{-1} + H_T^t \cdot R^{-1} \cdot H_T - H_T^t \cdot R^{-1} \cdot H_q \cdot \quad (\text{A.4})$$

$$(B_q^{-1} + H_q^t \cdot R^{-1} \cdot H_q)^{-1} \cdot H_q^t \cdot R^{-1} \cdot H_T \quad (\text{A.5})$$

Soit $B_2 = H_T^t \cdot R^{-1} \cdot H_q$. avec R diagonale, B_2 serait identique dans le cas d'une contamination liée à l'incertitude sur le profil de température T . Alors :

$$A'_T{}^{-1} = A^{-1} + H_T^t \cdot R^{-1} \cdot H_T - B_2 \cdot (B_q^{-1} + H_q^t \cdot R^{-1} \cdot H_q)^{-1} \cdot B_2^t \quad (\text{A.6})$$

Finalement :

$$= A_T^{-1} - B_2 \cdot A_q \cdot B_2^t \quad (\text{A.7})$$

Restitution simultanée- Lorsqu'on considère la restitution simultanée de T et q , on a la matrice de covariance d'erreur de restitution égale à :

$$A_{T+q}{}^{-1} = \begin{pmatrix} B_T{}^{-1} & 0 \\ 0 & B_q{}^{-1} \end{pmatrix} + \begin{pmatrix} H_T^t & H_q^t \end{pmatrix} \cdot R^{-1} \cdot \begin{pmatrix} H_T \\ H_q \end{pmatrix} \quad (\text{A.8})$$

$$A_{T+q}^{-1} = \begin{pmatrix} B_T^{-1} + H_T^t \cdot R^{-1} \cdot H_T & H_T^t \cdot R^{-1} \cdot H_q \\ H_q^t \cdot R^{-1} \cdot H_q & B_q^{-1} + H_q^t \cdot R^{-1} \cdot H_T \end{pmatrix} \quad (\text{A.9})$$

$$A_{T+q}^{-1} = \begin{pmatrix} A_T^{-1} & B_2 \\ B_2^t & Q_q^{-1} \end{pmatrix} \quad (\text{A.10})$$

Alors par l'inverse par block de la matrice $A_{T+q}^{-1} =$

$$\begin{pmatrix} (A_T^{-1} - B_2 \cdot A_q \cdot B_2^t)^{-1} & -A_T \cdot B_2 \cdot (A_q^{-1} - B_2 \cdot A_T \cdot B_2^t)^{-1} \\ -A_q \cdot B_2^t \cdot (A_T^{-1} - B_2 \cdot A_q \cdot B_2^t) & (A_q^{-1} - B_2 \cdot A_T \cdot B_2^t)^{-1} \end{pmatrix}^{-1} \quad (\text{A.11})$$

$$A_{T+q}^{-1} = \begin{pmatrix} A_T' & -A_T \cdot B_2 \cdot A_q \\ -A_q \cdot B_2^t \cdot A_T & A_q' \end{pmatrix}^{-1} \quad (\text{A.12})$$

La matrice de covariance d'erreur sur la restitution de T est ici équivalente à A_T' (première approche avec effets de contamination) si et seulement si les éléments diagonaux hors bloc ($A_T \cdot B_2 \cdot A_q$ et $-A_q \cdot B_2^t \cdot A_T$) sont négligeables. Cela signifie que la première approche de contamination n'est valide que si la corrélation entre les variables à restituer (p. ex. T et q) est négligeable. La première approche ne prend en compte que le bruit induit par l'incertitude sur les autres variables.

Annexe B

Autre méthode d'extrapolation spatiale de la contrainte hydrologique avec régression statistique et indice pixelaire

Principe du Closure Correction Model CCM

Dans cette section, un modèle (statistique) de correction de fermeture (CCM) est introduit pour corriger chaque composante du cycle de l'eau indépendamment, sur la base des résultats de l'intégration SW+PF. Cette méthode a été développée par Munier et al. (2014). Le modèle de correction de fermeture corrige directement la précipitation, sans l'utilisation des autres composantes (E, R et ΔS). Le modèle de correction de fermeture est représenté par des transformations affines simples définies par un facteur d'échelle a , un offset b et un coefficient c permettant de traiter le cas où la variable est nulle afin de ne pas avoir des corrections négatives.

$$Var_{new} = a \cdot Var + b(1 - e^c \cdot Var). \quad (B.1)$$

Où Var est une des composantes du cycle de l'eau. Les paramètres a et b du modèle sont calculés grâce aux régressions linéaires entre les estimations Var_{SW} et les composantes intégrées Var_{SW+PF} .

Afin d'étalonner le modèle (calcul de a , b et c) plusieurs options sont possibles :

- Un modèle est établi pour chaque sous-bassin particulier CCM(bassin)
- Le modèle est établi pour l'ensemble des sous-bassins CCM(global)

Alors que la CCM(global) tend à modéliser une correction "moyenne" sur l'ensemble des sous-bassins, cette méthode n'est pas optimale pour représenter l'hétérogénéité spatiale du cycle de l'eau sur le pourtour méditerranéen (cf Figure V.6). D'autre part, la solution locale CCM(bassin) assure une meilleure prise en compte des caractéristiques locales : chaque modèle est établi suivant le cycle de l'eau sur un sous-bassin particulier. Mais cette solution ne permet pas d'assurer une continuité spatiale d'un sous-bassin à un autre.

Indice pixelaire pour la fermeture Munier et Aires (2017)

Les erreurs de produits satellites ne sont pas homogènes à l'échelle du globe et le CCM devrait idéalement dépendre de différents facteurs tels que les conditions hydro-

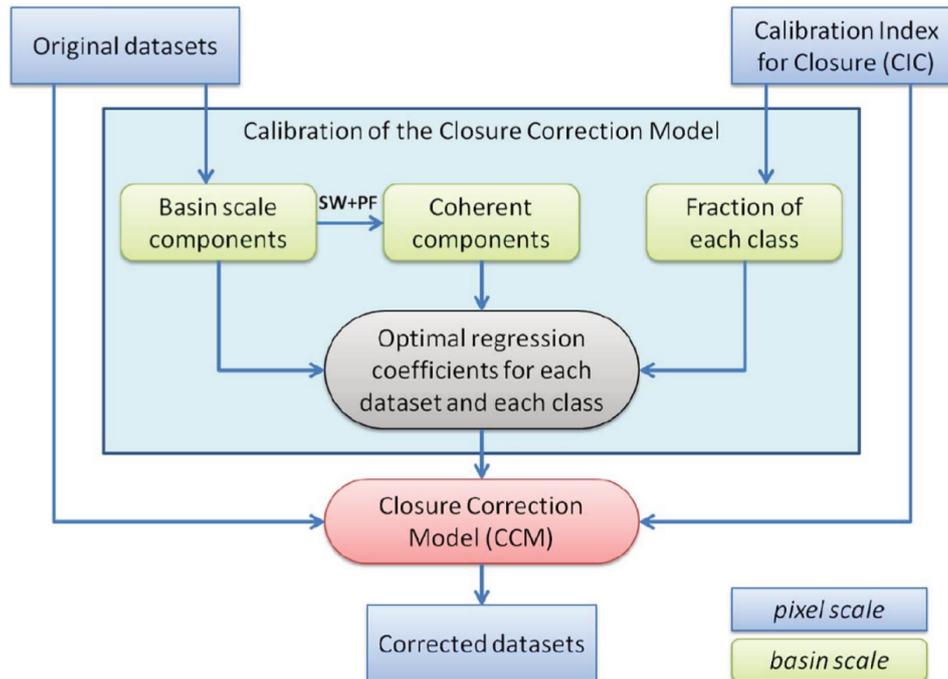


Figure B.1. Ordinoigramme de la méthode CCM et CCI d'après Munier et Aires (2017).

météorologiques. Dans Munier et Aires (2017), les auteurs ont proposé de rendre le CCM dépendant des conditions hydrométéorologiques en utilisant un indice de surface pour classifier les pixels et calibrer le CCM sur différentes classes.

Le but de cette section est de présenter un indice pixelaire d'étalonnage pour la fermeture (Calibration Index for Closure, CIC) tel que développé par Munier et Aires (2017) qui sera utilisé pour caractériser la surface à l'échelle du pixel. On peut alors corriger chaque pixel d'une classe en utilisant la CCM étalonnée sur les bassin caractéristiques de cette classe.

Dans Munier et Aires (2017), l'indice pixelaire est statique et issu de deux variables de surface : $NDVI$ et la moyenne climatologique de $\overline{P-E}$ normalisée. Les résultats de Munier et Aires (2017) indiquent que ces deux variables sont les facteurs qui ont eu le plus d'impact sur l'amélioration du bilan hydrique après CCM.

La Figure B.2 montre l'indice pixelaire utilisant $NDVI$ et $\overline{P-E}$ à l'échelle du globe. Le graphique encapsulé montre la distribution des 11 bassins considérés dans Munier et Aires (2017), dans un diagramme $\overline{P-E}/NDVI$. Ce graphique a été utilisé pour créer un nouvel indice, selon les quatre classe définies arbitrairement. En supposant que les 11 bassins considérés sont représentatifs des différentes conditions hydrométéorologiques, les auteurs étalonnent alors quatre CCM basées sur deux, trois ou plus de bassins. On étalonne alors les modèles CCM(CIC) sur chaque classe de pixel (voir Figure B.1). Cette méthode prend en compte des spécificité local de surface comme proxy de déséquilibre du bilan hydrique ce qui permet de corriger les produits satellites en prenant en compte leur hétérogénéité spatiale.

Contrairement à d'autres systèmes de classification climatique bien connus (comme, par exemple, la classification Köppen (1936)), la classification CIC est quantitative et chaque pixel peut alors être assigné par pondération à deux classes. Cela permet d'éviter les discontinuités aux limites des classes.

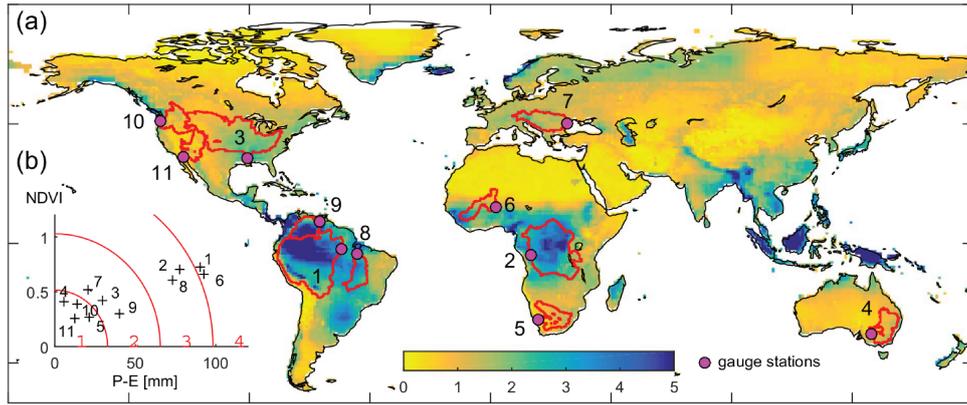


Figure B.2. (a) Graphique de NDVI (multiplié par 100) avec les précipitations nettes ($P-E$) pour les bassins considérés dans cette étude. Chaque bassin est représenté par son numéro sur la carte. Les lignes rouges représentent les limites entre les classes numérotées en rouge. (b) Carte de l'indice pixelaire (CIC). Les limites des bassins sont représentées par des lignes rouges et les stations de mesures par des cercles verts. Munier et Aires (2017)

Nouveau développement pour le CIC

Dans le cadre de cette thèse, nous avons cherché à améliorer l'indice pixelaire afin d'obtenir une CCM(CIC) plus flexible et plus locale permettant une amélioration du bilan hydrique à l'échelle du sous-bassin.

Une première étude a été menée afin de trouver les variables climatiques les plus pertinentes comme proxy du déséquilibre en eau. L'humidité du sol (SM) avec la température de surface ($Temp$) et l'excédent en eau $P - E$, ont été sélectionnés au cours d'une régression séquentielle où l'on cherchait à reconstruire les erreurs du modèle standard CCM(global).

À la différence de Munier et Aires (2017), l'indice n'est plus statique mais varie au cours de la saison pour le pixel. En utilisant les valeurs mensuelles de SM , $Temp$ et $P - E$ entre 2004-2009 sur l'ensemble des 6 sous-bassins continentaux, on obtient le nuage de $6 \cdot 12 \cdot 6 = 432$ points de la Figure B.3 à gauche. L'utilisation de l'algorithme des k-plus proches voisins permet alors la définition de 6 classes (voir plus loin pour la discussion autour du nombre de classes).

La Table B.1 donne des informations hydrologiques quant à la nature de chaque classe.

	P	E	ΔS	ΔWV	DIV	Obs
Class1	73.0	24.0	29.3	-1.8	-47.1	strong precip,
Class2	46.1	36.3	6.9	1.4	-12.6	low Temp
Class3	69.7	33.7	11.2	0.63	-36.7	mountain forcing
Class4	53.0	72.9	-25.6	2.3	17.6	strong evap
Class5	46.3	34.2	7.4	-1.1	-10.9	high Temp
Class6	19.3	37.0	-16.8	-0.6	19.3	low precip

TABLE B.1 – K-mean classes characteristics

L'indice pixelaire varie au cours de la saison climatologique. La Figure B.4 montre

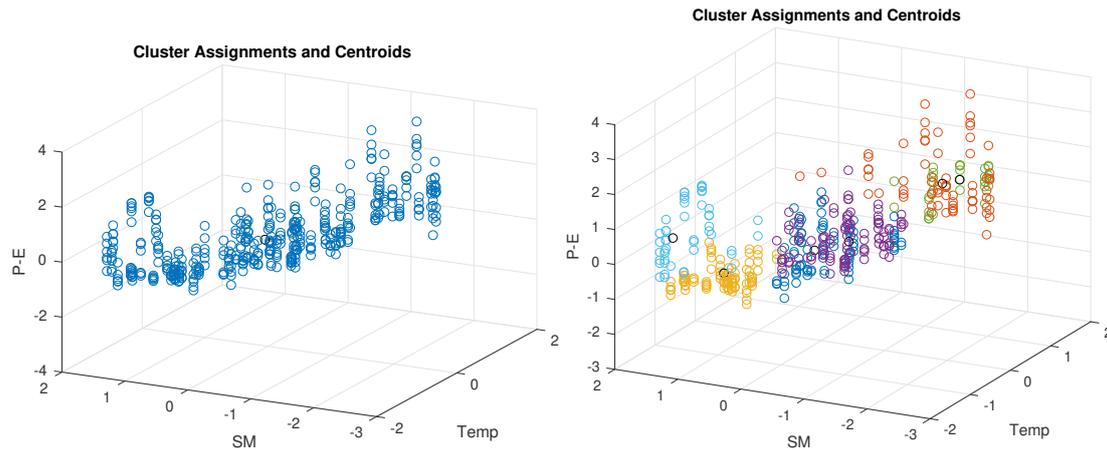


Figure B.3. Illustration du calcul des classes avec la méthode *k-means* à partir des trois variables choisies : Température *Temp*, humidité du sol *SM* et $P - E$.

les indice pixelaire sur le pourtour méditerranéen au mois de janvier et juillet.

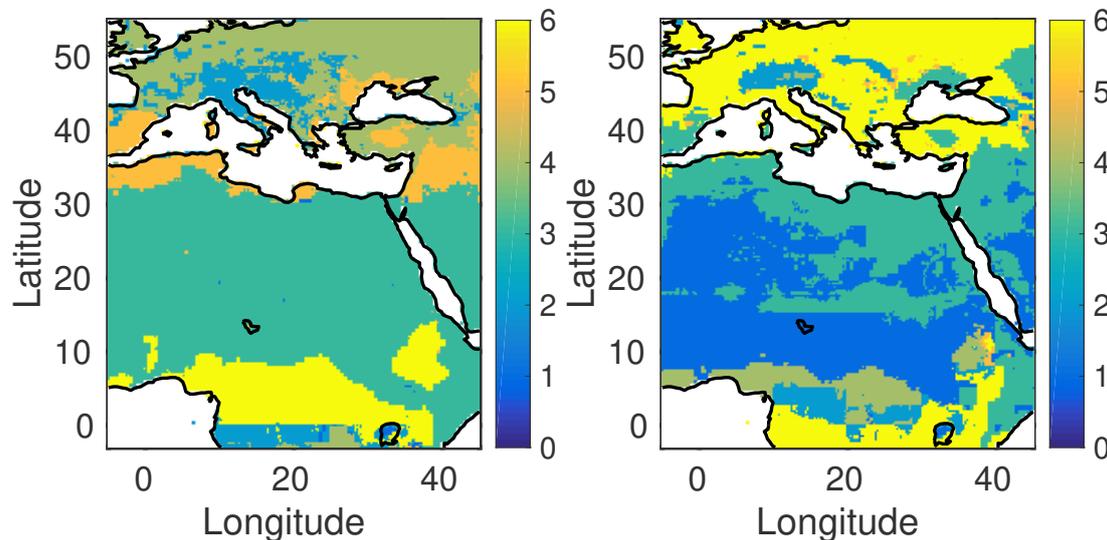


Figure B.4. 6 classes définies dans le CIC en utilisant l'humidité du sol, la température et $P-E$ pour janvier (gauche) et juillet (droite).

Validation et discussion

Bilan hydrique

La Figure B.6 montre la RMS du résidu du bilan hydrique continental (à gauche) et atmosphérique (à droite) obtenus avec les différentes stratégies SW, SW+PF, CCM(Global), CCM(bassin) CCM(köppen) ou un modèle de calibration pour la fermeture est étalonné sur les classes de köppen et CCM(CIC) basé sur l'indice mis en place avec *SM*, *Temp* et $P - E$.

L'étalonnage CCM par bassin (CCM(bassin)) est le meilleur consensus possible obtenu avec la méthodologie de la CCM et réduit fortement les résidus des bilans hy-

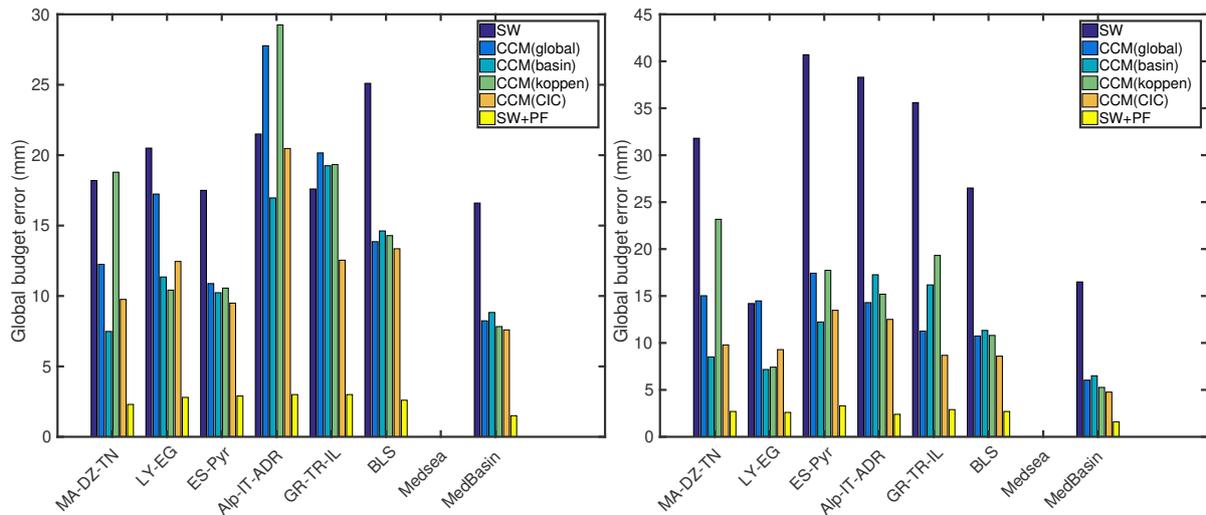


Figure B.5. *RMS de résidus du bilan hydrique de surface (gauche) et du bilan hydrique atmosphérique (droite) pour les différentes méthodes développées dans ce travail de recherche.*

driques. Cela met en évidence la pertinence de la méthode de régression CCM pour approcher la référence SW+PF. L'amélioration est plus faible pour la composante continentale du bilan hydrique que atmosphérique. Cela peut s'expliquer par des variations moins linéaires dans le changement du contenu en eau dans le sol, difficilement approximable linéairement par la CCM. En outre l'amélioration est relative au sous-bassin. Si le CCM(bassin) réduit jusqu'à 50% du résidu continental sur le sous-bassin BLS, il est moins performant sur le sous-bassin Alp-IT-ADR, là aussi la particularité locale de ce sous-bassin montagneux avec des précipitation neigeuse et un fort albédo en hiver rend moins linéaires les effets de la fermeture. Le sous-bassin GR-TR-IL est le seul sur lequel le modèle CCM(bassin) n'arrive pas à améliorer la solution SW en terme de résidu continental. Ce sous-bassin présente une orographie particulière ou la présence de nombreuses îles dans la partie orientale de la méditerranée tend à largement bruyter les estimations du contenu en eau dans le sol ce qui peut être à l'origine de bruit non-linéaires. Un modèle CCM plus sophistiqué pourrait réduire encore davantage les résidus. Cependant, aucune CCM simple ne peut corriger des erreurs fondamentales dans les observations satellites originales. La méthode CCM(global) montre globalement de meilleures performances que SW avec une réduction moyenne de 25%. Les bassins affichant les pires performances sont ceux dont les caractéristiques sont les plus particulières dans le pourtour méditerranéen : Alp-IT-ADR et GR-TR-IL. La prise en compte des caractéristiques hydrométéorologiques avec le CIC réduit significativement les résidus pour tous les bassins avec des performances proches de CCM(bassin).

Le nombre de classes pour le CIC

Le nombre de classes pixelaires à définir dans le cadre d'un modèle CCM(CIC) est un problème lié au dilemme de biais-variance. Moins il y a de classes, plus le nombre de points nécessaires à l'apprentissage statistique de CCM sur cette classe est élevée et meilleur est cet apprentissage en terme de décart type. Mais plus l'échantillonnage est représentatif d'une moyenne (fort biais) et ne représentera pas de spécificité local. Avec une seule classe, l'apprentissage est bon mais le CCM ne rend compte de la fermeture

qu'à grande échelle. Avec 6 classes, le nombre de points par classe diminue et l'erreur d'étalonnage augmente en terme d'écart type en revanche, les spécificités locales (biais) sont mieux prise en compte. Il faut alors sélectionner un nombre de classes pour faire un compromis judicieux entre le biais et l'écart-type de l'erreur. La Figure B.6 montre la performance de CCM(CIC) en termes de résidus de bilan hydrique continental (gauche) et atmosphérique (droite) pour un nombre croissant de classes et pour trois apprentissages différents. Le résidu des bilans diminue avec un nombre croissant de classes (meilleur biais dans l'estimation) mais à partir de 6 classes la variabilité de l'apprentissage est trop grande et vient impacter les performance de CCM(CIC). D'après cette étude, nous avons fixé à six le nombre de classes dans notre approche CCM(CIC).

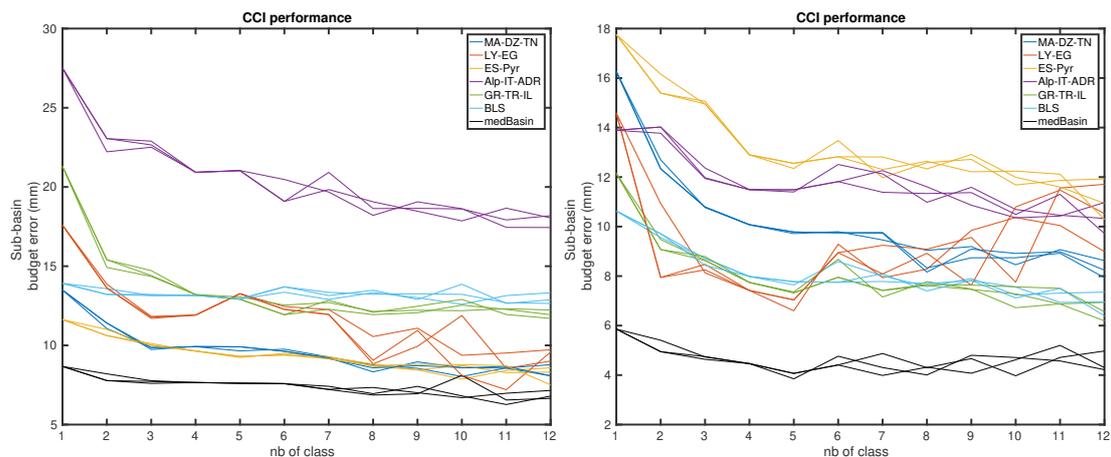


Figure B.6. RMS des résidus du bilan hydrique de surface (gauche) et atmosphérique (droite) en utilisant la méthodologie CIC avec un nombre croissant de classes (3 tirages pour montrer la variabilité de la calibration en fonction du nombre de classes définies).

Conclusion sur l'approche CCM

CCM(global) est performante à l'échelle du bassin méditerranéen entier mais ne permet pas la prise en compte de spécificités locales. Outre les résultats présentés par Munier et Aires (2017), les nouveaux développements d'un indice CIC sur six classes, variables dans la saison, basé sur trois sources d'informations hydrométéorologiques SM , $Temp$ et $P - E$ permet la mise en place d'un modèle CCM qui réduit quantitativement les résidus du bilan hydrique. Néanmoins cette méthode s'éloigne de la notion du bassin versant pour se concentrer sur la correction d'erreur de produits satellites à partir d'une information de "surface" et un bassin peut être alors dans une classe ou une autre selon le mois de l'année. Si cette méthode permet l'extrapolation spatiale de la correction à l'extérieur des sous-bassin, elle n'est pas optimale à l'intérieur. CCM(bassin) est optimal à l'intérieur des sous-bassins mais son extrapolation n'est pas évidente. La région méditerranéenne est une région naturellement bordée, par le désert au sud et l'atlantique au nord. Les bassins du pourtour sont mitoyen et l'utilisation de la distance géographique est une bonne piste pour améliorer la spatialisation et l'extrapolation spatiale de notre solution SW+PF ce qui permet de conserver la notion de sous-bassin. En outre la calibration saisonnière que permet le modèle CCM(CIC)

peut être directement implémentée dans le CCM(bassin) mais avec le risque que le dilemme biais-variance ne permette pas une bonne calibration en vu du nombre de points (72) par bassin.

Table des figures

.1	<i>Maximilien Pellet</i> , l'observation du cycle de l'eau depuis les satellites. Monotype, 2017	i
I.1	Schéma de génération d'un produit satellite à partir de la mesure satellite. Modifié à partir de Sun et al. (2018)	4
I.2	Tendance de la croissance mondiale du volume de données climatiques 2010-2030 (Overpeck et al., 2011) (à gauche). Le détail du volume de données climatiques avec, pour la partie observations satellites, le volume en fonction de la plateforme satellite sur 2006-2020 (document de la NOAA) (à droite).	5
I.3	Schéma illustrant l'assimilation d'observations satellites au sein d'un modèle.	7
I.4	Capacité d'échantillonnage vertical de l'atmosphère pour un sondeur traditionnel infrarouge à 19 canaux, le High-resolution Infrared Radiation Sounder (HIRS/4) (à droite) et pour le sondeur hyperspectral infrarouge IASI et ses 8461 canaux (à gauche).	9
II.1	Sur le graphique, la variation du coefficient d'extinction atmosphérique est représentée en bleu. La transmission atmosphérique est en vert, et la fonction de poids résultante est représentée en rouge. Ces courbes sont données à titre d'exemple, elles ne représentent pas une longueur d'onde particulière (Paul, 2013).	15
II.2	Schéma des différents rayonnements reçus par le satellite au sommet de l'atmosphère : l'émission de la surface, le rayonnement montant émis par l'atmosphère et le rayonnement descendant émis par l'atmosphère puis réfléchi par la surface (Paul, 2013).	16
II.3	Représentation du spectre électromagnétique et correspondance entre les différentes unités. Les utilisations les plus communes des différentes gammes de longueur d'onde sont représentées. Les instruments IASI (Infrared Atmospheric Sounding Interferometer) et AMSU (Advanced Microwave Sounding Unit) sont également représentés (Paul, 2013). . .	17
II.4	La transmittance relative de l'atmosphère pour un rayonnement à différentes longueurs d'onde. Les zones blanches désignent les fenêtres atmosphériques, dans lesquelles le rayonnement n'interagit pas beaucoup avec les molécules d'air et n'est donc pas absorbé. (http://geoinfo.amu.edu.pl)	18

II.5	Principe du sondage satellite. À gauche, spectre d'émission atmosphérique dans l'infrarouge entre 650 et 800 cm^{-1} . Ce spectre donne la température de brillance en fonction du nombre d'ondes (l'inverse de la longueur d'onde). A droite profil de la température atmosphérique avec les fonctions de poids de quatre canaux (Phulpin et al., 2003).	20
II.6	Comparaison des fonctions de poids liées aux canaux du sondeur micro-onde AMSU (à droite) et à ceux du sondeur hyperspectral infrarouge IASI (à gauche).	21
II.7	Exemple de représentation graphique d'un nuage de points de dimension 2. Elle peut être décrite par l'abscisse et l'ordonnée d'origine ou bien les nouveaux axes verts. Finalement, la projection sur l'unique axe PCA-1 st Dimension semble suffisante pour rendre compte de la variabilité du nuage.	24
II.8	Représentation schématique des BCs stricts (a) et projetés (b). De gauche à droite, les observations bruitées \tilde{y} , les BCs $\tilde{\eta} \approx y_s$, et le spectre reconstruit \tilde{y} . Les matrices P et Q sont les deux régressions linéaires entre ces trois vecteurs.	36
II.9	Exemple de corrélation entre les BC-projetés (version linéaire) sur l'axe des y et les canaux de température de Brillance (TB) (axe des x) pour un instrument hyperspectral (251 canaux) micro-onde.	37
II.10	Corrélation croisée entre le profil de température et celui de l'humidité le long de l'atmosphère. Les axes vertical et horizontal représentent la pression atmosphérique des niveaux verticaux (hPa).	39
II.11	Jacobiens en Température (gauche) et humidité (droite) pour les 251 canaux choisis dans les bandes d'absorption à 60, 118, 183, 325, 425, and 448 GHz. L'axe horizontal n'est pas continu en fréquence.	41
II.12	Évolution de l'erreur de dé-bruitage (i.e., $y - \bar{y}$), avec un nombre croissant de composantes (pour la PCA et NAPCA) ou de BCs stricts ou projetés (linéaires ou non). (A) représente l'erreur de dé-bruitage moyen sur tout le spectre d'observation, et (B) représente l'erreur de dé-bruitage maximal atteint sur les plus subtiles canaux. Le bruit instrumental moyen comprenant l'erreur du modèle direct est représenté par sa déviation standard à des fins de comparaison (en noir).	42
II.13	Erreur de dé-bruitage ($y - \bar{y}$), pour les six bandes spectrales (à gauche), et autour de la bande 60 GHz (au milieu), et autour de la bande 183 GHz (à droite). Les différentes méthodes de compression sont présentées, de haut en bas : PCA, NAPCA, BC-stricts, BC-Projetés/Lin et BC-Projetés/non-Lin pour un nombre différent de composantes ou BCs. Le bruit instrumental avec l'erreur du modèle est représenté en noir à des fins de comparaison.	44
II.14	Erreur de dé-bruitage pour un instrument micro-onde $STD(y - \bar{y})$ où $\bar{y} = Q \cdot P \cdot \tilde{y}$, est une régression linéaire, ou bien $\bar{y} = NN_2(NN_1(\tilde{y}))$, une régression utilisant un réseau de neurones; avec 16 (A) ou 100 (B) BCs utilisés.	45

II.15	De haut en bas : fréquence de canaux sélectionnés par BC-Projeté/Lin, MRE et ER. De gauche à droite : sélection de 50 canaux dans la bande O_2 pour la restitution de T uniquement (à gauche), sélection de 50 canaux dans la bande H_2O pour la restitution de q uniquement (au milieu), 100 canaux dans les bande O_2 et H_2O pour la restitution combinée de $T + q$ (à droite).	47
II.16	Pour les PCs/BCs/canaux obtenus : Jacobien en température T (K/K) (les deux colonnes de gauche) et Jacobien en humidité q (K/Kg/Kg) (les deux colonnes de droite). De haut en bas : NAPCA, BC-Projeté/Lin, MRE et ER. Pour MRE et ER, les Jacobiens T sont affichés pour la sélection pure- T ou $T + q$, et les Jacobiens q sont affichés pour la configuration pur- q ou $T + q$. Chaque fois, les sept premiers PCs/BCs/canaux sont indiqués en gras et leurs fréquences sont indiquées, les autres PCs/BCs/canaux d'ordre supérieur sont représentés par une ligne en pointillés.	48
II.17	Taux d'amélioration (en %) de l'erreur de restitution par rapport à l' <i>a priori</i> pour la température (en haut) et l'humidité (en bas) lors de l'utilisation de 20 canaux sélectionnés dans le micro-onde : 10 canaux sélectionnés pour T plus 10 canaux pour q (ligne grise en pointillés) ou 20 canaux simultanément pour T et q (ligne noire continue).	49
II.18	Taux d'amélioration (en %) de l'erreur de restitution par rapport à l' <i>a priori</i> pour température (en haut) et l'humidité (en bas) lors de l'utilisation de 20 (à gauche) ou 100 (à droite) PCs/BCs/canaux, pour les méthodes : ER, MRE, BC/LIN et NAPCA. La sélection a été effectuée simultanément pour T et q	50
II.19	Spectre en température de brillance de IASI simulé (en haut) avec le bruit instrumental correspondant (en bas).	52
II.20	Distribution spatiale des situations dans la base de données TIGR, la couleur du pixel représente la température de surface (T_s) en K	53
II.21	De haut en bas : Jacobiens des canaux de l'instrument IASI en température (T), humidité (q), et ozone (O_3), donnés pour une situation particulière dans TIGR.	54
II.22	Évolution des erreurs de projection, $E[(\eta_s - TB)^2]$. En utilisant 100 BC, avec un nombre croissant de canaux voisins pour les approches BC-Localisé/linéaire et BC-Localisé/non-linéaire, comparées à l'erreur de projection des BC-projetés standards.	55
II.23	De haut en bas : ER-selection et BC-contraints.	56
II.24	Erreur de débruitage ($y - \hat{y}$) pour un nombre croissant de PCs ou BCs. Le bruit instrumental moyen est représenté par une ligne noire en pointillés.	57
II.25	Jacobien des canaux (colonne à gauche) ou des BCs (colonne au milieu) et amélioration de la restitution, par rapport à l' <i>a priori</i> pour les différentes approches (colonne de droite), en température (en haut), humidité (au milieu) et ozone (en bas).	59
II.26	Déviation standard (STD) de l'erreur de restitution NN de la température (à gauche) et de l'humidité (à droite), pour le NAPCA, le ER, la sélection Collard (2007) et BC-projeté/lin/contraints, avec 30 (lignes pointillées) et 270 composants/canaux (lignes continues).	61

II.27	Corrélation entre les 50 premières composantes NAPCA issues du spectre d'IASI et les profils T (en haut), q (au milieu) et O_3 (en bas).	63
III.1	Diagramme schématique du cycle de l'eau méditerranéen.	67
III.2	Ordinogramme de la restitution des précipitations de produits satellites multi-sources (Sun et al., 2018).	70
III.3	Estimation de précipitation pour le mois de janvier 2004 des produits TMPA 3B42-V7 (haut, gauche) et GPCP-2 (haut, droite) PERSIANN (bas, gauche) et CMORPH (bas, droite). On peut remarquer la très grande disparité de ces estimations, issues d'algorithmes différents, même à l'échelle du mois. Cela fera l'objet d'étude et de commentaires au Chapitre IV.	72
III.4	Estimation du taux d'évaporation pour le mois de janvier 2004 des produits OAFflux (à gauche) et SEAFflux (à droite). À la différence des produits satellites de précipitation, les produits d'évaporation, incorporent des sortie de ré-analyse et une paramétrisation qui permet une certaine cohérence entres les deux estimations.	77
III.5	Estimation du taux d'évapotranspiration pour le mois de janvier 2004 des produits GLEAM (à gauche), MOD16 (au centre) et NTSG (à droite). Deux d'entre eux ne donnent pas d'estimation évapotranspiration pour la zone désertique du Sahara. Outre ce constat pour cette zone désertique, on observe de forte différences entre le produit GLEAM et MOD16 au dessus des reliefs Alpains. La paramétrisation des différents produits au dessus des régions complexes (désert ou montagne) est problématiques ce qui met en évidence les limites de l'estimation satellite de l'évapotranspiration.	81
III.6	Schéma explicatif de la mesure des variations de masse d'eau sur terre depuis les deux instruments GRACE (Japan Agency for Marine-Science and Technology, JAMSTEC ; Nagoya University).	82
III.7	Estimation des changements de masse d'eau entre décembre 2003 et janvier 2004 des produits "land-only" JPL (en haut, à gauche), CSR (en haut, au centre), GFZ (en haut, à droite) et du produit GRGS (en bas). Alors que trois produits sont très semblables, le quatrième, GRGS, diffère dans sa structure spatiale.	83
III.8	Estimation des changements de masse d'eau atmosphérique entre décembre 2003 et janvier 2004 du produit GlobVapour. On observe pour ce mois particulier, une estimation très bruitée.	84
III.9	Visualisation des côtes dans les produits CEFREM-V2 et ORCHIDEE au niveau de l'Italie (gauche). Débit d'eau annuel entrant dans la Méditerranée entre 1980 et 2009 pour les deux produits CEFREM-V2 et ORCHIDEE (droite).	85
III.10	Écoulement net à Gibraltar (en $mm.mois^{-1}$) avant (bleu) et après (rouge) l'ajustement de biais (Jordà et al., 2017). L'aire totale de la Méditerranée incluant la mer Noire est de $3.5 \times 10^6 km^2$	86

III.11	Distribution spatiale des jauges utilisées dans le produit EOBS (Haylock et al., 2008). On remarque la répartition hétérogène des jauges de précipitation. L'ensemble des données <i>in situ</i> pour la région nord-africaine est relativement épars ce qui limite la résolution du produits EOBS pour cette région. En outre aucune estimation n'est obtenue pour la mer Méditerranée.	87
III.12	Distribution spatiale des sites présents dans la base de données FLUX-NET. Les différentes zones de Köppen sont précisées sur la carte (http://flux-net.fluxdata.org/sites/).	88
III.13	Réduction de l'erreur de prévision ("Forecast Error", FE) en pourcentage par assimilation de différents types d'observations satellites dans ERA-I à deux dates (English et al., 2013).	89
III.14	Estimation de la divergence d'humidité atmosphérique verticalement intégrée pour le mois de janvier 2004 dans le produit de ré-analyse ERA-I.	90
III.15	Estimation des changements de masse d'eau atmosphérique entre décembre 2003 et janvier 2004 dans le produit de ré-analyse ERA-I. Contrairement au produit satellite bruité (voir Figure III.8), l'utilisation d'un modèle permet ici d'obtenir une estimation plus continue spatialement.	90
IV.1	La mer méditerranée et sa région (www.earth.google.fr)	93
IV.2	Ressources en eau naturelles renouvelables et ressources exploitables, par habitant, dans les pays méditerranéens (pour l'année 2000, Margat (2004))	96
IV.3	Organisation des différents groupes de travail (WG pour Working Group) autour du projet HyMeX.	98
IV.4	Sous-bassins de l'aire de drainage méditerranéenne considérés dans cette étude. Les sous-bassins ont été définis à l'aide d'un modèle hydrologique (Wu et al., 2011), et les rivières proviennent de HydroShed (http://www.hydrosheds.org/). Voir le texte pour la définition des sous-bassins.	100
IV.5	Climatologie saisonnière des composantes du cycle de l'eau (en lignes) sur les différents sous-bassins (en colonnes) dans la région méditerranéenne, à partir des observations satellites et de la ré-analyse.	103
IV.6	Moyenne mensuelle pour la précipitation via SW (première rangée) et ERA-I (deuxième rangée). De gauche à droite, pour : Janvier, avril, juillet et octobre. La dernière colonne représente la moyenne annuelle.	110
IV.7	Figure IV.6 mais pour l'évapotranspiration	111
IV.8	De haut en bas, les moyennes mensuelles pour : le ruissellement, les variations du contenu en eau sur terre et sur mer, les variations de la vapeur d'eau atmosphérique et la divergence de l'humidité estimés via SW. De gauche à droite, pour : Janvier, avril, juillet et octobre. La dernière colonne représente la moyenne annuelle.	113
IV.9	Écoulement net à Gibraltar après l'ajustement de biais (Jordà et al., 2017) entre 2004 et 2016 et sa climatologie saisonnière en $mm.mois^{-1}$. L'aire totale de la Méditerranée avec la mer Noire est de $3.5 \cdot 10^6 km^2$	115

IV.10	Cycle de l'eau annuel moyen vu par les produits satellites pour les six sous-bassins, la partie océanique et l'ensemble du bassin méditerranéen : précipitations (rouge), évapotranspiration (bleu), ruissellement (rose), divergence (orange), résidus du bilan hydrique au niveau de la surface (vert) et dans l'atmosphère (jaune). Les moyennes annuelles sont exprimées en $mm.mois^{-1}$	116
V.1	Illustration du théorème de Bayes en deux dimensions. (Rodgers, 2000)	122
V.2	Fonctionnement séquentiel du filtre de Kalman. A un instant donné k , nous disposons de x_k^f (point bleu) une prévision faite grâce au modèle à partir de l'état analysé au temps précédent, (équivalent à $\hat{x}_{k k-1}$ dans le texte) de sa matrice de covariance d'erreur P_k^f (symbolisée par l'ellipse bleue) et d'une observation y_k (triangle vert). La prévision est corrigée à l'aide de l'observation pour produire l'analyse x_k^a (point rouge, équivalent à $\hat{x}_{k k}$ dans le texte) avec pour matrice de covariance d'erreur P_k^a (symbolisée par l'ellipse rouge). (Bonan, 2014)	128
V.3	Cycle du 3D-Var. A chaque cycle c , le modèle d'évolution propage l'ébauche du système x_b jusqu'à l'instant $t_{k,c}$ pour produire $x_b(t_{k,c})$ (courbe bleue). L'assimilation par 3D-VAR consiste ensuite à utiliser les observations (triangles verts) pour produire une analyse $x_a(t_{k,c})$. Pour cela, il faut calculer la distance entre les observations (triangles verts) et les sorties du modèle $H_k(x)$. Après la minimisation, $x_a(t_{k,c})$ peut être utilisé comme condition initiale pour le cycle suivant. (Bonan, 2014).	131
V.4	Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'une des intégration (en rouge) ne considère comme contrainte seulement l'équation continental, l'autre (en jaune) contraint le cycle de l'eau dans sa globalité. La fermeture de la partie continentale seule ne permet pas de fermer le cycle de l'eau sur sa partie océanique et atmosphérique.	138
V.5	Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'intégration est effectué sans (en rouge) ou avec (en jaune) décomposition fréquentielle. La contrainte sans décomposition fréquentielle engendre des hautes fréquences dans la correction du signal de l'évapotranspiration.	140
V.6	Série temporelle entre 2004 et 2009 du résidus du bilan hydrique au niveau du sous-bassin BLS avec une correction issue d'une fermeture seulement à l'échelle du bassin méditerranéen total (CCM) et une fermeture à l'échelle du sous-bassin BLS (SW+PF).	141
V.7	Série temporelle entre 2004 et 2009 des différentes composantes du cycle de l'eau avant et après intégration. L'une des intégration (en rouge) ne considère comme contrainte seulement le cycle de l'eau à l'échelle du bassin méditerranéen entier, l'autre (en jaune) contraint le cycle de l'eau dans sa globalité et à l'échelle du sous-bassin et de façon interdépendante entres les sous-bassins. La fermeture inter-dépendante des sous-bassins avec le bassin total et la partie océanique assure un équilibre du bilan hydrique à grande échelle.	143

V.8	Etapes de la spatialisation de la contrainte hydrologique depuis SW jusqu'à INT pour la précipitation (à gauche) et l'évapotranspiration (à droite), pour juillet 2008. Les unités sont en $mm.mois^{-1}$	146
V.9	Comparaison des estimations des composantes du cycle de l'eau et de la densité de probabilité des résidus du bilan hydrique (en ligne). Les estimations concernent la période 2004-2009 sur les 6 sous-bassins continentaux, la partie océanique et le bassin méditerranéen total (en colonne) issus des différentes méthodologies présentées dans l'étude : SW, SW+PF, INT et CAL.	151
V.10	Validation du contenu en eau sur mer des produits SW, INT et CAL par rapport à l'altimétrie moins le contenu thermique.	153
V.11	Gauche : Diagramme de dispersion de la RMSD entre la station FLUXnet et les produits SW, INT et CAL, pour les précipitations. Les points sous la ligne 1:1 (verts) indiquent une amélioration et les points au-dessus de la ligne (rouges) indiquent une dégradation. La figure encapsulée montre la distribution des différences avec les estimations Fluxnet. Droite : Emplacement des stations FLUXnet utilisées pour la validation : les points verts indiquent une amélioration pour INT et CAL par rapport à SW, les points magenta indiquent une amélioration uniquement pour CAL, et les points bleus uniquement pour INT. Les points rouges sont ceux où aucune amélioration n'est observée.	154
V.12	Même Figure que V.11 pour l'évapotranspiration.	155
V.13	Flux annuels moyens ($km^3.an^{-1}$) du cycle de l'eau en Méditerranée et incertitudes associées dans SW (petite police) et INT (grande police) au cours de la période 2004-2009.	156
V.14	Tendance linéaire mensuelle des composantes du flux du cycle de l'eau estimée via CAL. De haut en bas : précipitations, évapo(transpi)ration, ruissellement et divergence de l'humidité. Les lignes zéro sont tracées pour faciliter la lecture. Les tendances des précipitations et d'évapo(transpi)ration de l'ERA-I sont également représentées à des fins de comparaison.	161
V.15	Tendance annuelle des précipitations dans CAL sur la période 1980-2012. La tendance avec une signifiante statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.	162
V.16	Tendance annuelle de l'évaporation (à droite) et de l'évapotranspiration (à gauche) dans CAL sur la période 1980-2012. La tendance avec une signifiante statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.	163
V.17	Tendance annuelle de la divergence de l'humidité dans CAL sur la période 1980-2012 à droite. Comme la signification de la tendance varie selon la valeur moyenne (positive ou négative) de la divergence, cette dernière est représentée à gauche. La tendance avec une signification statistique de $p=0,1$ (basée sur le test de Mann-Kendall) est indiquée par des points noirs.	163
V.18	Estimation de biais des produits satellites décrivant les différentes composantes du cycle de l'eau. La moyenne du produits INT est pris comme référence pour chacune des variables.	164

V.19	Même que Figure que V.18 mais pour les erreurs aléatoires (incertitudes) des produits satellites.	166
VI.1	Cartographie des tensions et conflits liées aux pénurie d'eau (ONU). . .	172
B.1	Ordinogramme de la méthode CCM et CCI d'après Munier et Aires (2017).	182
B.2	(a) Graphique de NDVI (multiplié par 100) avec les précipitations nettes (P-E) pour les bassins considérés dans cette étude. Chaque bassin est représenté par son numéro sur la carte. Les lignes rouges représentent les limites entre les classes numérotées en rouge. (b) Carte de l'indice pixelaire (CIC) . Les limites des bassins sont représentées par des lignes rouges et les stations de mesures par des cercles verts. Munier et Aires (2017)	183
B.3	Illustration du calcul des classes avec la méthode k-means à partir des trois variables choisies : Température <i>Temp</i> , humidité du sol <i>SM</i> et <i>P-E</i> . 184	184
B.4	6 classes définies dans le CIC en utilisant l'humidité du sol, la température et P-E pour janvier (gauche) et juillet (droite).	184
B.5	RMS de résidus du bilan hydrique de surface (gauche) et du bilan hydrique atmosphérique (droite) pour les différentes méthodes développées dans ce travail de recherche.	185
B.6	RMS des résidus du bilan hydrique de surface (gauche) et atmosphérique (droite) en utilisant la méthodologie CIC avec un nombre croissant de classes (3 tirages pour montrer la variabilité de la calibration en fonction du nombre de classes définies).	186

Liste des tableaux

I.1	Quelques exemples de satellites/instruments utilisés pour l'étude du cycle de l'eau.	3
II.1	Caractéristiques instrumentales des sondeurs hyperspectraux infrarouge AIRS, IASI, CrIS, IASI-NG. FTS : "Fourier Transforms Sounder". LWIR : "Long Wave InfraRed", MWIR : "Middle Wave InfraRed", SWIR : "Short Wave InfraRed".	18
II.2	Caractéristiques instrumentales des sondeurs micro-ondes SSMIS, AMSU/MHS, ATMS, MWS. La quatrième ligne indique la résolution spatiale au nadir pour les canaux de la raie d'absorption de l' O_2 à 57 GHz (T) et de H_2O à 183 GHz (q).	20
II.3	Notations pour la réduction de dimension des données hyperspectrales.	23
II.4	Description des six bandes d'absorption associées aux 251 canaux simulés. Pour chaque bande, la résolution spectrale est indiquée avec le nombre de canaux et le bruit instrumental choisi.	40
II.5	Résumé des variables géophysiques relatives à la mesure de l'IASI	51
II.6	Statistiques de restitution(s) de la température T et l'humidité q , pour les ER, BCs-constraints, NAPCA et COL (Collard, 2007), lorsqu'on utilise 270 composants/canaux. La moyenne, l'écart-type et l'erreur quadratique moyenne ("Root Mean Square Error" RMS, en anglais) sont fournis.	62
II.7	Comparaison des propriétés des techniques de compression, BC et de sélection de canaux.	65
III.1	Aperçu des différents produits utilisés dans cette étude. Leur période de couverture temporelle commune, sur laquelle le bilan du cycle de l'eau est estimé, est 2004-2009.	69
IV.1	RMS des résidus du bilan hydrique continental ou océanique (partie supérieure) et atmosphérique (partie inférieure) sur les sous-bassins méditerranéens en utilisant la climatologie saisonnière de six produits de précipitations et la climatologie de référence pour les autres composantes du cycle de l'eau. Voir le texte pour la notion de climatologie de référence. Les résidus ont été calculés à l'échelle mensuelle. Les meilleurs résultats en terme de fermeture du cycle de l'eau sont indiqués en gras.	106

IV.2	Statistiques sur les résidus mensuels des bilans hydriques continentaux, océaniques et atmosphériques, calculées sur les sous-bassins méditerranéens pour la période 2004-2009, et pour les deux méthodes : OS et SW. Les résidus des bilans hydriques pour la ré-analyse de ERA-I sont également montrés.	109
V.1	Les trois principales initiatives en matière de contrainte de fermeture du cycle de l'eau ("Water Cycle" WC en anglais) et leurs différences. Dans la troisième colonne, les caractères en gras indiquent les nouvelles caractéristiques de la méthodologie présentée dans ce travail de thèse. Les indices sont : <i>l</i> pour la terre, <i>o</i> pour l'océan, les deux incluent l'atmosphère au-dessus de leurs surfaces. Toutes les notations sont résumées dans la Table V.2.	133
V.2	Notations utilisées dans la section	136
V.3	Incertitudes pour les composantes continentales du cycle de l'eau, en termes d'écart-type d'erreur mensuelle et de pourcentage d'incertitude totale sur le résidu du bilan hydrique. Les pondérations associées à une variable (calculées comme le rapport entre la variabilité de la variable particulière et celle du résidu du bilan hydrique $\sigma_i^2 / \sum_i \sigma^2$) sont exprimées en pourcentage.	137
V.4	RMS des résidus du bilan hydrique mensuel(en <i>mm</i>) sur les sous-bassins en utilisant la solution OS, SW et SW+PF, pour la période 2004-2009. Le pourcentage d'amélioration du RMS des résidus, par rapport à la solution SW, des méthodes d'intégration est également indiqué. À des fins de comparaison, les résultats obtenus avec ERA-I sont également représentés.	144
V.5	Principales caractéristiques des cinq méthodes développées dans cette thèse : EO signifie "Earth Observation" (satellite d'observation de la Terre), * ne tient pas compte de la période de GRACE. La dernière colonne montre la capacité de la méthodologie à fermer le bilan hydrique. Le terme "- -" signifie une très mauvaise fermeture, "-" signifie une mauvaise fermeture, "+" signifie une bonne fermeture et "++" signifie une très bonne fermeture.	149
V.6	RMS des résidus du bilan hydrique (en <i>mm.mois⁻¹</i>) calculés sur le sous-bassin en utilisant les solution OS, SW, SW+PF, INT et CAL et pour la période 2004-2009. Le pourcentage d'amélioration du RMS des résidus de la solution SW aux méthodes contraintes est également indiqué. À des fins de comparaison, les résultats obtenus à l'aide de l'ensemble de données ERA-I sont également représentés.	150
V.7	Comparaison des solutions de précipitation SW, SW+PF, INT et CAL avec l'ensemble de données EOBS, en termes de corrélation, de RMSD (en <i>mm.mois⁻¹</i>) et de pourcentage d'amélioration de la RMSD par rapport à la solution SW. Ces scores sont calculés sur chaque sous bassin continental et la région entière couverte par le produit EOBS (dernière colonne).	152

V.8	Comparaison des flux moyens annuels et de leur variabilité associée (en $mm.an^{-1}$) pour la mer Méditerranée (sans la mer Noire) dans la littérature. La variabilité des produits est ici l'écart-type des valeurs annuelles, sauf pour l'incertitude associée à la moyenne des modèles climatiques régionaux qui est l'écart entre les modèles. La période d'analyse des différentes études est rappelée.	157
B.1	K-mean classes characteristics	183

Bibliographie

- Adler, R. F., G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, et E. Nelkin
2003. The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979Present). *J. Hydrometeorol.*, 4(6) :1147–1167. 69, 71, 73, 117
- Aires, F.
2014. Combining Datasets of Satellite-Retrieved Products. Part I : Methodology et Water Budget Closure. *J. Hydrometeorol.*, 15(4) :1677–1691. 105, 117, 124, 126, 132, 133, 134, 135, 137, 139, 144
- Aires, F., O. Aznay, C. Prigent, M. Paul, et F. Bernardo
2012. Synergetic multi-wavelength remote sensing versus a posteriori combination of retrieved products : Application for the retrieval of atmospheric profiles using MetOp measurements. *J. Geophys. Res.*, 117(D18304).
- Aires, F., M. Paul, C. Prigent, B. Rommen, et M. Bouvet
2011b. Measure et exploitation of multisensor et multiwavelength synergy for remote sensing : 2. Application to the retrieval of atmospheric temperature et water vapor from MetOp. *J. Geophys. Res. Atmos.*, 116(D2) :D02302. 31
- Aires, F., V. Pellet, C. Prigent, et J.-L. Moncet
2015. Dimension reduction of satellite observations for remote sensing, Part I : A comparison of compression, channel selection, et bottleneck channel approaches. *qjrms*, Pp. 1–14. 11, 34, 40, 41
- Aires, F., W. B. Rossow, N. A. Scott, et A. Chédin
2002c. Remote sensing from the infrared atmospheric sounding interferometer instrument 2. Simultaneous retrieval of temperature, water vapor, et ozone atmospheric profiles. *J. Geophys. Res. Atmos.*, 107(D22) :ACH 7—1—ACH 7—12. 22, 51
- Allen, R. G., L. S. Pereira, D. Raes, M. Smith, et W. Ab
1998. Crop evapotranspiration - Guideline for computing crop water requirements. *FAO Irrig. Drain. Pap.*, Pp. 1–15. 78, 140
- Antonelli, P., H. E. Revercomb, L. A. Sromovsky, W. L. Smith, R. O. Knuteson, D. C. Tobin, R. K. Garcia, H. B. Howell, H. L. Huang, et F. A. Best
2004. A principal component noise filter for high spectral resolution infrared measurements. *J. Geophys. Res.*, 109 :D23102—23124.
- Arkin, P. A. et B. N. Meisner
1987. The Relationship between Large-Scale Convective Rainfall and Cold Cloud

- over the Western Hemisphere during 1982-84. *Mon. Weather Rev.*, 115(1) :51–74. 71, 73
- Ashouri, H., K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, O. P. Prat, H. Ashouri, K.-L. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, et O. P. Prat
2015. PERSIANN-CDR : Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bull. Am. Meteorol. Soc.*, 96(1) :69–83. 69, 72
- Atkinson, N. C., F. I. Hilton, S. M. Illingworth, J. R. Eyre, et T. Hultberg
2010. Potential for the use of reconstructed IASI radiances in the detection of atmospheric trace gases. *Atmos. Meas. Tech.*, 3(4) :991–1003. 22
- Atlas, R., R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, D. Gombos, R. Atlas, R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, et D. Gombos
2011. A Cross-calibrated, Multiplatform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications. *Bull. Am. Meteorol. Soc.*, 92(2) :157–174. 76
- Ba, M. B. et A. Gruber
2001. GOES Multispectral Rainfall Algorithm (GMSRA). 71
- Bahurel, P., M. Ocean, P. coordinator PierreBahurel, M.-o. MBell, et P. Le Traon
2012. Ocean Monitoring and Forecasting in the European MyOcean GMES Marine Initiative. 84
- Beck, H. E., A. I. J. M. van Dijk, V. Levizzani, J. Schellekens, D. G. Miralles, B. Martens, et A. de Roo
2017. MSWEP : 3-hourly 0.25deg ; global gridded precipitation by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.*, 21(1) :589–615. 79
- Bernstein, L., P. Bosch, O. Canziani, Z. Chen, R. Christ, O. Davidson, W. Hare, S. Huq, D. Karoly, V. Kattsov, Z. Kundzewicz, J. Liu, U. Lohmann, M. Manning, T. Matsuno, B. Menne, B. Metz, M. Mirza, N. Nicholls, L. Nurse, R. Pachauri, J. Palutikof, M. Parry, D. Qin, N. Ravindranath, A. Reisinger, J. Ren, K. Riahi, C. Rosenzweig, M. Rusticucci, S. Schneider, Y. Sokona, S. Solomon, P. Stott, R. Stouffer, T. Sugiyama, R. Swart, D. Tirpak, C. Vogel, G. Yohe, T. Barker, A. Allali, R. Bojariu, S. Diaz, I. Elgizouli, D. Griggs, D. Hawkins, O. Hohmeyer, B. Pateh Jallow, K. Kajfez, N. Leary, H. Lee, et D. Wratt
2014. Core Writing Team Climate Change 2014 : Synthesis Report Synthesis Report An Assessment of the Intergovernmental Panel on Climate Change. 97
- Bettadpur, S.
2012. GRACE 327-742 (CSR-GR-12) GRAVITY RECOVERY AND CLIMATE EXPERIMENT UTCSR Level-2 Processing Standards Document. 69, 81
- Biancale, R., J.-M. Lemoine, S. Bruinsma, et S. Bourgoigne
2005. 10-day time series of the geoid from GRACE and LAGEOS data Evaluation of solutions Processing standards Mean field and 10-day solutions Noise estimation over global oceans. 69, 81

- Blackwell, W. J.
2005. A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data. *IEEE Trans. Geosci. Remote Sens.*, 43(11) :2535–2546. 25
- Bonan, B.
2014. *Assimilation de données pour l'initialisation et l'estimation de paramètres d'un modèle d'évolution de calotte polaire*. PhD thesis. 120, 128, 131, 194
- Bosilovich, M. G., J. Chen, F. R. Robertson, R. F. Adler, M. G. Bosilovich, J. Chen, F. R. Robertson, et R. F. Adler
2008. Evaluation of Global Precipitation in Reanalyses. *J. Appl. Meteorol. Climatol.*, 47(9) :2279–2299. 77
- Brown, P. J. et C. D. Kummerow
2013. An Assessment of Atmospheric Water Budget Components over Tropical Oceans. *J. cli*, 27. 160
- Brunke, M. A., X. Zeng, et S. Anderson
2002. Uncertainties in sea surface turbulent flux algorithms and data sets. *J. Geophys. Res.*, 107(C10) :3141. 75
- Buehler, S. A., N. Courcoux, et V. O. John
2006. Radiative transfer calculations for a passive microwave satellite sensor : Comparing a fast model and a line-by-line model. *J Geophys Res-Atmos*, 111(D20304).
- Buehler, S. A., P. Eriksson, T. Kuhn, A. von Engel, et C. Verdes
2005. ARTS, the atmospheric radiative transfer simulator. *J. Quant. Spectrosc. Radiat. Transf.*, 91(1) :65–93.
- Cayla, F., B. Tournier, et P. Hebert
1995. Performance budgets of IASI options. 51
- Chachine, M. T., T. S. Pagano, H. Aumann, R. Atlas, C. Barnet, et al.
2006. AIRS : Improving Weather Forecasting and Providing New Data on Greenhouse Gases. *Bull. Am. Meteorol. Soc.*, 87(7) :911–926. 18
- Chedin, A., N. A. Scott, C. Wahiche, P. Moulinier, A. Chedin, N. A. Scott, C. Wahiche, et P. Moulinier
1985. The Improved Initialization Inversion Method : A High Resolution Physical Method for Temperature Retrievals from Satellites of the TIROS-N Series. *J. Clim. Appl. Meteorol.*, 24(2) :128–143. 52
- Chevallier, F., F. Chéruy, N. A. Scott, A. Chédin, F. Chevallier, F. Chéruy, N. A. Scott, et A. Chédin
1998. A Neural Network Approach for a Fast and Accurate Computation of a Long-wave Radiative Budget. *J. Appl. Meteorol.*, 37(11) :1385–1397. 52
- Chou, S.-H., R. M. Atlas, C.-L. Shie, J. Ardizzone, S.-H. Chou, R. M. Atlas, C.-L. Shie, et J. Ardizzone
1995. Estimates of Surface Humidity and Latent Heat Fluxes over Oceans from SSM/I Data. *Mon. Weather Rev.*, 123(8) :2405–2425. 76

- Clayson, C. A., A. S. Bogdanoff, C. A. Clayson, et A. S. Bogdanoff
2013. The Effect of Diurnal Sea Surface Temperature Warming on Climatological AirSea Fluxes. *J. Clim.*, 26(8) :2546–2556. 77
- Clerbaux, C., P. F. Coheur, D. Hurtmans, et S. Turquety
2009. Measurements of total and tropospheric ozone from IASI : comparison with correlative satellite, ground-based and ozonesonde observations. *Atmos Chem . . .* 51
- Collard, A., B. Candy, et F. Hilton
2010. The importance of sounding instruments. *Weather*, 65(8) :227. 4
- Collard, A. D.
2007. Selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 133(629) :1977–1991. 6, 22, 26, 30, 31, 34, 50, 51, 60, 61, 62, 191, 197
- Collard, A. D. et A. P. McNally
2009. Sounding interferometer radiances at ECMWF. *qjrms*, 135 :1044–1058. 22
- Crevoisier, C., C. Clerbaux, V. Guidard, T. Phulpin, R. Armante, B. Barret, C. Camy-Peyret, J.-P. Chaboureau, P.-F. Coheur, L. Crépeau, G. Dufour, L. Labonnote, L. Lavanant, J. Hadji-Lazaro, H. Herbin, N. Jacquinet-Husson, S. Payan, E. Péquignot, C. Pierangelo, P. Sellitto, et C. Stubenrauch
2014. Towards IASI-New Generation (IASI-NG) : impact of improved spectral resolution and radiometric noise on the retrieval of thermodynamic, chemistry and climate variables. *Atmos. Meas. Tech.*, 7(12) :4367–4385. 4, 8, 18
- Cruzado, A.
1985. *Chemistry of Mediterranean waters*. 99
- Curry, J. A., A. Bentamy, M. A. Bourassa, D. Bourras, E. F. Bradley, M. Brunke, S. Castro, S. H. Chou, C. A. Clayson, W. J. Emery, L. Eymard, C. W. Fairall, M. Kubota, B. Lin, W. Perrie, R. A. Reeder, I. A. Renfrew, W. B. Rossow, J. Schulz, S. R. Smith, P. J. Webster, G. A. Wick, X. Zeng, J. A. Curry, A. Bentamy, M. A. Bourassa, D. Bourras, E. F. Bradley, M. Brunke, S. Castro, S. H. Chou, C. A. Clayson, W. J. Emery, L. Eymard, C. W. Fairall, M. Kubota, B. Lin, W. Perrie, R. A. Reeder, I. A. Renfrew, W. B. Rossow, J. Schulz, S. R. Smith, P. J. Webster, G. A. Wick, et X. Zeng
2004. SEAFLUX. *Bull. Am. Meteorol. Soc.*, 85(3) :409–424. 69, 74
- Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, et K.-H. Neumayer
2013. GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005. 69, 81
- Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kállberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, et F. Vitart
2011. The ERA-Interim reanalysis : configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, 137(656) :553–597. 8, 69, 89, 90

- Dimet, F.-X. L. et O. Talagrand
1986. Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A*, 38A(2) :97–110. 132
- Drobinski, P., V. Ducrocq, P. Arbogast, C. Basdevant, S. Bastin, L. Beguery, S. Belamari, K. Béranger, O. Bock, M.-N. Bouin, B. Boudevillain, O. Bousquet, C. Bouvier, I. Braud, J.-C. Calvet, C. Champollion, a. Chanzy, M. Claude, P. Cocquerez, L. Coppola, S. Coquillat, D. Courault, E. Defer, J. Delanoë, G. Delrieu, a. Doerenbecher, a. Dorfliger, M. Estèves, C. Estournel, C. Flamant, N. Fourrié, O. Garrouste, E. Gaume, H. Giordani, L. Gomes, G. Roberts, H. Jourde, L. Labatut, D. Lambert, J. L. Coz, W. Ludwig, C. Lutoff, J.-F. Mahfouf, E. Martin, L. Mastrorillo, N. Mathys, Y. Michel, M. Nuret, F. Orain, S. Perez, C. Prigent, O. Radakovitch, K. Ramage, E. Richard, J.-L. Roujeau, O. Roussot, I. Ruin, F. Saïd, a. Schwarzenboeck, K. Sellegri, S. Somot, I. Taupier-Letage, P. Testor, J. V. Baelen, et J.-P. Vandervaere
2013. HyMeX, le cycle de l'eau méditerranéen à la loupe. *La Météorologie*, 8(80) :23. 98
- English, S., A. McNally, N. Borman, K. Salonen, M. Matricardi, A. Horányi, M. Rennie, M. Janisková, S. Di Michele, A. Geer, E. Di Tomaso, C. Cardinali, P. de Rosnay, J. Sabater, M. Bonavita, C. Albergel, R. Engelen, et J.-N. Thépaut
2013. Impact of Satellite Data. *Tech. Memorandum ECMWF*, (October) :46. 6, 89, 193
- Eriksson, P., S. A. Buehler, C. P. Davis, C. Emde, et O. Lemke
2011. ARTS, the atmospheric radiative transfer simulator, Version 2. *J. Quant. Spectrosc. Radiat. Transf.*, 112(10) :1551–1558.
- Evensen, G.
1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. 128
- Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, J. B. Edson, C. W. Fairall, E. F. Bradley, J. E. Hare, A. A. Grachev, et J. B. Edson
2003. Bulk Parameterization of AirSea Fluxes : Updates and Verification for the COARE Algorithm. *J. Clim.*, 16(4) :571–591. 75
- Falge, E., M. Aubinet, P. Bakwin, D. Baldocchi, P. Berbigier, et al.
2017. FLUXNET Research Network Site Characteristics, Investigators, and Bibliography, 2016. 69, 87
- Fenoglio-Marc, L., J. Kusche, et M. Becker
2006. Mass variation in the Mediterranean Sea from GRACE and its validation by altimetry, steric and hydrologic fields. *Geophys. Res. Lett.*, 33(19) :1–5. 84
- Good, S. A., M. J. Martin, et N. A. Rayner
2013. EN4 : Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Ocean.*, 118(12) :6704–6716. 84
- Guo, H., S. Chen, A. Bao, J. Hu, A. Gebregiorgis, X. Xue, et X. Zhang
2015. Inter-Comparison of High-Resolution Satellite Precipitation Products over Central Asia. *Remote Sens.*, 7(12) :7181–7211. 102

- Hamon, M., J. Beuvier, S. Somot, J.-M. Lellouche, E. Greiner, G. Jordà, M.-N. Bouin, T. Arsouze, K. Béranger, F. Sevault, C. Dubois, M. Drevillon, et Y. Drillet
2016. Design and validation of MEDRYS, a Mediterranean Sea reanalysis over the period 1992-2013. *Ocean Sci.*, 12(2) :577–599. 84
- Han, Y., H. Revercomb, M. Crompton, D. Gu, D. Johnson, D. Mooney, D. Scott, L. Strow, G. Bingham, L. Borg, Y. Chen, D. Deslovers, M. Esplin, D. Hagan, X. Jin, R. Knuteson, H. Motteler, J. Predina, L. Suwinski, J. Taylor, D. Tobin, D. Tremblay, C. Wang, L. Wang, L. Wang, et V. Zavyalov
2013. Suomi NPP CrIS measurements, sensor data record algorithm, calibration and validation activities, and record data quality. *J. Geophys. Res. Atmos.*, 118(12) :734–742. 18
- Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, et M. New
2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *J. Geophys. Res.*, 113. 69, 87, 193
- Hesselbjerg Christensen, J. et Meteorologisk Institut
1996. *The HIRHAM4 regional atmospheric climate model*. Copenhagen : DMI. 157
- Hilton, F., R. Armante, T. August, C. Barnet, A. Bouchard, C. Camy-Peyret, V. Capelle, L. Clarisse, C. Clerbaux, P.-F. Coheur, A. Collard, C. Crevoisier, G. Dufour, D. Edwards, F. Faijan, N. Fourrié, A. Gambacorta, M. Goldberg, V. Guidard, D. Hurtmans, S. Illingworth, N. Jacquinet-Husson, T. Kerzenmacher, D. Klaes, L. Lavanant, G. Masiero, M. Matricardi, A. McNally, S. Newman, E. Pavelin, S. Payan, E. Péquignot, S. Peyridieu, T. Phulpin, J. Remedios, P. Schlüssel, C. Serio, L. Strow, C. Stubenrauch, J. Taylor, D. Tobin, W. Wolf, et D. Zhou
2012. Hyperspectral Earth Observation from IASI : Five Years of Accomplishments. <http://dx.doi.org/10.1175/BAMS-D-11-00027.1>. 4, 8, 18
- Hirschi, M., P. Viterbo, et S. I. Seneviratne
2006. Basin-scale water-balance estimates of terrestrial water storage variations from ECMWF operational forecast analysis. *Geophys. Res. Lett.*, 33(21) :L21401. 109
- Holm, E. V. et T. Kral
2012. Flow-dependent, geographically varying background error covariance for 1D-var applications in MTG-IRS L2 processing. Technical report.
- Hsu, K.-I., X. Gao, S. Sorooshian, H. V. Gupta, K.-I. Hsu, X. Gao, S. Sorooshian, et H. V. Gupta
1997. Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteorol.*, 36(9) :1176–1190. 73
- Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, J. Susskind, G. J. Huffman, R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, et J. Susskind
2001. Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. *J. Hydrometeorol.*, 2(1) :36–50. 77

- Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, et E. F. Stocker
2007. The TRMM Multisatellite Precipitation Analysis (TMPA) : Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.*, 8(1) :38–55. 69, 70, 71, 73
- Ide, K., P. Courtier, M. Ghil, et A. C. Lorenc
1997. Unified Notation for Data Assimilation : Operational, Sequential and Variational (gtSpecial Issue>Data Assimilation in Meteorology and Oceanography : Theory and Practice). *J. Meteorol. Soc. Japan. Ser. II*, 75(1B) :181–189.
- Ishii, M., M. Kimoto, et M. Kachi
2003. Historical Ocean Subsurface Temperature Analysis with Error Estimates. *Mon. Weather Rev.*, 131(1) :51–73. 84
- Jordà, G. et D. Gomis
2006. On the interpretation of the steric and mass components of sea level variability : The case of the Mediterranean basin. *J. Geophys. Res. Ocean.*, 118 :953–963. 84
- Jordà, G., A. Sánchez-Román, and D. Gomis
2016. Reconstruction of transports through the Strait of Gibraltar from limited observations. *Clim. Dyn.* 69, 86, 106, 108, 158, 160
- Jordà, G., K. Von Schuckmann, S. A. Josey, G. Caniaux, J. García-Lafuente, S. Sammartino, E. Özsoy, J. Polcher, G. Notarstefano, P. M. Poulain, F. Adloff, J. Salat, C. Naranjo, K. Schroeder, J. Chiggiato, G. Sannino, and D. Macías
2017. The Mediterranean Sea heat and mass budgets : Estimates, uncertainties and perspectives. *Prog. Oceanogr.*, 156(July) :174–208. 86, 115, 156, 157, 158, 159, 160, 192, 193
- Joyce, R. J., J. E. Janowiak, P. A. Arkin, et P. Xie
2004. CMORPH : A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. 69, 71, 74
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, et D. Joseph
1996a. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, 77(3) :437–471. 76
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins,

- J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, et D. Joseph
1996b. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, 77(3) :437–471. 80
- Köppen, W.
1936. Das geographische System der Klimate. *Handb. der Klimatologie*, (c) :7–30. 94, 147, 182
- Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, T. T. Wilheit, C. Kummerow, Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D.-B. Shin, et T. T. Wilheit
2001. The Evolution of the Goddard Profiling Algorithm (GPROF) for Rainfall Estimation from Passive Microwave Sensors. *J. Appl. Meteorol.*, 40(11) :1801–1820. 73
- Kummerow, C. D., D. L. Randel, M. Kulie, N. Y. Wang, R. Ferraro, S. Joseph Munchak, et V. Petkovic
2015. The evolution of the goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Ocean. Technol.*, 32(12) :2265–2280.
- Kummerow, C. D., S. Ringerud, J. Crook, D. Randel, et W. Berg
2011. An observationally generated a priori database for microwave rainfall retrievals. *J. Atmos. Ocean. Technol.*, 28(2) :113–130.
- Lacombe, H. et P. Tchernia
1972. Caractères hydrologiques et circulation des eaux en Méditerranée. *Mediterr. Sea A Nat. Sediment. Lab.*, Pp. 25–36. 157
- L'Ecuyer, T. S., H. K. Beaudoin, M. Rodell, W. Olson, B. Lin, S. Kato, C. A. Clayson, E. Wood, J. Sheffield, R. Adler, G. Huffman, M. Bosilovich, G. Gu, F. Robertson, P. R. Houser, D. Chambers, J. S. Famiglietti, E. Fetzer, W. T. Liu, X. Gao, C. A. Schlosser, E. Clark, D. P. Lettenmaier, K. Hilburn, T. S. L'Ecuyer, H. K. Beaudoin, M. Rodell, W. Olson, B. Lin, S. Kato, C. A. Clayson, E. Wood, J. Sheffield, R. Adler, G. Huffman, M. Bosilovich, G. Gu, F. Robertson, P. R. Houser, D. Chambers, J. S. Famiglietti, E. Fetzer, W. T. Liu, X. Gao, C. A. Schlosser, E. Clark, D. P. Lettenmaier, et K. Hilburn
2015. The Observed State of the Energy Budget in the Early Twenty-First Century. *J. Clim.*, 28(21) :8319–8346. 123, 133
- Levizzani, V., S. Laviola, et E. Cattani
2011. Detection and measurement of snowfall from space. *Remote Sens.*, 3(1) :145–166. 102
- Li, L. Z. X.
2006. Atmospheric GCM response to an idealized anomaly of the Mediterranean sea surface temperature. *Clim. Dyn.*, 27(5) :543–552. 114
- Liang, X., D. P. Lettenmaier, E. F. Wood, et S. J. Burges
1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, 99(D7) :14415. 7, 129

- Lionello, P., P. Malanotte-Rizzoli, et R. R. Boscolo
2006. *Mediterranean climate variability*. Elsevier. 97, 112
- Lipton, A. E.
2003. Satellite sounding channel optimization in the microwave spectrum. *IEEE Trans. Geosci. Remote Sens.*, 41(4) :761–781. 34
- Lipton, A. E., J. Moncet, S. A. Boukabara, G. Uymin, et K. J. Quinn
2009. Fast and Accurate Radiative Transfer in the Microwave With Optimum Spectral Sampling. *IEEE Trans. Geosci. Remote Sens.*, 47(7) :1909–1917. 52
- Liu, W. T., K. B. Katsaros, et J. A. Businger
1979. Bulk Parameterization of Air-Sea Exchanges of Heat and Water Vapor Including the Molecular Constraints at the Interface. *J. Atmos. Sci.*, 36(9) :1722–1735. 74, 75, 77
- Llasat, M. C., R. Marcos, M. Turco, J. Gilabert, et M. Llasat-Botija
2016. Trends in flash flood events versus convective precipitation in the Mediterranean region : The case of Catalonia. *J. Hydrol.*, 541 :24–37.
- Longuevergne, L., B. R. Scanlon, et C. R. Wilson
2010. GRACE Hydrological estimates for small basins : Evaluating processing approaches on the High Plains Aquifer. *Water Resour. Res.*, 46. 104
- Ludwig, W., E. Dumont, M. Meybeck, et S. Heussner
2009. River discharges of water and nutrients to the Mediterranean and Black Sea : Major drivers for ecosystem changes during past and future decades ? *Prog. Oceanogr.*, 80(3-4) :199–217. 69, 85, 94, 99, 102
- Luus, K. A., Y. Gel, J. C. Lin, R. E. J. Kelly, et C. R. Duguay
2013. Pan-Arctic linkages between snow accumulation and growing-season air temperature, soil moisture and vegetation. *Biogeosciences*, 10 :7575–7597. 79
- Maheras, P., K. Tolika, C. Anagnostopoulou, M. Vafiadis, I. Patrikas, et H. Flocas
2004. On the relationships between circulation types and changes in rainfall variability in Greece. *Int. J. Climatol.*, 24(13) :1695–1712.
- Mahfouf, J.-F., C. Birman, F. Aires, C. Prigent, E. Orlandi, et M. Milz
2014. Information content on temperature and water vapour from an hyper-spectral microwave sensor. *Q. J. Roy. Meteor. Soc.*, Pp. 1–33.
- Margat, J.
2004. Programme des Nations Unies pour l'environnement plan d'action pour la Méditerranée : situation et perspectives MAP Technical Report Series No . 158 PNUE / PAM. (158). 94, 95, 96, 97, 193
- Mariotti, A.
2010. Recent changes in the mediterranean water cycle : A pathway toward long-term regional hydroclimatic change ? *J. Clim.*, 23(6) :1513–1525. 98, 162
- Mariotti, A., M. V. Struglia, N. Zeng, et K.-M. Lau
2002. The Hydrological Cycle in the Mediterranean Region and Implications for the

- Water Budget of the Mediterranean Sea. *J. Clim.*, 15(13) :1674–1690. 7, 94, 98, 157, 158, 159, 160, 167
- Mariotti, A., N. Zeng, J.-H. Yoon, V. Artale, A. Navarra, P. Alpert, et L. Z. X. Li
2008. Mediterranean water cycle changes : transition to drier 21st century conditions in observations and CMIP3 simulations. *Environ. Res. Lett.*, 3(4) :044001. 99
- Martens, B., D. G. Miralles, H. Lievens, R. van der Schalie, R. A. M. de Jeu, D. Fernández-Prieto, H. E. Beck, W. A. Dorigo, et N. E. C. Verhoest
2016. GLEAM v3 : satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev. Discuss.*, Pp. 1–36. 69, 79, 111
- Matricardi, M.
2010. A principal component based version of the RTTOV fast radiative transfer model. *Q. J. R. Meteorol. Soc.*, 136(652) :1823–1835.
- MED-HYCOS, p. .-.
2001. The Mediterranean hydrological cycle observing system. Technical Report 0. 97
- Melsheimer, C., C. Verdes, S. A. Buehler, C. Emde, P. Eriksson, D. G. Feist, S. Ichizawa, V. O. John, Y. Kasai, G. Kopp, N. Koulev, T. Kuhn, O. Lemke, S. Ochiai, F. Schreier, T. R. Sreerekha, M. Suzuki, C. Takahashi, S. Tsujimaru, et J. Urban
2005. Intercomparison of General Purpose Clear Sky Atmospheric Radiative Transfer Models for the Millimeter/Submillimeter Spectral Range. *Radio Sci.*, 40(1).
- Miralles, D. G., R. A. M. De Jeu, J. H. Gash, T. R. H. Holmes, et A. J. Dolman
2011. Magnitude and variability of land evaporation and its components at the global scale. *Hydrol. Earth Syst. Sci.*, 15 :967–981. 79
- Moncet, J., G. Uymin, et A. Lipton
2008. *OSS Radiative Transfer Method Performance in Real Time Atmosphere Characterization from Satellite Sounding and Imaging Data*, volume 3. IEEE. 34
- Moncet, J.-L., G. Uymin, A. E. Lipton, et H. E. Snell
2010. Infrared Radiance Modeling by Optimal Spectral Sampling. <http://dx.doi.org/10.1175/2008JAS2711.1>. 34, 52
- Mu, Q., M. Zhao, and S. W. Running
2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.*, 115(8) :1781–1800. 69, 80
- Mueller, B., M. Hirschi, et S. I. Seneviratne
2011. New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data. *Hydrol. Process.*, 25(7) :996–1008. 99
- Munier, S. et F. Aires
2017. A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water cycle budget. *Remote Sens. Environ.*, in revisio. 9, 133, 134, 135, 145, 181, 182, 183, 186, 196

- Munier, S., F. Aires, S. Schläffer, C. Prigent, F. Papa, P. Maisongrande, et M. Pan
2014. Combining data sets of satellite-retrieved products for basin-scale water balance study : 2. Evaluation on the Mississippi Basin and closure correction model. *J. Geophys. Res.*, 119(21) :12,100–12,116. 117, 133, 134, 135, 137, 139, 144, 181
- Overpeck, J. T., G. A. Meehl, S. Bony, et D. R. Easterling
2011. Climate Data Challenges in the 21st Century. *Science (80-.)*, 331(6018) :700–702. 5, 189
- Pan, M., A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, Wood, et E. F
2012. Multisource estimation of long-term terrestrial water budget for major global river basins. *J. Clim.*, 25(9) :3191–3206. 133, 135
- Pan, M. et E. F. Wood
2006. Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. *J. Hydrometeorol.*, 7(3) :534–547. 7, 117, 126, 129, 130, 132, 133, 134, 137, 139, 167
- Paul, M.
2013. *Synergie infrarouge et micro-onde pour la restitution atmosphérique*. PhD thesis. 11, 15, 16, 17, 189
- Paul, M. et F. Aires
2014. Using Shannon's entropy to sample heterogeneous and high-dimensional atmospheric datasets. *Q. J. R. Meteorol. Soc.* 38
- Pellet, V. et F. Aires
2018. Bottleneck Channels algorithm for satellite data dimension reduction : A case study for IASI. *IEEE T. Geosci. Remote Sens.* 11, 50
- Pellet, V., F. Aires, A. Mariotti, et D. Fernandez
2018. Analyzing the Mediterranean water cycle via satellite data integration. *Pure Appl. Geophys.* 93, 98, 119, 144
- Pellet, V., F. Aires, S. Munier, G. Jordá, D. F. Prieto, A. Dorigo, J. Polcher, et L. Brocca
2018b. Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle - Application to the Mediterranean region. *HESS Rev.*, Pp. 1–38. 98, 119
- Pellet, V., F. Aires, et C. Prigent
2015. Dimension reduction of satellite observations for remote sensing, Part II : A comparison of compression, channel selection, and bottleneck channel approaches. *qjrms*, Pp. 1–11. 11, 34, 38
- Philandras, C. M., P. T. Nastos, J. Kapsomenakis, K. C. Douvis, G. Tselioudis, et C. S. Zerefos
2011. Long term precipitation trends and variability within the Mediterranean region. *Nat. Hazards Earth Syst. Sci.*, 11(12) :3235–3250. 94
- Phulpin T., F. Rabier
2003. Le sondage satellitaire et son évolution. *Nat. Hazards Earth Syst. Sci. La Météorologie*, 80–85. 19, 20, 190

- Pike, J.
1964. The estimation of annual run-off from meteorological data in a tropical climate. *J. Hydrol.*, 2(2) :116–123. 85
- Plummer, D. A., D. Caya, A. Frigon, H. Côté, M. Giguère, D. Paquin, S. Biner, R. Harvey, R. de Elia, D. A. Plummer, D. Caya, A. Frigon, H. Côté, M. Giguère, D. Paquin, S. Biner, R. Harvey, et R. de Elia
2006. Climate and Climate Change over North America as Simulated by the Canadian RCM. *J. Clim.*, 19(13) :3112–3132. 157
- Polcher, J., B. Mcavaney, P. Viterbo, M.-A. Gaertner, A. Hahmann, J.-F. Mahfouf, J. Noilhan, T. Phillips, A. Pitman, C. A. Schlosser, J.-P. Schulz, B. Timbal, D. Verseghy, et Y. Xue
1998. A proposal for a general interface between land surface schemes and general circulation models. *Glob. Planet. Change*, 19 :261–276. 69, 85
- Priestley, C. and R. Taylor
1972. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Mon. Weather Rev.*, 100(2) :81–92. 79
- Rabier, F., N. F. 'e, D. Chafa, et P. Prunet
2002. Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteorol. Soc.*, 128 :1011–1027. 33
- Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, M. G. Schlax, R. W. Reynolds, T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, et M. G. Schlax
2007. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *J. Clim.*, 20(22) :5473–5496. 76
- Roberts, J. B., C. A. Clayson, F. R. Robertson, et D. L. Jackson
2010. Predicting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach. *J. Geophys. Res.*, 115(D19) :D19113. 77
- Rodell, M., H. Beaudoin, T. L'Ecuyer, W. Olson, J. Famiglietti, P. Houser, R. Adler, M. Bosilovich, C. Clayson, D. Chambers, E. Clark, E. Fetzer, X. Gao, G. Gu, K. Hilburn, G. Huffman, D. Lettenmaier, W. Liu, F. Robertson, C. Schlosser, J. Sheffield, et E. Wood
2015. The Observed State of the Water Cycle in the Early 21st Century. *J. Clim.*, Pp. 8289–8318. 106, 116, 117, 123, 124, 132, 133, 134, 135, 137, 157, 158, 159, 160, 167
- Rodgers, C. D.
1996a. Information content and optimisation of high spectral resolution measurements. In *Proc. SPIE*, P. B. Hays et J. Wang, eds., Pp. 136–147. Univ. of Oxford, United Kingdom, SPIE. 33, 34, 58
- Rodgers, C. D.
2000. Inverse Methods for Atmospheric Sounding - Theory and Practice. *Inverse Methods Atmos. Sound. - Theory Pract. Ser. Ser. Atmos. Ocean. Planet. Phys.*, 2. 27, 28, 122, 123, 167, 194

- Sahoo, A. K., M. Pan, T. J. Troy, R. K. Vinukollu, J. Sheffield, et E. F. Wood
2011. Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sens. Environ.*, 115(8) :1850–1865. 133, 135, 137
- Sanchez-Gomez, E., S. Somot, S. A. Josey, C. Dubois, N. Elguindi, et M. Déqué
2011. Evaluation of Mediterranean Sea water and heat budgets simulated by an ensemble of high resolution regional climate models. *Clim. Dyn.*, 37(9-10) :2067–2086. 98, 157, 158, 159, 160, 167
- Sapiano, M. R. P., P. A. Arkin, M. R. P. Sapiano, et P. A. Arkin
2009. An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data. *J. Hydrometeorol.*, 10(1) :149–166. 71
- Sasaki, B. Y.
1955. A Fundamental Study of the Numerical Prediction based on Study of the Numerical Prediction the Variational Principle. 130
- Scannell, C., D. Hurtmans, et A. Boynard
2012. Antarctic ozone hole as observed by IASI/MetOp for 2008–2010. *Atmos.*
- Schneider, N., M. Schröder, R. Lindstrot, R. Preusker, M. Stengel, et E. D. G. ESA DUE GlobVapour Consortium
2013. ESA DUE GlobVapour water vapor products : Validation. volume 1531, Pp. 484–487. 69, 83
- Seager, R. et N. Henderson
2013. Diagnostic computation of moisture budgets in the ERA-interim reanalysis with reference to analysis of CMIP-archived atmospheric model data. *J. Clim.*, 26(20) :7876–7901. 89, 136
- Senay, G. B., K. Asante, et G. Artan
2009. Water balance dynamics in the Nile Basin. *Hydrol. Process.*, 23(26) :n/a–n/a. 98
- Seneviratne, S. I., D. Lüthi, M. Litschi, et C. Schär
2006. Landatmosphere coupling and climate change in Europe. *Nature*, 443(7108) :205–209. 114
- Sheffield, J., C. R. Ferguson, T. J. Troy, E. F. Wood, et M. F. McCabe
2009. Closing the terrestrial water budget from satellite remote sensing. *Geophys. Res. Lett.*, 36(7) :1–5. 9, 133
- Simpson, J., C. Kummerow, W. K. Tao, et R. F. Adler
1996. On the Tropical Rainfall Measuring Mission (TRMM). *Meteorol. Atmos. Phys.*, 60(1-3) :19–36. 72
- Soto-Navarro, J., F. Criado-Aldeanueva, J. García-Lafuente, et A. Sánchez-Román
2010. Estimation of the Atlantic inflow through the Strait of Gibraltar from climatological and in situ data. *J. Geophys. Res.*, 115(C10) :C10023. 114
- Stopa, J. E. et K. F. Cheung
2014. Intercomparison of wind and wave data from the ECMWF Reanalysis Interim

- and the NCEP Climate Forecast System Reanalysis. *Ocean Model.*, 75 :65–83. 89, 160
- Struglia, M. V., A. Mariotti, et A. Filograsso
2004. River discharge into the Mediterranean sea : Climatology and aspects of the observed variability. *J. Clim.*, 17(24) :4740–4751. 99
- Sun, B., L. Yu, et R. A. Weller
2003. Comparisons of Surface Meteorology and Turbulent Heat Fluxes over the Atlantic : NWP Model Analyses versus Moored Buoy Observations*. *J. Clim.*, 16(4) :679–695. 69, 74
- Sun, Q., C. Miao, Q. Duan, H. Ashouri, S. Sorooshian, et K.-L. Hsu
2018. A Review of Global Precipitation Data Sets : Data Sources, Estimation, and Intercomparisons. *Rev. Geophys.*, 56(1) :79–107. 4, 70, 189, 192
- Trenberth, K. E. et J. T. Fasullo
2013. Regional Energy and Water Cycles : Transports from Ocean to Land. *J. Clim.*, 26. 116, 117
- Trenberth, K. E., J. T. Fasullo, et J. Mackaro
2011. Atmospheric Moisture Transports from Ocean to Land and Global Energy Flows in Reanalyses. *J. Clim.*, 24. 7, 106
- Trenberth, K. E., L. Smith, T. Qian, A. Dai, et J. Fasullo
2007. Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data. *J. Hydrometeorol.*, 8. 106
- Tucker, C. J., J. E. P. Molly, E. Brown, D. A. Slayback, E. W. Pak, R. Mahoney, et E. F. Vermote
2005. An extended AVHRR 8-kni NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *Int. J. Remote Sens.*, 26(20) :4485–98. 80
- Turk, J. T., G. V. Mostovoy, et V. Anantharaj
2010. The NRL-Blend High Resolution Precipitation Product and its Application to Land Surface Hydrology. In *Satell. Rainfall Appl. Surf. Hydrol.*, Pp. 85–104. Dordrecht : Springer Netherlands. 72
- Uppala, S. M., P. W. Killberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, et J. Woollen
2005. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, 131(612) :2961–3012. 76
- Ventress, L. et A. Dudhia
2014. Improving the selection of IASI channels for use in numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 140(684) :2111–2118. 31

- Vincensini, A.
2013. *Contribution du sondeur IASI à l'estimation des paramètres des surface continentales pour la Prévision Numerique du Temps*. PhD thesis. 11
- Watkins, M. M. et D.-N. Yuan
2014. GRACE Gravity Recovery and Climate Experiment JPL Level-2 Processing Standards Document For Level-2 Product Release 05.1. 69, 81
- Wu, H., J. S. Kimball, N. Mantua, et J. Stanford
2011. Automated upscaling of river networks for macroscale hydrological modeling. *Water Resour. Res.*, 47(3) :n/a–n/a. 99, 100, 193
- Yilmaz, M. T., T. DeSole, et P. R. Houser
2011. Improving Land Data Assimilation Performance with a Water Budget Constraint. *J. Hydrometeorol.*, 12(5) :1040–1055. 129, 130, 137, 139
- Zhang, K., J. S. Kimball, R. R. Nemani, et S. W. Running
2010. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resour. Res.*, 46(9) :n/a–n/a. 69, 80
- Zhang, Y., M. Pan, J. Sheffield, A. Siemann, C. Fisher, M. Liang, H. Beck, N. Wanders, R. MacCracken, P. R. Houser, T. Zhou, D. P. Lettenmaier, Y. Ma, R. T. Pinker, J. Bytheway, C. D. Kummerow, et E. F. Wood
2017. A Climate Data Record (CDR) for the global terrestrial water budget : 1984-2010. *Hydrol. Earth Syst. Sci. Discuss.*, (April) :1–40. 133, 137
- Zhang, Y., M. Pan, et E. F. Wood
2016. On Creating Global Gridded Terrestrial Water Budget Estimates from Satellite Remote Sensing. *Surv. Geophys.*, 37(2) :1–20. 135

**Sujet : Télédétection satellite du cycle de l'eau -
De l'optimisation du traitement des observations à
l'intégration des produits restitués**

Résumé : L'observation satellite offre l'opportunité unique d'une couverture globale et d'un suivi temporel continu en météorologie et climatologie. On rencontre toutefois de nombreuses difficultés dans leur exploitation, ce qui requiert l'utilisation de techniques d'optimisation dans les diverses étapes de leur traitement. On traitera dans cette thèse deux questions : (1) au niveau du traitement des données satellites, en amont de la chaîne de restitution, et (2) en aval, pour l'intégration des multiples produits satellites disponibles. On propose ainsi une nouvelle approche de compression/sélection de canaux (les "Bottlenecks Channels") pour réduire la dimension des données hyper-spectrales (dans l'infrarouge avec l'instrument IASI et dans les micro-ondes pour un nouveau concept de sondeur). On développe également un nouveau schéma d'intégration de multiples produits satellites pour l'étude du cycle de l'eau (surface, océan et atmosphère). Cette approche d'intégration permet de rendre plus cohérentes les composantes du cycle de l'eau. Elle est présentée ici dans le contexte du cycle de l'eau méditerranéen.

Mots clés : Télédétection, cycle de l'eau, observations satellites, optimisation, réduction de dimension, intégration

**Subject : Satellite remote sensing of the water cycle - From
the optimization of the observations to the integration of
the retrieved products**

Abstract : Satellite observation offers the unique opportunity for global coverage and continuous temporal monitoring in meteorology and climatology. However, there are many difficulties in their exploitation, which requires the use of optimization techniques at the different steps of their processing. This thesis will deal with two questions : (1) the processing of satellite data before the restitution, and (2) the integration of the multiple satellite products available. A new approach for channel compression/selection (Bottlenecks Channels) is proposed to reduce the dimension of hyperspectral data (in the infrared with the IASI instrument and in microwaves for a new sounder concept). A new scheme is also developed to integrate multiple satellite products for the study of the water cycle (surface, ocean and atmosphere). This integration approach makes the components of the water cycle more coherent to each other. It is presented here in the context of the Mediterranean water cycle.

Keywords : Remote sensing, satellite observations, optimization, dimension reduction, integration, water cycle