

Restitution de l'eau liquide nuageuse

Chapitre V

V.1 Introduction

Ce chapitre est consacré au développement et à la validation d'un modèle d'inversion d'eau liquide nuageuse. Après avoir décrit les différentes bases de données synthétiques, nous présentons maintenant la mise en œuvre du réseau Multi-Expert (ME) décrit dans le chapitre 2. Il sera référencé dans la suite par « ME_Iclw » (ME Integreated Cloud Liquid Water). Dans un premier temps, nous décrirons le modèle ME_Iclw (§ V.2), son architecture, et plus particulièrement la séparation effectuée par l'algorithme Multi-Expert. Dans un deuxième temps, nous comparerons les performances de ce modèle à ceux d'un réseau de neurones classique (PMC) et d'une régression « log-linéaire » (§ V.3). Pour l'eau liquide nuageuse la validation est délicate par manque de données in situ en quantité suffisante. Nous effectuerons des comparaisons avec des données SSM/I sur plusieurs cas d'études entre le modèle ME_Iclw et les algorithmes de la littérature (§ V.4). Nous nous attacherons à comparer, les performances des algorithmes dans la détection des situations sans nuage puis la vraisemblance des structures nuageuses à l'aide d'images infrarouge.

V.2 Description du modèle ME.

V.2.1 Architecture

La précision du modèle neuronal dépend de son architecture, du nombre de données utilisées pour la détermination de ses poids et de la qualité des données utilisées pour son

apprentissage. Dans de nombreux cas de figure, plusieurs architectures neuronales permettent l'obtention de résultats similaires et il faut donc simplement déterminer une architecture suffisamment complexe pour bien représenter le problème. En règle générale, on choisit l'architecture la plus simple. La base d'apprentissage est très importante, elle permet de bien représenter statistiquement la relation entre les températures de brillance mesurées et la quantité d'eau liquide nuageuse qui lui est associée.

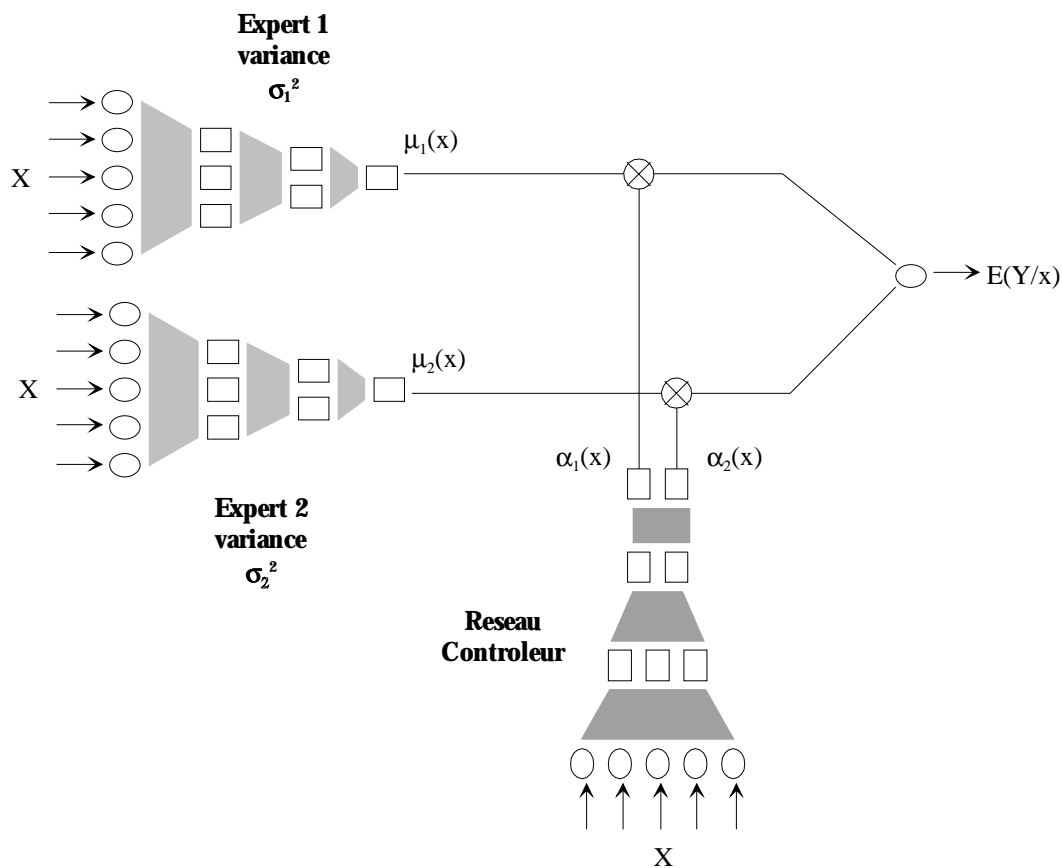


Figure V-1. Architecture du réseau ME_Iclw. X représente le vecteur d'entrée, composé des 5 températures de brillance (Tb_{19v} , Tb_{19h} , Tb_{22v} , Tb_{37v} et Tb_{37h}). μ_1 et μ_2 sont les sorties des experts 1 et 2. α_1 et α_2 sont les sorties du réseau contrôleur et sont appelées coefficients de mélanges. La sortie globale du ME notée

$$E(Y/x) = \sum_{k=1}^2 \alpha_k(x) \mu_k(x) \text{ est équivalent à l'espérance mathématique de } Y \text{ sachant } X.$$

L'ensemble des données que nous avons utilisées pour l'étalonnage du modèle ME_Iclw est la base SMSP_app (Section précédente). Les paramètres d'apprentissages sont décrits dans le Tableau V-1. De manière pratique, différentes architectures ont été testées. Dans un premier temps, nous avons fait varier le nombre d'experts, afin de déterminer le nombre de

régimes que pouvait détecter le modèle ME. Quelle que soit l'architecture des experts et du contrôleur testée, seulement deux régimes ont été détectés. A partir de cette constatation, différentes architectures d'experts et de contrôleurs ont été testées, de façon à déterminer l'architecture optimale associée à la base SMSP_app. Nous avons fait varier le nombre de couches cachées (entre 1 et 2) et le nombre de neurones dans les couches cachées (entre 2 et 6) pour les deux experts et le réseau contrôleur. L'architecture optimale est définie comme l'architecture la plus simple qui minimise en mieux l'erreur sur la base de test.

Nombre d'itérations	5x10 ⁶
Poids initiaux	Aléatoire entre [-1.0, +1.0]
Normalisation des entrées	[-1.0, +1.0]
Pas d'apprentissage initial (ϵ)	0.01

Tableau V-1. Paramètres d'apprentissage.

Le modèle ME_Iclw se compose donc d'un réseau de type ME estimant l'eau liquide nuageuse. Ce réseau est composé de 2 réseaux experts et d'un réseau contrôleur.

La couche d'entrée des Experts et du Contrôleur se compose de 5 neurones correspondant à Tb_{19v} , Tb_{19h} , Tb_{22v} , Tb_{37v} et Tb_{37h} . Les deux experts ont une architecture similaire, composée de deux couches cachées et d'une couche de sortie comportant un neurone unique (de fonction de transition linéaire) estimant la quantité d'eau liquide nuageuse. Les deux couches cachées comportent respectivement 3 et 2 neurones, de fonction transitoire sigmoïde. Le réseau contrôleur est composé de deux couches cachées comportant respectivement 3 et 2 neurones et d'une couche de sortie comportant deux neurones, un pour chaque expert. La Figure V-1 présente l'architecture du réseau ME_Iclw. Le réseau possède ainsi 71 poids qui ont été estimés lors de la phase d'apprentissage.

V.2.2 Spécialisation des experts

V.2.2.1 Description

Le réseau ME a donc finalement détecté 2 régimes différents. Chaque expert c'est spécialisé dans une région de l'espace d'entrée. Les valeurs des coefficients de mélange déterminent le domaine de validité de chaque expert. Le coefficient de mélange α_i peut être compris comme la probabilité à posteriori de l'expert i sachant X . Le domaine de validité de l'expert i correspond au domaine où le coefficient de mélange α_i est égal à 1. Le domaine de validité de l'expert 1 est $[0, 0.3]$ kg/m² et celui de l'expert 2 est $[0.5, 2]$ kg/m². Dans chacun de ces domaines, la sortie globale du réseau ($\alpha_1\mu_1 + \alpha_2\mu_2$) est identique à l'une des sorties des deux experts.

Les Figure V-3 et Figure V-4 présentent les performances de chaque expert sur la base d'apprentissage. La répartition quasiment binaire des coefficients de mélange montre la très bonne séparation des experts effectuée par le réseau contrôleur. Les figures de dispersion des sorties de chacun des experts montrent que les données sont très bien distribuées le long de la bissectrice pour chacun des domaines de validation.

Pour l'expert 1, par exemple, on constate que dès que le contrôleur indique une probabilité 1 ($\alpha_1 = 1$), la sortie de l'expert 1 (μ_1) correspond bien à la valeur d'eau liquide désiré, notée $Iclw_0$. Inversement, lorsque que la sortie de l'expert 1 n'est pas correcte ($\mu_1 \neq Idw_0$), cet expert n'est pas utilisé ($\alpha_1 = 0$). L'expert 2 se comporte de manière similaire. L'expert 2 donne des valeurs d'Iclw correctes lorsque $Iclw < 0.3$ kg/m²; le contrôleur lui attribue alors une probabilité de 1. Par contre, l'expert 2 donne de très mauvais résultats lorsque $Iclw > 0.5$ kg/m² mais n'est pas utilisé, car la probabilité α_2 est alors nulle.

Il existe une zone intermédiaire où $\alpha_1 (= 1 - \alpha_2)$ est différent de 1. Cette zone correspond à des valeurs d'Iclw comprises entre 0.3 et 0.5 kg/m². La sortie globale du réseau est alors égale à la somme des sorties de chaque expert pondérées par leurs coefficients de mélange. La quantité moyenne ainsi calculée a un sens si, et seulement si, les sorties de chaque expert sont « proches », ce qui est le cas ici.

La Figure V-2 présente la dispersion des I_{clw} estimés sur la base d'apprentissage en fonction des valeurs désirées (I_{clw_0}). Le coefficient de corrélation R^2 , égal à 0.99, indique un excellent accord entre le modèle et les données.

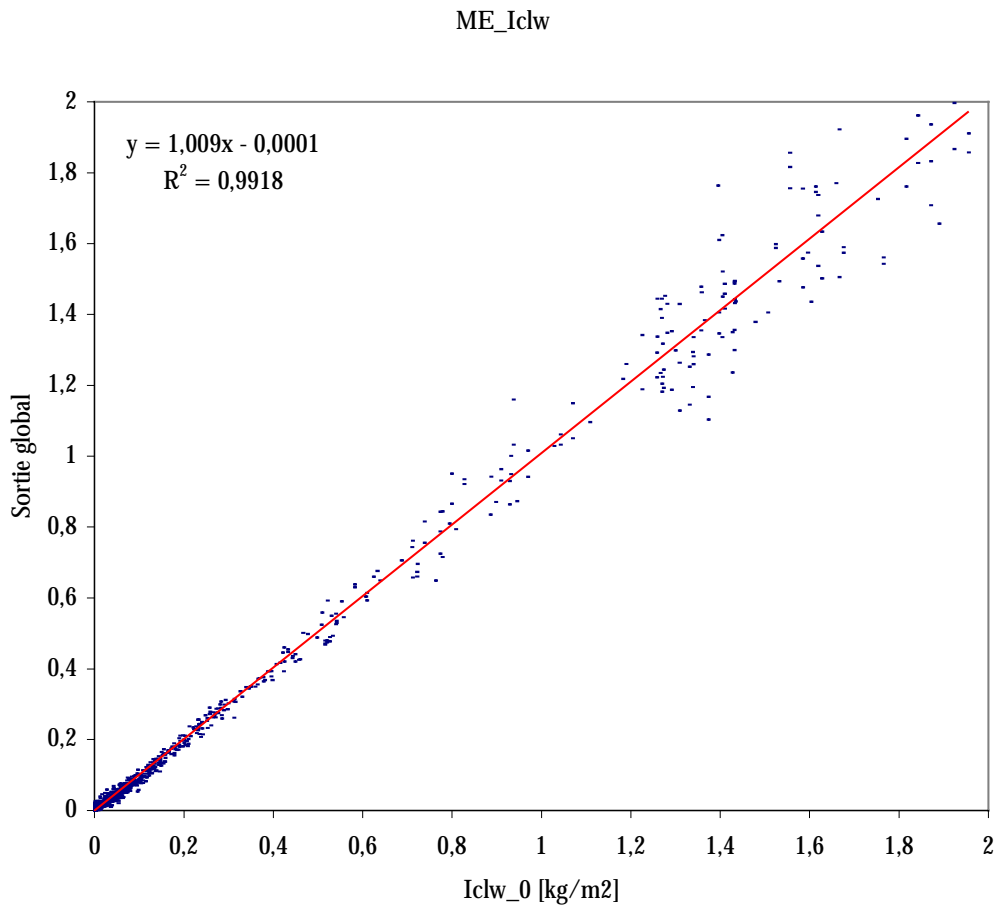


Figure V-2. Valeurs d'Iclw estimées par le modèle ME_Iclw en fonction des valeurs « vraies » d'Iclw sur l'ensemble d'apprentissage. Unité en kg/m².

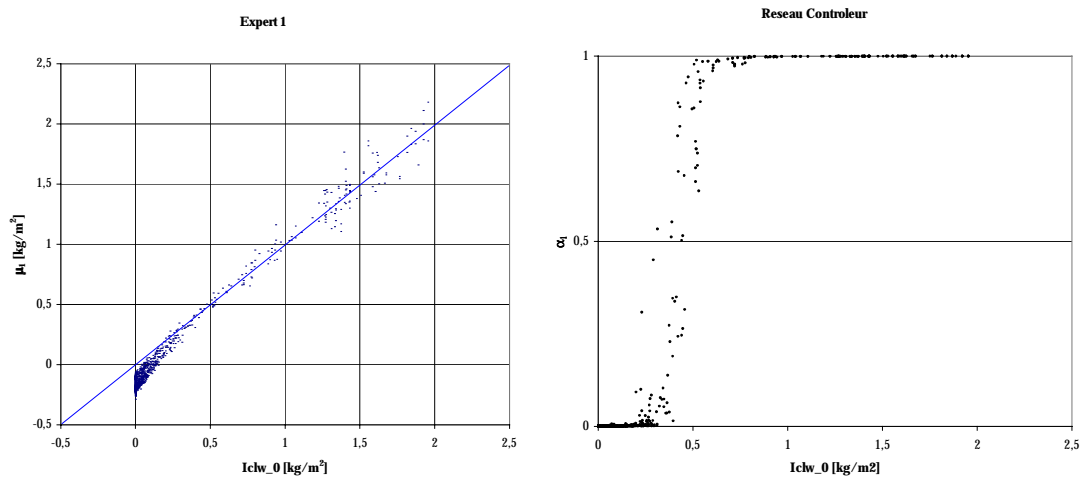


Figure V-3. (A gauche) Sortie de l'expert 1 (μ_1) en fonction de la quantité d'eau liquide nuageuse désirée. (A droite) Valeur du coefficient de mélange dédié à l'expert 1 (α_1) en fonction de la quantité d'eau liquide nuageuse désirée.

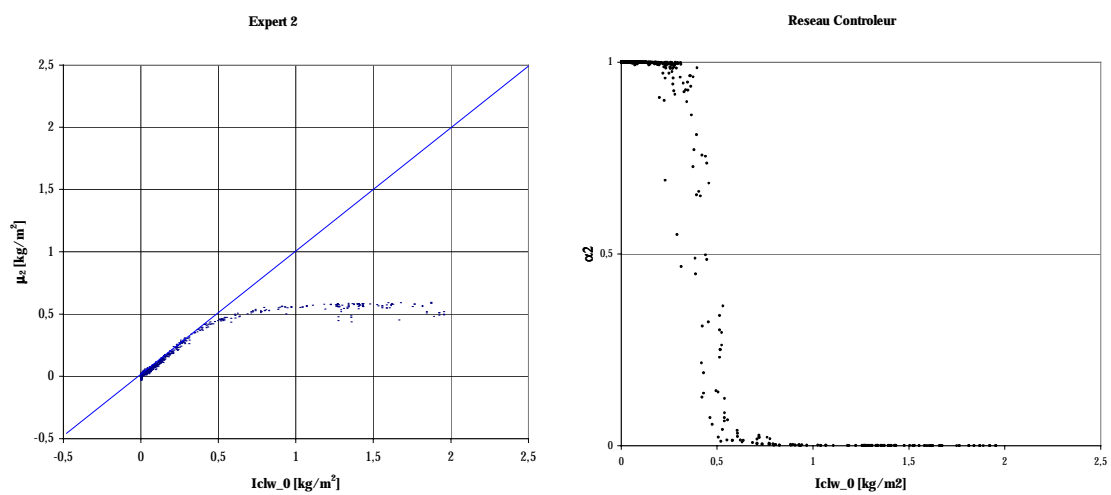


Figure V-4. Idem Figure V-3 pour l'expert 2.

V.2.2.2 Interprétation

Contrairement à ce que l'on pouvait s'attendre à trouver, nous n'avons pas de partition en ciel clair ($I_{clw}=0$) et ciel nuageux ($I_{clw}\neq 0$), mais en faible et forte valeurs d'eau liquide nuageuse.

Dans le modèle ME_Iclw, la partition est effectuée par le réseau contrôleur uniquement à partir de la connaissance des données d'entrée, soit les températures de brillance. En visualisant cette partition dans l'espace d'entrée (Figure V-5), après avoir effectué une ACP (Analyse en Composante Principale), on observe deux zones bien distinctes, correspondant chacune au domaine d'application des experts (en bleu pour l'expert 1 et en vert pour l'expert 2). La partition de l'expert 1 est caractérisée par des situations météorologiques avec de faibles quantités d' Iclw ($< 0.3 \text{ kg/m}^2$) et de larges variations de vapeur d'eau, pouvant aller de 0.5 à 7.0 kg/m^2 (Figure V-6). La partition de l'expert 2 est, quant à elle, caractérisée par de fortes quantités d'Iclw ($> 0.5 \text{ kg/m}^2$), et de plus faibles variations de vapeur d'eau (entre 2.5 et 7.0 kg/m^2). Comme le montre la Figure V-6, la variance des températures de brillance est principalement expliquée par les deux quantités atmosphériques que sont l'eau liquide nuageuse et la vapeur d'eau.

Nombre de variables	5	
Nombre de facteurs	2	
Nombre de points	2598	
	Magnitude	Variance proposée
1 ^{ère} valeur	4.651	0.930
2 ^{ème} valeur	0.320	0.064

Tableau V-2. Caractéristiques de l'ACP effectuée sur la base SMSP.

La relation entre les températures de brillance aux fréquences SSM/I (excepté à 85.5 GHz) et la quantité d'eau liquide est biunivoque lorsque l'on fixe tous les autres paramètres atmosphériques. La difficulté provient de la présence de paramètres cachés, telle que la vapeur d'eau, qui ont une influence très forte sur les températures de brillance comme le montre la Figure V-6.

La grande variation de la quantité de vapeur d'eau pour une même quantité d'eau liquide nuageuse provoque une dispersion des températures de brillance, plus ou moins importante suivant la fréquence considérée. La dispersion est plus importante à 19 GHz qu'à 37 GHz.

Quelle que soit la fréquence, la dispersion des températures de brillance est directement liée aux variations de vapeur d'eau dans l'atmosphère.

	σ [kg/m ²]	σ_{relatif} [%]
Expert 1	0.01	29.0
Expert 2	0.12	9.2

Tableau V-3. Ecart type des réseaux experts.

Autrement dit, le niveau de bruit est directement lié aux variations possibles de la quantité de vapeur d'eau. Chaque expert s'est spécialisé dans une région, caractérisée par un niveau de bruit différent. Les valeurs relatives des déviations standard des gaussiennes associées à chaque expert en est caractéristique (Tableau V-3).

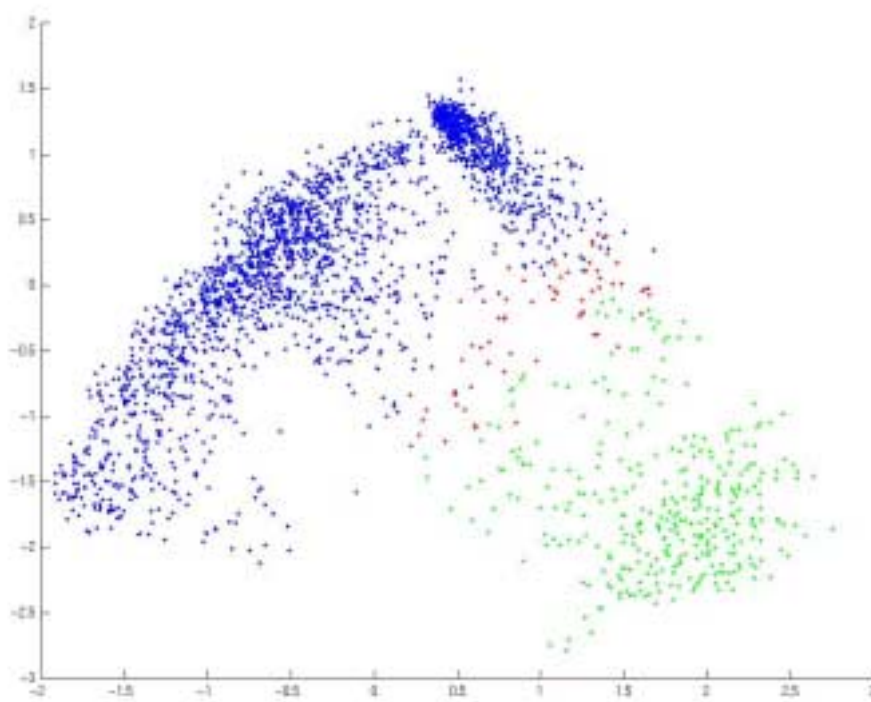


Figure V-5. Représentation de la partition du modèle ME_Icw dans l'espace d'entrée réduit après ACP. Les caractéristiques de l'ACP sont présentées dans le Tableau V-2. En bleu (resp. en vert), les données telles que $\alpha_1 > 0.9$ (resp. $\alpha_2 > 0.9$). En rouge, les données telles que $\alpha_1 \in [0.1, 0.9]$.

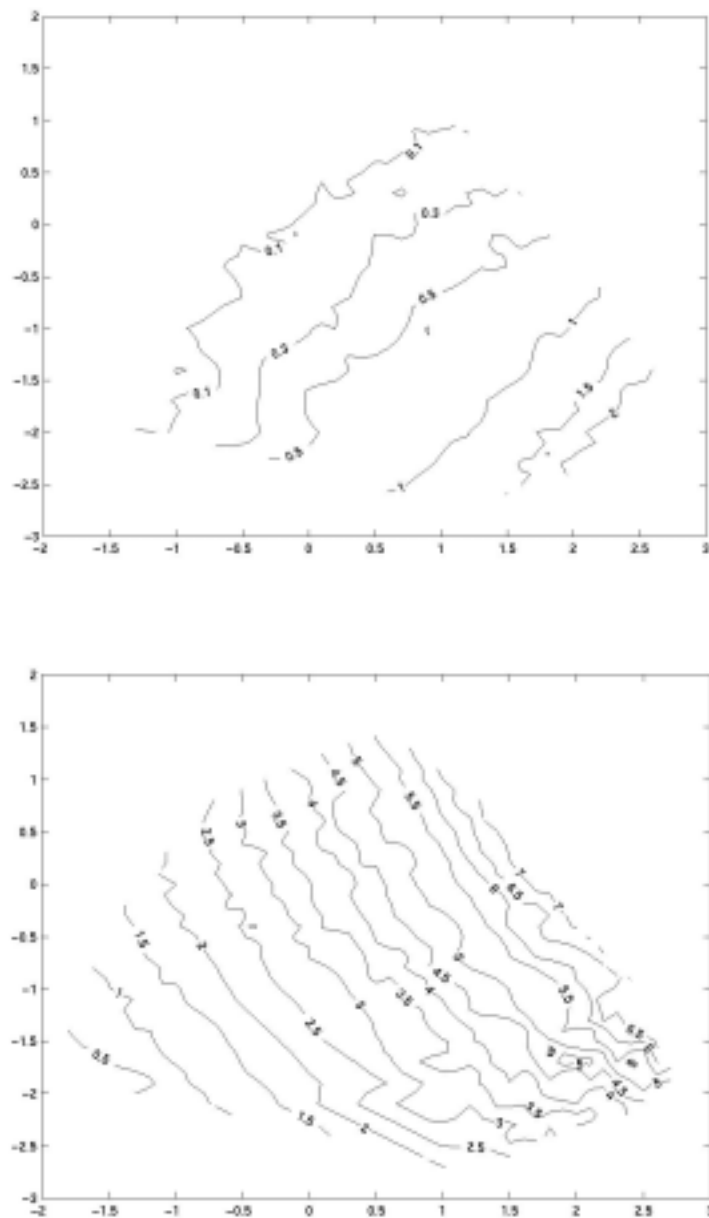


Figure V-6. Contours en fonction de l'Idw (en haut) et de la vapeur d'eau (en bas), des Tb de la base SMSP dans le repère réduit après ACP.

V.3 Comparaison avec d'autres algorithmes statistiques

L'objectif de ce paragraphe est de mettre en évidence les avantages des modèles modulaires par rapport à d'autres modèles statistiques couramment employés pour l'estimation de $Iclw$. Pour ce faire, nous avons développé, en parallèle, à partir de la base de donnée SMSP_app, deux algorithmes statistiques : un algorithme log-linéaire et un algorithme neuronal de type PMC. Ces deux algorithmes sont développés avec la même base de données pour une comparaison la plus juste possible. La forme du modèle log-linéaire est la suivante :

$$Iclw = a_0 + \sum_{i=1}^5 a_i \log(280 - Tb_i) \quad (48)$$

où Tb représente les canaux SSM/I aux 5 premières fréquences et a_i les coefficients de régression. La forme (48) utilisé par Gérard (1996), Lojou (1990), linéarise la relation $Iclw-Tb$ dans le cas d'atmosphère non diffusante. D'autres algorithmes de même forme, mais utilisant uniquement deux ou trois canaux (Grody et Ferraro, 1992) sont aussi employés. Nous avons choisi d'utiliser 5 canaux comme pour le modèle ME_Iclw. Le modèle log-linéaire sera dorénavant appelé LL_Iclw.

Différentes architectures de perceptron multicouches (PMC) ont été testées, composées de une à deux couches cachées. Le nombre de neurones d'entrée est égal à 5, comme pour le modèle modulaire, un pour chacune des 5 températures de brillance. Les meilleurs performances ont été obtenues avec une architecture composée de deux couches cachées comprenant respectivement 6 et 4 neurones sigmoïdes et un neurone linéaire dans la couche de sortie. Les données d'entrée sont normalisées et les poids du réseau sont initialisés à des valeurs aléatoires comprises entre -1 et 1. Ce modèle sera dorénavant appelé PCM_Iclw.

Pour effectuer la comparaison, les trois modèles (ME_Iclw, PMC_Iclw et LL_Iclw) sont appliqués sur une base de données complètement indépendante (notée SMSP_val). On entend ici par indépendante une base qui n'a jamais été utilisée pendant la phase d'apprentissage, que ce soit pour apprendre les poids ou pour choisir l'architecture.

Les performances en termes de biais et de déviation standard sont décrits dans le Tableau V-4.

Algorithmes	Biais [kg/m²]	Ecart type [kg/m²]
ME_Iclw	-3,36 E-04	4,40 E-02
PMC_Iclw	1,70 E-03	3,70 E-02
LL_Iclw	3,20 E-03	5,80 E-02

Tableau V-4. Performances globales des trois algorithmes statistiques sur la base SMSP_val.

Les deux algorithmes neuronaux (ME_Iclw et PMC_Iclw) ont des performances meilleures à tous points de vue, que l'algorithme LL_Iclw. En ce qui concerne les algorithmes neuronaux, leurs performances sont très semblables, le modèle ME_Iclw a un biais relativement meilleur mais un écart type un peu plus importante. La Figure V-7 présente le biais et la déviation standard des trois algorithmes par classe d'eau liquide nuageuse. La première classe (resp. la troisième) représente le domaine de validité de l'expert 1 (resp. expert 2). La deuxième classe représente la zone de partition des experts où la sortie globale du réseau modulaire est une moyenne des sorties de chaque experts.

On obtient les résultats suivants :

Le biais est constamment plus faible pour les deux algorithmes neuronaux et légèrement meilleur pour le modèle ME_Iclw.

La déviation standard des deux modèles est identique pour les deux premières classes, ce qui correspond à des valeurs de Iclw comprises entre 0 et 0.6 kg/m². Le modèle PMC_Iclw se distingue quant à lui par une déviation standard plus faible pour les fortes valeurs de Iclw.

Ces conclusions confirment celles obtenues par Jung et al. (1998) sur l'estimation de l'eau liquide nuageuse. Celui-ci montre qu'un PMC est plus performant qu'un algorithme de forme log-linéaire, tout deux utilisant en entrée les températures de brillances à 22 et 37 GHz en polarisation verticale.

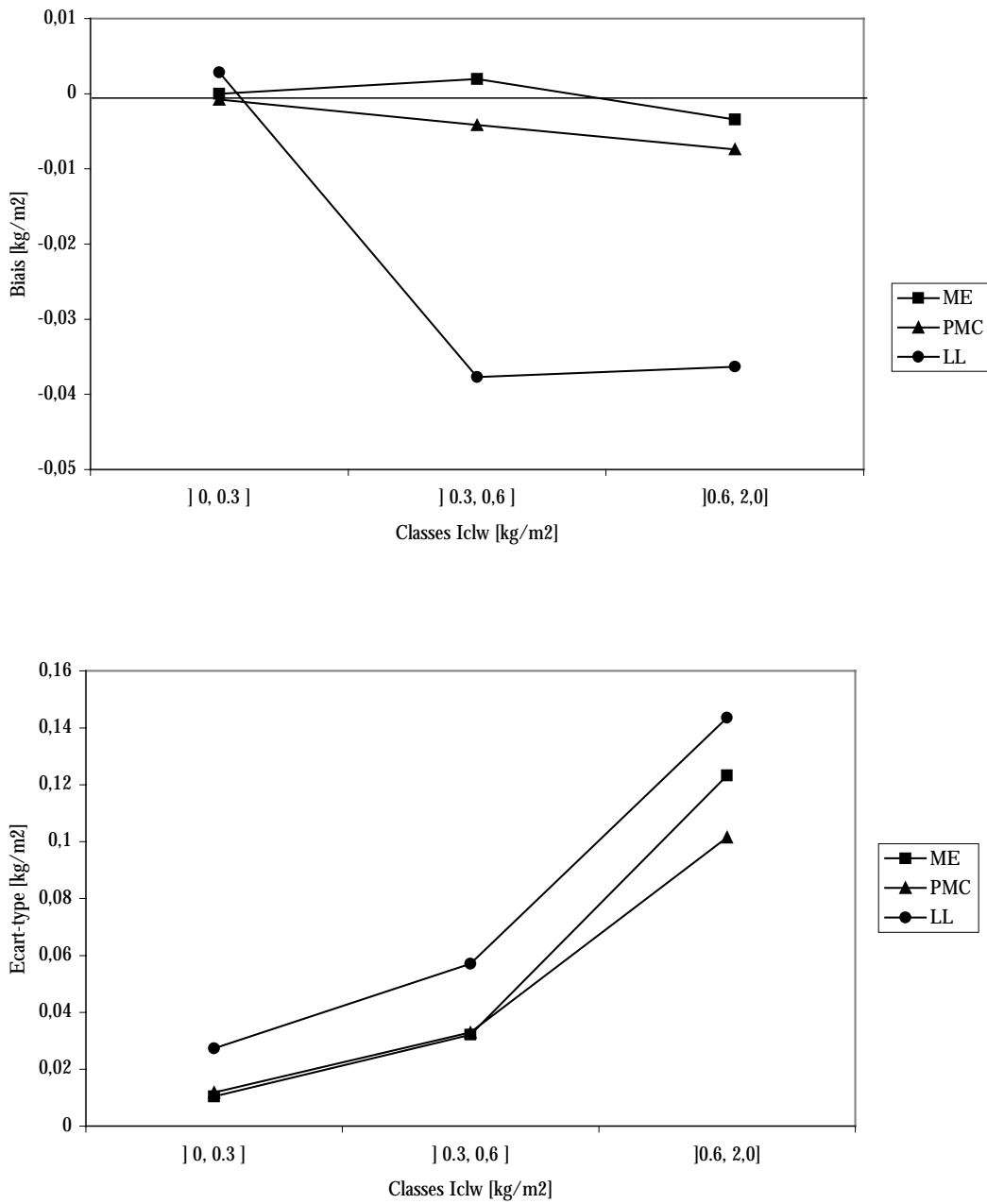


Figure V-7. Performances des trois algorithmes sur la base SMSP_val, par classes Iclw. La première (resp. la troisième) classe correspond au domaine de validité de l'expert 1 (resp. l'expert 2). En haut, le biais et en bas, l'écart-type.

Le choix du « meilleur » algorithme n'est alors pas évident et peut se faire uniquement sur des considérations de performances, d'autant plus que les erreurs sont globalement très faibles. Le problème le plus important des modèles PMC est qu'ils sont in-interprétables. En effet, les valeurs des poids synaptiques ne peuvent être directement reliées au problème physique sous-jacent. Autrement dit, comment être sûr que la fonction de transfert trouvée par le réseau est « physiquement » valable.

Les modèles modulaires ont l'avantage de pouvoir remédier partiellement à ce problème. Ils ont l'avantage indéniable, de par leur nature, de pouvoir effectuer des partitions de l'espace d'entrée sur des critères physiques (Figure V-5). Dans le cas du modèle ME_Iclw, chacune des partitions a donc une interprétation possible et peut être reliée à un ensemble de situations météorologiques. Partiellement seulement parce que chaque expert n'en reste pas moins in-interprétable, mais leurs domaines d'actions, eux, le sont.

Les réseaux modulaires fournissent les écarts types associées à chaque experts. Ce qui revient à fournir les barres d'erreurs des valeurs estimées suivant qu'elles sont issues du premier ou du deuxième expert. Dans l'état actuel des choses, le modèle ME employé dans cette étude ne permet pas de fournir des écarts types dépendantes des températures de brillance, mais seulement dépendantes des experts.

Finalement, à performance équivalente, le modèle ME_Iclw a l'avantage d'être *relativement* interprétable. C'est donc ce dernier que nous avons utilisé pour effectuer une comparaison, sur des données SSM/I, avec des modèles « classiques » issus de la littérature.

V.4 Validation

V.4.1 Introduction

Nous disposons d'un algorithme neuronal d'eau liquide nuageuse pour le radiomètre SSM/I, l'objet de l'étude qui suit est la validation de cet algorithme.

La difficulté de cette étude provient du fait qu'il n'existe pas de mesures intégrées d'eau liquide nuageuse représentatives d'une large résolution spatiale (20kmx20 km). Des mesures aéroportées d'eau liquide nuageuse sont effectuées durant des campagnes scientifiques (Semaphore par exemple); cependant, ces données, spatialement très localisées, sont difficilement utilisables pour notre étude. De façon générale, la validation d'algorithmes spatiaux d'eau liquide nuageuse ne peut se faire que qualitativement. Cependant, nous avons essayé de dégager quelques critères objectifs de validation.

Les valeurs d'eau liquide nuageuse sont comparées à celles provenant des algorithmes de la littérature, sur quatre scènes géographiques. Cette comparaison permet de mettre en évidence l'excellent comportement de notre algorithme dans la détection du ciel clair.

La cohérence des structures nuageuses restituées est contrôlée en utilisant des données infrarouges provenant de l'instrument OLS (Operational Linescan System) sur le satellite DMSP. Le cas du Février 1993 de l'expérience TOGA-COARE (Tropical Ocean Global Atmosphere Coupled Ocean Atmosphere Response Experiment) a été choisi pour cette étude.

V.4.2 Description des algorithmes de la littérature

Parmi les nombreux algorithmes de la littérature, nous en avons choisi cinq (Tableau V-5). Ils ont en commun (1) d'utiliser uniquement les températures de brillance comme paramètres d'entrée, (2) d'avoir été construits à partir d'une base de données, et (3) de pouvoir être implémenté facilement. Cependant, la nature des bases de données est très variable comme le montre le Tableau V-5.

Les trois algorithmes de Gérard (1996), Petty (1990) et Wentz(1997) ont été construits de façon semblable en utilisant une base de donnée composée de paramètres atmosphériques et de températures de brillance simulées. Seul le nombre de canaux et la forme des algorithmes sont différents. L'algorithme de Gérard est relativement proche de celui que nous avons développé dans cette étude. La méthode statistique employée est différente, mais les bases de données utilisées sont d'origine identique (Gérard, 1996); les profils atmosphériques sont issus du modèle ECMWF.

Les algorithmes de Grody et Ferraro (1992) et de Alishouse (1990) ont la particularité d'avoir été construits respectivement à partir d'une base de donnée « semi-mesurée » et « totalement mesurée ». Ils ont aussi la particularité d'utiliser peu de canaux en paramètres d'entrée.

L'algorithme de Grody et Ferraro (1992) repose sur une base de données composée de radiosondages. Si les mesures en surface ne sont pas disponibles, elles sont empruntées à la climatologie ou extrapolées à partir des valeurs des profils les plus proches du sol. Pour compléter la gamme de paramètres nécessaires au calcul du transfert radiatif, il reste à déterminer le profil d'eau liquide nuageuse. Nombreuses sont les méthodes d'estimation de l'eau liquide nuageuse à partir des profils atmosphériques : eau liquide aléatoire (Lojou, 1990), eau liquide adiabatique ou encore eau liquide empirique définie par le type de nuage, sa base et son sommet.

L'algorithme de Alishouse (1990) repose sur un vaste jeu de données composé de températures de brillance SSM/I mesurées dans plusieurs canaux et de contenus en eau liquide déduits de radiosondages, en coïncidence dans le temps et l'espace.

Auteurs des algorithmes	Type de méthode d'inversion.	Origine des Paramètres atmosphériques & Températures de brillance	Domaine d'application	Domaine de validité (kg/m ²)
Moreau <i>et al.</i>	Réseaux de neurones : Multi-Expert	Modèle météo. & modèle de TR	global	0 – 2.0
Gérard 1996	Régression « log-linéaire »	Modèle météo. & modèle de TR	global	0 – 0.8
Alishouse et al 1990	Régression linéaire.	Mesure & observation	global	0 – 0.3
Grody et Ferraro 1992	Régression « log-linéaire »	Mesure & modèle de TR	Latitudes moyennes, tropicales	0 – 0.5
Petty 1990	Régression empirique	Modèle météo. & modèle de TR	global	0 – 0.5
Wentz 1997	Régression polynomiale (2)	Modèle météo. & modèle de TR	global	0 – 0.3

Tableau V-5. Caractéristiques générales des cinq algorithmes de la littérature pour le radiomètre SSM/I et de l'algorithme ME_Icw.

Auteurs des algorithmes	Canaux utilisés				
	19V	19H	22V	37V	37H
Gérard 1996	X	X	X	X	X
Alishouse 1990		X	X	X	
Grody et Ferraro 1992	X			X	
Petty 1990	X	X	X	X	X
Wentz 1997	X	X	X	X	X

Tableau V-6. Paramètres d'entrée de cinq algorithmes pour le radiomètre SSM/I.

On remarque de plus la grande différence des domaines de validité de ces cinq algorithmes, avec des valeurs d'eau liquide maximales pouvant aller de 0.3 kg/m^2 à 0.8 kg/m^2 . L'algorithme neuronal que nous avons développé a un domaine de validité théorique plus important. Tous les algorithmes présentés ici estiment le contenu en eau liquide non précipitante. Comme nous allons le montrer ultérieurement, la distinction entre une situation fortement nuageuse et une situation faiblement pluvieuse est difficile à faire. Physiquement, cette séparation est relativement floue et, en terme de température de brillance, elle est inexistante. Physiquement, la différence est basée sur le fait que, dans les nuages non précipitant, les particules d'eau sont de faibles dimensions ($<100 \mu\text{m}$) et restent donc en suspension dans l'atmosphère, alors que dans les nuages précipitants les particules sont de dimensions supérieures. La distinction est faite en fonction de la taille des particules d'eau liquide. Reste que dans les cas peu précipitants, les lois dimensionnelles accordent une grande importance aux particules de faibles tailles. Il en résulte que les températures de brillance associées à ces deux types de situations sont très proches, au moins aux fréquences inférieures à 40 GHz.

La quantité maximale d'eau liquide estimée par les différents algorithmes est directement liée aux différentes approximations faites par les auteurs quant à la limite au-dessus de laquelle la pluie commence à être générée par le nuage. D'après Prabhakara *et al.* (1992), la présence de précipitation est identifiable lorsque la quantité d'eau liquide est supérieure à 0.2 kg/m^2 et d'après Gérard (1996), la limite supérieure est de l'ordre de 0.8 kg/m^2 .

V.4.3 Comparaison avec les algorithmes de la littérature

V.4.3.1 Détection 'air clair'

Tout d'abord, nous comparons les différents algorithmes présentés ci-dessus dans la détection des situations sans nuage dites de 'ciel clair'. Pour ce faire, nous appliquons les algorithmes sur 4 jeux de données SSM/I situées autour du Japon et autour de l'Australie (Tableau V-7, Figure V-8)

Dates	Zone géographique	
	Latitudes	Longitudes
22/02/95	15S-45S	90E 120E
07/05/95	40N 60N	140 E 160E
09/06/95	30N 50N	120E 145E
22/10/95	55S 30S	110E 130E

Tableau V-7. Cas d'études SSM/I

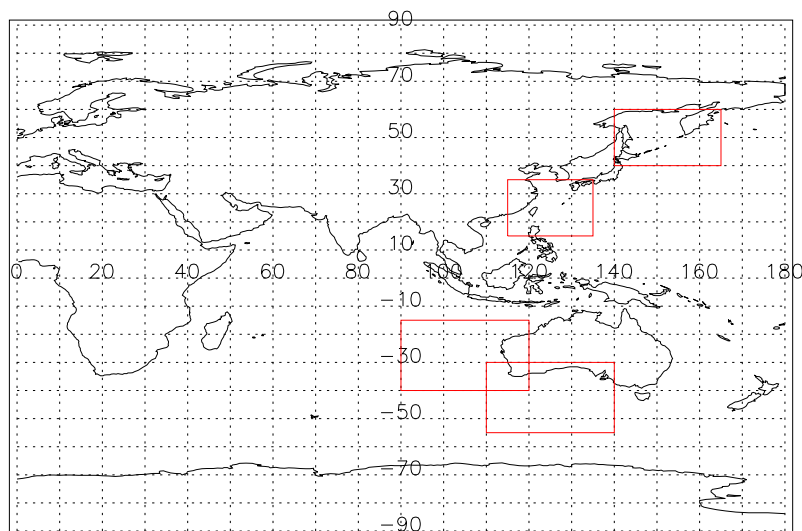


Figure V-8. Zones géographiques d'étude (en rouge).

Comme le montre la Figure V-9 représentant la quantité d'eau liquide estimée par les cinq algorithmes de la littérature par rapport à celle estimée par l'algorithme neuronal développé dans cette étude, les comportements relatifs des 5 algorithmes sont très différents. Notre algorithme a un comportement proche de celui de Wentz (1997) et de Gérard (1996). Vers les faibles contenus en eau liquide ($< 0.5 \text{ kg/m}^2$), l'algorithme ME_Iclw sous-estime

systématiquement les contenus par rapport aux algorithmes de Gérard (1996) ainsi que de Grody et Ferraro (1992) et de Petty (1990).

En ce qui concerne les fortes valeurs d'eau liquide, la comparaison paraît difficile étant donné les domaines de validité restreints des autres algorithmes. En comparant avec l'algorithme de Gérard (1996), qui a été validé jusqu'à 0.8 kg/m^2 , on observe une sous-estimation des forts contenus en eau liquide par rapport à l'algorithme ME_Iclw. Cette tendance pourrait être liée à la forme de l'algorithme de Gérard (1996). Comme nous l'avons montré dans le paragraphe (V.3), la forme « log-linéaire » est peu adaptée pour les forts contenus en eau liquide nuageuse. Cette tendance est confirmée par les travaux effectués par Gérard (1996), qui montrent que la forme « log-linéaire » provoque une sous-estimation systématique des contenus en eau liquide nuageuse supérieures à 0.6 kg/m^2 de l'ordre de 5%, par rapport à une méthode variationnelle.

Quand à l'algorithme de Alishouse (1990), la dispersion des points est telle qu'aucune conclusion en terme de tendance ne peut être avancée. L'algorithme de Petty, par rapport au notre, surestime de façon systématique la quantité d'eau liquide nuageuse et cela dans tous les cas.

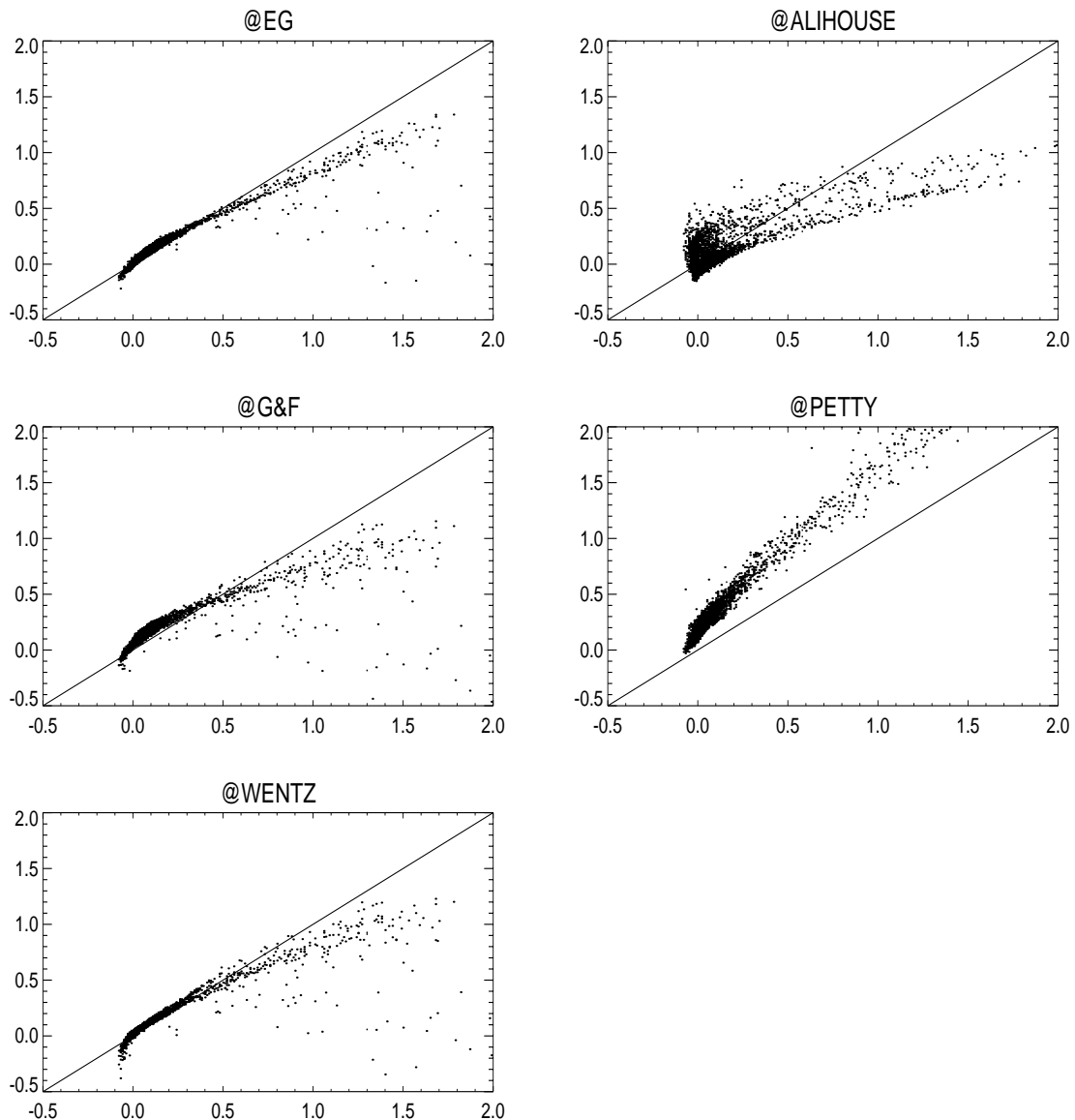


Figure V-9. Comparaison de l'algorithme neuronale d'eau liquide avec les algorithmes de la littérature (EG pour E.Gérard et G&F pour Grody et Ferraro) appliqués sur les 4 zones géographiques (6432 points). En abscisse, le contenu en eau liquide estimé par l'algorithme ME_Iclw et en ordonnée celui estimé par les algorithmes de la littérature. Unités en kg/m^3 .

Revenons sur la détection du ciel clair, c'est-à-dire des situations sans eau liquide nuageuse. La Figure V-10 permet de mettre en évidence le comportement de chacun des algorithmes en présentant les histogrammes de probabilités de l'eau liquide pour les 4 zones d'études réunies. Rappelons que la comparaison présentée ici est uniquement relative. Tout d'abord, on remarque la présence de valeurs négatives d'eau liquide, en quantité importante dans le cas des algorithmes de Alishouse (~25%) et de Gérard (~15%). Si d'un point de vue opérationnel, ces valeurs ne posent pas de problème car elles peuvent être

systématiquement remises à zéro, d'un point de vue physique, elles mettent en évidence un problème dans la fonction de transfert de chacun de ces algorithmes au voisinage du zéro. Les algorithmes de Grody et Ferraro (1992) et de Petty (1990) présentent des maximum de probabilités, respectivement autour de 0.1 kg/m^2 et de 0.2 kg/m^2 . Si ces maximum peuvent sembler suspects, les très faibles probabilités associées aux situations de ciel clair le sont d'autant plus.

En ce qui concerne l'algorithme de Wentz (1997) et l'algorithme multi-expert, leurs histogrammes de probabilités apparaissent relativement semblables avec une probabilité maximale pour les situations d'air clair.

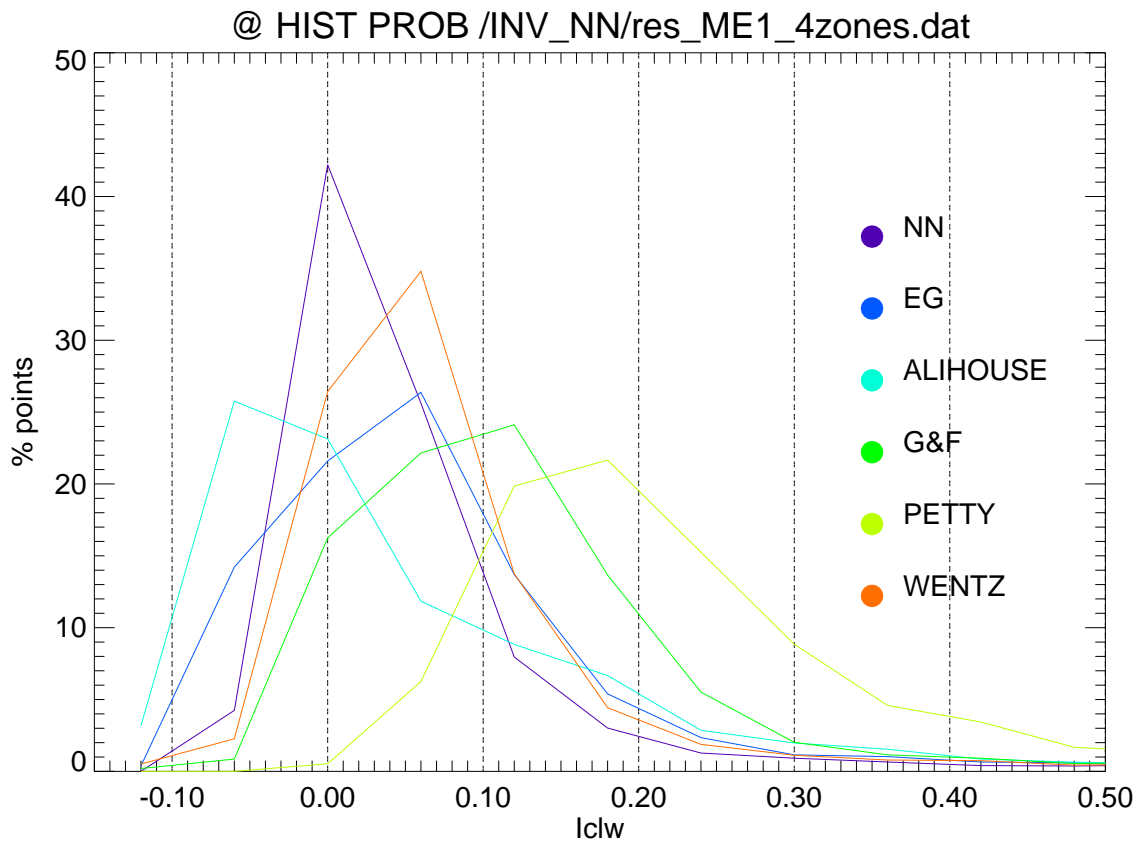


Figure V-10. Histogrammes de probabilité de l'eau liquide des 5 algorithmes de la littérature et de l'algorithme ME_Iclw sur les 4 zones d'études.

V.4.3.2 Vraisemblance des champs d'eau liquide restitués

Ne pouvant pas effectuer de validation directe par manque de mesures d'eau liquide, l'idée est de comparer les champs d'eau liquide restitués avec une image infrarouge.

Nous nous sommes intéressés au cas du 6 Février 1993 sur la région de TOGA-COARE. Pour ce cas d'étude, nous disposons à la fois de mesures de températures de brillance effectuées par le radiomètre SSM/I et de mesures infrarouges effectuées par le radiomètre OLS. Ces deux instruments sont embarqués sur la même plate-forme militaire DMSP.

Le cas du 6 février 1993 correspond au cyclone « Oliver ». Ce cyclone tropical se déplace vers le Sud à une vitesse d'environ 9 km/h, avec des vents maximum de plus de 150 km/h, créant une forte zone de convergence autour des îles Salomon. A l'Est du cyclone se trouve une large zone avec plusieurs structures nuageuses très localisée (Figure V-12).

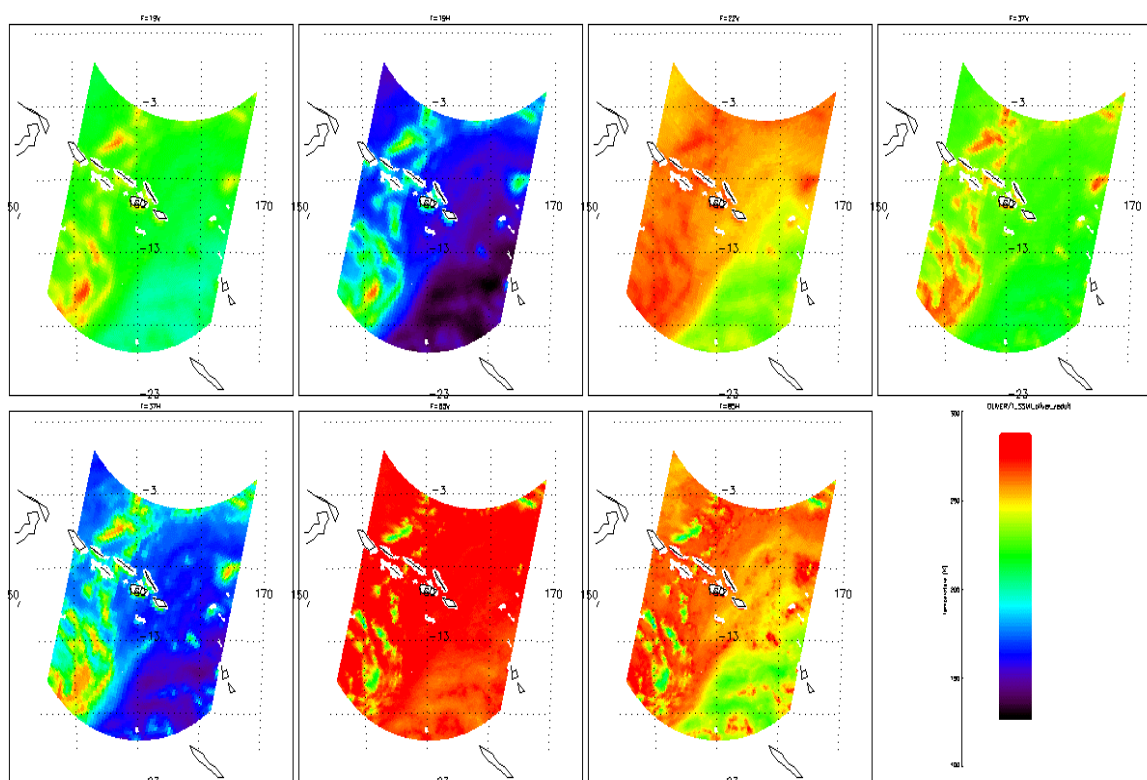


Figure V-11. Températures de brillance aux 7 fréquences SSM/I du 6 février 1993 sur la zone TOGA-COARE. De gauche à droite et de haut en bas, les fréquences :19V, 19H, 22V, 37V, 37H, 85V, 85H.

L'algorithme d'eau liquide neuronal, comme les algorithmes de la littérature, est appliqué sur les températures de brillance mesurées par le radiomètre SSM/I (Figure V-11). La Figure V-13 présente le produit de l'algorithme neuronal et la Figure V-14 ceux des algorithmes de la littérature.

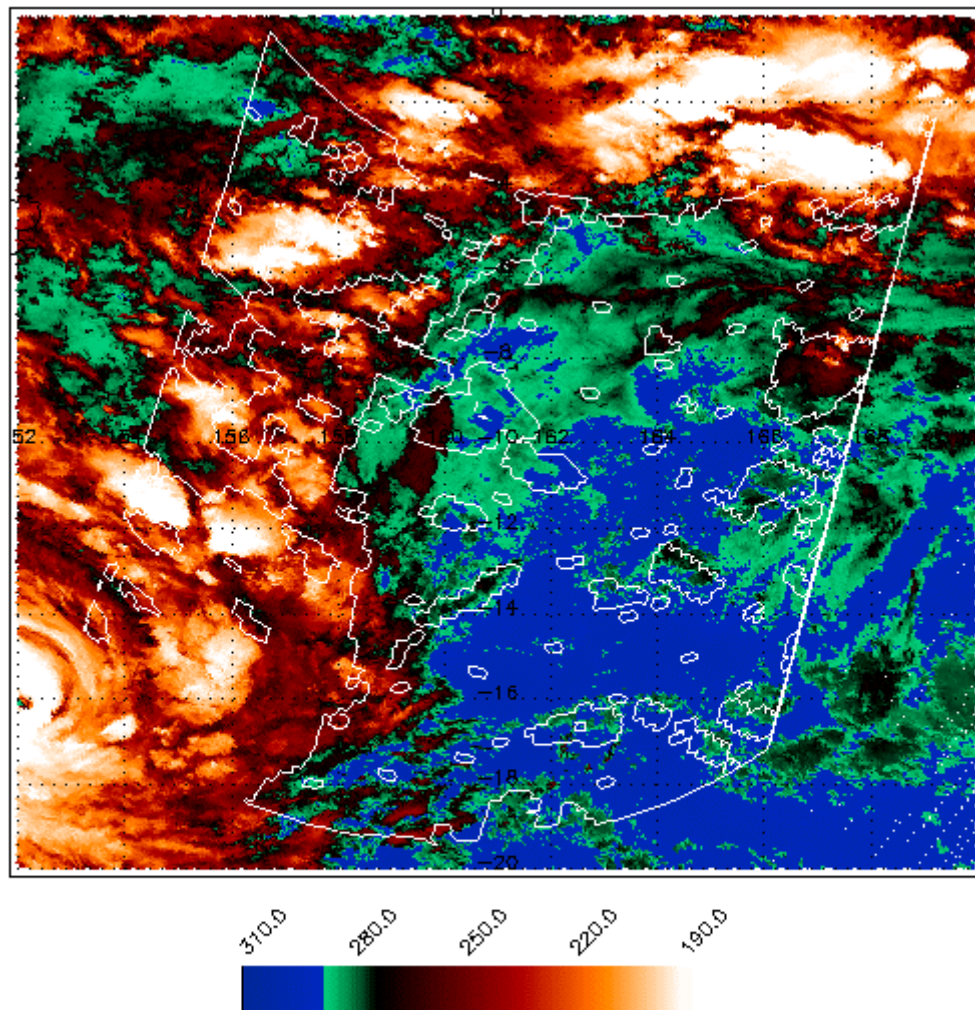


Figure V-12. Températures infrarouge mesurées par l'OLS le 6 Février 1993 sur la région de TOGA-COARE et superposition des contours des nuages restitués par l'algorithme ME_Icdw ($I_{cdw} > 0.02 \text{ kg/m}^3$)

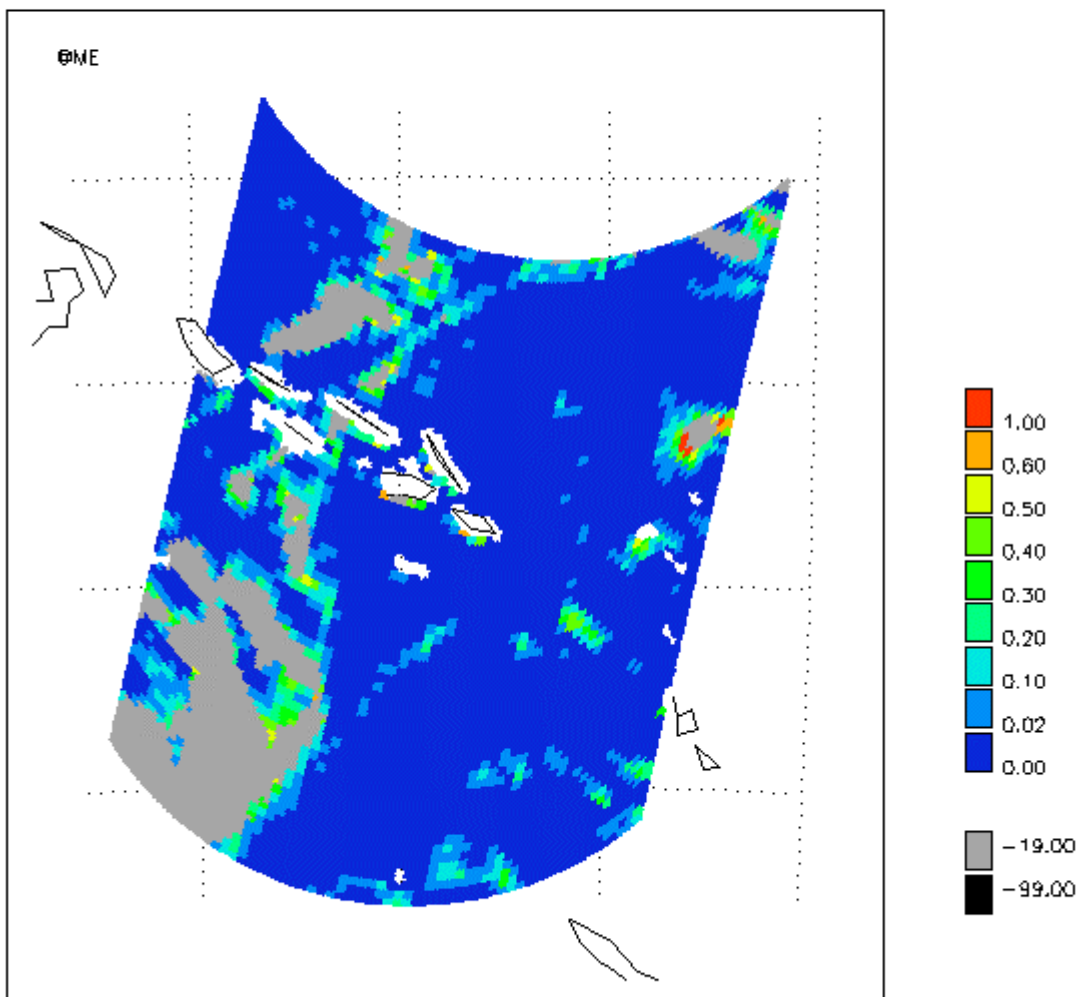


Figure V-13. Produit de l'algorithme neuronal d'eau liquide nuageuse pour le radiomètre SSM/I le 6 février 1993. Le code couleur correspond à la quantité d'eau liquide nuageuse en kg/m^3 . L'indice (-19) correspond aux zones précipitantes (d'après Liu et al. 1993). L'indice (-99) correspond aux zones en dehors du domaine de validité de l'algorithme. Les valeurs négatives de l'eau liquide ont été mises à zéro.

Afin de rester dans le domaine de validité des algorithmes d'eau liquide nuageuse, un masque de précipitation a été ajouté aux restitutions. Le masque choisi est celui de Liu *et al.* (1993). Il est basé sur le calcul d'un coefficient, nommé intensité du signal micro-onde, qui dépend de la dépolarisation du signal à 19 GHz et du PCT (Spencer *et al.* 1989), la polarisation corrigée de la température. En fonction de la valeur de ce coefficient, le pixel considéré est dit précipitant ou non précipitant.

On retrouve les principales structures nuageuses observées sur les données infrarouge, avec le cyclone Oliver au Sud-Ouest et une longue bande nuageuse, qui s'étire jusqu'à l'extrémité Nord-Est de la zone. Ces structures, caractérisées par des températures infrarouge élevées, confirment la présence de précipitation détectée par le masque de Liu (1993). Au bord de ces structures de pluie, l'algorithme neuronal détecte la présence de nuages avec des contenus en eau liquide qui croissent lorsque l'on se rapproche des cellules de pluies. Les contenus maximaux en eau liquide restitués aux bords de ces cellules sont typiquement supérieurs à 0.5 kg/m^2 , et peuvent même aller jusqu'à 1.0 kg/m^2 .

Dans le quart Sud-Est de la zone d'étude, on retrouve plusieurs structures nuageuses très localisées géographiquement. Ces structures classifiées comme non précipitantes présentent des contenus en eau liquide maximale pouvant aller de 0.2 kg/m^2 à 0.6 kg/m^2 . Les valeurs simulées supérieures à 1.0 kg/m^2 ne correspondent pas à des situations réalistes, elles évitent que la limitation ne proviennent de la base d'apprentissage. L'algorithme ainsi obtenu peut prétendre à un domaine de validité aussi grand que possible.

Globalement, le champ nuageux restitué est en bon accord avec l'image infrarouge, toutes les structures nuageuses restituées via l'algorithme neuronal (contours blanc dans la Figure V-12) sont présentes sur l'image infrarouge. Inversement, certaines structures nuageuses présentes sur l'image infrarouge, ne le sont pas sur le champ restitué. Par exemple, la structure nuageuse au Nord de la zone, de forme filamenteuse dans la direction Est-Ouest. On remarque que cette structure est également absente des restitutions réalisées avec les algorithmes de la littérature, est quasi inexistante au regard des 5 températures de brillance basse fréquence.

Le décalage spatial entre les données IR et les contours de nuage restitués par l'algorithme ME_Iclw est vraisemblablement dû à des différences de navigations et d'angles de vues entre les instruments.

La Figure V-14 présente les champs restitués sur le cas du 6 février 1993 par les algorithmes de la littérature. On observe que :

-- Les structures nuageuses importantes sont globalement bien restituées par tous les algorithmes (sauf Alihouse (1990)).

-- Les algorithmes de Grody et Ferraro (1992), de Petty (1990) et de Wentz (1997) produisent d'importantes structures nuageuses de très faibles contenus ($< 0.1 \text{ kg/m}^2$) qui ne sont pas présentes avec les autres algorithmes, y compris le nôtre. Il apparaît difficile de pouvoir porter un jugement objectif sur la vraisemblance de ces structures. Néanmoins il paraît clairement que l'algorithme neuronal et celui de Gérard (1996) produisent des structures nuageuses plus « piquées », avec des gradients plus importants que les autres algorithmes.

-- Les faibles domaines de validité des algorithmes de Alihouse (1990), de Grody et Ferraro (1992), de Petty (1990) et de Wentz (1997) sont une limitation importante pour la restitution de contenus en eau liquide dans un nombre important de situations et en particulier aux bords des structures de pluie.

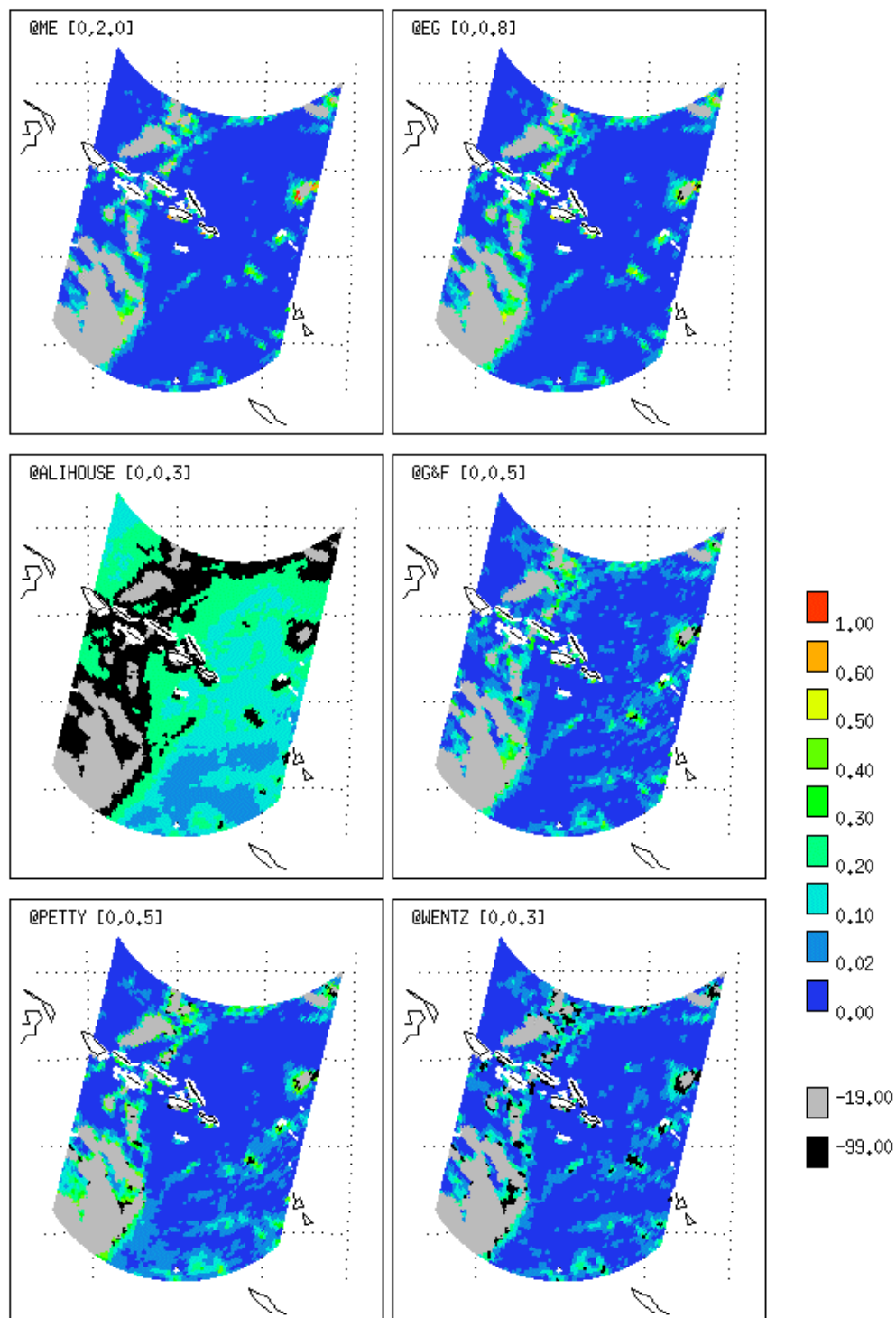


Figure V-14. Idem Figure V-13 pour les algorithmes de littérature. Le domaine de validité de chaque algorithme est porté en haut de chaque figure.

V.4.3.3 Résumé

L'algorithme d'Alihouse, bien que pionnier en matière de restitution d'Iclw pour SSM/I, est le plus « critiquable » des algorithmes que nous avons considérés dans cette étude, avec des structures nuageuses incohérentes avec celles de l'image infrarouge OLS. C'est le seul algorithme, parmi les cinq sélectionnés, à avoir été développé à partir d'une base de données composée uniquement de co-localisation entre des températures de brillance SSM/I et des contenus en eau liquide déduits de radiosondages, et à utiliser une régression linéaire.

Il ressort de la comparaison sur les 4 zones d'études que l'algorithme de Petty (1990) fournit le plus d'eau liquide, il est suivi ensuite de l'algorithme de Grody et Ferraro (1992).

Pour les faibles contenus en eau liquide, notre algorithme a un comportement très proche de ceux de Wentz (1997), de Gérard (1996) et de Grody et Ferraro (1992). On observe néanmoins une légère sous-estimation des contenus restitués par notre algorithme par rapport aux autres. Cependant, on peut penser que le comportement de notre algorithme est le plus proche de la réalité. En effet, il semble avoir le meilleur comportement au voisinage de zéro.

Pour les forts contenus en eau liquide, la comparaison est d'autant plus restreinte que les domaines de validité de quatre des cinq algorithmes sont inférieurs à 0.5 kg/m^2 . Les études menées par Gérard (1996) tendent à montrer que son algorithme sous-estime les contenus en eau liquide et donc que l'algorithme ME ne les surestime pas. La tendance de la forme log-linéaire à sous-estimer les forts contenus en eau liquide confirme ce résultat.

V.5 Conclusion

A partir de la base de donnée SMSP décrite dans le chapitre précédent, nous avons simulé les température de brillance émergentes en haut de l'atmosphère aux cinq fréquences SSM/I, en utilisant le modèle de transfert radiatif que nous avons développé (Chapitre II). Trois algorithmes ont été développés à partir de ce jeu de donnée et nous avons montré que les deux algorithmes neuronaux (PMC et ME) étaient beaucoup mieux adaptés à la restitution de l'eau liquide nuageuse que l'algorithme classiquement utilisé, de forme « log-linéaire ». A performance équivalente, nous avons choisi, pour effectuer la validation, l'algorithme multi-expert, qui a l'avantage d'être relativement interprétable. La validation de l'algorithme a consisté en une comparaison avec cinq algorithmes de la littérature sur des données SSM/I, correspondant à plusieurs situations météorologiques.

Des diverses comparaisons entre l'algorithme d'eau liquide pour SSM/I, il ressort que l'algorithme ME_Iclw : (1) fournit une bonne détection des faibles contenus, (2) restitue des champs de nuage en bon accord avec ceux issus d'images infrarouge et (3) possède un domaine de validité supérieur aux autres algorithmes (entre 0 et 1.0 kg/m²).

Ces comparaisons ont permis, de plus, non seulement, de vérifier la qualité du modèle de transfert radiatif, mais aussi la cohérence globale de notre approche.

