

Modélisation des covariances inter-canaux des erreurs d'observation du sondeur IASI

FORAY Anouck
OHAYON William

Projet modélisation effectué du 4 au 22 janvier 2010 et du 1 au 12
février 2010

Dans l'équipe CNRM/GMAP/OBS
Sous l'encadrement de Vincent Guidard et Aurélie Bouchard

RESUME

Une des difficultés de l'assimilation est de déterminer les matrices de covariance d'erreurs de mesure. Actuellement les matrices de covariances d'erreurs d'observation, notées R , sont empiriques et diagonales, ce qui signifie notamment que l'on néglige les corrélations entre les erreurs de mesure des différents canaux d'un même instrument.

Dans leur article de 2005 G.Desroziers et al. établissent une formule qui permet de calculer la matrice R à partir d'informations issues de cycles d'assimilation. Un premier calcul de R fait pour des mesures sans nuage et sur mer des données IASI, montre un certain nombre de corrélations à prendre en compte pour améliorer l'assimilation. Un second calcul fait sur terre et toujours sans nuage montre une structure différente, avec des corrélations plus fortes et plus difficiles à prendre en compte.

Le but étant de conserver les corrélations qui améliorent significativement l'assimilation, la nullité éventuelle des coefficients non diagonaux a été testé statistiquement. L'opération a été renouvelée sur d'autres zones géographiques toujours en ciel clair. (Antarctique, côtes...) Les matrices des différentes zones géographiques ont été comparées entre elles pour identifier les différences entre les corrélations en fonction de la zone géographique.

Pour chaque zone géographique, des données prises en un point où un radiosondage a été réalisé, ont été relevées. Ces données ont été utilisées dans un modèle 1D-VAR pour restituer des profils de température et d'humidité qui ont été comparés aux profils obtenus par radiosndage. Ces profils ont aussi été comparés à ceux obtenus par le 1D-VAR utilisant une matrice R de référence, diagonale. Les résultats sont assez convaincants sur mer, cela encourage de futures études modifiant le code du 4D-VAR du modèle ARPEGE pour prendre en compte les covariances intercanaux de IASI.

REMERCIEMENTS

Nous remercions l'équipe du le GMAP pour son accueil.
Nous remercions Gérald Desroziers pour son cours sur l'assimilation qui nous a été fort utile, et pour sa *méthode* élément clef de ce projet.
Et en particulier Vincent Guidard et Aurélie Bouchard, nos encadrants pour ce projet, qui ont su nous apporter une aide efficace.

Table des matières

Glossaire	5
Introduction	6
1 CONTEXTE GENERAL	8
1.1 L'assimilation 4D-VAR	8
1.2 L'instrument IASI	9
1.3 La méthode de Desroziers	12
2 APPLICATION DE LA METHODE DE DESROZIER A IASI	14
2.1 Description des données de base pour appliquer la méthode de Desroziers	14
2.2 Choix des données	14
3 LE CALCUL DES MATRICES DIAGNOSTIQUES	16
3.1 Le calcul de la matrice diagnostique	16
3.2 Comparaison des matrices diagnostiques des différentes zones :	22
4 IMPACT DES MATRICES DIAGNOSTIQUES	25
4.1 Le 1D-VAR	25
4.2 La colocalisation des données	25
4.3 L'estimation de l'efficacité de la matrice	26
4.4 Les résultats sur mer :	28
4.5 Les résultats sur terre	36
Conclusions et perspectives	42
Bibliographie	43
Annexes	44

GLOSSAIRE

1D-VAR : 1 Dimensional VARIational assimilation

4D-VAR : 4 Dimensional VARIational assimilation

AIRS : Atmospheric Infrared Sounder

ARPEGE : Action Recherche Petite Echelle Grande Echelle

BDM : Base de Données Météorologiques

CNES : Centre National d'Etudes Spatiales

CRIS : Cross-track Infrared Sounder

EUMETSAT : European Organisation for the Exploitation of Meteorological Satellites

IASI : Interféromètre Atmosphérique de Sondage Infrarouge.

PNT : Prévision Numérique du Temps

RMS : Root Mean Square, ou racine carrée de la moyenne quadratique

INTRODUCTION

La prévision météorologique nécessite de connaître l'état de l'atmosphère, à un instant donné, et les lois physiques qui en gouvernent l'évolution. La connaissance de l'état de l'atmosphère est déterminée à l'aide de mesures, appelées observations, de paramètres classiques comme la température, la pression, le vent etc... A cela s'ajoutent, depuis l'utilisation des satellites, les mesures des températures de brillances au sommet de l'atmosphère, et ceci pour différentes longueurs d'onde. Quelque soit l'observation effectuée, la mesure faite est toujours entachée d'erreurs. L'état de l'atmosphère ne peut donc être connu de façon exacte, il faut analyser les différentes observations, estimer une approximation de l'état réel de l'atmosphère, en minimisant les erreurs commises. Cette minimisation passe par l'évaluation des erreurs d'observations. Ce processus d'optimisation qui prend en compte les observations et leurs erreurs est appelé assimilation.

Les erreurs sont considérées comme des variables aléatoires gaussiennes. Le satellite Metop A utilise pour mesurer des températures de brillance au sommet de l'atmosphère, un sondeur nommé IASI dont chaque canal correspond à une longueur d'onde. Des tests réalisés sur IASI en laboratoire donnent une estimation des écarts types des erreurs, des corrélations entre les erreurs, et le cas échéant de leur biais (de leur espérance). Ces valeurs mesurées en laboratoire sont purement instrumentales, or les données utilisées pour l'analyse de l'atmosphère sont prétraitées et d'autres erreurs apparaissent, il faut aussi estimer les paramètres des erreurs (écarts types, corrélations et biais) pour l'utilisation qui en faite dans la prévision du temps.

Ce sont les écarts types (noté σ_o) et les corrélations des erreurs de mesures de l'instrument IASI qui font l'objet de cette étude. Actuellement, dans les modèles et en particulier dans ARPEGE ici, l'assimilation des données satellitaires se fait avec une matrice de covariances d'erreurs de mesures entre les différents canaux de IASI diagonale, donc on ne tient pas compte des corrélations d'erreurs intercanaux.

La prise en compte de ces corrélations devrait permettre d'améliorer l'assimilation des données IASI. C'est la méthode de Desroziers qui a été utilisée dans cette étude pour estimer les corrélations des erreurs de mesures. Cette méthode a fourni des matrices de corrélations qui ont ensuite été testées, comparées à la matrice utilisée actuellement dans la version opérationnelle d'ARPEGE. Enfin une première estimation des corrélations qui améliorent significativement l'analyse a été réalisée.

On trouvera tout d'abord dans ce rapport une description du processus d'assimilation 4D-VAR opérationnel à Météo France actuellement, du sondeur IASI, et de la méthode Desroziers. Les deux chapitres suivants sont consacrés aux calculs des matrices de covariances, avec le détail des données utilisées pour les calculs, les choix qui ont conduit aux calculs de plusieurs matrices et les méthodes mises en place pour évaluer la qualité des matrices. Enfin les résultats obtenus sont présentés et commentés.

Chapitre 1

CONTEXTE GENERAL

1.1 L'assimilation 4D-VAR

L'assimilation de données est un processus permettant d'obtenir une représentation aussi fidèle que possible de l'état de l'atmosphère à partir de mesures, de lois statistiques, d'une ébauche et de lois physiques.

Notons, comme utilisé classiquement dans la littérature, x l'état de l'atmosphère décrit par le modèle, x^b un état a priori donné de l'atmosphère (climatologie, ou une prévision à courte échéance dans notre cas), x^a l'état analysé de l'atmosphère issu d'un processus de l'assimilation. Le vecteur d'observations y est défini dans l'espace des observations. Le vecteur x^t est la réalité exprimée dans l'espace des variables du modèle.

Les sources d'information x^b et y sont entachées d'erreurs, notées respectivement ϵ^b et ϵ^o où $\epsilon^b = x^b - x^t$ et $\epsilon^o = y - H(x^t)$ avec H l'opérateur permettant de passer des variables du modèle à l'espace des observations. L'opérateur H peut contenir des interpolations horizontales ou verticales ; ou un modèle de transfert radiatif qui permet de simuler des observations satellitaires.

En général on considère que les erreurs sont des variables aléatoires indépendantes qui suivent des lois gaussiennes. Une des difficultés de l'assimilation est l'estimation de ces erreurs, puisque la réalité n'est jamais connue.

La matrice R est la matrice des variances-covariances des erreurs d'observation :

$$E(\epsilon^o \epsilon^{oT}) = R$$

La matrice B est la matrice de covariance des erreurs de prévision : $B = E(\epsilon^b \epsilon^{bT})$ où $E()$ est l'espérance mathématique et T la transposition. x^a est l'état de l'atmosphère qui minimise la fonction quadratique

$$J(x) = \frac{1}{2}(x^b - x)^T B^{-1}(x^b - x) + \frac{1}{2}(y - H(x))^T R^{-1}(y - H(x))$$

Le vecteur d'innovation d_b^o vaut $y - H(x^b)$ tandis que le vecteur des résidus d_a^o vaut

$y - H(x^a)$. Le premier mesure donc l'écart entre l'observation faite de l'atmosphère à un instant donné et l'état de celle-ci prévu par l'ébauche, la seconde mesure l'écart entre l'observation et l'analyse issue de l'assimilation au même instant, dans les deux cas dans l'espace des observations.

Le 4D-VAR est une réponse à la prise en compte des observations à différents instants. C'est une méthode d'assimilation variationnelle quadridimensionnelle. Cette méthode ne vise pas à obtenir l'état optimal à un instant donné, mais la trajectoire optimale sur une fenêtre de temps donné. Le 4D-Var apporte un aspect temporel car il propage l'information apportée par les observations sur toute la fenêtre d'assimilation. De ce fait, l'analyse obtenue doit permettre au modèle d'évolution d'avoir la trajectoire la plus proche possible de l'ensemble des observations utilisées, et de l'ébauche.

1.2 L'instrument IASI

IASI est un interféromètre de Michelson sondant l'atmosphère, il est décrit dans l'article de Phulpin et al. de 2003. Il est composé de 8461 canaux, il a une résolution spectrale de $0,25 \text{ cm}^{-1}$ (le canal numéro 1 correspond au nombre d'onde 645 cm^{-1}) et une résolution spatiale de 12 km au nadir qui augmente sur les bords de la fauchée. La technique de mesure est basée sur la télédétection infrarouge passive utilisant un spectromètre à transformée de Fourier calibré avec précision travaillant dans la gamme spectrale de $3.7 - 15.5 \text{ }\mu\text{m}$. La quantité d'information apportée par cet instrument est très importante. Afin de limiter le coût numérique, une sélection de l'information utilisée dans les modèles de PNT est faite en n'assimilant qu'un petit nombre de canaux : 77 dans la version pré-opérationnelle actuelle d'ARPEGE (cf. Figure 1.1). La figure 1.2 montre à quelle zone de l'atmosphère et à quels paramètres physiques sont sensibles ces 77 canaux. La figure 1.3 illustre d'une autre façon les zones de l'atmosphère auxquelles ces canaux sont sensibles.

IASI a été conçu pour des sondages météorologiques opérationnels avec une très grande précision (spécifications de la précision de la température : 1K pour 1 km et 10 pourcents pour l'humidité) dédié à l'amélioration des prévisions météorologiques à moyen terme. Il est aussi conçu pour la chimie de l'atmosphère dans le but d'estimer et de suivre les gaz traces tels que l'ozone, le méthane ou le monoxyde de carbone à l'échelle mondiale. La mesure est contaminée par la présence de nuage. La prise en compte de l'effet des nuages n'étant pas facile dans l'assimilation, des méthodes de détection nuageuse sont mises en place pour identifier les canaux contaminés par les nuages et les rejeter de l'assimilation.

IASI - spectre complet

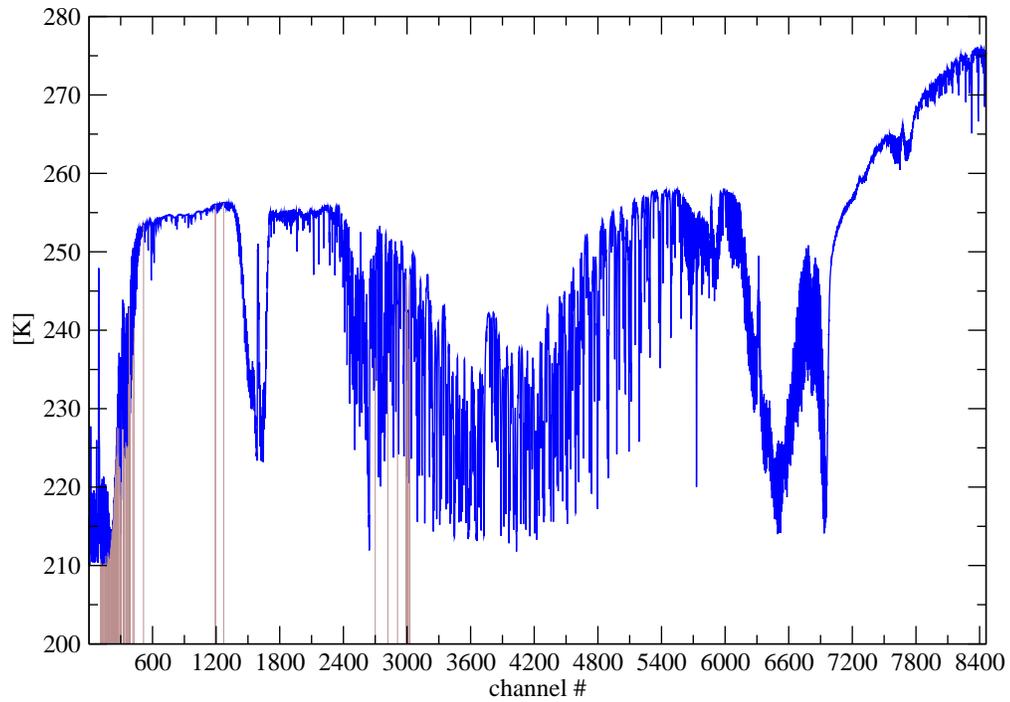


FIG. 1.1 – Spectre des températures de brillance pour l'ensemble des canaux IASI (en bleu). Les 77 canaux étudiés dans ce rapport sont identifiés par les barres verticales marrons

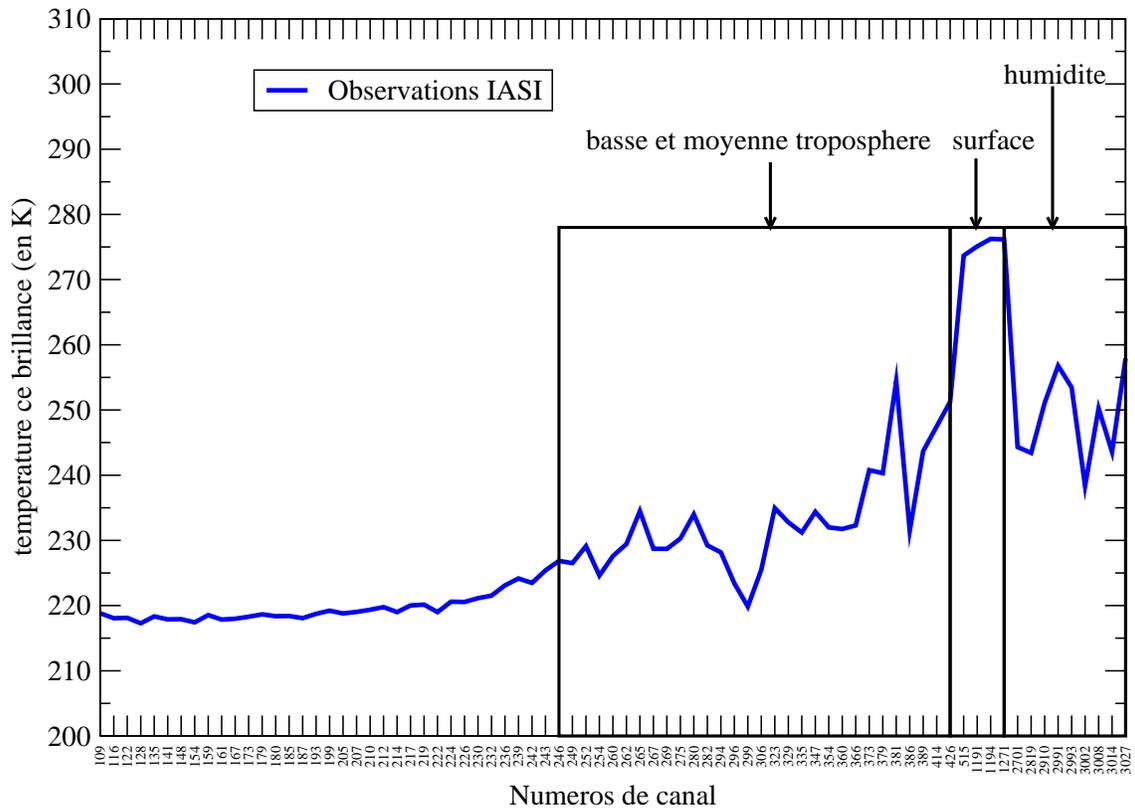


FIG. 1.2 – spectre des canaux I.A.S.I. assimilés en opérationnel avec les zones de sensibilité

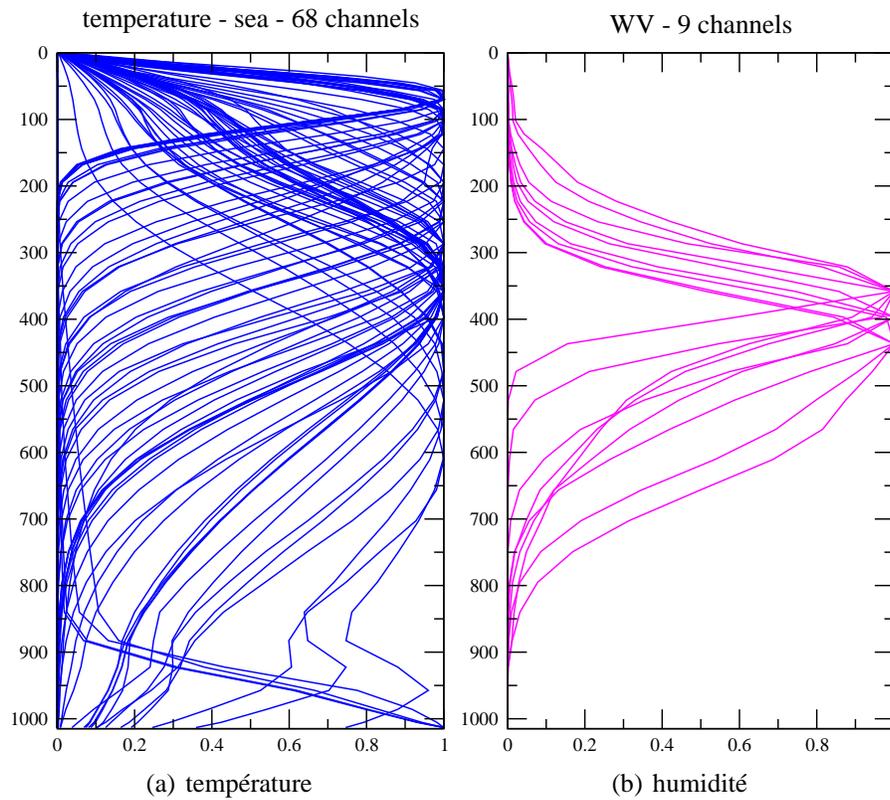


FIG. 1.3 – Fonctions de poids des 77 canaux IASI étudiés dans ce rapport (sensibilité [nombre sans unité entre 0 et 1] en abscisse ; altitude [hPa] en ordonnée)

1.3 La méthode de Desroziers

Cette méthode est une estimation a posteriori de la matrice R . L'état réel de l'atmosphère n'étant pas connu, les matrices de covariances d'erreurs d'ébauche et d'observations sont évaluées de façon empirique.

Desroziers et al. (2005) montrent que, si les conditions suivantes sont réalisées :

- le vecteur des erreurs d'ébauche est gaussien
- le vecteur des erreurs d'observation est gaussien
- la matrice de covariance des erreurs d'ébauche est "juste"¹
- la matrice de covariance des erreurs d'observation est "juste"¹

¹matrice "juste" au sens où l'on considère pour une mesure donnée que l'on connaît parfaitement la loi de probabilité du vecteur gaussien d'erreurs

alors on peut par le calcul retrouver ces matrices ; cela sans avoir besoin de l'état réel de l'atmosphère, uniquement à partir des observations, de l'ébauche et de l'analyse .

Ainsi :

$$\begin{aligned} E[d_b^a(d_b^o)^T] &= \mathbf{H}B\mathbf{H}^T \\ E[d_a^o(d_b^o)^T] &= R \end{aligned}$$

avec d_b^a l'écart de l'analyse à l'ébauche dans l'espace des observations et \mathbf{H} la matrice de l'opérateur d'observation linéarisé.

Ces matrices *diagnostiques* permettent de savoir si les matrices prescrites a priori, avant analyse, sont bien spécifiées ou non. Les matrices obtenues par ce calcul, qui ne peut se faire qu'après analyse, peuvent à leur tour être utilisées lors de l'assimilation. Il faut noter que les matrices utilisées lors de la première analyse n'étant pas "*justes*", les nouvelles matrices ne le sont pas non plus. Mais on peut espérer que les matrices *diagnostiques* soient de meilleures approximations. Une méthode itérative peut être une solution pour optimiser les matrices de covariances d'erreurs.

²L'analyse étant une estimation de l'état de l'atmosphère résultant de l'assimilation

Chapitre 2

APPLICATION DE LA METHODE DE DESROZIERS A IASI

En utilisant la méthode de Desroziers, appliquée aux données IASI, les prévisions ARPEGE pourraient être améliorées grâce à la description et la prise en compte des corrélations entre les erreurs des différents canaux dans le processus d'assimilation.

2.1 Description des données de base pour appliquer la méthode de Desroziers

Cette étude a été faite avec les données des 77 canaux assimilés de IASI datant du 1^{er} au 27 décembre 2009. Chaque fichier contient les 6 heures de données IASI comprises entre deux assimilations ARPEGE (toutes les 6 heures). Les 108 fichiers se présentent sous la forme de tableaux contenant de nombreuses données telles que le lieu, le numéro du canal, l'innovation, le résidu, la présence ou l'absence de nuage...

2.2 Choix des données

Un tri des données a été réalisé (à l'aide d'un programme codé en Fortran 90).
Ont été conservées les données suivantes :

- données effectivement assimilées par ARPEGE dans sa version opérationnelle
- en ciel clair, de façon à ce que l'effet des nuages ne soit pas à modéliser dans le transfert radiatif

L'influence du type de surface sur le bilan radiatif, conduit à une nouvelle séparation des données, tout d'abord, de la façon suivante :

- les zones au delà du 60°NORD et du 60°SUD ont été éliminées pour éviter l'impact sur le bilan radiatif des glaces de mer et des terres gelées ou enneigées
- zones maritimes (la part de terre du point de grille sondé inférieure à 0.2) avec 162197 points de mesure utilisables
- zones terrestres (la part de terre du point de grille sondé supérieure à 0.8) avec 281548 données utilisables
- zones côtières (la part de terre du point de grille sondé comprise entre 0.2 et 0.8) avec 49443 points de mesure utilisables

Enfin une surface polaire a été retenue (non pas pour être utilisée pour notre étude directement, mais pour préparer les données pour la campagne Concordiasi cf. Bibliographie 4.5) :

- zone antarctique (zone terrestre sous le 60°SUD) avec 42471 points de mesure utilisables

Chapitre 3

LE CALCUL DES MATRICES DIAGNOSTIQUES

3.1 Le calcul de la matrice diagnostique

A l'aide de la formule décrite au paragraphe 1.1 chaque point de mesure permet le calcul d'une matrice diagnostique. La matrice obtenue pour un seul point de mesure n'a pas de sens statistique. Il convient de réaliser des moyennes sur un grand nombre de point de mesures, de façon à pouvoir appliquer la loi des grands nombres, qui assure que la moyenne empirique des matrices converge vers l'espérance.

Le programme de calcul a été développé par étapes :

Les premiers calculs ont été effectués sur un fichier de données (par jour et tranche de 6 heures) puis élargis à une journée entière et enfin effectués sur l'ensemble des données du 1^{er} au 27 décembre 2009. Les 27 matrices de covariances R, une par jour, sont calculées pour chaque type de surface et avec les données triées de la façon précédemment décrites.

Les matrices de covariances obtenues ont été séparées en matrices de corrélations d'une part et écarts types¹ notées σ_o d'autre part. Cela permet de comparer les σ_o ainsi diagnostiqués à ceux de la matrice diagonale utilisée en opérationnel pour voir l'impact de la valeur de ces coefficients en sortie du 1D-VAR (cf. Figure 3.1). Les σ_o de la matrice diagnostique sur mer sont plus faibles que ceux de la matrice diagnostique sur terre pour les canaux supérieurs à 400 ce qui correspond aux canaux de surface et d'humidité.

Dans la suite R_{cor} désigne la matrice de corrélations.

La matrice R_{cor} théorique étant symétrique mais la matrice diagnostique ne l'étant pas forcément, on a utilisé la décomposition unique d'une matrice $M=S+A$, où S est une matrice symétrique et A une matrice antisymétrique² pour la symétriser. D'autres façons

¹racines carrées des valeurs de la diagonale

²avec $S = \frac{M+M^T}{2}$ et $A = \frac{M-M^T}{2}$

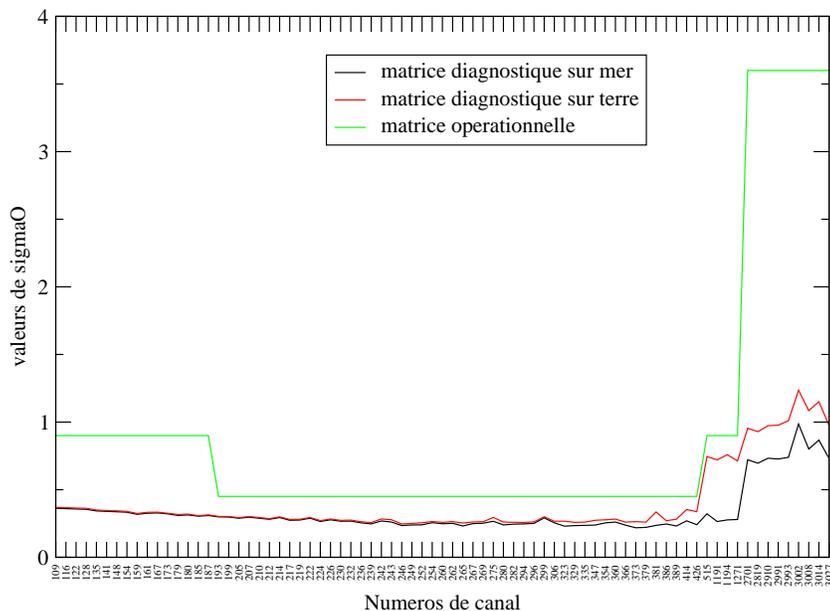


FIG. 3.1 – Valeurs des σ_o des différentes matrices

de symétriser la matrice issue directement du calcul existant, mais ce choix se justifie à nouveau par la loi des grands nombres.

Illustration justifiant la symétrisation des matrices :

D'après les images de matrices tracées sur mer pour le mois de décembre 2009 (cf. Figure 3.2), on remarque bien que la matrice obtenue est quasi symétrique. Il en est de même pour tous les jours.

Le reste de l'étude pourra donc être fait avec la partie symétrique des matrices obtenues.

L'assimilation opérationnelle ayant des contraintes de temps, et étant amenée à prendre en compte plus de 77 canaux dans l'avenir, nous avons choisi de déterminer les coefficients qui peuvent statistiquement être nuls de façon à alléger les calculs. Cela a été réalisé à l'aide d'un "t-test" du logiciel "R" sur les 27 matrices issues du calcul ainsi que sur les 27 matrices symétrisées, sur la surface mer, pour chaque jour de données disponible. Le test portait donc pour chaque coefficient sur une série de 27 nombres. Le test, ainsi réalisé en première approche, comporte un certain nombre de défauts. D'une part il est fait sur 27 valeurs ce qui n'est pas un très grand nombre, mais chacune de ces valeurs représente une moyenne faite sur environ trois à six mille points selon les jours. D'autre part ces coefficients moyens n'étant pas tous calculés avec le même nombre de

points de mesures, il conviendrait de les pondérer. Les résultats obtenus sont les suivants : 1510 coefficients sur 5929 (= 77×77) peuvent être nuls pour les matrices calculées et 1532 coefficients sur 5929 peuvent être nuls pour les matrices symétriques déduites, dans les deux cas les résultats ne changent pas que le niveau du test soit 0,05 ou 0,01 (cf. Figure 3.3).

Remarque : compte tenu des valeurs obtenues il est vraisemblable que le même test aurait permis la validation statistique de l'annulation d'encore moins de coefficients si on avait pris en compte les 162197 matrices calculées sur mer.

Une autre approche a donc été retenue pour se ramener à une matrice quasi-diagonale par blocs (certains blocs étant eux-mêmes diagonaux). Le but est de tester la corrélation des erreurs inter-canaux, et l'impact de leur prise en compte dans l'assimilation. Or plus les niveaux de sondage de deux canaux sont proches, plus la probabilité que la corrélation soit importante est élevée. Pour mettre en relief les différents niveaux de corrélations, les matrices ont été seuillées tous les 0,05 de 0 à 1.

Chaque matrice de corrélation seuillée fournit une matrice de covariance, recalculée avec les σ_o qui peut servir pour l'assimilation.

Elles ont des formes proches dans tous les cas (cf. Figure 3.5) : diagonale par blocs. Avec un bloc carré de fortes corrélations sur les 9 canaux vapeur d'eau en haut à droite sur la figure, puis en dessous un bloc de 4 canaux de surface corrélés plus ou moins fortement suivant la surface, ensuite un bloc de 29 canaux moyennement corrélés, et enfin quelques blocs de 2 canaux (cf. Figure 1.2).

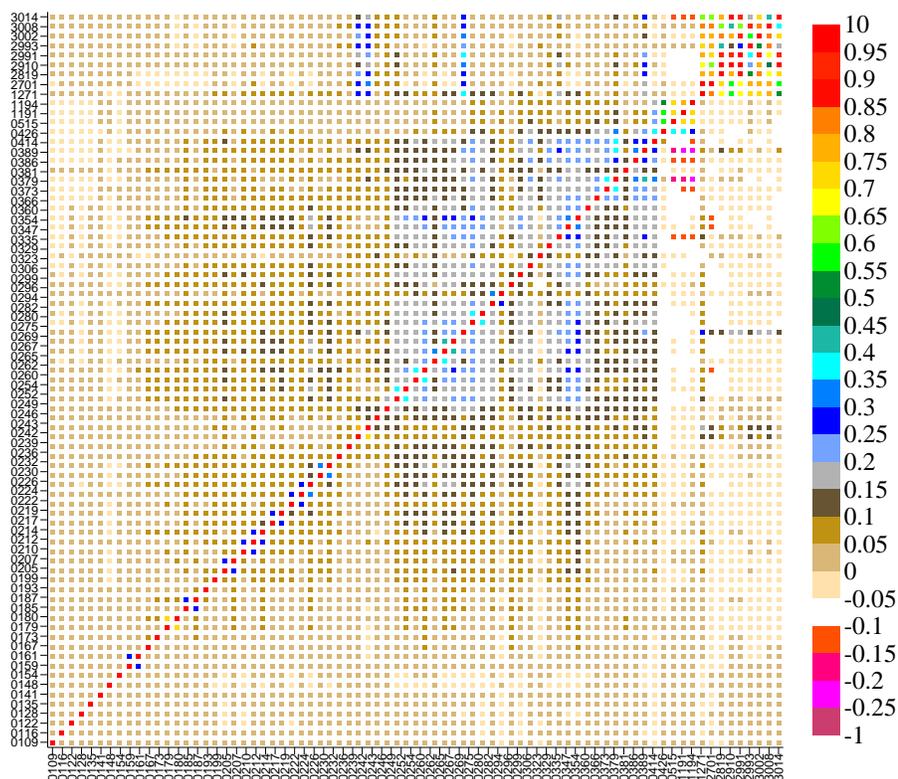
Ces blocs de corrélations s'expliquent de la façon suivante :

- Les fonctions de poids des 9 canaux vapeur d'eau pointent à des niveaux proches et se recoupent fortement (cf. Figure 1.3)
- Les fonctions de poids des 4 canaux de surface pointent à la surface et se recoupent
- les 29 canaux de basse et moyenne troposphère doivent leur corrélation aux éventuelles traces nuageuses pour les nuages non détectés³ ce qui masque la surface et remonte le niveau auquel les fonctions de poids des canaux des basses couches pointent (cf. Figure 1.3). Et pour les surfaces autres que la mer il y a l'orographie⁴, le type de sol, sable, terre, forêt, neige, glace et bien d'autres ont des comportements radiatifs différents.

³la détection automatique des nuages par le modèle de transfert radiatif est parfois erronée, par exemple dans le cas où l'environnement et le nuage ont des températures proches

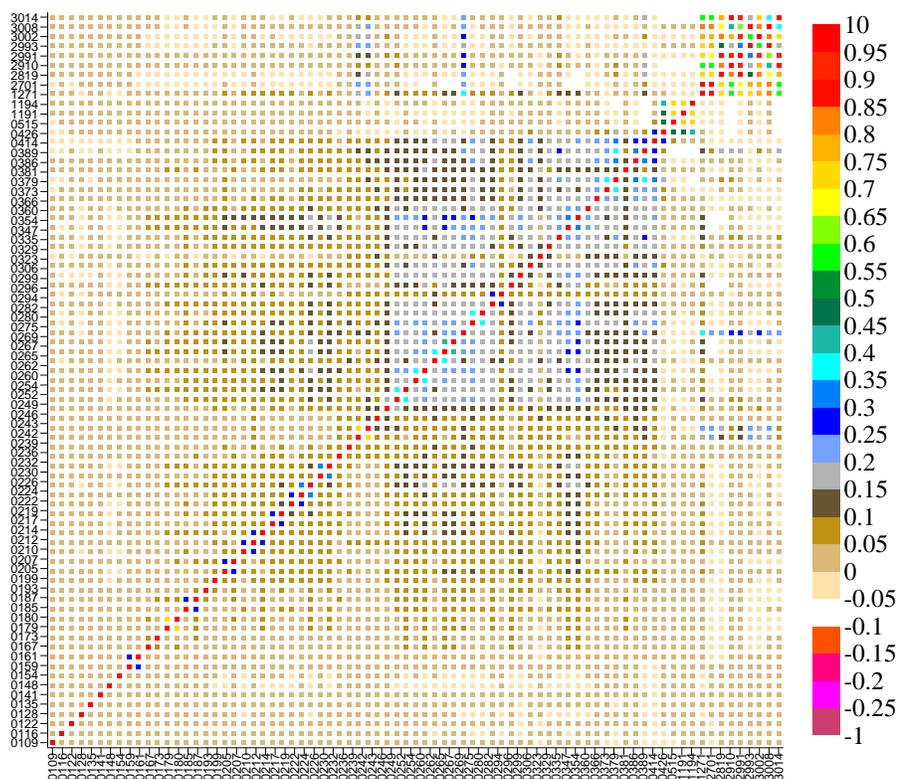
⁴en altitude toutes les fonctions de poids des canaux de surface, basse et moyenne troposphère deviennent sensibles aux mêmes zones de l'atmosphère.

Matrice R - partie corrélation



(a) matrice brute

Matrice R - partie corrélation



(b) matrice symétrique

FIG. 3.2 – Image de la matrice obtenue sur mer pour décembre 2009

Matrice R - partie corrélation

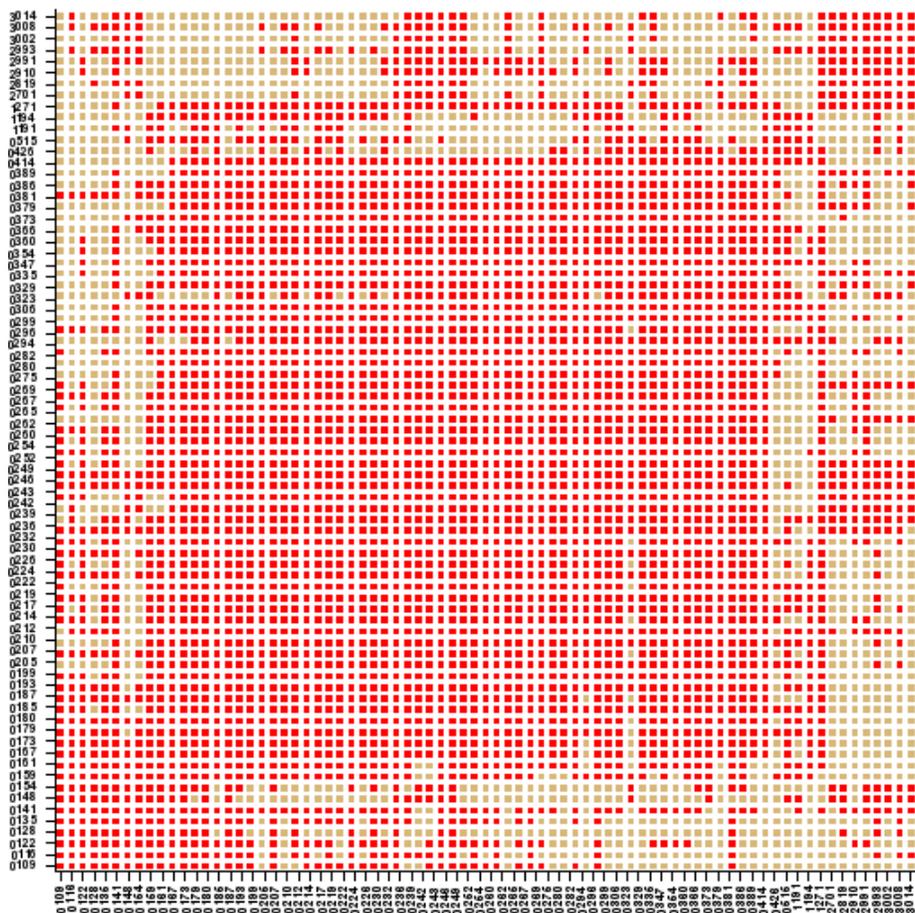


FIG. 3.3 – Matrice des nullités statistiques des coefficients sur mer (en beige : coefficients pouvant être nuls d'après nos tests)

Illustration du seuillage des matrices :

En comparant directement les matrices symétriques établies sur tout le mois de décembre 2009 sur mer et sur terre, on se rend compte que les coefficients de corrélation sont plus élevés sur terre, mais également qu'il y a plus de coefficients nuls ou proches de 0. En comparant les écarts types trouvés à ceux de la matrice utilisée pour l'assimilation opérationnelle, on constate que l'ensemble des valeurs prescrites en opérationnel sont surestimées sur mer et plus proches de la valeur optimale sur Terre (cf. Figure 3.1). D'après le tracé des matrices de corrélations seuillées, il apparaît que certains coefficients sont plus élevés (cf. Figure 3.4). Ces coefficients correspondent aux corrélations des canaux sondant les mêmes niveaux de l'atmosphère et sensibles à l'humidité (numéros de canaux compris entre 2000 et 3000) d'une part et, d'autre part, les canaux sensibles à la température sondant les très basses couches de la troposphère voire la surface (numéros de canaux compris entre 450 et 1300).

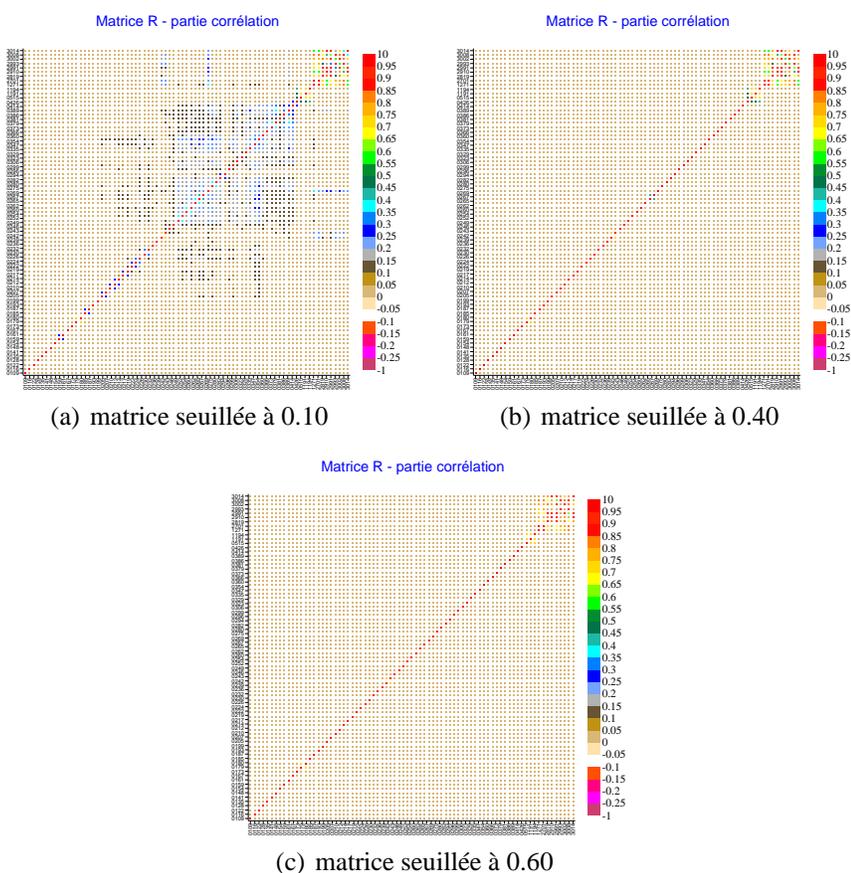


FIG. 3.4 – Image de la matrice symétrique seuillée sur mer pour décembre 2009

3.2 Comparaison des matrices diagnostiques des différentes zones :

Les matrices de chacune des zones étudiées ont la même forme générale (cf. Figure 3.5). Les zones terre et Antarctique présentent davantage de canaux corrélés entre eux que les zones maritimes et côtières, ce sont les canaux compris entre les numéros 360 et 415. Les valeurs des corrélations intercanaux varient selon les zones, celles-ci peuvent donc être classées suivant ces valeurs.

Dans l'ordre croissant des corrélations on a : la zone maritime, la zone côtière, la zone terrestre et la zone Antarctique.

En effet, selon la nature de la surface étudiée, les corrélations sont différentes. Sur terre, les corrélations sont plus fortes que sur mer car le niveau de sondage de chaque canal est modifié par le relief et que les effets des paramètres de surface (température et émissivité) sont plus forts. Les niveaux de sondages se recoupant davantage sur terre les valeurs des corrélations y sont plus importantes que sur mer. De même si ces mesures avaient été menées sur des zones nuageuses, les coefficients auraient été plus élevés ; cela s'explique par le fait que les nuages ont tendance à élever en altitude le niveau sur lequel pointe la fonction de poids d'un canal donné. Ainsi les différents niveaux de sondages sont proches les uns des autres, et les corrélations en sont augmentées comparativement à celles obtenues avec une atmosphère sans nuage.

Les matrices à utiliser semblent être différentes suivant la zone d'application. On remarque cependant une structure globale commune : diagonale par blocs. (cf. Figure 3.5) Sur l'Antarctique, l'altitude moyenne est élevée. Les canaux qui, sur mer, sont sensibles à la moyenne et à la basse troposphère se trouvent plus fortement influencés par la surface. Par ailleurs, la détection nuageuse est moins facile dans les atmosphères quasi-isothermes froides qui s'accompagnent d'une surface, elle aussi, froide. Des contaminations nuageuses résiduelles sont donc possibles dans cet échantillon.

Conclusion

L'utilisation de 77 canaux sur les 8461 de IASI, a permis d'utiliser les matrices non seuillées dans le 1D VAR sans que le temps de calcul ne soit trop important (environ 2 minutes par jeu de données). Pour chaque surface, la matrice utilisée est la matrice moyenne sur le mois de décembre, encore une fois ce choix est justifié par la loi des grands nombres. Il conviendrait de tester les différents seuillages et d'évaluer la perte d'information acceptable si l'on veut utiliser une matrice seuillée, et donc simplifiée, avec éventuellement un ajustement des écarts types σ_o , ce qui pourrait faire le sujet d'un autre projet si besoin.

L'article "Estimates of spatial and inter-channel observation error characteristics for

current sounder radiances for NWP" de Niels Bormann, Andrew Collard and Peter Bauer, présente trois méthodes différentes pour estimer et analyser les erreurs d'observation et leur corrélation en ciel clair et ceci pour différents instruments dont IASI, dans le système d'assimilation du Centre Européen pour les prévisions météorologiques à moyen terme.

Ces trois méthodes sont :

- La méthode d'Holligsworth-Lönnerberg
- Une méthode de soustraction des erreurs d'ébauches et des covariances
- Le diagnostique de Desroziers

Les sondages effectués par AIRS et IASI suggèrent que la température de la moyenne troposphère à la stratosphère ne présente pas ou peu de corrélation intercanale ou spatiale des erreurs d'observation. Les canaux les plus sensibles à la surface ont une erreur d'observation plus importante et parfois corrélée spatialement et entre les différents canaux. Les trois méthodes s'accordent sur ces résultats.

Les estimations faites sur l'humidité sont plus complexes à interpréter car les erreurs d'observation sont bien plus importantes.

Ces conclusions sont comparables à nos résultats.

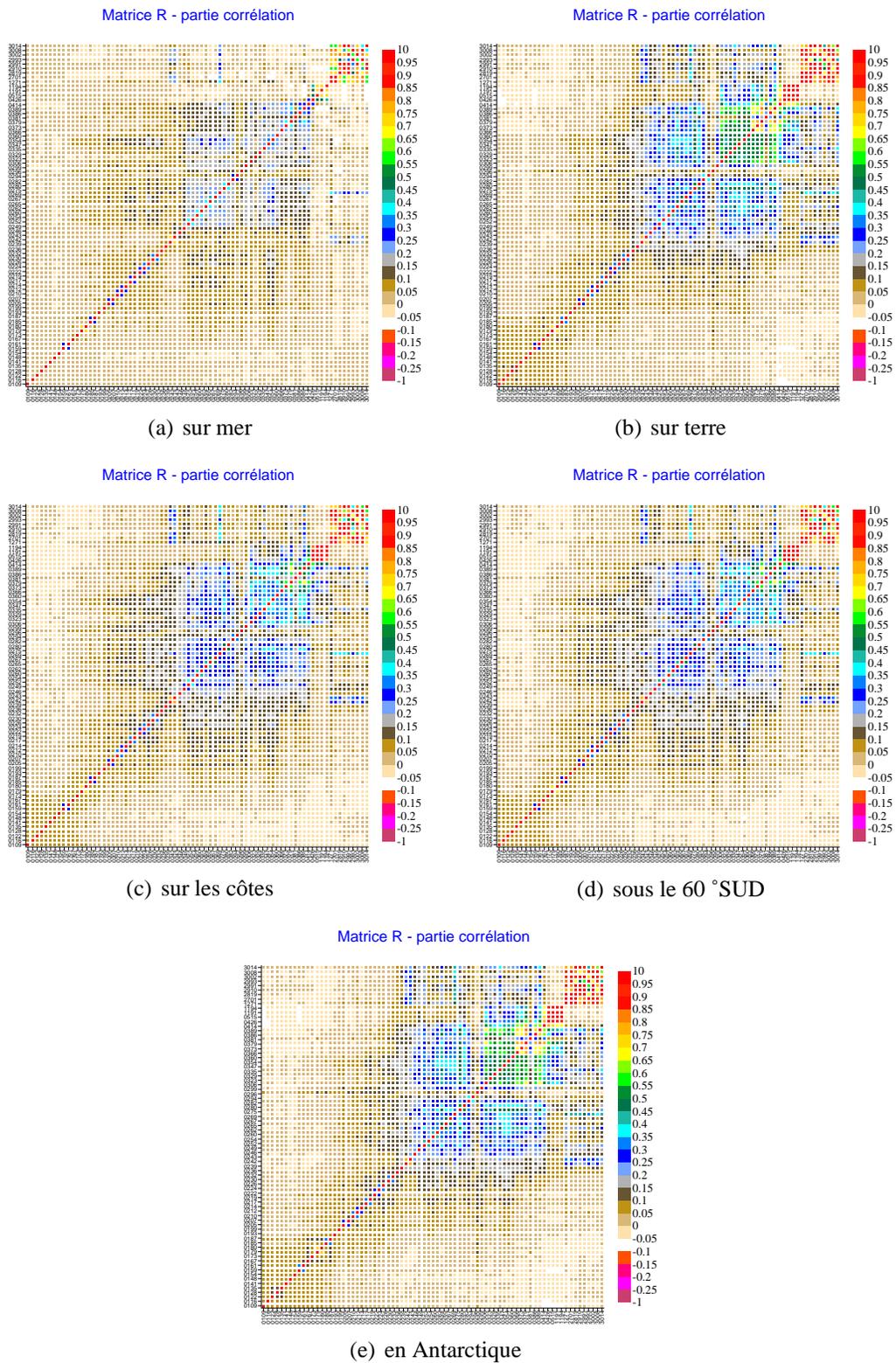


FIG. 3.5 – Images des matrices symétriques obtenues pour décembre 2009

Chapitre 4

EVALUATION DE L'IMPACT DES MATRICES DIAGNOSTIQUES SUR L'ANALYSE

4.1 Le 1D-VAR

Le 1D-VAR est une forme simplifiée du 4D-VAR. C'est un outil fourni par le Met Office dans le cadre d'un programme d'EUMETSAT. Il permet de tester rapidement d'éventuelles modifications à apporter aux entrées du modèle opérationnel pour en améliorer les résultats. A la différence du 4D-VAR, l'assimilation ne se fait pas sur toute l'atmosphère, mais en un lieu donné sur une colonne verticale ; la dimension horizontale n'est pas prise en compte. L'utilisation du 1D-VAR pour tester les matrices de covariance d'erreurs d'observation nécessite donc de le faire en un lieu précis. De plus l'aspect temporel n'est pas non plus pris en compte, les observations utilisées en entrée du 1D-VAR doivent donc être simultanées pour que leur assimilation ait un sens météorologique. Les données d'entrée du 1D-VAR sont les suivantes : la matrice B est celle utilisée en opérationnel (reformatée pour être entrée dans le modèle 1D-VAR), les profils d'ébauche proviennent des prévisions à 6 heures du modèle ARPEGE pour les dates du 7 au 17 janvier 2010.

4.2 La colocalisation des données

L'évaluation de la qualité d'une matrice peut se faire en comparant les sondages de température et d'humidité restitués par l'analyse du modèle 1D-VAR et un radiosondage issu du réseau synoptique (00 et 12 UTC) qui sert de référence. La comparaison porte sur les profils de température et d'humidité. Les radiosondages effectués entre le 7 et le 17 janvier 2010 qui ont servi pour cette évaluation, ont été extraits des bases utilisées

dans les modèles ARPEGE opérationnel ou pré-opérationnel.

En entrée le modèle 1D-VAR a besoin des observations IASI faites en un lieu et à un moment donné. Pour que la comparaison soit significative, il faut que le radiosondage et l'observation IASI soient représentatifs d'une même masse d'air. Les critères retenus sont les suivants :

- une distance entre le radiosondage et l'observation IASI inférieure à 50 km
- un écart temporel de moins de 30 minutes
- les critères correspondants à la surface traitée devant être réalisés pour le radiosondage et l'observation IASI.

Le programme de colocalisation (FORTRAN) réalisé a permis d'obtenir pour les surfaces mer et terre, en ciel clair et entre 60°Nord et 60°Sud, quelques concordances : 16 pour la mer et 6 pour la terre. Ce petit nombre est dû à des conditions fortes sur la colocalisation. Les points sur terre obtenus sont essentiellement côtiers, leur petit nombre est passé de 6 à 16 en élargissant la distance à 100 km. Ces conditions de colocalisation élargies mais encore assez fortes. Autoriser des distances supérieures se heurte, sur terre, au changement de type de surface dont l'influence sur le bilan radiatif est loin d'être négligeable, on pourrait alors élargir le critère d'écart maximum de temps entre un radiosondage et une mesure IASI pour obtenir plus de points. Les dix points supplémentaires colocalisés sont à nouveau des points côtiers, et les résultats n'ont pas apporté d'informations supplémentaires. Les points obtenus ont été cartographiés grâce à l'outil Metview.(cf. Figure 4.1 et Figure 4.8)

Une fois que l'ensemble des données a été récupéré et mis au format pour être traité dans le 1D-VAR, les simulations ont été lancées.

4.3 L'estimation de l'efficacité de la matrice

L'efficacité de la matrice diagnostique d'une surface donnée, est évaluée en effectuant, pour chaque point de concordance défini au paragraphe 4.2 entre un radiosondage et une mesure IASI, trois analyses avec la modèle 1D-VAR. Pour chacune de ces analyses, seule la matrice R de covariance des mesures des 77 canaux IASI assimilés change, les trois matrices R utilisées sont les suivantes :

- la matrice diagnostique symétrisée non seuillée
- la matrice diagonale utilisée en assimilation opérationnelle
- la matrice diagonale contenant les σ_o^2 de la matrice diagnostique symétrisée non seuillée

Cela permet de confronter les sondages verticaux issus de l'analyse 1D-VAR faite avec les réglages de l'opérationnel d'ARPEGE à ceux de l'analyse faite avec la matrice diagnostique, et de les comparer au radiosondage. La matrice diagonale de l'opérationnel contient des coefficients volontairement surestimés. A cela deux raisons : la première

est qu'il faut garder un équilibre entre l'ébauche et les observations, une petite valeur σ_o^2 entraînerait une prépondérance de l'observation sur l'ébauche ; la seconde est la prise en compte de façon indirecte des corrélations intercanaux qui sont *moralement* sommées sur la diagonale. Les variances de la matrice diagnostique sont significativement plus faibles que celles de la matrice opérationnelle (cf. Figure 3.1). Il convient donc de tester si les différences obtenues sur profils restitués avec chacune des matrices diagnostique et opérationnelle sont imputables à la réduction des variances pour l'essentiel, ou à la prise en compte des corrélations. Ce test se fait par la comparaison des deux profils déjà décrits, avec celui du radiosondage et, celui issu de l'analyse faite avec la matrice diagonale contenant les σ_o^2 de la matrice diagnostique.

Les profils ont été tracés avec le logiciel XmGrace. Sur un même graphique (cf. Figures 4.15 à 4.30 en annexes) on a, pour un point donné, les profils verticaux de températures ou d'humidités du radiosondage, de l'ébauche et des analyses faites avec les trois matrices (opérationnelle, diagnostique, diagonale des σ_o^2 diagnostiques).

4.4 Les résultats sur mer :

Les 16 profils verticaux de température sur mer retrouvés avec la matrice diagnostique sont plus proches des radiosondages correspondants que les profils retrouvés avec la matrice opérationnelle. La matrice diagnostique permet donc d'avoir en sortie du 1D-VAR des résultats plus proches d'une réalité indépendante dans la zone maritime. La tropopause est mieux placée, l'inversion de température est plus précise (quand il y en a une).

Concernant les 16 profils verticaux d'humidité, les résultats sont variables : certains profils retrouvés avec la matrice diagnostique sont meilleurs que ceux retrouvés avec la matrice opérationnelle, surtout en basses couches (cf. Figure 4.2 et Figure 4.3) et d'autres sont par contre moins bons (cf. Figure 4.18 et Figure 4.25 en annexes).

On peut également noter que sur certains points (cf. Figure 4.2), le profil retrouvé avec la matrice opérationnelle marquait un pic d'humidité à 200 hPa qui était absent du radiosondage, le profil retrouvé avec la matrice diagnostique marque toujours ce pic mais il a une amplitude nettement moins importante.

Le nombre de données n'est pas suffisant pour tirer des conclusions définitives sur l'efficacité de la matrice diagnostique pour les profils verticaux sur mer. Néanmoins c'est suffisant pour poursuivre les investigations, les résultats sont encourageants.

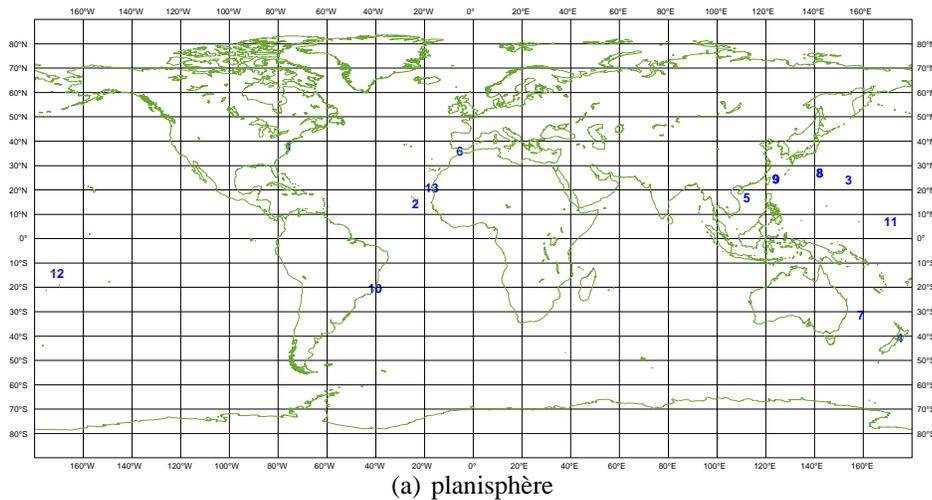
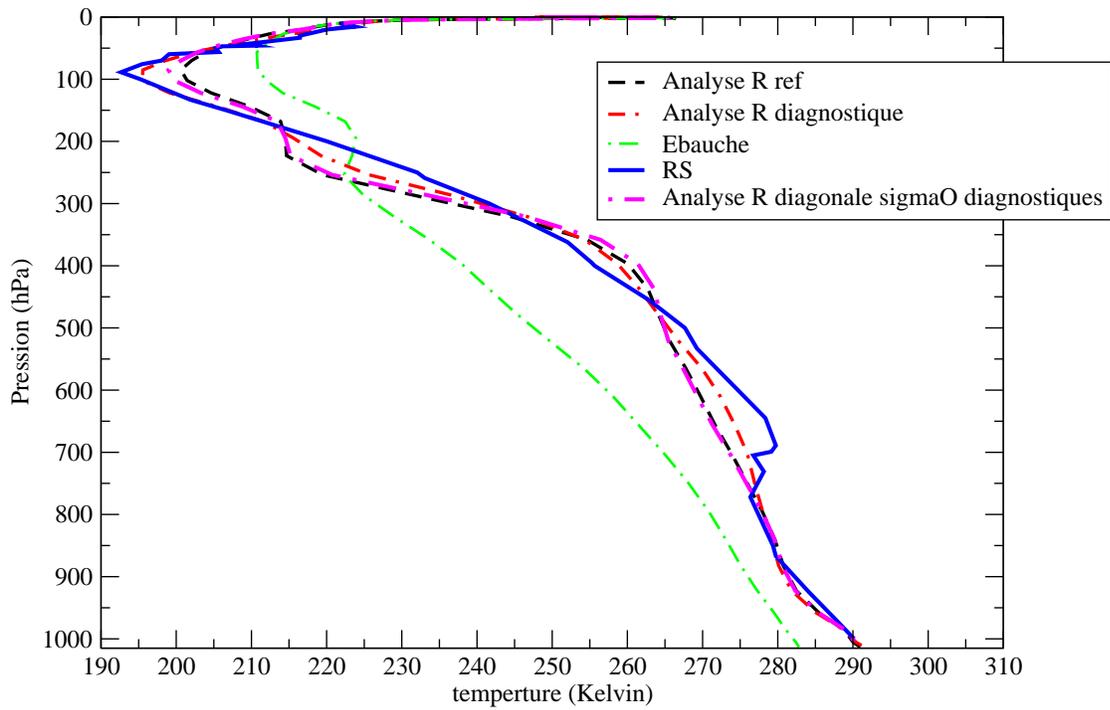
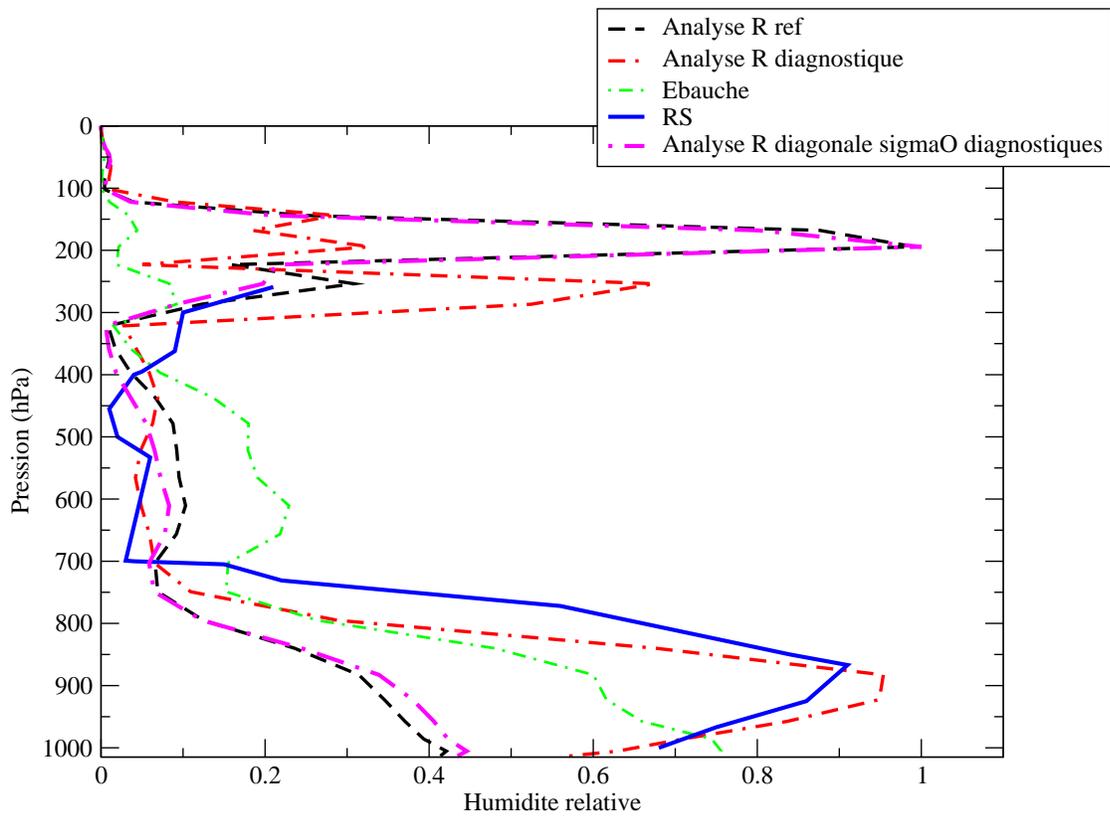


FIG. 4.1 – Planisphère de localisation des radiosondages colocalisés sur mer



(a) profil vertical de température



(b) profil vertical d'humidité

FIG. 4.2 – du 17/01/2010 à 12 UTC sur les Iles Hatoma noté 9 sur la Figure 4.1

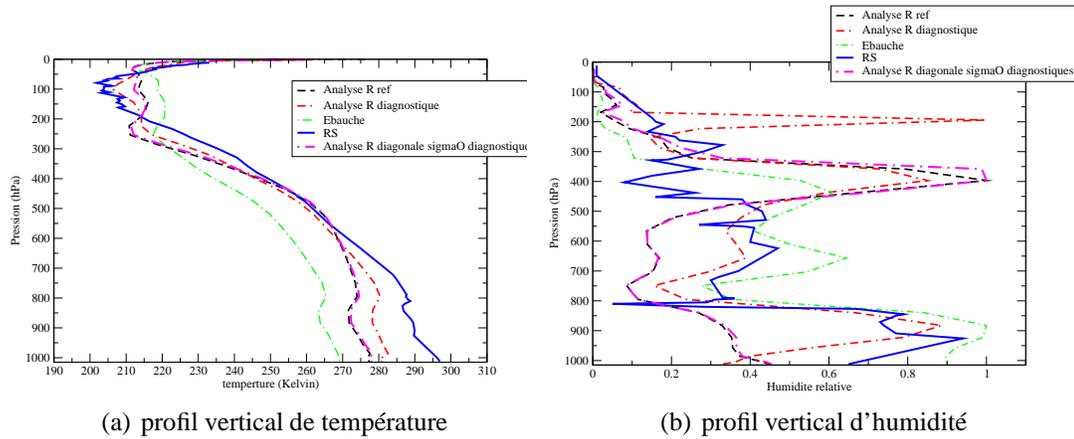


FIG. 4.3 – du 10/01/2010 à 00 UTC sur l'Ile Lord Howe noté 7 sur la Figure 4.1

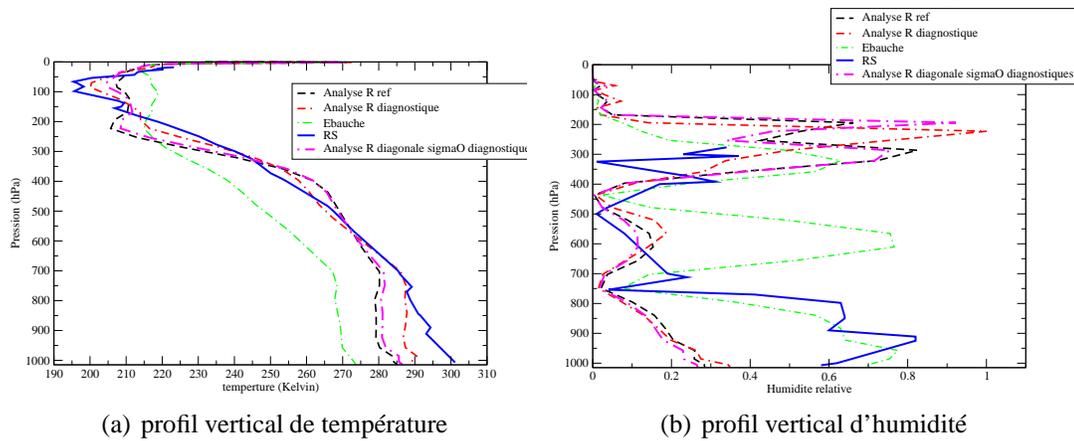


FIG. 4.4 – du 11/01/2010 à 12 UTC à Espirito Santo noté 10 sur la Figure 4.1

En comparant les profils verticaux retrouvés à partir des matrices diagnostiques et diagonales contenant les σ_o diagnostiques, il est clair que les améliorations sont pour l'essentiel dues aux coefficients de corrélation intercanaux.(cf. Figures 4.15 à 4.30 en annexes)

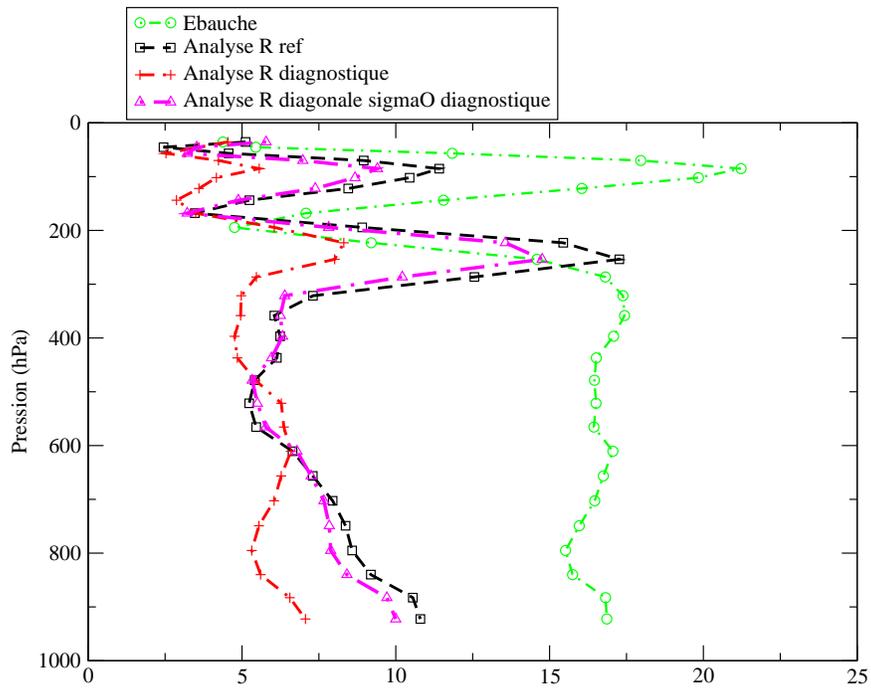
Quelques statistiques ont été réalisées malgré le faible nombre de données. Des scores, ici les RMS¹ des écarts entre les différents profils verticaux restitués et le radiosondage pour les 16 points ont été calculés, ainsi que les moyennes de ces mêmes écarts qui évaluent le biais des profils restitués. Cela a nécessité une interpolation linéaire des valeurs des RS sur les niveaux du modèle, les valeurs statistiques obtenues ne sont donc qu'indicatives. Ces statistiques ont été calculées entre les niveaux de 922 hPa (pour ne pas subir les influences orographiques car les points obtenus sont situés sur de petites îles) et 30 hPa pour les profils de température et 300 hPa pour les profils d'humidité (30 hPa et 300 hPa ont été imposées par les données des radiosondages). Les résultats soulignent que la prise en compte des corrélations intercanaux améliore nettement l'analyse (cf. Figures 4.5 et 4.6).

Pour ce qui est du biais, sur la zone maritime, on remarque une légère sous-estimation de la température et une surestimation de l'humidité en basses couches et à partir de 600 hPa. Les résultats sont tout de même meilleurs qu'avec la matrice opérationnelle actuelle.

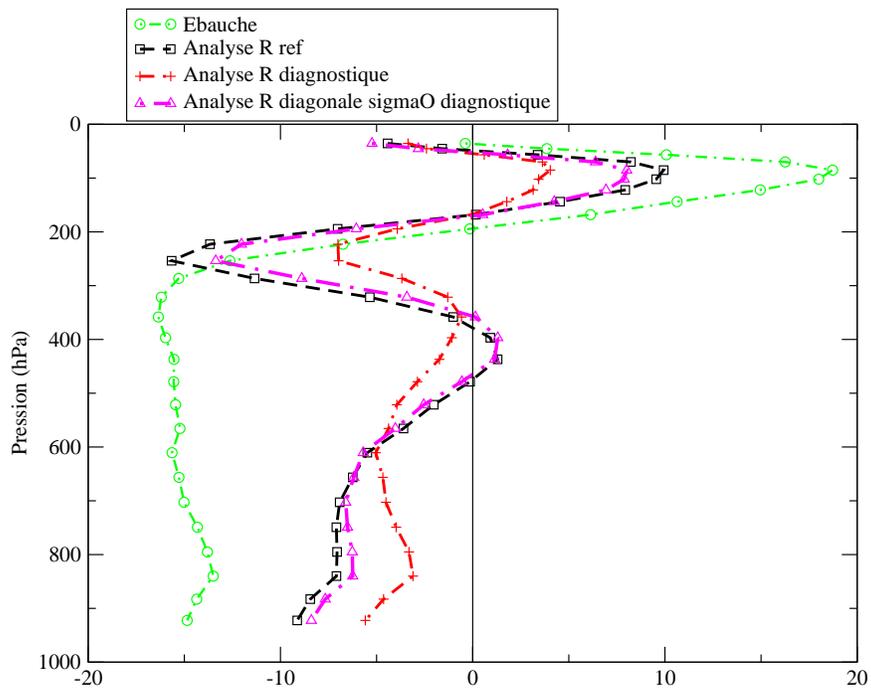
¹Root Mean Square Error

Le rôle des corrélations par canal

Les résultats statistiques sur les températures de brillance montrent que (cf. Figure 4.7 les RMS courbe (a)), contrairement aux suppositions faites en début d'étude, malgré leurs faibles valeurs, les corrélations des premiers canaux (n°109 à 236) jouent un rôle important dans la restitution des températures de brillance lors de l'analyse. Pour toutes les matrices, le signe du biais des températures de brillance restituées par l'analyse (cf. Figure 4.7) est le même pour chacun des canaux, à un ou deux près. On constate tout de même que l'analyse utilisant la matrice diagnostique réduit significativement ce biais en valeur absolue (la courbe "Analyse R diagnostique" est proche de l'axe des abscisses), en particulier pour les canaux n° 109 à 187 et pour le canal n° 3002. Pour ces derniers canaux une réduction de 50 % sur le biais obtenu avec la matrice opérationnelle est imputable à la réduction des valeurs des σ_o , comme le montre la courbe "Analyse R diagonale sigmaO diagnostique", mais la courbe "Analyse R diagnostique" montre un facteur en général très inférieur à 75 %, c'est la contribution des corrélations intercanaux qui réduit significativement le biais. Le biais sur les températures de brillance des canaux n° 199 à 242 est lui aussi fortement réduit, en valeur absolue, mais là, la réduction est uniquement dûe aux corrélations.

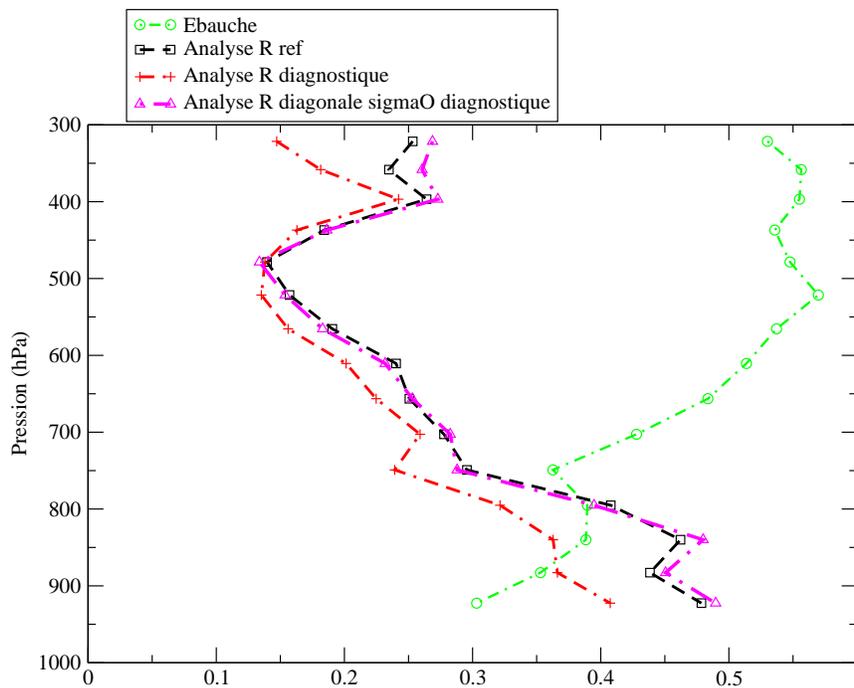


(a) RMS des écarts

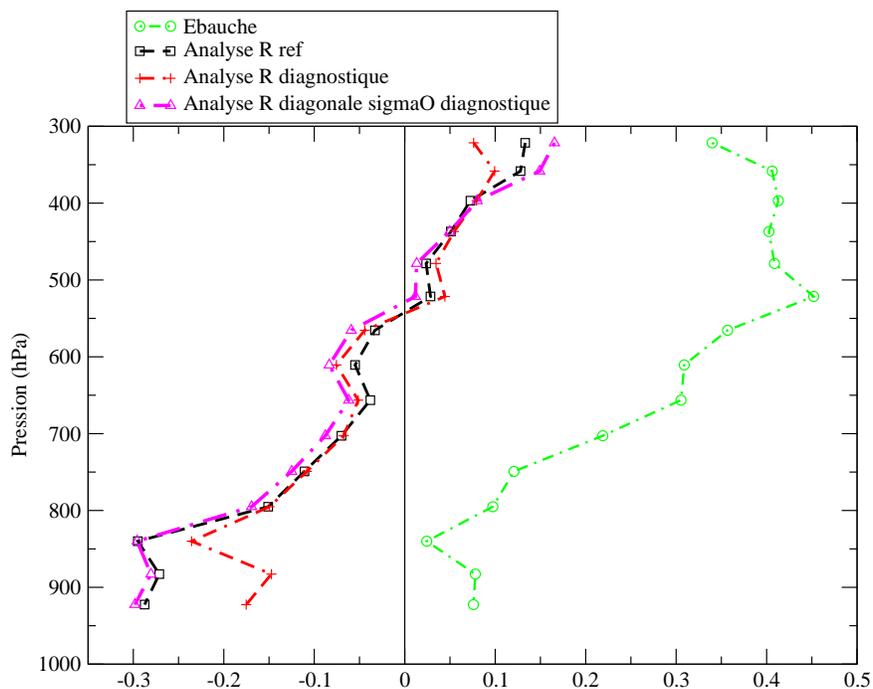


(b) moyenne des écarts (biais)

FIG. 4.5 – Statistiques en température (en K) par niveaux de pression sur mer

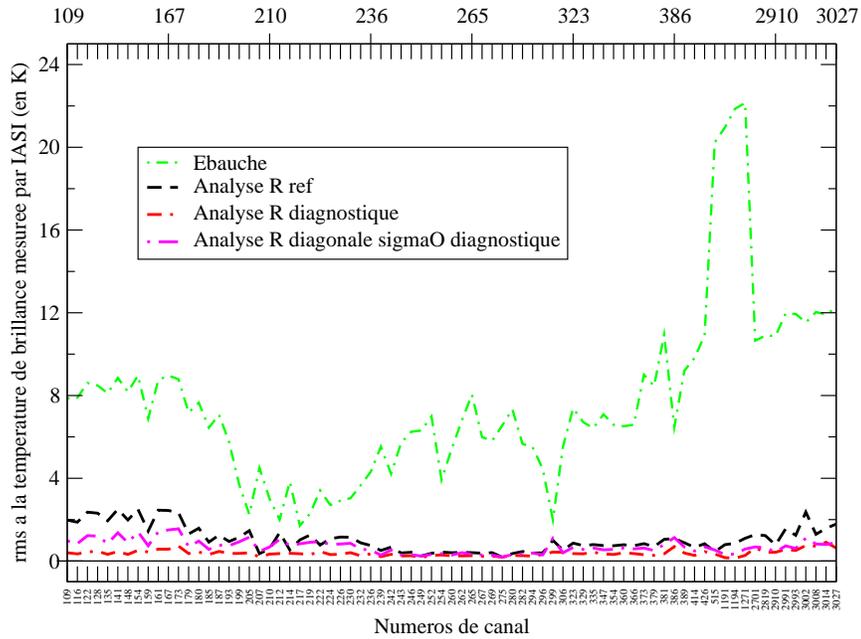


(a) RMS des écarts

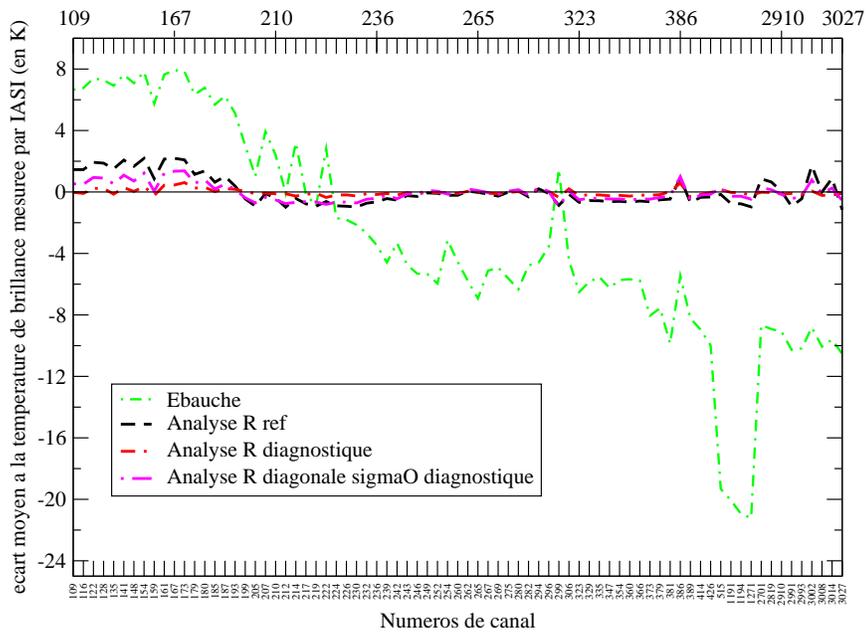


(b) moyenne des écarts (biais)

FIG. 4.6 – Statistiques en parts d'humidité par niveaux de pression sur mer



(a) RMS



(b) biais

FIG. 4.7 – Ecart entre les températures de brillance simulées lors du 1Dv-VAR et les températures de brillance mesurées par IASI pour chaque canal, sur les 16 valeurs

4.5 Les résultats sur terre

Les 6 profils verticaux de température et d'humidité sur terre (cf. Figures 4.8 à 4.11) retrouvés avec la matrice diagnostique ne donnent pas d'améliorations notables. Ceci peut s'expliquer par :

- les différentes natures que peuvent avoir les sols sur les zones terrestres et qui ont une influence sur les profils verticaux aussi bien en température qu'en humidité.
- le fait que l'ensemble des données IASI sont situés aux abords des côtes et que les radiosondages associés aient pu avoir une trajectoire déviant vers la mer.

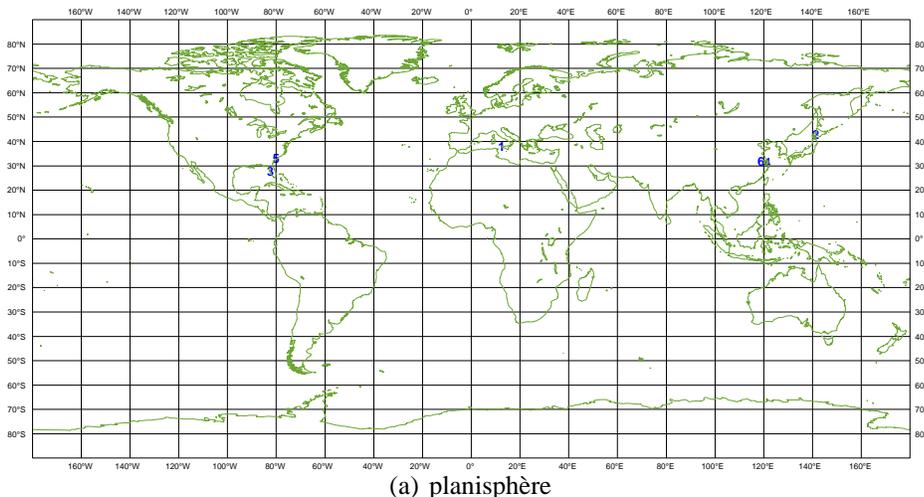


FIG. 4.8 – Planisphère de localisation des radiosondages colocalisés sur terre

Le nombre de points de concordances entre un radiosondage et une mesure IASI est petit : 6. Pour avoir d'avantage de points le critère de distance a été augmenté de 50 km il est passé à 100 km, 10 points supplémentaires ont été relevés, mais toujours près des côtes. Les profils restitués avec ces nouveaux points sont semblables à ceux obtenus avec les 6 points originaux. Cela ne permet toutefois pas une évaluation correcte de la matrice diagnostique "terre" puisque le nombre de points est relativement faible, mais aussi et surtout parce que se sont des points côtiers. Pour évaluer la matrice terre, il faudrait des points situés à l'intérieur des terres. On pourrait élargir le jeu des données qui sert à relever les points concordances pour les tests, et n'autoriser que les points situés à 50 km à l'intérieur des terres par exemple.

Le résultat de l'évaluation de la matrice diagnostique "terre", n'étant pas concluant, et les points retenus pour cette évaluation étant côtiers, ils ont été réutilisés pour une analyse avec la matrice diagnostique "mer". Cette analyse n'a pas non plus apporté ap-

porté d'amélioration sur les profils restitués par l'analyse (cf. Figure 4.12). On constate même que les valeurs fournies par les deux analyses sont égales. Ceci appelle deux expériences qu'il faudrait réaliser :

1. utiliser ces points côtiers pour évaluer la matrice diagnostique "côte"²
2. utiliser la matrice diagnostique "terre" pour analyser les points utilisés pour l'évaluation de la matrice diagnostique "mer", cela afin de comparer les deux matrices.

Les courbes obtenues sur terre semblent montrer que la matrice diagnostique n'améliore pas l'analyse, les profils restitués sont même moins bons que ceux obtenus avec la matrice opérationnelle.

Les mêmes statistiques que pour les résultats sur mer ont été calculées, mais sur 6 points cela ne permet en aucun cas de valider ces résultats.

Les résultats RMS confirment que la matrice diagnostique semble donner de moins bons résultats que la matrice opérationnelle. Les moyennes montrent que la température est en général sous-estimée au dessus de 400 hPa et légèrement surestimée en altitude, et que l'humidité est très légèrement surestimée au dessus de 700 hPa, mais que ceci semble surtout être lié aux faibles valeurs de σ_o plutôt qu'aux coefficients de corrélation intercanaux (cf. Figures 4.13 et 4.14).

² on peut s'attendre à des résultats analogues à ceux obtenus avec les matrices diagnostiques "terre" et "mer", puisque les corrélations diagnostiques des erreurs intercanaux qui apparaissent pour les côtes sont intermédiaires entre celles pour la mer et pour la terre

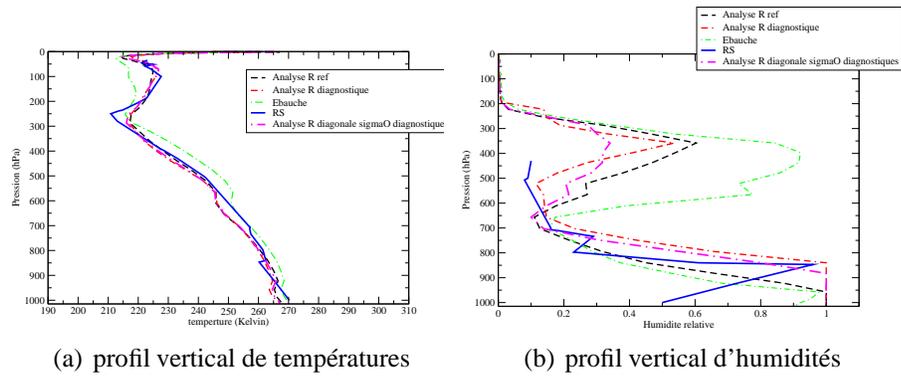


FIG. 4.9 – du 08/01/2010 à 00 UTC à Sapporo noté 2 sur la Figure 4.8

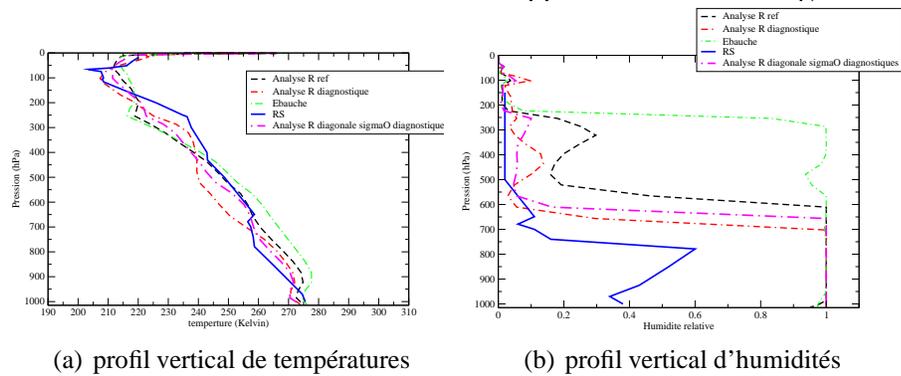


FIG. 4.10 – du 12/01/2010 à 12 UTC à Shanghai noté 4 sur la Figure 4.8

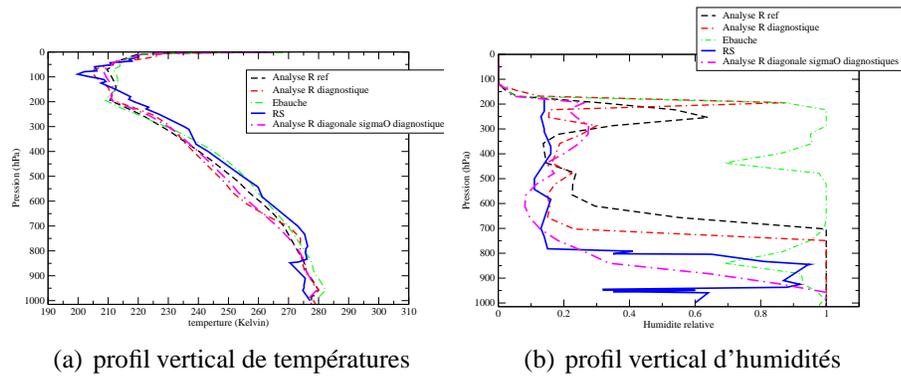
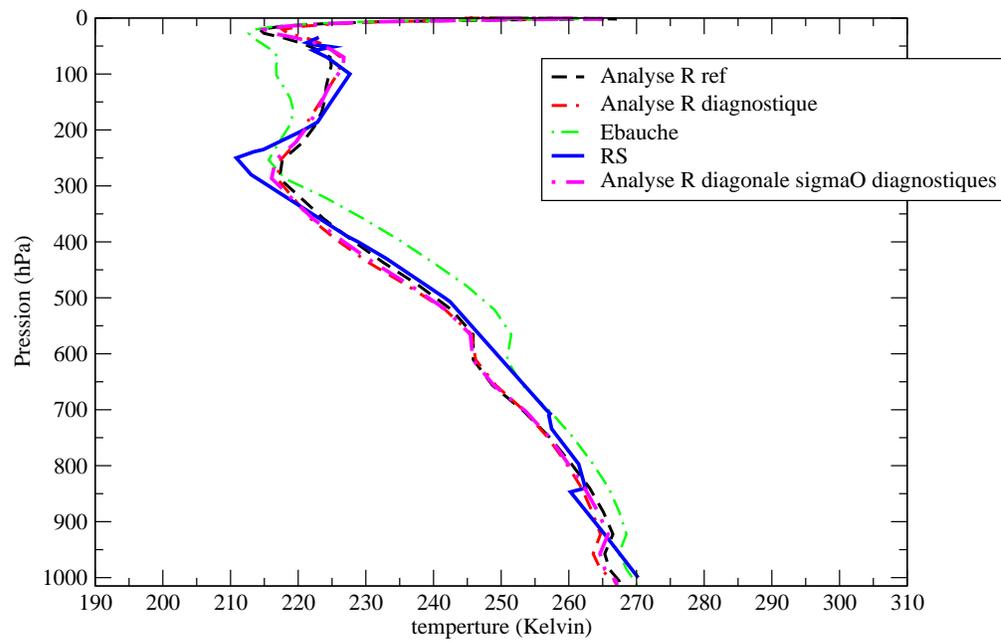
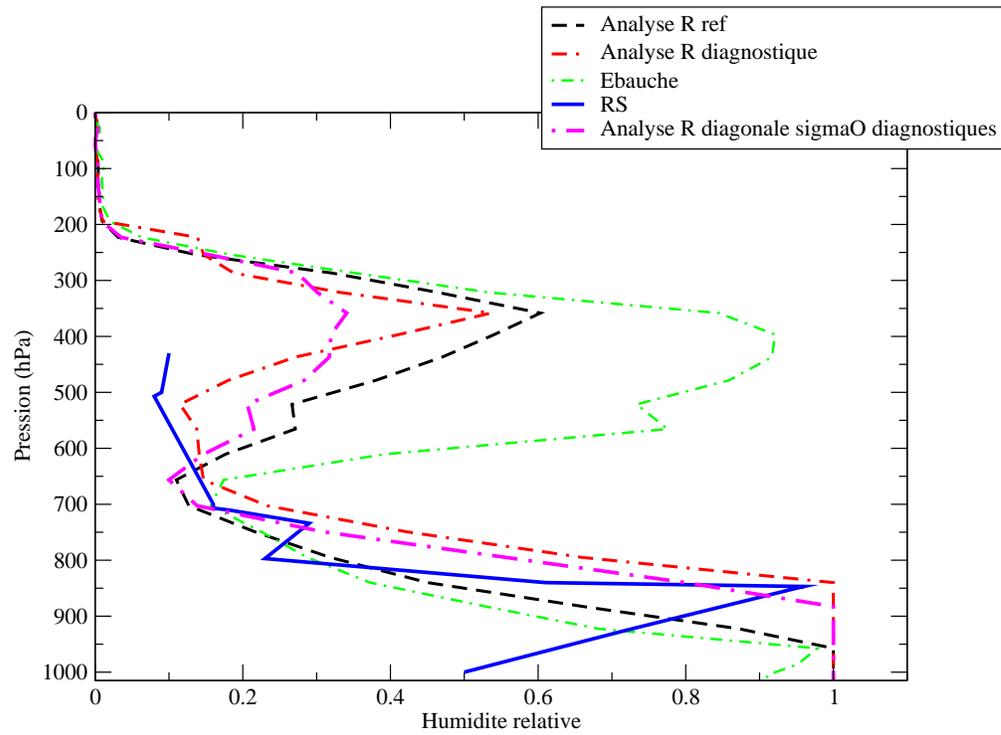


FIG. 4.11 – du 16/01/2010 à 12 UTC à Nanjing noté 6 sur la Figure 4.8

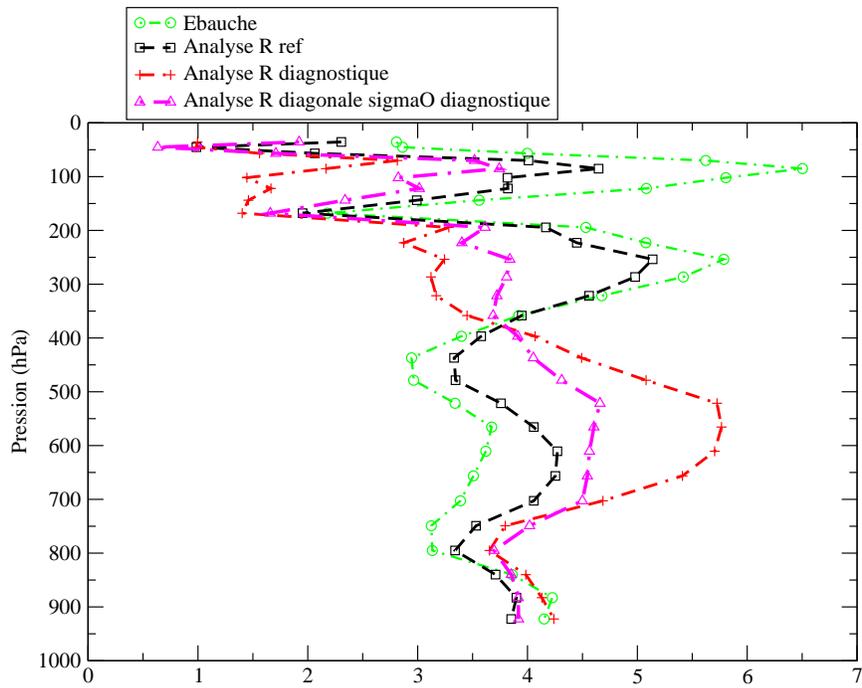


(a) profil vertical de températures

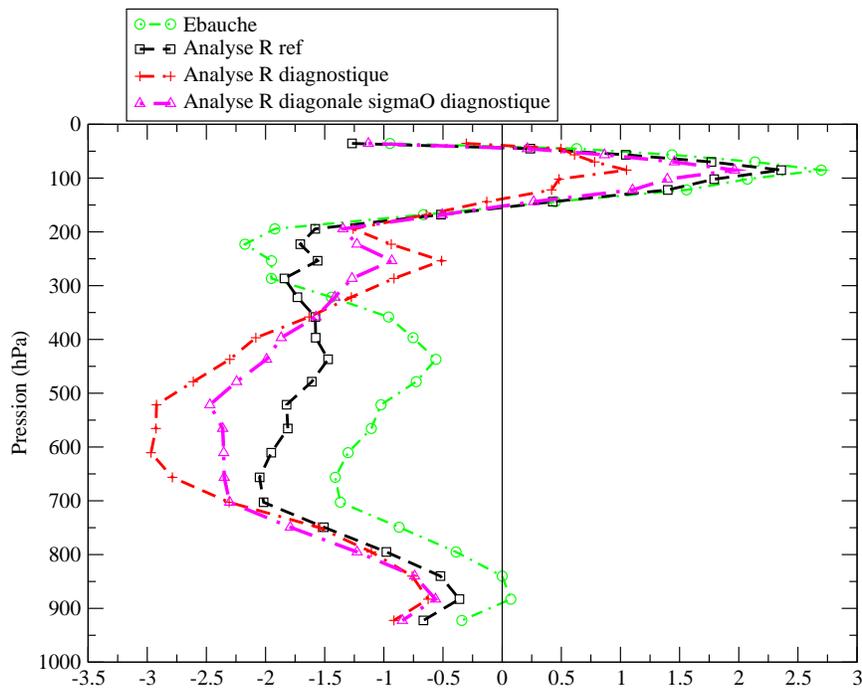


(b) profil vertical d'humidités

FIG. 4.12 – du 08/01/2010 à 00 UTC

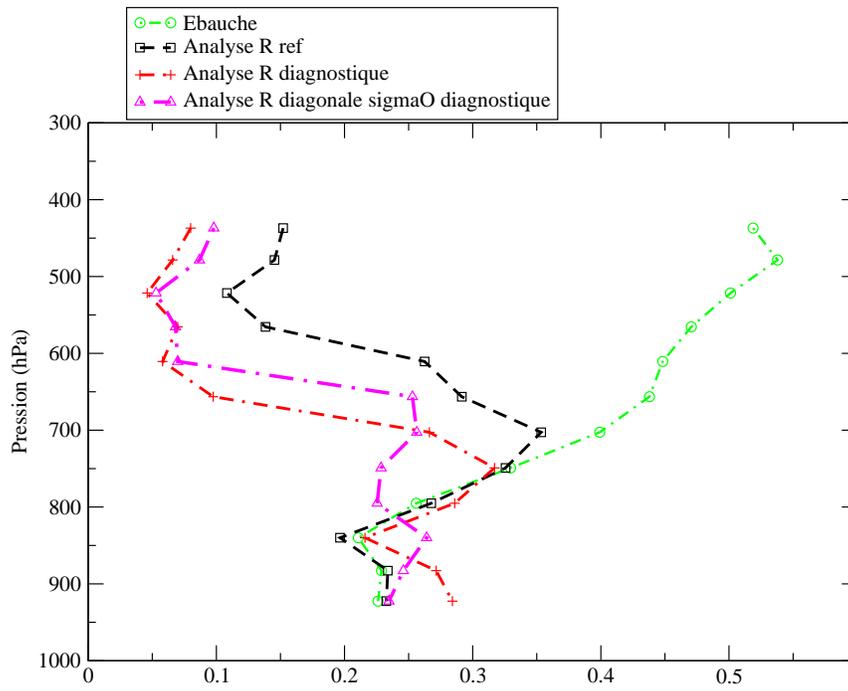


(a) RMS des écarts

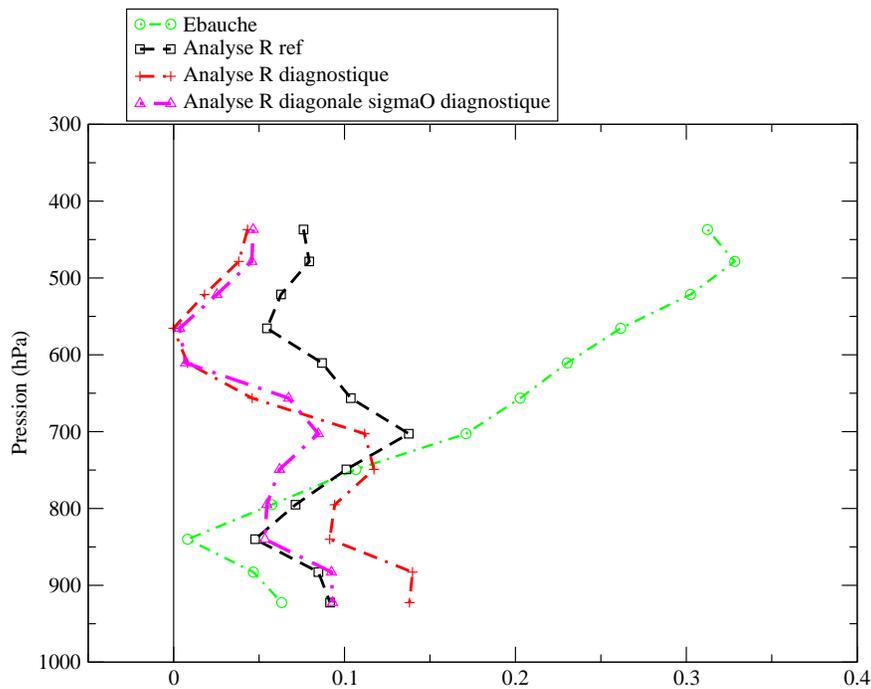


(b) moyenne des écarts (biais)

FIG. 4.13 – Statistiques en température (en K) par niveaux de pression sur terre



(a) RMS des écarts



(b) moyenne des écarts (biais)

FIG. 4.14 – Statistiques en parts d'humidité par niveaux de pression sur terre

CONCLUSIONS ET PERSPECTIVES

La méthode de Desroziers nous a permis d'arriver à des résultats qui semblent pour l'instant prometteurs. En effet, les résultats sur mer, en ciel clair, prouvent l'importance de la prise en compte des coefficients de corrélation intercanaux. Cette étude doit maintenant se poursuivre sur le reste de la surface terrestre où la nature des sols est différente (déserts, mers de glace, reliefs...) et également dans les cas nuageux.

Une amélioration de la matrice diagnostique pourrait être obtenue en utilisant la méthode de Desroziers sur les analyses faites avec la matrice diagnostique utilisée pour cette étude, l'idée étant d'obtenir un résultat par itérations successives. Un essai a été réalisé avec les résultats sur mer pour les 16 analyses réalisées, mais la matrice fournie était grossièrement mauvaise puisque les corrélations atteignaient des valeurs allant jusqu'à 64 et des variances négatives. Cela peut tenir au faible nombre de points de mesure : 16 (contre 162197 pour la matrice de l'étude). Cela peut aussi venir du fait que la méthode itérative diverge, est mal conditionnée ou que les 2 systèmes (1D-VAR et 4D-VAR) n'ont pas les mêmes caractéristiques. Il faudrait donc procéder aux itérations dans le 4D-VAR ARPEGE pour prendre en compte toutes les observations alors que le 1D-VAR ne prend en compte que les données satellitaires des 77 canaux.

Les simplifications éventuelles des matrices diagnostiques qui seraient (seront) créées pour l'opérationnel doivent tenir compte des corrélations qui entraînent une amélioration. Mais comme nous l'avons vu au paragraphe 4.4 ce n'est pas la valeur de la corrélation qu'il faut prendre en compte mais son impact sur la restitution des profils verticaux et des températures de brillance. Ainsi une simplification pourrait consister à annuler certaines covariances correspondant à des corrélations importantes, quitte à augmenter artificiellement les variances qui sont situées sur la même ligne et celles qui sont situées sur la même colonne, pour ne pas donner un poids trop important à l'observation faite par ce canal.

Dans l'attente des données issues d'une campagne de mesures par sondes lâchées depuis des ballons dérivant à l'altitude de 60 hPa qui permettront de confirmer les améliorations attendues sur mer avec une matrice tenant compte des coefficients de corrélation intercanaux, nous pouvons d'ores et déjà dire qu'une réflexion sur des modifications dans le code d'ARPEGE va être faite afin d'améliorer les prévisions.

BIBLIOGRAPHIE

"Diagnosis of observation, background and analysis-error statistics in observation space" de G.Desroziers, L.Berre, B.Chapnik et P.Poli (2005)

"NWPSAF Met Office 1D-Var User Manual" : manuel d'utilisation du 1D-VAR

"Estimates of spatial and inter-channel observation error characteristics for current sounder radiances for NWP" de Niels Bormann, Andrew Collard and Peter Bauer (2009)

"Le sondage satellitaire et son évolution" de T.Phulpin, F.Rabier, L.Lavanant dans La Météorologie Spécial Observation Volume II 8^{me} série - N°40 - février 2003

"Support de cours d'assimilation de données" enseigné par G.Desroziers à l'ENM

"Concordiasi pour des mesures de haut vol" de L.Charpentier dans Météo - le magazine - N°4 - décembre 2008

ANNEXES

Graphiques

Ci-dessous, les profils verticaux sur mer entre les 07 et 17 janvier 2010 :

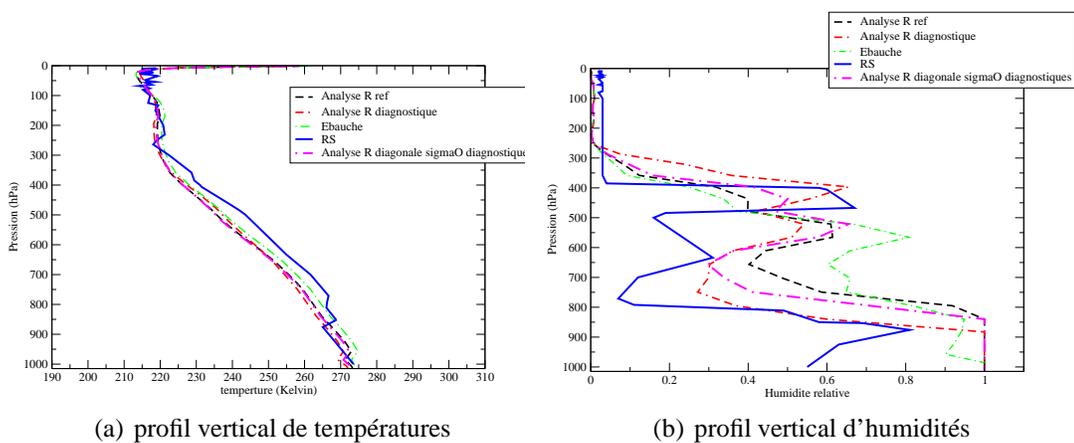


FIG. 4.15 – du 07/01/2010 à 00 UTC sur la côte Atlantique des USA noté 1 sur la Figure 4.1

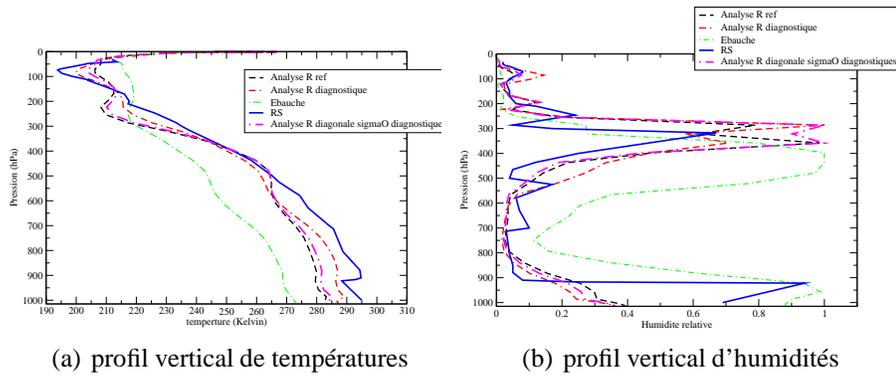


FIG. 4.16 – du 07/01/2010 à 12 UTC sur l'Atlantique noté 2 sur la Figure 4.1

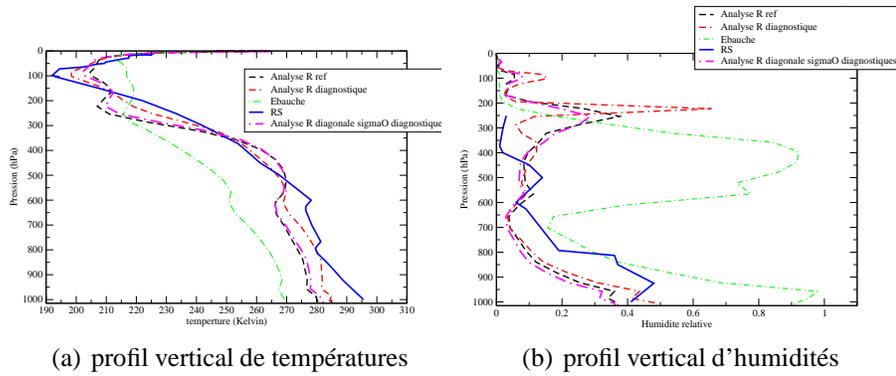


FIG. 4.17 – du 08/01/2010 à 00 UTC sur le Pacifique noté 3 sur la Figure 4.1

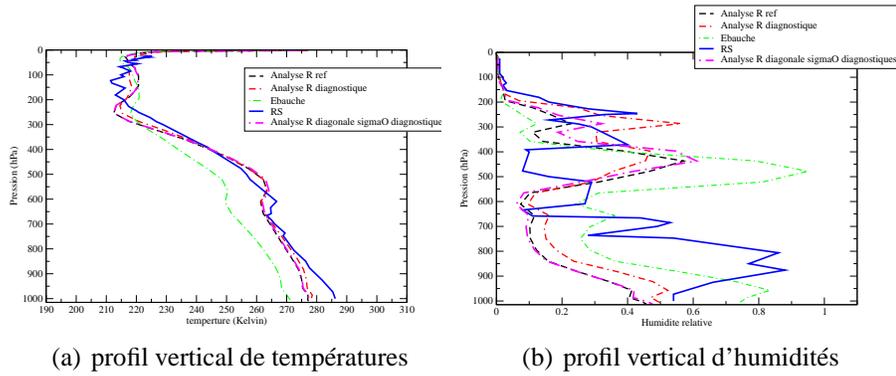


FIG. 4.18 – du 08/01/2010 à 12 UTC en Nouvelle Zélande noté 4 sur la Figure 4.1

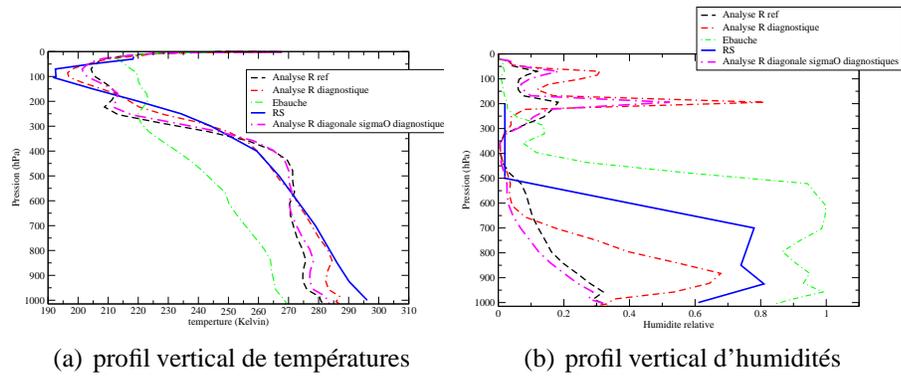


FIG. 4.19 – du 09/01/2010 à 00 UTC sur les Iles Paracel noté 5 sur la Figure 4.1

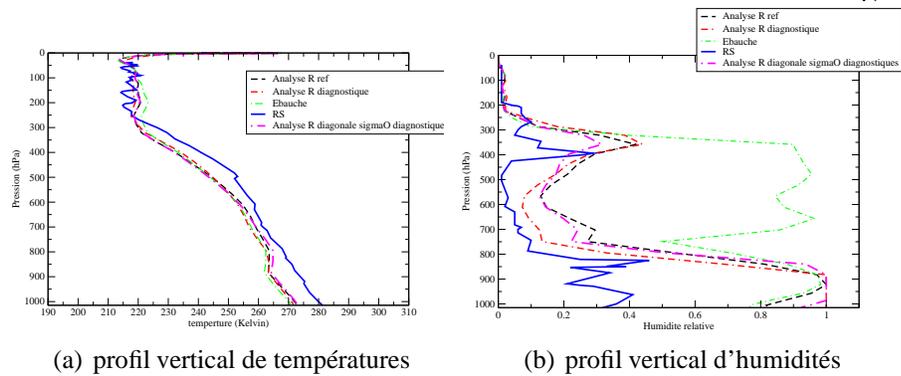


FIG. 4.20 – du 09/01/2010 à 12 UTC à Ceuta noté 6 sur la Figure 4.1

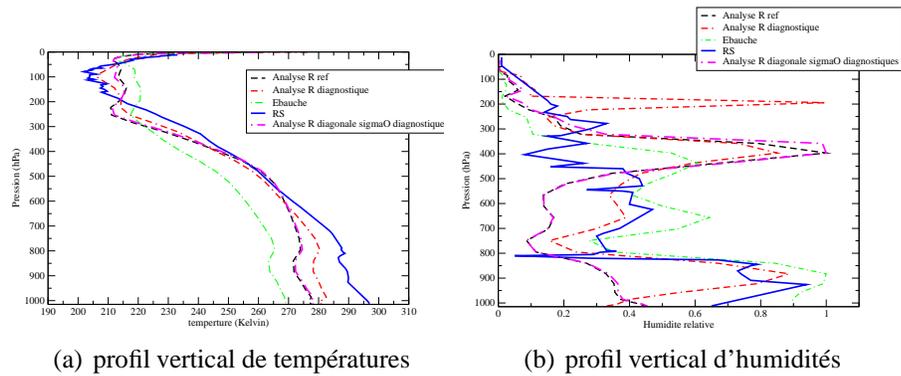


FIG. 4.21 – du 10/01/2010 à 00 UTC sur l'Ile Lord Howe noté 7 sur la Figure 4.1

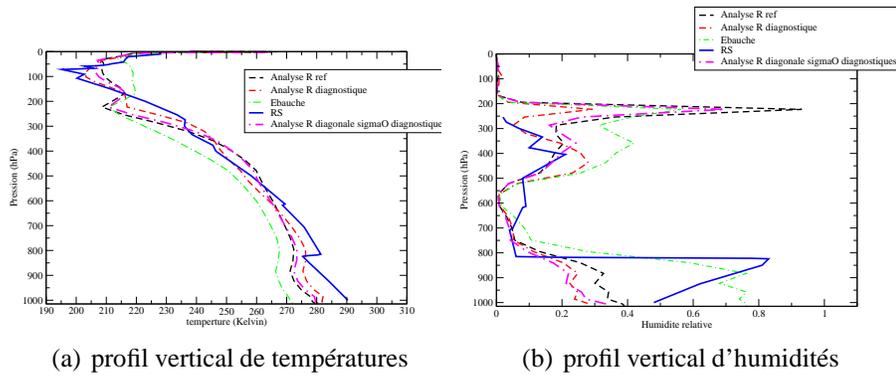


FIG. 4.22 – du 10/01/2010 à 12 UTC à Ogasawara noté 8 sur la Figure 4.1

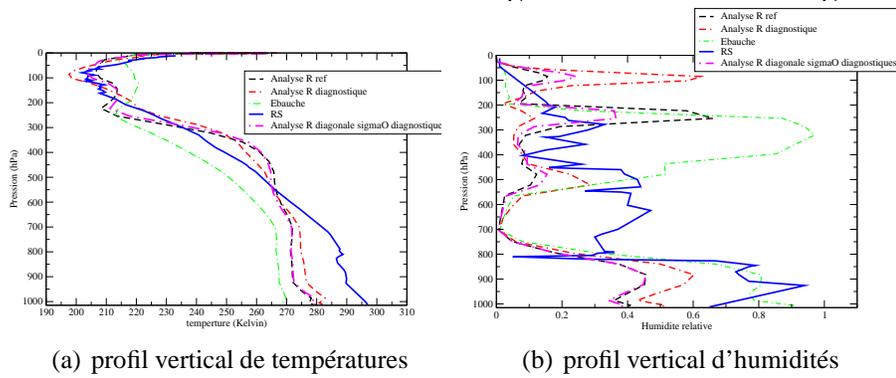


FIG. 4.23 – du 11/01/2010 à 00 UTC sur les Iles Hatoma noté 9 sur la Figure 4.1

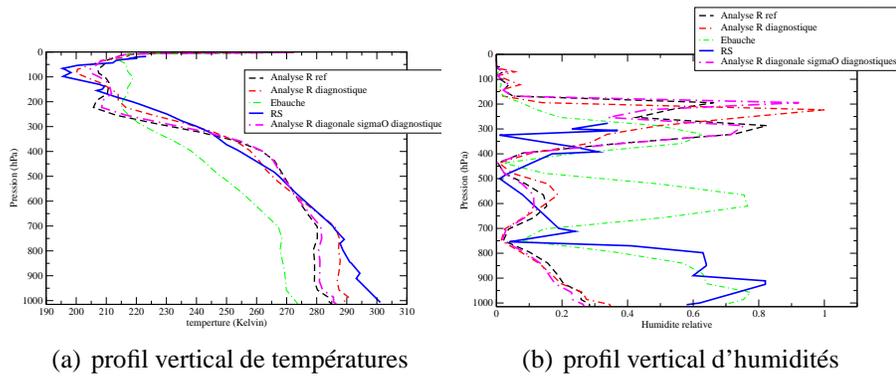


FIG. 4.24 – du 11/01/2010 à 12 UTC à Espirito Santo noté 10 sur la Figure 4.1

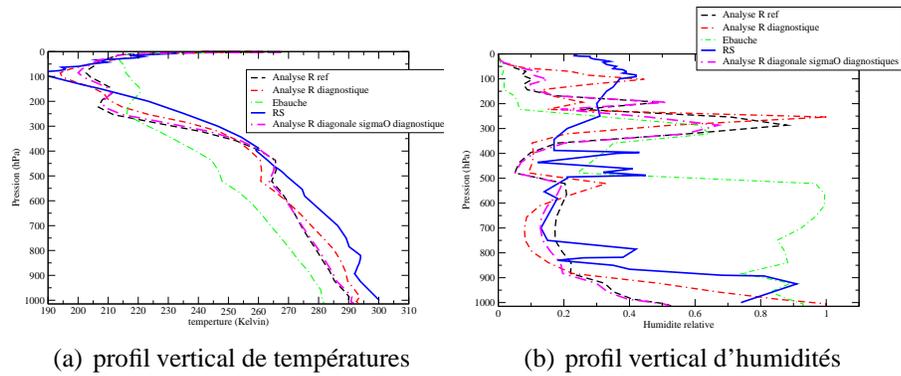


FIG. 4.25 – du 13/01/2010 à 12 UTC sur les Iles Marshall noté 11 sur la Figure 4.1

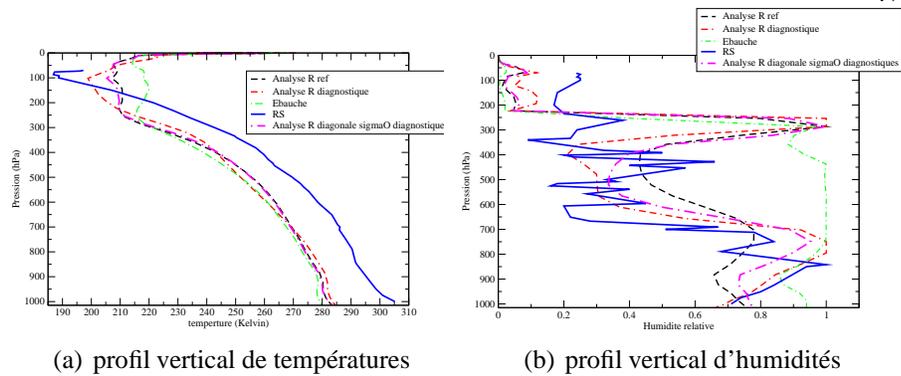


FIG. 4.26 – du 14/01/2010 à 00 UTC sur les Iles Samoa noté 12 sur la Figure 4.1

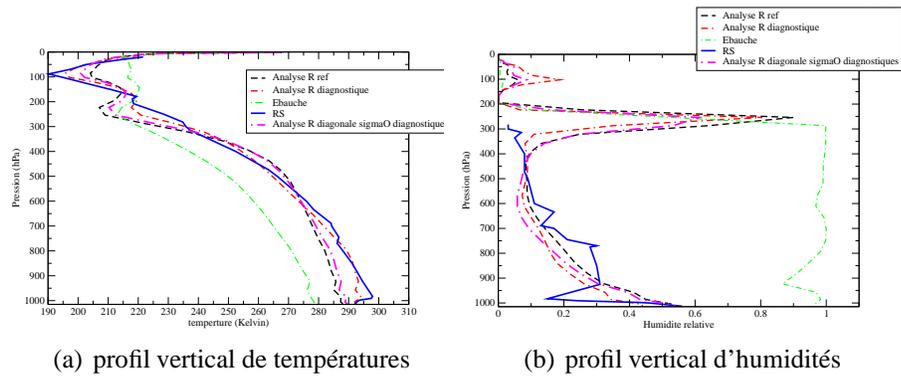


FIG. 4.27 – du 14/01/2010 à 12 UTC sur l'Atlantique noté 13 sur la Figure 4.1

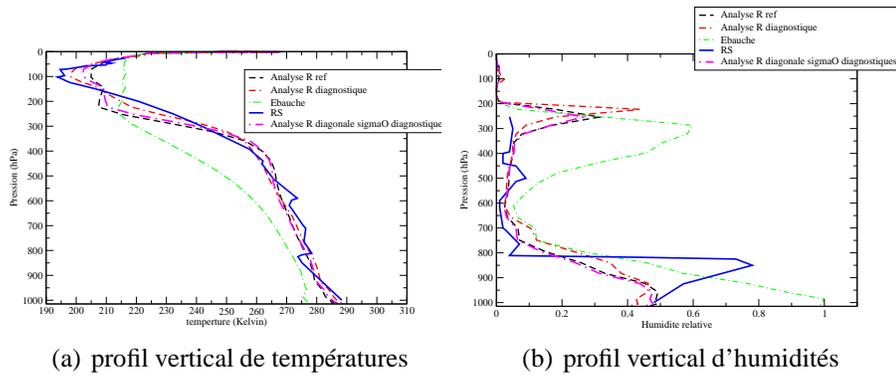


FIG. 4.28 – du 15/01/2010 à 00 UTC à Ogasawara noté 8 sur la Figure 4.1

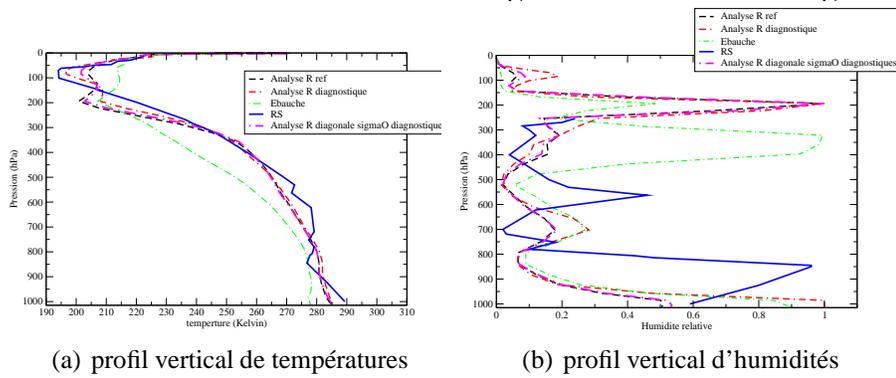


FIG. 4.29 – du 15/01/2010 à 12 UTC sur les Iles Hatoma noté 9 sur la Figure 4.1

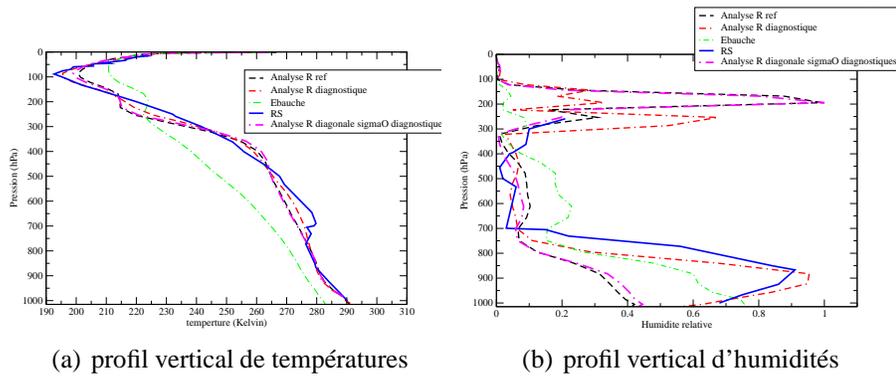


FIG. 4.30 – du 17/01/2010 à 12 UTC sur les Iles Hatoma noté 9 sur la Figure 4.1