

Vers une implantation opérationnelle du filtre de Kalman d'ensemble.

Peter Houtekamer

Martin Charron, Chantal Côté, Bin He, Herschel Mitchell, Gérard Pellerin et
Lubos Spacek

Après un travail de recherche et de développement de huit ans, nous sommes presque rendus à la première livraison du filtre de Kalman d'ensemble (EnKF) aux opérations. L'EnKF livrera alors l'ensemble des conditions initiales pour notre système de prévisions d'ensemble.

Je vais vous présenter où nous en sommes exactement dans notre travail de recherche et de développement.

Contenu de la présentation

- Introduction
- État de la recherche:
 - “l’ erreur du modèle”
 - corrélations verticales étroites
- Ancienne et nouvelle configurations du système de prévisions d’ensemble
- Résultats des expériences de comparaison:
 - le cycle d’assimilation d’ été
 - les prévisions d’ été
 - les problèmes du cycle d’ assimilation d’ hiver
- Conclusions

“L’ erreur du modèle”

$$\Psi_i^f(t+1) = M(\Psi_i^a) + q_i$$
$$Q \equiv \lim_{N \rightarrow \infty} \frac{1}{N-1} \sum_{i=1}^N (q_i - \bar{q}_i)(q_i - \bar{q}_i)^T$$

Ψ : L’ état du modèle,

M : Le modèle de prévision,

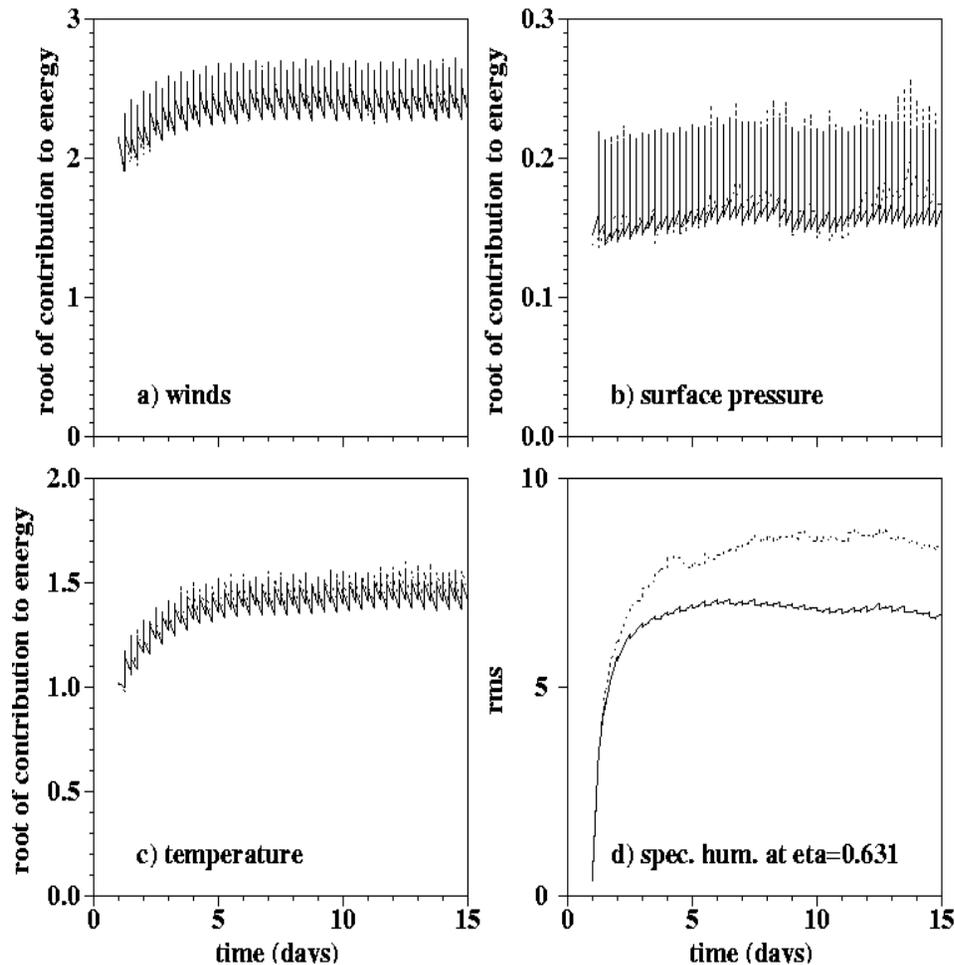
q_i : L’ erreur du modèle simulée pour le membre i ,

Q : La matrice de covariance pour “l’ erreur du modèle”.

Avec un ensemble d’ intégrations $M(\Psi_i^a)$, nous transportons les statistiques de l’ erreur d’ analyse 6 heures dans le temps. Il y a alors une évolution avec la dynamique du modèle. Pour tenir compte des différences entre le modèle et l’ atmosphère, on ajoute une perturbation q_i .

Les méthodes d’ assimilation 4D (4D-Var et EnKF) sont limitées par l’ amplitude de “l’ erreur du modèle”.

L' amplitude de "l' erreur du modèle"



En réalité, il y a un équilibre entre la croissance avec "l' erreur du modèle" et la décroissance grâce à l' assimilation des observations.

Source: "Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations".

Houtekamer, Mitchell, Pellerin, Buehner, Charron, Spacek et Hansen. En révision par Monthly Weather Review.

L'origine de "l'erreur du modèle"

Toute erreur ou incohérence qui n'est pas simulée directement dans l'EnKF va apparaître comme une "erreur du modèle".

Des sources potentielles:

- l'interpolation vers les observations (e.g. les observations sont présumées valides aux heures synoptiques),
- les statistiques des erreurs d'observation (e.g. observations indépendantes),
- les champs de surface (e.g. HS climatologique),
- la dynamique du modèle (e.g. pas de temps de 45 minutes),
- la physique du modèle (e.g. le cycle diurne),
- les interpolations entre le modèle et l'analyse (e.g. pour les vents),
- réinitialisation de certaines variables du modèle qui ne sont pas assimilées (e.g. cloud water et recalcul de GZ),
- ...

Nous ignorons l'ordre des faiblesses menant à "l'erreur du modèle". Nous pourrions commencer par faire un meilleur couplage entre le modèle et l'analyse. Le modèle ne devrait pas montrer de "spin-up" après l'analyse.

Interpolation verticale avec les polynômes cubiques de Lagrange

Avec l' EnKF, nous observons des corrélations verticales étroites en haut de 200 hPa.

Avec le modèle tangent linéaire de GEM, des problèmes similaires ont été observés près de la surface. Ces problèmes avaient alors été contrôlés en utilisant des polynômes cubiques de Lagrange pour l' interpolation verticale au lieu des splines cubiques.

Nous avons fait un cycle d' assimilation avec les polynômes cubiques de Lagrange pour le modèle GEM. Nous observons:

- 1) une réduction des oscillations verticales;
- 2) une réduction des écarts-types en haut de 200 hPa;
- 3) une amélioration de la qualité de la moyenne d' ensemble pour les vents en haut de 200 hPa;
- 4) un meilleur accord avec les observations AMSU-A.

Corrélations verticales

Après un cycle d'assimilation de deux semaines, nous observons pour la température les corrélations suivantes avec la température au niveau 6 ($\eta=0.101$).

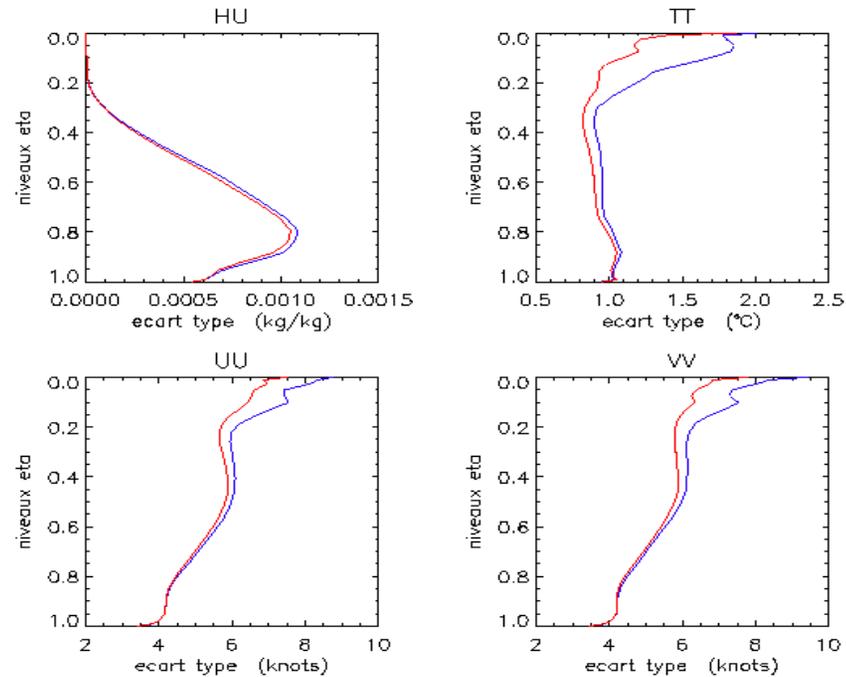
région	niveau	splines cub.	cub. Lagrange
Globe	5 ($\eta=0.075$)	-0.19	0.15
Globe	6 ($\eta=0.101$)	1.00	1.00
Globe	7 ($\eta=0.127$)	-0.01	0.34
Hémisphère Sud ($\phi < -30^\circ$)	5 ($\eta=0.075$)	-0.09	0.30
Hémisphère Sud ($\phi < -30^\circ$)	6 ($\eta=0.101$)	1.00	1.00
Hémisphère Sud ($\phi < -30^\circ$)	7 ($\eta=0.127$)	0.08	0.49
Tropiques ($-30^\circ < \phi < 30^\circ$)	5 ($\eta=0.075$)	-0.32	0.04
Tropiques ($-30^\circ < \phi < 30^\circ$)	6 ($\eta=0.101$)	1.00	1.00
Tropiques ($-30^\circ < \phi < 30^\circ$)	7 ($\eta=0.127$)	-0.10	0.23
Hémisphère Nord ($\phi > 30^\circ$)	5 ($\eta=0.075$)	-0.04	0.23
Hémisphère Nord ($\phi > 30^\circ$)	6 ($\eta=0.101$)	1.00	1.00
Hémisphère Nord ($\phi > 30^\circ$)	7 ($\eta=0.127$)	0.12	0.42

L'usage des polynômes cubiques de Lagrange semble filtrer des oscillations en haut de 200 hPa.

Déviations standard

Référence: interpolation avec splines cubiques

Expérience: avec les polynômes cubiques de Lagrange



P0 1.16880 (hPa)

P0 1.16300 (hPa)

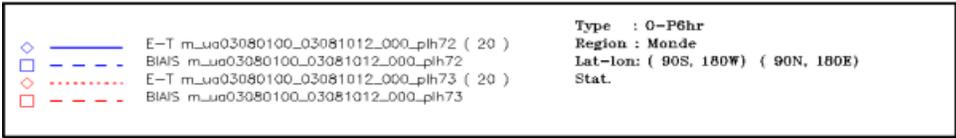
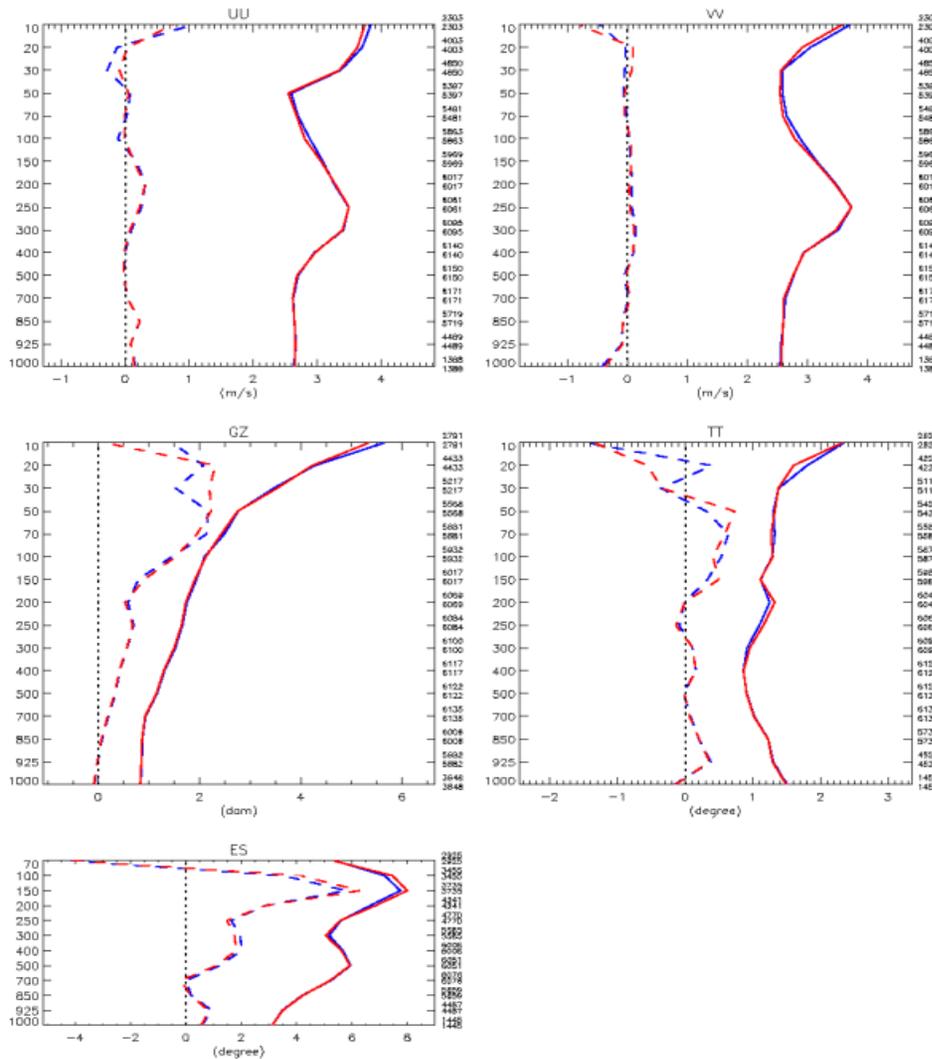
[forrms.2003081012_plh72.dat](#)

[forrms.2003081012_plh73.dat](#)

TG 0.872830 (°C)

TG 0.874430 (°C)

0-P6hr



Vérification de la moyenne d'ensemble contre observations

Référence: interpolation avec splines cubiques

Expérience: avec les polynômes cubiques de Lagrange

Avec les polynômes cubiques de Lagrange, on améliore la qualité de l'analyse EnKF en haut de 200 hPa.

Verification contre les observations AMSU-A

Avec les polynômes cubiques de Lagrange, les valeurs O-P sont légèrement meilleures. Nous observons les déviations standards suivantes (uniquement pour NOAA 15):

canal	splines cub.	cub. Lagrange
3	1.8648	1.8616
4	0.4382	0.4360
5	0.2455	0.2439
6	0.2137	0.2122
7	0.2237	0.2213
8	0.2652	0.2610
9	0.3693	0.3652
10	0.5067	0.5037

Nous avons décidé d'utiliser les polynômes cubiques de Lagrange pour toutes les expériences avec l' EnKF.

Le système opérationnel de prévisions d'ensemble

Nous tournons 8 cycles d'assimilation avec le modèle SEF et la procédure d'Interpolation Optimale (OI). L'OI utilise toujours les "observations" d'épaisseur au lieu des données TOVS de radiance. On peut dire qu'on utilise la configuration 1994 de l'analyse globale.

Un ensemble bi-modèle de 16 membres, dont 8 avec le modèle SEF et 8 avec le modèle GEM, est intégré jusqu'au jour 10.

Afin de pouvoir partir 16 intégrations, n'ayant que 8 conditions initiales des cycles SEF/OI, nous utilisons une procédure spéciale pour doubler la taille de l'ensemble. Cette procédure peut causer des problèmes d'équilibre pour les intégrations.

Nous souhaitons moderniser toute la procédure pour obtenir les conditions initiales.

Le filtre de Kalman d' ensemble proposé

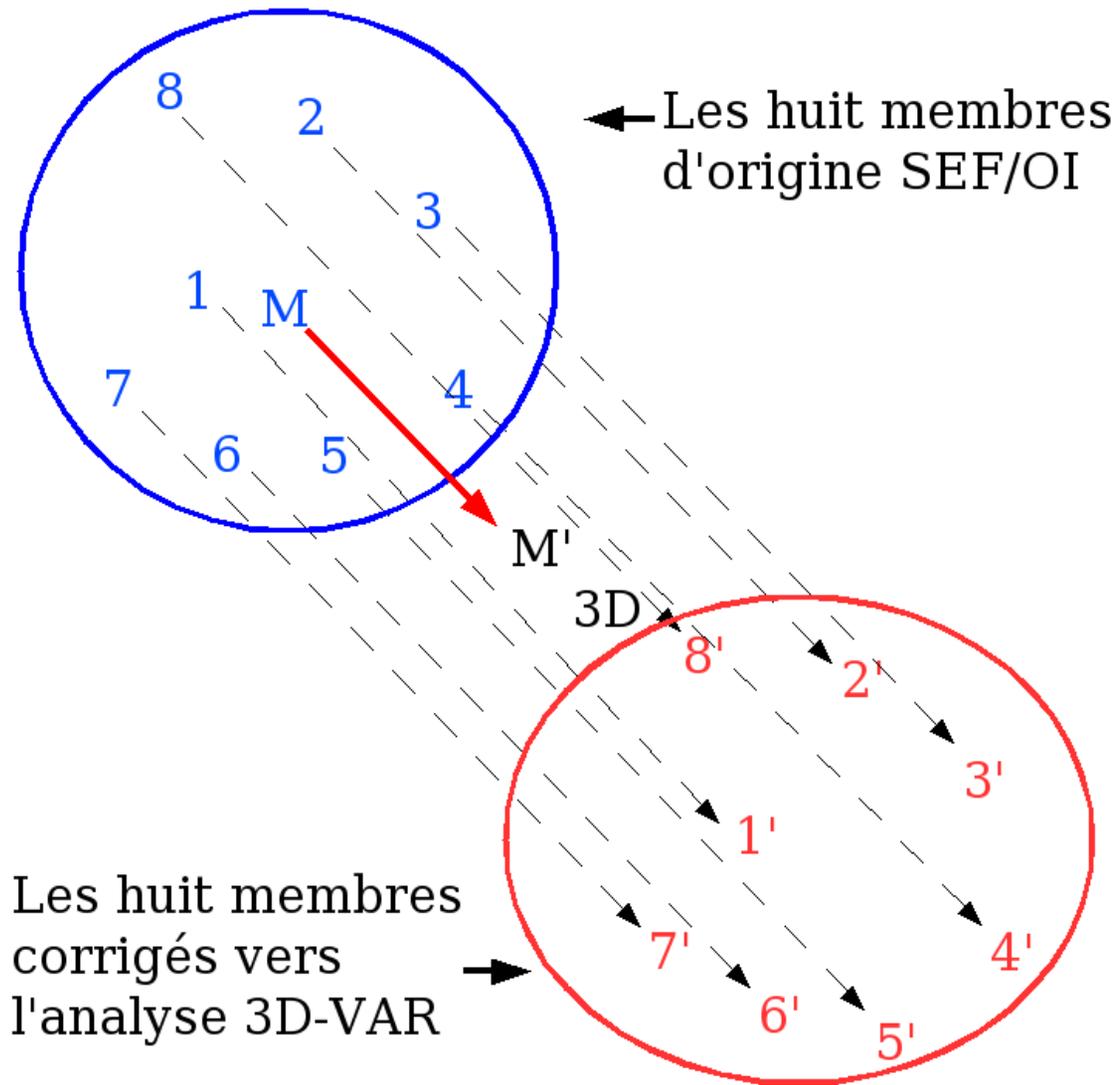
Nous utilisons la version 3.1.1 du modèle GEM-DM. La grille horizontale est de taille $300 * 150$. Le toit du modèle est à 10 hPa. Les pôles du modèle sont localisés aux pôles géographiques. La version du modèle est quasi-identique au modèle global $400 * 200$. L' ensemble compte 96 membres.

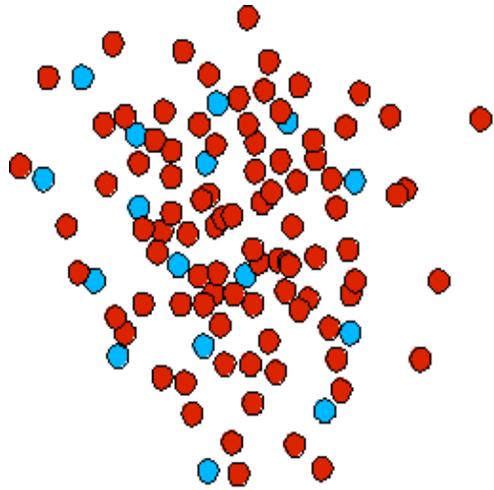
Nous assimilons presque toutes les observations qui sont actuellement assimilées par le 3D-Var avec le modèle $400 * 200$, incluant les radiances AMSU-A et AMSU-B. Exceptions: les vents de surface et l' humidité de surface.

Nous localisons l' effet des observations à 2800 km (distance horizontale) et à deux unités de log p.

Nous n' utilisons pas de filtre digitale.

La procédure "KICK" pour doubler la taille de l'ensemble



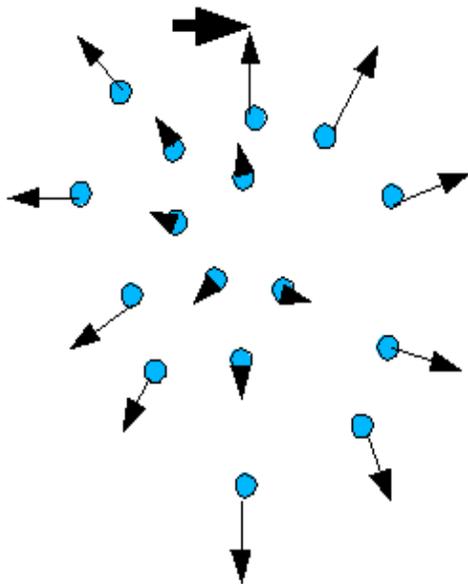


La recette pour obtenir 16 conditions initiales à partir d'un ensemble de 96 membres:

1) Éliminer $96-16=80$ membres d'ensemble en conservant la moyenne d'ensemble.

2) Multiplier la déviation standard avec 1.5.

3) Corriger pour sursaturation ou humidité négative.



Dépendances

OI/SEF:

Pour les observations, le système OI/SEF est indépendant du système d'analyse déterministe 3D-VAR/GEM 400*200. Nous avons nos propres observations (incluant les données SATEM) et notre propre contrôle de qualité.

Pour la procédure "KICK", il faut attendre que l'analyse 3D-VAR soit disponible.

EnKF:

Nous utilisons les observations du fichier postal du 3D-VAR et profitons ainsi du contrôle de qualité du 3D-VAR.

Il faut attendre que l'analyse déterministe soit disponible.

Il faut réviser la procédure à chaque fois qu'il y a des changements à la procédure pour l'analyse déterministe (pour ce qui concerne le traitement des observations).

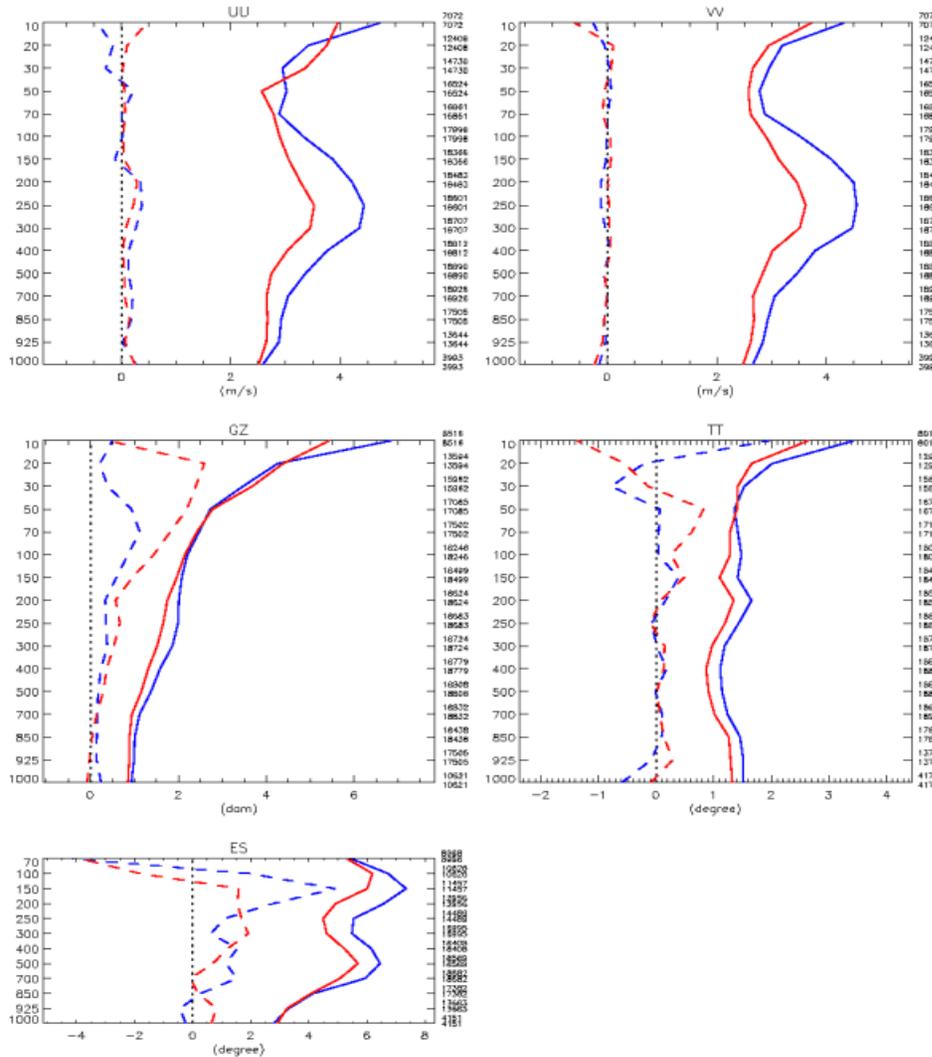
Vers une proposition de l' EnKF à CPOP

En préparation pour une proposition à CPOP, nous avons fait un cycle d' été et un cycle d' hiver.

Le cycle d' été choisi est celui du 27 juillet 2003 au 31 août 2003. Nous éliminons les 4 premières journées et nous regardons les résultats pour août 2003.

Le cycle d' hiver devait être celui du 7 janvier 2004 au 12 février 2004. Malheureusement, le cycle a avorté le 2 février avec un comportement pathologique impliquant les données de radiance. Nous avons également noté des incréments très localisés pour la température au niveau $\eta = 0.993$. Nous verrons plus loin les correctifs proposés.

O-P6hr

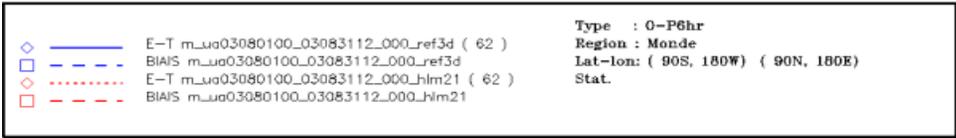
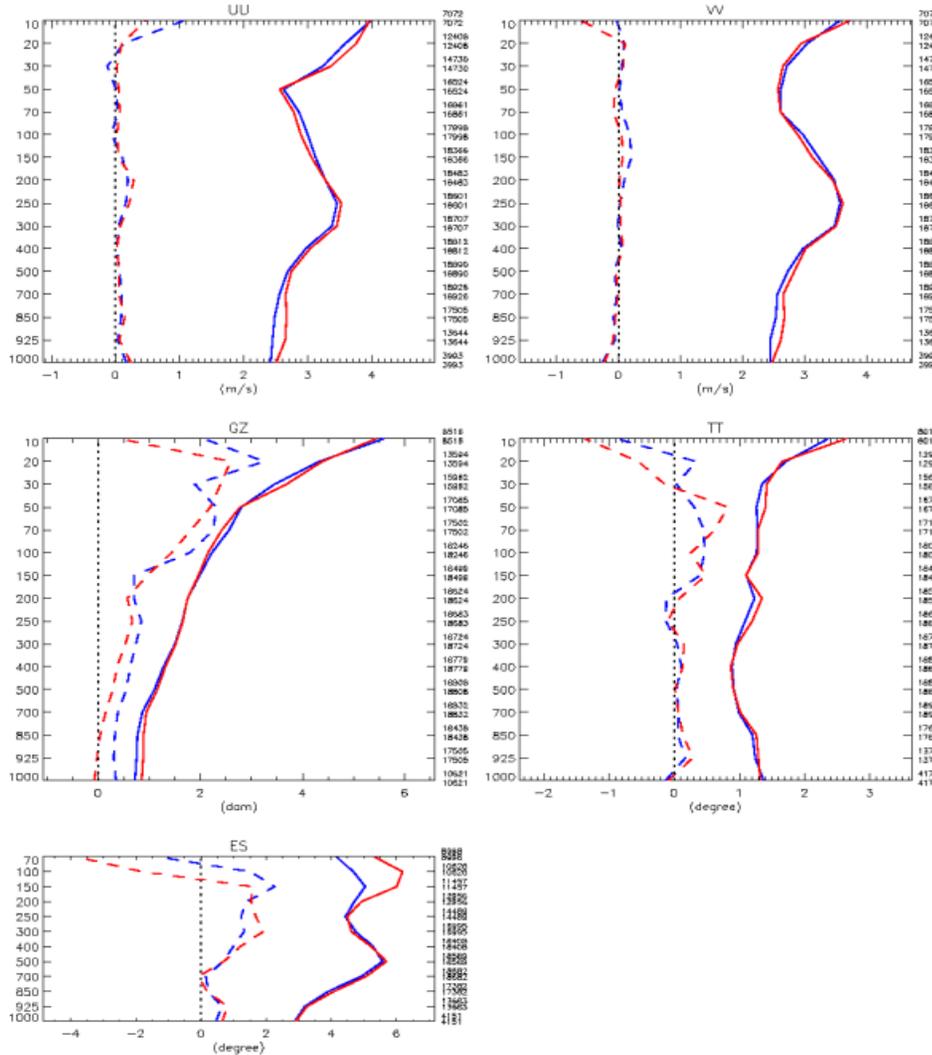


Comparaison des O-P pour l' OI et l' EnKF

En mode assimilation, L' EnKF est supérieur a l' OI. Nous avons observé de bonnes différences pour toutes les régions.

En mode prévision, le système OI profite de la correction vers l' analyse 3D-VAR que nous n' avons par pour l' EnKF. Il est alors aussi important de comparer avec les O-P du 3D-VAR.

O-P6hr

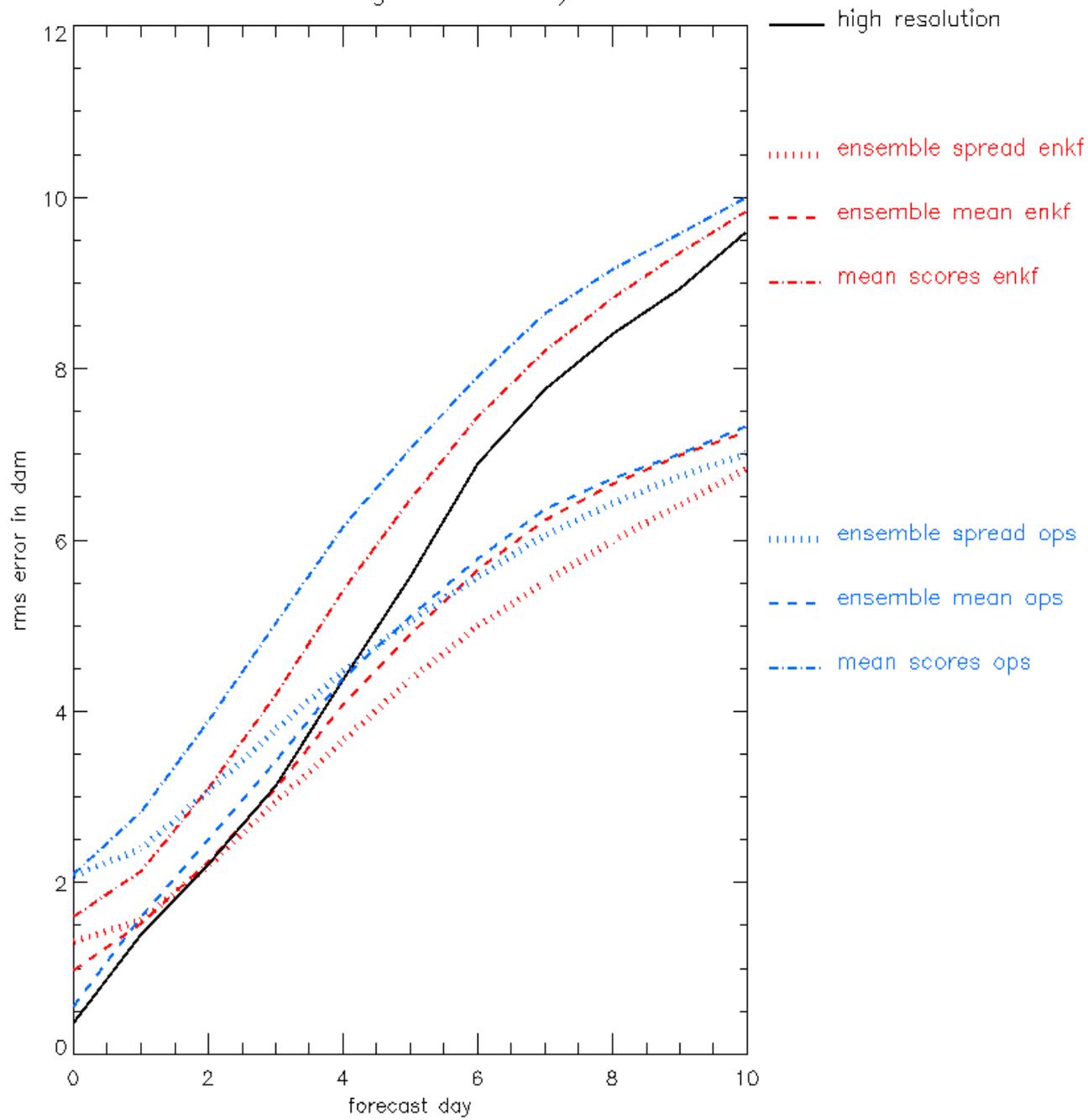


Comparaison des O-P pour le 3D-VAR et l' EnKF

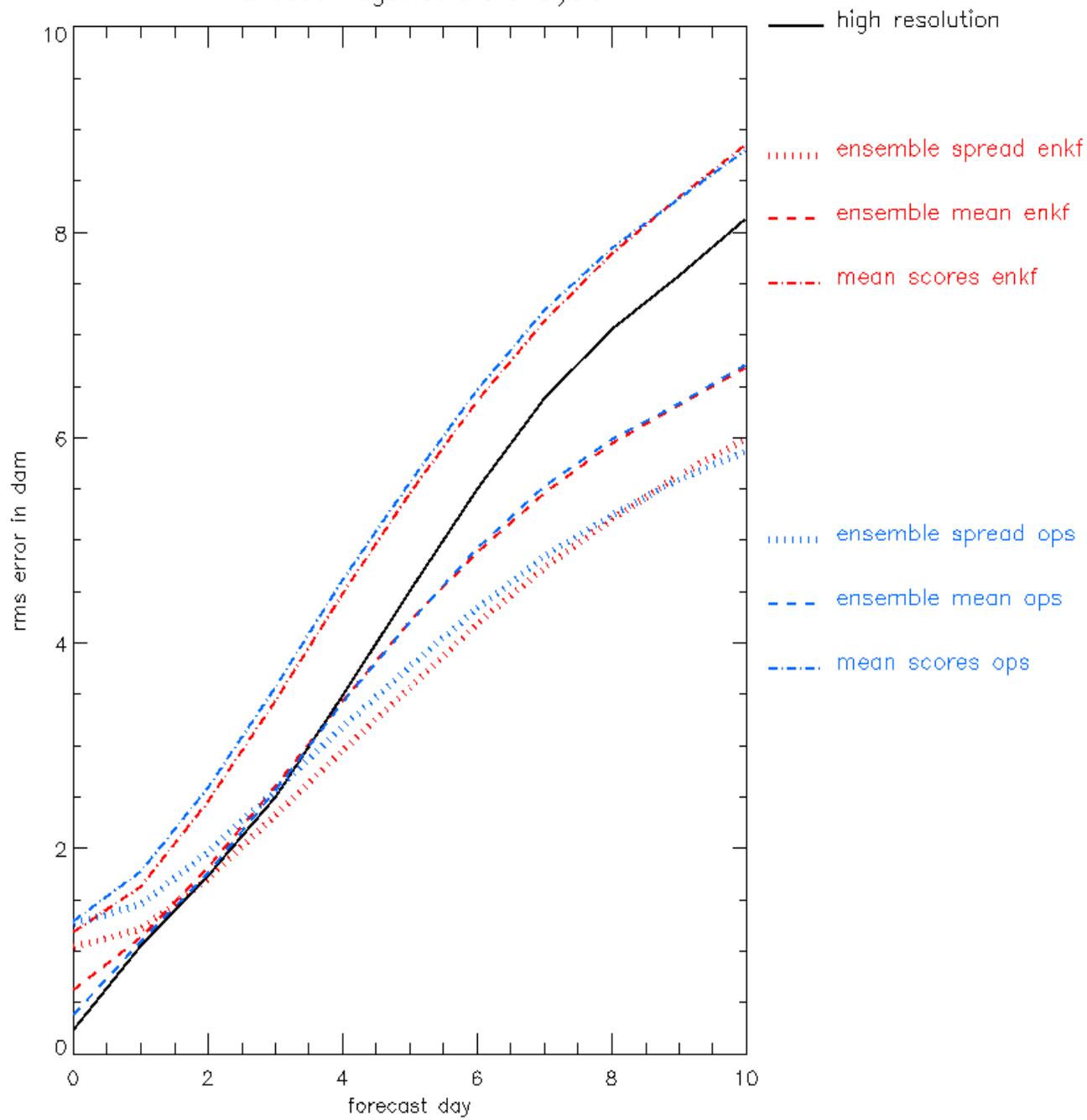
Le 3D-VAR opérationnel utilise une grille 400 * 200. L' EnKF utilise 96 membres et une grille 300 * 150.

À première vue le 3D-VAR et l' EnKF ont une qualité comparable (pour la moyenne). Il n' est alors pas évident qu' un système de prévisions d' ensemble avec l' EnKF sera supérieur, en mode prévision à un système avec l' OI et une correction vers le 3D-VAR.

quality of the ensemble forecast
ops_enkf_AOUT 2003 sef_gem
global area 500.000 mb
validation against the analysis



quality of the ensemble forecast
ops_enkf_AOUT 2003 sef_gem
northern extratropics 500.000 mb
validation against the analysis



Résumé des vérifications contre analyse

Nous avons fait des prévisions de 10 jours à partir des conditions initiales d' août 2003.

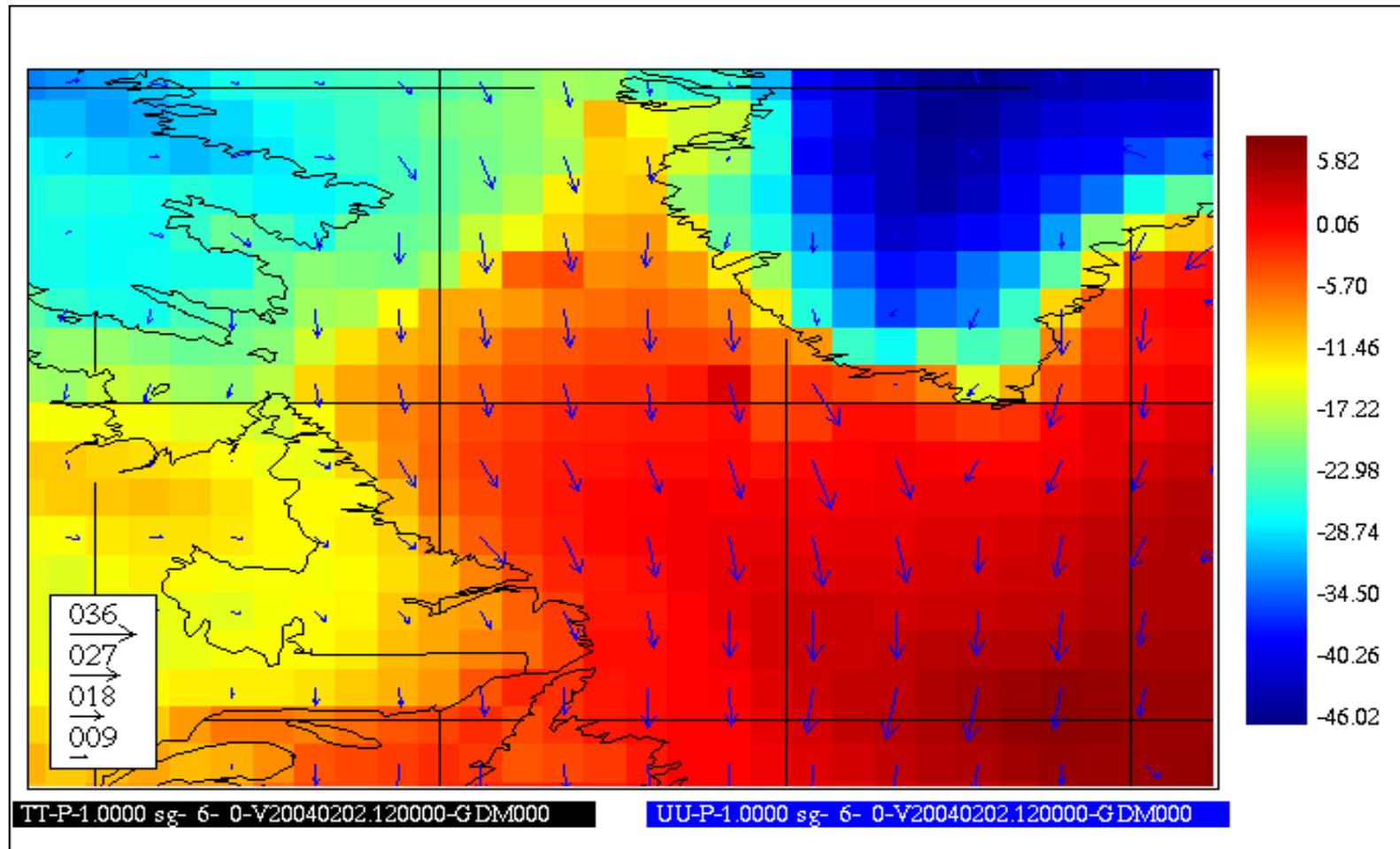
Avec l' EnKF:

- Nous obtenons des meilleurs scores pour la moyenne d'ensemble. Surtout pour l' hémisphère sud.
- Nous avons moins d écarts-types au début de la prévision.
- La croissance des écarts-types est plus rapide.
- Il y a un meilleur accord entre l' amplitude de l' erreur de la moyenne et les écarts-types de l' ensemble.
- Les scores pour les membres individuels sont meilleurs.

Problème avec le cycle d'hiver

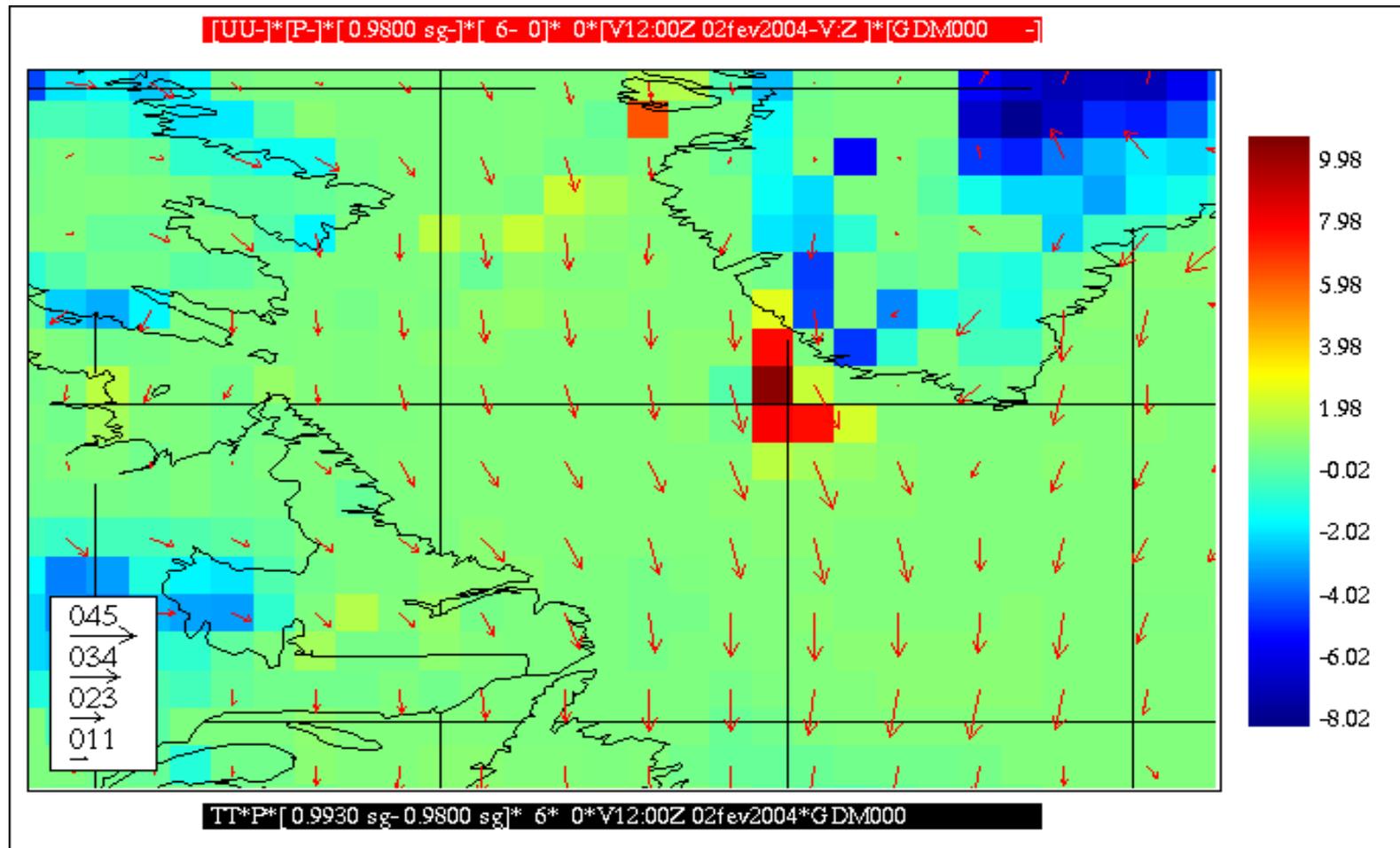
Nous avons constaté un problème lorsqu'un vent très froid souffle sur l'eau. Le flux de chaleur sensible est alors très important.

La situation météorologique typique:



Profil sur-adiabatique

À cause de nos pas de temps de 45 minutes et un traitement explicite, le flux de chaleur peut se ramasser au niveau $\eta = 0.993$. Le profil de température peut alors devenir très sur-adiabatique à des points.

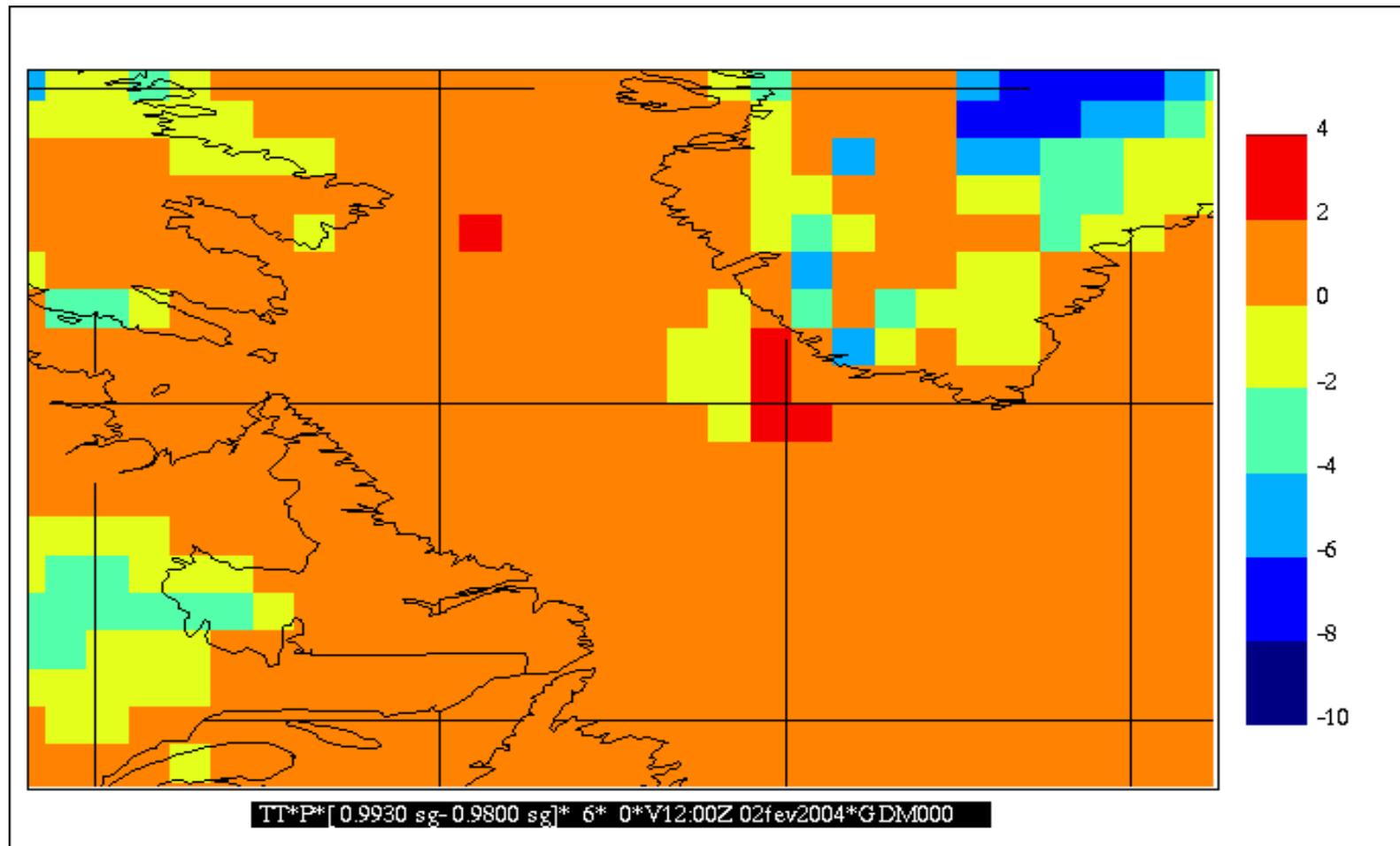


L' EnKF distribue les points problématiques entre les membres. Le filtre ne saute pas, mais le résultat n' est pas acceptable.

Corrections des profils

Pour corriger le problème, nous aurions pu réduire le pas de temps de façon significative (jusqu'à 15 minutes) ou utiliser une première couche plus profonde. Yves Delage nous a proposé une procédure pour corriger les profils de température et d'humidité.

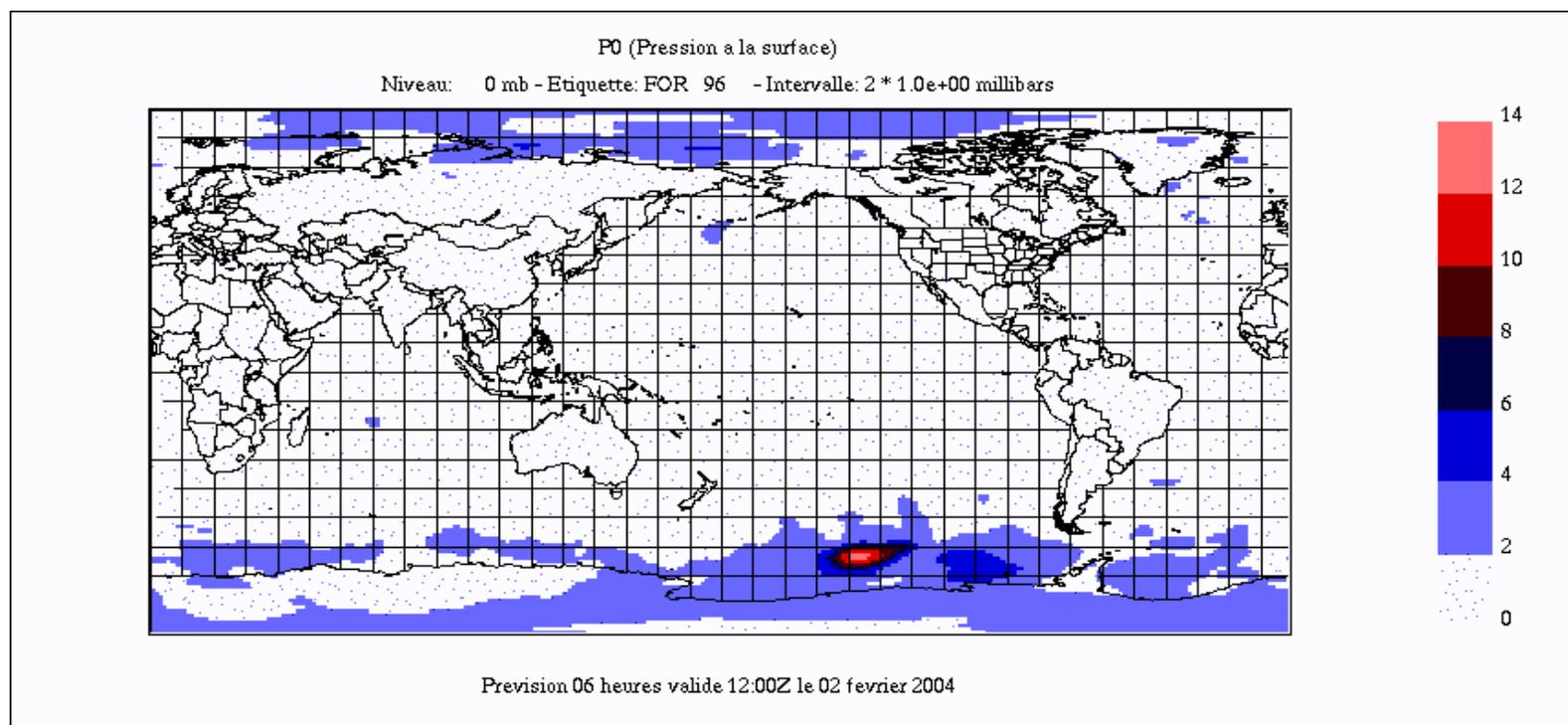
Après la correction:



Divergence du cycle d' assimilation

Le cycle d' hiver a divergé au 2 février 2004. On avait alors des écarts-types de l' ordre 100 hPa et 100 m/s dans une région sur l' océan Pacifique Sud. L' analyse avait alors fait augmenter les écarts-types, qui étaient déjà grands, avec un facteur 6 avant de “planter” à cause du caractère non météorologique des champs.

Les écarts-types pour la pression de surface en hPa avant l' analyse:



Sensibilité de l'opérateur RTTOV-7 aux vents de surface

L'opérateur RTTOV-7 est très sensible pour des vents de surface W forts pour canal-30. Il s'agit de la modélisation de l'émissivité F due à l'écume de l'océan:

$$F = 0.19510^{-4}W^{2.55} \quad (\text{Stephen English UKMO})$$

Pour mieux coller aux observations du canal-30, l'analyse peut augmenter le vent, peu importe la direction initiale du vent. Les écarts-types de l'ensemble vont alors augmenter et l'analyse va donner encore plus de poids aux observations suivantes du même type. Avec la dépendance du type $W^{2.55}$, nous arrivons rapidement dans un régime ($W > 30m/s$) pour lequel le paramétrage n'a pas été développé. Après quoi, nous arrivons à des valeurs non physiques pour F à $W = 70m/s$. L'analyse arrête peu après ($W = 100m/s$).

Pour éliminer toute possibilité de récurrence, nous allons limiter le vent W vu par RTTOV-7 à 10 m/s et nous allons assimiler les observations de vent de surface (comme le fait déjà le 3D-VAR).

Le statut actuel du 16 juin 2004

Nous avons recommencé le cycle d'hiver avec la modification au modèle pour le flux de surface, avec la modification pour les observations AMSU-A et en incluant les observations de vent de surface. Nous allons également reprendre le cycle d'été avec les mêmes modifications.

Une version précédente du filtre a été pré-transférée aux opérations. Cette version tourne aux opérations depuis le 5 mars 2004. Il y a eu un seul pépin. À cause des différences mineurs dans l'environnement, nous n'avons pas été capable de reproduire le problème en développement.

Suite au transfert, nous avons entamé une réécriture de nos scripts visant à faciliter des transferts futurs avec la possibilité de vérifier bit-par-bit. Nous pourrions pré-transférer la version actuelle.

Nous étudions encore la procédure pour initialiser le modèle SEF à partir des conditions initiales en coordonnées η du modèle GEM.

Conclusions

1. Nous sommes presque rendus à une première livraison de l' EnKF. Malheureusement, nous devons reprendre nos expériences pour corriger un problème avec des flux élevés dans le modèle GEM global et un problème avec la sensibilité aux vents de surface de l' opérateur RTTOV-7.
2. Avec l' EnKF nous modernisons la composante "cycles d' assimilation" du système de prévisions d' ensemble.
3. Nous pouvons partager beaucoup de quincaillerie avec l' équipe variationnelle.
4. Les vérifications (préliminaires) des prévisions de 10 jours montrent des effets positifs sur les caractéristiques de l' ensemble.
5. Nous observons que les erreurs dans le cycle d' assimilation sont beaucoup plus grandes que prévues par l' EnKF. Pour expliquer l' amplitude des erreurs, nous avons besoin d' une composante "erreur modèle" très importante. Nous devons essayer de mieux comprendre l' origine de cette erreur.
6. Il y a beaucoup de potentiel pour l' amélioration de nos analyses.