

# Le filtre de Kalman d'ensemble révision de l' algorithme

Peter Houtekamer

Depuis le 12 janvier 2005, le filtre de Kalman d'ensemble (EnKF) est utilisé aux opérations pour fournir les conditions initiales au système de prévision d' ensemble. C' est une première mondiale.

Nous avons alors réussi à remplacer les cycles d' assimilation avec l' interpolation optimale (OI) et le modèle SEF par des cycles avec l' EnKF et le modèle GEM.

Toutefois, les différences en faveur du système avec l' EnKF n' étaient pas grandes. Nous considérons alors que la qualité du système actuel n' est pas satisfaisante.

Qu'est-ce qu'on peut faire mieux?

## Contenu de la présentation

- Introduction
- Théorie du filtre de Kalman et de l'EnKF
- La version opérationnelle
- Manque d'équilibre
- Interpolation temporelle
- Problème de biais
- Convergence avec le nombre de membres
- Ajout de l'erreur du modèle
- Conclusions

## Le filtre de Kalman

Le filtre de Kalman donne la solution optimale (variance minimale) au problème d'assimilation étant donné une dynamique linéaire et des statistiques d'erreur gaussienne (Maybeck, 1979, *Stochastic Models, Estimation and Control*).

$$\begin{aligned}x^f(t) &= Mx^a(t-1) \\ P^f(t) &= MP^a(t-1)M^T + Q \\ K(t) &= P^f(t)H^T(HP^f(t)H^T + R(t))^{-1} \\ P^a(t) &= (I - KH)P^f(t) \\ x^a(t) &= x^f(t) + K(t)(y^o(t) - Hx^f(t))\end{aligned}$$

$y^o$	les observations,	R	matrice de covariance pour $y^o$ ,
$x^f$	le champ d'essai,	$P^f$	matrice de covariance pour $x^f$ ,
$x^a$	l'analyse,	$P^a$	matrice de covariance pour $x^a$ ,
M	modèle linéaire de prévision,	Q	matrice de covariance pour M,
H	matrice d'interpolation vers $y^o$ ,	K	matrice de "gain".

## Caractéristiques du filtre de Kalman

1. La solution est optimale. Toutefois, pour y arriver, il faut respecter des conditions sévères. En particulier, les erreurs du modèle et des observations doivent être sans biais. Leurs matrices de covariances  $Q$  and  $R$  doivent être connues.
2. Si l'erreur du modèle est balancée, l'analyse sera également balancée.
3. Il n'y a pas de perte d'information entre deux cycles d'assimilation.
4. Les opérateurs étant linéaires, la saturation des erreurs ne peut pas être décrite.
5. Les coûts d'une évaluation numérique sont élevés.

## Les équations du filtre de Kalman d'ensemble

Les équations de l'EnKF sont motivées par celles du filtre de Kalman. Un petit ensemble de  $N \approx O(100)$  membres est utilisé pour estimer les covariances.

$$\begin{aligned}x_i^f &= M(t, t-1)x_i^a(t-1) + q_i, \quad i = 1, \dots, N \\q_i &\sim N(0, Q)\end{aligned}$$

$$P^f H^T \equiv \frac{1}{N-1} \sum_{i=1}^N (x_i^f - \overline{x_i^f})(Hx_i^f - \overline{Hx_i^f})^T$$

$$HP^f H^T \equiv \frac{1}{N-1} \sum_{i=1}^N (Hx_i^f - \overline{Hx_i^f})(Hx_i^f - \overline{Hx_i^f})^T$$

$$K = P^f H^T (HP^f H^T + R)^{-1},$$

$$y_i^o = y^o + r_i, \quad i = 1, \dots, N$$

$$r_i \sim N(0, R)$$

$$x_i^a = x_i^f + K(y_i^o - Hx_i^f), \quad i = 1, \dots, N$$

# Assimilation séquentielle

Exemple:

$$f : N(0, 1), \quad o1 : N(0, 1), \quad o2 : N(0, 1)$$

$$a = k_f f + k_1 o1 + k_2 o2$$

$$k_f = \frac{\sigma_f^{-2}}{\sigma_f^{-2} + \sigma_{o1}^{-2} + \sigma_{o2}^{-2}} = k_1 = k_2 = \frac{1}{3}$$

étape 1 de l'assimilation séquentielle:

$$f1 : N(0, 1), \quad o1 : N(0, 1)$$

$$f2 = k_{f1} f1 + k_1 o1$$

$$k_{f1} = k_1 = \frac{1}{2}$$

étape 2 de l'assimilation séquentielle:

$$f2 : N(0, 0.5), \quad o2 : N(0, 1)$$

$$a = k_{f2} f2 + k_2 o2$$

$$k_{f2} = \frac{2}{2 + 1}, k_2 = \frac{1}{3}$$

$$a = \frac{1}{3}(f1 + o1 + o2)$$

## Caractéristiques du filtre de Kalman d'ensemble

1. L'usage d'un petit ensemble de  $N \approx O(100)$  membres et l'application de l'assimilation séquentielle ramènent les coûts à un niveau acceptable.
2. Généralement les opérateurs non-linéaires (comme  $M(t, t - 1)$  et  $H$ ) sont utilisés. En se servant du modèle non-linéaire de prévision nous pouvons correctement décrire la saturation des erreurs.
3. Si l'erreur du modèle est balancée et si toutes les perturbations sont petites, l'analyse sera également balancée.
4. Avec une taille de l'ensemble  $N$  plus grande et des opérateurs linéaires  $M(t, t - 1)$  et  $H$ , l'EnKF converge vers le filtre de Kalman.
5. La solution n'est pas optimale. Lorenc (2003, QJRMS) montre un exemple où l'analyse est même pire que le champs d'essai.

## Besoin de localiser

L'EnKF fournira uniquement un incrément d'analyse pour les directions où l'ensemble de champs d'essai montre une incertitude. Pour d'autres directions, les champs d'essai sont en apparence déjà parfaits. Conséquemment, avec 100 membres d'ensemble, l'incrément d'analyse sera uniquement dans l'espace qui est défini par les différences entre les champs d'essai. Comme l'atmosphère a sûrement plus de 100 degrés de liberté, ceci est un problème.

Des mesures artificielles, sans fondement dans les équations du filtre de Kalman, sont alors nécessaires pour augmenter la dimension de l'ensemble.

Avec ces mesures, l'incrément d'analyse ne restera pas dans l'espace balancé qui est défini par l'ensemble.

Nous devons alors trouver un compromis entre les désirs d'avoir (i) une analyse qui est bien balancée et (ii) une analyse qui est proche des  $O(100\ 000)$  observations.

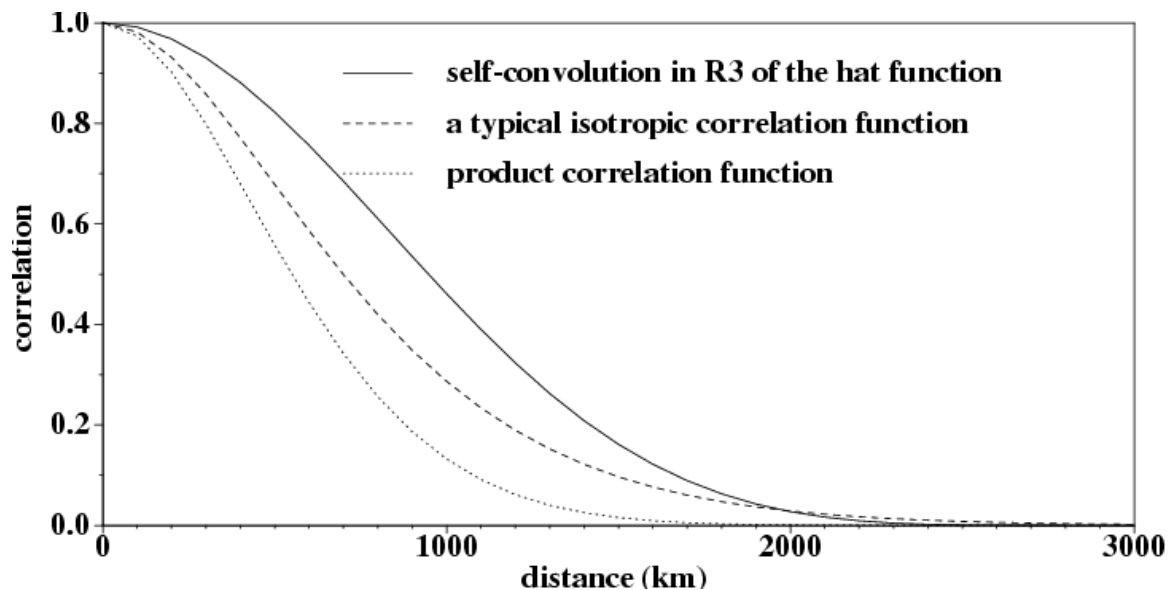
## Produit de Hadamard

Idée (obtenue) de Richard Ménard (Gaspari & Cohn, 1999, QJRMS):

$$P^f(r_i, r_j) = P_{ensemble}^f(r_i, r_j) \rho(r, L).$$

Pour filtrer les covariances à longue distance, on utilise le produit de Hadamard (multiplier deux matrices élément par élément). Ceci devrait donner une matrice  $P^f$  positive qui n'a pas de problème de rang.

Avec plus de membres, on pourra prendre  $L$  plus grand.



## La configuration opérationnelle

Nous utilisons le modèle GEM avec une grille horizontale de  $300 \times 150$  et 28 niveaux  $\eta$ .

Nous avons un ensemble de  $2 \times 48$  membres (la configuration de l' EnKF avec deux ensembles).

Nous essayons d' utiliser toutes les observations utilisées par le 4D-Var. Nous pouvons alors profiter des procédures de contrôle de qualité comme le qc-var et le “background check”.

L' effet des observations est forcé vers zéro à une distance horizontale de 2800 km et à une distance verticale de deux unités de  $\ln p$ .

Il n' y a pas de procédure explicite, comme le filtre digital, pour assurer un équilibre.

L' application opérationnelle est la preuve que nous avons maintenant une qualité acceptable. Houtekamer et al. (MWR, 2005) montrent également que l' EnKF et le 3D-Var ont une qualité semblable.

## Similitude du 3D-Var et EnKF

Nous devons essayer de comprendre comment une méthode 3D (le 3D-Var) et 4D (l' EnKF) peuvent avoir une qualité similaire. Constatons d'abord que les algorithmes ont beaucoup de choses en commun:

- usage des mêmes observations,
- avec les mêmes statistiques d'erreur (sans biais et avec la matrice  $R$  diagonale),
- et le même contrôle de qualité,
- nous présumons aussi que les observations sont valides au moment central de l'analyse (il est à noter que Hunt et al. (2004, Tellus) utilisent une version de l' EnKF avec interpolation temporelle),
- usage du même modèle de prévision,
- avec une matrice  $Q$ , pour l'erreur du modèle, qui ressemble à la matrice  $B$  du 3D-Var.

Ces choix étaient convenables pour le développement. Par contre, il pourrait y avoir un effet négatif sur la qualité de nos analyses.

## Longueur des incréments

Une norme d'énergie (Ehrendorfer and Errico, 1995, JAS) est utilisée pour quantifier l'amplitude des différentes opérations dans l'EnKF:

$$E = \frac{1}{2S} \int_S \int_0^1 \left[ u^2 + v^2 + \frac{c_p}{T_r} T^2 + R_a T_r \left( \frac{p_s}{p_r} \right)^2 \right] d\eta dS$$

Pour une journée spécifique et un membre spécifique, nous avons obtenu une longueur de

- 4.62 m/s pour  $F$ , la prévision,
- 1.39 m/s pour l'erreur du modèle,
- 2.65 m/s pour  $A$ , l'incrément d'analyse,
- 0.35 m/s pour  $I$ , l'incrément de l'initialisation (si on l'applique et si on fait la mesure à 6 heures).

L'EnKF respecte alors la condition de Hollingsworth et al. (1986, MWR) que  $F > A > I$ .

Notons, par contre, l'erreur de 2.4 m/s due à l'absence d'interpolation temporelle aux limites de la fenêtre d'assimilation. Cette erreur devrait être plus petite que l'incrément de l'analyse.

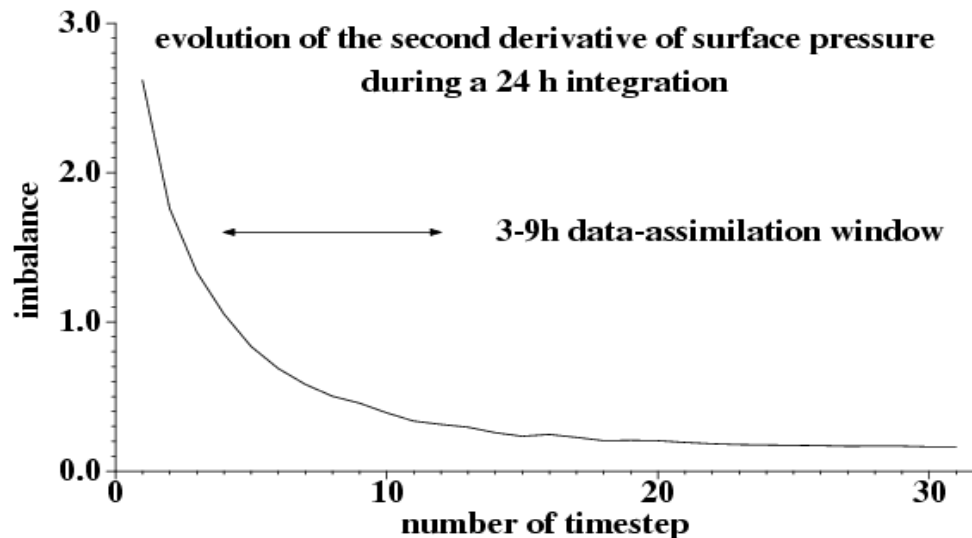
## Déséquilibre dans les conditions initiales

À fin de pouvoir quantifier le déséquilibre, nous utilisons la deuxième dérivée  $p_s''$  de la pression de surface,  $p_s$ , dans une serie temporelle obtenue d' une intégration de 24 heures.

$$p_s'' = p_s(t + 45\text{min}) - 2p_s(t) + p_s(t - 45\text{min})$$

Une norme globale  $\| p_s'' \|$  est definie:

$$\| p_s'' \|^2 = (p_s'', p_s'') = \frac{1}{S} \int_S p_s'' p_s'' dS$$



## Manque d' équilibre et localisation

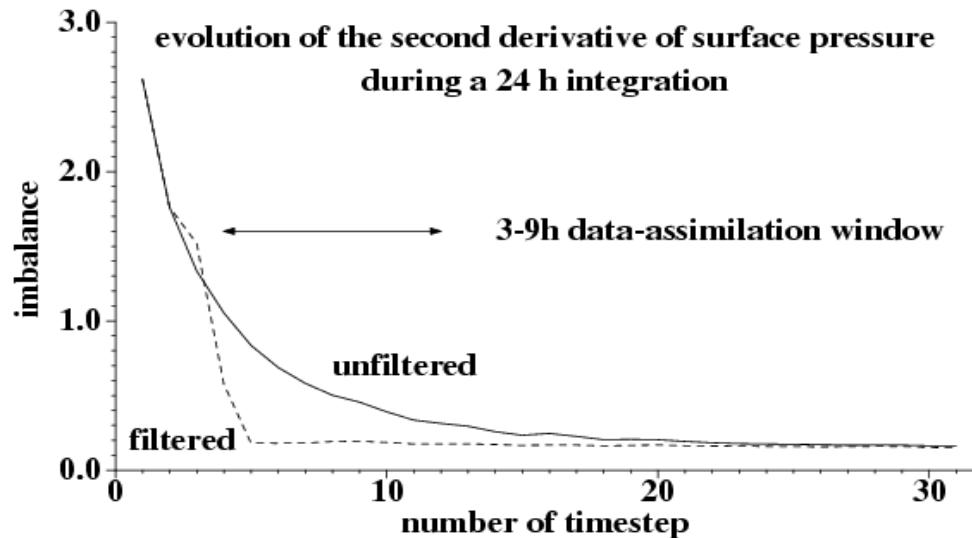
$r_h(km)$	déséquilibre initial $\  p_s''(45min) \ $				
	$r_z = 2$	$r_z = 4$	$r_z = 6$	$r_z = 8$	$r_z = 100$
2800	2.618	2.913	3.100	3.236	n.d.
5600	2.600	2.725	2.783	2.815	2.925
8400	2.744	2.520	2.381	2.345	2.423
11200	2.883	2.420	2.316	2.133	2.145
280000	1.996	1.558	1.223	1.153	0.454

La deuxième dérivée de la pression de surface estimée pour différentes valeurs des paramètres de localisation  $r_h$  et  $r_z$ . Les corrélations horizontales sont forcées à zéro dans une distance de  $r_h$  km et les corrélations verticales dans  $r_z$  unités de  $\ln p$ . Des valeurs non disponibles sont indiquées avec n.d.

Relâcher la localisation a peu d' effet sur le déséquilibre. Dans l' absence totale de localisation les conditions initiales sont balancées.

## Application d' un filtre digital

Avec une finalisation par filtre digital (Fillion et al., 1995, Tellus A) nous pouvons obtenir une intégration qui est en équilibre après 4 pas de temps (3 heures).



Avec le filtre digital, nous pouvons envisager d' ajouter une interpolation temporelle vers les observations (comme dans le 4D-Var).

L' application du filtre digital dans un cycle d' assimilation qui n' a pas d' interpolation temporelle a également un effet positif sur nos vérifications.

Toutefois, il serait préférable de traiter les problèmes d' équilibre à la source.

## L' équilibre géostrophique

Avec le produit de Hadamard, nous localisons les incréments de masse. Avec des structures de masse plus étroites, on pourrait s' attendre à des vents plus intenses.

Nous localisons également les incréments pour les vents avec le produit de Hadamard. Nous diminuons alors les vents là où nous devrions les augmenter et les vents deviennent alors sous-géostrophiques par rapport aux incréments de masse (Lorenc, 2003, QJRMS).

À mesure que l' EnKF assimile une plus grande fraction des observations, le déséquilibre augmente. Nous craignons que ceci résulte en une assimilation sous-optimale pour les derniers groupes d' observations.

Il est également désolant de voir que nous perdons un des grands avantages de l' analyse 4D (l' équilibre).

## La variable de contrôle

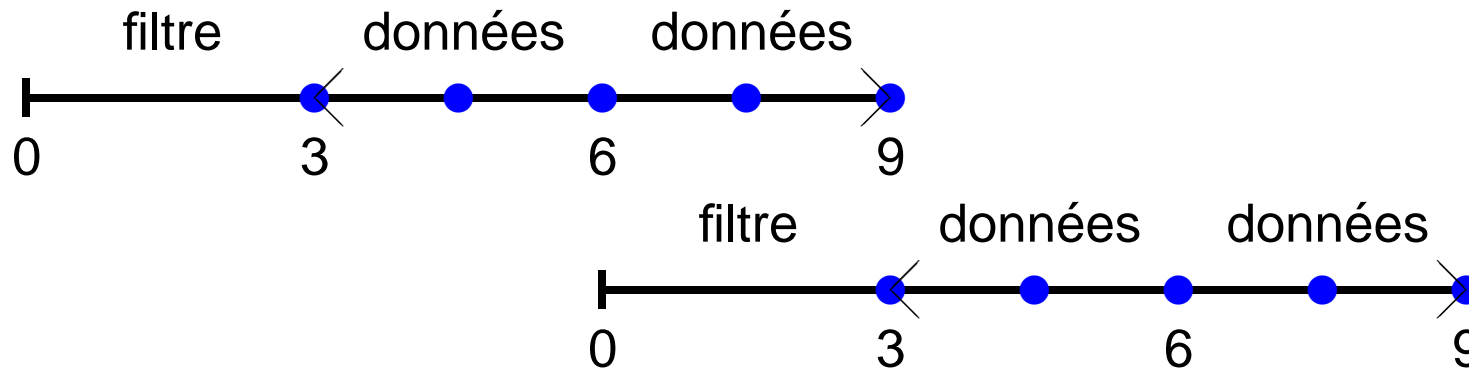
Pour résoudre le problème de déséquilibre, nous allons modifier la variable de contrôle (actuellement nous utilisons  $(u, v, T, q, p_s)$ ).

Suivant Jeff Kepert (Prague, 2005, poster) nous voulons utiliser  $(\Psi, \chi', T', q, p'_s)$  comme variable de contrôle. Tout l'équilibre géostrophique est alors contenu dans la variable  $\Psi$  et une localisation d'un incrément de  $\Psi$  ne devrait pas automatiquement causer une perte de l'équilibre géostrophique.

Remarquons que le 3D-Var utilise également une séparation entre une variable balancée  $\Psi$  et des variables non-balancées  $(\chi', T', \ln q, \ln p'_s)$ .

## La fenêtre d'assimilation

Avec l'EnKF, nous voulons assimiler toutes les observations qui sont disponibles dans la fenêtre de 6-h (comme dans le 4D-Var).



Pour l'interpolation temporelle, nous avons besoin de l'état du modèle aux moments  $t = 3h$ ,  $t = 6h$  et  $t = 9h$  (ou plus souvent). L'intégration subséquente part de l'analyse valide à  $t = 6h$ .

## Expansion de la variable de contrôle

Il est possible d'ajouter le vecteur  $Hx$  à la variable de contrôle  $x$  (Tarantola 1987, Anderson, MWR, 2001, Gauthier 2005) qui devient alors  $(x, Hx)$ .

L'avantage classique semble être que  $H$  peut être un opérateur complexe existant à l'extérieur de l'algorithme d'assimilation (comme un algorithme pour la convection).

Nous voulons utiliser la méthode pour d'abord calculer  $Hx$  en utilisant une trajectoire complète du modèle et ensuite faire l'analyse uniquement pour l'état du modèle au moment central. Le coût de l'algorithme:

$$N_{ens} * (N_{mod} + N_{obs}) * N_{obs} \approx N_{ens} * N_{mod} * N_{obs}$$

Toute l'information temporelle demeure dans les corrélations entre  $Hx(t = t_{obs})$  et  $x(t = 6h)$ . Nous pouvons alors tenir compte de l'évolution temporelle pour très peu de coûts informatiques. En réalité l'algorithme sera probablement plus rapide parce que nous n'aurons plus besoin de considérer les halo's.

Alternativement, il faudrait assimiler les observations dans l'ensemble de trajectoires (Hunt et al. 2004). Le coût de l'algorithme alternatif:

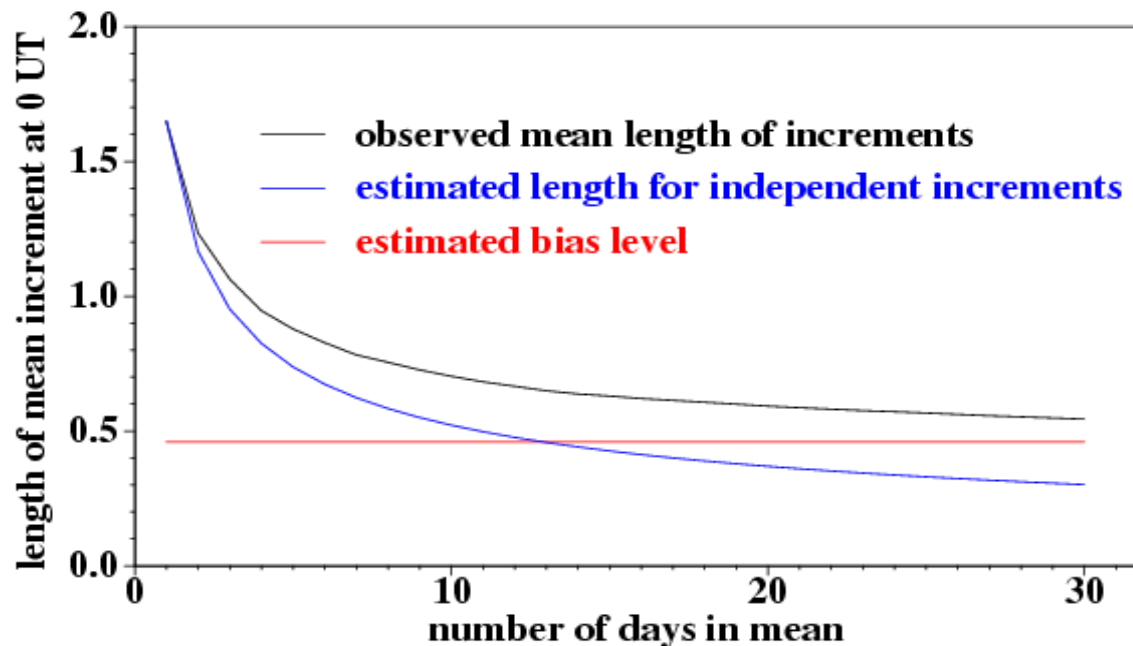
$$N_{ens} * N_{mod} * N_{obs} * N_{pas}$$

## Algorithme pour l'assimilation

actuel	futur
obtenir les champs d'essai $(u, v, T, q, p_s)$	
champs d'essai valide à 6 h	champs d'essai à 3 h, 4h30m, 6 h, 7h30, 9h
	précalculer $Hx(t_{obs})$
	transformation de $x(t = 6h)$ vers $\psi, \chi', T', q, p'_s$
variable de contrôle	
$(u, v, T, q, p_s)$	$(\psi, \chi', T', q, p'_s, Hx(t_{obs}))$
boucle sur l'ensemble d'observations:	
sélectionner un groupe d'observations	
calculer $Hx(t = 6h)$	trouver $Hx(t_{obs})$
calculer l'innovation $y^o - Hx$	
calculer la matrice de gain pour les observations	
mettre à jour $(u, v, T, q, p_s)$	mettre à jour $(\psi, \chi', T', q, p'_s, Hx(t_{obs}))$
	transformation de $x(t = 6h)$ vers $(u, v, T, q, p_s)$

## Présence de biais

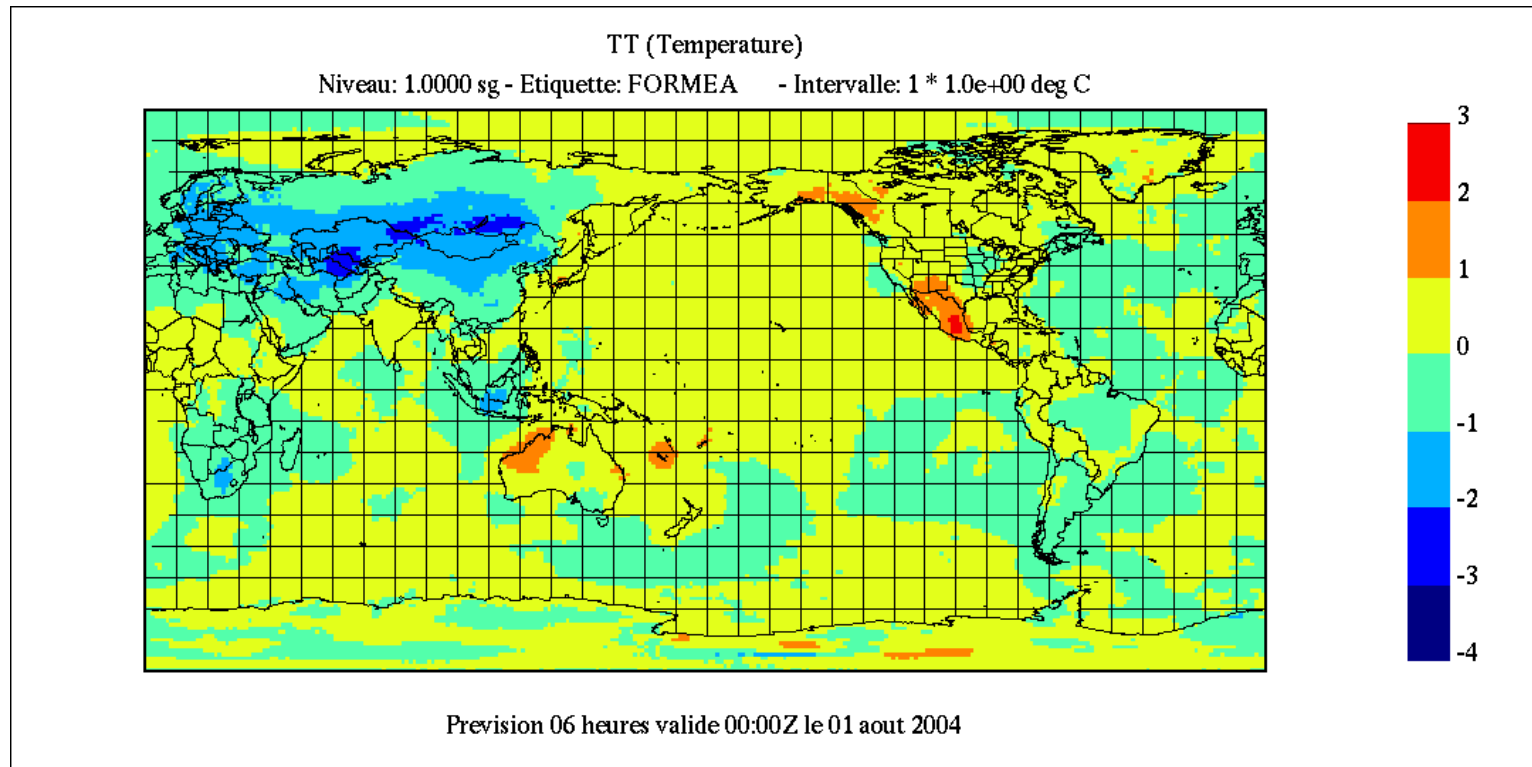
Dans le filtre de Kalman (et également dans l' EnKF), il est habituel de présumer que le modèle et les observations n' ont pas de biais systématique. On s' attend alors à ce que la moyenne de plusieurs incréments d' analyse convergera vers zéro comme la racine du nombre de cas utilisés pour la moyenne.



La composante systématique, mesurée avec norme d' énergie, est plus grande que la composante due à l' initialisation. Il est inquiétant qu' une des hypothèses derrière le filtre de Kalman ne soit pas mieux respectée.

## Biais à la surface

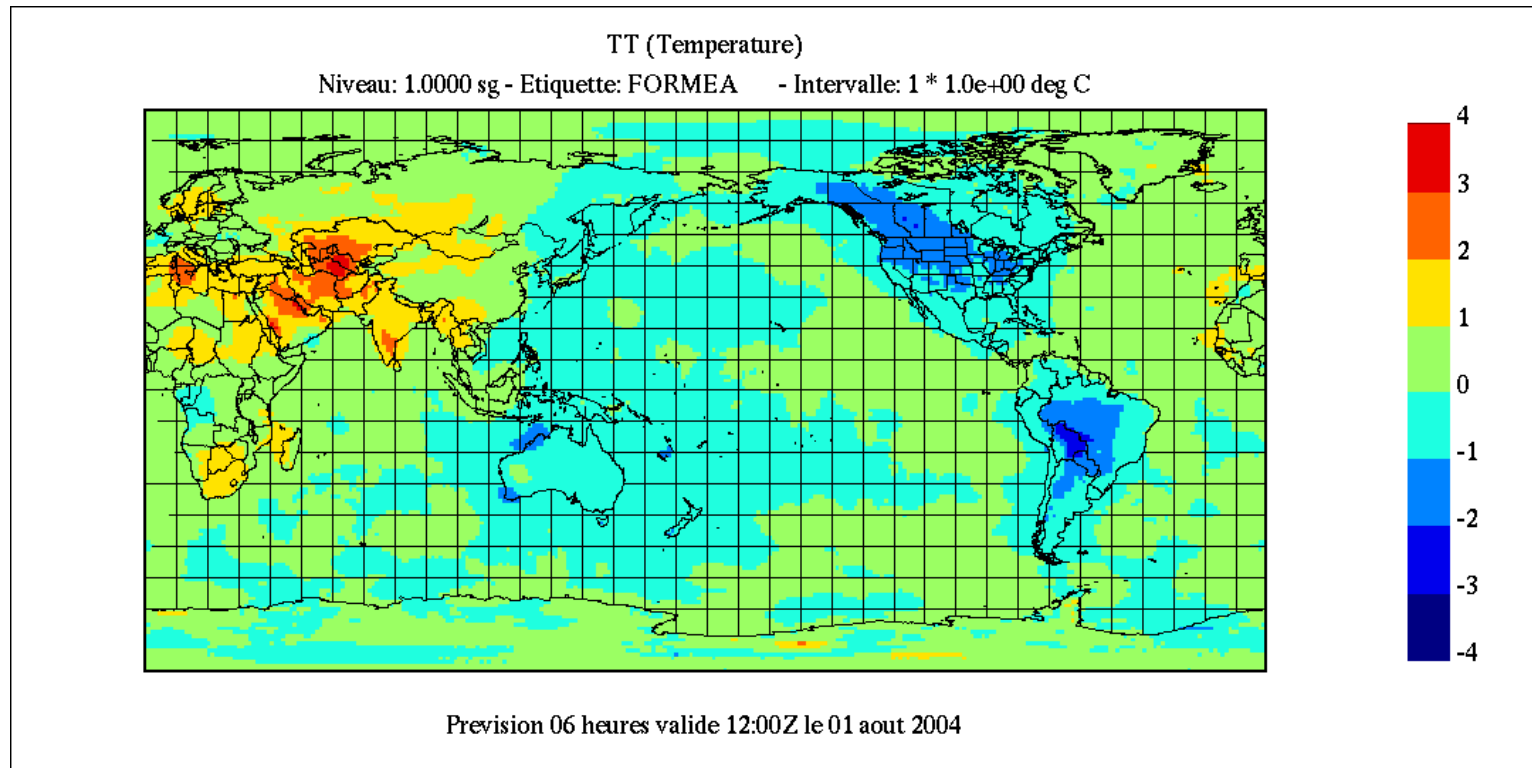
Martin Charron a trouvé un cycle diurne dans les incréments d'analyse.



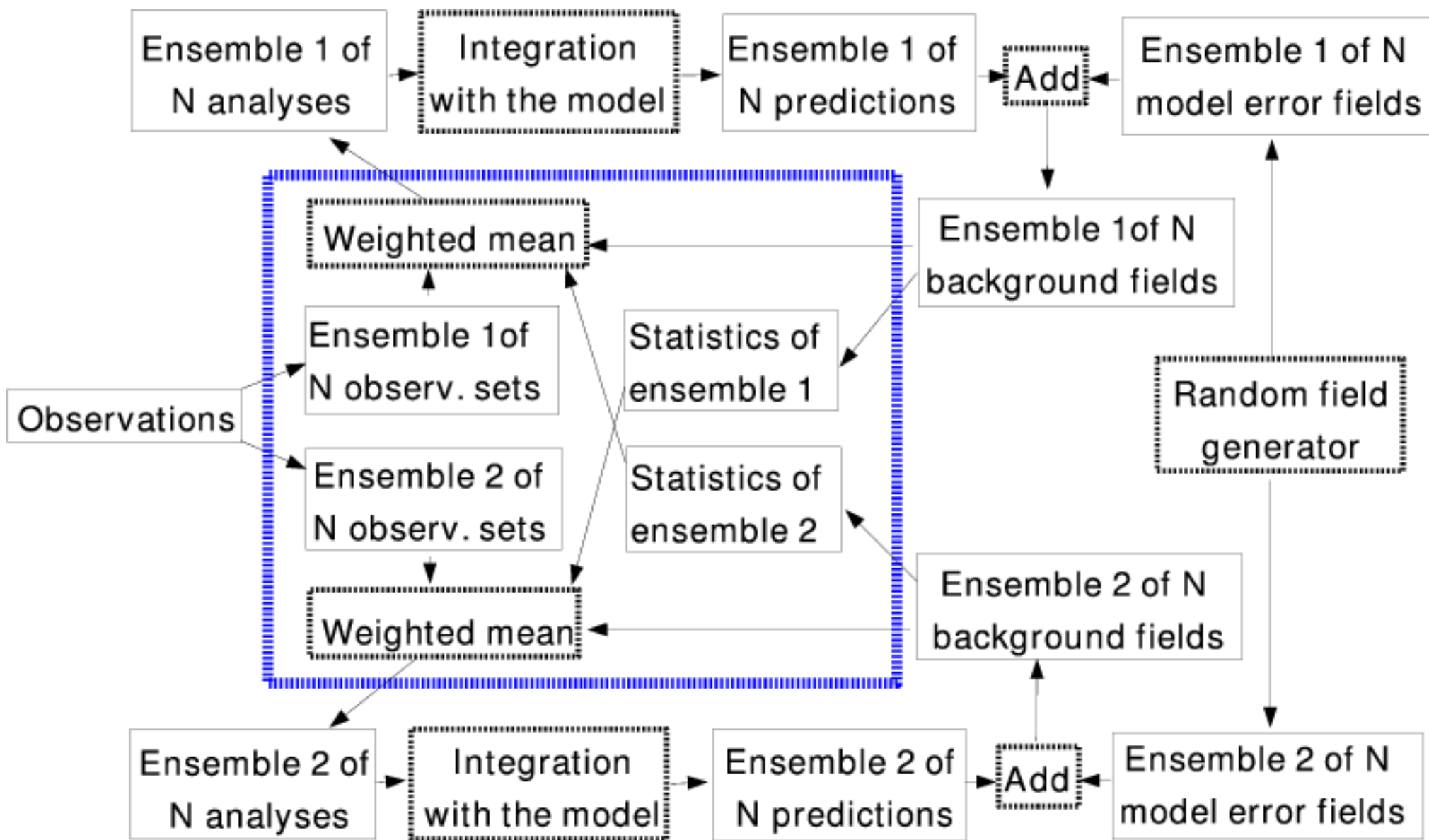
À chaque analyse de 00 UTC, nous intensifions le cycle diurne.

## Biais à la surface

À chaque analyse de 12 UTC, nous intensifions le cycle diurne.



L'analyse est un outil sous-optimal pour corriger un biais. Il serait préférable de corriger le problème à la source. Nous aimerions d'abord répéter les expériences avec ISBA et par la suite, si nécessaire, faire une correction appropriée.



## Convergence

Nous avons fait une séquence d'assimilations pour mesurer l'erreur due à la taille limitée de l'ensemble (toujours avec une localisation aussi sévère).

taille	erreur (m/s)
$2 \times 12$	0.89
$2 \times 24$	0.63
$2 \times 48$	0.44
$2 \times 96$	0.31
$2 \times 192$	0.22

En très bonne approximation, les erreurs diminuent comme la racine de 1 sur  $N$ . Évidemment nous sommes limités par la puissance de nos ordinateurs.

Rendu à  $2 \times 48$  membres, d'autres facteurs semblent dominer sur l'erreur d'échantillonnage.

## Interprétation de l'erreur du modèle

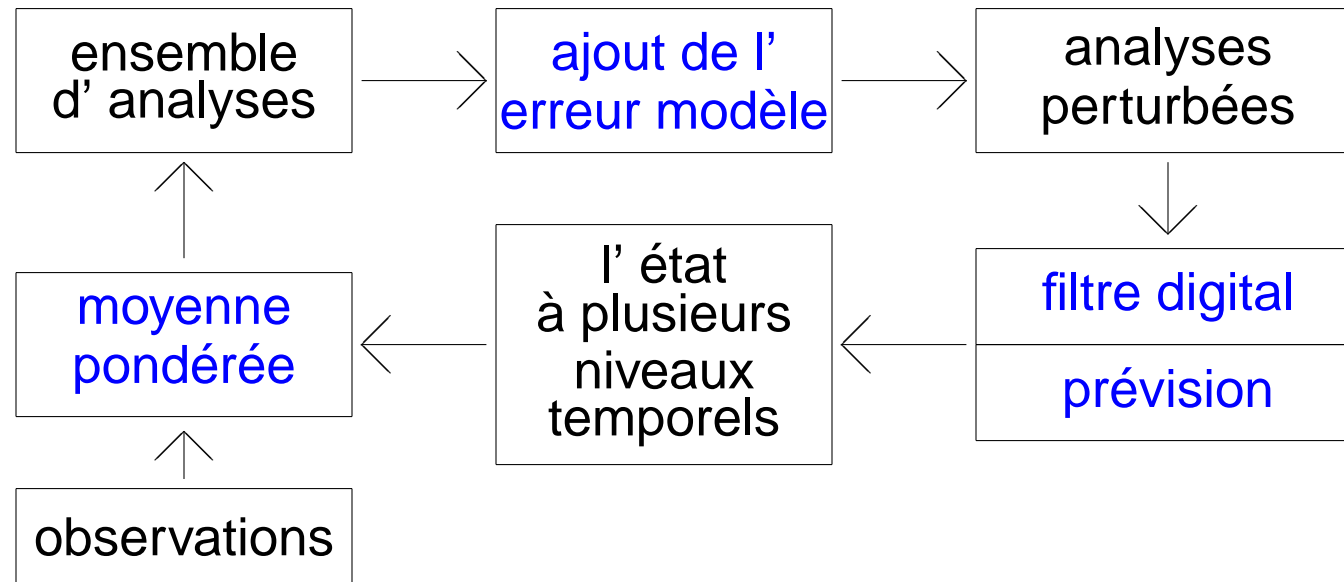
L'erreur paramétrisée du modèle a une longueur de 1.39 m/s. Avec un ajout régulier d'une erreur modèle de cette taille, nous sommes capables de maintenir suffisamment d'écart entre les membres de l'ensemble.

Avec une intégration de 6 heures sans physique, nous avons obtenu une différence de 1.52 m/s. Il est difficile de croire que la physique est 90 % en erreur.

Alternativement, nous pouvons argumenter que chaque imperfection dans le cycle d'assimilation doit éventuellement être corrigée par une erreur du modèle plus grande. Alors, l'absence d'interpolation temporelle mènera par exemple à une erreur du modèle,  $Q$ , plus grande.

Il est alors justifiable d'ajouter l'erreur du modèle après l'analyse (au lieu de l'ajouter après l'intégration du modèle).

## La configuration révisée de l' EnKF



Nous avons obtenu des meilleurs résultats en déplaçant l' ajout de l' erreur du modèle et en ajoutant un filtre digital. Avec cette séquence d' opérations, nous pourrions faire une interpolation temporelle.

Après le changement de la variable du contrôle, nous allons réévaluer le besoin d' avoir un filtre digital.

## Conclusions

Nous avons étudié le comportement de l'EnKF qui tourne aux opérations et constaté certains défauts.

1. Pour corriger un problème d'équilibre, nous utiliserons une autre variable de contrôle.
2. Nous tiendrons compte de l'évolution temporelle à l'intérieur de la fenêtre d'assimilation de 6 heures. Nous ajouterons alors les valeurs de  $Hx_i^f$  à la variable de contrôle.
3. Le cycle diurne n'est pas assez intense. Une correction au modèle ou aux champs de surface serait souhaitable.

Nous pensons alors pouvoir corriger les problèmes que nous avons décelés.

C'est à suivre.

Merci de votre patience.