Assimilation de données

Modélisation des covariances d'erreur R & B

- Correction de biais
- Covariances d'erreur d'obs R
- Covariances d'erreur d'ébauche B
- Lien entre B et les fonctions de structure

Rappel: notations

x état du modèle

y vecteur d'observations

H opérateur d'observation: H(x) sont les valeurs d'obs simulées à partir de l'état du modèle x.

x_b ébauche de l'analyse

x_a analyse à calculer

x_t réalité (inaccessible)

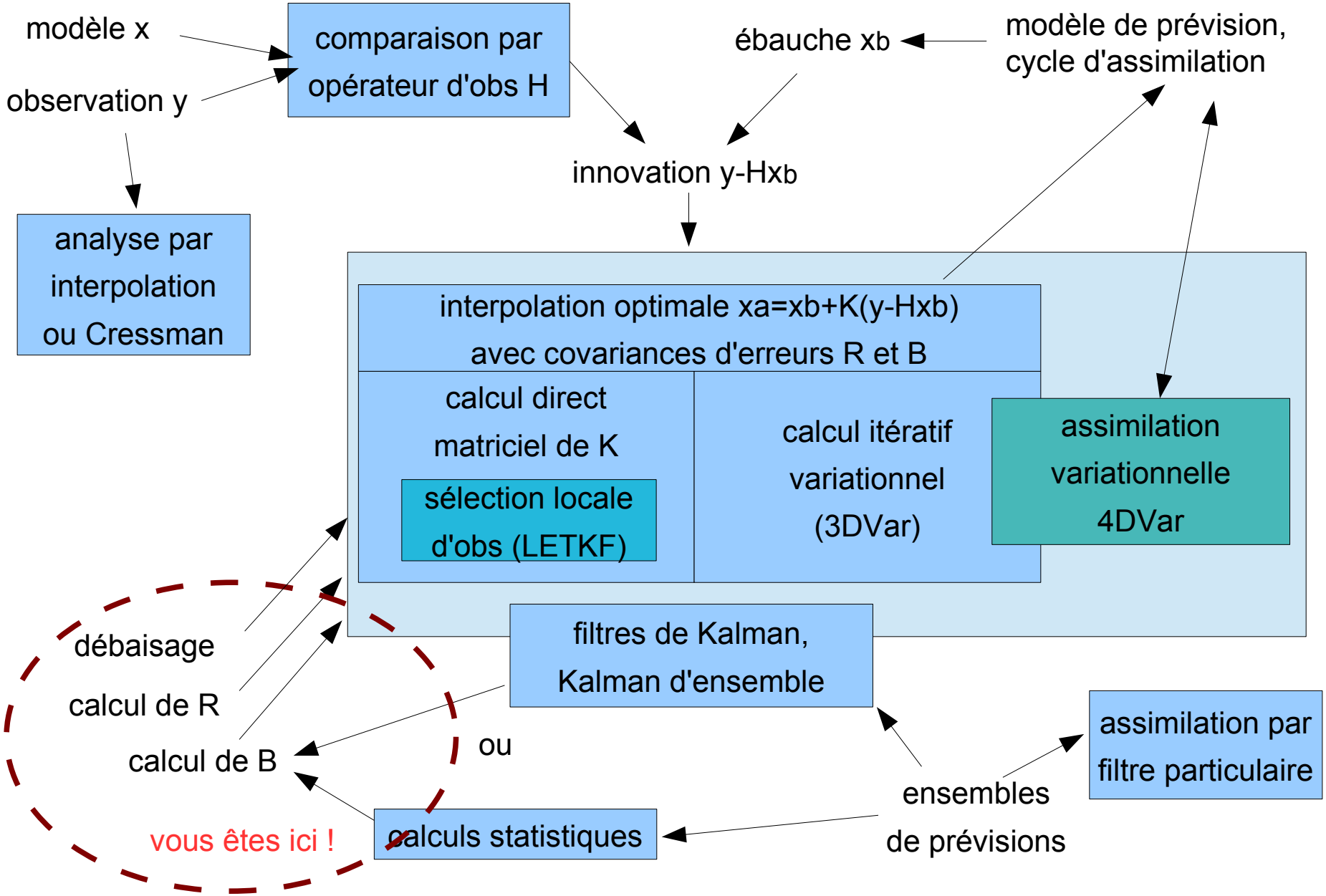
B covariances d'erreurs d'ébauche $=\mathrm{cov}(\, x_b - x_t \,,\, x_b - x_t \,)$

A covariances d'erreurs d'analyse $=\mathrm{cov}(\, x_a - x_t \,,\, x_a - x_t \,)$

R covariances d'erreur d'observations $=\mathrm{cov}(\, y - H(x_t) \,,\, y - H(x_t) \,)$

en entrée: x_b,y,H,B,R en sortie: x_a,A

Rappel: organisation du cours



Rappel: équations du BLUE

L'analyse optimale est $\mathbf{x_a = x_b + K\,(\,y - H[x_b]\,)}$

avec $\mathbf{K = B\,H^T\,(H\,B\,H^T + R\,)^{-1}}$

Forme variationnelle: x_a s'obtient aussi par $\mathbf{x_a = Arg\ min\ (J)}$

avec $\mathbf{J(x) = (x - x_b)^T\,B^{-1}\,(x - x_b) + (y - H[x])^T\,R^{-1}\,(y - H[x])}$

Questions:

- comment construire R & B ? Elles doivent représenter les erreurs $(y - Hx_t)$ et $(x_b - x_t)$

- comment gérer les biais ?

Principes d'estimation des statistiques: biais, R, B

Ce sont des propriétés statistiques à un instant donné d'erreurs non-mesurables (= moyennes d'ensembles sur des réalités alternatives).

• méthode 1: **estimer les statistiques à partir d'échantillons approximatifs**: archives, climatologies, prévisions anciennes + hypothèses simplificatrices pour généraliser les moyennes:

- *ergodicité:* moyenne dans le temps = moyenne à 1 instant
- *homogénéité:* moyenne spatiale = moyenne en 1 point
- *isotropie:* moyenne directionnelle = moyenne 2D (de covariances)
- *valables uniquement si le système modélisé se comporte de manière homogène dans les échantillons ainsi définis.*

• *méthode 2:* **simuler directement les erreurs:** construction d'*ensembles* x_i dont les variances, corrélations, etc approximent les écarts x_i-x_t. Méthodes: **prévision d'ensemble**, **filtre de Kalman d'ensemble**, **filtre à particules**. *Complexes et chères.*

Correction de biais

déf. d'un biais d'obs ou d'ébauche: erreur systématique et significative (par rapport aux variances d'erreur). Cela produit un biais dans l'analyse BLUE.

Exemples:

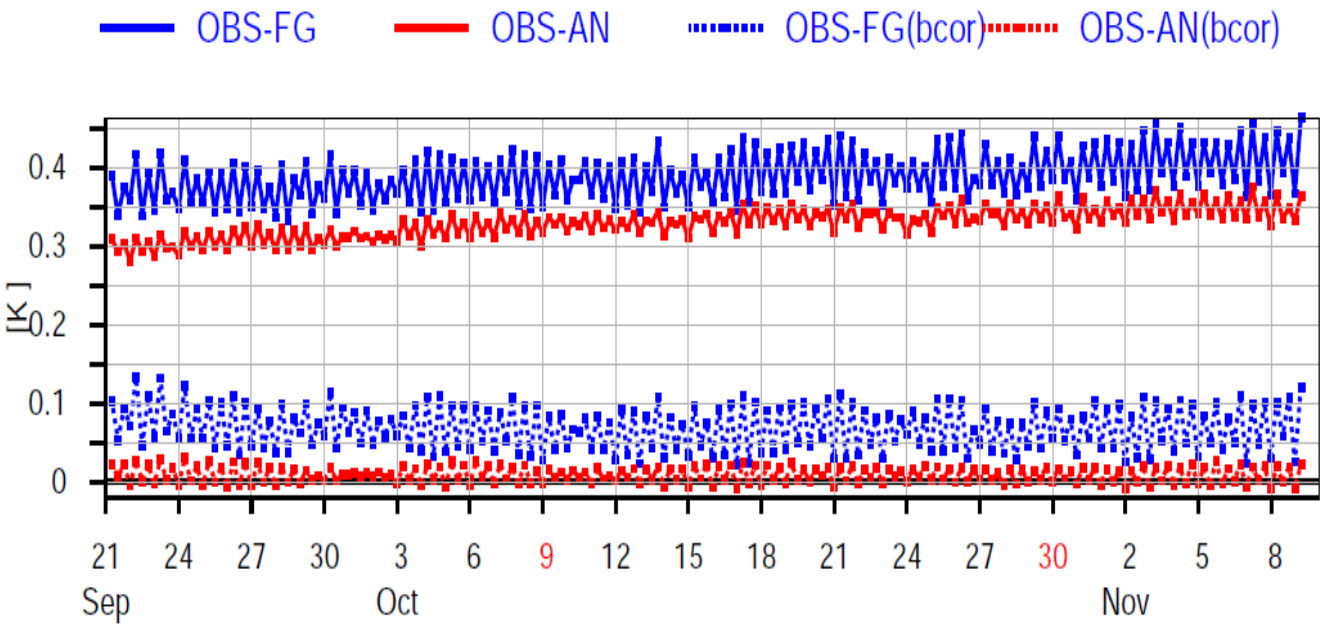
- biais d'obs: **erreur instrumentale** systématique (ex.: altitude d'un baromètre), ou **erreur dans l'opérateur H** (ex.: transfert radiatif pour simuler des radiances)
- biais d'ébauche: **dérive systématique du modèle** de prévision

Méthodes d'estimations habituelles:

- hypothèse de persistence: on suppose que le biais évolue lentement, calcul des corrections de biais par moyenne glissante
- il faut moyenner sur un échantillon assez grand pour isoler la composante systématique des erreurs, et filtrer les erreurs aléatoires.
- regarder les moyennes d'**innovations $y-Hx_b$, incréments x_a-x_b, écarts entre obs.**

Correction de biais: exemple (radiances satellitaires dans un modèle météo)

Statistics for RADIANCES from NOAA-15/AMSUA (Global)
Channel =10, All data [time step = 6 hours]
Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0 (over All_surfaces)
EXP = 0001



y-H(xb) moyen toutes les 6h

y-H(xa)

ŷ-H(xb) après débiaisage de y

ŷ-H(xa)

On peut: remplacer l'obs brute y par sa version débiaisée ŷ = y - moy[y-H(xb)]

Correction de biais

Cas du biais d'un type d'obs donné: calcul par moyenne sur un jeu de données d'apprentissage:

$$\text{moy}[\ y - H(x_b)\ ] = b$$

Pour les obs de ce type-là, supposé de biais homogène, on remplace dans l'analyse y par $\hat{y}=y-b$ qui respecte l'hypothèse de nullité du biais:

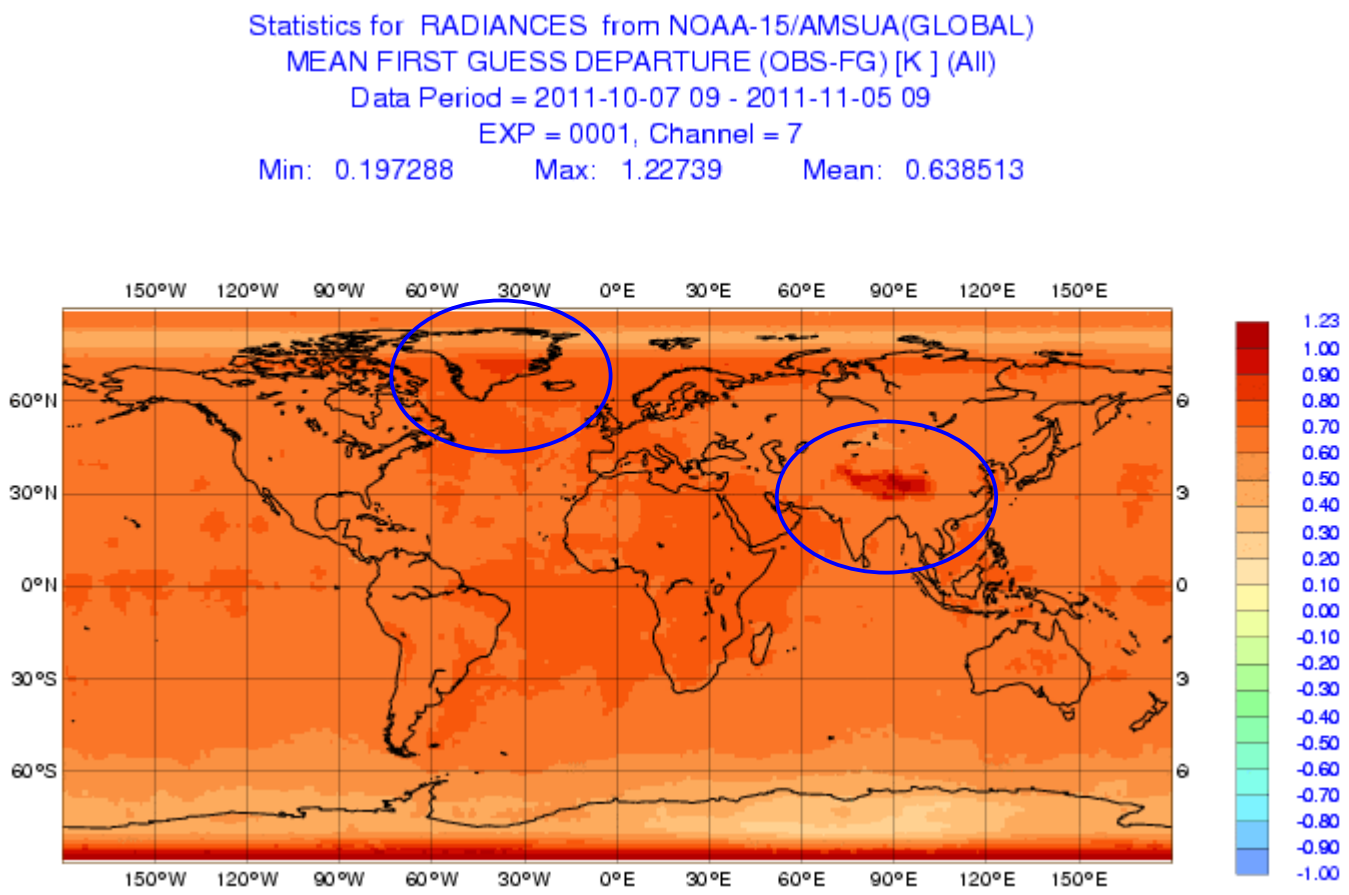
$$\text{moy}[\ \hat{y} - H(x_b)\ ] = 0$$

Ce qui revient à n'utiliser que les variations du signal observé autour de sa moyenne

On peut aussi quelquefois identifier les raisons physiques du biais, et le modéliser dans H.

Correction de biais

ex: carte des biais y-H(x_b) en moyenne sur une longue période



Estimation des covariances d'erreur d'observation R

Ce sont les **covariances de y-H(x_t)**

rappel: **cov(a_i,a_j)= m{ [a_i-m(a_i)][a_j-m(a_j)] }** donc après débiaisage: **cov(a_i,a_j)= m(a_i a_j)** pour tous les couples d'obs **(y_i , y_j)**

• coeffs diagonaux: **variances d'erreurs d'obs**, généralement *fonction du type d'instrument*, ex: $\sigma_o \approx$1m/s pour le vent, 0.5K pour T, 0.1hPa pour psol...

• on peut aussi utiliser le ***variogramme*** de y-H(x_b): cf. plus loin

• coeffs non-diagonaux: proportionnels aux **corrélations** **ρ(a_i,a_j)**, souvent supposées nulles (si les instruments de mesure sont indépendants)

Estimation des variances d'erreur par colocalisation (=validation croisée)

1) Avec une liste de couples d'obs (a,b) colocalisées qui observent le même paramètre avec 2 instruments A et B: la variance des différences entre obs est

$$\mathrm{var}(a - b) = \mathrm{var}(a - Hx_t + Hx_t - b)$$
$$= \mathrm{var}(a - Hx_t) + \mathrm{var}(b - Hx_t) + \mathrm{cov}(a-Hx_t, b-Hx_t)$$

- $\mathrm{cov}(a-Hx_t, b-Hx_t)=0$ en supposant les erreurs d'obs non corrélées
- **donne la variance d'erreur d'obs de l'instrument a si celle de b est connue**
- **si B est bien meilleur que A, var(a-b) ~ var(a - Hx_t)**

2) En comparant des obs et des ébauches:

$$\mathrm{var}(y - Hx_b) = \mathrm{var}(y - Hx_t) + H^T BH + \mathrm{cov}(y- Hx_t, Hx_b - Hx_t)$$

- $\mathrm{cov}(y - Hx_t, Hx_b - Hx_t)=0$ en supposant que les erreurs d'obs et d'ébauche sont non corrélées entre elles
- **donne la variance d'erreur d'obs si celle d'ébauche est connue (et vice versa)**

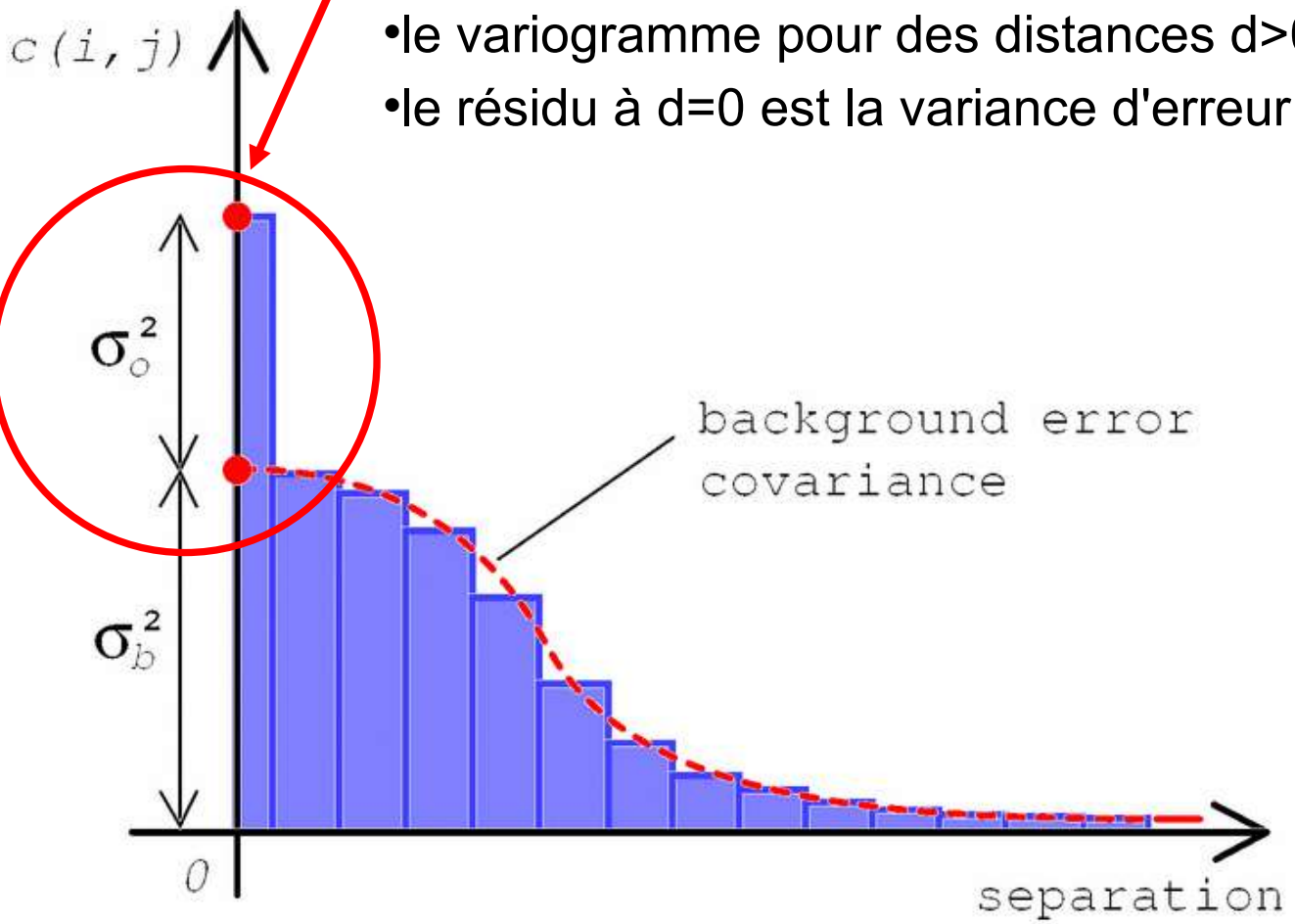
(attention les biais doivent être nuls)

Estimation de R et de B par variogramme

- on construit un variogramme des écarts obs - prévisions:

$$cov(y-Hx_b) \sim cov(y-Hx_t)+cov(Hx_t-Hx_b) = HBH^T + R$$

- on suppose que R n'est pas corrélée spatialement, et que B l'est, pour séparer HBH^T de R:
- le variogramme pour des distances d>0 est celui de HBH^T
- le résidu à d=0 est la variance d'erreur d'obs



Résumé pratique sur l'estimation de R et B

- Commencer par estimer les biais, en regardant les écarts obs-modèle et prévision-analyses: soustraire les biais des obs et des ébauches autant que possible.
- Erreurs d'obs:
  - les erreurs instrumentales souvent connues par ailleurs
  - erreur de représentativité et corrélation d'erreur d'obs: à estimer si vraiment significatives

- Erreurs d'ébauche B:
  - méthodes avancées: colocations d'obs, variogrammes d'erreur de prévision (estiment à la fois R et B)
  - cf méthodes plus pratiques ci-dessous: modèles physiques de B, et ensembles

interprétation physique de $B H^T$ et $H B H^T$

Une bonne compréhension de la physique du système à analyser permet de simplifier la construction de B:

- B contient les ***covariances d'erreurs d'ébauche*** (càd de prévision) entre les variables du modèle x.
- H est un opérateur qui convertit (interpole) l'état du modèle en variables observées. C'est un "simulateur de mesures".

Donc:

- **BH^T** est une matrice B dont les lignes ont été converties par H. C'est une matrice de covariances d'erreur d'ébauche entre toutes les variables du modèle, et toutes les variables observées.
- **HBH^T** est une matrice de covariance d'erreur d'ébauche entre tous les couples de variables observées.
- modéliser B pour des variables modèle non-observées nécessite des **hypothèses sur la structure des erreurs d'ébauche x_b-x_t.**

Propriétés fondamentales de B

Covariances des vecteurs (x_b - x_t) pour tous les couples de variables du modèle **(x_i,x_j)**

- c'est une matrice de taille dimx.dimx
- <u>coeffs diagonaux:</u> **variances d'erreurs de l'ébauche**, elles sont **plus élevées dans les situations peu prévisibles**
- <u>coeffs non-diagonaux:</u> proportionnels aux **corrélations** **ρ(x_i,x_j)**, reflètent des propriétés physiques du modèle:
    - caractère plus ou moins **lisse** des champs (et donc de leurs erreurs)
    - **équilibre entre différentes variables ('balance')**: non-divergence, équilibres hydrostatiques et géostrophiques, pompage d'Ekman.

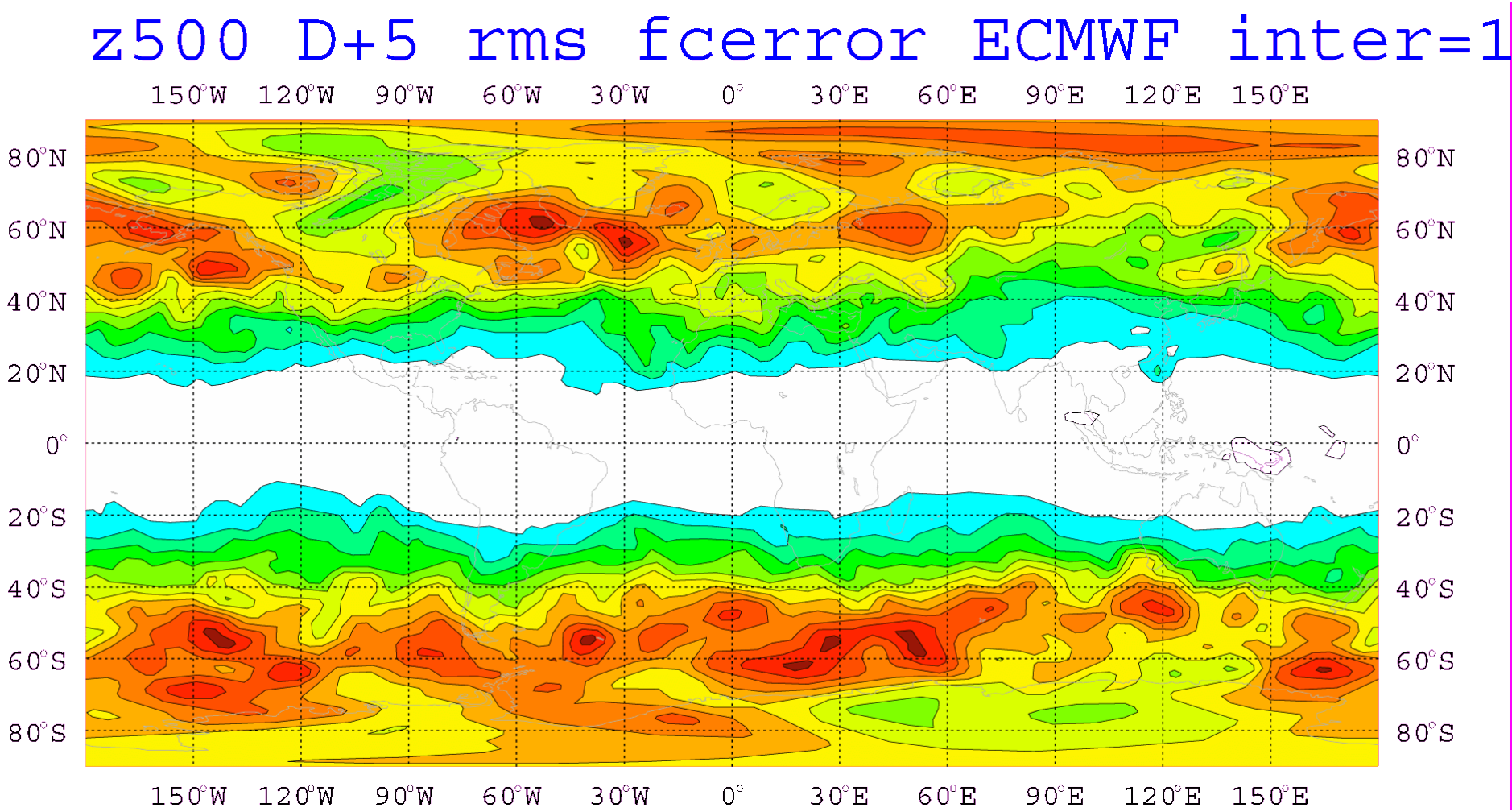
Dans un fluide chaotique, les corrélations adoptent la structure des modes d'instabilités locaux = erreurs d'ébauche les plus probables, que l'analyse devra corriger

Un bon modèle numérique peut simuler ces caractéristiques, donc aide à spécifier une grande partie de B.

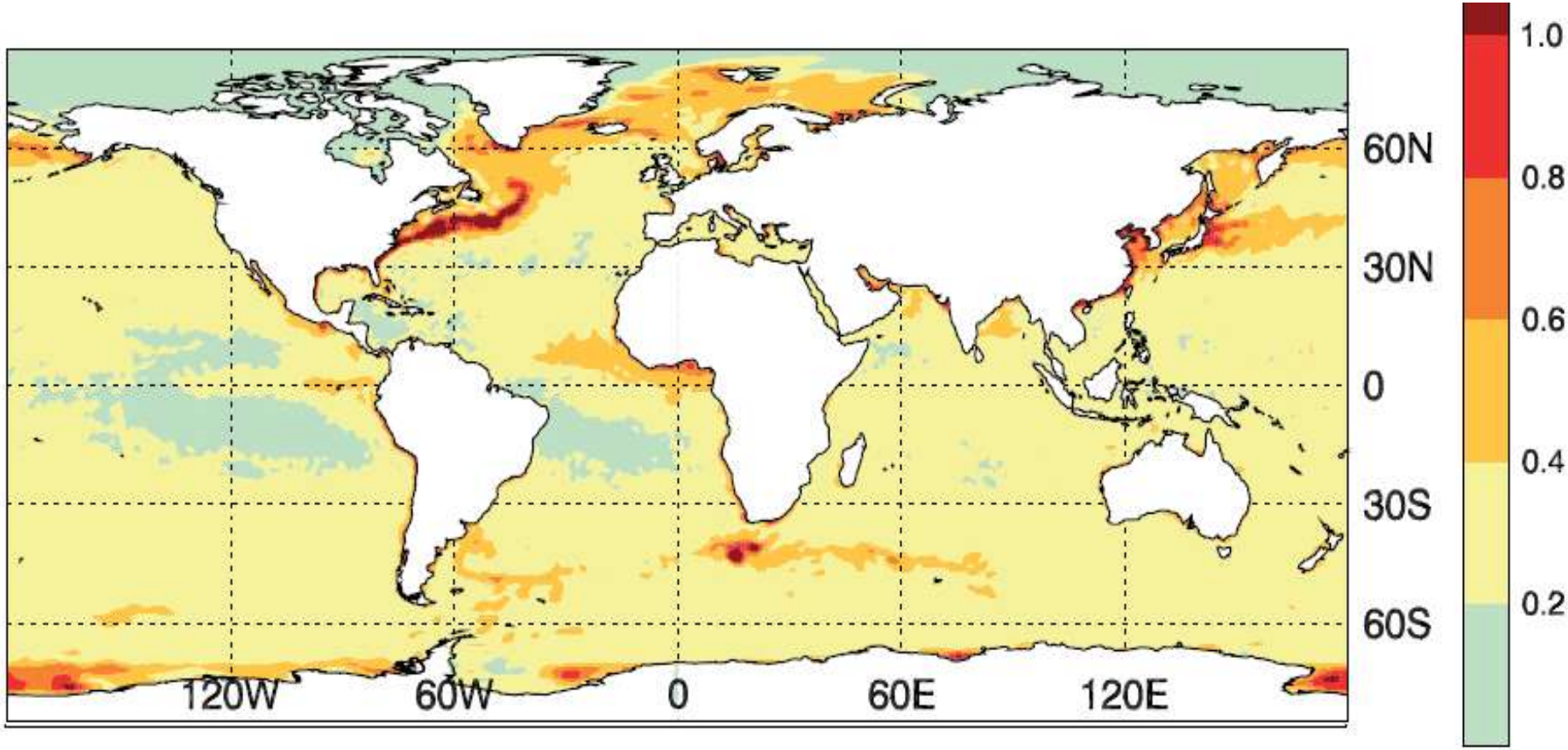
estimation des variances d'erreur d'ébauche dans l'atmosphère avec des statistiques d'erreur de prévision

Variances moyennes d'erreurs de prévision x(t+5jours) - x_a

On peut supposer qu'elles sont proportionnelles aux erreurs d'ébauche (=prévisions à échéance 6h)



estimation des variances d'erreur d'ébauche dans l'océan avec des stats d'écarts obs-ébauche



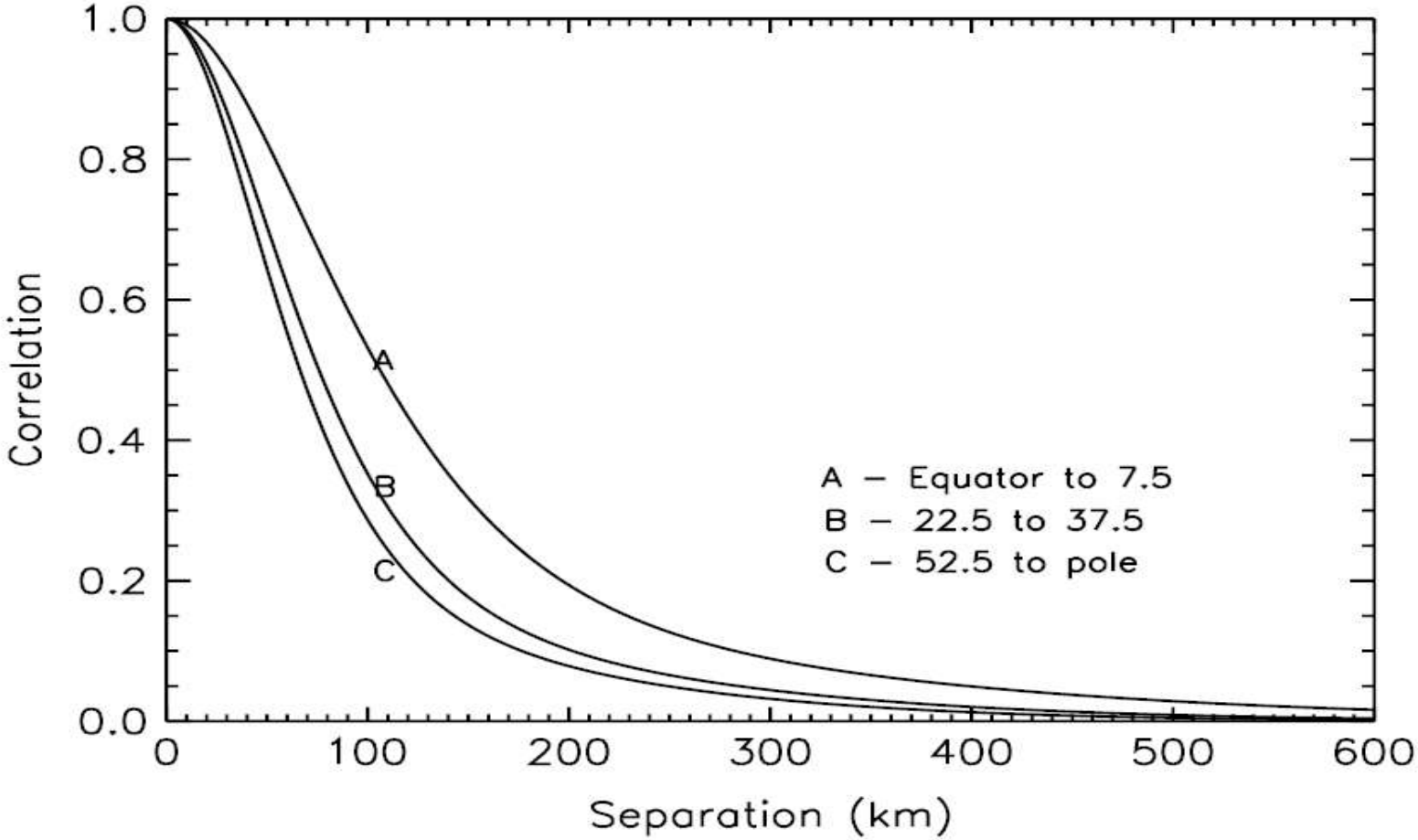
Estimated background standard deviations (K) of SST for January to March three years of data.

Estimation de corrélations d'erreur d'ébauche de SST à partir d'écarts obs-prévision

Aici observations satellitaires

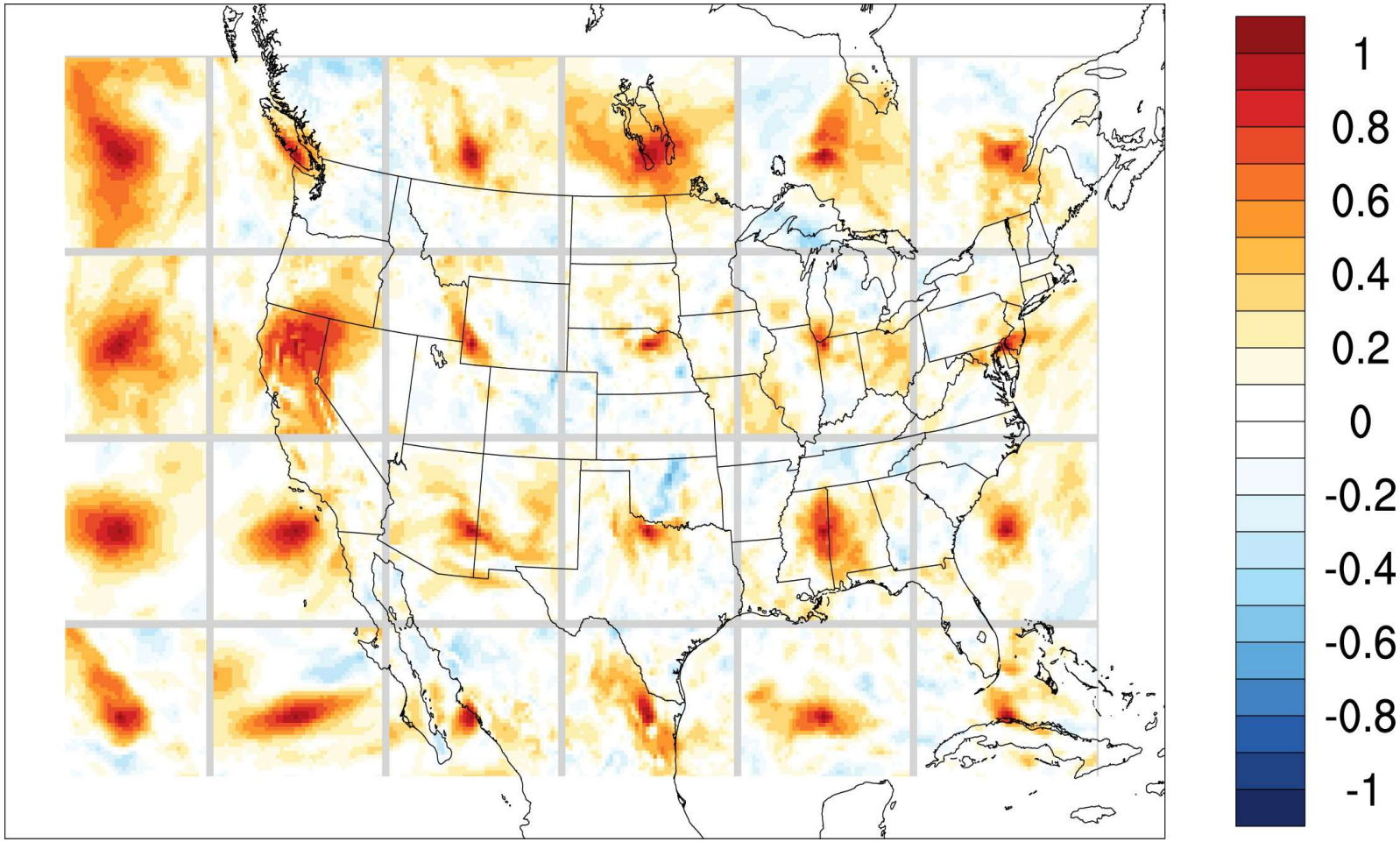
Bvariogramme sous hypothèse d'isotropie & homogénéité par domaine

ajusté par des fonctions analytiques définies positives (auto-régressives)



Estimation de corrélations d'erreur d'ébauche dans l'atmosphère

(corrélations horizontales de température)



Principaux équilibres dans l'atmosphère et l'océan:

- Dépendent des échelles de temps & d'espace
- Pas exactement respectés, mais presque
- Induisent des corrélations multivariées

conservation de la masse: pour toute espèce quasi-conservative, div(rU)~0 (r=concentration volumique)

équilibre hydrostatique: $dp/dz \sim - \rho g$

équilibre géostrophique: $U = 1/f\ k \wedge \nabla (gz)$

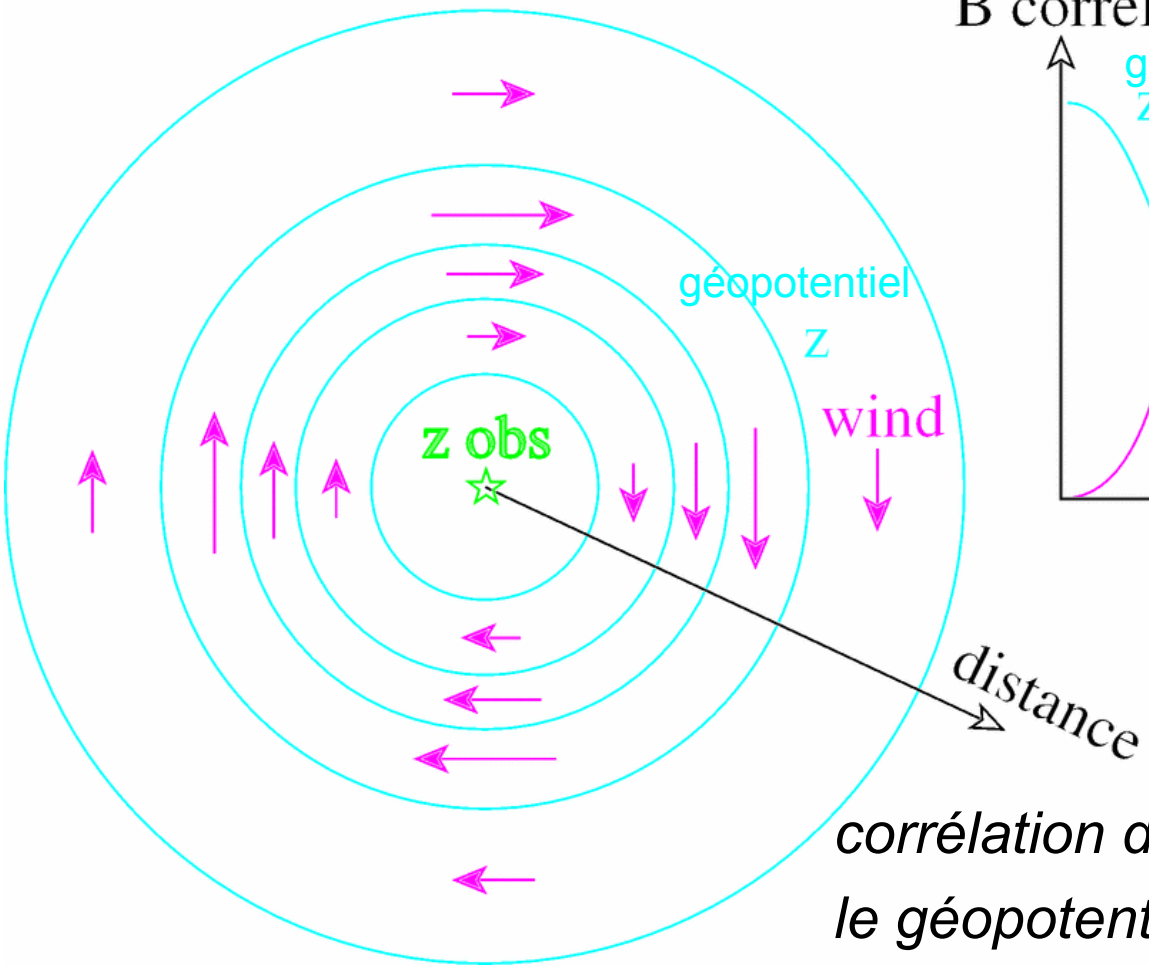
- k vecteur vertical
- z géopotentiel (altitude d'isobare)

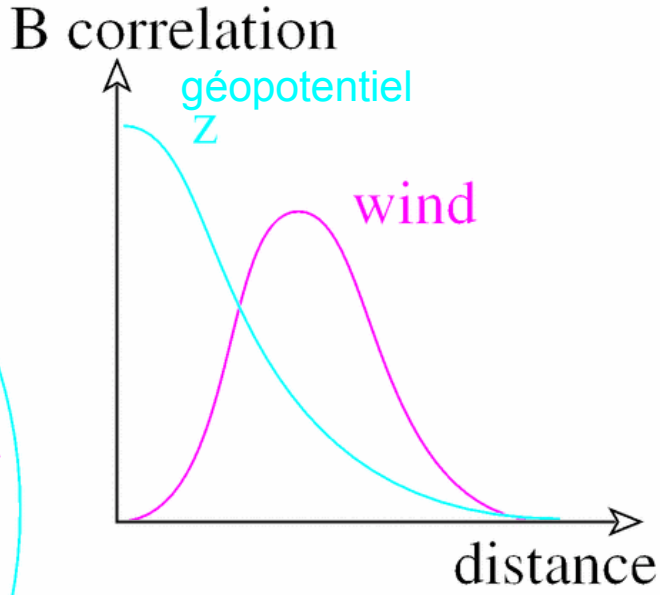
pompage d'Ekman: convergence de vent et humidité de basses couches dans les dépressions.

Estimation de corrélations entre variables en météo & océano: notion de 'balance'

l'équilibre géostrophique à grande échelle impose des corrélations entre courant et masse = f(pression, salinité, T, géopotentiel)

Hémisphère Nord: géopotentiel élevé = anticyclone = rotation horaire (en général)





corrélation de vent & de z avec le géopotentiel en 1 point

relation entre matrice B et fonctions de structure

Fonction de structure d'une analyse: forme de *l'incrément* (x_a - x_b) produit séparément par chaque obs y.

Equation du BLUE:

$$x_a = x_b + K\,(y - H[x_b]) \text{ avec } K = B\,H^T\,(H\,B\,H^T + R)^{-1}$$

<u>S'il n'y a qu'une obs</u>: dimy=1, dim($H\,B\,H^T + R$)=1x1: alors,

(x_a-x_b) est proportionnel à $B\,H^T$

- **les fonctions de structures sont les covariances de B** (plus exactement, les colonnes de la matrice BH^T)
- en l'absence de corrélation d'erreur d'obs, l'incrément d'analyse est une combinaison linéaire entre fonctions de structure élémentaires. Les poids de cette combinaison dépendent de la qualité des obs et de l'ébauche.

cela permet de voir l'effet physique de B sur le système d'analyse

Analyse avec 1 seule obs: cartes d'incréments

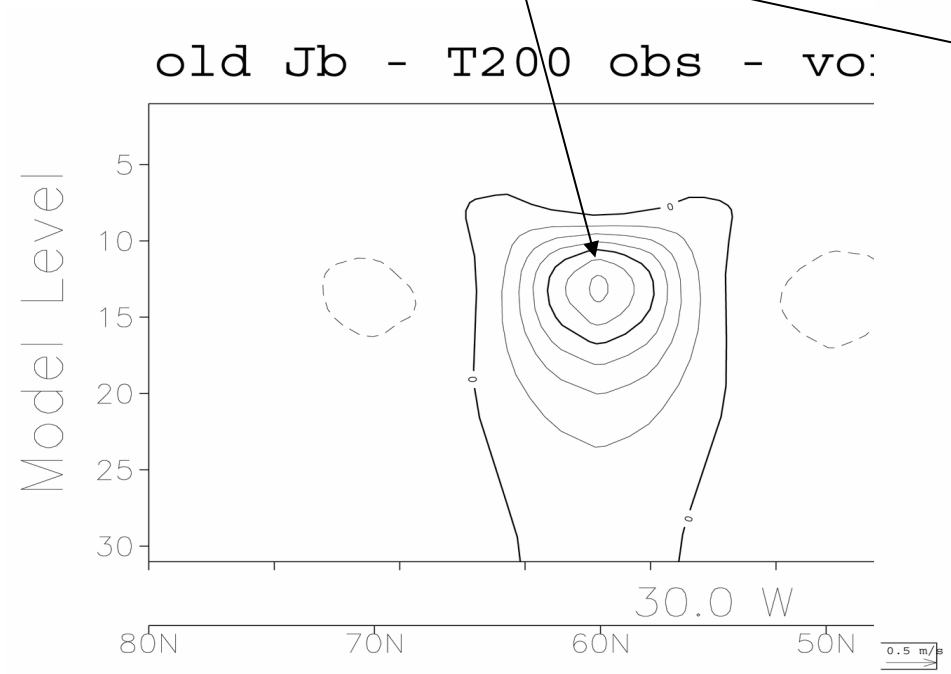
$x_a - x_b = BH^T$ = colonnes de la matrice B:

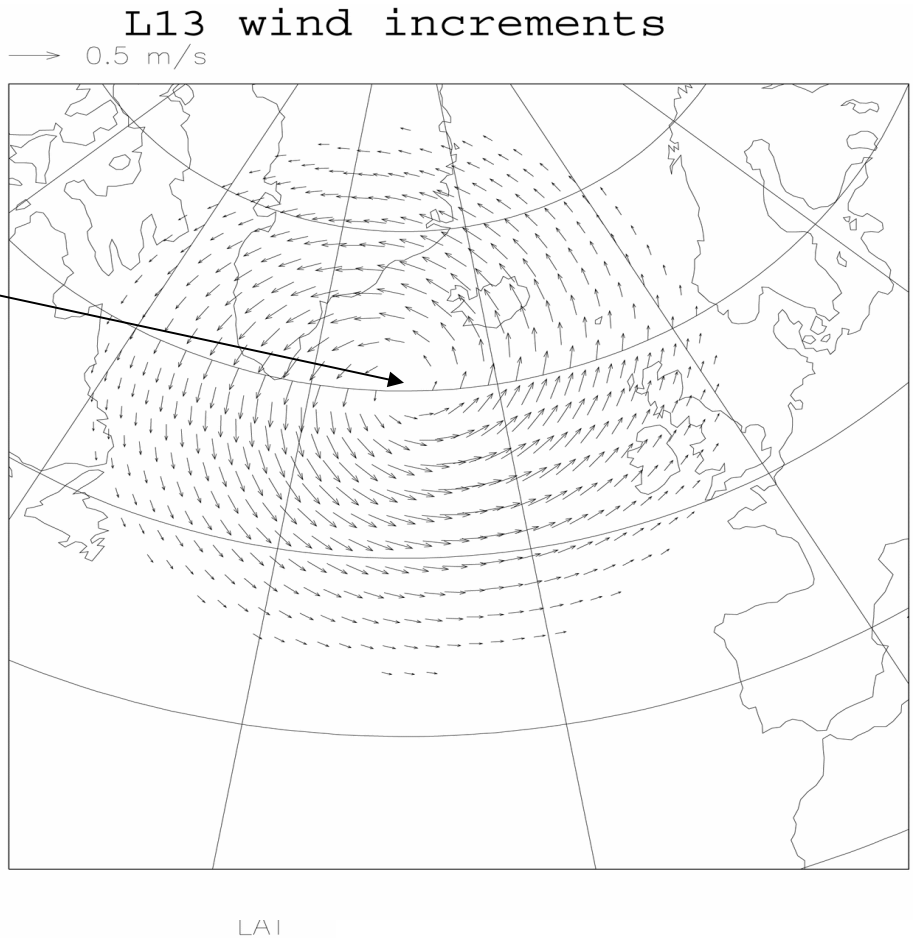
on voit les autocorrélations de B

obs avion: augmente T,

diminue le géopotentiel,

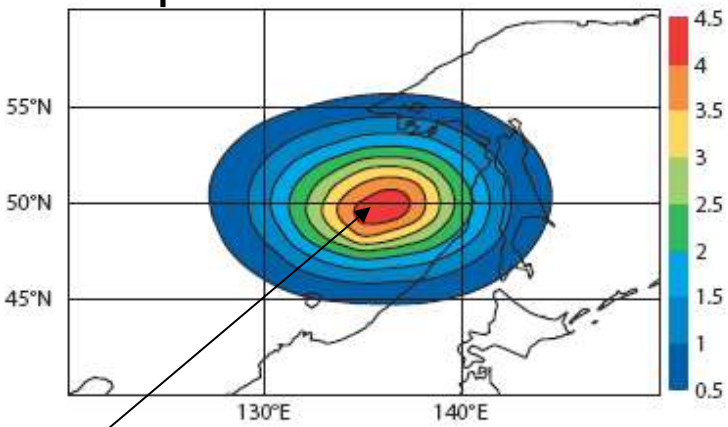
crée un tourbillon cyclonique



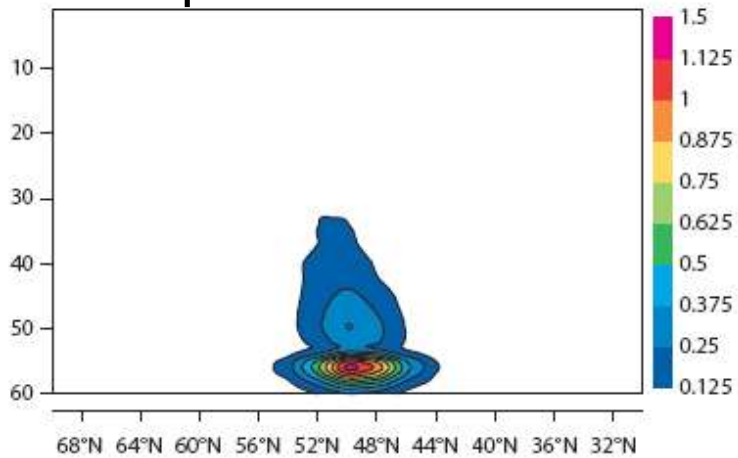


Exemple d'incrément produit par une obs chimique simulée (analyse GEMS de formaldéhyde)

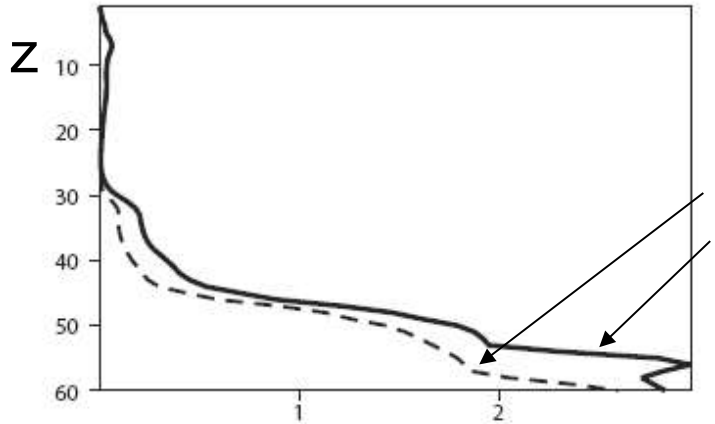
coupe horizontale



coupe verticale



1 obs d'intégrale verticale d'HCHO



analyse et ébauche, profil vertical

Figure 15: Total column HCHO analysis increment (top left) in 10^{15} mol/cm^2, vertical cross section of analysis increment at 136.4° E in ppb (top right), and HCHO analysis (solid) and first-guess (dashed) profiles in ppb (bottom) from a single HCHO observation placed at 49.4° N, 136.4° E on 20060701, at 01:31:18 hours. The observation has a value of 30x10^{15} mol/cm^2 and an error of 20%, and is 15.7x10^{15} mol/cm^2 higher than the background.

Modélisation des covariances d'ébauche B dans de grands modèles: dimx >>1000

- B trop grosse pour être stockée, on n'a besoin que de coder l'opérateur $x \rightarrow Bx$
- B doit être symétrique définie positive:

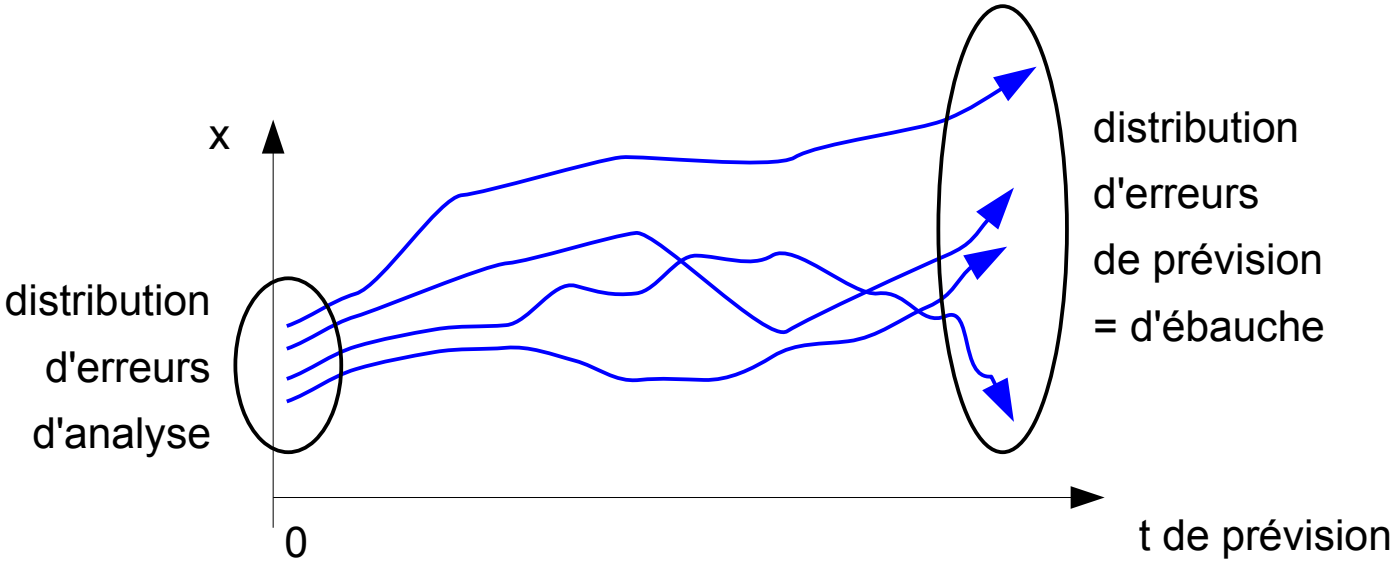
  pour tout $x \neq 0$, tout y: $x^T Bx > 0$ et $x^T By = y^T Bx$
- on construit B comme produit d'opérateurs linéaires simples:
  - variance = modulation d'amplitude
  - corrélations 1D/2D = opérateurs de lissage = filtres digitaux
  - corrélations 3D = produits d'opérateurs 1D et 2D
  - corrélations multivariées = opérateurs « de balance », transformations linéarisées des variables manipulées
- on peut construire différents B dans des sous-espaces orthogonaux

exemples:

- *lissage par des filtres digitaux*
- *balance géostrophique entre vent et pression= décomposition des erreurs en équilibre géostrophique et résidu:* $\delta p = B(\delta u, \delta v) + \delta p_u$

Méthodes "stochastiques"
= Estimation de B par des ensembles

- méthode plus pratique que le calcul de B à partir d'observations
- Idée: **construire un échantillon d'états de modèle x_i ("ensemble") dont la loi de probabilité est semblable à celle de x_b-x_t**
- les covariances des différences entre membres de l'ensemble: $cov(x_i-x_j)$ sont alors identiques à celles de B, sauf qu'il faut diviser les variances par 2.
- on peut poser $B = {}^1/_{2(n-1)} \sum_i (x_i - \bar{x}_i) (x_i - \bar{x}_i)^T + \ldots$



Merci pour votre attention