



**METEO
FRANCE**

Optimisation conjointe de la localisation et de l'hybridation pour filtrer des covariances échantillonnées

Benjamin Ménérier

Colloque National d'Assimilation de données

Grenoble - 2 décembre 2016

Introduction

Contexte :

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.
- Un bruit d'échantillonnage affecte cette estimation car la taille de l'ensemble est limitée.

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.
- Un bruit d'échantillonnage affecte cette estimation car la taille de l'ensemble est limitée.
- Question : comment réduire ce bruit d'échantillonnage ?

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.
- Un bruit d'échantillonnage affecte cette estimation car la taille de l'ensemble est limitée.
- Question : comment réduire ce bruit d'échantillonnage ?

Méthodes usuelles :

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.
- Un bruit d'échantillonnage affecte cette estimation car la taille de l'ensemble est limitée.
- Question : comment réduire ce bruit d'échantillonnage ?

Méthodes usuelles :

- Localisation de covariances
 - amortissement à longue distance des covariances

Introduction

Contexte :

- L'assimilation de données repose en général sur des covariances d'erreur d'ébauche.
- Celles-ci peuvent être estimées à partir d'un ensemble.
- Un bruit d'échantillonnage affecte cette estimation car la taille de l'ensemble est limitée.
- Question : comment réduire ce bruit d'échantillonnage ?

Méthodes usuelles :

- Localisation de covariances
→ amortissement à longue distance des covariances
- Hybridation de covariances
→ combinaison linéaire de plusieurs matrices de covariance

Introduction

Questions :

Introduction

Questions :

1. Peut-on optimiser la localisation avec une méthode :

Introduction

Questions :

1. Peut-on optimiser la localisation avec une méthode :
 - n'utilisant **que les données de l'ensemble**,

Introduction

Questions :

1. Peut-on optimiser la localisation avec une méthode :
 - n'utilisant **que les données de l'ensemble**,
 - applicable pour des **systèmes de grande dimension**.

Introduction

Questions :

1. Peut-on optimiser la localisation avec une méthode :
 - n'utilisant **que les données de l'ensemble**,
 - applicable pour des **systèmes de grande dimension**.

2. Peut-on optimiser **conjointement** la localisation et l'hybridation de covariances ?

Echantillonnage de covariances

Un ensemble de N prévisions $\{\tilde{\mathbf{x}}_p^b\}$ est utilisé pour échantillonner la matrice de covariances $\tilde{\mathbf{B}}$:

$$\tilde{\mathbf{B}} = \frac{1}{N-1} \sum_{p=1}^N \delta\tilde{\mathbf{x}}^b (\delta\tilde{\mathbf{x}}^b)^T$$

où :

$$\delta\tilde{\mathbf{x}}_p^b = \tilde{\mathbf{x}}_p^b - \langle \tilde{\mathbf{x}}^b \rangle \quad \text{et} \quad \langle \tilde{\mathbf{x}}^b \rangle = \frac{1}{N} \sum_{p=1}^N \tilde{\mathbf{x}}_p^b$$

Echantillonnage de covariances

Un ensemble de N prévisions $\{\tilde{\mathbf{x}}_p^b\}$ est utilisé pour échantillonner la matrice de covariances $\tilde{\mathbf{B}}$:

$$\tilde{\mathbf{B}} = \frac{1}{N-1} \sum_{p=1}^N \delta\tilde{\mathbf{x}}^b (\delta\tilde{\mathbf{x}}^b)^T$$

où :

$$\delta\tilde{\mathbf{x}}_p^b = \tilde{\mathbf{x}}_p^b - \langle \tilde{\mathbf{x}}^b \rangle \quad \text{et} \quad \langle \tilde{\mathbf{x}}^b \rangle = \frac{1}{N} \sum_{p=1}^N \tilde{\mathbf{x}}_p^b$$

Comportement asymptotique : si $N \rightarrow \infty$, alors $\tilde{\mathbf{B}} \rightarrow \mathbf{B}$

Echantillonnage de covariances

Un ensemble de N prévisions $\{\tilde{\mathbf{x}}_p^b\}$ est utilisé pour échantillonner la matrice de covariances $\tilde{\mathbf{B}}$:

$$\tilde{\mathbf{B}} = \frac{1}{N-1} \sum_{p=1}^N \delta\tilde{\mathbf{x}}^b (\delta\tilde{\mathbf{x}}^b)^T$$

où :

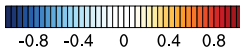
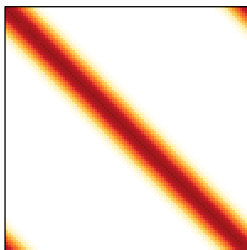
$$\delta\tilde{\mathbf{x}}_p^b = \tilde{\mathbf{x}}_p^b - \langle \tilde{\mathbf{x}}^b \rangle \quad \text{et} \quad \langle \tilde{\mathbf{x}}^b \rangle = \frac{1}{N} \sum_{p=1}^N \tilde{\mathbf{x}}_p^b$$

Comportement asymptotique : si $N \rightarrow \infty$, alors $\tilde{\mathbf{B}} \rightarrow \mathbf{B}$

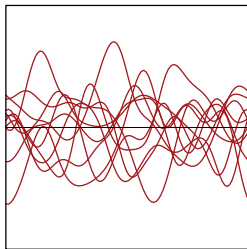
En pratique, $N < \infty \Rightarrow$ bruit d'échantillonnage $\tilde{\mathbf{B}}^e = \tilde{\mathbf{B}} - \mathbf{B}$

Echantillonnage de covariances

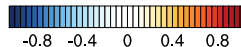
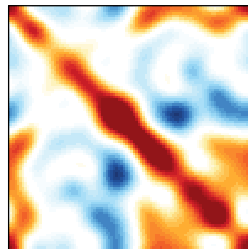
Specified B



Randomized perturbations



Sampled B (N = 10)

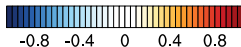
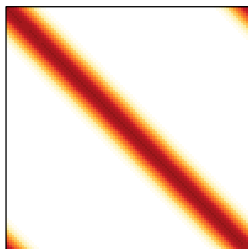


Randomisation de **B** :

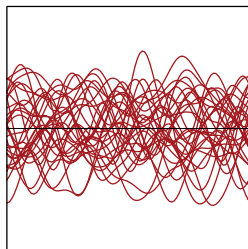
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

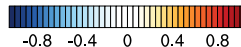
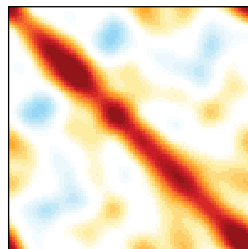
Specified B



Randomized perturbations



Sampled B (N = 30)

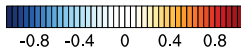
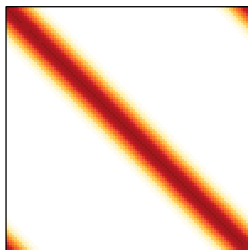


Randomisation de **B** :

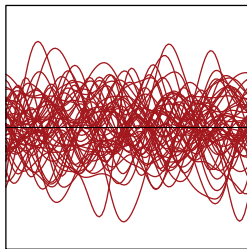
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

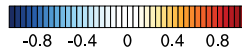
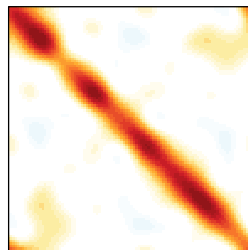
Specified B



Randomized perturbations



Sampled B (N = 50)

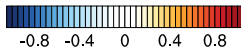
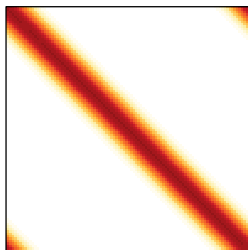


Randomisation de **B** :

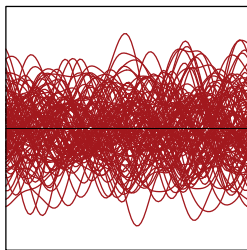
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

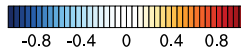
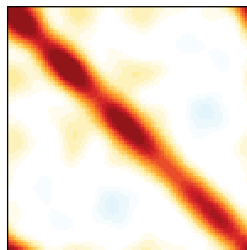
Specified B



Randomized perturbations



Sampled B (N = 100)

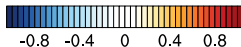
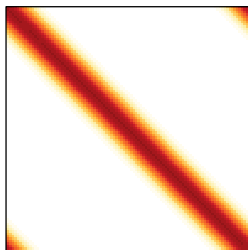


Randomisation de **B** :

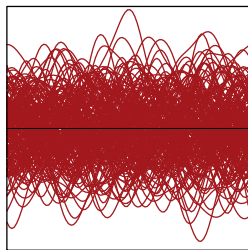
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

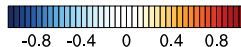
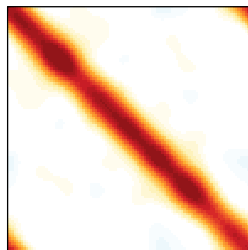
Specified B



Randomized perturbations



Sampled B (N = 200)

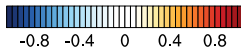
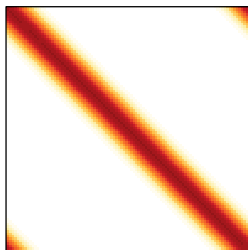


Randomisation de **B** :

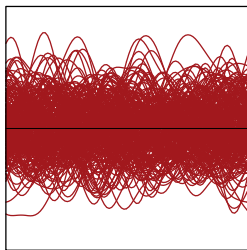
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

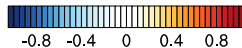
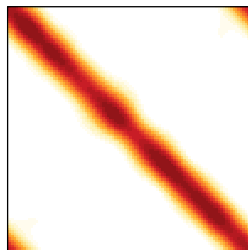
Specified B



Randomized perturbations



Sampled B (N = 500)

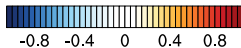
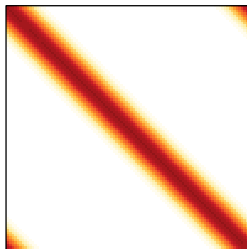


Randomisation de **B** :

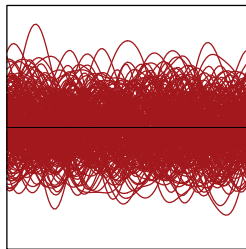
$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Echantillonnage de covariances

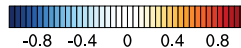
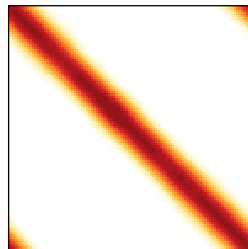
Specified B



Randomized perturbations



Sampled B (N = 1000)



Randomisation de **B** :

$$\delta \tilde{\mathbf{x}}_p^b = \mathbf{U} \boldsymbol{\eta}_p \quad \text{avec} \quad \mathbf{U} \mathbf{U}^T = \mathbf{B} \quad \text{et} \quad \boldsymbol{\eta}_p \sim \mathcal{N}(0, \mathbf{I}).$$

Plan

Introduction

Optimisation objective de la localisation

Implémentation des diagnostics

Optimisation conjointe de la localisation et de l'hybridation

Conclusions

Plan

Introduction

Optimisation objective de la localisation

Implémentation des diagnostics

Optimisation conjointe de la localisation et de l'hybridation

Conclusions

Localisation de covariance

Localisation = produit de Schur avec une matrice de localisation **L**:

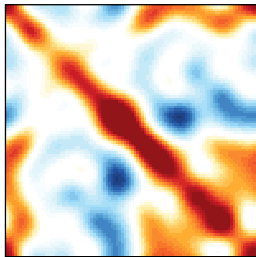
$$\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

Localisation de covariance

Localisation = produit de Schur avec une matrice de localisation **L**:

$$\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

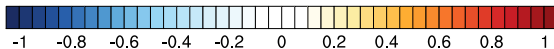
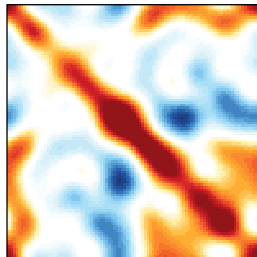
Sampled B (N = 10)



No localization



Localized B



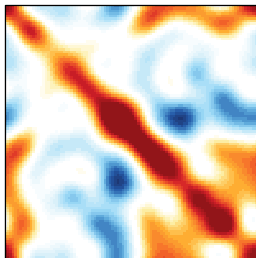
Pas d'impact

Localisation de covariance

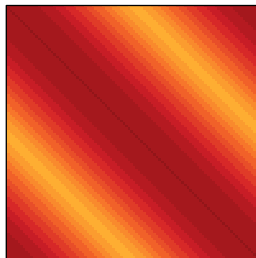
Localisation = produit de Schur avec une matrice de localisation L :

$$\hat{B} = L \circ \tilde{B} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

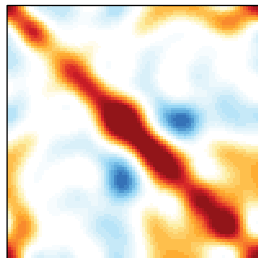
Sampled B (N = 10)



Localization (L = 0.4)



Localized B



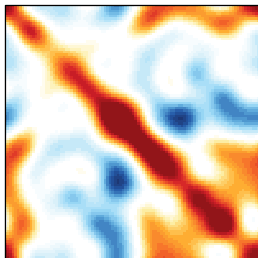
Réduction du bruit d'échantillonnage...

Localisation de covariance

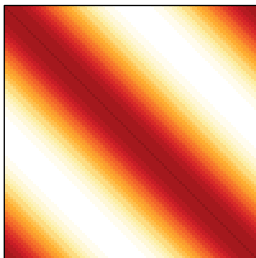
Localisation = produit de Schur avec une matrice de localisation **L**:

$$\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

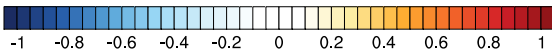
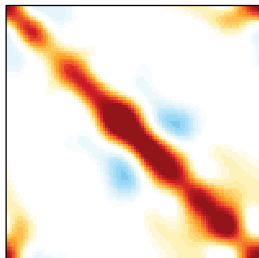
Sampled B (N = 10)



Localization (L = 0.2)



Localized B



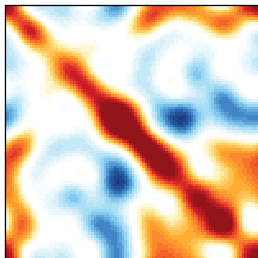
Réduction du bruit d'échantillonnage...

Localisation de covariance

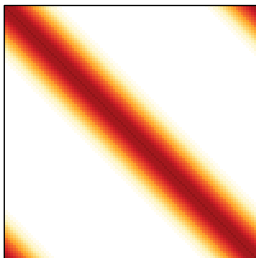
Localisation = produit de Schur avec une matrice de localisation L :

$$\hat{B} = L \circ \tilde{B} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

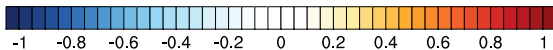
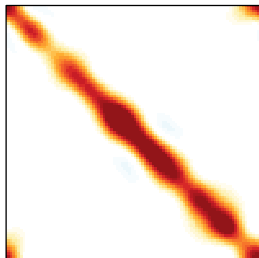
Sampled B (N = 10)



Localization (L = 0.1)



Localized B



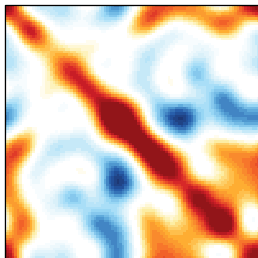
Réduction du bruit d'échantillonnage...

Localisation de covariance

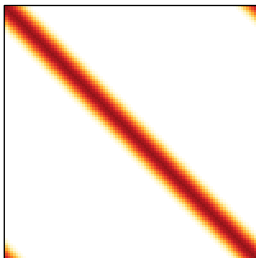
Localisation = produit de Schur avec une matrice de localisation L :

$$\hat{B} = L \circ \tilde{B} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

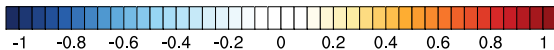
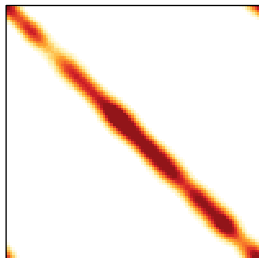
Sampled B (N = 10)



Localization (L = 0.05)



Localized B



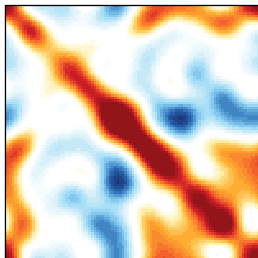
Réduction du bruit d'échantillonnage... mais aussi du signal !

Localisation de covariance

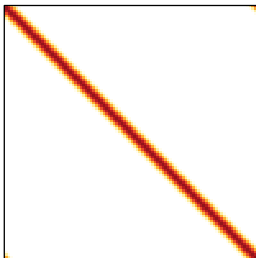
Localisation = produit de Schur avec une matrice de localisation L :

$$\hat{B} = L \circ \tilde{B} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

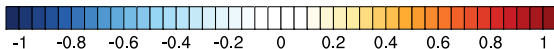
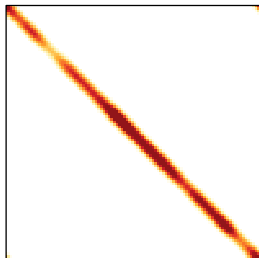
Sampled B (N = 10)



Localization (L = 0.025)



Localized B



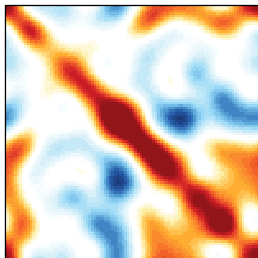
Réduction du bruit d'échantillonnage... mais aussi du signal !

Localisation de covariance

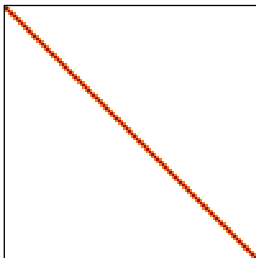
Localisation = produit de Schur avec une matrice de localisation **L**:

$$\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}} \quad \Leftrightarrow \quad \hat{B}_{ij} = L_{ij} \tilde{B}_{ij}$$

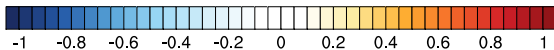
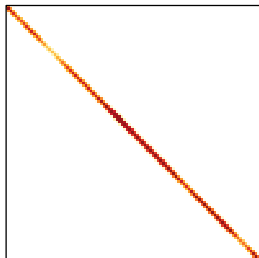
Sampled B (N = 10)



Localization (L = 0.01)



Localized B



Réduction du bruit d'échantillonnage... mais aussi du signal !

Localisation de covariance

Question : qu'est-ce qu'une localisation optimale ?

Localisation de covariance

Question : qu'est-ce qu'une localisation optimale ?

Définition *possible* : la localisation optimale est celle qui minimise

$$e = \mathbb{E} \left[\|\mathbf{L} \circ \tilde{\mathbf{B}} - \mathbf{B}\|_F^2 \right]$$

où :

- $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ est la matrice de covariance localisée,
- $\mathbf{B} = \lim_{N \rightarrow \infty} \tilde{\mathbf{B}}$ est la cible,
- $\|\cdot\|_F$ est la norme de Frobenius.

Localisation de covariance

Question : qu'est-ce qu'une localisation optimale ?

Définition *possible* : la localisation optimale est celle qui minimise

$$e = \mathbb{E} \left[\|\mathbf{L} \circ \tilde{\mathbf{B}} - \mathbf{B}\|_F^2 \right]$$

où :

- $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ est la matrice de covariance localisée,
- $\mathbf{B} = \lim_{N \rightarrow \infty} \tilde{\mathbf{B}}$ est la cible,
- $\|\cdot\|_F$ est la norme de Frobenius.

En annulant les dérivées partielles de e :

$$\frac{\partial e}{\partial L_{ij}} = 0 \quad \Leftrightarrow \quad L_{ij} = \frac{\mathbb{E} \left[\tilde{B}_{ij} B_{ij} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]}$$

Localisation de covariance

Question : qu'est-ce qu'une localisation optimale ?

Définition *possible* : la localisation optimale est celle qui minimise

$$e = \mathbb{E} \left[\|\mathbf{L} \circ \tilde{\mathbf{B}} - \mathbf{B}\|_F^2 \right]$$

où :

- $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ est la matrice de covariance localisée,
- $\mathbf{B} = \lim_{N \rightarrow \infty} \tilde{\mathbf{B}}$ est la cible,
- $\|\cdot\|_F$ est la norme de Frobenius.

En annulant les dérivées partielles de e :

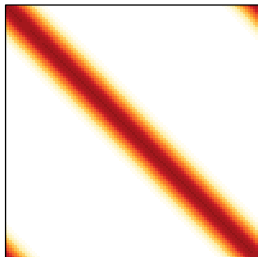
$$\frac{\partial e}{\partial L_{ij}} = 0 \quad \Leftrightarrow \quad L_{ij} = \frac{\mathbb{E} \left[\tilde{B}_{ij} B_{ij} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]} \quad \leftarrow \text{calculable ?}$$

\leftarrow calculable à partir de l'ensemble

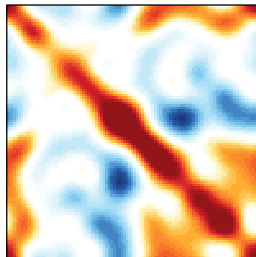
Propriétés du bruit d'échantillonnage

Variance homogène / portée homogène

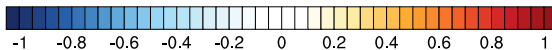
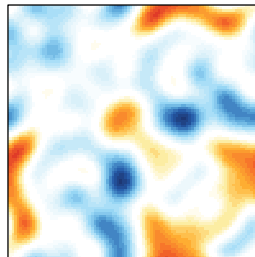
Asymptotic B ($N \rightarrow \infty$)



Sampled B ($N = 10$)



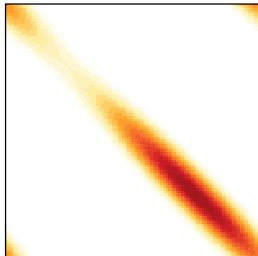
Sampling noise



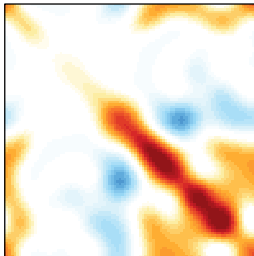
Propriétés du bruit d'échantillonnage

Variance hétérogène / portée homogène

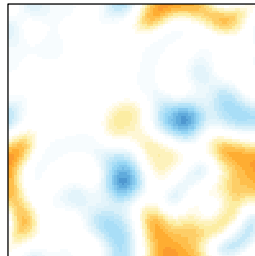
Asymptotic B ($N \rightarrow \infty$)



Sampled B ($N = 10$)



Sampling noise

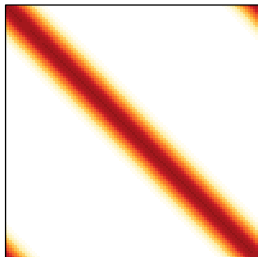


L'amplitude du bruit d'échantillonnage est liée aux variances asymptotiques.

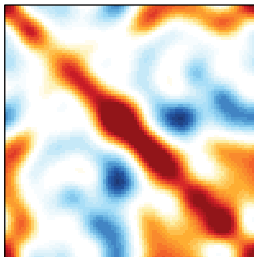
Propriétés du bruit d'échantillonnage

Variance homogène / portée homogène

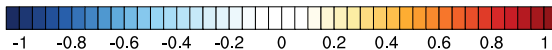
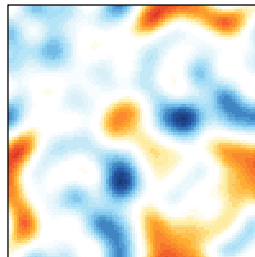
Asymptotic B ($N \rightarrow \infty$)



Sampled B ($N = 10$)



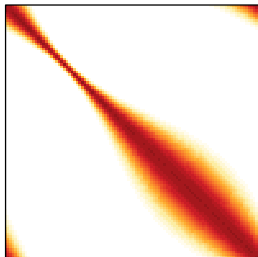
Sampling noise



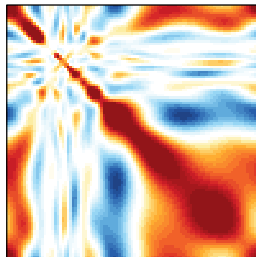
Propriétés du bruit d'échantillonnage

Variance homogène / portée hétérogène

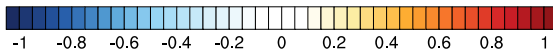
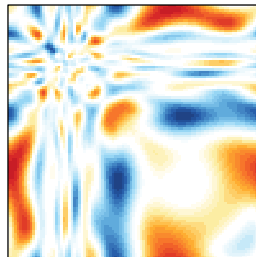
Asymptotic B ($N \rightarrow \infty$)



Sampled B ($N = 10$)



Sampling noise



La portée du bruit d'échantillonnage est liée
aux portées asymptotiques.

Théorie du bruit d'échantillonnage

Conclusion intuitive : il existe une relation statistique entre bruit d'échantillonnage et covariances asymptotiques.

Théorie du bruit d'échantillonnage

Conclusion intuitive : il existe une relation statistique entre bruit d'échantillonnage et covariances asymptotiques.

Les covariances échantillonnées sont sans biais : $\mathbb{E}[\tilde{B}_{ij}] = \mathbb{E}[B_{ij}]$

Théorie du bruit d'échantillonnage

Conclusion intuitive : il existe une relation statistique entre bruit d'échantillonnage et covariances asymptotiques.

Les covariances échantillonnées sont sans biais : $\mathbb{E}[\tilde{B}_{ij}] = \mathbb{E}[B_{ij}]$

Il existe une relation linéaire dépendant de la taille de l'ensemble :

$$\begin{pmatrix} \mathbb{E} \begin{bmatrix} B_{ij} & B_{kl} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} B_{ik} & B_{jl} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} B_{il} & B_{jk} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} \Xi \\ ijkl \end{bmatrix} \end{pmatrix} = \mathbf{A}(N) \begin{pmatrix} \mathbb{E} \begin{bmatrix} \tilde{B}_{ij} & \tilde{B}_{kl} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} \tilde{B}_{ik} & \tilde{B}_{jl} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} \tilde{B}_{il} & \tilde{B}_{jk} \end{bmatrix} \\ \mathbb{E} \begin{bmatrix} \tilde{\Xi} \\ ijkl \end{bmatrix} \end{pmatrix}$$

où $\tilde{\Xi}$ et Ξ sont les moments d'ordre 4 échantillonné et asymptotique. $\mathbf{A}(N)$ est connue !

Retour à la localization

Seule hypothèse : indépendance faible du bruit d'échantillonnage et des covariances asymptotiques :

$$\mathbb{E}\left[\left(\tilde{B}_{ij} - B_{ij}\right)B_{kl}\right] = 0 \quad \Leftrightarrow \quad \mathbb{E}\left[\tilde{B}_{ij}B_{kl}\right] = \mathbb{E}\left[B_{ij}B_{kl}\right]$$

Retour à la localisation

Seule hypothèse : indépendance faible du bruit d'échantillonnage et des covariances asymptotiques :

$$\mathbb{E} \left[\left(\tilde{B}_{ij} - B_{ij} \right) B_{kl} \right] = 0 \quad \Leftrightarrow \quad \mathbb{E} \left[\tilde{B}_{ij} B_{kl} \right] = \mathbb{E} \left[B_{ij} B_{kl} \right]$$

Résultat final pour la localisation :

$$L_{ij} = \frac{\mathbb{E} \left[\tilde{B}_{ij} B_{ij} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]} = \frac{(N-1)^2}{N(N-3)} + \frac{N-1}{N(N-2)(N-3)} \frac{\mathbb{E} \left[\tilde{B}_{ii} \tilde{B}_{jj} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]} - \frac{N}{(N-2)(N-3)} \frac{\mathbb{E} \left[\tilde{\Xi}_{ijj} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]}$$

Retour à la localisation

Seule hypothèse : indépendance faible du bruit d'échantillonnage et des covariances asymptotiques :

$$\mathbb{E} \left[\left(\tilde{B}_{ij} - B_{ij} \right) B_{kl} \right] = 0 \quad \Leftrightarrow \quad \mathbb{E} \left[\tilde{B}_{ij} B_{kl} \right] = \mathbb{E} \left[B_{ij} B_{kl} \right]$$

Résultat final pour la localisation :

$$L_{ij} = \frac{\mathbb{E} \left[\tilde{B}_{ij} B_{ij} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]} = \frac{(N-1)^2}{N(N-3)} + \frac{N-1}{N(N-2)(N-3)} \frac{\mathbb{E} \left[\tilde{B}_{ii} \tilde{B}_{jj} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]} - \frac{N}{(N-2)(N-3)} \frac{\mathbb{E} \left[\tilde{\Xi}_{ijj} \right]}{\mathbb{E} \left[\tilde{B}_{ij}^2 \right]}$$

Localisation optimale exprimée sans quantités asymptotiques !

Plan

Introduction

Optimisation objective de la localisation

Implémentation des diagnostics

Optimisation conjointe de la localisation et de l'hybridation

Conclusions

Hypothèse d'ergodicité

En pratique, on ne dispose que d'un seul ensemble à la fois, donc d'une seule réalisation de $\tilde{\mathbf{B}}$ et $\tilde{\mathbf{E}}$.

Les espérances $\mathbb{E}[\cdot]$ doivent être estimées par une hypothèse d'ergodicité :

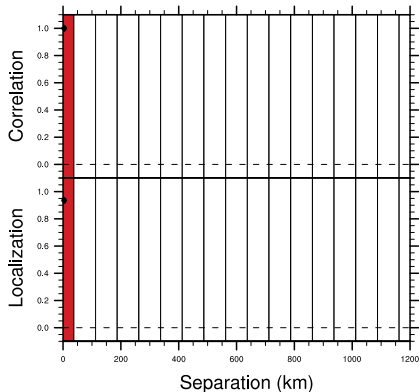
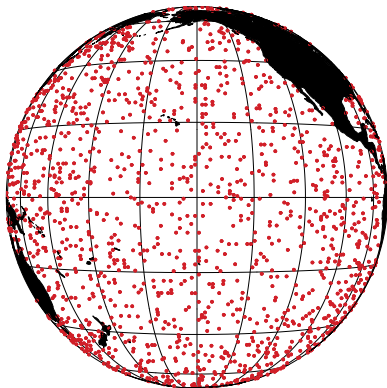
- Diagnostics horizontaux : moyenne angulaire et spatiale,
- Diagnostics verticaux : moyenne spatiale horizontale.

Dans les deux cas :

- Moyennes globales sur tout le domaine.
- Moyennes masquées sur un sous-domaine.
- Moyennes locales robustes (moyennes tronquées) utilisant les plus proches voisins.

Echantillonnage global

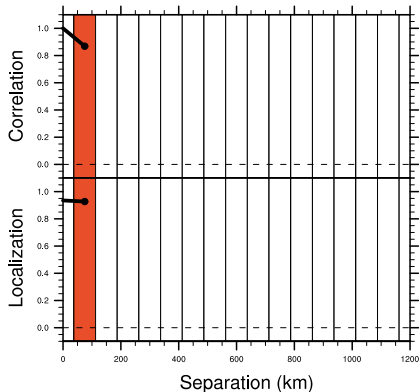
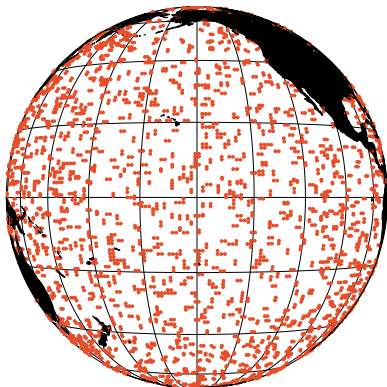
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

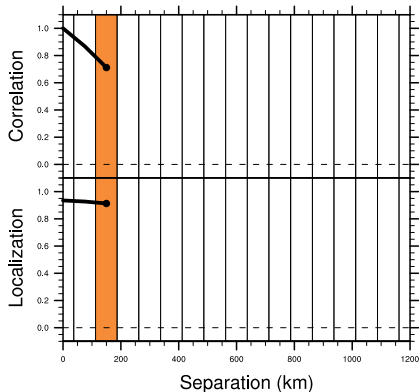
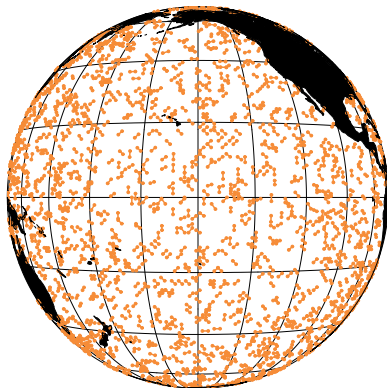
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

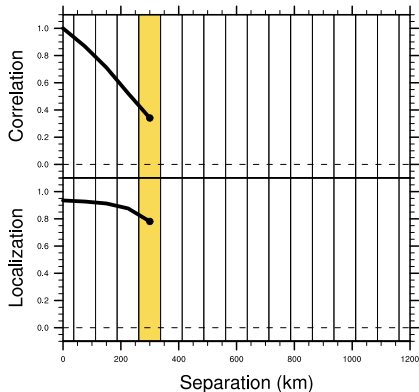
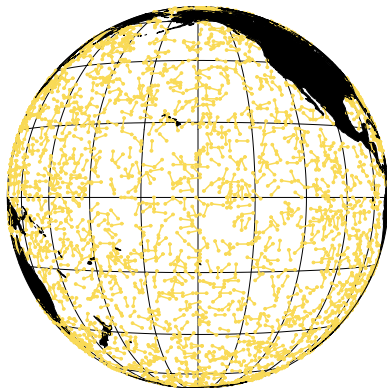
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

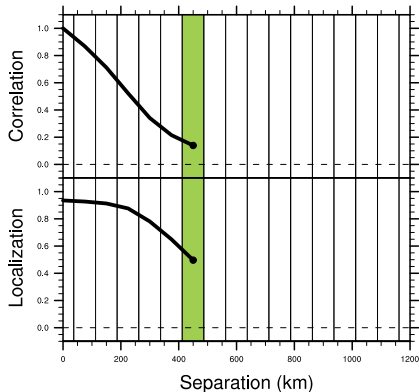
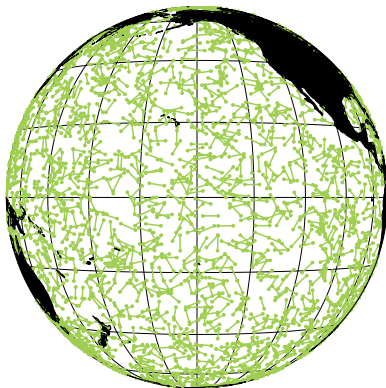
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

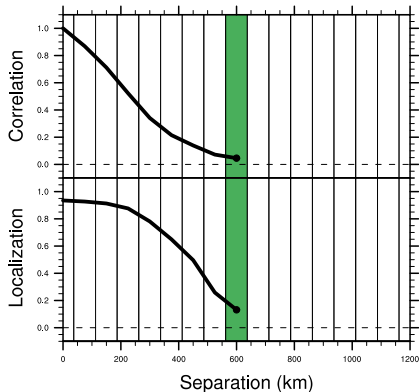
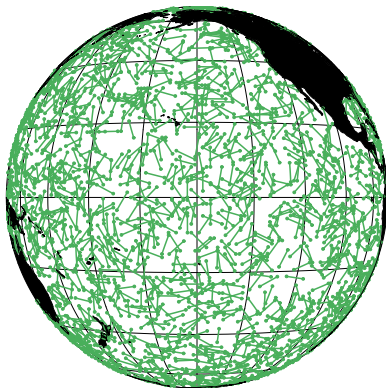
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

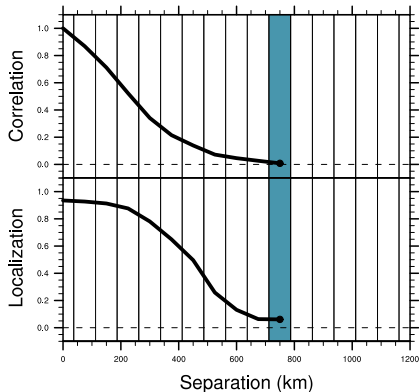
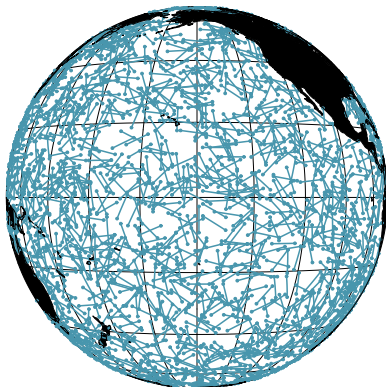
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

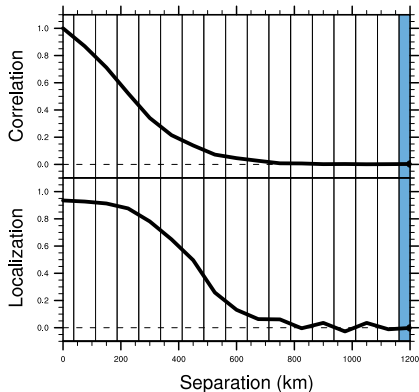
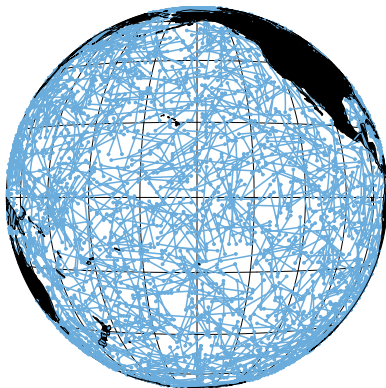
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage global

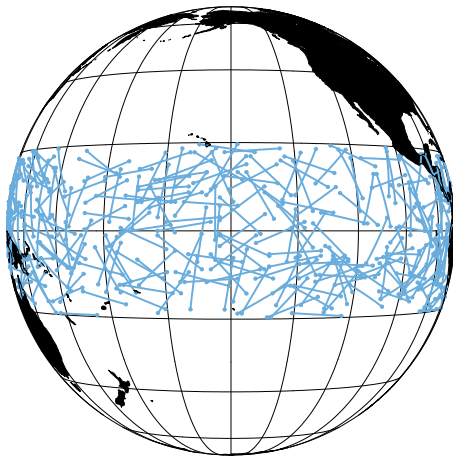
Exemple pour le modèle NEMO :



Corrélation et localisation (25 membres, température à 5 m de profondeur)
 Données de l'ensemble : Anthony Weaver (CERFACS).

Echantillonnage masqué

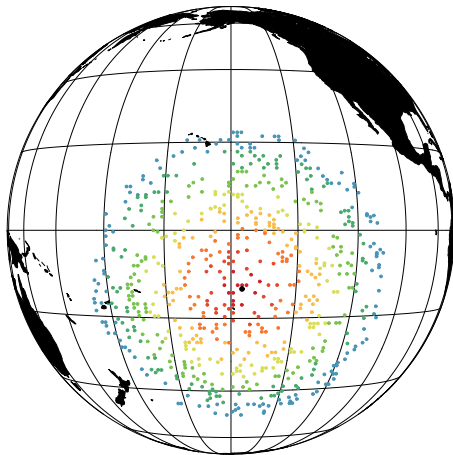
Exemple pour le modèle NEMO :



Domaine masqué pour ne garder que la bande inter-tropicale.

Echantillonnage local

Exemple pour le modèle NEMO :

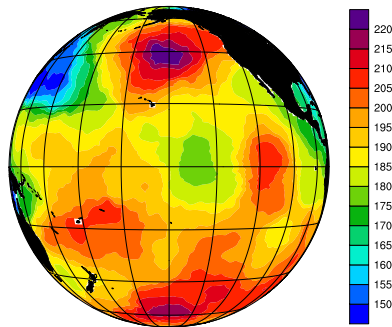


Plus proches voisins pour des distances croissantes.

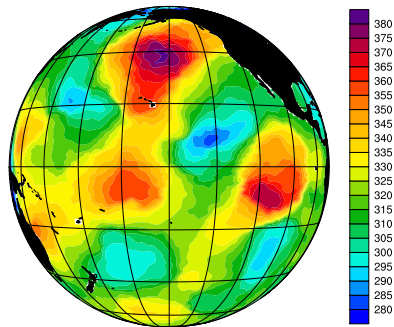
Diagnostics locaux

Exemple pour le modèle NEMO :

Correlation (km)



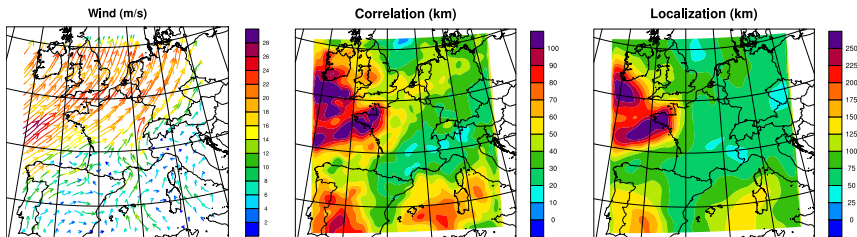
Localisation (km)



Portées de corrélation et de localisation
 (25 membres, température à 5 m de profondeur)

Diagnostics locaux

Exemple pour le modèle AROME :



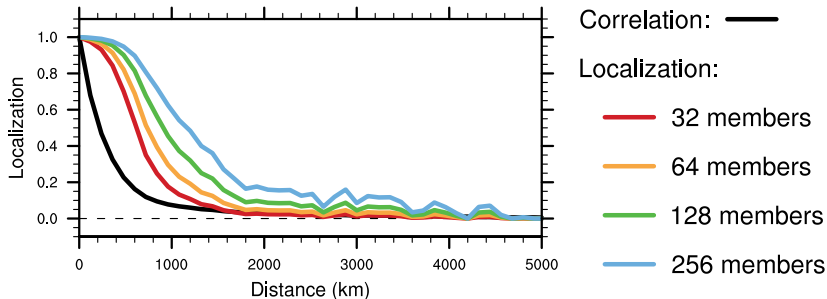
Vent et portées de corrélation et de localisation (température à 950 hPa)

Données de l'ensemble : Yann Michel (Météo-France).

Dépendance au flux météorologique des diagnostics :
 intrusion de longues portées depuis le modèle coupleur.

Impact de la taille d'ensemble

Exemple pour le modèle GEM :



Corrélation et localisation (température à 875 hPa)

Données de l'ensemble : Jean-François Caron (Environnement Canada).

Dépendance cohérente à la taille de l'ensemble :
 localisation s'élargissant avec la taille de l'ensemble.

Plan

Introduction

Optimisation objective de la localisation

Implémentation des diagnostics

Optimisation conjointe de la localisation et de l'hybridation

Conclusions

Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

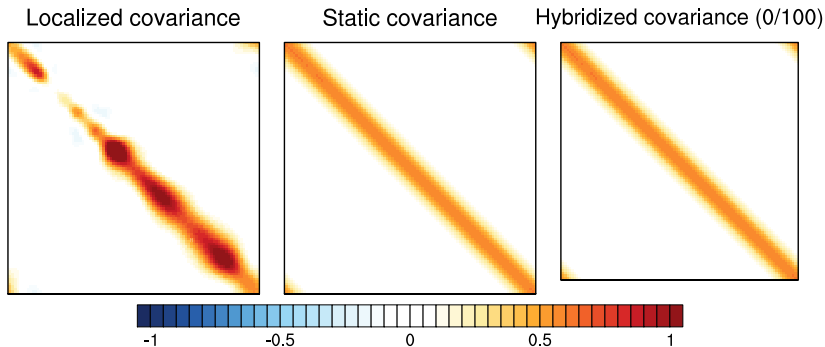
$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Pour un \mathbf{B}^c homogène :



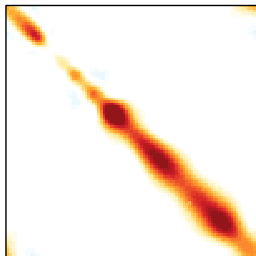
Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

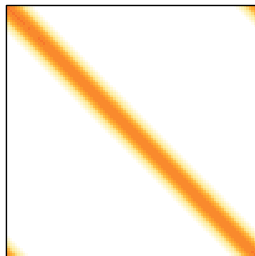
$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Pour un \mathbf{B}^c homogène :

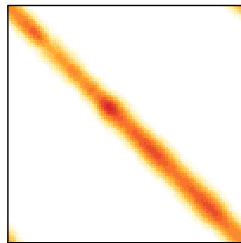
Localized covariance



Static covariance



Hybridized covariance (25/75)

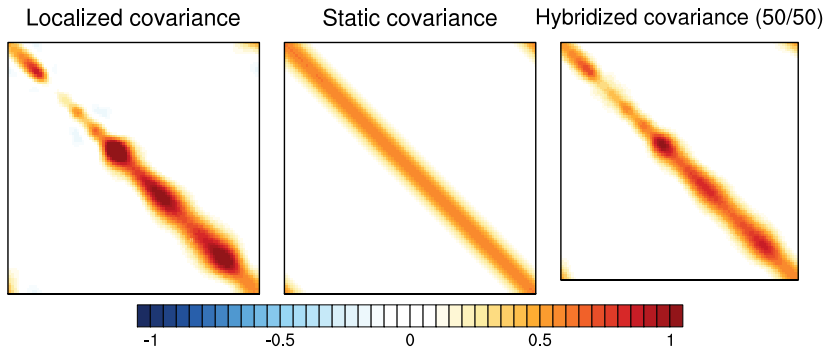


Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Pour un \mathbf{B}^c homogène :



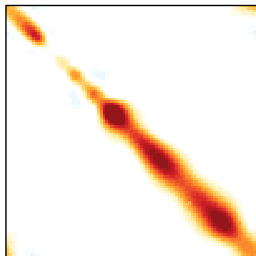
Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

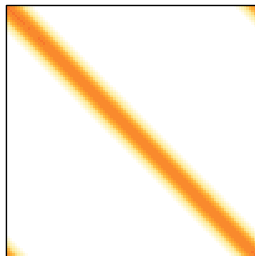
$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Pour un \mathbf{B}^c homogène :

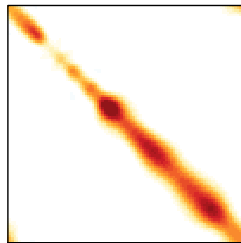
Localized covariance



Static covariance



Hybridized covariance (75/25)

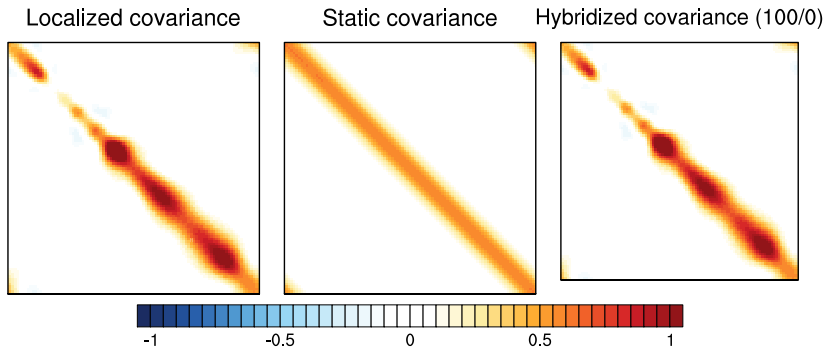


Covariances hybridées

Hybridation = combinaison linéaire de la matrice de covariance localisée $\hat{\mathbf{B}} = \mathbf{L} \circ \tilde{\mathbf{B}}$ et d'une matrice de covariance statique \mathbf{B}^c :

$$\hat{\mathbf{B}}^h = \beta^{e2} \hat{\mathbf{B}} + \beta^{c2} \mathbf{B}^c \quad \Leftrightarrow \quad \hat{B}_{ij}^h = \beta^{e2} \hat{B}_{ij} + \beta^{c2} B_{ij}^c$$

Pour un \mathbf{B}^c homogène :



Formalisme du filtrage linéaire

Localisation par L + hybridation avec B^c :

$$\hat{B}^h = \beta^{e2} L \circ \tilde{B} + \beta^{c2} B^c$$

Formalisme du filtrage linéaire

Localisation par L + hybridation avec B^c :

$$\hat{B}^h = \beta^{e2} L \circ \tilde{B} + \beta^{c2} B^c$$

$$\hat{B}_{ij}^h = \underbrace{\beta^{e2} L_{ij}}_{\text{Gain } L_{ij}^h} \tilde{B}_{ij} + \underbrace{\beta^{c2} B_{ij}^c}_{\text{Offset}}$$

Formalisme du filtrage linéaire

Localisation par L + hybridation avec B^c :

$$\hat{B}^h = \beta^{e2} L \circ \tilde{B} + \beta^{c2} B^c$$

$$\hat{B}_{ij}^h = \underbrace{\beta^{e2} L_{ij}}_{\text{Gain } L_{ij}^h} \tilde{B}_{ij} + \underbrace{\beta^{c2} B_{ij}^c}_{\text{Offset}}$$

Localisation + hybridation = filtrage linéaire de \tilde{B}
 L^h et β^{c2} doivent être optimisées conjointement

Formalisme du filtrage linéaire

Localisation par \mathbf{L} + hybridation avec \mathbf{B}^c :

$$\widehat{\mathbf{B}}^h = \beta^{e2} \mathbf{L} \circ \widetilde{\mathbf{B}} + \beta^{c2} \mathbf{B}^c$$

$$\widehat{B}_{ij}^h = \underbrace{\beta^{e2} L_{ij}}_{\text{Gain } L_{ij}^h} \widetilde{B}_{ij} + \underbrace{\beta^{c2} B_{ij}^c}_{\text{Offset}}$$

Localisation + hybridation = filtrage linéaire de $\widetilde{\mathbf{B}}$
 \mathbf{L}^h et β^{c2} doivent être optimisées conjointement

La localisation et l'hybridation optimales doivent minimiser :

$$e^h = \mathbb{E} \left[\|\mathbf{L}^h \circ \widetilde{\mathbf{B}} + \beta^{c2} \mathbf{B}^c - \mathbf{B}\|_F^2 \right]$$

Résultat de l'optimisation

A l'optimalité:

$$\frac{\partial e^h}{\partial L_{ij}^h} = 0 \quad \text{et} \quad \frac{\partial e^h}{\partial \beta^{c2}} = 0$$

Résultat de l'optimisation

A l'optimalité:

$$\frac{\partial e^h}{\partial L_{ij}^h} = 0 \quad \text{et} \quad \frac{\partial e^h}{\partial \beta^{c2}} = 0$$

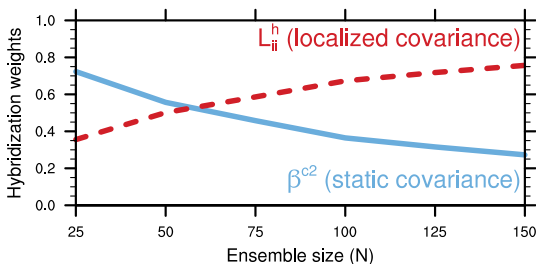
Avec des hypothèses identiques :

$$\beta^{c2} = \frac{\sum_{ij} (1 - L_{ij}) \mathbb{E}[\tilde{B}_{ij}] B_{ij}^c}{\sum_{ij} \frac{\text{Var}[\tilde{B}_{ij}]}{\mathbb{E}[\tilde{B}_{ij}^2]} B_{ij}^{c2}} \quad \text{and} \quad L_{ij}^h = L_{ij} - \frac{\mathbb{E}[\tilde{B}_{ij}]}{\mathbb{E}[\tilde{B}_{ij}^2]} \beta^{c2} B_{ij}^c$$

où L_{ij} est la localisation optimisée seule (sans hybridation).

Coefficients d'hybridation

Exemple pour le modèle ARPEGE :



Coefficients d'hybridation (température à 930 hPa)
 Données de l'ensemble : Vincent Chabot (Météo-France).

Comme attendu :

- $\beta^{e2} = L_{ii}^h$ augmente avec la taille de l'ensemble,
- β^{c2} diminue avec la taille de l'ensemble.

Plan

Introduction

Optimisation objective de la localisation

Implémentation des diagnostics

Optimisation conjointe de la localisation et de l'hybridation

Conclusions

Conclusions

1. La localisation peut être optimisée avec une méthode :

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,
 - efficace pour des **systèmes de grande dimension**,

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,
 - efficace pour des **systèmes de grande dimension**,
 - s'attaquant au bruit d'échantillonnage uniquement.

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,
 - efficace pour des **systèmes de grande dimension**,
 - s'attaquant au bruit d'échantillonnage uniquement.

2. La localisation et l'hybridation peuvent être optimisées conjointement (filtrage linéaire de covariances).

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,
 - efficace pour des **systèmes de grande dimension**,
 - s'attaquant au bruit d'échantillonnage uniquement.
2. La localisation et l'hybridation peuvent être optimisées conjointement (filtrage linéaire de covariances).
3. Ces méthodes sont générales, indépendantes du modèle et de sa géométrie.

Conclusions

1. La localisation peut être optimisée avec une méthode :
 - basée sur les propriétés **de l'ensemble seulement**,
 - efficace pour des **systemes de grande dimension**,
 - s'attaquant au bruit d'échantillonnage uniquement.
2. La localisation et l'hybridation peuvent être optimisées conjointement (filtrage linéaire de covariances).
3. Ces méthodes sont générales, indépendantes du modèle et de sa géométrie.
4. Référence : Ménétrier, B. and T. Auligné, 2015: Optimized Localization and Hybridization to Filter Ensemble-Based Covariances, *Monthly Weather Review*, **143(10)**, 3931-3947

Code disponible

Code open-source de diagnostic de localisation / hybridation :

Code disponible

Code open-source de diagnostic de localisation / hybridation :

- interfaces pour 9 modèles déjà implémentées (AROME, ARPEGE, GEM, GEOS, GFS, IFS, MPAS, NEMO et WRF),

Code disponible

Code open-source de diagnostic de localisation / hybridation :

- interfaces pour 9 modèles déjà implémentées (AROME, ARPEGE, GEM, GEOS, GFS, IFS, MPAS, NEMO et WRF),
- ajout d'un nouveau modèle facile, quelle que soit sa géométrie,

Code disponible

Code open-source de diagnostic de localisation / hybridation :

- interfaces pour 9 modèles déjà implémentées (AROME, ARPEGE, GEM, GEOS, GFS, IFS, MPAS, NEMO et WRF),
- ajout d'un nouveau modèle facile, quelle que soit sa géométrie,
- documentation importante et support aux utilisateurs !

Code disponible

Code open-source de diagnostic de localisation / hybridation :

- interfaces pour 9 modèles déjà implémentées (AROME, ARPEGE, GEM, GEOS, GFS, IFS, MPAS, NEMO et WRF),
- ajout d'un nouveau modèle facile, quelle que soit sa géométrie,
- documentation importante et support aux utilisateurs !

Dépôt : opensource.cnrm-game-meteo.fr/projects/hybrid_diag

Contact : benjamin.menetrier@meteo.fr

Code disponible

Code open-source de diagnostic de localisation / hybridation :

- interfaces pour 9 modèles déjà implémentées (AROME, ARPEGE, GEM, GEOS, GFS, IFS, MPAS, NEMO et WRF),
- ajout d'un nouveau modèle facile, quelle que soit sa géométrie,
- documentation importante et support aux utilisateurs !

Dépôt : opensource.cnrm-game-meteo.fr/projects/hybrid_diag

Contact : benjamin.menetrier@meteo.fr

Merci pour votre attention.