BUMP, a generic tool for background error covariance modeling

Benjamin Menetrier, IRIT, Toulouse, France
benjamin.menetrier@irit.fr
Funded by the JCSDA

5th JEDI Academy - November 2020

Basic facts

- Background error covariance (aka B matrix) is a key aspect of variational DA systems.
- In the practical implementation, the B matrix itself is not required, only its effect on a state vector.
- B must be symmetric and positive, so we generally build its square-root U: $\mathrm{B} = \mathrm{U}\mathrm{U}^{\mathrm{T}}$.
- Usual covariance models are:
  - static B: a sequence of parametrized operators
  - ensemble-based B: a localized sample covariance matrix
  - hybrid B: a linear combination of previous models
- All covariance models require a smoother to spread the innovation information. In most implementations, smoothers are grid-specific (e.g. spectral transform, recursive filters).

BUMP library

- BUMP: "Background error on Unstructured Mesh Package"
- It is designed to work on any grid (gaussian, cubed-sphere, unstructured, limited-area, ocean).
- It can diagnose parameters for all the usual covariance models.
- It implements a generic smoother, NICAS ("Normalized Interpolated Convolution from an Adaptive Subgrid").
- The code is written in Fortran 90, around 25.000 lines.
- It is part of the SABER repository ("System-Agnostic Background Error Representation"), which will include other covariance modeling libraries in the future.
- It is fully interfaced with OOPS, for both diagnostic and application parts.

BUMP credo: subsampling

High-resolution grids are great for models dynamics, but very costly and not really useful for background error covariance modeling.

In general, the size of the background error structures that we can represent is significantly larger than the model grid cell size.

The credo of BUMP is:

You shall cleverly subsample the model grids.

You shall perform costly operations on the subsampled mesh.

You shall interpolate your final results on the model grid.

-The BUMP book

Outline

Outline

Static B basics

- The static B is a robust and well-conditioned model, based on a sequence of parametrized operators.
- The parameters can be defined using ensemble data over a long period, giving a climatological estimate.
- They can also be estimated over a shorter sliding window, and updated at every cycle.
- The most common static B model is:

$$\mathrm{B}^s = \mathrm{K}_p \boldsymbol{\Sigma} \mathrm{C} \boldsymbol{\Sigma}^{\mathrm{T}} \mathrm{K}_p^{\mathrm{T}}$$

  where
  - C is a correlation matrix
  - $\boldsymbol{\Sigma}$ is a diagonal matrix of standard deviations
  - K_p is a multivariate balance operator

Static B with BUMP

BUMP can be used for all these operators:

- The correlation length-scales of C are estimated globally or locally, and used to set up the NICAS smoother, which is exactly normalized ($C_{ii} = 1$).
- The standard-deviations in Σ are estimated locally and potentially filtered, to remove the sampling noise (objective filtering of Ménétrier *et al.*, 2015a,b).
- A vertical balance operator computing regressions between variables can be diagnosed and applied with BUMP (still under development), as part of a more complex K_p.

In the OOPS framework, the static B components can come from BUMP or from your own model, and be combined as you wish.

Outline

Sample covariance

An ensemble of N forecasts $\{\mathrm{x}_p^b\}$ is used to estimate the sample covariance matrix $\widetilde{\mathrm{B}}$:

$$\widetilde{\mathrm{B}} = \frac{1}{N-1} \sum_{p=1}^{N} \delta\mathrm{x}_p^b \delta\mathrm{x}_p^{b\mathrm{T}}$$

where $\delta\mathrm{x}_p^b$ is the p^{th} ensemble perturbation:

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

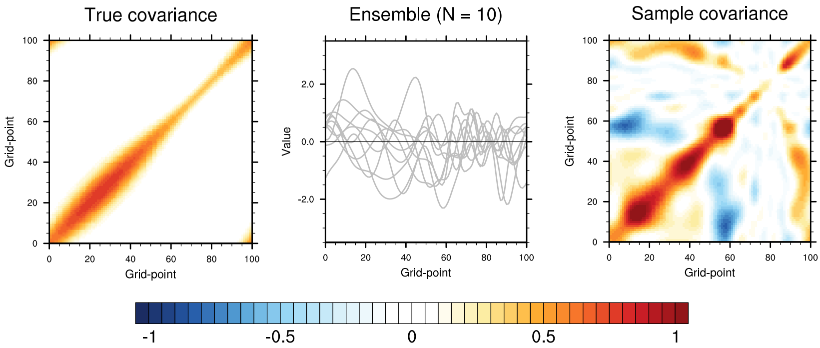
Asymptotic sample covariance: $\mathrm{B} = \lim\limits_{N \to \infty} \widetilde{\mathrm{B}}$

Since the ensemble size $N < \infty$, $\widetilde{\mathrm{B}}$ is affected by sampling noise:

$$\widetilde{\mathrm{B}}^e = \widetilde{\mathrm{B}} - \mathrm{B}$$

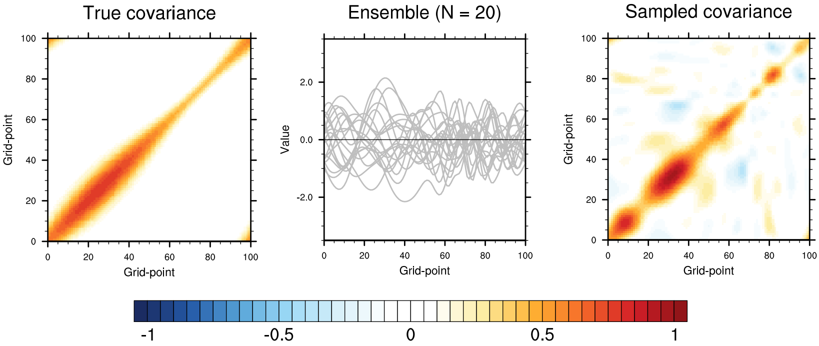
Sample covariance

Sampling noise strongly depends on the ensemble size:



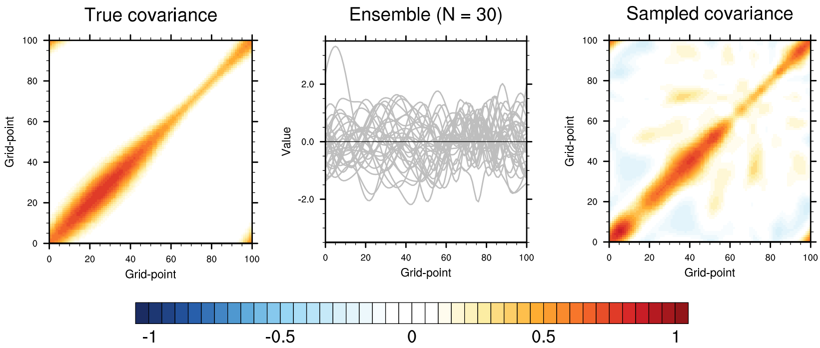
Sample covariance

Sampling noise strongly depends on the ensemble size:



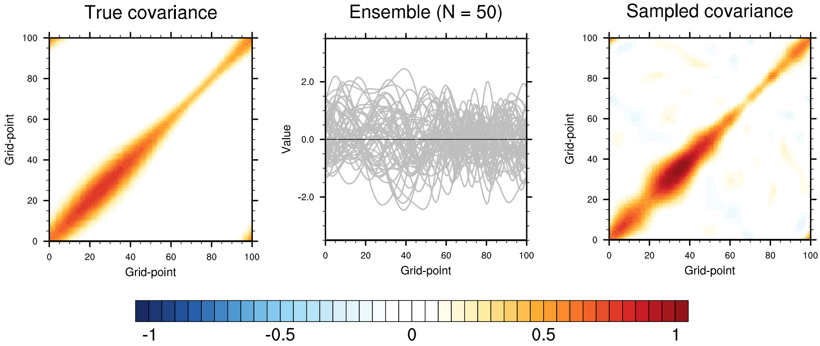
Sample covariance

Sampling noise strongly depends on the ensemble size:



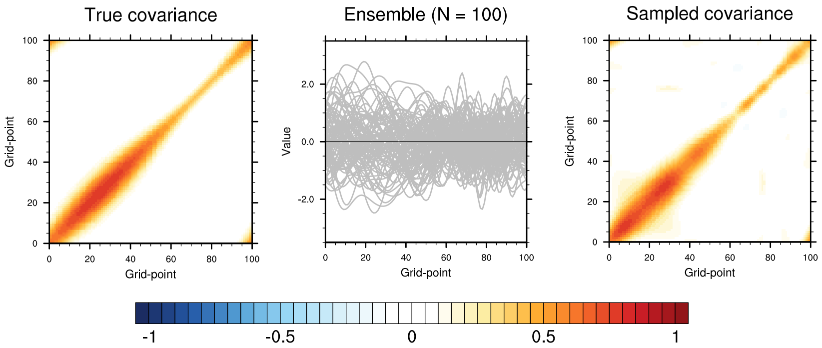
Sample covariance

Sampling noise strongly depends on the ensemble size:



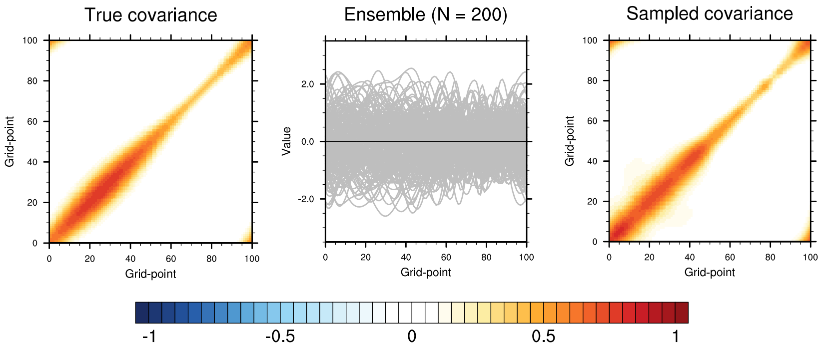
Sample covariance

Sampling noise strongly depends on the ensemble size:



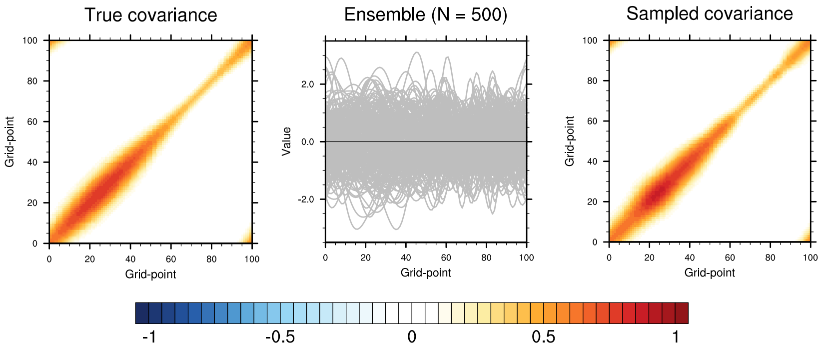
Sample covariance

Sampling noise strongly depends on the ensemble size:



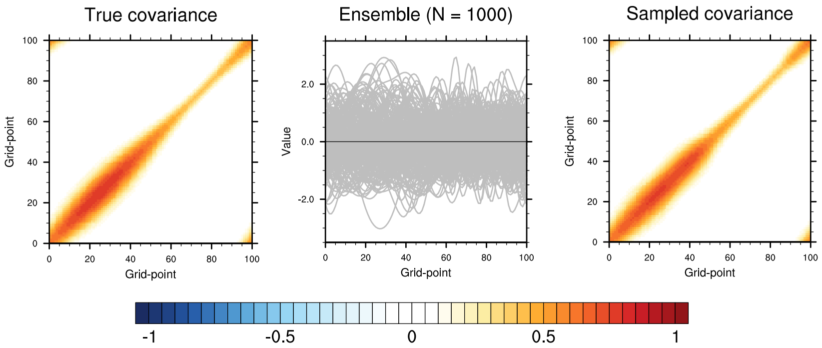
Sample covariance

Sampling noise strongly depends on the ensemble size:



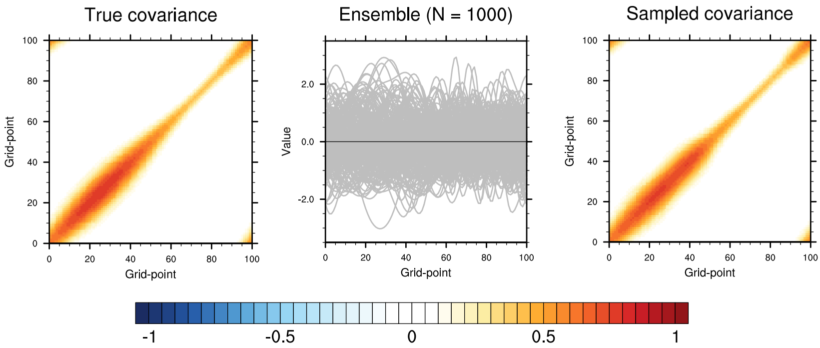
Sample covariance

Sampling noise strongly depends on the ensemble size:



Sample covariance

Sampling noise strongly depends on the ensemble size:



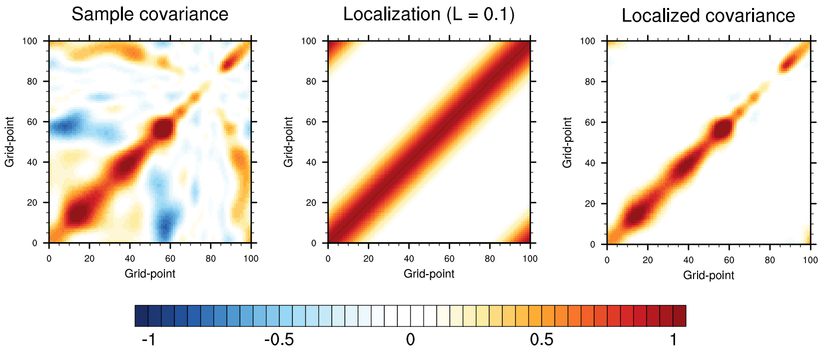
We don't have oil, but we have ideas!

Localized covariance

Sampling noise on $\widetilde{\mathrm{B}}$ can be damped via a Schur product (element-by-element) with a localization matrix L:

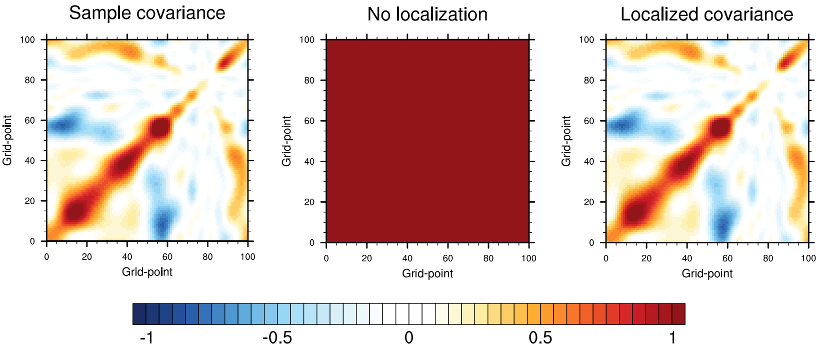
$$\widehat{\mathrm{B}} = \mathrm{L} \circ \widetilde{\mathrm{B}} \quad \Leftrightarrow \quad \widehat{B}_{ij} = L_{ij} \widetilde{B}_{ij}$$

In practice, L damps the long-distance correlations that are small and more affected by sampling noise (hence the "localization").



Localization: what is the optimal length-scale?

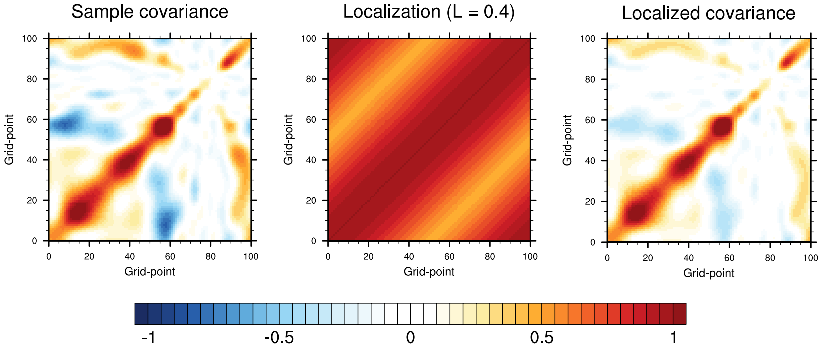
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



No impact

Localization: what is the optimal length-scale?

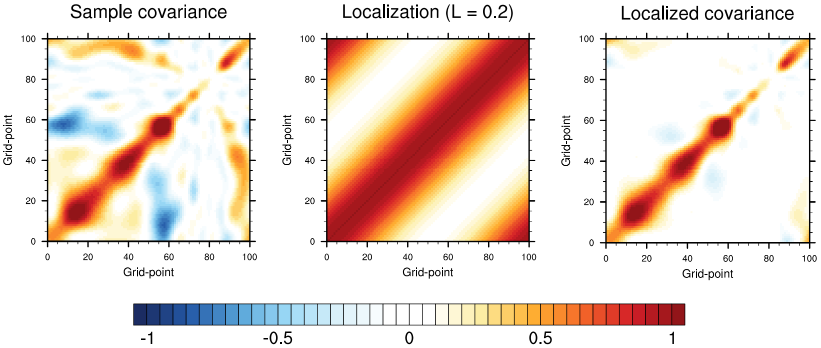
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



Start reducing the sampling noise...

Localization: what is the optimal length-scale?

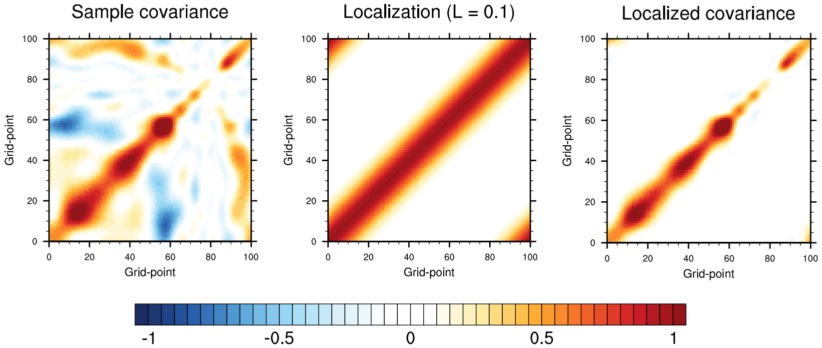
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



Less and less sampling noise...

Localization: what is the optimal length-scale?

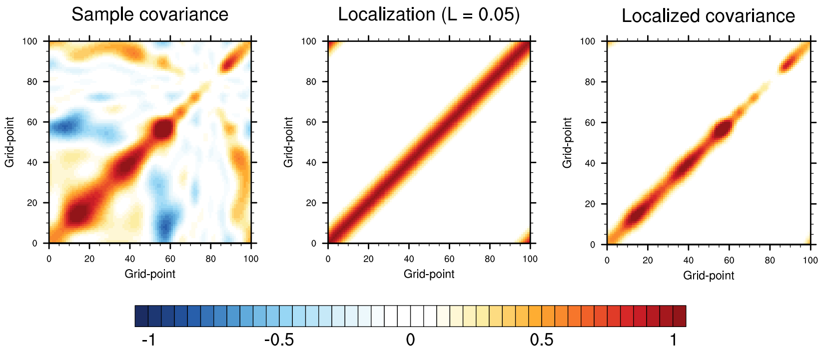
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



Good ! Almost no sampling noise anymore...

Localization: what is the optimal length-scale?

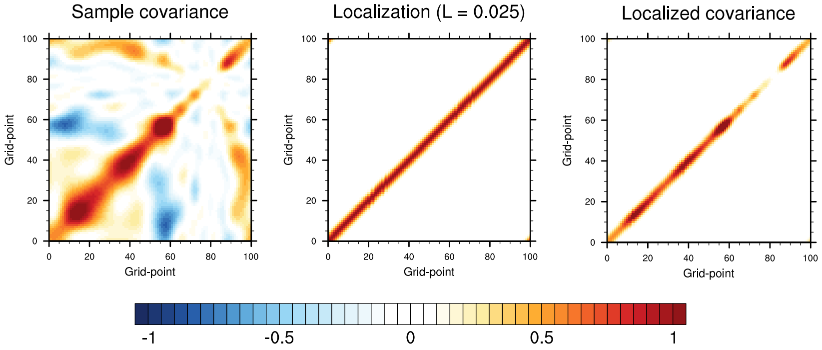
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



Well, we are loosing some signal now...

Localization: what is the optimal length-scale?

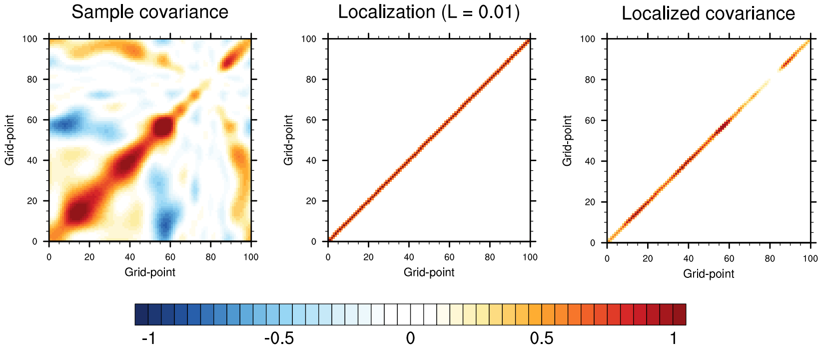
The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



Hey, stop loosing signal !

Localization: what is the optimal length-scale?

The localization length-scale is critical to remove the sampling noise while keeping the relevant covariance signal:



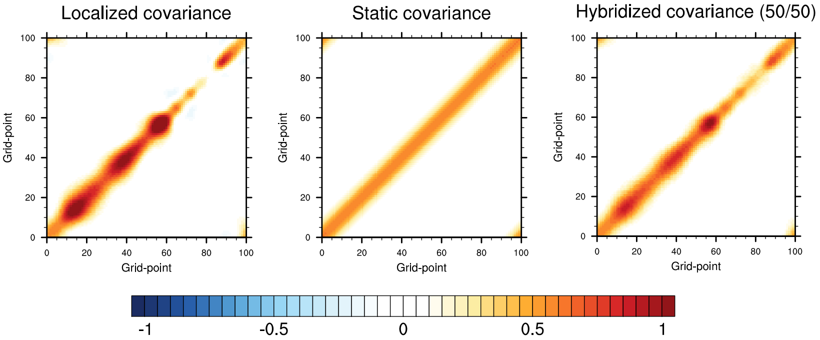
No more signal !

Localized and hybridized covariance

Deficiencies of the localized covariance matrix $\widehat{\mathrm{B}}$ can be corrected via a hybridization with a static covariance matrix:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\,\widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2}B^s_{ij}$$

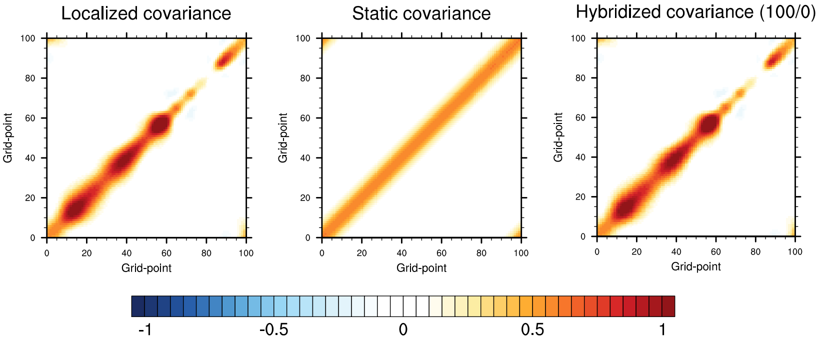
For a homogeneous B^s:



Hybridization: what are the optimal weights

Localization + hybridation:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\ \widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2}B^s_{ij}$$

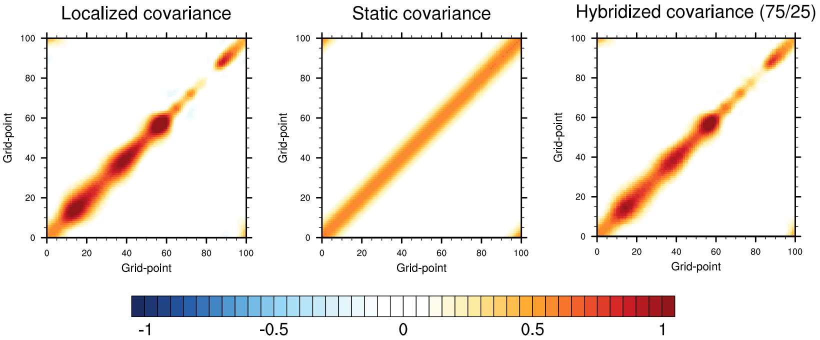


100% on the localized covariance matrix

Hybridization: what are the optimal weights

Localization + hybridation:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\ \widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2}B^s_{ij}$$

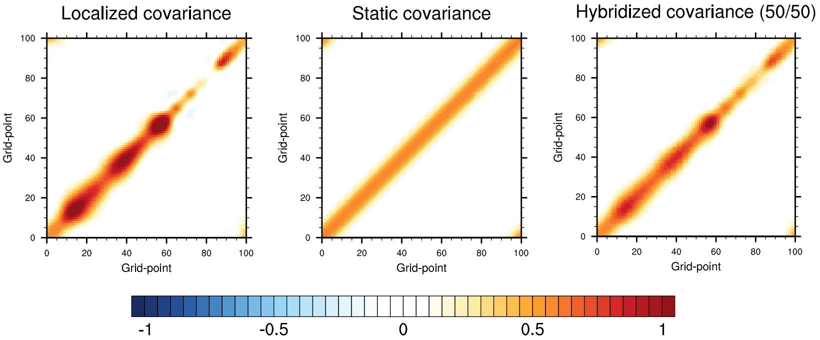


75% on the localized covariance matrix

Hybridization: what are the optimal weights

Localization + hybridation:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\ \widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2}B^s_{ij}$$

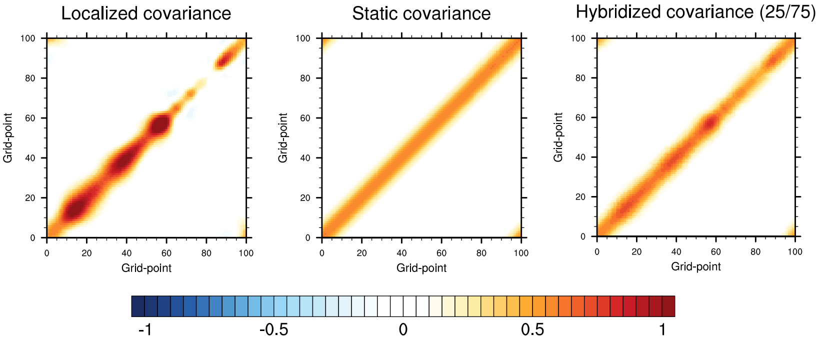


Equal weights on both covariance matrices

Hybridization: what are the optimal weights

Localization + hybridation:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\ \widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2}B^s_{ij}$$

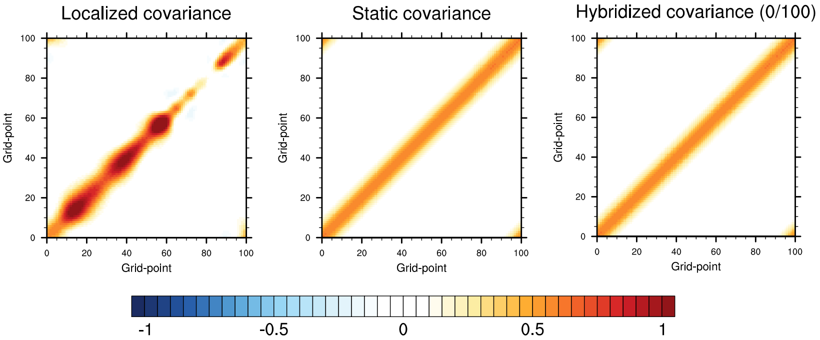


75% on the static covariance matrix

Hybridization: what are the optimal weights

Localization + hybridation:

$$\widehat{\mathrm{B}}^h = \beta^{e2}\ \widehat{\mathrm{B}} + \beta^{s2}\mathrm{B}^s \quad \Leftrightarrow \quad \widehat{B}^h_{ij} = \beta^{e2}\widehat{B}_{ij} + \beta^{s2} B^s_{ij}$$



100% on the static covariance matrix

How to optimize localization and hybridization?

Existing methods are empirical and costly (e.g. OSSE, brute-force optimization). We need a new method that:

- uses only ensemble members,
- is affordable for high-dimensional systems,
- can be generic enough to be run with all kinds of grids.

Principle :

$$\underbrace{\widehat{\mathrm{B}}^h}_{\text{Localized/hybridized covariance}} = \underbrace{\beta^{e2}}_{\text{Ensemble weight}} \underbrace{\mathrm{L}}_{\text{Localization matrix}} \circ \underbrace{\widetilde{\mathrm{B}}}_{\text{Sampled covariance}} + \underbrace{\beta^{s2}}_{\text{Static weight}} \underbrace{\mathrm{B}^s}_{\text{Static covariance}}$$

Localization + hybridization = linear filtering of $\widetilde{\mathrm{B}}$

How to optimize localization and hybridization?

Asymptotic sample covariance: $\mathrm{B} = \lim_{N \to \infty} \widetilde{\mathrm{B}}$

Residual noise (after localization/hybridization): $\widehat{\mathrm{B}}^h - \mathrm{B}$

Objectives:

- Express $\beta^{e2}\mathrm{L}$ and β^{s2} minimizing the error $\mathbb{E}\left[\|\widehat{\mathrm{B}}^h - \mathrm{B}\|^2\right]$.

  $\rightarrow$ Linear filtering theory.

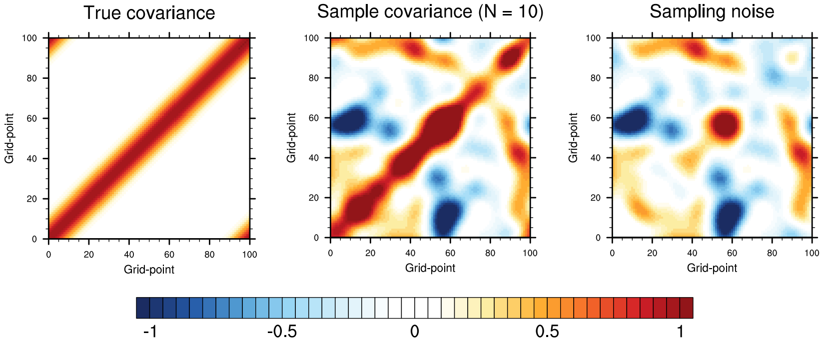
  Some statistics involve the asymptotic sample covariance.

- Express statistics on asymptotic quantities (unknown) with expected sample quantities (knowable).

  $\rightarrow$ Centered moments sampling theory (non-Gaussian case).

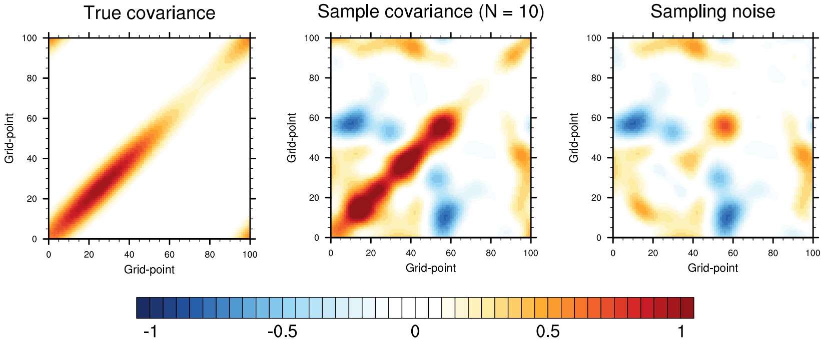
Sampling noise properties

Homogeneous variance / length-scale



Sampling noise properties

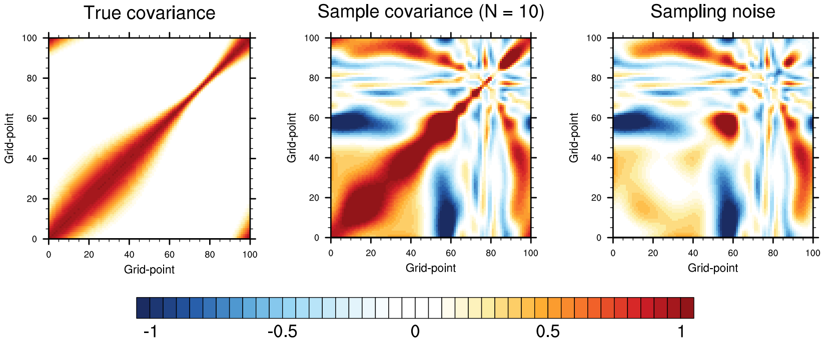
Heterogeneous variance / homogeneous length-scale



Sampling noise amplitude related to the asymptotic variance

Sampling noise properties

Homogeneous variance / heterogeneous length-scale



Sampling noise length-scale related to the asymptotic length-scale

How to optimize localization and hybridization?

Optimal localization alone (without hybridization):

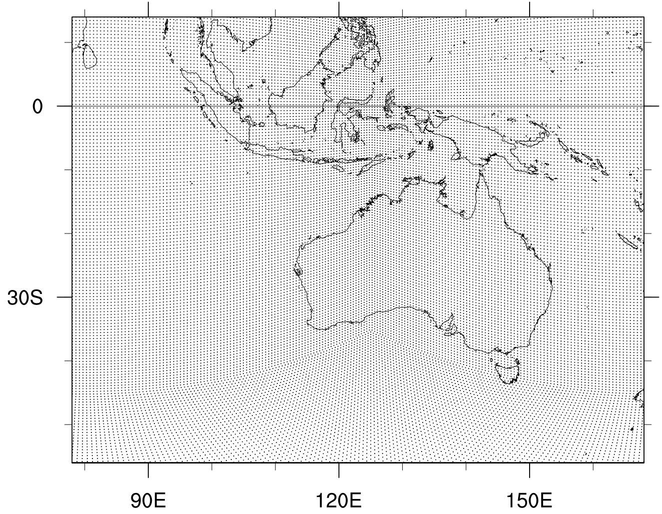
$$L'_{ij} = \frac{\mathbb{E}\left[B_{ij}^2\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]}$$

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

where $\widetilde{\Xi}$ is the sample fourth-order centered moment.

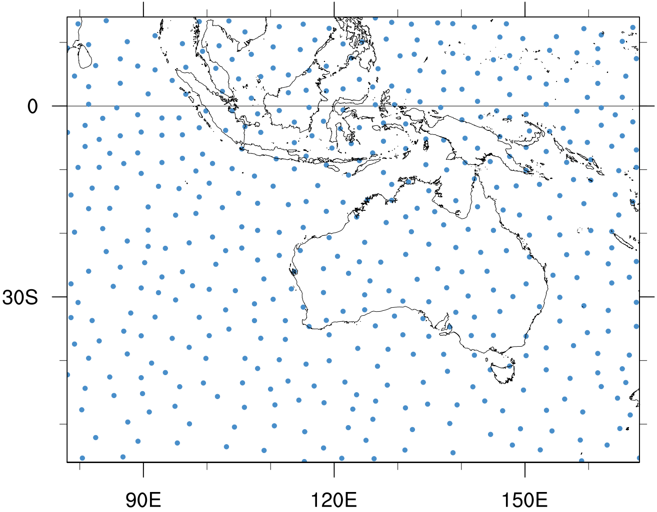
$$\beta^{s2} = \frac{\sum_{ij} \left(1 - L'_{ij}\right)\mathbb{E}\left[\widetilde{B}_{ij}\right] B_{ij}^s}{\sum_{ij} \frac{\mathrm{Var}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} B_{ij}^{s2}} \quad \text{and} \quad \beta^{e2} L_{ij} = L'_{ij} - \frac{\mathbb{E}\left[\widetilde{B}_{ij}\right]}{\mathbb{E}\left[\widetilde{B}_{ij}^2\right]} \beta^{s2} B_{ij}^s$$

Homogenous and isotropic sampling



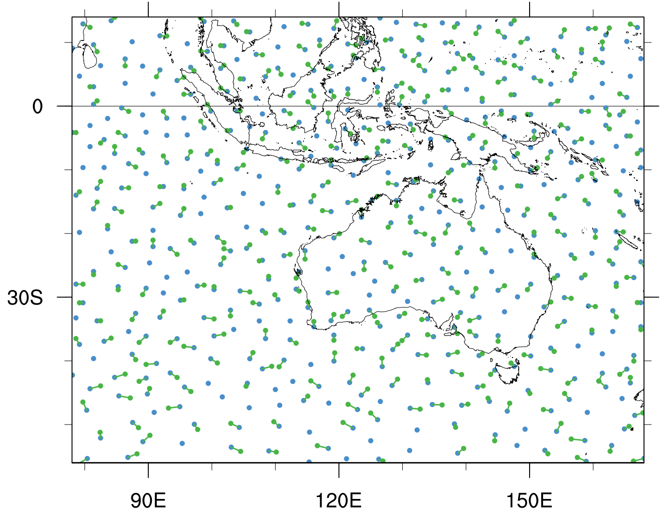
Full model grid (FV3 cubed-sphere grid - C180)

Homogenous and isotropic sampling



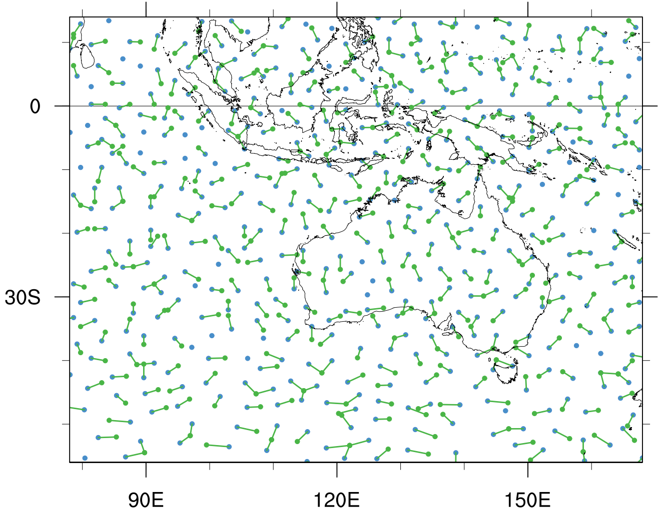
Homogenous subsampling for origin points

Homogenous and isotropic sampling



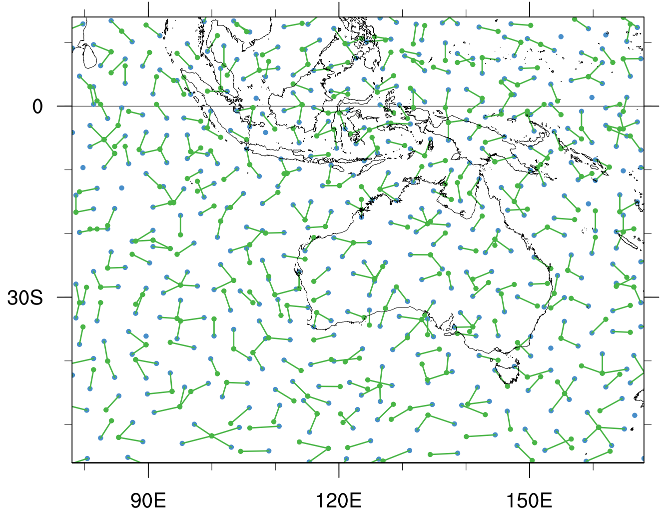
Isotropic subsampling for distant points (increasing distance class)

Homogenous and isotropic sampling



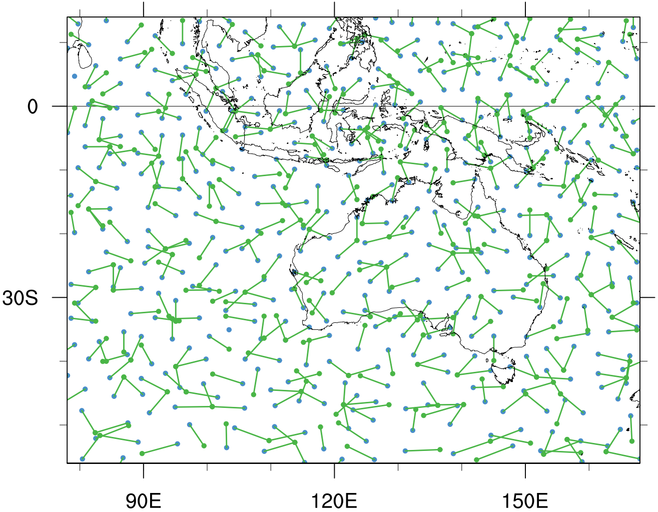
Isotropic subsampling for distant points (increasing distance class)

Homogenous and isotropic sampling



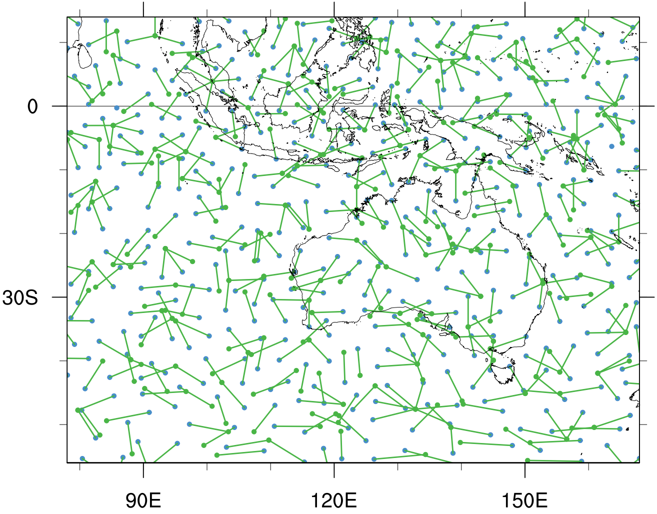
Isotropic subsampling for distant points (increasing distance class)

Homogenous and isotropic sampling



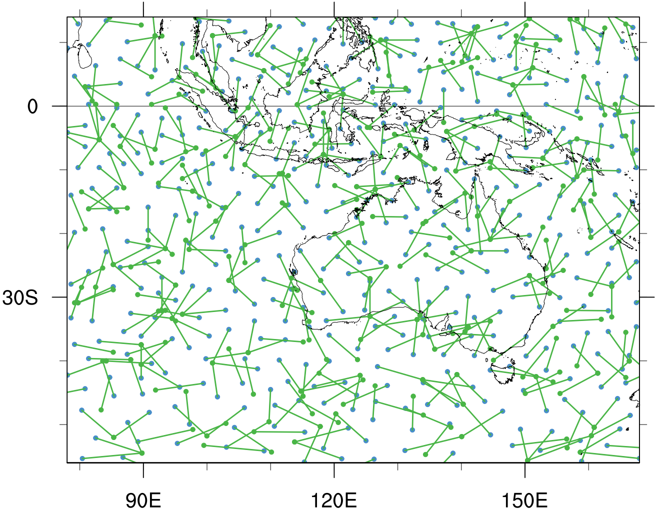
Isotropic subsampling for distant points (increasing distance class)

Homogenous and isotropic sampling



Isotropic subsampling for distant points (increasing distance class)

Homogenous and isotropic sampling



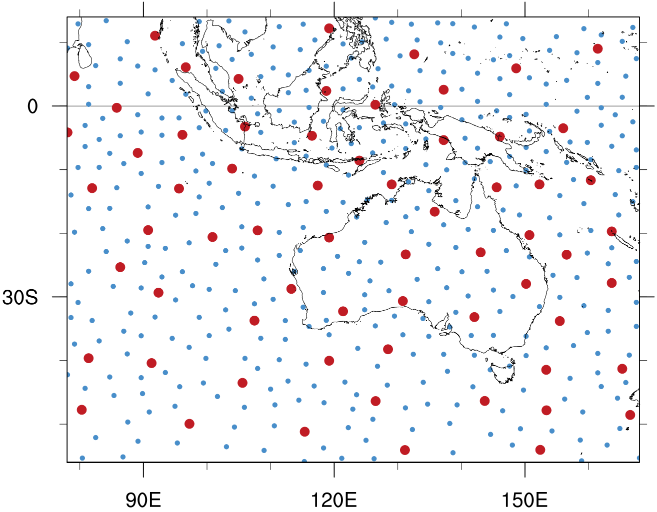
Isotropic subsampling for distant points (increasing distance class)

Ergodicity assumption

For each couple of points, BUMP uses the ensemble to estimate:

- the sample variances $\widetilde{B}_{ii}$ and $\widetilde{B}_{jj}$
- the sample covariance $\widetilde{B}_{ij}$
- the sample fourth-order centered moment $\widetilde{\Xi}_{ijij}$

Local averaging



Local averaging centers (red points)

Ergodicity assumption

For each couple of points, BUMP uses the ensemble to estimate:

- the sample variances $\widetilde{B}_{ii}$ and $\widetilde{B}_{jj}$
- the sample covariance $\widetilde{B}_{ij}$
- the sample fourth-order centered moment $\widetilde{\Xi}_{ijij}$

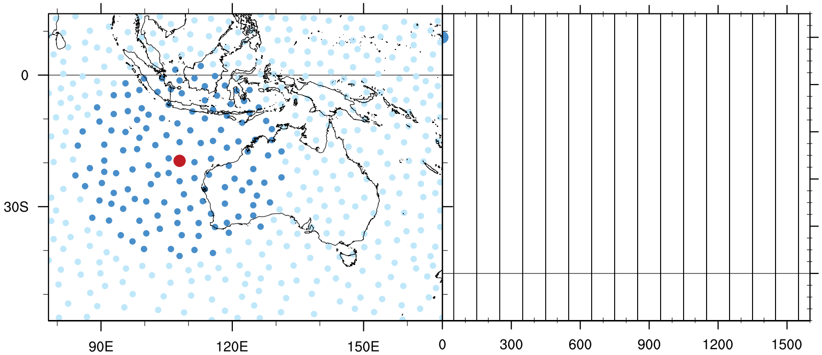
With the spatial and angular ergodicity assumption, these values are averaged locally for each distance class, to get estimations of the following expectations:

$$\mathbb{E}\left[\widetilde{B}_{ii}\widetilde{B}_{jj}\right] \quad , \quad \mathbb{E}\left[\widetilde{B}_{ij}^2\right] \quad \text{and} \quad \mathbb{E}\left[\widetilde{\Xi}_{ijij}\right]$$

These quantities are useful to compute the localization function and hybridization weights with the previous formulas.

Local averaging

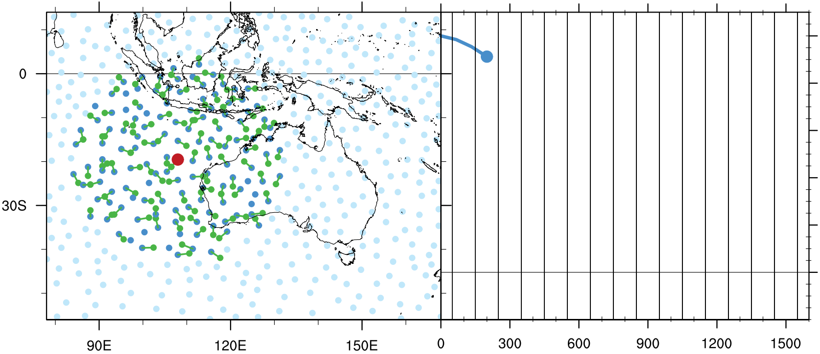
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

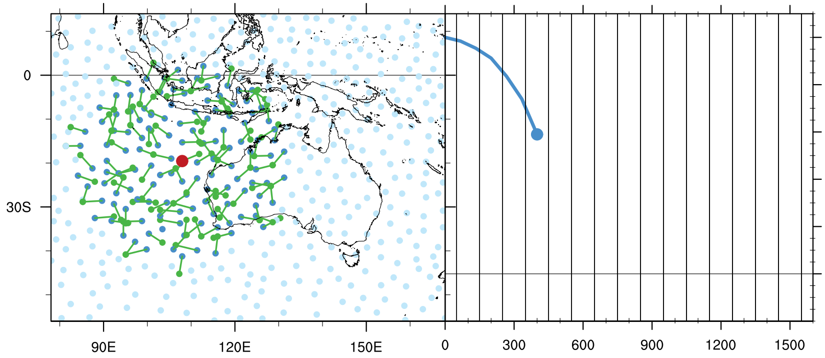
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

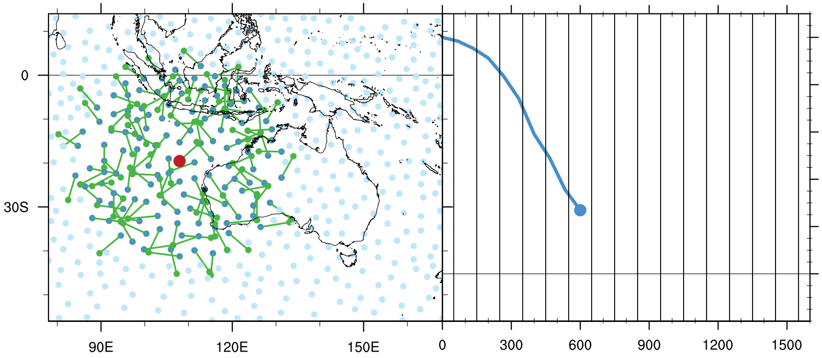
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

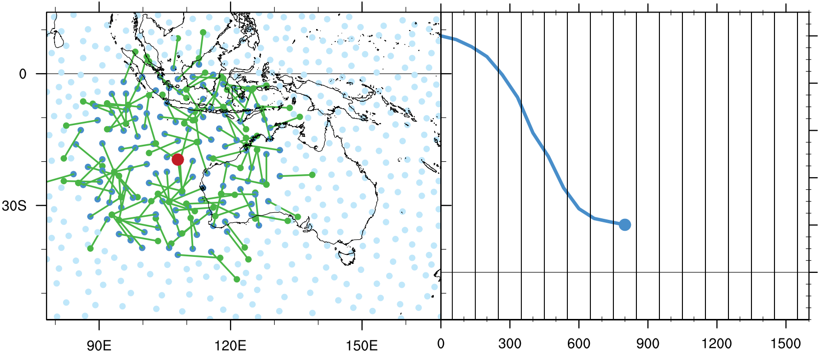
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

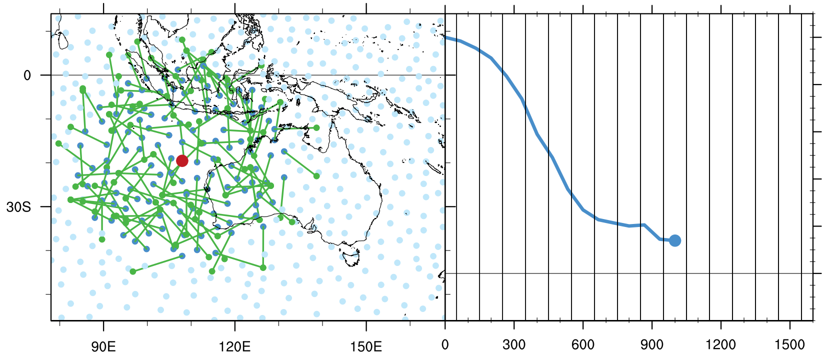
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

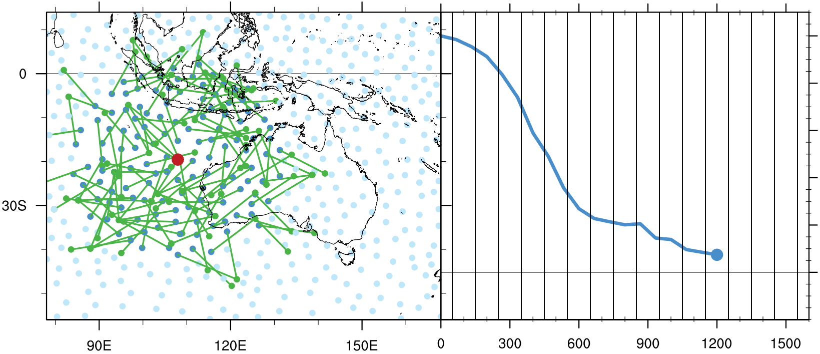
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

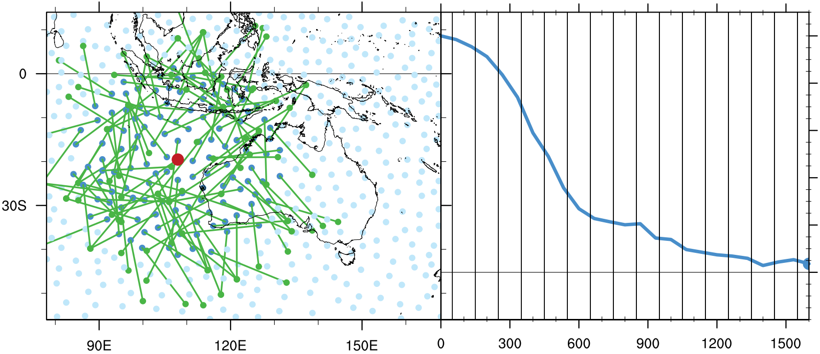
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Local averaging

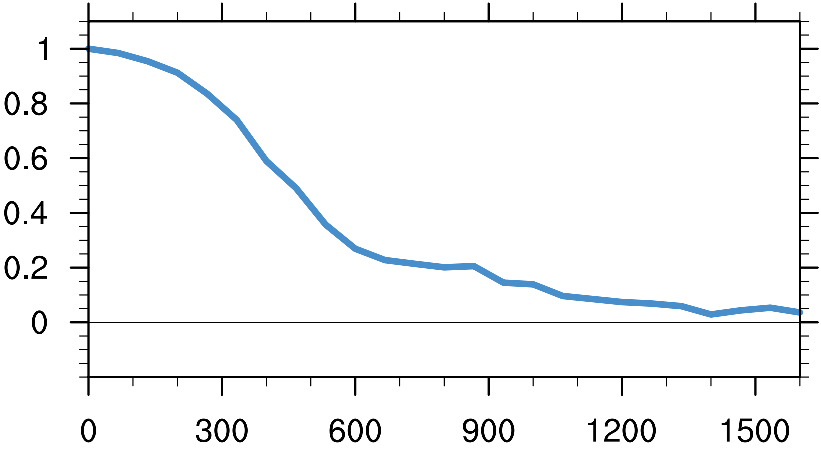
For a given local averaging center:



Averaged couples for a given local averaging center (left)
Diagnosed localization as a function of distance in km (right)

Localization function fit

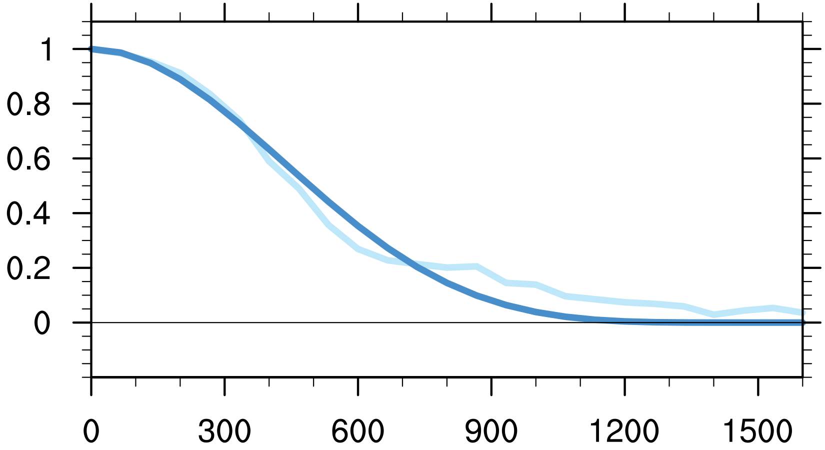
For a given local averaging center:



Raw localization as a function of distance (km)

Localization function fit

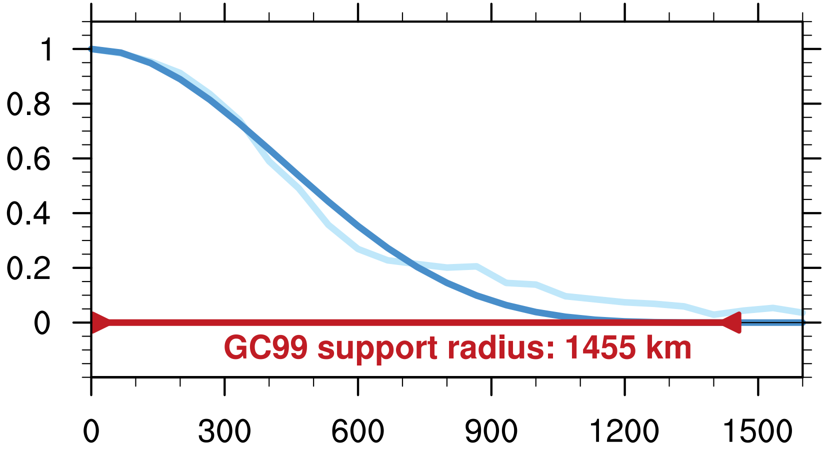
For a given local averaging center:



Fit with a Gaspari-Cohn (1999) function

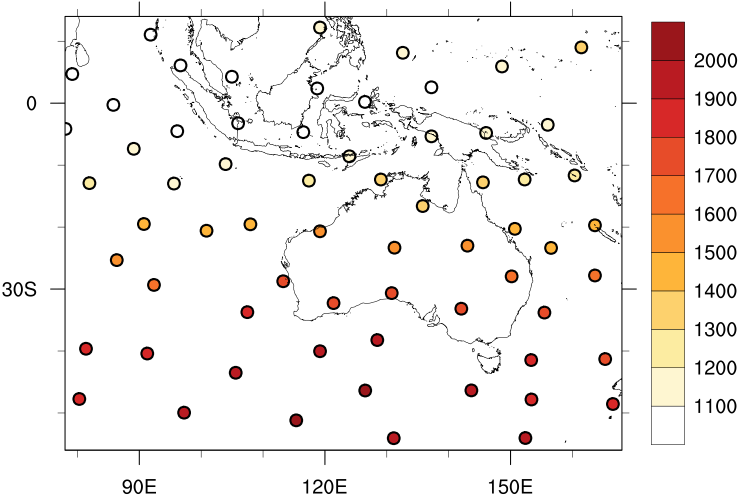
Localization function fit

For a given local averaging center:



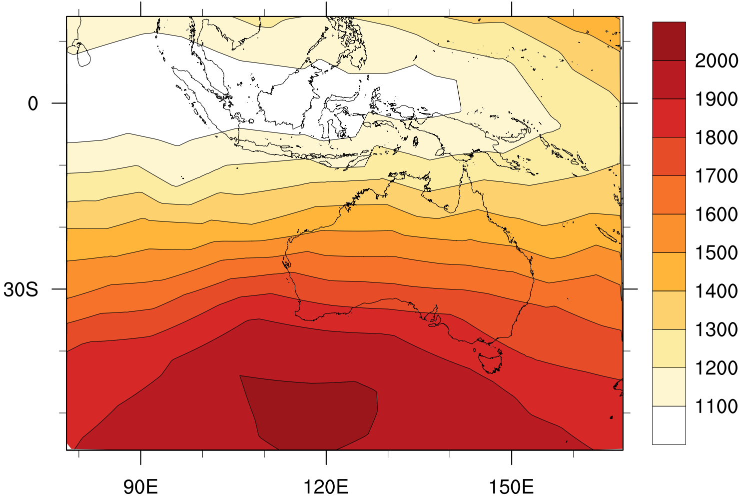
Localization length-scale (support radius actually, km)

Support radius interpolation



Localization support radius (km) at local averaging centers

Support radius interpolation

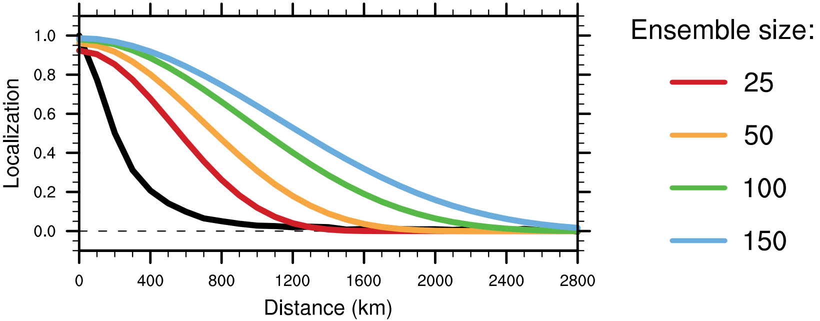


Localization support radius (km) interpolated on the model grid

HDIAG process

1. Define a homogeneous subsampling of the model grid (origin points).
2. For each origin point, find distant points for a series of distance classes.
3. For each local averaging center (a subset of origin points):
   - Average diagnostic values over a given radius.
   - Compute the localization and hybridization weights using the Ménétrier et al. (2015) formula.
   - Fit the raw localization function with a parametrized function (e.g. Gaspari-Cohn, 1999) to get the localization support radius.
4. Interpolate the localization support radius and hybridization weights over the model grid.

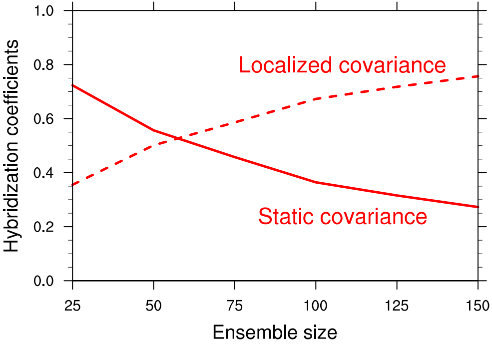
Ensemble size sensitivity



Correlation (black) et localization (colors) for various ensemble sizes

Localization length-scale increases as the ensemble size increases (less sampling noise to remove)

Ensemble size sensitivity



Ensemble and static weights as a function of the ensemble size

Less weight on the static covariance as the ensemble size increases (less deficiencies to correct).

Summary

- The sample covariance is affected by sampling noise.
- This sampling noise decreases if the ensemble size increases.
- Using a huge ensemble is too costly for operational applications.
- Localization and hybridization can be used to filter out the sampling noise.
- The asymptotic sample covariance is the filtering target.
- We combine the linear filtering theory and the centered moments sampling theory.
- This leads to practicable formulas for optimal localization and hybridization weights.
- We compute robust estimates using ergodicity assumptions.
- These statistics can be computed globally, or locally.
- Thus, we get global, or local estimates of the optimal localization and hybridization weights.

BUMP, a generic tool for background error covariance modeling

Benjamin Menetrier, IRIT, Toulouse, France
benjamin.menetrier@irit.fr

Funded by the JCSDA

5th JEDI Academy - November 2020