CERFACS

CENTRE EUROPÉEN DE RECHERCHE ET DE FORMATION AVANCÉE EN CALCUL SCIENTIFIQUE

Characterizing and modelling background error

Anthony Weaver
CERFACS, Toulouse

With thanks to Marcin Chrust, Oliver Guillet and Benjamin Ménétrier

(anonymous authors)

"An accurate specification of the statistics of background errors is a fundamental prerequiste of any effective data assimilation scheme..."

"The success of a data assimilation system relies heavily on the characterization of the background errors statistics..."

"Specification of the background error covariances is complex and its importance cannot be overestimated..."

⋮

$\mathbf{B}$ basics

$\mathbf{B}$ modelling

Using ensembles to define a flow-dependent $\mathbf{B}$

Concluding remarks

$\mathbf{B}$ basics

$\mathbf{B}$ modelling

Using ensembles to define a flow-dependent $\mathbf{B}$

Concluding remarks

The importance of $\mathbf{B}$ illustrated algebraically

- Consider the inner loop problem of incremental variational DA:

$$\min_{\delta\mathbf{x}} J[\delta\mathbf{x}] = \underbrace{\frac{1}{2}\,\delta\mathbf{x}^{\mathrm{T}}\,\mathbf{B}^{-1}\,\delta\mathbf{x}}_{J_{\mathrm{b}}} + \underbrace{\frac{1}{2}\,(\mathbf{G}\,\delta\mathbf{x} - \mathbf{d}\,)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{G}\,\delta\mathbf{x} - \mathbf{d}\,)}_{J_{\mathrm{o}}}$$

where $\mathbf{d} = \mathbf{y} - \mathcal{G}(\mathbf{x}^{\mathrm{b}})$ is the p-dimensional innovation vector.

- The exact minimizing solution is

$$\delta\mathbf{x}^{\mathrm{a}} = \mathbf{B}\mathbf{G}^{\mathrm{T}}\underbrace{\left(\mathbf{G}\mathbf{B}\mathbf{G}^{\mathrm{T}} + \mathbf{R}\right)^{-1}\mathbf{d}}_{\boldsymbol{\beta}}$$

- If $\mathbf{r}_i$ are the columns of $\mathbf{B}\mathbf{G}^{\mathrm{T}}$ then

$$\delta\mathbf{x}^{\mathrm{a}} = \mathbf{B}\mathbf{G}^{\mathrm{T}}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{r}_1 & \dots & \mathbf{r}_i & \dots & \mathbf{r}_p \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_p \end{pmatrix} = \sum_{i=1}^{p} \beta_i \mathbf{r}_i .$$

- The solution space is spanned by the columns of $\mathbf{B}\mathbf{G}^{\mathrm{T}} \approx E[\,\boldsymbol{\epsilon}^{\mathrm{b}}(\mathbf{G}\boldsymbol{\epsilon}^{\mathrm{b}})^{\mathrm{T}}\,]$.

What are the issues when specifying $\mathbf{B}$ in Earth System DA?

- $\mathbf{B}$ is a huge matrix ($O(10^{14}) - O(10^{18})$ elements) that is impossible to store and manipulate explicitly.
- There is not enough reliable information to specify all its elements anyway.
- $\mathbf{B}$ is inhomogeneous, anisotropic, multivariate and flow-dependent, which complicates its specification considerably.
- Ensemble DA is a practical and effective way to obtain flow-dependent information about $\mathbf{B}$.
- Simplifying assumptions and methods are essential to reduce sampling error in ensemble $\mathbf{B}$ estimates and to allow computationally efficient implementations of $\mathbf{B}$ in DA algorithms.

What are the issues when specifying $\mathbf{B}$ in Earth System DA?

- Practical methods to construct $\mathbf{B}$ must preserve its symmetric and positive (semi-)definite (SPD) attributes.
- This can be ensured numerically by constructing $\mathbf{B}$ as a product $\mathbf{U}\mathbf{U}^{\mathrm{T}}$ where $\mathbf{U}$ is generally a rectangular matrix (Parrish & Derber 1992).
- In variational DA, $\mathbf{U}$ and $\mathbf{U}^{\mathrm{T}}$ are specified as operators, through a control variable transform and the adjoint of this transform (Lorenc 2003a):

$$\delta\mathbf{x} = \mathbf{U}\mathbf{v} \quad \text{and} \quad \mathbf{v}^{*} = \mathbf{U}^{\mathrm{T}}\delta\mathbf{x}^{*}$$

- $\mathbf{U}$ is used for preconditioning and randomization applications.
- $\mathbf{B}$ plays a fundamental role in preconditioning variational DA minimization algorithms (Gürol *et al.* 2014).
- Most $\mathbf{B}$-preconditioned minimization algorithms do not require (thankfully!) specification of $\mathbf{B}^{-1}$.

- Consider $\mathbf{B}$ for a coupled system (e.g., atmosphere (a) and ocean (o)):

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\mathrm{aa}} & \mathbf{B}_{\mathrm{ao}} \\ \mathbf{B}_{\mathrm{ao}}^{\mathrm{T}} & \mathbf{B}_{\mathrm{oo}} \end{pmatrix}$$

- Specifying the within-component covariances ($\mathbf{B}_{\mathrm{aa}}$, $\mathbf{B}_{\mathrm{oo}}$) is hard.
- Specifying the cross-component covariances ($\mathbf{B}_{\mathrm{ao}}$) is even harder!
- These are the covariances that would allow an observation from one component to influence the analysis of the other component.
- How to deal with different scales, different cross-component balance relationships, and different grids?
- A coupled model ensemble can give you information about $\mathbf{B}_{\mathrm{ao}}$, but how do you localize the covariances across the interface?

Representing 4D covariances implicitly with 4D-Var

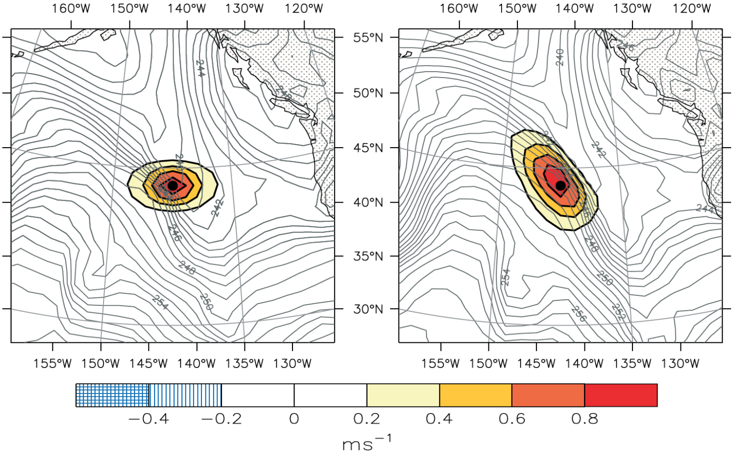
- In 4D-Var, $\mathbf{B}$ is implicitly propagated within the assimilation window by the linearized model operator $\mathbf{M}$ and its adjoint $\mathbf{M}^{\mathrm{T}}$.
- The 4D background error covariance matrix, on the interval t_0 to t_n, is implicitly given by (Lorenc 2003a):

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\,\mathbf{B}\underline{\mathbf{M}}^{\mathrm{T}}$$

where

$$\underline{\mathbf{M}} = \begin{pmatrix} \mathbf{I} \\ \mathbf{M}(t_0, t_1) \\ \vdots \\ \mathbf{M}(t_0, t_n) \end{pmatrix}$$

Zonal wind increment produced from a single zonal wind observation at the start (left panel) and at the end (right panel) of a 6hr 4D-Var window.



(From Clayton *et al.* 2013)

Representing 4D covariances implicitly with 4D-Var

- In 4D-Var, $\mathbf{B}$ is implicitly propagated within the assimilation window by the linearized model operator $\mathbf{M}$ and its adjoint $\mathbf{M}^{\mathrm{T}}$.
- The 4D background error covariance matrix, on the interval t_0 to t_n, is implicitly given by (Lorenc 2003a):

$$\underline{\mathbf{P}} = \underline{\mathbf{M}}\,\mathbf{B}\underline{\mathbf{M}}^{\mathrm{T}}$$

where

$$\underline{\mathbf{M}} = \begin{pmatrix} \mathbf{I} \\ \mathbf{M}(t_0, t_1) \\ \vdots \\ \mathbf{M}(t_0, t_n) \end{pmatrix}$$

- If $\mathbf{M}$ is a linearized coupled model then the 4D covariances will be strongly coupled even if $\mathbf{B}$ has no cross-component covariance ($\mathbf{B}_{\mathrm{ao}} = \mathbf{0}$).
- E.g., if $n = 1$ and $\mathbf{M} = \mathbf{M}(t_0, t_1)$ then

$$\underline{\mathbf{P}} = \begin{pmatrix} \mathbf{B} & \mathbf{B}\mathbf{M}^{\mathrm{T}} \\ \mathbf{M}\mathbf{B} & \mathbf{M}\mathbf{B}\mathbf{M}^{\mathrm{T}} \end{pmatrix}$$

$\mathbf{B}$ basics

$\mathbf{B}$ modelling

Using ensembles to define a flow-dependent $\mathbf{B}$

Concluding remarks

- A standard technique to account for multivariate covariances in variational DA is to transform the model state variables $(\mathbf{x})$ into a new set of (analysis) variables $(\mathbf{w})$ whose cross-covariances are much weaker than those of the original state variables (Derber & Bouttier 1999).
- The remaining cross-covariances are usually neglected so that $\mathbf{B}$ becomes block-diagonal (strictly univariate) with respect to $\mathbf{w}$.
- The **balance operator** (K_{bal}) is the inverse transformation from analysis variables back to state variables.
- It is constructed such that when linearized it forms a lower triangular matrix $(\mathbf{K}_{\mathrm{bal}})$:

$$\mathbf{B} = \mathbf{K}_{\mathrm{bal}} \underbrace{\mathbf{\Sigma}\,\mathbf{C}\,\mathbf{\Sigma}}_{\mathbf{B}_{(\mathbf{w})}} \mathbf{K}_{\mathrm{bal}}^{\mathrm{T}}$$

Example of a nonlinear balance operator for ocean DA

$$\left.\begin{array}{lclcl} T &=& T & & \\ S &=& K_S(T) &+& S_{\mathrm{U}} \\ \eta &=& K_\eta(T,S) &+& \eta_{\mathrm{U}} \\ u &=& K_u(T,S,\eta) &+& u_{\mathrm{U}} \\ v &=& K_v(T,S,\eta) &+& v_{\mathrm{U}} \end{array}\right\} \quad \mathbf{x} = K_{\mathrm{bal}}(\mathbf{w})$$

- Water masses (temperature, salinity) are approximately conserved (K_S):

$$S \approx S(T) \qquad (\textit{nonlinear T-S relation})$$

- Sea-surface height is related to density variations below the surface (K_η):

$$\eta \approx -\int_{z_{\mathrm{ref}}}^{0} (\rho/\rho_0)\,\mathrm{d}z \qquad (\textit{dynamic height})$$

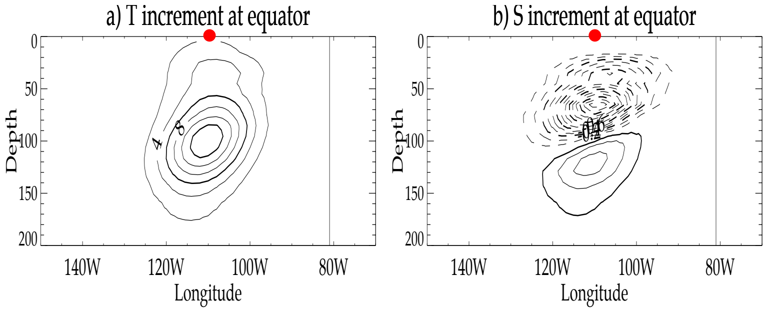
$$\rho = \rho(T,S) \qquad (\textit{nonlinear equation of state})$$

- Ocean currents are related to horizontal pressure gradients (K_u, K_v):

$$\mathbf{u} \approx \frac{1}{\rho_0 f}\,\hat{\mathbf{k}} \times \nabla p \qquad (\textit{geostrophic balance})$$

$$p(z) = -\int_{z'=z}^{0} \rho\, g \mathrm{d}z' + \rho_0 g \eta \qquad (\textit{hydrostatic balance})$$

3D-Var DA of a single +ve SSH innovation (●) at the equator



(From Weaver *et al.* 2005)

- To increase SSH, heat is added at the level of the thermocline, and salt is removed/added above/below the salinity maximum to approximately preserve water mass.
- The balance operator produces increments that are multivariate, anisotropic and flow dependent.

Specifying the correlation block-matrix operators in $\mathbf{C}$

- Applying the correlation operator $\mathbf{C}$ is the most computationally demanding component of the $\mathbf{B}$ model.
- We need to be able to perform efficient matrix-vector products with the block-matrix components of $\mathbf{C}$.
- Standard techniques in atmospheric variational DA:
  - Spectral/wavelet transform (Fisher 2003)
  - Recursive filter (Purser *et al.* 2003)
- **Issues**: regular grid required; complex boundaries difficult to handle; scalability
- Other (grid-point) techniques:
  - Diffusion operator (Weaver & Courtier 2001)
  - Explicit convolution (Gaspari & Cohn 1999; B. Ménétrier, unpublished material)
- **Issues**: cost; memory

Specifying the correlation block-matrix operators in $\mathbf{C}$

- Applying the correlation operator $\mathbf{C}$ is the most computationally demanding component of the $\mathbf{B}$ model.
- We need to be able to perform efficient matrix-vector products with the block-matrix components of $\mathbf{C}$.
- Standard techniques in atmospheric variational DA:
  - Spectral/wavelet transform (Fisher 2003)
  - Recursive filter (Purser *et al.* 2003)
- **Issues**: regular grid required; complex boundaries difficult to handle; scalability
- Other (grid-point) techniques.
  - **Diffusion operator** (Weaver & Courtier 2001)
  - Explicit convolution (Gaspari & Cohn 1999; B. Ménétrier, unpublished material)
- **Issues**: cost; memory

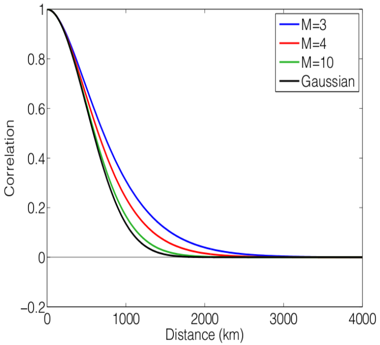
- Define $\mathbf{C}\,\psi_0$, for some ψ_0, to be the solution of the elliptic equation

$$(1 - \nabla \cdot \boldsymbol{\kappa}\nabla)^M \boldsymbol{\psi}_M = \boldsymbol{\psi}_0 \quad (1)$$

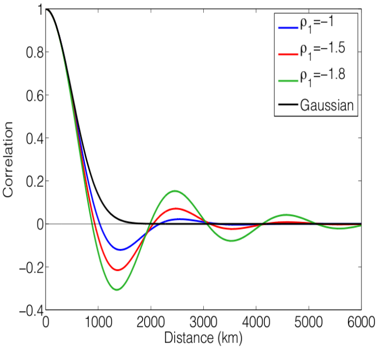
  where M is a positive integer and $\boldsymbol{\kappa}$ is a scale tensor.
- Solutions to Eq. (1) on $\mathbb{R}^d$ or $\mathbb{S}^2$, subject to appropriate boundary conditions, are covariance operators.
- The associated covariance functions are closely related to those from the **Matérn class** (Guttorp and Gneiting 2006).
- Eq. (1) can be interpreted as an **implicitly formulated diffusion operator** acting over M pseudo-time steps of unit length, and with $\boldsymbol{\kappa}$ the diffusion tensor (Weaver & Mirouze 2013).
- Covariance operators (stochastic PDEs) based on Eq. (1) have been studied independently in fields outside DA:
  - Spatial statistics (Lindgren *et al.* 2010)
  - Seismic inversion (Bui-Thanh *et al.* 2013)
  - Uncertainty Quantification (Gmeiner *et al.* 2017)
- More general PDEs can be used to represent a wider class of covariance functions (e.g., oscillatory).

Correlation functions $c(r)$ with constant $\boldsymbol{\kappa} = L^2 \boldsymbol{I}$.

$$\left(1 - L^2\nabla^2\right)^M c(r) = \gamma\,\delta(r)$$



$$\left(1 - \rho_1 L^2\nabla^2 + L^4\nabla^4\right)^2 c(r) = \gamma\,\delta(r)$$



(From Weaver & Mirouze 2013)

- In factored form, the correlation matrix is

$$\mathbf{C} = \mathbf{N}\,\mathbf{L}^{1/2}\mathbf{W}^{-1}\left(\mathbf{L}^{1/2}\right)^{\mathrm{T}}\mathbf{N}$$

  - $\mathbf{W}$ contains grid-dependent weights.
  - $\mathbf{L}^{1/2} \longleftrightarrow \left(1 - \nabla \cdot \boldsymbol{\kappa}\nabla\right)^{-M/2}$ (self-adjoint w.r.t. $\mathbf{W}$ inner product).
  - $\mathbf{N}$ is a diagonal normalization matrix.
- How to estimate the elements of $\mathbf{N}$ accurately and cheaply when $\boldsymbol{\kappa}$ is flow dependent? Different methods have been proposed, but none completely satisfactory.
- We can evaluate $\mathbf{L}^{1/2}\boldsymbol{\psi}_0$ by solving in sequence $M/2$ sparse linear SPD systems:

$$\left.\begin{array}{rcl} \mathbf{A}\boldsymbol{\psi}_1 & = & \boldsymbol{\psi}_0 \\ \mathbf{A}\boldsymbol{\psi}_2 & = & \boldsymbol{\psi}_1 \\ & \vdots & \\ \mathbf{A}\boldsymbol{\psi}_{M/2} & = & \boldsymbol{\psi}_{M/2-1} \end{array}\right\}$$

- Possible solvers: direct; multigrid; 1D approach (cf. recursive filter); iterative (polynomial-based).

Matrix formulation of implicit diffusion correlation models

- In factored form, the correlation matrix is

$$\mathbf{C} = \mathbf{N}\,\mathbf{L}^{1/2}\mathbf{W}^{-1}\left(\mathbf{L}^{1/2}\right)^{\mathrm{T}}\mathbf{N}$$

  - $\mathbf{W}$ contains grid-dependent weights.
  - $\mathbf{L}^{1/2} \longleftrightarrow \left(1 - \nabla \cdot \boldsymbol{\kappa}\nabla\right)^{-M/2}$ (self-adjoint w.r.t. $\mathbf{W}$ inner product).
  - $\mathbf{N}$ is a diagonal normalization matrix.
- How to estimate the elements of $\mathbf{N}$ accurately and cheaply when $\boldsymbol{\kappa}$ is flow dependent? Different methods have been proposed, but none completely satisfactory.
- We can evaluate $\mathbf{L}^{1/2}\boldsymbol{\psi}_0$ by solving in sequence $M/2$ sparse linear SPD systems:

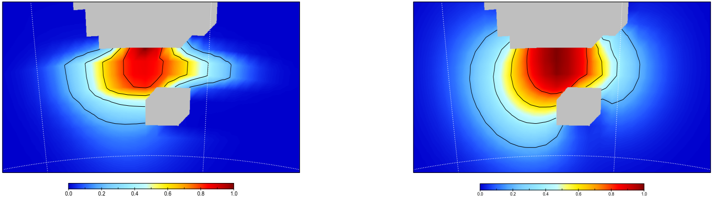
$$\left.\begin{array}{lcl} \mathbf{A}\boldsymbol{\psi}_1 & = & \boldsymbol{\psi}_0 \\ \mathbf{A}\boldsymbol{\psi}_2 & = & \boldsymbol{\psi}_1 \\ & \vdots & \\ \mathbf{A}\boldsymbol{\psi}_{M/2} & = & \boldsymbol{\psi}_{M/2-1} \end{array}\right\}$$

- Possible solvers: direct; multigrid; **1D approach (cf. recursive filter)**; iterative (polynomial-based).

- Approximate the 2D/3D operator ($\mathbf{L}$) as a product of simpler 1D operators ("recursive filters"):

$$\mathbf{L}^{1/2} \approx \mathbf{L}_1^{1/2}\mathbf{L}_2^{1/2}\mathbf{L}_3^{1/2}$$

- The resulting algorithm involves small, sparse SPD matrices that can be inverted efficiently using Cholesky decomposition.
- Drawbacks with the method: gets complex when accounting for general covariances; can produce numerical artefacts near complex boundaries; scales poorly on massively parallel machines.



(a) $2 \times 1\text{D}$ (b) 2D

(From Weaver *et al.* 2016)

Matrix formulation of implicit diffusion correlation models

- In factored form, the correlation matrix is

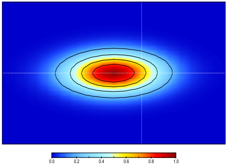
$$\mathbf{C} = \mathbf{N}\,\mathbf{L}^{1/2}\mathbf{W}^{-1}\left(\mathbf{L}^{1/2}\right)^{\mathrm{T}}\mathbf{N}$$

  - $\mathbf{W}$ contains grid-dependent weights.
  - $\mathbf{L}^{1/2} \longleftrightarrow \left(1 - \nabla\cdot\boldsymbol{\kappa}\nabla\right)^{-M/2}$ (self-adjoint w.r.t. $\mathbf{W}$ inner product).
  - $\mathbf{N}$ is a diagonal normalization matrix.
- How to estimate the elements of $\mathbf{N}$ accurately and cheaply when $\boldsymbol{\kappa}$ is flow-dependent? Different methods have been proposed, but none completely satisfactory.
- We can evaluate $\mathbf{L}^{1/2}\boldsymbol{\psi}_0$ by solving in sequence $M/2$ sparse linear SPD systems:

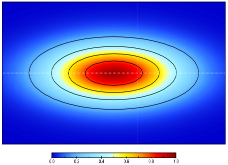
$$\left.\begin{array}{rcl} \mathbf{A}\boldsymbol{\psi}_1 & = & \boldsymbol{\psi}_0 \\ \mathbf{A}\boldsymbol{\psi}_2 & = & \boldsymbol{\psi}_1 \\ & \vdots & \\ \mathbf{A}\boldsymbol{\psi}_{M/2} & = & \boldsymbol{\psi}_{M/2-1} \end{array}\right\}$$

- Possible solvers: direct; multigrid; "1D approach" (recursive filter); **iterative (polynomial-based)**.

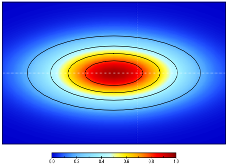
- The **Chebyshev iteration** (CI) is a linear solver that requires a guess and the extreme eigenvalues of $\mathbf{A}$ as input.
- For implicit diffusion, CI has similar convergence properties to conjugate gradients (Weaver *et al.* 2016).
- **No global MPI communications** (scalability bottleneck) required.
- It can also be used to solve a nonsymmetric version of the linear system to allow **parallelization in "time"** (M) (Weaver *et al.* 2018).
- It can be applied with a fixed number of iterations (K) to enforce **machine precision symmetry** (using the solver and its adjoint).
- A strict convergence criterion (large K) is not required to get an adequate solution.



(c) $K = 4$ ($\epsilon = 10^{-1}$)



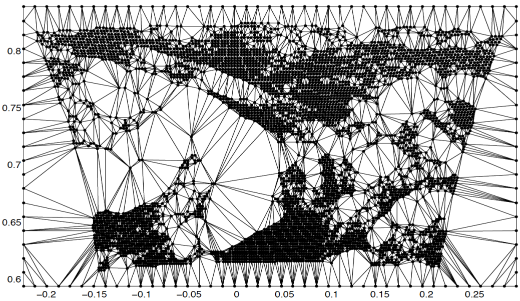
(d) $K = 13$ ($\epsilon = 10^{-3}$)



(e) $K = 43$ ($\epsilon = 10^{-10}$)

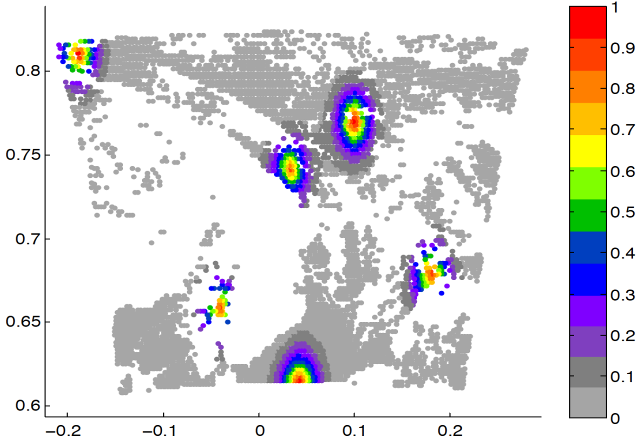
(From Weaver *et al.* 2016)

- Solve the diffusion equation using a **finite-element method**.
- A method appropriate for modelling correlations in $\mathbf{R}$ as well as $\mathbf{B}$.
- The **unstructured mesh** below is constructed from satellite observation locations (SEVIRI).



(From O. Guillet, PhD thesis CERFACS/Météo-France)

Diffusion-based correlations evaluated on an unstructured mesh using a finite-element method.



(From O. Guillet, PhD thesis CERFACS/Météo-France)

Specifying the correlation block-matrix operators in $\mathbf{C}$

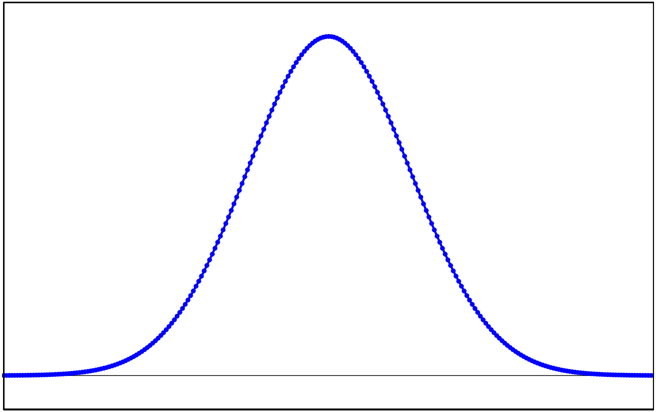
- Applying the correlation operator $\mathbf{C}$ is the most computationally demanding component of the $\mathbf{B}$ model.
- We need to be able to perform efficient matrix-vector products with the block-matrix components of $\mathbf{C}$.
- Common techniques in atmospheric variational DA:
  - Spectral/wavelet transform (Fisher 2003)
  - Recursive filter (Purser *et al.* 2003)
- **Issues**: regular grid required; complex boundaries difficult to handle; scalability
- Other (grid-point) techniques.
  - Diffusion operator (Weaver & Courtier 2001)
  - **Explicit convolution** (Gaspari & Cohn 1999; B. Ménétrier, unpublished material)
- **Issues**: cost; memory

- The computational cost of evaluating an explicit convolution integral is prohibitive on a high-resolution grid (cost is $O(n^2)$ where n is the number of grid points).
- The cost remains high even with compactly supported (space-limited) correlation functions (Gaspari & Cohn 1999).
- To limit the computational cost, we can approximate $\mathbf{C}$ on a subgrid; i.e., using a subset of $n^s \ll n$ points of the model grid.
- We define the correlation matrix as

$$\mathbf{C} = \mathbf{N}\,\mathbf{S}\,\mathbf{C}^s\,\mathbf{S}^{\mathrm{T}}\mathbf{N}$$

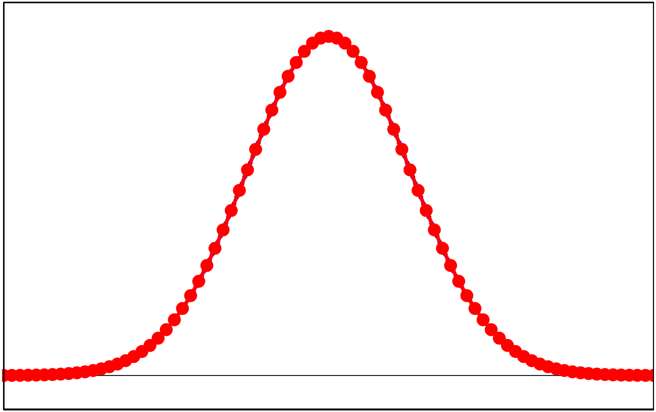
  - $\mathbf{S}$ is an **interpolation** from the subgrid to the model grid.
  - $\mathbf{C}^s = \mathbf{U}^s\mathbf{U}^{s\mathrm{T}}$ is a **convolution matrix** on the subgrid.
  - $\mathbf{N}$ is a diagonal **normalization matrix** ($\mathbf{S}\,\mathbf{C}^s\,\mathbf{S}^{\mathrm{T}}$ does not have diagonal elements equal to one, even if $\mathbf{C}^s$ does!).
- The convolution function will be distorted if the subgrid density is too coarse compared to the convolution length-scale.

Convolution function on model grid



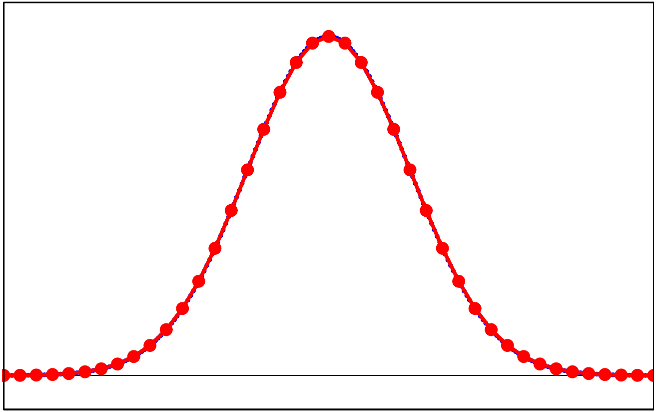
Model grid (blue)

Subsampling: 1 point over 3



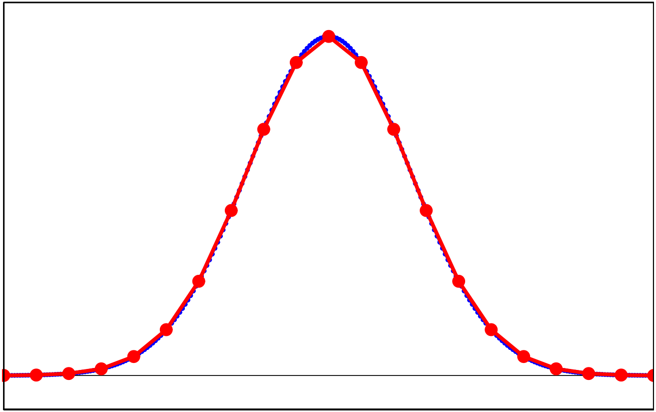
Model grid (blue) and subgrid (red)

Subsampling: 1 point over 6



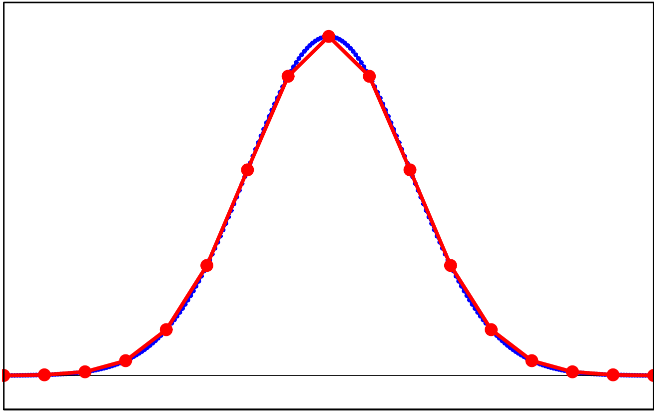
Model grid (blue) and subgrid (red)

Subsampling: 1 point over 12



Model grid (blue) and subgrid (red)

Subsampling: 1 point over 15



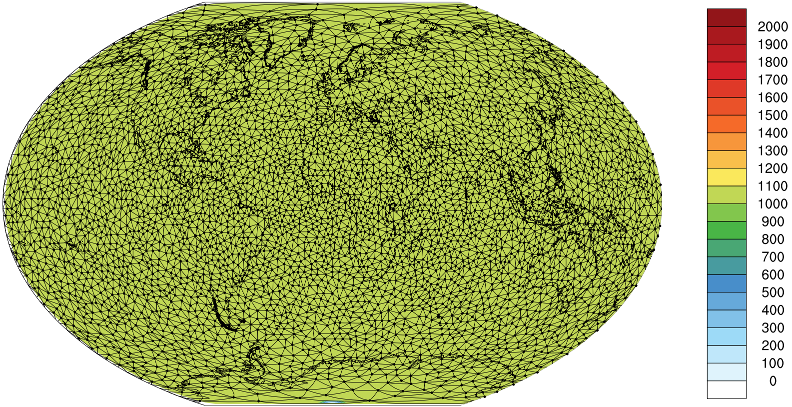
Model grid (blue) and subgrid (red)

Normalized Interpolated Convolution from an Adaptive Subgrid (NICAS)[1]

- Use the standard **compactly supported convolution function** of Gaspari & Cohn (1999) (Eq. (4.10)) with heterogeneous normalized distances:
  - $d'_{ij} = \dfrac{d_{ij}}{\sqrt{(r_i^2 + r_j^2)/2}}$ (distance-based)
  - $\widetilde{d}'_{ij} = \displaystyle\sum_{k=i}^{j-1} d'_{k,k+1}$ (network-based)
- The subgrid is locally adapted to account for the convolution length-scale.
- Local MPI communications are performed on the subgrid (no global MPI communications).
- A particularly efficient method when the convolution length-scales are large (e.g., as typically the case with localization).
- Complementary to the diffusion approach, which is more efficient for small length-scales.

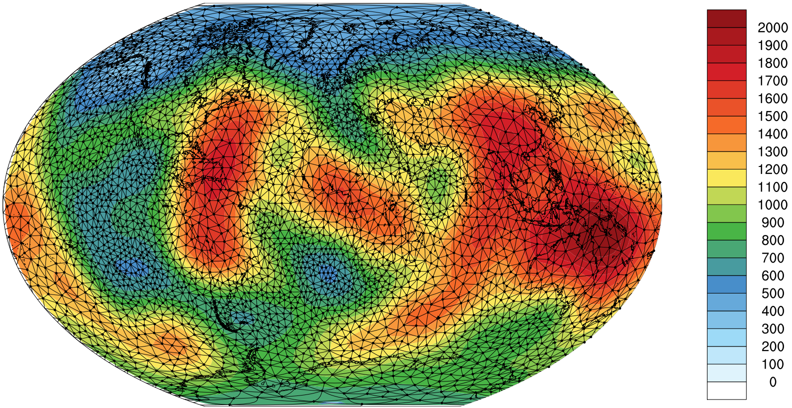
[1] Part of the "Background error covariance on Unstructured Mesh Package" (**BUMP**); B. Ménétrier, https://github.com/benjaminmenetrier/bump

Homogeneous convolution length-scale → homogeneous subgrid



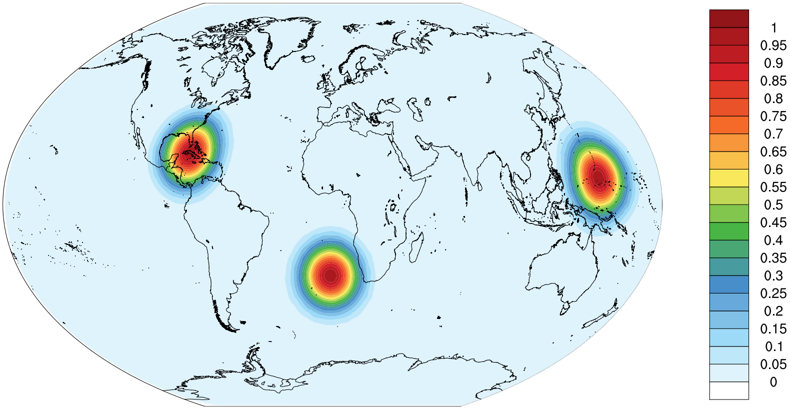
(Courtesy of B. Ménétrier)

Heterogeneous convolution length-scales → heterogeneous subgrid



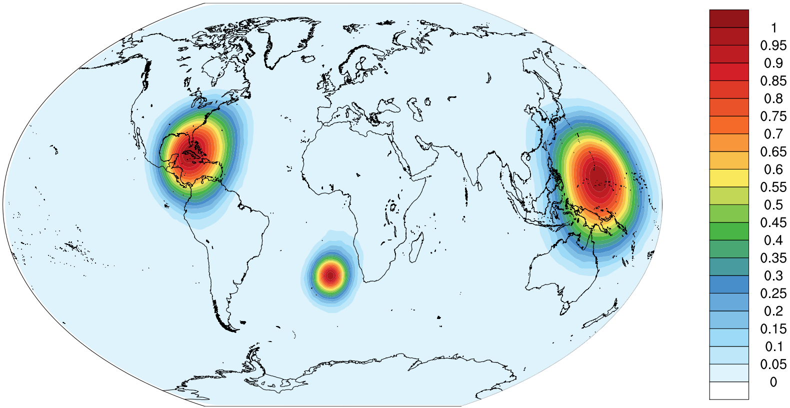
(Courtesy of B. Ménétrier)

Convolution with a homogeneous length-scale



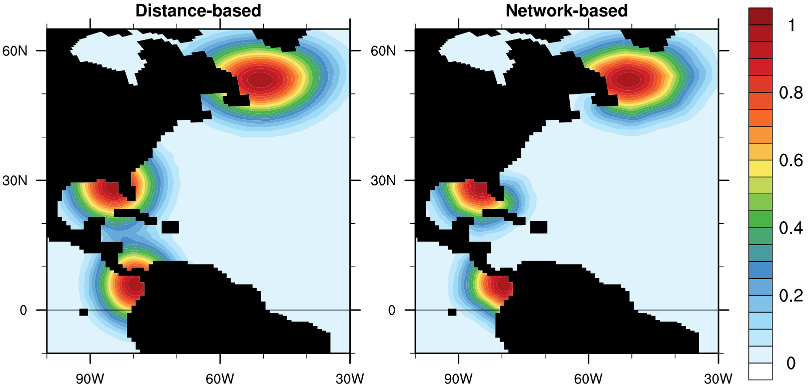
(Courtesy of B. Ménétrier)

Convolution with a heterogeneous length-scale



(Courtesy of B. Ménétrier)

Convolution functions in the presence of complex boundaries



Exact and cheap normalization is possible with both approaches.

(Courtesy of B. Ménétrier)

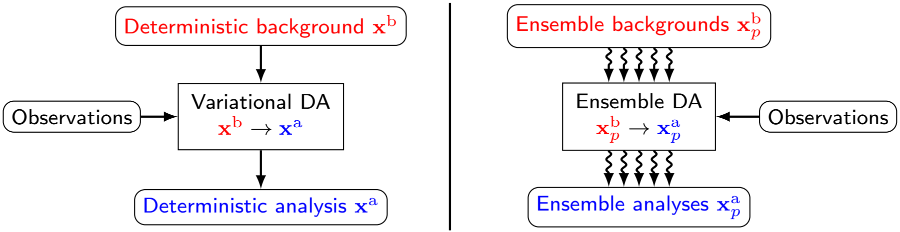
$\mathbf{B}$ basics

$\mathbf{B}$ modelling

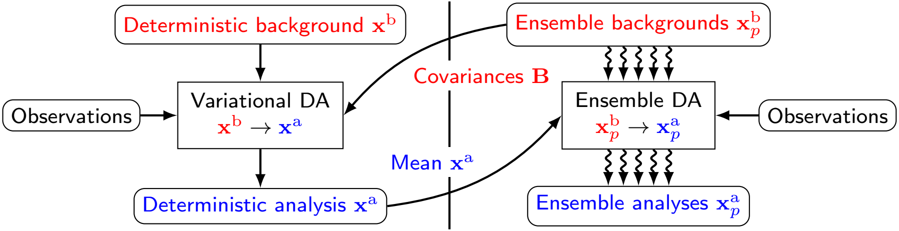
Using ensembles to define a flow-dependent $\mathbf{B}$

Concluding remarks

An ensemble of backgrounds is transformed into an ensemble of analyses



An ensemble of backgrounds is transformed into an ensemble of analyses



- Ensemble DA perturbations simulate deterministic errors.
- Deterministic and ensemble DA can use different algorithms, observations and model grids.
- Ensemble DA can also use variational methods.
- There are scientific and practical advantages in keeping the determinstic and ensemble DA systems as consistent as possible.

There are many methods to transform an ensemble of backgrounds into an ensemble of analyses that have been developed for NWP:

- Sequential filters (stochastic and deterministic)
  - Ensemble Kalman Filters (EnKF; Houtekamer *et al.* 1998)
  - Local Ensemble Transform Kalman Filter (LETKF; Hunt *et al.* 2007)
  - Ensemble Square Root Filter (EnSRF; Whitaker and Hamill 2002)
  - Ensemble Adjustment Kalman Filter (EAKF; Anderson 2003)
- Variational methods
  - Ensemble of Data Assimilations (EDA; Isaksen *et al.* 2010)
  - Mean-Pert method (Lorenc *et al.* 2016)
  - Var-EnKF (Buehner *et al.* 2017)
  - EVIL methods (Aulginé *et al.* 2017)

- How to define a huge matrix of $O(10^{14})$-$O(10^{18})$ elements from a small number ($O(10)$-$O(100)$) of ensemble perturbations?

- How to define a huge matrix of $O(10^{14})$-$O(10^{18})$ elements from a small number ($O(10)$-$O(100)$) of ensemble perturbations?
- The different approaches developed for variational DA can be divided into two classes:

- How to define a huge matrix of $O(10^{14})$-$O(10^{18})$ elements from a small number ($O(10)$-$O(100)$) of ensemble perturbations?

- The different approaches developed for variational DA can be divided into two classes:

1. Approaches involving estimating **covariance model parameters**, and possibly combining them with climatological or empirical estimates (hybrid parameters).
    - Sampling error is substantially reduced but (optimal) filtering of the ensemble-estimated parameters is still desirable.

Using ensembles to define a flow-dependent $\mathbf{B}$

- How to define a huge matrix of $O(10^{14})$-$O(10^{18})$ elements from a small number ($O(10)$-$O(100)$) of ensemble perturbations?

- The different approaches developed for variational DA can be divided into two classes:

1. Approaches involving estimating **covariance model parameters**, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Sampling error is substantially reduced but (optimal) filtering of the ensemble-estimated parameters is still desirable.
2. Approaches involving an ensemble-based **sample covariance matrix**.
   - Sampling error is huge and must be suppressed (inflation/filtering/localization).

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)

"The use of flow-dependent variances alone is able to introduce a significant degree of flow dependency in the analysis increments" (Bonavita *et al.* 2012)

ζ variances from EDA (upper) and quasi-static (lower)

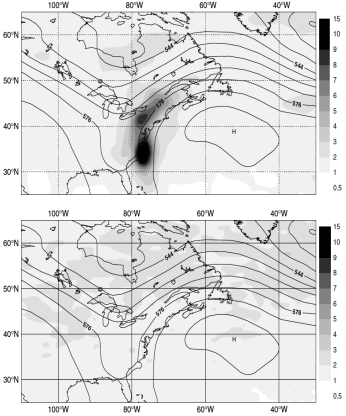


Figure 10. Geopotential at 500 hPa and MSLP background fields valid on 30 September 2009, 2100 UTC (first row). Geopotential at 500 hPa and vorticity background error standard deviations estimated from the ECMWF EDA (second row). Geopotential at 500 hPa and vorticity background error standard deviations computed with the randomization method (third row).

T (upper) and ζ (lower) increments from a single T observation for quasi-static (left) and EDA (right)

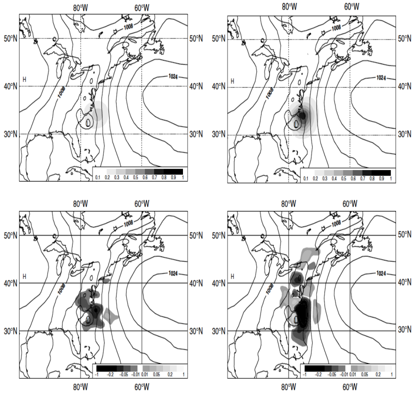
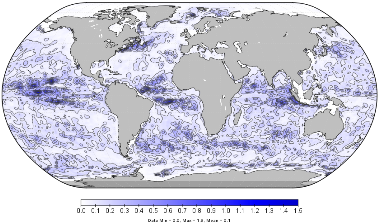


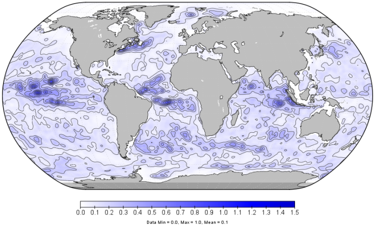
Figure 11. Single observation (observation departure $\Delta T = +1$ K at 900 hPa) analysis increments for temperature at model level 81 (first row) and vorticity at model level 78 (second row) valid on 30 September 2009, 2100 UTC. The first column refers to an assimilation experiment making use of the 'randomization' error estimates, the second column to an assimilation experiment making use of estimates from the ECMWF EDA. Scale intervals are in degrees for the temperature increments, in 10^{-5} s^{-1} for the vorticity increments.

Estimating flow-dependent variances from ensembles

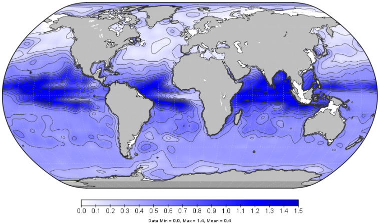
Background T error standard deviations at 100 m (11-member ensemble)



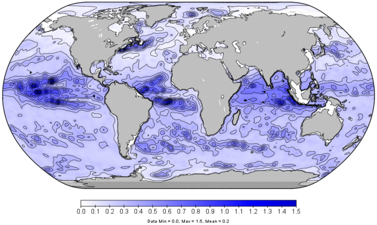
(a) Raw



(b) Objectively filtered (Ménétrier *et al.* 2015)



(c) Parameterized



(d) Hybrid

(Example from the NEMOVAR system; Weaver *et al.* 2018)

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
   - Vertical EOF-based correlations (Storto *et al.* 2018)

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
   - Vertical EOF-based correlations (Storto *et al.* 2018)
   - Local correlation tensor (Weaver & Mirouze 2013; Michel *et al.* 2016)

- The LCT ($\boldsymbol{H}$) is related to the diffusion tensor $\boldsymbol{\kappa}$ of an implicit diffusion-based correlation model (Weaver and Mirouze 2013):

$$\boldsymbol{\kappa}^{-1} \propto -\nabla\nabla^{\mathrm{T}} c_d\big|_{r=0} = \boldsymbol{H}$$

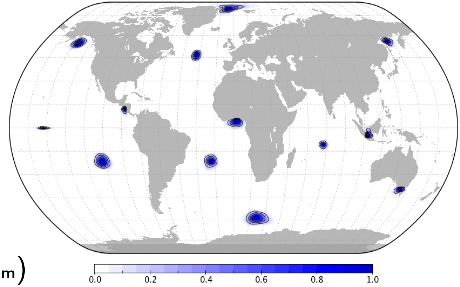
where c_d is a Matérn (Gaussian-like) correlation function in $\mathbb{R}^d$.

- $\boldsymbol{H}$ can be approximated locally using derivatives of ensemble perturbations $\epsilon(\mathbf{z})$ (Belo Pereira & Berre 2006; Michel *et al.* 2016):

$$\boldsymbol{H}(\mathbf{z}) \approx \overline{\nabla\widetilde{\epsilon}(\mathbf{z})\,(\nabla\widetilde{\epsilon}(\mathbf{z}))^{\mathrm{T}}} \quad \text{where} \quad \widetilde{\epsilon}(\mathbf{z}) = \epsilon(\mathbf{z})/\sigma(\mathbf{z})$$

- Filtering $\boldsymbol{H}$ is tricky to keep it positive definite (Michel *et al.* 2016).

Diffusion-modelled SSH correlations with $\boldsymbol{H}$ estimated from a 20-member ensemble.



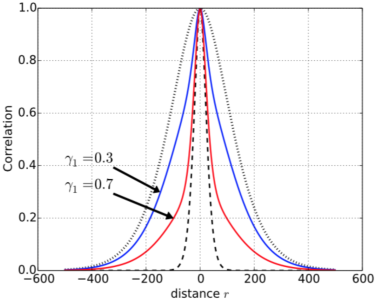
(Example from the NEMOVAR system)

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
   - Vertical EOF-based correlations (Storto *et al.* 2018)
   - Local correlation tensor (Weaver & Mirouze 2013; Michel *et al.* 2016)
   - Multiple scale covariance model parameters (Mirouze *et al.* 2016)

Estimating parameters of a multiple scale covariance model

- Linearly combine correlation models $\mathbf{C}_{\mathrm{m}_i}$, $i = 1, \ldots, n_{\mathrm{s}}$, with different LCTs to represent scale-dependent background errors.
- E.g., a "two-scale" model ($n_{\mathrm{s}} = 2$) can be used to separate background errors associated with the mesoscale (10-100 km) from those associated with larger scales (> 100 km) (Martin *et al.* 2007; Mirouze *et al.* 2016):

$$\mathbf{C}_{\mathrm{m}} = \gamma_1 \mathbf{C}_{\mathrm{m}_1} + (1 - \gamma_1)\mathbf{C}_{\mathrm{m}_2}$$



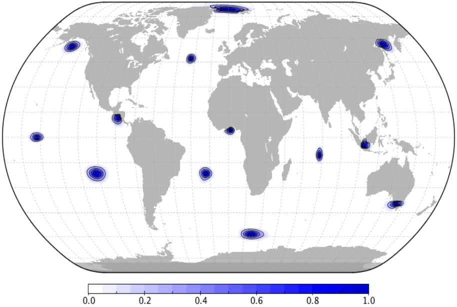
(From Mirouze *et al.* 2016)

Estimating parameters of a multiple scale covariance model

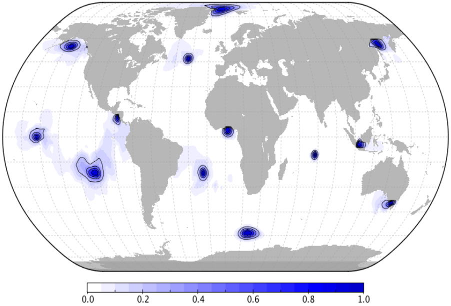
- A multiple scale covariance model can be fit locally to sample correlations from an ensemble (e.g., using BUMP).
- The estimated parameters (weights and LCTs) can be used with a diffusion-based correlation model (Weaver *et al.* 2016; 2018).

Diffusion-modelled SSH correlations with parameters estimated from a 20-member ensemble.

"One-scale" model



"Two-scale" model



(Example from the NEMOVAR system)

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
   - Vertical EOF-based correlations (Storto *et al.* 2018)
   - Local correlation tensor (Weaver & Mirouze 2013; Michel *et al.* 2016)
   - Multiple scale covariance model parameters (Mirouze *et al.* 2016)

2. Approaches involving an ensemble-based sample covariance matrix.

Using ensembles to define a flow-dependent $\mathbf{B}$

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
   - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
   - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
   - Vertical EOF-based correlations (Storto *et al.* 2018)
   - Local correlation tensor (Weaver & Mirouze 2013; Michel *et al.* 2016)
   - Multiple scale covariance model parameters (Mirouze *et al.* 2016)

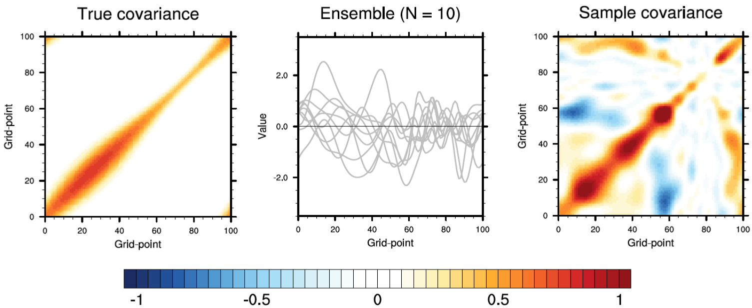
2. Approaches involving an ensemble-based sample covariance matrix.
   - Localized sample covariance matrix (EnVar) (Lorenc 2003b; Buehner 2005;...)

- Construct a low-rank sample covariance matrix $\widetilde{\mathbf{B}} = \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}}$ where

$$\widetilde{\mathbf{X}} = \frac{1}{\sqrt{N_{\mathrm{e}} - 1}} \left(\mathbf{x}_1 - \overline{\mathbf{x}}, \quad \ldots, \quad \mathbf{x}_{N_{\mathrm{e}}} - \overline{\mathbf{x}} \right)$$

- Remove perceived spurious correlations by forming the Schur product ($\circ$) of $\widetilde{\mathbf{B}}$ with a localization matrix $\mathbf{L}$:

$$\mathbf{B}_{\mathrm{e}} = \mathbf{L} \circ \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}} \quad \Longleftrightarrow \quad \left(B_{\mathrm{e}}\right)_{ij} = L_{ij}\,\widetilde{B}_{ij}$$

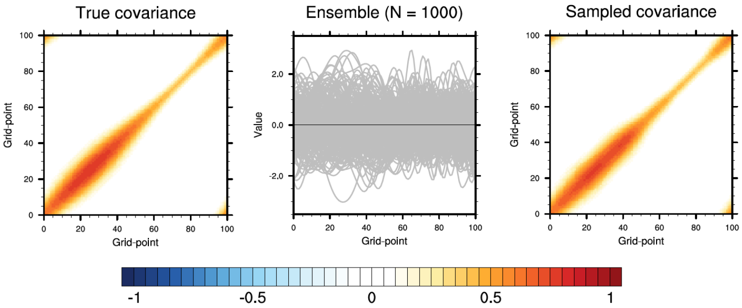


- Construct a low-rank sample covariance matrix $\widetilde{\mathbf{B}} = \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}}$ where

$$\widetilde{\mathbf{X}} = \frac{1}{\sqrt{N_{\mathrm{e}} - 1}} \left(\mathbf{x}_1 - \overline{\mathbf{x}}, \quad \ldots, \quad \mathbf{x}_{N_{\mathrm{e}}} - \overline{\mathbf{x}} \right)$$

- Remove perceived spurious correlations by forming the Schur product ($\circ$) of $\widetilde{\mathbf{B}}$ with a localization matrix $\mathbf{L}$:

$$\mathbf{B}_{\mathrm{e}} = \mathbf{L} \circ \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}} \qquad \Longleftrightarrow \qquad \left(B_{\mathrm{e}}\right)_{ij} = L_{ij}\,\widetilde{B}_{ij}$$

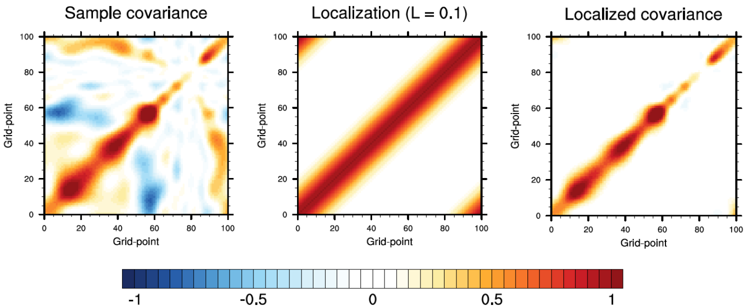


- Construct a low-rank sample covariance matrix $\widetilde{\mathbf{B}} = \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}}$ where

$$\widetilde{\mathbf{X}} = \frac{1}{\sqrt{N_{\mathrm{e}} - 1}} \left(\; \mathbf{x}_1 - \overline{\mathbf{x}}, \quad \ldots, \quad \mathbf{x}_{N_{\mathrm{e}}} - \overline{\mathbf{x}} \; \right)$$

- Remove perceived spurious correlations by forming the Schur product ($\circ$) of $\widetilde{\mathbf{B}}$ with a localization matrix $\mathbf{L}$:

$$\mathbf{B}_{\mathrm{e}} \; = \; \mathbf{L} \circ \widetilde{\mathbf{X}}\,\widetilde{\mathbf{X}}^{\mathrm{T}} \qquad \Longleftrightarrow \qquad \left(B_{\mathrm{e}}\right)_{ij} = L_{ij}\,\widetilde{B}_{ij}$$



- More generally, we can define (Clayton *et al.* 2013)

$$\mathbf{B}_{\mathrm{e}} \;=\; \mathbf{T}\big(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^{\mathrm{T}}\big)\mathbf{T}^{\mathrm{T}} \tag{2}$$

where $\widehat{\mathbf{X}} = \widehat{\mathbf{T}}^{-1}\widetilde{\mathbf{X}}$ and $\mathbf{T} \approx \widehat{\mathbf{T}}$.

- More generally, we can define (Clayton *et al.* 2013)

$$\mathbf{B}_\mathrm{e} \;=\; \mathbf{T}\big(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^\mathrm{T}\big)\mathbf{T}^\mathrm{T} \tag{2}$$

where $\widehat{\mathbf{X}} = \widehat{\mathbf{T}}^{-1}\widetilde{\mathbf{X}}$ and $\mathbf{T} \approx \widehat{\mathbf{T}}$.

- Some examples:
  - $\widehat{\mathbf{T}} \equiv \widetilde{\mathbf{\Sigma}}$ $\implies$ Localize the sample **correlation** matrix
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_\mathrm{bal}$ $\implies$ Localize the sample **covariance** matrix of the **unbalanced** variables
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_\mathrm{bal}\widetilde{\mathbf{\Sigma}}$ $\implies$ Localize the sample **correlation** matrix of the **unbalanced** variables

- More generally, we can define (Clayton *et al.* 2013)

$$\mathbf{B}_{\mathrm{e}} = \mathbf{T}\left(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^{\mathrm{T}}\right)\mathbf{T}^{\mathrm{T}} \tag{2}$$

where $\widehat{\mathbf{X}} = \widehat{\mathbf{T}}^{-1}\widetilde{\mathbf{X}}$ and $\mathbf{T} \approx \widehat{\mathbf{T}}$.

- Some examples:
  - $\widehat{\mathbf{T}} \equiv \widetilde{\mathbf{\Sigma}}$ $\Longrightarrow$ Localize the sample **correlation** matrix
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_{\mathrm{bal}}$ $\Longrightarrow$ Localize the sample **covariance** matrix of the **unbalanced** variables
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_{\mathrm{bal}}\widetilde{\mathbf{\Sigma}}$ $\Longrightarrow$ Localize the sample **correlation** matrix of the **unbalanced** variables

- Applying Eq. (2) in variational DA is made computationally feasible using the operator form

$$\left(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^{\mathrm{T}}\right)\mathbf{v} = \sum_{p=1}^{N_{\mathrm{e}}} \widehat{\mathbf{x}}'_p \circ \mathbf{L}\left(\widehat{\mathbf{x}}'_p \circ \mathbf{v}\right)$$

- More generally, we can define (Clayton *et al.* 2013)

$$\mathbf{B}_\mathrm{e} \;=\; \mathbf{T}\big(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^\mathrm{T}\big)\mathbf{T}^\mathrm{T} \qquad (2)$$

where $\widehat{\mathbf{X}} = \widehat{\mathbf{T}}^{-1}\widetilde{\mathbf{X}}$ and $\mathbf{T} \approx \widehat{\mathbf{T}}$.

- Some examples:
  - $\widehat{\mathbf{T}} \equiv \widetilde{\boldsymbol{\Sigma}}$ $\Longrightarrow$ Localize the sample **correlation** matrix
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_\mathrm{bal}$ $\Longrightarrow$ Localize the sample **covariance** matrix of the **unbalanced** variables
  - $\widehat{\mathbf{T}} \equiv \mathbf{K}_\mathrm{bal}\widetilde{\boldsymbol{\Sigma}}$ $\Longrightarrow$ Localize the sample **correlation** matrix of the **unbalanced** variables

- Applying Eq. (2) in variational DA is made computationally feasible using the operator form

$$\big(\mathbf{L} \circ \widehat{\mathbf{X}}\,\widehat{\mathbf{X}}^\mathrm{T}\big)\mathbf{v} = \sum_{p=1}^{N_\mathrm{e}} \widehat{\mathbf{x}}'_p \circ \mathbf{L}\big(\widehat{\mathbf{x}}'_p \circ \mathbf{v}\big)$$

- $\mathbf{L} = \mathbf{U}_\mathrm{L}\mathbf{U}_\mathrm{L}^\mathrm{T}$ can be formulated in different ways (with different costs): univariate, multivariate, scale-dependent, ...

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

- Separate the perturbations $\widehat{\mathbf{x}}'_p$ into $m = 1, \ldots, n_\mathrm{s}$ scales, from smallest to largest.

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

- Separate the perturbations $\widehat{\mathbf{x}}'_p$ into $m = 1, \ldots, n_\mathrm{s}$ scales, from smallest to largest.
- This can be done using a spatial filter $\mathbf{F}_{L_m}$ with filtering scale L_m, which we assume increases from $m = 1$ to $m = n_\mathrm{s} - 1$.

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

- Separate the perturbations $\widehat{\mathbf{x}}'_p$ into $m = 1, \dots, n_\mathrm{s}$ scales, from smallest to largest.
- This can be done using a spatial filter $\mathbf{F}_{L_m}$ with filtering scale L_m, which we assume increases from $m = 1$ to $m = n_\mathrm{s} - 1$.
- Define $\mathbf{f}_p^m = \mathbf{F}_{L_m}\widehat{\mathbf{x}}'_p$, from $m = 0, \dots, n_\mathrm{s}$, where $\mathbf{F}_{L_0} \equiv \mathbf{I}$ and $\mathbf{F}_{L_{n_\mathrm{s}}} \equiv \mathbf{0}$.

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

- Separate the perturbations $\widehat{\mathbf{x}}'_p$ into $m = 1, \ldots, n_\mathrm{s}$ scales, from smallest to largest.
- This can be done using a spatial filter $\mathbf{F}_{L_m}$ with filtering scale L_m, which we assume increases from $m = 1$ to $m = n_\mathrm{s} - 1$.
- Define $\mathbf{f}_p^m = \mathbf{F}_{L_m} \widehat{\mathbf{x}}'_p$, from $m = 0, \ldots, n_\mathrm{s}$, where $\mathbf{F}_{L_0} \equiv \mathbf{I}$ and $\mathbf{F}_{L_{n_\mathrm{s}}} \equiv \mathbf{0}$.
- Construct the scale-separated perturbations as

$$\mathbf{e}_k^{n_\mathrm{s}-m+1} = \mathbf{f}_k^{m-1} - \mathbf{f}_k^m,$$

so that $\sum_{m=1}^{n_\mathrm{s}} \mathbf{e}_k^{n_\mathrm{s}-m+1} = \widehat{\mathbf{x}}'_p$ (the original perturbation!).

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

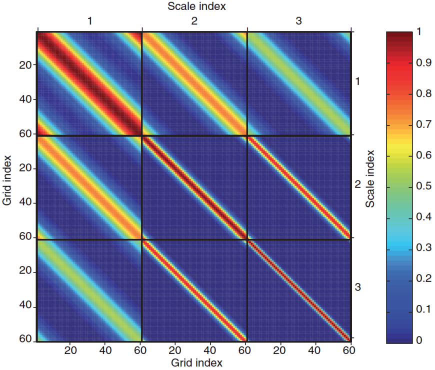
- Separate the perturbations $\widehat{\mathbf{x}}'_p$ into $m = 1, \ldots, n_\mathrm{s}$ scales, from smallest to largest.
- This can be done using a spatial filter $\mathbf{F}_{L_m}$ with filtering scale L_m, which we assume increases from $m = 1$ to $m = n_\mathrm{s} - 1$.
- Define $\mathbf{f}_p^m = \mathbf{F}_{L_m}\widehat{\mathbf{x}}'_p$, from $m = 0, \ldots, n_\mathrm{s}$, where $\mathbf{F}_{L_0} \equiv \mathbf{I}$ and $\mathbf{F}_{L_{n_\mathrm{s}}} \equiv \mathbf{0}$.
- Construct the scale-separated perturbations as

$$\mathbf{e}_k^{n_\mathrm{s}-m+1} = \mathbf{f}_k^{m-1} - \mathbf{f}_k^{m},$$

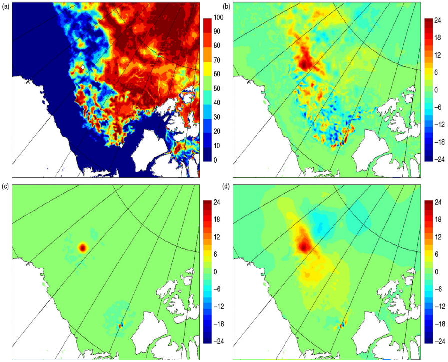
  so that $\sum_{m=1}^{n_\mathrm{s}} \mathbf{e}_k^{n_\mathrm{s}-m+1} = \widehat{\mathbf{x}}'_p$ (the original perturbation!).
- Apply a localization matrix $\mathbf{U}_m\mathbf{U}_n^\mathrm{T}$ to each of the sub-matrices $\mathbf{e}_k^m\left(\mathbf{e}_k^n\right)^\mathrm{T}$, $m, n = 1, \ldots, n_\mathrm{s}$, of an augmented sample covariance matrix.

Scale-dependent localization (SDL; Buehner and Shlyaeva *et al.* 2015)

SDL matrix with 3 scale-separation



Sea-ice concentration (two observations at ◇)



(From Buehner & Shylaeva 2015)

- SDL introduces a between-scale component in the localization matrix, with amplitude less than one.
- The SIC example shows the difference between large-scale localization only (top right), small-scale localization only (bottom left), and SDL (bottom right).

Using ensembles to define a flow-dependent $\mathbf{B}$

- How to define a huge matrix of $O(10^{14})$-$O(10^{18})$ elements from a small number ($O(10)$-$O(100)$) of ensemble perturbations?
- The different approaches developed for variational DA can be divided into two classes:

1. Approaches involving estimating covariance model parameters, and possibly combining them with climatological or empirical estimates (hybrid parameters).
    - Variances (Bonavita *et al.* 2012; Raynaud *et al.* 2011; Daget *et al.* 2009)
    - Wavelet-based correlations (Bonavita *et al.* 2016; Berre *et al.* 2015)
    - Vertical EOF-based correlations (Storto *et al.* 2018)
    - Local correlation tensor (Weaver & Mirouze 2013; Michel *et al.* 2016)
    - Multiple scale covariance model parameters (Mirouze *et al.* 2016)
2. Approaches involving an ensemble-based sample covariance matrix.
    - Localized sample covariance matrix (EnVar) (Lorenc 2003b; Buehner 2005;...)
    - Hybrid $\mathbf{B}$ (Hamill & Snyder 2000; Lorenc 2003b;...)

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_{\mathrm{m}}^2 \, \mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2 \, \mathbf{B}_{\mathrm{e}}$$

where β_{m}^2 and β_{e}^2 are weighting coefficients.

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_{\mathrm{m}}^2 \, \mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2 \, \mathbf{B}_{\mathrm{e}}$$

  where β_{m}^2 and β_{e}^2 are weighting coefficients.

- $\mathbf{B}_{\mathrm{m}}$ usually contains "climatological" covariance information and is intended to provide robustness to $\mathbf{B}$.

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2\,\mathbf{B}_{\mathrm{e}}$$

  where β_{m}^2 and β_{e}^2 are weighting coefficients.

- $\mathbf{B}_{\mathrm{m}}$ usually contains "climatological" covariance information and is intended to provide robustness to $\mathbf{B}$.
- How to estimate β_{m}^2, β_{e}^2 and $\mathbf{L}$?

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2\,\mathbf{B}_{\mathrm{e}}$$

where β_{m}^2 and β_{e}^2 are weighting coefficients.

- $\mathbf{B}_{\mathrm{m}}$ usually contains "climatological" covariance information and is intended to provide robustness to $\mathbf{B}$.
- How to estimate β_{m}^2, β_{e}^2 and $\mathbf{L}$?
- One approach (Ménétrier and Auligné 2015) is to absorb β_{e}^2 into the localization $\mathbf{L}$ and determine L_{ij} and β_{m}^2 (neglecting systematic error!) by solving

$$\min_{\beta_{\mathrm{m}}^2,\,L_{ij}} \mathbb{E}\left[\|\beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} + \mathbf{B}_{\mathrm{e}}(L_{ij}) - \widetilde{\mathbf{B}}^{\star}\|_{\mathrm{F}}^2\right]$$

where $\widetilde{\mathbf{B}}^{\star} = \lim_{N_{\mathrm{e}}\to\infty} \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{T}}$.

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} + \beta_{\mathrm{e}}^2\,\mathbf{B}_{\mathrm{e}}$$

  where β_{m}^2 and β_{e}^2 are weighting coefficients.

- $\mathbf{B}_{\mathrm{m}}$ usually contains "climatological" covariance information and is intended to provide robustness to $\mathbf{B}$.
- How to estimate β_{m}^2, β_{e}^2 and $\mathbf{L}$?
- One approach (Ménétrier and Aulígné 2015) is to absorb β_{e}^2 into the localization $\mathbf{L}$ and determine L_{ij} and β_{m}^2 (neglecting systematic error!) by solving

$$\min_{\beta_{\mathrm{m}}^2, L_{ij}} \mathbb{E}\left[\|\beta_{\mathrm{m}}^2\,\mathbf{B}_{\mathrm{m}} + \mathbf{B}_{\mathrm{e}}(L_{ij}) - \widetilde{\mathbf{B}}^{\star}\|_{\mathrm{F}}^2\right]$$

  where $\widetilde{\mathbf{B}}^{\star} = \lim_{N_{\mathrm{e}}\to\infty} \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{T}}$.

- Fit the localization estimates to a correlation function (Matérn, Gaussian,...) and model $\mathbf{L}$ with a diffusion operator or spectral transform.

- Linearly combine a modelled and localized sample covariance matrix:

$$\mathbf{B} = \beta_\mathrm{m}^2 \, \mathbf{B}_\mathrm{m} + \beta_\mathrm{e}^2 \, \mathbf{B}_\mathrm{e}$$

where β_m^2 and β_e^2 are weighting coefficients.

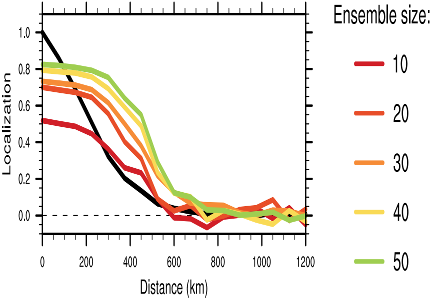
- $\mathbf{B}_\mathrm{m}$ usually contains "climatological" covariance information and is intended to provide robustness to $\mathbf{B}$.
- How to estimate β_m^2, β_e^2 and $\mathbf{L}$?
- One approach (Ménétrier and Auligné 2015) is to absorb β_e^2 into the localization $\mathbf{L}$ and determine L_{ij} and β_m^2 (neglecting systematic error!) by solving

$$\min_{\beta_\mathrm{m}^2, L_{ij}} \mathbb{E}\left[\| \beta_\mathrm{m}^2 \, \mathbf{B}_\mathrm{m} + \mathbf{B}_\mathrm{e}(L_{ij}) - \widetilde{\mathbf{B}}^\star \|_\mathrm{F}^2 \right]$$

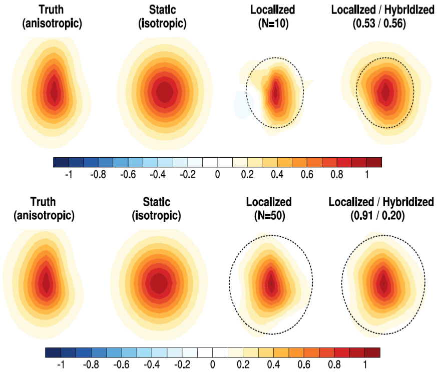
where $\widetilde{\mathbf{B}}^\star = \lim_{N_\mathrm{e} \to \infty} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^\mathrm{T}$.

- Fit the localization estimates to a correlation function (Matérn, Gaussian,...) and model $\mathbf{L}$ with a diffusion operator or spectral transform.
- Alternatively, fit them to a compactly supported correlation function (Gaspari and Cohn 1999) and use direct convolution (e.g., as done in BUMP).

Estimated localization functions

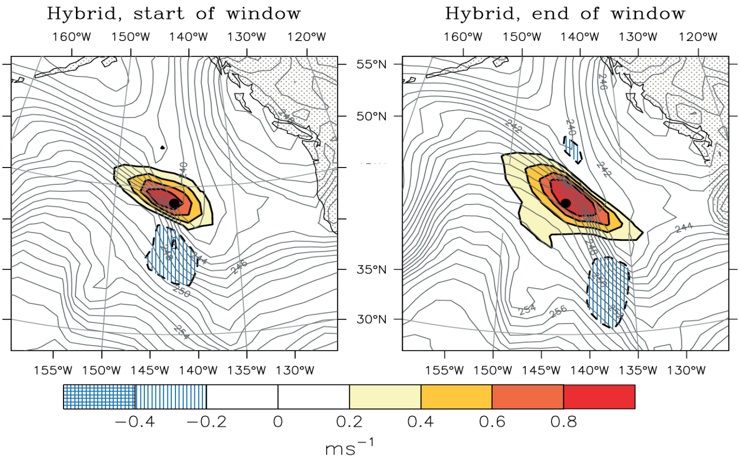


Hybrid covariances



- Localization scales increase with ensemble size.
- Hybrid weight β_e^2 (amplitude) increases with ensemble size.

Zonal wind increment produced from a single zonal wind observation at the start (left panel) of a 6hr 4D-Var window, using a hybrid $\mathbf{B}$.



(From Clayton *et al.* 2013)

$\mathbf{B}$ basics

$\mathbf{B}$ modelling

Using ensembles to define a flow-dependent $\mathbf{B}$

Concluding remarks

- Continue to improve the use of ensembles in flow-dependent formulations of $\mathbf{B}$.
  - Better covariance models (scale-dependent, anisotropic tensors, ...); flexible localization operators (scale-dependent, multivariate, adaptive, ...); hybrid $\mathbf{B}$; algorithms for estimating and filtering covariance parameters.

- Continue to improve the use of ensembles in flow-dependent formulations of $\mathbf{B}$.
  - Better covariance models (scale-dependent, anisotropic tensors, ...); flexible localization operators (scale-dependent, multivariate, adaptive, ...); hybrid $\mathbf{B}$; algorithms for estimating and filtering covariance parameters.
- Continue to improve the computational aspects of correlation operators (for covariance models and localization).
  - Methods suitable for unstructured meshes (these will be needed for both $\mathbf{B}$ and $\mathbf{R}$); accurate and efficient normalization procedures; better solvers and parallelization strategies.

- Continue to improve the use of ensembles in flow-dependent formulations of $\mathbf{B}$.
  - Better covariance models (scale-dependent, anisotropic tensors, ...); flexible localization operators (scale-dependent, multivariate, adaptive, ...); hybrid $\mathbf{B}$; algorithms for estimating and filtering covariance parameters.
- Continue to improve the computational aspects of correlation operators (for covariance models and localization).
  - Methods suitable for unstructured meshes (these will be needed for both $\mathbf{B}$ and $\mathbf{R}$); accurate and efficient normalization procedures; better solvers and parallelization strategies.
- Estimating, understanding and modelling coupled (cross-component) covariances (this work has only just begun!).

- Continue to improve the use of ensembles in flow-dependent formulations of $\mathbf{B}$.
  - Better covariance models (scale-dependent, anisotropic tensors, ...); flexible localization operators (scale-dependent, multivariate, adaptive, ...); hybrid $\mathbf{B}$; algorithms for estimating and filtering covariance parameters.
- Continue to improve the computational aspects of correlation operators (for covariance models and localization).
  - Methods suitable for unstructured meshes (these will be needed for both $\mathbf{B}$ and $\mathbf{R}$); accurate and efficient normalization procedures; better solvers and parallelization strategies.
- Estimating, understanding and modelling coupled (cross-component) covariances (this work has only just begun!).
- Development of flexible software infrastructure (refactored codes, OOPS, BUMP,...) is essential.

1. Anderson J. L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**: 634–642.
2. Auligné T., Ménétrier B., Lorenc A., Buehner M. 2016. Ensemble-Variational Integrated Localized Data Assimilation. *Mon. Wea. Rev.*, **144**: 3677–3696.
3. Bui-Thanh T., Ghattas O., Martin J., Stadler G. 2013. A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion. *SIAM J. Sci. Comput.*, **35**: A2494–A2523.
4. Belo Pereira M., Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.*, **134**: 2466–2489.
5. Berre L., Varella H., Desroziers G. 2015. Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Q. J. Roy. Meteorol. Soc.*, **141**: 2803–2812.
6. Bonavita M., Hólm E., Isaksen L., Fisher M. 2016: The evolution of the ECMWF hybrid data assimilation system. *Q. J. Roy. Meteorol. Soc.*, **142**: 287-303.
7. Bonavita M., Isaksen L., Hólm E. 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Q. J. Roy. Meteorol. Soc.*, **138**: 1540-1559.
8. Buehner M., McTaggart-Cowan R., Heilliette S. 2017. An Ensemble Kalman Filter for Numerical Weather Prediction based on variational data assimilation: VarEnKF. *Mon. Wea. Rev.*, **145**: 617–635.
9. Buehner M., McTaggart-Cowan R., Beaulne A., Charette C., Garand L., Heilliette S., Lapalme E., Laroche S., Macpherson S. R., Morneau J., Zadra A. 2015: Implementation of determinstic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: the global system. *Mon. Wea. Rev.*, **143**: 2532–2559.
10. Buehner M., Shlyaeva A. 2015: Scale-dependent background-error covariance localisation. *Tellus*, **67A**: 28027.
11. Buehner M. 2005. Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, **131**: 1013–1043.

12. Clayton A. M., Lorenc A. C., Barker D. M. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.*, **139**: 1445–1461.
13. Daget N., Weaver A. T., Balmaseda M. A. 2009. Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean. *Q. J. R. Meteorol. Soc.*, **135**: 1071–1094.
14. Derber J. C., Bouttier F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**: 195–221.
15. Fisher M. 2003. Background error covariance modelling. In Proceedings of Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean: 45–63 ECMWF: Reading, UK.
16. Gaspari G., Cohn S. E. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**: 723–757.
17. Gmeiner B., Drzisga D, Rüde U., Scheichl R., Wohlmuth B. 2017. Scheduling massively parallel multigrid for multilevel Monte Carlo methods, *SIAM J. Sci. Comput.*, **39**: S873–S897.
18. Gürol S., Weaver A. T., Moore A. M., Piacentini A., Arango H., Gratton S. 2014. Preconditioned minimization algorithms for variational data assimilation. *Q. J. R. Meteorol. Soc.*, **140**: 539–556.
19. Guttorp P., Gneiting T. 2006. Miscellanea studies in the history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika*, **93**: 989–995.
20. Gustafsson N. 2007. Discussion on "4D-Var or EnKF?". *Tellus*, **59A**: 774–777.
21. Hamill T. M., Snyder C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**: 2905–2919.
22. Houtekamer P. L , Mitchell H. L, Pellerin G., Buehner M. Charron M., Spacek L., Hansen B. 2005. Atmospheric data assimilation with the ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.
23. Hunt B. R., Kostelich E. J., Szunyogh I. 2007. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman Filter. *Physica D: Nonlinear Phenomena*, **230 (12)**, 112–126.
24. Isaksen L., Bonavita M., Buizza R., Fisher M., Haseler J., Leutbecher M., Raynaud L. 2010. Ensemble Data Assimilations at ECMWF. ECMWF Tech. Memo. 636.

25. Lindgren F., Rue H., Lindström J. 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. Roy. Stat. Soc.: Series B Stat. Method.*, **73**: 423–498.
26. Lorenc A. C., Bowler N. E., Clayton A. M., Pring S. R., Fairbairn D. 2015. Comparison of Hybrid-4DEnVar and Hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212–229.
27. Lorenc A. C., Jardek M., Payne T., Bowler N. E., Wlasak M. A. 2016. Computing an ensemble of variational data assimilations using its mean and perturbations. *Q. J. R. Meteorol. Soc.*.
28. Lorenc, A. C. 2003a. Modelling of error covariances by four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **129**, 3167–3182.
29. Lorenc A. C. 2003b. The potential of the ensemble Kalman filter for NWP - A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, **129**: 3183–3203.
30. Martin M. J., Hines A., Bell M. J. 2007. Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. *Q. J. R. Meteorol. Soc.*, **133**: 981–995.
31. Ménétrier B., Auligné T. 2015. Optimized localization and hybridization to filter ensemble-based covariances. *Mon. Wea. Rev.*, **143**: 3931–3947.
32. Ménétrier B., Montmerle T., Michel Y., Berre L. 2015a. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Wea. Rev.*, **143**: 1622–1643.
33. Michel Y., Ménétrier B., Montmerle T. 2016. Objective filtering of the local correlation tensor. *Q. J. R. Meteorol. Soc.*, **142**: 2314–2323.
34. Mirouze I., Blockley E. W., Lea D. J., Martin M. J., Bell M. J. 2016. A multiple length scale correlation operator for ocean data assimilation. *Tellus A*, **68**: 29 744.
35. Parrish D. F., Derber J. C. 1992. The National Meteorological Center's Spectral Statistical-Interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
36. Purser R. J., Wu W.S., Parrish D. F., Roberts N. M. 2003a. Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**: 1524–1535.

25. Raynaud L., Berre L., Desroziers G. 2011. An extended specification of flow-dependent background-error variances in the Météo-France global 4D-Var system. *Q. J. R. Meteorol. Soc.*, **137**: 607–619.
26. Storto A., Oddo P., Cipollone A., Mirouze I., Lemieux B. 2018. Extending an oceanographic variational scheme to allow for affordable hybrid and four-dimensional data assimilation. *Ocean Model.*, **128**: 67–86.
27. Weaver A. T., Chrust M., Ménétrier B., Piacentini A., Tshimanga J., Yang Y., Gürol S. and Zuo H. 2018. Using ensemble-estimated background error variances and correlation scales in the NEMOVAR system. CERFACS Technical Report TR/PA/18/15 (report prepared for the ERA-CLIM2 EU-FP7 project). Available at https://cerfacs.fr/wp-content/uploads/2018/01/TR-PA-18-15.pdf
28. Weaver A. T., Gürol S., Tshimanga J., Chrust M., Piacentini A. 2018. "Time"-parallel diffusion-based correlation operators. *Q. J. R. Meteorol. Soc.*, in press, https://doi.org/10.1002/qj.3302.
29. Weaver A. T., Tshimanga J., Piacentini, A. 2016. Correlation operators based on an implicitly formulated diffusion equation solved with the Chebyshev iteration. *Q. J. R. Meteorol. Soc.*, **142**: 455–471.
30. Weaver A. T., Mirouze I. 2013. On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation. *Q. J. R. Meteorol. Soc.*, **139**: 242–260.
31. Weaver A. T., Deltel C., Machu E., Ricci S., Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.*, **131**: 3605–3625.
32. Weaver A. T., Courtier P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.*, **127**: 1815–1846.
33. Whitaker J. S., Hamill T. M. 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**: 1913–1924.
34. Yaremchuk M., Nechaev D. 2013. Covariance localization with the diffusion-based correlations models. *Mon. Wea. Rev.*, **141**: 848–860.