Using ensembles with a diffusion equation to define background-error correlations in variational (ocean) data assimilation

Anthony Weaver and Isabelle Mirouze

CERFACS, Toulouse

Outline

1. Using enembles to specify the background-error covariances in VarDA
2. Representing correlation functions via a diffusion equation
3. Estimating the diffusion tensor from ensemble statistics
4. Summary

Outline

1. Using enembles to specify the background-error covariances in VarDA
2. Representing correlation functions via a diffusion equation
3. Estimating the diffusion tensor from ensemble statistics
4. Summary

Using ensembles to specify **B** in VarDA

There are different ways to use ensembles to specify **B** in variational data assimilation.

1. Estimate **B** directly from an ensemble (as in EnKF).
   - Multivariate, inhomogeneous and anisotropic by construction.
   - Straightforward to implement in VarDA, e.g., using the α-control variable (Lorenc 2003).
   - Effective number of degrees of freedom is much less than the number of background variables. *Can lead to problems fitting the data.*
   - Covariance localization is necessary to reduce the effects of sampling error. *Somewhat ad hoc. Can disrupt dynamical balance.*

2. Use an ensemble to calibrate a parametric model for **B**.
   - A flexible and efficient model is needed for describing the correlations of the analysis variables. *Computationally challenging*.
   - Balance operators are used for the multivariate part. *Multivariate covariance information in the ensemble is neglected*.
   - Ensembles are typically used to estimate variances and parameters of the correlation model (length-scales or spectral/wavelet coefficients).
   - Fewer degrees of freedom to estimate.
   - The covariances are localized by construction.

3. Use a linear combination of the two **B** models above.
   - Usually referred to as *hybrid* data assimilation (e.g. *Wang et al. 2008*).
   - The parameterized **B** model in hybrid DA is usually based on a simplified (isotropic, homogeneous), static formulation.
   - Requires tuning of empirical weighting coefficients.

Using ensembles to specify **B** in VarDA: basic approaches

2. Use an ensemble to calibrate a parametric model for **B**.
    - A flexible and efficient model is needed for describing the correlations of the analysis variables. *Computationally challenging*.
    - Balance operators are used for the multivariate part. *Multivariate covariance information in the ensemble is neglected*.
    - Ensembles are typically used to estimate variances and parameters of the correlation model (length-scales or spectral/wavelet coefficients).
    - Fewer degrees of freedom to estimate.
    - The covariances are localized by construction.
3. Use a linear combination of the two **B** models above.
    - Usually referred to as *hybrid* data assimilation (e.g. *Wang et al. 2008*).
    - The parameterized **B** model in hybrid DA is usually based on a simplified (isotropic, homogeneous), static formulation.
    - Requires tuning of empirical weighting coefficients.

Outline

Representation of correlations via a diffusion equation

- Correlation operators for large VarDA problems can be conveniently modelled using a differential operator (grid-point filter) derived from the explicit or implicit solution of a diffusion equation (*Derber and Rosati 1989; Egbert et al.!994; Weaver and Courtier 2001; Pannekoucke and Massart 2008; Mirouze and Weaver 2010; Carrier and Ngodock 2010*...).
- Especially convenient in complex boundary domains (implementation of BCs straightforward).
- Widely used in ocean VarDA.
- Most ocean VarDA applications with the diffusion equation tend to use rather simple correlation structures (quasi-isotropic) and subjective estimates of the length-scales.
- Here the purpose is to outline:
  1. how the diffusion equation can be used to represent **anisotropic** and **inhomogeneous** correlation functions; and
  2. how the correlation structures can be calibrated using ensembles.

Accounting for anisotropy using tensors: some definitions

- *Aspect tensor* $\boldsymbol{A}$: For a correlation function $C(\widetilde{r})$ that depends on the (non-dimensional) distance $\widetilde{r}$ between locations $\mathbf{x}$ and $\mathbf{x}'$ then

$$\widetilde{r} = \|\widetilde{\mathbf{r}}\|_{\boldsymbol{A}^{-1}} = \sqrt{(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}\boldsymbol{A}^{-1}(\mathbf{x}-\mathbf{x}')}$$

  - Isotropic case: $\boldsymbol{A} = L^2\,\mathbf{I}$ and $\widetilde{r} = |\mathbf{x}-\mathbf{x}'|\,/L$.

- *Correlation Hessian* $\boldsymbol{H}$ and "*Daley" tensor* $\boldsymbol{D}$:

$$\boldsymbol{H} = \boldsymbol{D}^{-1} = -\,\nabla\nabla^{\mathrm{T}}C(\widetilde{r})\big|_{\widetilde{r}=0}$$

  - Isotropic case: $\boldsymbol{D} = D^2\,\mathbf{I}$ where D is the Daley length-scale.

- *Diffusion tensor* $\boldsymbol{\kappa}$:

$$\frac{\partial\eta}{\partial t} - \nabla\cdot\boldsymbol{\kappa}\nabla\eta = 0$$

  - Rescaled tensor: $\boldsymbol{L} = \Delta t\,\boldsymbol{\kappa}$ after "temporal" discretization.

All these tensors are assumed to be **symmetric** and **positive definite** (and hence invertible).

Why are these different tensors of interest?

- The normalized kernel of a diffusion operator with *constant* $\boldsymbol{\kappa}$ is a correlation function $C(\widetilde{r})$ with known analytical form.
  - The diffusion kernel with an explicit scheme approximates a Gaussian.
  - The diffusion kernel with an M-step implicit scheme is a member of the Whittle-Matérn or Matérn correlation family (see later).

- Link to ensemble estimation.
  - The Hessian $\boldsymbol{H}$, and hence $\boldsymbol{D}$, can be estimated from ensemble statistics (see later).
  - $\boldsymbol{H}$ can in turn be related to the aspect tensor $\boldsymbol{A}$ of the Gaussian and Matérn functions.
  - $\boldsymbol{A}$ can in turn be related to $\boldsymbol{\kappa}$ (or $\boldsymbol{L}$) of the explicit or implicit diffusion operator.

- Estimating $\boldsymbol{H}(\mathbf{x})$ at each grid-point $\mathbf{x}$ and using it to define $\boldsymbol{\kappa}(\mathbf{x})$ in the diffusion operator allows us to model **anisotropic** and **inhomogeneous** correlation functions.

Consider the 2D diffusion equation

$$\frac{\partial \eta}{\partial t} - \nabla \cdot \boldsymbol{\kappa} \nabla \eta = 0$$

where $\boldsymbol{\kappa}$ is an anisotropic (but constant) **diffusion tensor**

$$\boldsymbol{\kappa} = \begin{pmatrix} \kappa_{xx} & \kappa_{xy} \\ \kappa_{yx} & \kappa_{yy} \end{pmatrix}$$

which is assumed symmetric $\kappa_{yx} = \kappa_{xy}$ and positive definite.

Note: t is a pseudo-time variable in this context.

The solution is a Gaussian covariance operator:

$$\eta(x,y,t) = \int_{\mathbb{R}^2} C(x,y,x',y')\,\eta(x',y',0)\,\mathrm{d}x'\mathrm{d}y'$$

where

$$C(x,y,x',y') = C(\widetilde{r}\,) = \gamma^{-1}\, e^{-\widetilde{r}^{\,2}/2},$$

$$\gamma = 2\pi\,|\boldsymbol{A}|^{1/2},$$

$$\widetilde{r} = \sqrt{(\mathsf{x}-\mathsf{x}')^{\mathrm{T}}\,\boldsymbol{A}^{-1}\,(\mathsf{x}-\mathsf{x}')},$$

$$\boldsymbol{A} = 2\,t\,\boldsymbol{\kappa},$$

$$\boldsymbol{H} = -\,\nabla\,\nabla^{\mathrm{T}} C\big|_{\widetilde{r}=0} = \boldsymbol{A}^{-1}.$$

Representing anisotropic correlations with an **explicit** scheme

The anisotropic Gaussian correlation operator can be approximated numerically by iterating the diffusion operator with an **explicit** scheme:

$$\eta(x, y, t) = \gamma \left(1 + \nabla \cdot \boldsymbol{L}\nabla\right)^M \eta(x, y, 0)$$

where $\boldsymbol{L} = \Delta t\, \boldsymbol{\kappa}$.

We can relate $\boldsymbol{L}$ to $\boldsymbol{D}$:

$$\boldsymbol{L} = \frac{2M\Delta t}{2M}\,\boldsymbol{\kappa} = \frac{2t}{2M}\,\boldsymbol{\kappa} = \frac{1}{2M}\,\boldsymbol{A} = \frac{1}{2M}\,\boldsymbol{H}^{-1}$$

or

$$\boxed{\boldsymbol{L} = \frac{1}{2M}\,\boldsymbol{D}.}$$

The scheme is *conditionally stable*. In the isotropic case $M > 2(D/\Delta x)^2$.

Consider the solution to the linear system

$$\gamma^{-1} \left(1 - \nabla \cdot \boldsymbol{L} \nabla\right)^{M} \eta(x, y, t) = \eta(x, y, 0)$$

where, with foresight,

$$\gamma = 4\pi (M - 1) \, |\boldsymbol{L}|^{1/2}.$$

This elliptic equation can be interpreted as the inverse of a diffusion operator resulting from **implicit** time-discretization with $\boldsymbol{L} = \Delta t \, \boldsymbol{\kappa}$.

The scheme is *unconditionally stable*, so M is a free parameter.

Representing anisotropic correlations with an **implicit** scheme

It can be shown that the formal solution is given by (Whittle 1963)

$$\eta(x,y,t) = \int_{\mathbb{R}^2} C(x,y,x',y')\,\eta(x',y',0)\,\mathrm{d}x'\mathrm{d}y'$$

where

$$C(x,y,x',y') = C(\widetilde{r}\,) = \frac{2^{2-M}}{(M-2)!}\,\widetilde{r}^{\,M-1}\,K_{M-1}(\widetilde{r}\,)$$

are members of the Whittle-Matérn correlation family, with

$$\widetilde{r} = \sqrt{(\mathsf{x}-\mathsf{x}')^{\mathrm{T}}\,\boldsymbol{A}^{-1}\,(\mathsf{x}-\mathsf{x}')}.$$

$$\boldsymbol{A} = \Delta t\,\boldsymbol{\kappa}$$

$$\boldsymbol{H} = -\,\nabla\,\nabla^{\mathrm{T}} C\big|_{\widetilde{r}=0} = (2M-4)\,\boldsymbol{A}^{-1}.$$

Representing anisotropic correlations with an **implicit** scheme

As with the explicit scheme we can relate $\boldsymbol{L}$ to $\boldsymbol{D}$:

$$\boldsymbol{L} = \Delta t\, \boldsymbol{\kappa} = \boldsymbol{A} = \frac{1}{2M-4}\, \boldsymbol{H}^{-1}$$

or

$$\boxed{\boldsymbol{L} = \frac{1}{2M-4}\, \boldsymbol{D}.}$$

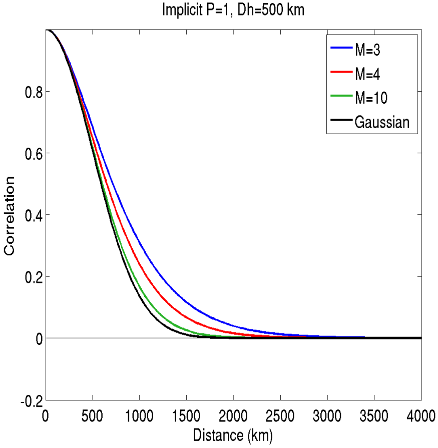
In $\mathbb{R}^d$, the d-dimensional implicit diffusion kernels are

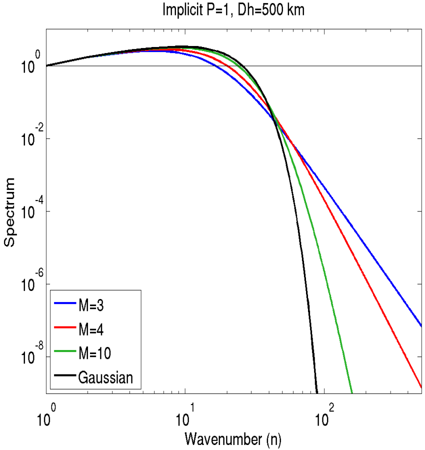
$$C(\widetilde{r}\,) = \frac{2^{1-M+d/2}}{\Gamma(M-d/2)}\, \widetilde{r}^{\,M-d/2}\, K_{M-d/2}(\widetilde{r}\,)$$

and

$$\boxed{\boldsymbol{L} = \frac{1}{2M-d-2}\, \boldsymbol{D}.}$$

Examples of 2D isotropic implicit-diffusion kernels





- A class of anisotropic and inhomogeneous correlation functions from the Matérn family is (Paciorek and Schervish 2006)

$$C(\mathsf{x},\mathsf{x}') = \widetilde{A}(\mathsf{x},\mathsf{x}')\,\frac{2^{1-M+d/2}}{\Gamma(M-d/2)}\,\widetilde{r}^{\,M-d/2}K_{M-d/2}(\widetilde{r}\,)$$

  where

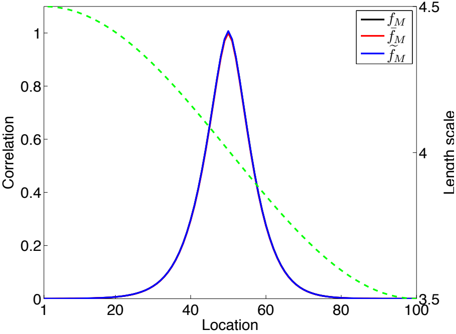
$$\widetilde{r} = \sqrt{(\mathsf{x}-\mathsf{x}')^{\mathrm{T}}\left(\frac{\boldsymbol{L}(\mathsf{x})+\boldsymbol{L}(\mathsf{x}')}{2}\right)^{-1}(\mathsf{x}-\mathsf{x}')}$$

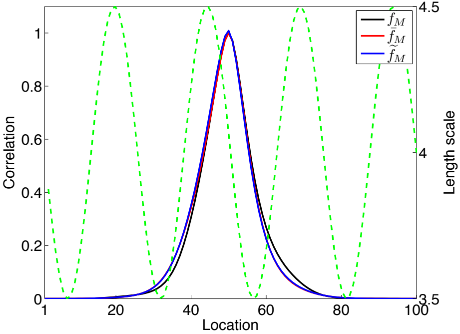
  and

$$\widetilde{A}(\mathsf{x},\mathsf{x}') = |\boldsymbol{L}(\mathsf{x})|^{1/4}\,|\boldsymbol{L}(\mathsf{x}')|^{1/4}\left|\frac{1}{2}\left(\boldsymbol{L}(\mathsf{x})+\boldsymbol{L}(\mathsf{x}')\right)\right|^{-1/2}.$$

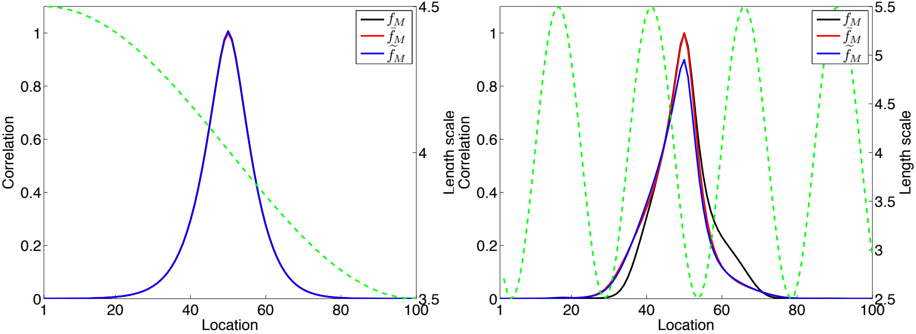
- These are the approximate kernels of the implicit form of the anisotropic diffusion operator when the aspect tensors vary *slowly* and *smoothly* in space.

1D example: inhomogeneous SOAR vs 2-step implicit-diffusion kernel





1D example: inhomogeneous SOAR vs 2-step implicit-diffusion kernel



Outline

Estimating the correlation Hessian tensor from ensemble statistics

- Assume the availability of a sample of simulated model-state errors ε (e.g. ensemble perturbations).
- Apply the inverse of the linearized balance operator to ε: $\mathsf{K}^{-1}\varepsilon = \epsilon$.
- Assume that the *covariance* function of ϵ is twice differentiable and that the *correlation* function $C(\widetilde{r})$ is homogeneous.
- Letting H_{xx}, H_{yy} and H_{xy} be the elements of $\boldsymbol{H} = -\nabla\nabla^{\mathrm{T}} C\big|_{\widetilde{r}=0}$ then it can be shown (e.g. *Belo Pereira and Berre 2006*),

$$
\begin{aligned}
H_{xx} &= \frac{E[(\partial\tilde{\epsilon}/\partial x)^2] - (\partial\sigma/\partial x)^2}{\sigma^2}, \\
H_{yy} &= \frac{E[(\partial\tilde{\epsilon}/\partial y)^2] - (\partial\sigma/\partial y)^2}{\sigma^2}, \\
H_{xy} &= \frac{E[(\partial\tilde{\epsilon}/\partial x)\,(\partial\tilde{\epsilon}/\partial y)] - (\partial\sigma/\partial x)\,(\partial\sigma/\partial y)}{\sigma^2}
\end{aligned}
$$

where $\tilde{\epsilon} = \epsilon - E[\epsilon]$ and $\sigma^2 = E[\tilde{\epsilon}^2]$.

- These formulae will be a good approximation of the Hessian tensor when the correlation function is approximately *locally* homogeneous.
- In compact form, the Hessian tensor estimated at each grid point x from sample statistics is

$$\boldsymbol{H}(\mathsf{x}) = \frac{\overline{\nabla\tilde{\epsilon}(\mathsf{x})\left(\nabla\tilde{\epsilon}(\mathsf{x})\right)^{\mathrm{T}}} - \nabla\widehat{\sigma}(\mathsf{x})\left(\nabla\widehat{\sigma}(\mathsf{x})\right)^{\mathrm{T}}}{\left(\widehat{\sigma}(\mathsf{x})\right)^{2}}$$

where

$$\overline{\nabla\tilde{\epsilon}(\mathsf{x})\left(\nabla\tilde{\epsilon}(\mathsf{x})\right)^{\mathrm{T}}} = \frac{1}{N_{\mathrm{e}}-1}\sum_{l=1}^{N_{\mathrm{e}}}\nabla\tilde{\epsilon}(\mathsf{x})\left(\nabla\tilde{\epsilon}(\mathsf{x})\right)^{\mathrm{T}},$$

$$\left(\widehat{\sigma}(\mathsf{x})\right)^{2} = \overline{\left(\tilde{\epsilon}(\mathsf{x})\right)^{2}} = \frac{1}{N_{\mathrm{e}}-1}\sum_{l=1}^{N_{\mathrm{e}}}\left(\tilde{\epsilon}(\mathsf{x})\right)^{2}.$$

Estimating the Hessian tensor from statistics: remarks

- From the local estimate of $\boldsymbol{H}(\mathsf{x})$, we invert it to obtain $\boldsymbol{D}(\mathsf{x})$, and specify the rescaled (2D) diffusion tensor according to

$$\boldsymbol{L}(\mathsf{x}) = \frac{1}{2M}\boldsymbol{D}(\mathsf{x}) \quad \text{(explicit)} \quad \text{or} \quad \boldsymbol{L}(\mathsf{x}) = \frac{1}{2M-4}\boldsymbol{D}(\mathsf{x}) \quad \text{(implicit)}.$$

- The number of elements to estimate is $3N$ (or $6N$ in 3D), where N is the number of grid points, so sampling errors will be similar to those of the variance estimation problem.
- As for the variance estimation problem, spatial averaging can be used to increase the effective sample size (*Raynaud et al. 2009; Berre and Desroziers 2010*).
- The approach has similarities to the 'hybrid' aspect tensor proposed at NCEP within the context of recursive filters (*Purser et al. 2003*; *Sato et al.* 2009):

$$\boldsymbol{A}_{\mathrm{ani}}^{-1}(\mathsf{x}) = \alpha \boldsymbol{A}_{\mathrm{iso}}^{-1}(\mathsf{x}) + \beta \frac{\overline{\nabla\tilde{\epsilon}(\mathsf{x})\left(\nabla\tilde{\epsilon}(\mathsf{x})\right)^{\mathrm{T}}}}{\left(\widehat{\sigma}(\mathsf{x})\right)^2}.$$

- $\boldsymbol{A}_{\mathrm{ani}}^{-1}(\mathsf{x})$ is equivalent to $\boldsymbol{H}(\mathsf{x})$ when $\alpha = 0$, $\beta = 1$ and $\widehat{\sigma}$ is constant.

Examples from an idealized numerical experiment

- Generate a set of random vectors $\boldsymbol{\epsilon}_l$, $l = 1, \ldots, N_e$, such that

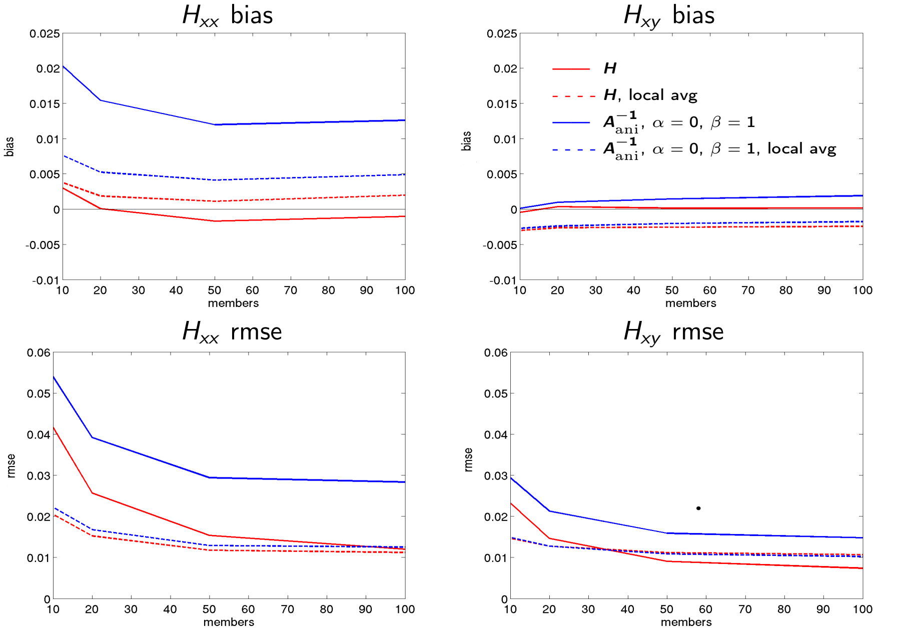
$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathsf{B}^\star)$$

  where $\mathsf{B}^\star$ is the 'true' covariance matrix.
- $\mathsf{B}^\star$ is Gaussian and constructed using a 2D explicit diffusion operator.
- The 'true' variances are spatially varying with a cosine dependence on $\mathsf{x} = (x, y)$.
- The 'true' anisotropic tensor of the diffusion operator is formulated as

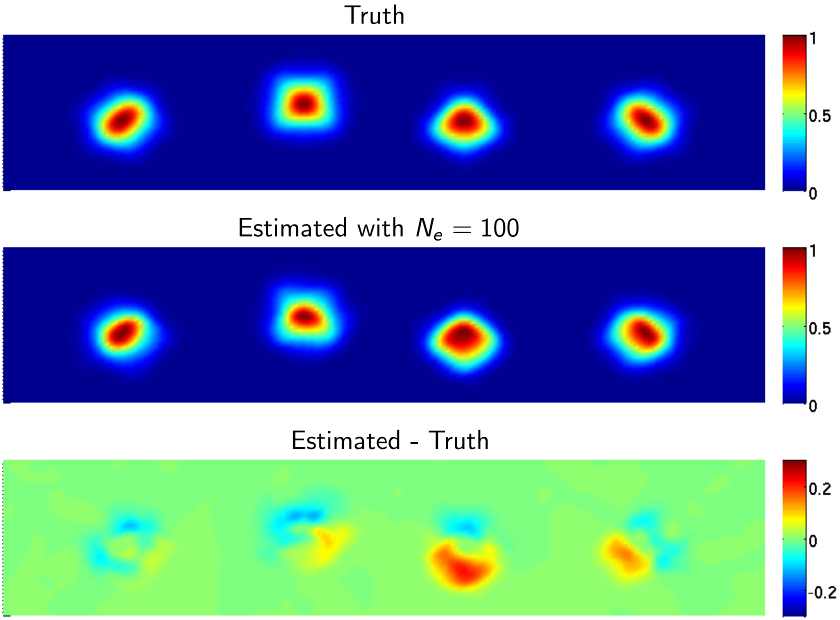
$$\boldsymbol{D}^\star(\mathsf{x}) = \boldsymbol{R}\,\overline{\boldsymbol{D}}(\mathsf{x})\,\boldsymbol{R}^{-1}$$

  where $\boldsymbol{R}$ is a constant rotation matrix and $\overline{\boldsymbol{D}}(\mathsf{x})$ is a diagonal tensor.
- The elements of $\overline{\boldsymbol{D}}(\mathsf{x})$ are spatially varying with a cosine dependence on x.
- The objective here is to try to reconstruct the tensor (and variances) of $\mathsf{B}^\star$ given the 'ensemble' perturbations $\boldsymbol{\epsilon}_l$.

Accuracy of the Hessian tensor elements versus ensemble size

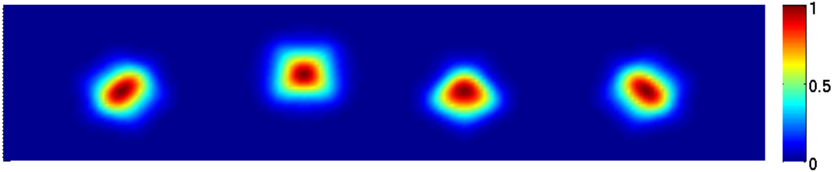


Sample correlations: estimated versus truth

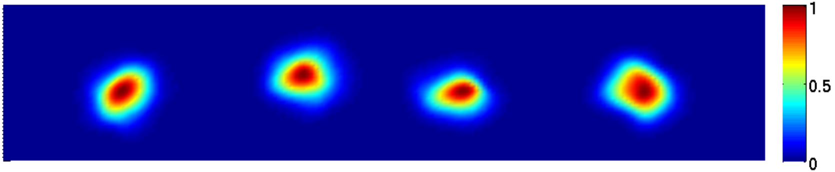


Sample correlations: estimated versus truth

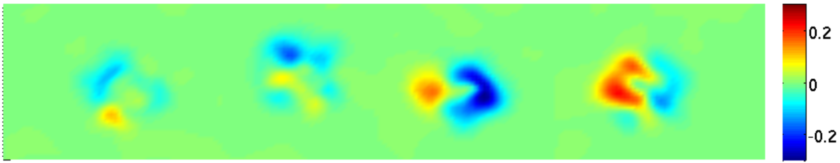
Truth



Estimated with $N_e = 10$ and local averaging



Estimated - Truth



Outline

Summary

- Increasing interest in using ensembles to improve the estimation of **B** in VarDA.
- The diffusion equation can be used to synthesize correlation information contained in an ensemble.
- Choice of explicit or implicit diffusion solver depends on the desired correlation function (Gaussian or Matérn) as well as computational issues.
  - Implicit schemes are more robust with general tensors, but require efficient solvers (CG, multigrid,...).
- The diffusion tensor and variance estimation problems are both $O(N)$.
- Local spatial or temporal averaging is beneficial with small ensemble sizes.
- These techniques are being explored with the NEMOVAR system.

Final remarks

- Negative-lobe or oscillatory correlations can be accounted for using generalized diffusion approaches but new parameters must be introduced and estimated.
- Other applications of the diffusion operator:
  - Grid-point covariance localization in hybrid En-Var.
  - Spatial filtering of ensemble-estimated variance and tensor elements.
  - Spatial filtering of randomized estimates of the normalization factors required by the diffusion-based correlation operator.