

A general hybrid formulation of the background-error covariance matrix for ensemble-variational ocean data assimilation

Anthony Weaver

Andrea Piacentini, Serge Gratton, Selime Gürol, Jean Tshimanga

CERFACS / SUC URA 1875, Toulouse

Acknowledgements:

LEFE-MANU

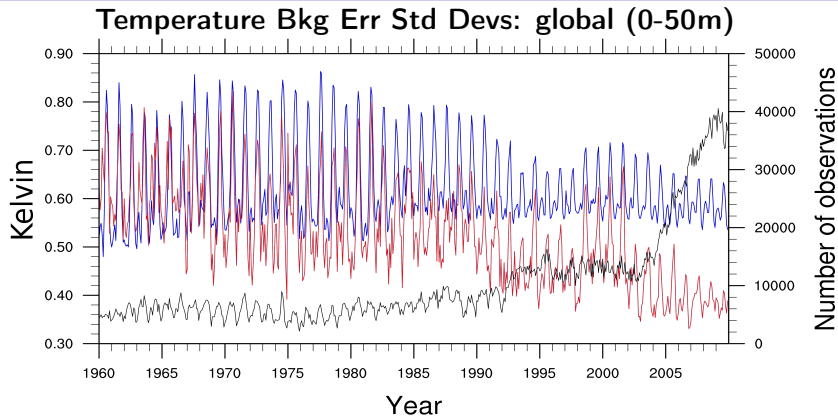
ERA-CLIM2 (FP7)

FILAOS and AVENUE (RTRA)

UK Met Office



- This work contributes to the development of NEMOVAR, a variational ocean data assimilation system for the NEMO model, with a focus on the global ocean.
- NEMOVAR is developed by a consortium consisting of CERFACS, ECMWF, INRIA and Met Office.
- Versions of NEMOVAR are used operationally at ECMWF (Balmaseda *et al.* 2013, QJRMS) and the Met Office (Waters *et al.* 2014, QJRMS).
- The ECMWF system produces an ensemble of ocean analyses but there is currently no coupling of the ensemble with the background error covariance matrix (\mathbf{B}).
- This presentation describes recent extensions to NEMOVAR to exploit ensemble perturbations for specifying a flow-dependent \mathbf{B} .



- Discrepancies between the **specified** and **expected** background-error standard deviations (**blue** and **red** curves, resp.) are a signal of suboptimality in the covariance specifications.
- There is evidence of missing flow dependence in \mathbf{B} , especially in response to the changing observing system (**black curve**).
- An ensemble-based \mathbf{B} should be able to capture this flow dependence.

- We consider the problem of estimating a $N \times N$ background error covariance matrix \mathbf{B} from an ensemble of $N_e \ll N$ error realizations.
- Let $\{\epsilon_p = \mathbf{x}_p - \mathbf{x}_0\}$ denote an ensemble of $p = 1, \dots, N_e$ state error realizations with respect to an unperturbed control member \mathbf{x}_0 .
- If a transformation of control variables is used (e.g., with a balance operator \mathbf{K}_{bal}) then we are interested here in the covariances of the transformed perturbations: $\epsilon_p \rightarrow \mathbf{K}_{\text{bal}}^{-1} \epsilon_p$.
- Let $\epsilon'_p = \epsilon_p - \bar{\epsilon}$ where $\bar{\epsilon} = \frac{1}{N_e} \sum_{p=1}^{N_e} \epsilon_p$, and let

$$\underline{\mathbf{X}}' = \frac{1}{\sqrt{N_e - 1}} \begin{pmatrix} \epsilon'_1 & \dots & \epsilon'_{N_e} \end{pmatrix}$$

- Let $\mathbf{D} = \mathbf{D}^{1/2} \mathbf{D}^{1/2} = \text{diag}(\underline{\mathbf{X}}' \underline{\mathbf{X}}'^T)$ be the matrix of ensemble estimated variances.
- Let $\hat{\underline{\mathbf{X}}}' = \mathbf{D}^{-1/2} \underline{\mathbf{X}}' = \begin{pmatrix} \hat{\epsilon}'_1 & \dots & \hat{\epsilon}'_{N_e} \end{pmatrix}$ be normalized perturbations.

$$\mathbf{B} = \alpha \mathbf{B}_{\text{mod}} + \beta \mathbf{B}_{\text{ens}}$$

where α and β are +ve weighting constants (often such that $\alpha + \beta = 1$).

$$\mathbf{B}_{\text{mod}} = \mathbf{K}_{\text{bal}} \mathbf{D}^{1/2} \mathbf{C}_{\text{mod}} \mathbf{D}^{1/2} \mathbf{K}_{\text{bal}}^T$$

- \mathbf{C}_{mod} is block-diagonal with respect to the transformed variables, where each block is defined by a **parametric correlation model**.
- \mathbf{K}_{bal} is a multivariate **balance operator**.

$$\mathbf{B}_{\text{ens}} = \mathbf{K}_{\text{bal}} \mathbf{D}^{1/2} \underbrace{\left(\underline{\hat{\mathbf{X}}} \underline{\hat{\mathbf{X}}}^T \circ \mathbf{C}_{\text{loc}} \right)}_{\mathbf{C}_{\text{ens}}} \mathbf{D}^{1/2} \mathbf{K}_{\text{bal}}^T$$

- \mathbf{C}_{ens} is a **sample-based correlation matrix** for the transformed variables.
- \mathbf{C}_{loc} is a **localization matrix** acting on the raw sample correlation matrix.

- \mathbf{C}_{loc} is applied as a Schur-product (\circ) to remove spurious long distance correlations from the raw sample estimate ($\mathbf{N} = \mathbf{L} \circ \mathbf{M} \implies N_{ij} = L_{ij} M_{ij}$).

- “Univariate” formulation:

$$\mathbf{C}_{\text{loc}}^{(1)} = \begin{pmatrix} \mathbf{C}_{\text{loc},T} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{C}_{\text{loc},V} \end{pmatrix}$$

- ▶ Localize each variable ξ separately, with possibly different $\mathbf{C}_{\text{loc},\xi}$ for each ξ .

- “Multivariate” formulation:

$$\mathbf{C}_{\text{loc}}^{(2)} = \begin{pmatrix} \mathbf{C}_{\text{loc},T} & \cdots & \mathbf{C}_{\text{loc},T} \mathbf{A}_{V \rightarrow T} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{V \rightarrow T}^T \mathbf{C}_{\text{loc},T} & \cdots & \mathbf{A}_{V \rightarrow T}^T \mathbf{C}_{\text{loc},T} \mathbf{A}_{V \rightarrow T} \end{pmatrix}$$

- ▶ Impose the same localization on the auto- and cross-variable correlations, with the use of an averaging operator $\mathbf{A}_{\xi_1 \rightarrow \xi_2}$ if ξ_1 and ξ_2 are defined at different points on the grid.

- The Schur product for each variable ξ can be rewritten as a sum of matrix operators:

$$\widehat{\mathbf{X}}'_\xi (\widehat{\mathbf{X}}'_\xi)^T \circ \mathbf{C}_{\text{loc},\xi} = \sum_{p=1}^{N_e} \mathbf{\Lambda}_{p,\xi} \mathbf{C}_{\text{loc},\xi} \mathbf{\Lambda}_{p,\xi}$$

- ▶ $\mathbf{\Lambda}_{p,\xi}$ is a diagonal matrix whose diagonal is the p th column of $\widehat{\mathbf{X}}'_\xi$.
- ▶ A computationally efficient filter can be used to define $\mathbf{C}_{\text{loc},\xi}$.
- The multivariate formulation can be conveniently factored as

$$\mathbf{C}_{\text{loc}}^{(2)} = \begin{pmatrix} \mathbf{I}_{N_T} \\ \vdots \\ \mathbf{A}_{v \rightarrow T}^T \end{pmatrix} \mathbf{C}_{\text{loc},T} \begin{pmatrix} \mathbf{I}_{N_T} & \cdots & \mathbf{A}_{v \rightarrow T} \end{pmatrix}$$

- ▶ Only one localization application is required per ensemble member.
- ▶ This reduces the cost of localization, but also the rank of \mathbf{B}_{ens} which is determined by the rank of $\mathbf{C}_{\text{loc},T}$.

- A hybrid \mathbf{B} involves a **sum** of operators:

$$\begin{aligned}\mathbf{B} &= \alpha \mathbf{B}_{\text{mod}} + \beta \mathbf{B}_{\text{ens}} \\ &= \begin{pmatrix} \alpha^{1/2} \mathbf{B}_{\text{mod}}^{1/2} & \beta^{1/2} \mathbf{B}_{\text{ens}}^{1/2} \end{pmatrix} \begin{pmatrix} \alpha^{1/2} \left(\mathbf{B}_{\text{mod}}^{1/2} \right)^{\text{T}} \\ \beta^{1/2} \left(\mathbf{B}_{\text{ens}}^{1/2} \right)^{\text{T}} \end{pmatrix} \\ &\equiv \underline{\mathbf{B}}^{1/2} \left(\underline{\mathbf{B}}^{1/2} \right)^{\text{T}}\end{aligned}$$

- Conventional 3D-Var/4D-Var requires preconditioning the cost function via a control variable transformation involving $\underline{\mathbf{B}}^{1/2}$.
- Large vectors are required in the minimization algorithm since $\underline{\mathbf{B}}^{1/2}$ is rectangular.
- The rectangular nature of $\underline{\mathbf{B}}^{1/2}$ increases with ensemble size due to localization.

- The “square-root” factorization of the localized correlation block corresponding to variable ξ is

$$\sum_{p=1}^{N_e} \Lambda_{p,\xi} \mathbf{C}_{loc,\xi} \Lambda_{p,\xi} = \left(\Lambda_{p,\xi} \mathbf{C}_{loc,\xi}^{1/2} \quad \cdots \quad \Lambda_{p,\xi} \mathbf{C}_{loc,\xi}^{1/2} \right) \begin{pmatrix} \left(\mathbf{C}_{loc,\xi}^{1/2} \right)^T \Lambda_{p,\xi} \\ \vdots \\ \left(\mathbf{C}_{loc,\xi}^{1/2} \right)^T \Lambda_{p,\xi} \end{pmatrix}$$

- This is a product of rectangular matrices of size:
 $N_\xi \times N_\xi = (N_\xi \times N_\xi N_e) \times (N_\xi N_e \times N_\xi)$.
- Preconditioning strategies that involve **B** rather than $\underline{\mathbf{B}}^{1/2}$ are preferable to avoid excessive memory requirements.

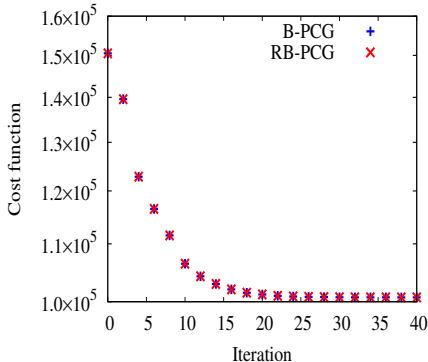
- 1 The possibility of using ensemble perturbations to calibrate parameters of the covariance model (\mathbf{B}_{mod}) and/or to define the localized sample covariance matrix (\mathbf{B}_{ens}).
- 2 The use of \mathbf{B} -preconditioned primal- and dual-space conjugate gradient algorithms (B-PCG and RB-PCG).
- 3 “Univariate” and “multivariate” formulations of the spatial localization operator in \mathbf{B}_{ens} .
- 4 The use of a diffusion operator for defining the correlation model in \mathbf{B}_{mod} and the spatial localization operator in \mathbf{B}_{ens} , and for filtering ensemble-estimated parameters in both \mathbf{B}_{mod} and \mathbf{B}_{ens} .
- 5 Revised numerical algorithms for implicitly formulated diffusion operators.
- 6 An algorithm for determining appropriate filtering and localization scales as a function of ensemble size.

(Gratton and Tshimanga 2009, QJRMS; Gürol *et al.* 2014, QJRMS)

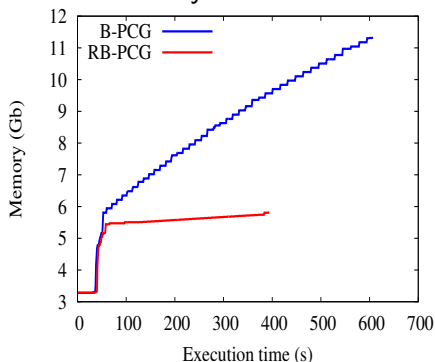
- **B-Preconditioned CG (B-PCG)** and **Restricted B-preconditioned CG (RB-PCG)** produce identical iterates within machine precision.
- Memory and CPU requirements can be significantly less with **RB-PCG** than with **B-PCG**, especially when reorthogonalization is used.

Example from a global ocean 3D-Var experiment with NEMOVAR

Cost function vs iteration



Memory vs wallclock time

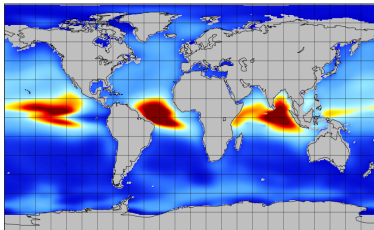


- Sampling error in the parameter estimates is very large with small (typical) ensemble sizes.
- Ensemble-estimated parameters must be filtered before being injected into **B**.

*Temperature variances at 100m estimated from a 10-member pseudo-ensemble (random samples drawn from a Gaussian distribution with specified **B**)*

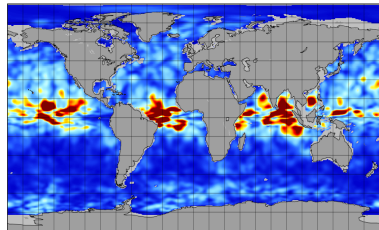
True variances

True var

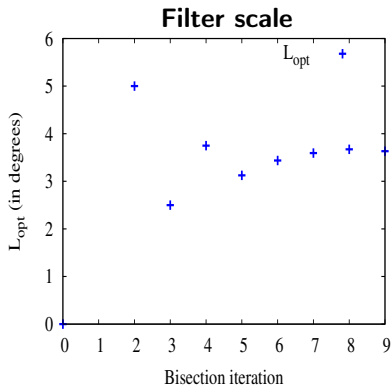
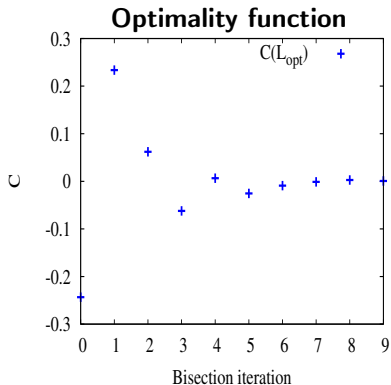


Raw estimate

Raw var sample estimate



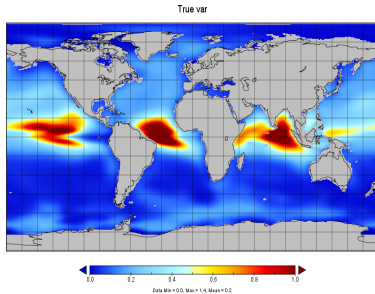
- For ensemble-parameter filtering, we use a diffusion-based method with the filtering scale determined by an algorithm derived from optimal filtering theory (Ménétrier 2014, PhD thesis; Ménétrier *et al.* 2014a,b, MWR).
- The algorithm requires finding the scale (L_{opt}^f) that corresponds to the unique zero-crossing of an optimality function $C(L^f)$ that depends on spatial averages of terms involving the filtered and raw variance.



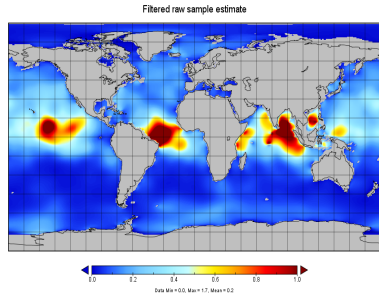
- Remarkably good variance estimates are obtained even with a small ensemble.

Temperature variances at 100m estimated from a 10-member pseudo-ensemble (random samples drawn from a Gaussian distribution with specified \mathbf{B})

True variances



Filtered estimate



- Raw length scale estimates are especially noisy since they are computed from the sample average of products of the spatial *derivatives* of the normalized ensemble perturbations:

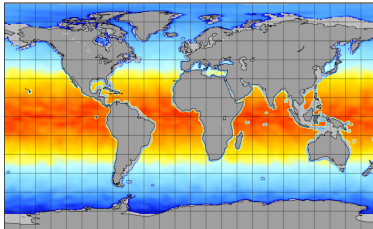
$$\mathbf{H}(\mathbf{x}) = \sum_{p=1}^{N_e} \nabla \hat{\epsilon}'_p(\mathbf{x}) (\nabla \hat{\epsilon}'_p(\mathbf{x}))^T$$

- The diffusion tensor is proportional to $(\mathbf{H}(\mathbf{x}))^{-1}$.

Zonal “diffusion” tensor elements (squared length scales D^2) for temperature at 100 m estimated from a 10-member pseudo-ensemble

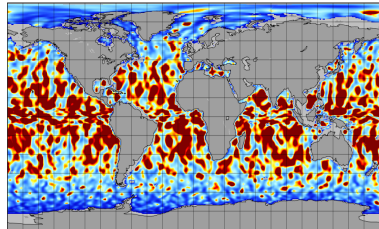
True zonal D^2 estimate

True K11



Raw zonal D^2 estimate

Raw K11 sample estimate

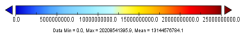
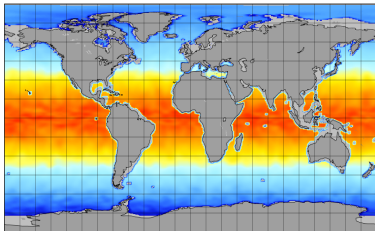


- Useful large-scale information can still be extracted using the diffusion operator with optimized filtering scale.

Zonal “diffusion” tensor elements (squared length scales D^2) for temperature at 100 m estimated from a 10-member pseudo-ensemble

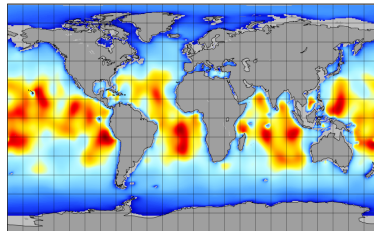
True zonal D^2 estimate

True K11



Filtered zonal D^2 estimate

Filtered K11 sample estimate

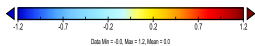
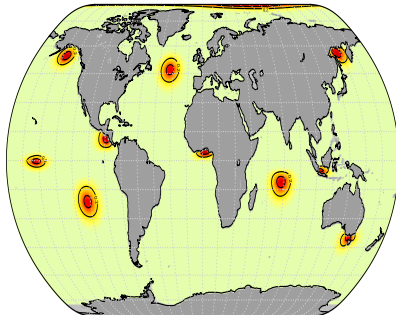


$$\mathbf{B} = \alpha \mathbf{B}_{\text{mod}} + \beta \mathbf{B}_{\text{ens}}$$

T-T covariances at 5 m at selected locations

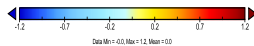
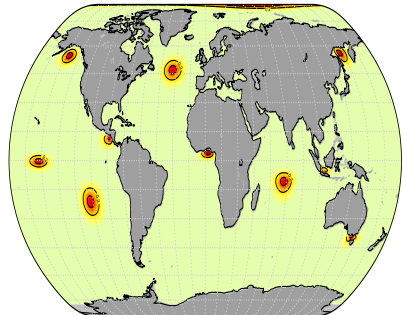
\mathbf{B}_{tru}

T-T covariances at 5m



\mathbf{B} with $\alpha = 1$ and $\beta = 0$

T-T covariance at 5m

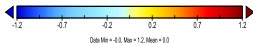
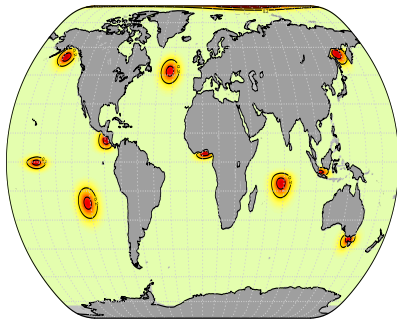


$$\mathbf{B} = \alpha \mathbf{B}_{\text{mod}} + \beta \mathbf{B}_{\text{ens}}$$

- An optimally-based technique involving the spatial average of the correlation Hessian tensor is used to determine the localization scale (B. Ménétrier, PhD thesis).

 \mathbf{B}_{tru}

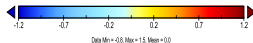
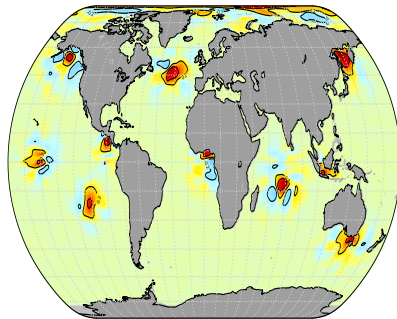
T-T covariances at 5m



Data Min = -0.8, Max = 1.2, Mean = 0.0

 \mathbf{B} with $\alpha = 0$ and $\beta = 1$

T-T covariance at 5m



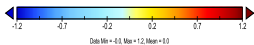
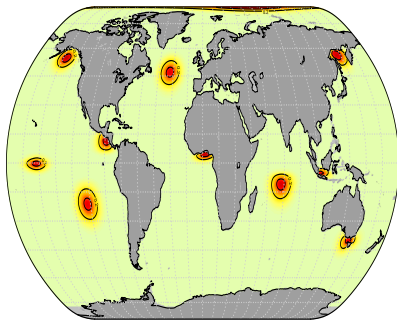
Data Min = -0.8, Max = 1.5, Mean = 0.0

$$\mathbf{B} = \alpha \mathbf{B}_{\text{mod}} + \beta \mathbf{B}_{\text{ens}}$$

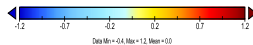
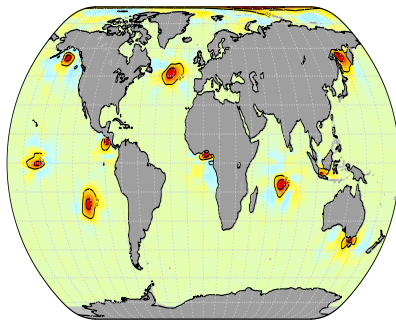
- \mathbf{B}_{mod} is more robust than \mathbf{B}_{ens} , but general anisotropy and complex multivariate relationships are easier to account for in \mathbf{B}_{ens} than \mathbf{B}_{mod} .

 \mathbf{B}_{tru}

T-T covariances at 5m


 \mathbf{B} with $\alpha = \beta = 0.5$

T-T covariance at 5m



(Mirouze and Weaver 2010, QJRMS; Weaver and Mirouze 2013, QJRMS)

- Similarities with correlation models based on the recursive filter (Purser *et al.* 2003) and smoothing norm splines (Wahba and Wendelberger 1982).
- The solution on \mathbb{S}^2 of the elliptic equation

$$(1 - L^2 \nabla^2)^M \psi = \gamma_m \widehat{\psi}(\lambda, \phi)$$

is a correlation operator

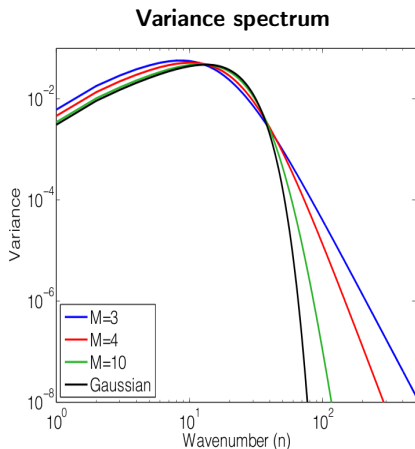
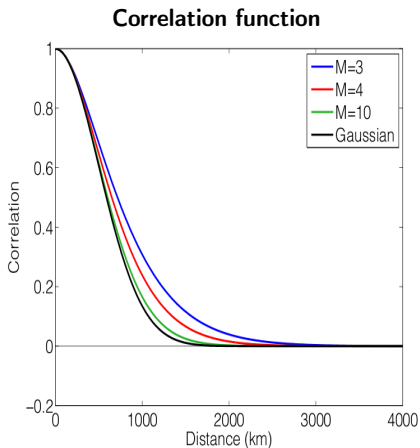
$$\psi(\lambda, \phi) = \int_{\mathbb{S}^2} C(r) \widehat{\psi}(\lambda', \phi') d\Sigma$$

- The kernel is a (SPD) correlation function

$$C(r) \approx C_m(r) \propto \left(\frac{r}{L}\right)^{M-1} K_{M-1}\left(\frac{r}{L}\right)$$

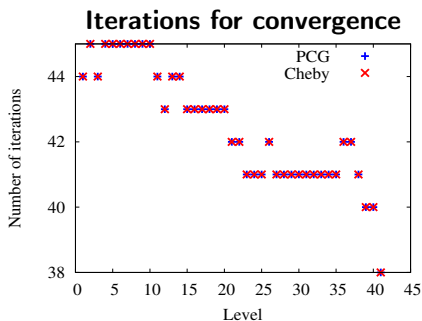
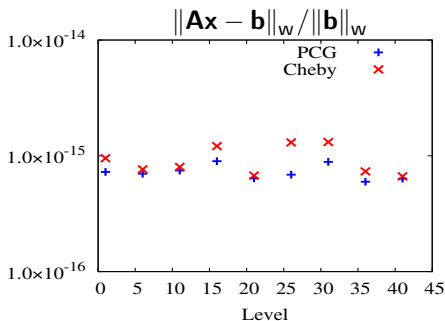
- K_{M-1} is the modified Bessel function of the 2nd kind of order $M - 1$, L is a scale parameter ($D = \sqrt{2M - 4} L$ is the length-scale), and M is a smoothness parameter.
- $C_m(r)$ is from the class of **Matérn functions** (Guttorp and Gneiting 2006).

- Correlation kernels of $(\mathcal{A}^M)^{-1}$ where $\mathcal{A} = 1 - L^2 \nabla^2$ for different M and fixed $D = \sqrt{2M - 4}L = 500$ km.



(Weaver and Mirouze 2013, QJRMS)

- The existing $2 \times 1D$ implicit diffusion (recursive filter-like) algorithm used to represent horizontal correlations in NEMOVAR has poor scalability properties.
- This algorithm has been replaced by *full* 2D implicit diffusion, solved with preconditioned conjugate gradient (PCG) or Chebyshev iteration.
- Our test results show Chebyshev to have similar convergence properties to PCG but since Cheby. does not require global communications (no scalar products!) it is better suited for massively parallel machines.



- A general hybrid \mathbf{B} matrix has been implemented in NEMOVAR with different ways of using ensembles to define covariances.
- A distinguishing feature of the formulation is the use of implicitly formulated diffusion operators for the correlation model, spatial localization operator, and parameter filtering.
- Future directions:
 - ▶ Further work on improving the computational efficiency and scalability of the diffusion algorithm (LEFE-MANU proposal 2015-16).
 - ▶ Further work on ensemble-parameter estimation, optimal filtering and localization (AVENUE).
 - ▶ Combining with 4D-Var (collab. INRIA).
 - ▶ Adaptation to operational systems and evaluation in an Ensemble Data Assimilation framework (collab. ECMWF, Met Office, ERA-CLIM2).