Covariance modelling and minimization for variational ocean data assimilation: developments in ADTAO

A. Weaver[1]
S. Gratton[2], S. Gürol[1,3], I. Mirouze[1,4], A. Moore[5,⋆],
A. Piacentini[1,⋆] and O. Titaud[1,6,⋆]

[1] CERFACS, Toulouse
[2] CERFACS / INPT-IRIT, Toulouse
[3] Now at ECMWF, Reading
[4] Now at Met Office, Exeter
[5] Uni. of California, Santa Cruz
[6] Now at CLS, Toulouse
⋆ Funded through the ADTAO and FILAOS projects (RTRA-STAE)

Outline

1. Context of the work
2. Minimization algorithms with the dual formulation of variational assimilation
3. Diffusion-based correlation operators
4. Conclusions and future directions

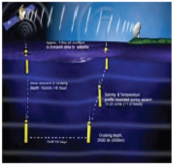
Outline

1. Context of the work
2. Minimization algorithms with the dual formulation of variational assimilation
3. Diffusion-based correlation operators
4. Conclusions and future directions

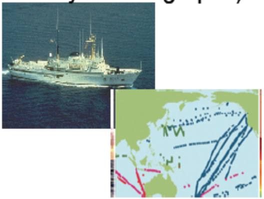
- Ocean data assimilation activities in ADTAO have focused on improving the NEMOVAR system for global applications.
- NEMOVAR is a variational ocean data assimilation system for the NEMO model (Nucleus for European Modelling of the Oceans).
- NEMO is developed by a consortium of European institutes for a variety of reserach and operational applications (regional and global).
- NEMOVAR is developed jointly by CERFACS, ECMWF, UK Met Office and INRIA (Grenoble), with a focus on global applications (Mogensen *et al.* 2009; Mogensen *et al.* 2012; Balmaseda *et al.* 2013).

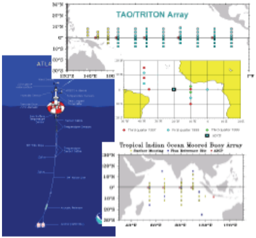
Moorings

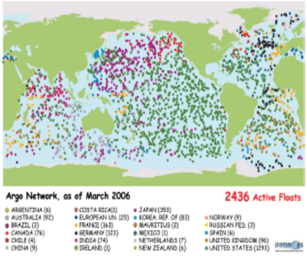
ARGO floats



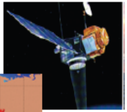
XBTs (eXpandable BathyThermographs)

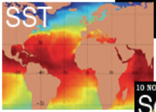


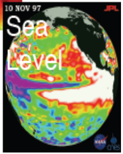




Satellites



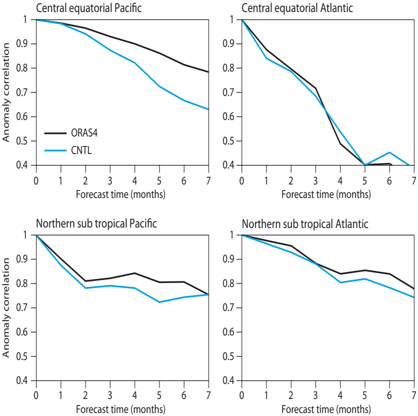




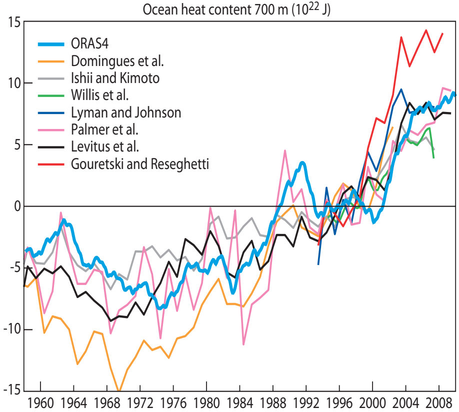
Elephant seals

- Seasonal forecasting and climate ocean reanalysis with the ECMWF Ocean ReAnalysis System (ORAS4) based on NEMOVAR (Balmaseda *et al.* 2013).

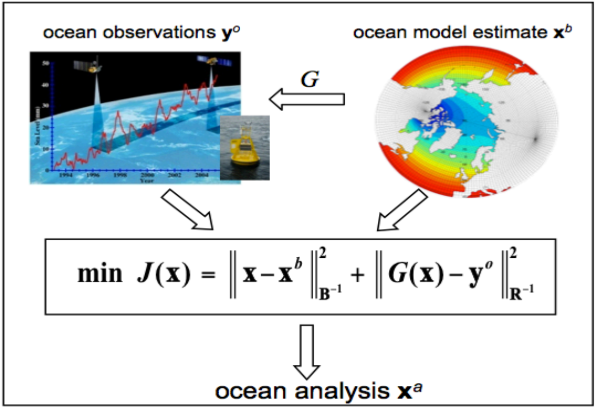
Seasonal (SST) forecast skill: ORAS4 vs Control (no DA) initialization



Time-evolution of global heat content anomalies: ORAS4 vs observational estimates



Developments in ADTAO



- NEMOVAR solves a large-scale nonlinear optimization problem using an outer/inner loop incremental (Truncated Gauss-Newton) algorithm.
- Developments in ADTAO have focused on two areas:
  - Krylov methods (CG and Lanczos) for solving the inner loop.
  - Covariance models for representing errors in **B** and **R**.

Outline

1. Context of the work
2. Minimization algorithms with the dual formulation of variational assimilation
3. Diffusion-based correlation operators
4. Conclusions and future directions

Characteristics of the inner-loop minimization

- Matrices are only available in operator form.
- Matrix-vector products are expensive.
  - Especially with $\mathbf{B}$ in 3D-Var, and $\mathbf{G}$ and $\mathbf{G}^{\mathrm{T}}$ in 4D-Var.
- $\mathbf{B}$ contains a wide range of eigenvalues.
  - First-order preconditioning by $\mathbf{B}$ is important.
- $\mathbf{B}^{-1}$ and $\mathbf{B}^{1/2}$ can be difficult to specify in practice.
  - CG or Lanczos methods requiring only $\mathbf{B}$ are desirable.
- The dimension (P) of observation space is much smaller than the dimension (N) of model-control space.
  - $P \sim O(10^5)$ compared to $N \sim O(10^6)$ or greater.
  - Dual formulations can be advantageous over primal formulations.

Characteristics of the inner-loop minimization

- Matrices are only available in operator form.
- Matrix-vector products are expensive.
  - Especially with $\mathbf{B}$ in 3D-Var, and $\mathbf{G}$ and $\mathbf{G}^{\mathrm{T}}$ in 4D-Var.
- $\mathbf{B}$ contains a wide range of eigenvalues.
  - First-order preconditioning by $\mathbf{B}$ is important.
- $\mathbf{B}^{-1}$ and $\mathbf{B}^{1/2}$ can be difficult to specify in practice.
  - CG or Lanczos methods requiring only $\mathbf{B}$ are desirable.
- The dimension (P) of observation space is much smaller than the dimension (N) of model-control space.
  - $P \sim O(10^5)$ compared to $N \sim O(10^6)$ or greater.
  - Dual formulations can be advantageous over primal formulations.

- The incremental cost function is

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

- The exact solution is $\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \delta \mathbf{x}^{a}$ where

$$\begin{aligned}
\underbrace{\delta \mathbf{x}^{\mathrm{a}}}_{N\times 1} &= \underbrace{\left(\mathbf{B}^{-1} + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\right)^{-1}}_{N\times N}\,\underbrace{\mathbf{G}^{\mathrm{T}}}_{N\times P}\,\underbrace{\mathbf{R}^{-1}}_{P\times P}\,\underbrace{\mathbf{d}}_{P\times 1} && \text{primal formulation}\\
&= \underbrace{\mathbf{B}}_{N\times N}\,\underbrace{\mathbf{G}^{\mathrm{T}}}_{N\times P}\,\underbrace{\left(\mathbf{G}\,\mathbf{B}\,\mathbf{G}^{\mathrm{T}} + \mathbf{R}\right)^{-1}}_{P\times P}\,\underbrace{\mathbf{d}}_{P\times 1} && \text{dual formulation}
\end{aligned}$$

$\mathbf{B}$-preconditioned primal and dual formulations

- Right $\mathbf{B}$-preconditioned **primal** formulation. Solve using CG or Lanczos (e.g., Chan *et al.* 1999)

$$\left(\mathbf{I}_N + \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{G}\,\mathbf{B}\right)\mathbf{z} \;=\; \mathbf{G}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}$$

with respect to the inner product $\mathbf{z}_1^{\mathrm{T}}\,\mathbf{B}\,\mathbf{z}_2$.

- Restricted $\mathbf{B}$-preconditioned **dual** formulation. Solve using CG or Lanczos (Gratton and Tshimanga 2009; Gürol *et al.* 2013)

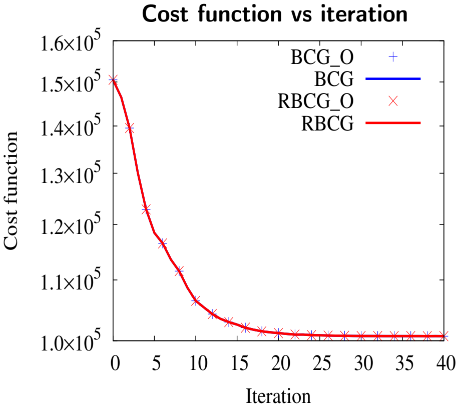
$$\left(\mathbf{R}^{-1}\,\mathbf{G}\,\mathbf{B}\,\mathbf{G}^{\mathrm{T}} + \mathbf{I}_p\right)\mathbf{w} \;=\; \mathbf{R}^{-1}\,\mathbf{d}$$

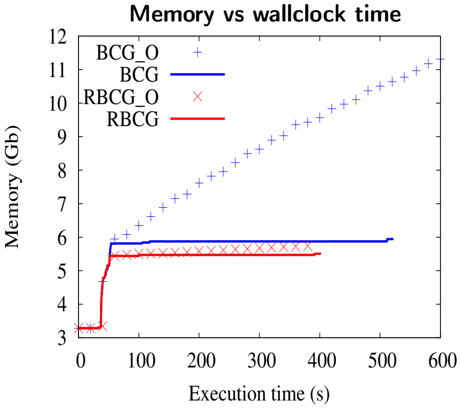
with respect to the inner product $\mathbf{y}_1^{\mathrm{T}}\,\mathbf{G}\,\mathbf{B}\,\mathbf{G}^{\mathrm{T}}\,\mathbf{y}_2$.

- If the first guess is $\mathbf{z} = \mathbf{w} = \mathbf{0}$, these algorithms only require $\mathbf{B}$, not its inverse $\mathbf{B}^{-1}$.
- They require the same number of matrix-vector products with $\mathbf{B}$, $\mathbf{R}^{-1}$, $\mathbf{G}$ and $\mathbf{G}^{\mathrm{T}}$.

Results with NEMO Global Ocean 3D-Var (Gürol *et al.* 2013)

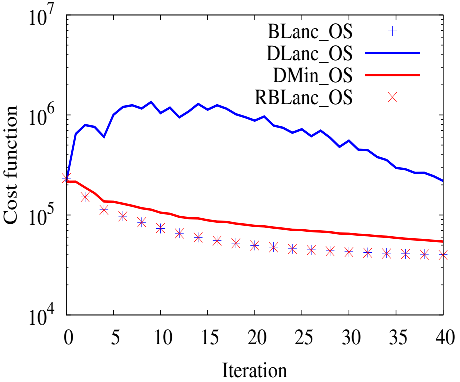
- **B**-preconditioned CG (BCG) and Restricted **B**-preconditioned CG (RBCG) produce identical iterates within machine precision.
- Memory and CPU requirements are significantly less with RBCG than with BCG, especially when reorthogonalization is used (curves labelled "_O").



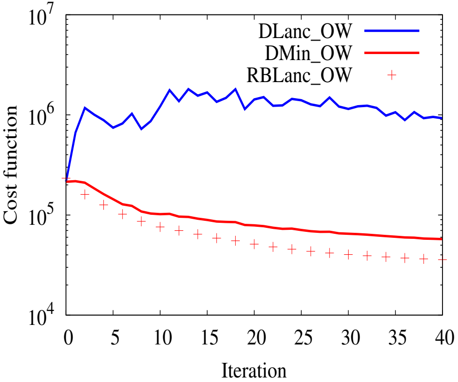


- **B**-preconditioned Lanczos (BLanczos) and Restricted **B**-preconditioned Lanczos (RBLanczos) produce identical iterates within machine precision.
- RBLanczos outperforms two other dual algorithms (dual-Lanczos and dual-MINRES) proposed in the data assimilation literature.

Strong-constraint 4D-Var ($N \sim O(10^6)$)



Weak-constraint 4D-Var ($N \sim O(10^7)$)



Outline

1 Context of the work

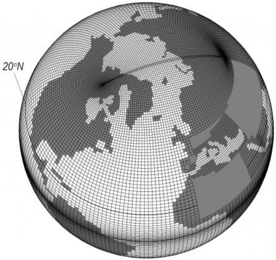
2 Minimization algorithms with the dual formulation of variational assimilation

3 Diffusion-based correlation operators

4 Conclusions and future directions

Characteristics and requirements of the B matrix

- Very large matrix that is difficult to estimate.
- Background errors are inhomogeneous, anisotropic and flow dependent.
- Simplifying assumptions are required to reduce the number of tunable parameters.
- Symmetric and positive-definite operators are required that are computationally efficient on massively parallel machines.
- Appropriate numerical methods are required to handle complex boundaries and complicated curvilinear grids on the sphere.



Diffusion-based correlation operators: theoretical basis

- The solution on $\mathbb{S}^2$ (ignoring solid boundaries) of the elliptic equation

$$\left(1 - L^2\nabla^2\right)^M \psi = \mathcal{A}^M \psi \;=\; \widehat{\psi}$$

is a correlation operator

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

- The kernel is a (SPD) correlation function (Weaver and Mirouze 2013)

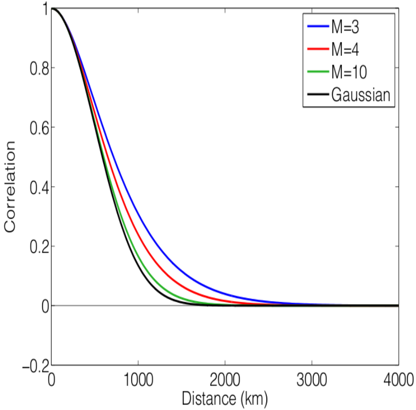
$$C(r) \;\approx\; \widetilde{C}(r) \;\propto\; \left(\frac{r}{L}\right)^{M-1} K_{M-1}\left(\frac{r}{L}\right)$$

- K_{M-1} is the modified Bessel function of the 2nd kind of order $M-1$, r is Euclidean distance, L (M) is a scale (smoothness) parameter.
- $\widetilde{C}(r)$ is from the class of Matérn functions (Guttorp and Gneiting 2006).
- $(\mathcal{A}^M)^{-1}$ can be interpreted as an M-step implicitly-formulated diffusion operator.

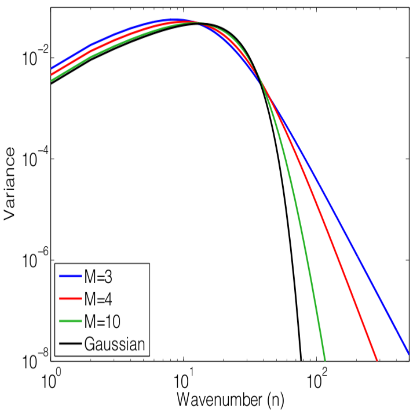
Examples of isotropic implicit-diffusion kernels on $\mathbb{S}^2$

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

Correlation function



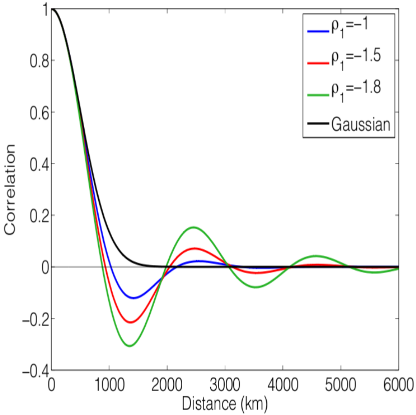
Variance spectrum



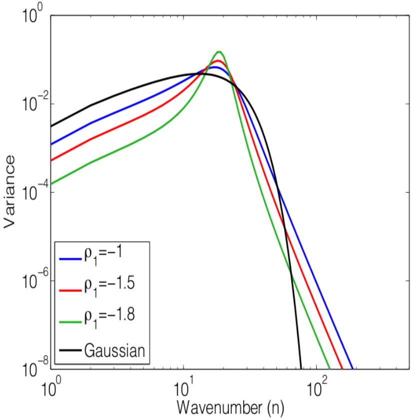
Examples of isotropic implicit-diffusion kernels on $\mathbb{S}^2$

- Correlation kernels of $(\mathcal{A}^M)^{-1}$ where $\mathcal{A} = 1 - \rho_1 L^2 \nabla^2 + L^4 \nabla^4$ for different ρ_1, and fixed $M = 2$ and $D = \sqrt{2M-4}L = 500\,\mathrm{km}$.

Correlation function



Variance spectrum



How to solve the large linear system $\mathbf{A}^M \boldsymbol{\psi} = \widehat{\boldsymbol{\psi}}$?

- Solve as a sequence of linear systems

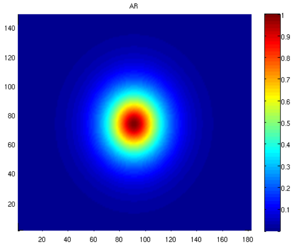
$$\left.\begin{array}{rcl} \mathbf{A}\psi_1 & = & \widehat{\boldsymbol{\psi}} \\ \mathbf{A}\psi_2 & = & \boldsymbol{\psi}_1 \\ & \vdots & \\ \mathbf{A}\psi_M & = & \boldsymbol{\psi}_{M-1} \end{array}\right\}$$

- One possibility is to split the 2D (or 3D) implicit diffusion operator into a self-adjoint product of simpler 1D implicit diffusion operators (Purser *et al.* 2003; Mirouze and Weaver 2010; Mirouze 2010).
- Then use a direct solver (e.g., Cholesky factorization) to solve each of the smaller 1D problems.
- For example, in 2D we can define

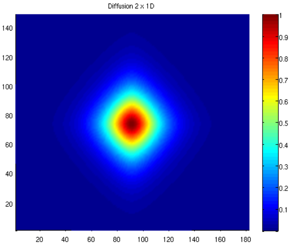
$$\mathbf{A}^M \;\leftarrow\; \mathbf{A}_x^{M/2}\mathbf{A}_y^{M/2}\left(\mathbf{A}_y^*\right)^{M/2}\left(\mathbf{A}_x^*\right)^{M/2}$$

where * denotes adjoint wrt area integration on the model grid.

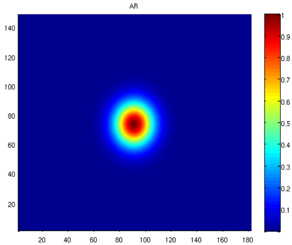
How many iterations should we use?



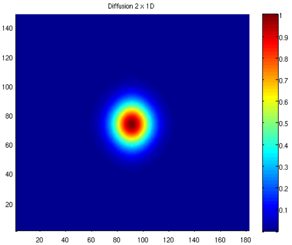
Matérn with $M = 2$



$2{\times}1$D implicit diffusion with $M = 2$

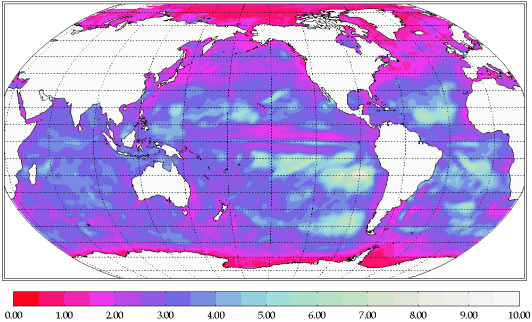


Matérn with $M = 10$

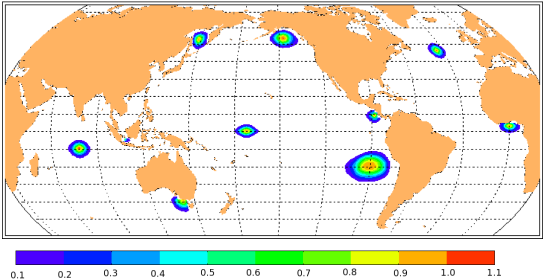


$2{\times}1$D implicit diffusion with $M = 10$

Zonal length scales (degs) for temperature (T) estimated from ensemble perturbations



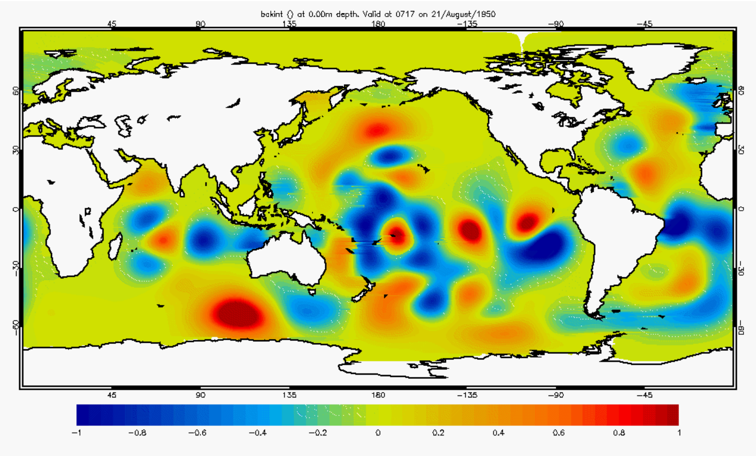
T-T correlations at selected points using 3×1D implicit diffusion



Limitations when length scale is large relative to local geometry

$$\mathbf{A}^M \;\leftarrow\; \mathbf{A}_x^{M/2}\mathbf{A}_y^{M/2}\left(\mathbf{A}_y^*\right)^{M/2}\left(\mathbf{A}_x^*\right)^{M/2}$$

Sea surface temperature analysis increments

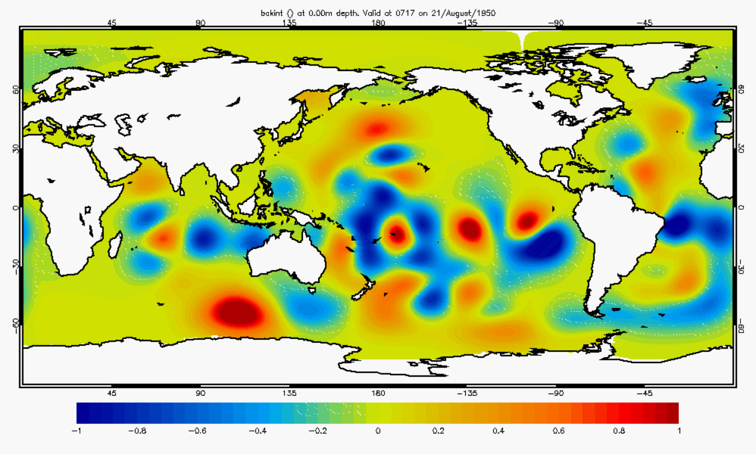


(Courtesy of James While, Met Office)

Impact of alternating the smoothing directions more frequently

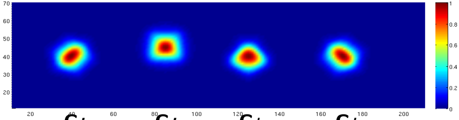
$$\mathbf{A}^M \;\leftarrow\; \left(\mathbf{A}_x\,\mathbf{A}_y\,\mathbf{A}_y^*\,\mathbf{A}_x^*\right)^{M/2}$$

Sea surface temperature analysis increments



(Courtesy of James While, Met Office)

"True" correlations at selected points



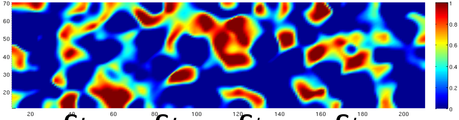
$\mathbf{c}_i \quad \mathbf{c}_j \quad \mathbf{c}_k \quad \mathbf{c}_l$

$$\mathbf{c}_i = \mathbf{C}_{\mathrm{tru}}\,\mathbf{e}_i$$

Sample correlation matrix

$$\mathbf{C}_{\mathrm{sam}} = \overline{(\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}})\,(\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}})^{\mathrm{T}}} \text{ where } \boldsymbol{\epsilon} = \mathbf{C}_{\mathrm{true}}^{1/2}\,\widehat{\boldsymbol{\epsilon}} \text{ and } \overline{\widehat{\boldsymbol{\epsilon}}\,\widehat{\boldsymbol{\epsilon}}^{\mathrm{T}}} \approx \mathbf{I}$$

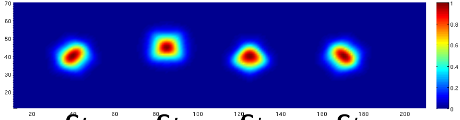
$N_{\mathrm{sam}} = 10$



$\mathbf{c}_i \quad \mathbf{c}_j \quad \mathbf{c}_k \quad \mathbf{c}_l$

$$\mathbf{c}_i = \overline{(\epsilon_i - \overline{\epsilon_i})\,(\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}})}$$

"True" correlations at selected points



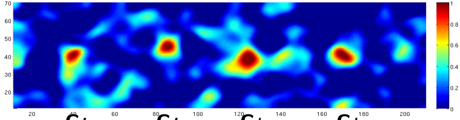
$\mathbf{c}_i$ $\mathbf{c}_j$ $\mathbf{c}_k$ $\mathbf{c}_l$

$$\mathbf{c}_i \;=\; \mathbf{C}_{\rm tru}\,\mathbf{e}_i$$

Sample correlation matrix

$$\mathbf{C}_{\rm sam} \;=\; \overline{(\boldsymbol{\epsilon}-\overline{\boldsymbol{\epsilon}})\,(\boldsymbol{\epsilon}-\overline{\boldsymbol{\epsilon}})^{\rm T}} \text{ where } \quad \boldsymbol{\epsilon} = \mathbf{C}_{\rm true}^{1/2}\,\widehat{\boldsymbol{\epsilon}} \quad \text{and} \quad \overline{\widehat{\boldsymbol{\epsilon}}\,\widehat{\boldsymbol{\epsilon}}^{\rm T}} \approx \mathbf{I}$$

$N_{\rm sam} = 100$

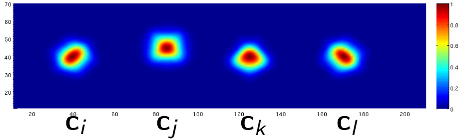


$\mathbf{c}_i$ $\mathbf{c}_j$ $\mathbf{c}_k$ $\mathbf{c}_l$

$$\mathbf{c}_i \;=\; \overline{(\epsilon_i-\overline{\epsilon_i})\,(\boldsymbol{\epsilon}-\overline{\boldsymbol{\epsilon}})}$$

Modelling ensemble correlations via diffusion (Weaver and Mirouze 2013)

"True" correlations at selected points

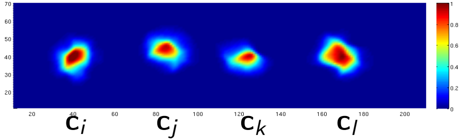


$$\mathbf{c}_i \;=\; \mathbf{C}_{\rm tru}\,\mathbf{e}_i$$

Estimation of the local diffusion tensor $\boldsymbol{\kappa}$

$$\boldsymbol{\kappa}^{-1} \;\propto\; \overline{\nabla\widetilde{\epsilon}\,(\nabla\widetilde{\epsilon})^{\rm T}} \text{ where } \widetilde{\epsilon} = \epsilon/\sigma \quad \text{and } \sigma^2 = \overline{(\epsilon)^2}$$

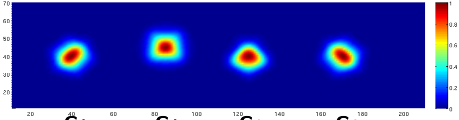
$N_{\rm sam} = 10$



$$\mathbf{c}_i \;=\; \mathbf{C}_{\rm dif}\,\mathbf{e}_i$$

Localizing ensemble correlations via diffusion (Weaver and Piacentini 2013)

"True" correlations at selected points



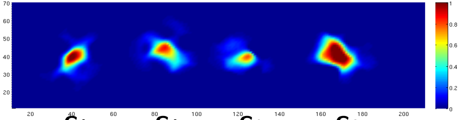
$\mathbf{c}_i$ $\mathbf{c}_j$ $\mathbf{c}_k$ $\mathbf{c}_l$

$$\mathbf{c}_i = \mathbf{C}_{\mathrm{tru}}\,\mathbf{e}_i$$

Compute Schur product of $\mathbf{C}_{\mathrm{sam}}$ with a localized correlation matrix

$$\mathbf{C}_{\mathrm{loc}} = \mathbf{C}_{\mathrm{sam}} \circ \mathbf{C}_{\mathrm{dif}} = \overline{\mathrm{diag}(\widetilde{\boldsymbol{\epsilon}})\,\mathbf{C}_{\mathrm{dif}}\,\mathrm{diag}(\widetilde{\boldsymbol{\epsilon}})}$$

$N_{\mathrm{sam}} = 10$



$\mathbf{c}_i$ $\mathbf{c}_j$ $\mathbf{c}_k$ $\mathbf{c}_l$

$$\mathbf{c}_i = \mathbf{C}_{\mathrm{loc}}\,\mathbf{e}_i$$

Outline

1. Context of the work
2. Minimization algorithms with the dual formulation of variational assimilation
3. Diffusion-based correlation operators
4. Conclusions and future directions

- Demonstration of the practical benefits of dual-formulated CG/Lanczos algorithms (RBCG and RBLanczos) for the inner-loop minimization in two operational ocean variational DA systems (NEMOVAR and ROMS 4D-Var).
- Improved theoretical understanding of diffusion-based correlation models and their applicability to background-error covariance estimation using ensembles.
- Development of diffusion-based correlation models for NEMOVAR using implicit numerical schemes.

- Combine ensemble (En) and variational (Var) assimilation methods with the goals of:
  1. Improving the specification of the background-error covariances;
  2. Improving the efficiency of the assimilation on massively parallel machines.
- Exploit and extend developments in ADTAO for NEMOVAR.
  1. Covariance modelling and localization using diffusion operators.
  2. Minimization algorithms with appropriate preconditioning, with applications to 1 (Gratton *et al.* 2013).
- Applications of EnVar to global ocean reanalysis with NEMOVAR.

Project-related references

- Balmaseda, M. A., Mogensen, K. and A. T. Weaver, 2013: Evaluation of the ECMWF Ocean Reanalysis ORAS4. *Q. J. R. Meteorol. Soc.*, **139**, 1132–1161.
- Gratton, S., Toint, P. L. and J. Tshimanga, 2013: Conjugate gradients versus multigrid solvers for diffusion-based correlation models in data assimilation. *Q. J. R. Meteorol Soc.*. DOI: 10.1002/qj.2050, In press.
- Gratton, S. and J. Tshimanga, 2009: An observation-space formulation of variational assimilation using a Restricted Preconditioned Conjugate-Gradient algorithm. *Q. J. R. Meteorol Soc.*. **135**, 1573–1585.
- Gürol, S., Weaver, A. T., Moore, A. M., Piacentini, A., Arango, H. and S. Gratton, 2013: **B**-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Q. J. R. Meteorol. Soc.*. DOI:10.1002/qj.2150. In press.
- Mirouze, I., 2010: "*Régularisation de problmes inverses à laide de léquation de diffusion généralisé*". PhD thesis, Université Paul Sabatier, September 2010.
- Mirouze, I. and A. T. Weaver, 2010: Representation of correlation functions in variational assimilation using an implicit diffusion operator. *Q. J. R. Meteorol. Soc.*, **136**, 1421–1443.
- Mogensen, K. S., Balmaseda, M. A., Weaver, A. T., Martin, M. and A. Vidard, 2009: 'NEMOVAR: a variational data assimilation system from the NEMO ocean model'. In ECMWF Newsletter 120 - Summer 2009, ECMWF, Reading, U. K., pp 17-21.

Project-related references cont.

- Mogensen, K., M. A. Balmaseda and A. T. Weaver, 2012: The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. ECMWF Tech. Memo., No. 668, 60 pp. Also registered as a CERFACS Technical Report No. TR-CMGC-12-30.
- Weaver, A. T. and I. Mirouze, 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.
- Weaver, A. T. and A. Piacentini, 2012: Representing ensemble covariances with a diffusion operator. Presentation at the International Conference on Ensemble Methods in Geophysical Sciences. Toulouse, France, 12–16 November 2012.

Other references

- Chan, T. F., Chow, E., Saad, Y. and Yeung, M. C., 1999: Preserving symmetry in preconditioned Krylov subspace methods. *SIAM J. Sci. Comput.*, **20**, 568–581.
- Guttorp, P. and T. Gneiting, 2006: Miscellanea studies in the history of probability an statistics XLIX: On the Matérn correlation family. *Biometrika*, **93**, 989–995.
- Purser, R. J., Wu, W. S., Parrish, D. F. and N. M. Roberts, 2003. Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. *Mon. Weather Rev.*, **131**, 1524–1535.