T2.3: Use of ensemble information in ocean analysis and development of efficient 4D-Var

A. Weaver[1], B. Ménétrier[1], J. Tshimanga[1]
and
A. Vidard[2]

[1]CERFACS, Toulouse
[2]INRIA/LJK, Grenoble

December 10, 2015

CERFACS

- NEMOVAR covariance code completely rewritten to facilitate the use of ensembles in defining the background-error covariance matrix (**B**).
- Diffusion-based filtering algorithms have been revised to improve their efficiency and scalability on MPP machines.
- This development has important implications for the ensemble-based **B** in the areas of:
  - correlation modelling
  - ensemble-covariance localization
  - filtering of ensemble-estimated parameters
- This work (not presented here) is described in a recent publication (Weaver, Tshimanga and Piacentini 2015, QJRMS).
- This talk focuses on other aspects of the hybrid **B** that have been developed for NEMOVAR.

- The NEMOVAR $\mathbf{B}$ formulation is quite general:

$$\mathbf{B} \;=\; \beta_{\mathrm{m}}^2 \underbrace{\left(\mathbf{B}_{\mathrm{m_1}} + \mathbf{B}_{\mathrm{m_2}} + \ldots\right)}_{\mathbf{B}_{\mathrm{m}}} + \beta_{\mathrm{e}}^2\, \mathbf{B}_{\mathrm{e}} + \beta_{\mathrm{E}}^2\, \mathbf{B}_{\mathrm{EOF}}$$

wnere β_{m}^2, β_{e}^2 and β_{E}^2 are constant weights or switches.

- Multiple covariance models for representing different "scales" (METO):

$$\mathbf{B}_{\mathrm{m}_i} \;=\; \mathbf{K}_{\mathrm{b}}\, \mathbf{D}_i^{1/2}\, \mathbf{C}_{\mathrm{m}_i}\, \mathbf{D}_i^{1/2}\, \mathbf{K}_{\mathrm{b}}^{\mathrm{T}}$$

- A localized ensemble-based correlation matrix:

$$\mathbf{B}_{\mathrm{e}} \;=\; \mathbf{K}_{\mathrm{b}}\, \mathbf{D}_{\mathrm{e}}^{1/2} \left(\mathbf{L} \circ \widetilde{\mathbf{X}}\, \widetilde{\mathbf{X}}^{\mathrm{T}}\right) \mathbf{D}_{\mathrm{e}}^{1/2}\, \mathbf{K}_{\mathrm{b}}^{\mathrm{T}}$$

where the columns of $\widetilde{\mathbf{X}} = \mathbf{D}_{\mathrm{e}}^{-1/2}\, \mathbf{K}_{\mathrm{b}}^{-1}\, \mathbf{X}$ are (transformed) ensemble perturbations.

- A large-scale EOF-based covariance matrix for assimilating sparse observations (METO):

$$\mathbf{B}_{\mathrm{EOF}} \;=\; \mathbf{P}\, \boldsymbol{\Lambda}\, \mathbf{P}^{\mathrm{T}}$$

Using ensemble perturbations to specify $\mathbf{B}$

CERFACS

- We have developed two ways of defining $\mathbf{B}$ from ensembles:
  1. Estimate variances and correlation scales of the covariance model $\mathbf{B}_{\mathrm{m}}$:
     $\Longrightarrow$ work described at last year's GA.
  2. Through the (Schur product) localized sample covariance matrix

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

     $\Longrightarrow$ new work described here.

- We also consider the **hybrid** variant:

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

  where $\mathbf{B}_{\mathrm{m}}$ employs climatological or modelled parameters.

- How to estimate the localization matrix $\mathbf{L}$ and hybridization weights β_{m} and β_{e}?

Localization and hybridization

Localization by $\mathbf{L}$ + hybridization with $\mathbf{B}_{\mathrm{m}}$ to combat sampling error:

$$\mathbf{B} = \underbrace{\beta_{\mathrm{e}}^2 \, \mathbf{L}}_{\text{Gain } \mathbf{L}^{\mathrm{h}}} \circ \widetilde{\mathbf{B}} + \underbrace{\beta_{\mathrm{m}}^2 \, \mathbf{B}_{\mathrm{m}}}_{\text{Offset}}$$

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

$\mathbf{L}^{\mathrm{h}}$ and β_{m}^2 have to be optimized together

Optimal localization/hybridization minimizes (Ménétrier & Aulígné 2015, MWR)

$$e^{\mathrm{h}} = \mathbb{E}\left[\| \mathbf{L}^{\mathrm{h}} \circ \widetilde{\mathbf{B}} + \beta_{\mathrm{m}}^2 \, \mathbf{B}_{\mathrm{m}} - \widetilde{\mathbf{B}}^{\star} \|_{\mathrm{F}}^2 \right]$$

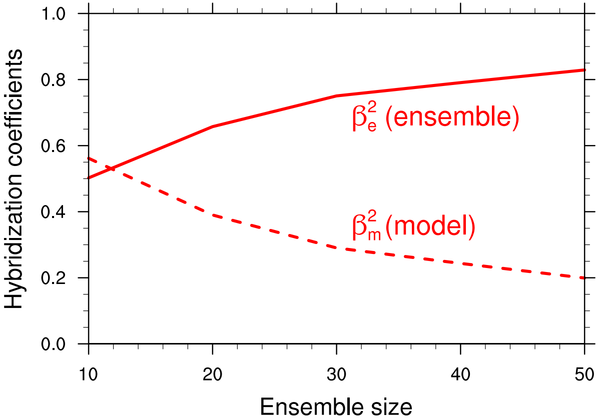
where $\widetilde{\mathbf{B}}^{\star} = \lim_{N_{\mathrm{e}} \to \infty} \widetilde{\mathbf{B}}$ is the target.

It can be shown that, with optimal parameters, whatever the static $\mathbf{B}_{\mathrm{m}}$:
Localization + hybridization is better than localization alone

CERFACS

Practical expressions can be derived for the optimal weights (M & A 2015).

Example from NEMOVAR



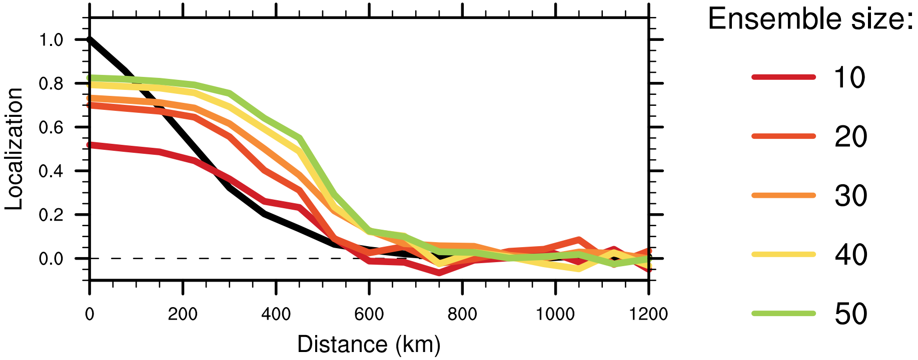
As expected:

- β_e^2 increases with the ensemble size,
- β_m^2 decreases with the ensemble size.

Optimal localization

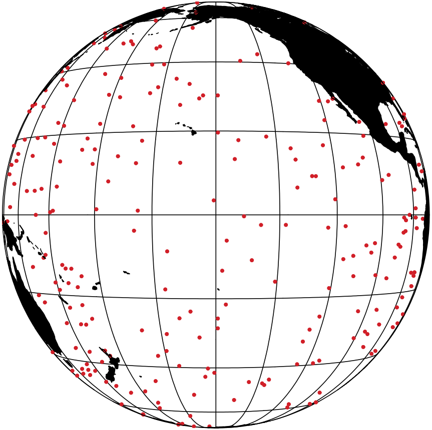
Localization and hybridization are optimized simultaneously.

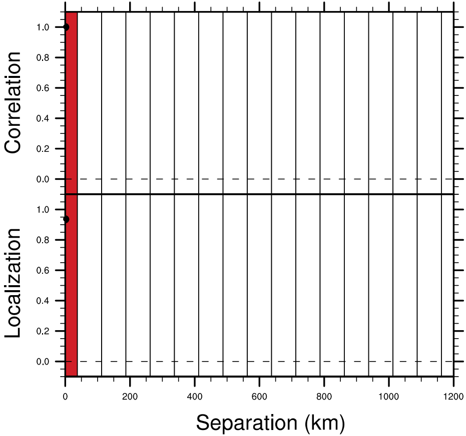
Example from NEMOVAR



Correlation (black) and localization (colors)

A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.

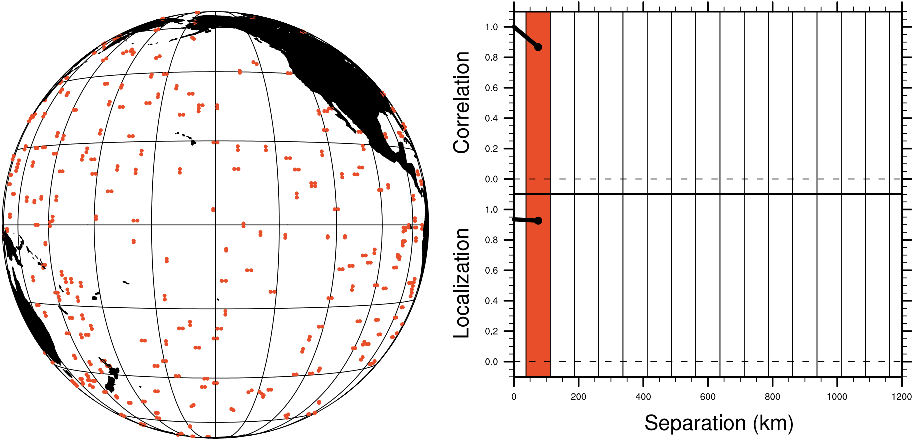




Estimation of correlation and localization (30 members, 5 m temperature)

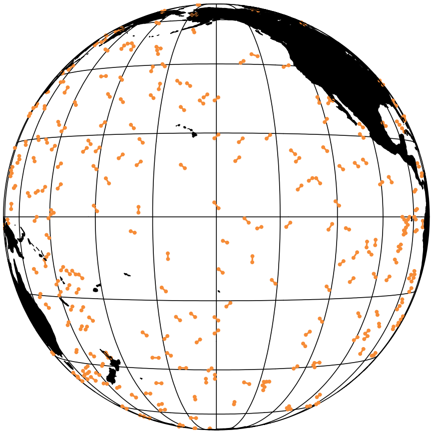
Aspects of the practical computation

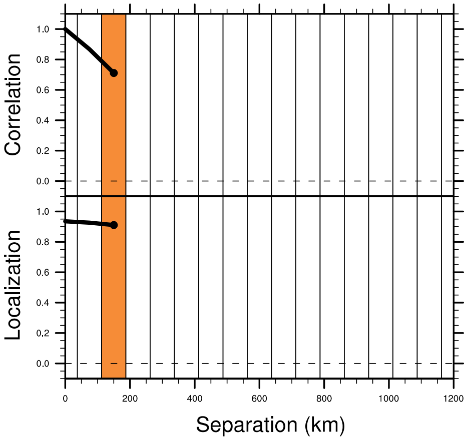
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.



Estimation of correlation and localization (30 members, 5 m temperature)

A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.

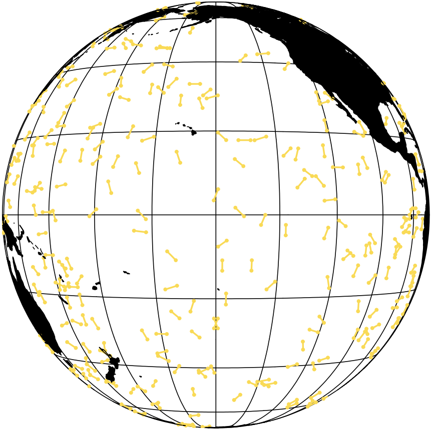


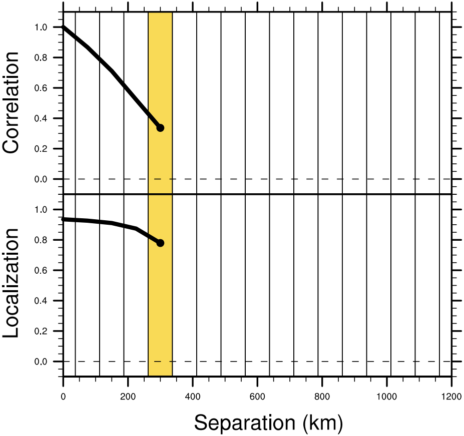


Estimation of correlation and localization (30 members, 5 m temperature)

Aspects of the practical computation

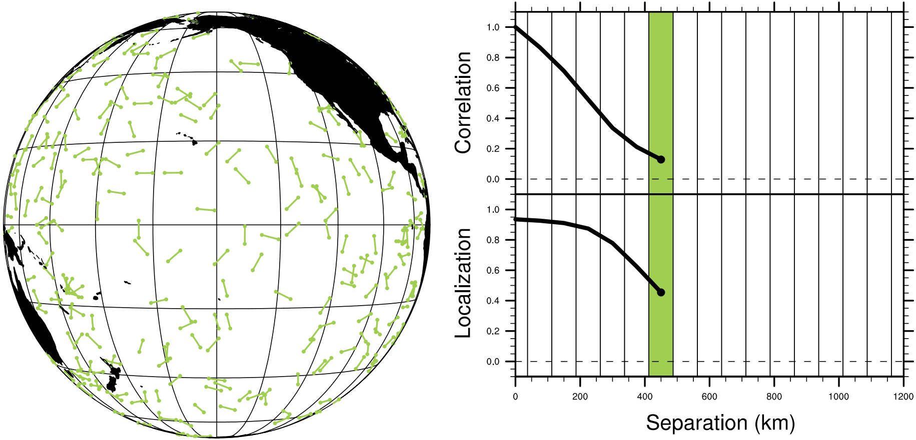
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.





Estimation of correlation and localization (30 members, 5 m temperature)

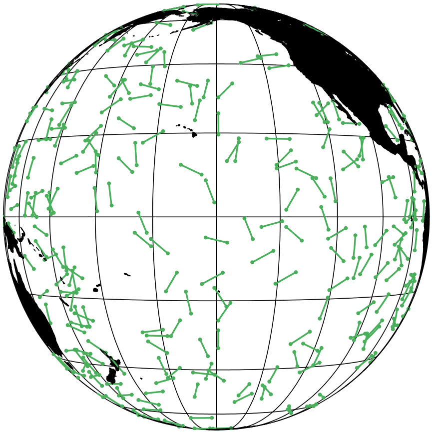
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.

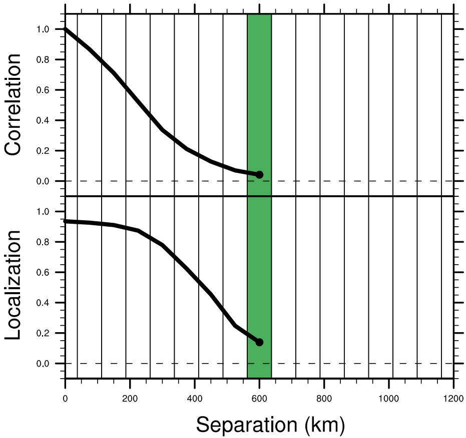


Estimation of correlation and localization (30 members, 5 m temperature)

Aspects of the practical computation

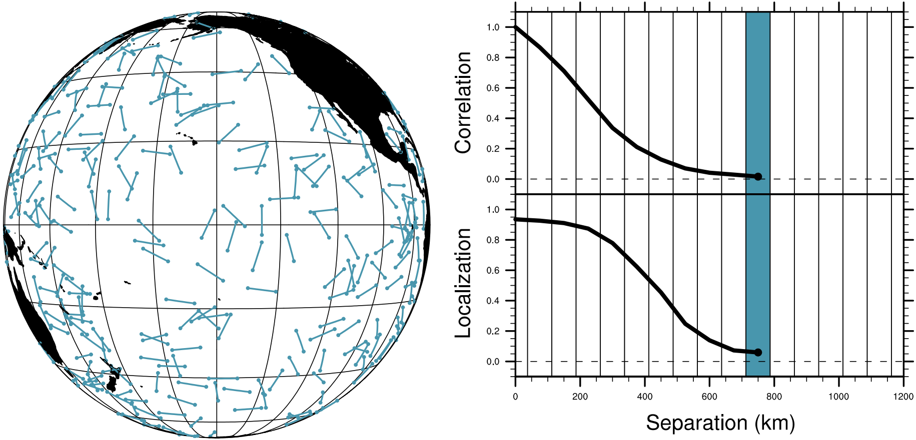
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.





Estimation of correlation and localization (30 members, 5 m temperature)

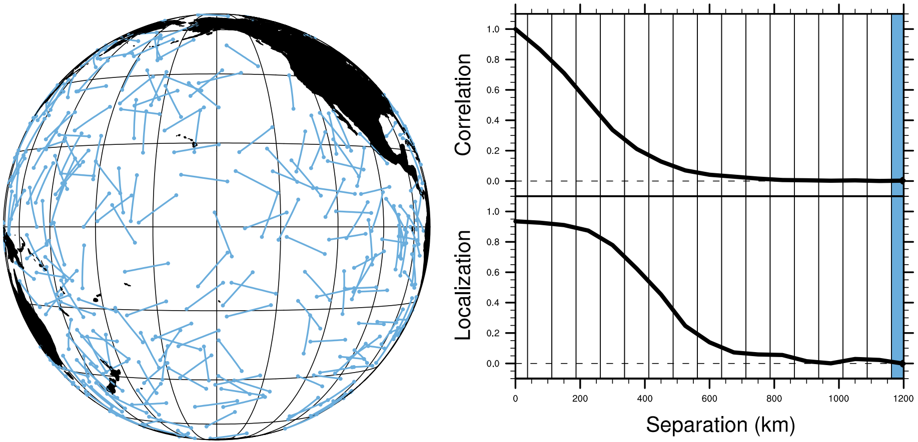
A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.



Estimation of correlation and localization (30 members, 5 m temperature)

Aspects of the practical computation

A spatial ergodicity assumption is required to estimate the statistical expectations $\mathbb{E}[\,\cdot\,]$ in the optimal formulae.

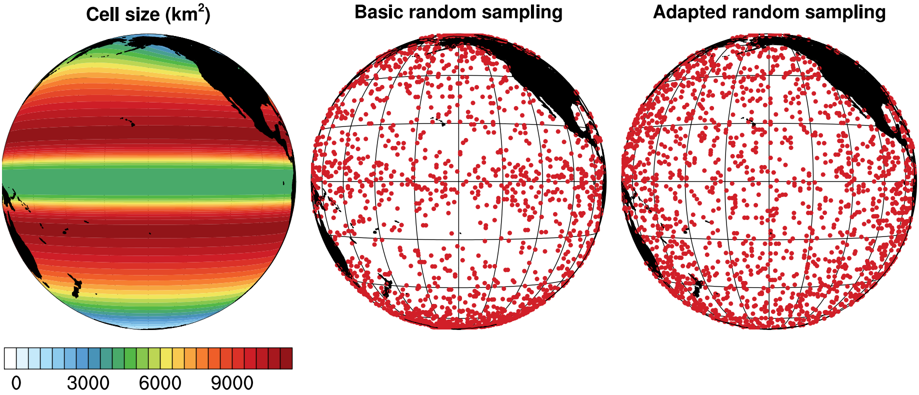


Estimation of correlation and localization (30 members, 5 m temperature)

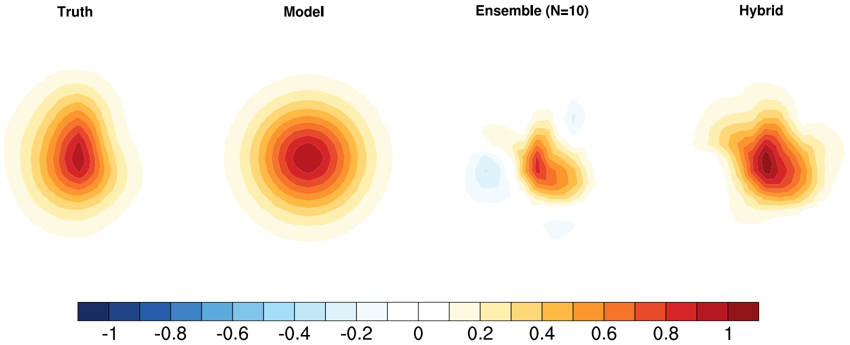
Homogeneity issue

Random points are drawn from the coordinates vector.

This sampling should take the grid structure into account.



Example of surface T-T correlations at a point in the North Atlantic



Fit the localization function to an Mth-order AR-function and model it with a diffusion operator ($\mathbf{L}$).

Optimized parameters: $\beta_{\rm e}^2 = 0.55$, $\beta_{\rm m}^2 = 0.56$, $L_{\rm loc} = 260$ km, $M = 6$.

Summary and plans

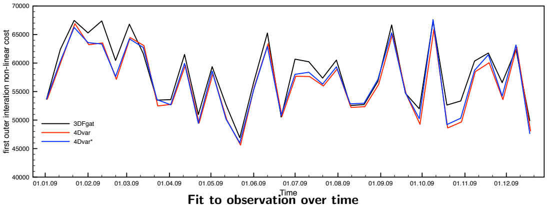
- More tuning and testing of the ensemble-estimation code within a simplified framework (i.e., using an ensemble based on a randomized $\mathbf{B}$).
- The code has been made available to ECMWF (via the Git source code repository at CERFACS).
- Integration of the code within an EDA framework - collaboration ECMWF and Met Office.

4D-Var in the ocean

The first goal of this task was to assess the feasibility and the added value of using 4D-Var in the ocean, compared to the current 3D-Var setting.
Feasibility was demonstrated last year...

Experiment with 1 year 3D-Var/4D-Var uncoupled 10-day window

- Generally smaller increments from 4D-Var compared to 3D-Var (which is good).
- Better fit to observations (which is good as well).
- However the improvement is rather limited despite a significant increase in cost (which is less good).

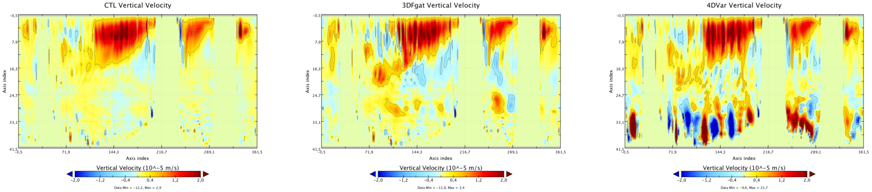


Fit to observation over time

4D-Var in the ocean

Vertical velocities are of importance for coupling with biogeochemical models.

- Assimilation is known to create spurious vertical velocities.
- With smaller increments one could hope that 4D-Var does a better job.
- But in general, the improvement exists but is barely noticeable, except in the equatorial band, where:



Mean analysed vertical velocities at the equator (No assimilation - 3D-Var - 4D-Var)

The strong and nasty signal is contained below 2000m (where there are no data) but it has an impact on sea surface elevation.
Currently under investigation.

The second part of this task is to improve the 4D-Var efficiency.

Plans

- This can be achieve using multigrid techniques (multi-incremental or FAS).
- However the definition of transfer operators (interpolation and simplification) is not trivial in the ocean due to complex boundaries.

So far

- A first version of the transfer operator is available
- On an academic rectangular configuration, multigrid seems quite efficient (4D-Var for the cost of 3D-Var, with the same level of quality)

Transfer operators

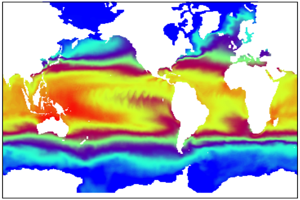
I_f^c may be defined through the general inverse of the interpolation I_c^f

$$\begin{aligned} I_f^c \mathbf{x}_f &= \text{Argmin}\left(\frac{1}{2}[I_c^f \mathbf{x}_c - \mathbf{x}]_f^T \ \mathbf{W} \ [I_c^f \mathbf{x}_c - \mathbf{x}_f]\right) \\ &= [(I_c^f)^T \ \mathbf{W} \ I_c^f]^{-1} (I_c^f)^T \mathbf{W} \mathbf{x}_f \end{aligned}$$

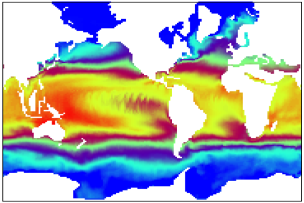
I_c^f being the interpolation operator, and $\mathbf{W}$ the diagonal matrix of volume elements
A cheap and approximate solution may be

$$\mathbf{x}_c \ \approx \ \mathbf{W}_c^{-1} \, (I_c^f)^T \, \mathbf{W} \, \mathbf{x}$$

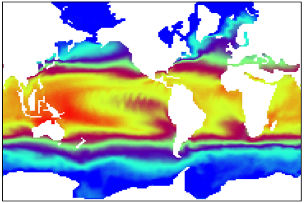
with $\mathbf{W}_c$ is the diagonal matrix of volume elements at low resolution



ORCA 1°



ORCA 2° full



ORCA 2° approx

Plans for the end of the project

Comparison 3D-Var / 4D-Var

- Investigate the issue of spurious vertical velocities deep at the equator
- Redo this comparison at higher resolution ($1/4^\circ$) where 4D-Var may be more beneficial

Multigrid

- First experiments with Orca $1/4^\circ$ grids show that the transfer operator (to/from Orca 1°) itself is a bit too expensive. A second version is on its way.
- Experiments on Orca$1/4^\circ$ / Orca1° and compare Orca$1/4^\circ$ alone