Constraining model biases in a global general circulation model with ensemble data assimilation methods

Canter Martin

GeoHydrodynamic and Environmental Research (GHER)
Department of Astrophysics, Geophysics and Oceanography (AGO)
University of Liège

5th April 2017

Overview

Thesis objective

Most **numerical models** suffer important errors due to poorly represented processes. This leads to a systematic error with a non-zero mean: **bias**.

Bias is considered to be the **main source** of errors in climatic model. It allows one only to study the variation of a model, not its absolute results (Zunz et al., 2013).

Thesis objective

Most **numerical models** suffer important errors due to poorly represented processes. This leads to a systematic error with a non-zero mean: **bias**.

Bias is considered to be the **main source** of errors in climatic model. It allows one only to study the variation of a model, not its absolute results (Zunz et al., 2013).

Objective: Develop a method aiming at correcting and coming closer to numerical model bias.

Model Bias

Bias definition: "It represents an inclination, predisposition or preference, towards a particular result, opinion, or tendency."

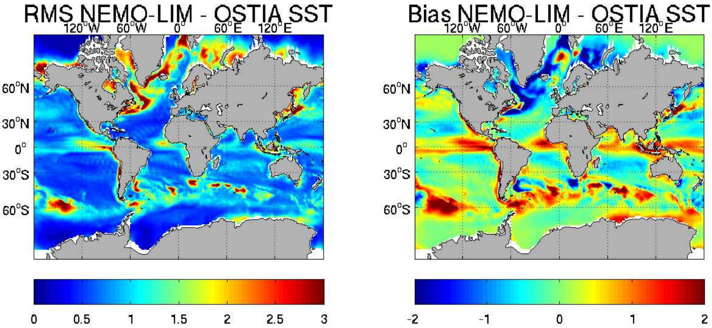
Examples: confirmation bias, cultural bias, media bias, publication bias, statistical bias.

Model Bias

Bias definition: "It represents an inclination, predisposition or preference, towards a particular result, opinion, or tendency."

Examples: confirmation bias, cultural bias, media bias, publication bias, statistical bias.

Numerical model bias



Model Bias

Dee, 2005: **Systematic** error with regard to the notion of the **average** of a model or estimator.

- Spatially variable
- Time dependence
- Depends on regime, situation

Model Bias

Dee, 2005: **Systematic** error with regard to the notion of the **average** of a model or estimator.

- Spatially variable
- Time dependence
- Depends on regime, situation

Origin:

- Poor parametrisation and representation of physical processes
- Bias in boundary and initial conditions
- Bias on observations

Original motivation

The need for a bias correction method arose during the PredAntar Project:

Understanding and predicting Antarctic sea ice variability at the decadal timescale.

Original motivation

The need for a bias correction method arose during the PredAntar Project:

Understanding and predicting Antarctic sea ice variability at the decadal timescale.

Characteristics and requirements of the project:

- Long term simulations
- Low resolution model: NEMO-LIM2
- Large uncertainties on the effects of small scale processes

Original motivation: example with PredAntar

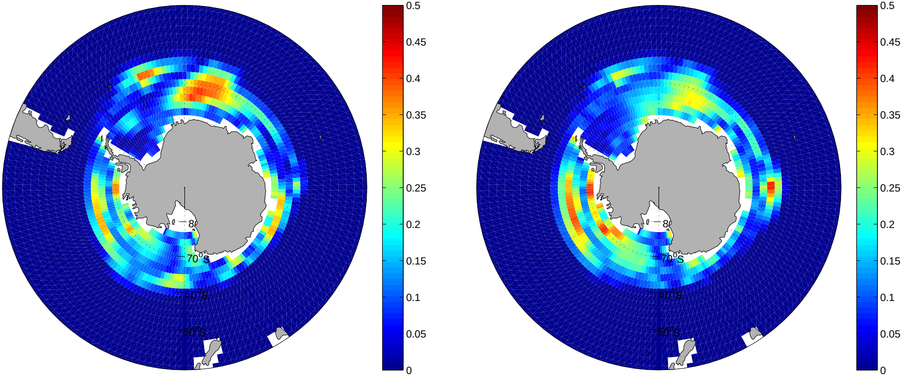
Comparison between 5^{th} Coupled Model Intercomparison Project (CMIP5) using the ORCA2 grid:

- CMCC-CM (Centro Euro-Mediterraneo sui Cambiamenti Climatici - Climate Model)
- CMCC-CMS
- NEMO-LIM2 Free run
- NEMO-LIM2 with data assimilation

and observational data from OSTIA (Operational SST and Sea Ice Analysis).

Original motivation: example with PredAntar

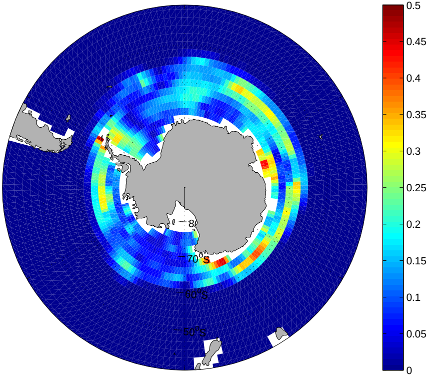
Antarctic ice coverage RMSE (in fraction) for period 1985-2005.



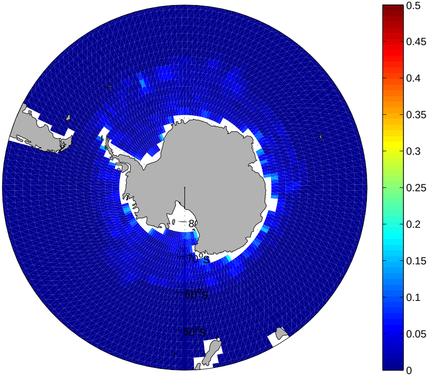
CMCC-CM CMCC-CMS

Original motivation: example with PredAntar

Antarctic ice coverage RMSE (in fraction) for period 1985-2005.



NEMO-LIM2 free run.



NEMO-LIM2 with assimilation of OSTIA observations.

State of the art: What is data assimilation ?

Data assimilation aims at determining the state of a model by combining heterogeneous and imperfect observations with model variables in an optimal way.

$$ \tag{2} $$

State of the art: What is data assimilation ?

Data assimilation aims at determining the state of a model by combining heterogeneous and imperfect observations with model variables in an optimal way.

$$\mathbf{x}_m = M(\mathbf{x}_{m-1}) + \eta_m, \tag{1}$$

(2)

Variables	Descriptions
m	Time index subscript.
$\mathbf{x}, \mathbf{y}$	State vector, observations.
$\mathbf{M}, \mathbf{H}$	Forward model and observation operators.
η, ϵ	model and observational error.

State of the art: What is data assimilation ?

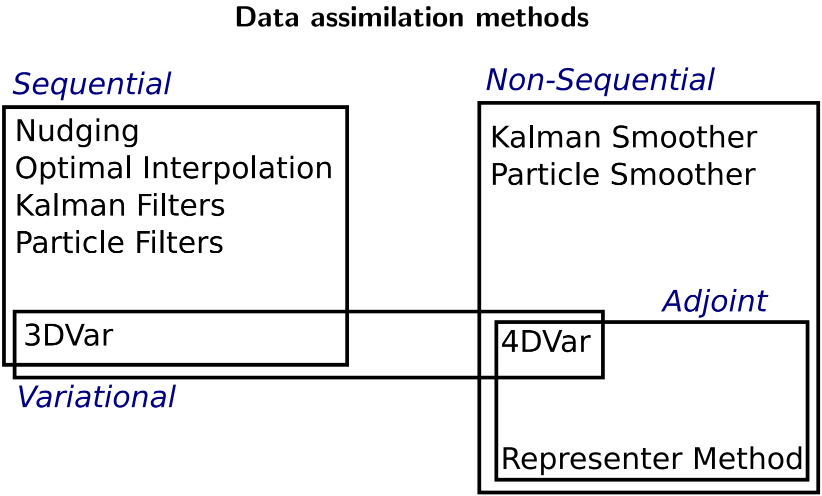
Data assimilation aims at determining the state of a model by combining heterogeneous and imperfect observations with model variables in an optimal way.

$$\mathbf{x}_m = M(\mathbf{x}_{m-1}) + \eta_m, \qquad (1)$$

$$\mathbf{y}_m = H(\mathbf{x}_m) + \epsilon_m, \qquad (2)$$

Variables	Descriptions
m	Time index subscript.
$\mathbf{x}, \mathbf{y}$	State vector, observations.
$\mathbf{M}, \mathbf{H}$	Forward model and observation operators.
η, ϵ	model and observational error.

State of the art: Data Assimilation



Data assimilation methods overview.

State of the art: Original and Extended Kalman Filter

Original Kalman Filter from Kalman (1960), in the Bayesian framework.
Computes background and error covariance matrix at each step.

State of the art: Original and Extended Kalman Filter

Original Kalman Filter from Kalman (1960), in the Bayesian framework.
Computes background and error covariance matrix at each step.

Assumptions:

- Errors are **Gaussian** distributed.
- Model and observation operators are **linear**.

State of the art: Original and Extended Kalman Filter

Original Kalman Filter from Kalman (1960), in the Bayesian framework. Computes background and error covariance matrix at each step.

Assumptions:

- Errors are **Gaussian** distributed.
- Model and observation operators are **linear**.

Extended Kalman Filter: Linearisation of non linear model and observation operators.

State of the art: Original and Extended Kalman Filter

Original Kalman Filter from Kalman (1960), in the Bayesian framework. Computes background and error covariance matrix at each step.

Assumptions:

- Errors are **Gaussian** distributed.
- Model and observation operators are **linear**.

Extended Kalman Filter: Linearisation of non linear model and observation operators.

However, error covariance matrix not computable: Size N_x^2, N_y^2, where $N_x > 10^6$, $N_y > 10^4$ in realistic models.

State of the art: Ensemble Transform Kalman Filter

Representation of a probability density function (PDF) using a sample of that PDF is known as a **Monte Carlo** algorithm.

State of the art: Ensemble Transform Kalman Filter

Representation of a probability density function (PDF) using a sample of that PDF is known as a **Monte Carlo** algorithm.

Ensemble Kalman filter: One aims to avoid using the full covariance matrix, but a representation through an **ensemble** (sample). The assimilation provides a linear combination of ensemble members.

In the linear case and for an infinite ensemble, it converges towards the Kalman filter.

State of the art: Ensemble Transform Kalman Filter

Representation of a probability density function (PDF) using a sample of that PDF is known as a **Monte Carlo** algorithm.

Ensemble Kalman filter: One aims to avoid using the full covariance matrix, but a representation through an **ensemble** (sample). The assimilation provides a linear combination of ensemble members.

In the linear case and for an infinite ensemble, it converges towards the Kalman filter.

Ensemble Transform Kalman filter: No perturbations on the observations.

State of the art: Bias correction

Data assimilation schemes:

- **Bias blind**: Ignores bias on the observations and in the model background estimate.
- **Bias aware**: Estimate model state and bias. Bias is isolated from other state vector variables.

State of the art: Bias correction

Data assimilation schemes:

- **Bias blind**: Ignores bias on the observations and in the model background estimate.
- **Bias aware**: Estimate model state and bias. Bias is isolated from other state vector variables.

Bias correction approaches:

- **Offline**: Bias estimated beforehand, from preliminary model run
- **Online**: Bias estimated and updated during the assimilation scheme.

State of the art: Bias correction

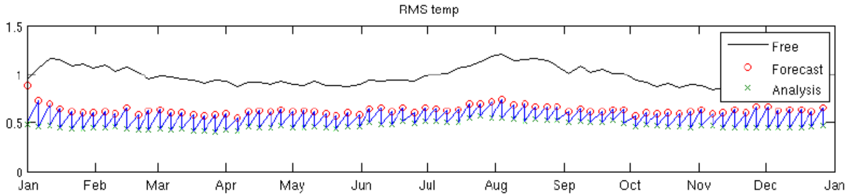
Data assimilation schemes:

- **Bias blind**: Ignores bias on the observations and in the model background estimate.
- **Bias aware**: Estimate model state and bias. Bias is isolated from other state vector variables.

Bias correction approaches:

- **Offline**: Bias estimated beforehand, from preliminary model run
- **Online**: Bias estimated and updated during the assimilation scheme.

Data assimilation with numerical model bias



State of the art: Bias correction

Dee and Da Silva (1998): **Online** bias correction in **sequential** data assimilation, through a separate assimilation step.

State of the art: Bias correction

Dee and Da Silva (1998): **Online** bias correction in **sequential** data assimilation, through a separate assimilation step.

Requirements for bias correction:

- Reference and **unbiased** data set from which an estimation can be provided.
- Bias model or characterisation with parameters.

State of the art: Bias correction

Dee and Da Silva (1998): **Online** bias correction in **sequential** data assimilation, through a separate assimilation step.

Requirements for bias correction:

- Reference and **unbiased** data set from which an estimation can be provided.
- Bias model or characterisation with parameters.

Crucial, for that an erroneous correction deteriorates the model even more than bias blind assimilation.

Bias correction: A new method

New bias correction method objective: **Estimate** a bias correction term through assimilation and **rerun** the corrected model.

Bias correction: A new method

New bias correction method objective: **Estimate** a bias correction term through assimilation and **rerun** the corrected model.

Notation of numerical model bias:

$$\mathbf{x}_m = M_m(\mathbf{x}_{m-1}), \tag{3}$$

(5)

Bias correction: A new method

New bias correction method objective: **Estimate** a bias correction term through assimilation and **rerun** the corrected model.

Notation of numerical model bias:

$$\mathbf{x}_m = M_m\left(\mathbf{x}_{m-1}\right), \tag{3}$$

$$\mathbf{x}_m^t = M_m\left(\mathbf{x}_{m-1}^t\right) + \boldsymbol{\beta}_m, \tag{4}$$

$$\tag{5}$$

Bias correction: A new method

New bias correction method objective: **Estimate** a bias correction term through assimilation and **rerun** the corrected model.

Notation of numerical model bias:

$$\mathbf{x}_m = M_m\left(\mathbf{x}_{m-1}\right), \tag{3}$$

$$\mathbf{x}_m^t = M_m\left(\mathbf{x}_{m-1}^t\right) + \boldsymbol{\beta}_m, \tag{4}$$

$$\boldsymbol{\beta}_m = \widetilde{\boldsymbol{\beta}}_m + \mathbf{b}. \tag{5}$$

Bias correction: A new method

New bias correction method objective: **Estimate** a bias correction term through assimilation and **rerun** the corrected model.

Notation of numerical model bias:

$$\mathbf{x}_m = M_m\left(\mathbf{x}_{m-1}\right), \tag{3}$$

$$\mathbf{x}_m^t = M_m\left(\mathbf{x}_{m-1}^t\right) + \boldsymbol{\beta}_m, \tag{4}$$

$$\boldsymbol{\beta}_m = \widetilde{\boldsymbol{\beta}}_m + \mathbf{b}. \tag{5}$$

One wants to correct the bias term **b** at every time step.

Bias correction: A new method

Start from the free and biased model trajectory: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{m_{\max}} \end{bmatrix}$.

Bias correction: A new method

Start from the free and biased model trajectory: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{m_{\max}} \end{bmatrix}$.

Generate an ensemble of bias correction estimators: $\widehat{\mathbf{b}}^{(i)}$.

Variables	Descriptions
i	Ensemble index superscript.

Bias correction: A new method

Start from the free and biased model trajectory: $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_{m_{\max}} \end{bmatrix}$.

Generate an ensemble of bias correction estimators: $\widehat{\mathbf{b}}^{(i)}$.

Run that ensemble to obtain an ensemble of forced runs: $\mathbf{x}^{(i)}$.

Variables	Descriptions
i	Ensemble index superscript.

Bias correction: A new method

State vector **augmentation** of the model trajectory with the bias correction estimator $\widehat{\mathbf{b}}^{(i)}$, **ensemble mean**, and assimilation scheme:

$$\mathbf{x}'^{(i)} = \begin{bmatrix} \mathbf{x}_1^{(i)} \\ \mathbf{x}_2^{(i)} \\ \vdots \\ \mathbf{x}_{m_{\max}}^{(i)} \\ \widehat{\mathbf{b}}^{(i)} \end{bmatrix}, \quad (6) \tag{9}$$

Bias correction: A new method

State vector **augmentation** of the model trajectory with the bias correction estimator $\widehat{\mathbf{b}}^{(i)}$, **ensemble mean**, and assimilation scheme:

$$\mathbf{x}'^{(i)} = \begin{bmatrix} \mathbf{x}_1^{(i)} \\ \mathbf{x}_2^{(i)} \\ \vdots \\ \mathbf{x}_{m_{\max}}^{(i)} \\ \widehat{\mathbf{b}}^{(i)} \end{bmatrix}, \qquad (6)$$

$$\mathbf{x}' = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}'^{(i)}, \qquad (7)$$

$$(9)$$

Bias correction: A new method

State vector **augmentation** of the model trajectory with the bias correction estimator $\widehat{\mathbf{b}}^{(i)}$, **ensemble mean**, and assimilation scheme:

$$\mathbf{x}'^{(i)} = \begin{bmatrix} \mathbf{x}_1^{(i)} \\ \mathbf{x}_2^{(i)} \\ \vdots \\ \mathbf{x}_{m_{\max}}^{(i)} \\ \widehat{\mathbf{b}}^{(i)} \end{bmatrix}, \quad (6)$$

$$\mathbf{x}' = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}'^{(i)}, \quad (7)$$

$$\mathbf{K}' = \mathbf{P}'^f \mathbf{H}'^T (\mathbf{H}'\mathbf{P}'^f \mathbf{H}'^T + \mathbf{R})^{-1}, \quad (8)$$

$$\mathbf{x}'^a = \mathbf{x}'^f + \mathbf{K}' \left(\mathbf{y}^o - \mathbf{H}'\mathbf{x}'^f \right). \quad (9)$$

Variables	Descriptions
^{f} ^{a}	Forecast and analysis superscripts.
$\mathbf{P}$	Covariance matrix of state vector.
$\mathbf{R}$	Observation error covariance matrix.
$\mathbf{K}$	Kalman gain.
N_e	Ensemble size.

Bias correction: Model time average

Observations are already time averages. To limit the model trajectory and state vector size, one can take the **time average** as follow

$$\mathbf{H}'\mathbf{x}' = \frac{1}{m_{max}} \sum_{m=1}^{m_{max}} \mathbf{H}\mathbf{x}_m = \mathbf{H}\overline{\mathbf{x}}. \tag{10}$$

Bias correction: Model time average

Observations are already time averages. To limit the model trajectory and state vector size, one can take the **time average** as follow

$$\mathbf{H}'\mathbf{x}' = \frac{1}{m_{max}} \sum_{m=1}^{m_{max}} \mathbf{H}\mathbf{x}_m = \mathbf{H}\overline{\mathbf{x}}. \tag{10}$$

Augmented state vector and observation operator become

$$\mathbf{x}'' = \begin{bmatrix} \overline{\mathbf{x}} \\ \widehat{\mathbf{b}} \end{bmatrix}, \tag{11}$$

$$\mathbf{H}''\mathbf{x}'' = \mathbf{H}\overline{\mathbf{x}}. \tag{12}$$

Bias correction: Rerun

Analysis with the average model state is equivalent to full trajectory, and is expressed as

$$\mathbf{x}''^{a} = \mathbf{x}''^{f} + \mathbf{K}'' \left(\mathbf{y}^{o} - \mathbf{H}''\mathbf{x}''^{f} \right) . \tag{13}$$

Bias correction: Rerun

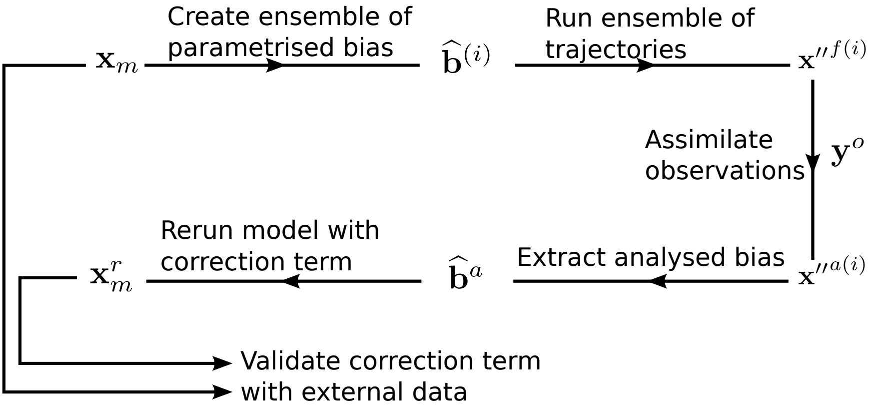
Analysis with the average model state is equivalent to full trajectory, and is expressed as

$$\mathbf{x}''^{a} = \mathbf{x}''^{f} + \mathbf{K}'' \left(\mathbf{y}^{o} - \mathbf{H}''\mathbf{x}''^{f} \right). \tag{13}$$

The bias correction term $\widehat{\mathbf{b}}^{a}$ is provided by the analysis $\mathbf{x}''^{a}$. One obtains the **corrected model rerun** with

$$\mathbf{x}_{m}^{r} = M_{m}\left(\mathbf{x}_{m-1}^{r}\right) + \widehat{\mathbf{b}}^{a}. \tag{14}$$

Bias correction: Method Schematic



General schematic of the bias correction method, from the initial model run $\mathbf{x}_m$ to the corrected model run $\mathbf{x}_m^r$.

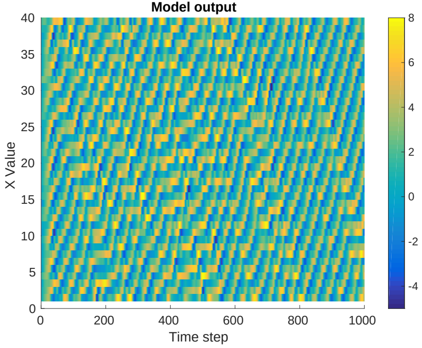
Lorenz '96 model: introduction

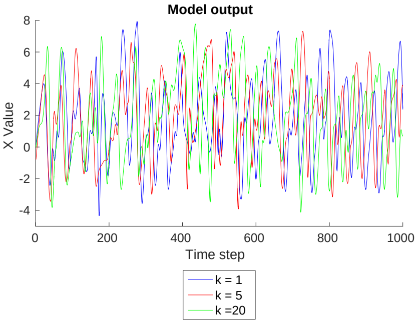
Formulated in 1996 ($K = 40$ variables), exhibits advection, diffusion, periodicity, and chaotic behaviour. With $k = 1, \ldots, K$, it is described by

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F. \tag{15}$$

Formulated in 1996 ($K = 40$ variables), exhibits advection, diffusion, periodicity, and chaotic behaviour. With $k = 1, \dots, K$, it is described by

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F. \tag{15}$$





Lorenz '96: modification

Modification of the forcing parameter by

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F_k. \tag{16}$$

Lorenz '96: modification

Modification of the forcing parameter by

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F_k. \tag{16}$$

Creates a spatially variable constant in time forcing term F_k with a mean **F** and a perturbation depending on covariance **P** defined by

$$P_{i,j} = 0.3\ e^{\frac{-(i-j)^2}{15}}. \tag{17}$$

Lorenz '96: modification

Modification of the forcing parameter by

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + F_k. \tag{16}$$

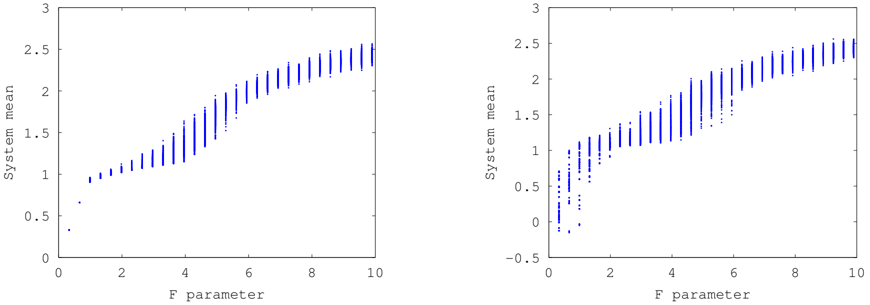
Creates a spatially variable constant in time forcing term F_k with a mean **F** and a perturbation depending on covariance **P** defined by

$$P_{i,j} = 0.3\, e^{\frac{-(i-j)^2}{15}}. \tag{17}$$

Model retains the same characteristics, but spatial F_k is more challenging to recover.

Lorenz '96 model: modification

Lorenz '96 model mean state



(a) Constant F.

(b) Spatially variable F_k.

Lorenz ’96 model: Single assimilation twin experiment

Twin experiment: One simulation is considered as **truth** or **reference**. Observations are extracted from that simulation.

Lorenz '96 model: Single assimilation twin experiment

Twin experiment: One simulation is considered as **truth** or **reference**. Observations are extracted from that simulation.

Parameters:

- Model mean $\overline{F^t} = 4$
- Initial conditions
  $l_{max} = 15$
- Time span $m_{max} = 1000$
- Ensemble size $N_e = 100$

Lorenz '96 model: Single assimilation twin experiment

Twin experiment: One simulation is considered as **truth** or **reference**. Observations are extracted from that simulation.

Parameters:

- Model mean $\overline{F^t} = 4$
- Initial conditions $l_{max} = 15$
- Time span $m_{max} = 1000$
- Ensemble size $N_e = 100$

Model mean obtained from average model trajectory by

$$X_k^t = \frac{1}{l_{max}} \sum_{l=1}^{l_{max}} \frac{1}{m_{max} - 199} \sum_{m=200}^{m_{max}} X_{k,l,m}^t. \quad (18)$$

Lorenz '96 model: Single assimilation twin experiment

Twin experiment: One simulation is considered as **truth** or **reference**. Observations are extracted from that simulation.

Parameters:

- Model mean $\overline{F^t} = 4$
- Initial conditions $l_{max} = 15$
- Time span $m_{max} = 1000$
- Ensemble size $N_e = 100$

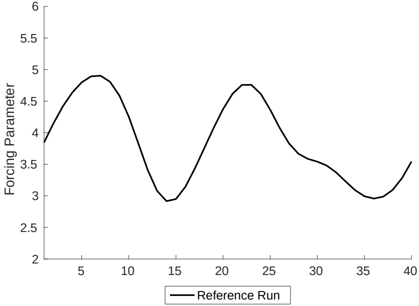
Model mean obtained from average model trajectory by

$$X_k^t = \frac{1}{l_{max}} \sum_{l=1}^{l_{max}} \frac{1}{m_{max} - 199} \sum_{m=200}^{m_{max}} X_{k,l,m}^t. \tag{18}$$

Observations are created by adding noise to the reference model run.

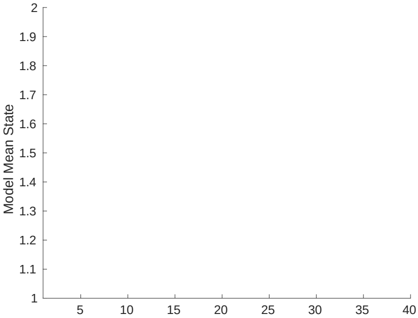
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

Output

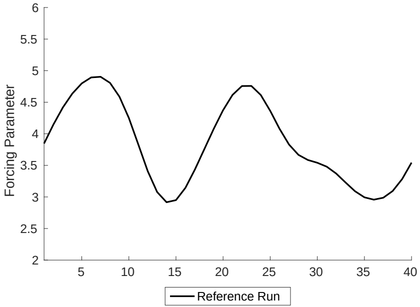


(b)

Reference parameter from twin experiment.

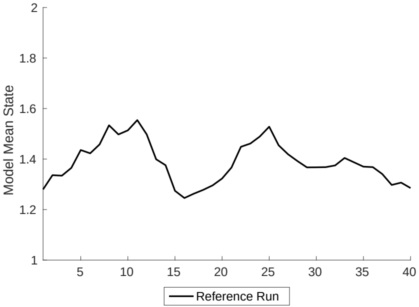
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

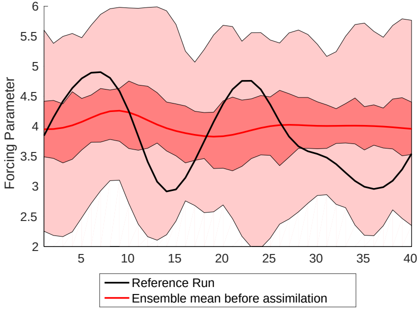
Output



(b)

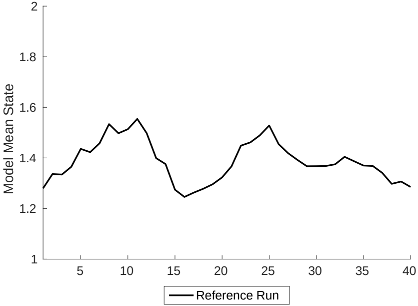
Reference parameter and corresponding model output

Forcing parameter



(a)

Output

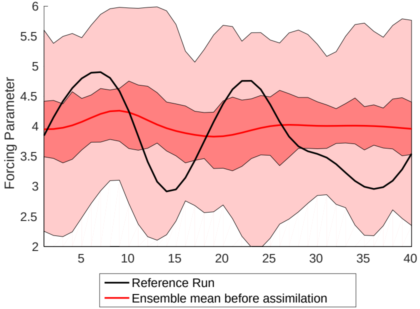


(b)

Ensemble of forcings is generated.

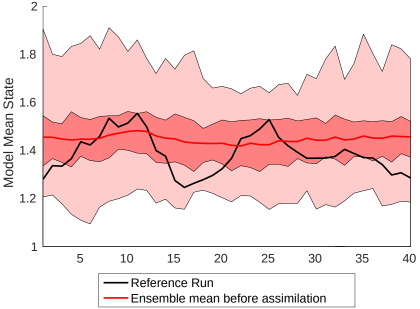
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

Output

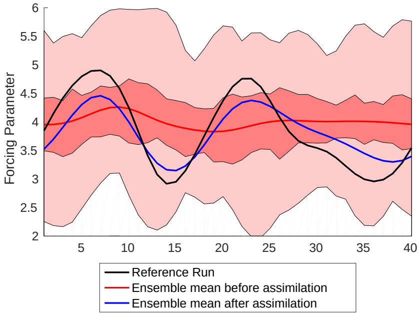


(b)

Ensemble is run to provide an ensemble of model outputs.

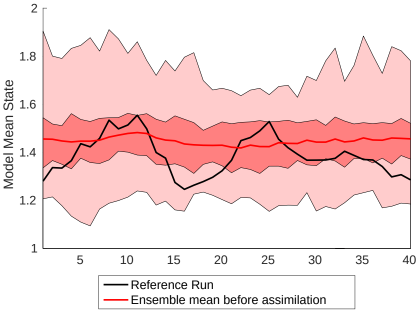
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

Output

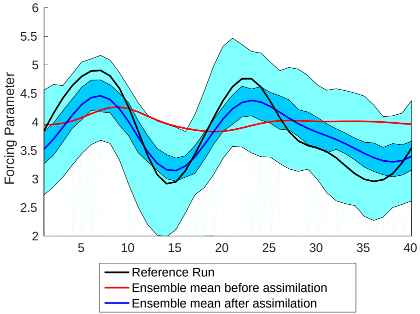


(b)

Observations are assimilated.

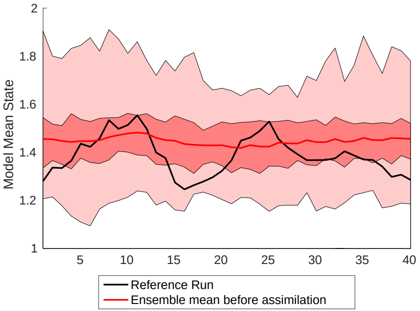
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

Output

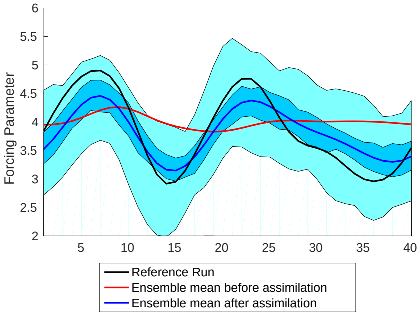


(b)

Ensemble of forcing parameters is corrected.

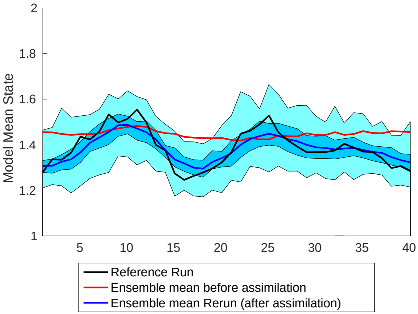
Lorenz '96 model: Single assimilation twin experiment

Forcing parameter



(a)

Output



(b)

Ensemble is rerun to obtain rerun model outputs.

Lorenz '96 model: iterative assimilation

To reduce nonlinear behaviour, one can **iterate** the assimilation procedure. Observation batches creation as follows

$$\mathbf{y}_2^o = \begin{pmatrix} \mathbf{y}^o \\ \mathbf{y}^o \end{pmatrix}, \quad (19) \qquad \mathbf{R}_2 = \begin{pmatrix} 2\mathbf{R} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{R} \end{pmatrix}, \quad (20) \qquad \mathbf{H}_2 = \begin{pmatrix} \mathbf{H} \\ \mathbf{H} \end{pmatrix}. \quad (21)$$

Lorenz '96 model: iterative assimilation

To reduce nonlinear behaviour, one can **iterate** the assimilation procedure. Observation batches creation as follows

$$\mathbf{y}_2^o = \begin{pmatrix} \mathbf{y}^o \\ \mathbf{y}^o \end{pmatrix}, \quad (19) \qquad \mathbf{R}_2 = \begin{pmatrix} 2\mathbf{R} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{R} \end{pmatrix}, \quad (20) \qquad \mathbf{H}_2 = \begin{pmatrix} \mathbf{H} \\ \mathbf{H} \end{pmatrix}. \quad (21)$$

In absence of correlation between subsets of data, iterative assimilation is **equivalent** to single assimilation.

$$(\mathbf{P}^a)^{-1}\mathbf{x}^a = (\mathbf{P}^f)^{-1}\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o = (\mathbf{P}^f)^{-1}\mathbf{x}^f + \mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{y}_2^o. \qquad (22)$$

Lorenz '96 model: iterative assimilation

To reduce nonlinear behaviour, one can **iterate** the assimilation procedure. Observation batches creation as follows

$$\mathbf{y}_2^o = \begin{pmatrix} \mathbf{y}^o \\ \mathbf{y}^o \end{pmatrix}, \quad (19) \qquad \mathbf{R}_2 = \begin{pmatrix} 2\mathbf{R} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{R} \end{pmatrix}, \quad (20) \qquad \mathbf{H}_2 = \begin{pmatrix} \mathbf{H} \\ \mathbf{H} \end{pmatrix}. \quad (21)$$

In absence of correlation between subsets of data, iterative assimilation is **equivalent** to single assimilation.

$$(\mathbf{P}^a)^{-1}\mathbf{x}^a = (\mathbf{P}^f)^{-1}\mathbf{x}^f + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o = (\mathbf{P}^f)^{-1}\mathbf{x}^f + \mathbf{H}_2^T\mathbf{R}_2^{-1}\mathbf{y}_2^o. \qquad (22)$$

However, for a **nonlinear** observation operator, iterations allow smaller steps by model rerun.

Lorenz '96 model: iterative assimilation

Suggested in Annan et al., 2005, and similar to running in place (RIP) algorithm (Kalnay and Yang, 2010).

Lorenz '96 model: iterative assimilation

Suggested in Annan et al., 2005, and similar to running in place (RIP) algorithm (Kalnay and Yang, 2010).

Experiment parameters:

- $\overline{F^t} = 5$
- $\overline{F^f} = 6$
- $l_{max} = 10$
- $m_{max} = 1000$
- $N_e = 50$
- $n_{iter}^{max} = 1, 4$
- $n_{iter} = 1, n_{iter}^{max}$

Lorenz '96 model: iterative assimilation

Suggested in Annan et al., 2005, and similar to running in place (RIP) algorithm (Kalnay and Yang, 2010).

Experiment parameters:

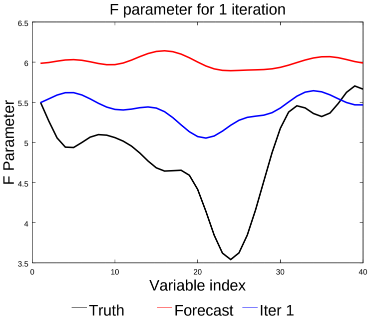
- $\overline{F^t} = 5$
- $\overline{F^f} = 6$
- $l_{max} = 10$
- $m_{max} = 1000$
- $N_e = 50$
- $n_{iter}^{max} = 1, 4$
- $n_{iter} = 1, n_{iter}^{max}$

Different $\overline{F^t}$ and $\overline{F^f}$ for readability. Spread is sufficient.

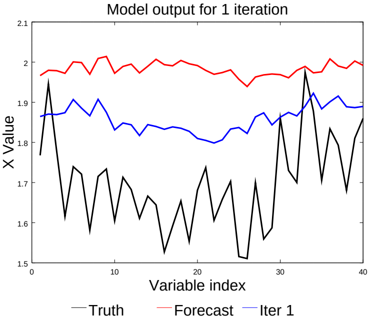
Same **initial condition** for every iteration experiment.

Increase in computational cost proportional to n_{iter}^{max}.

Lorenz '96 model: iterative assimilation

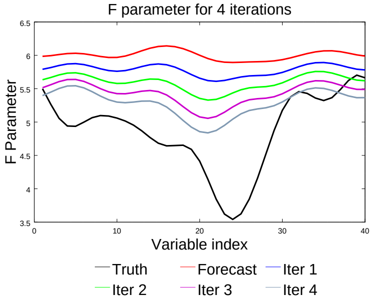


(a) F_k for a single assimilation.

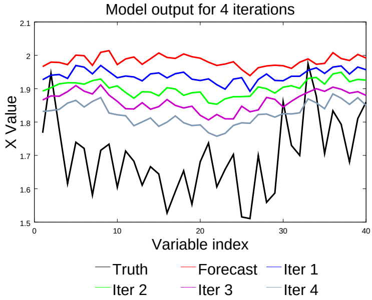


(b) Corresponding time averaged model state.

Lorenz '96 model: iterative assimilation



(a) F_k for an iterated assimilation.



(b) Corresponding time averaged model state.

Lorenz '96 model: iterative assimilation

Exact RMSE values on the forcing parameter:

Background RMSE		Analysed RMSE			
n_{iter}^{max}	Background	$n_{iter} = 1$	$n_{iter} = 2$	$n_{iter} = 3$	$n_{iter} = 4$
1	1.270	0.726			
2	1.270	0.915	0.663		
3	1.270	1.007	0.799	0.639	
4	1.270	1.060	0.887	0.737	0.619

Table : RMSE on F_k

Lorenz '96 model: iterative assimilation

Exact RMSE values of the corrected model state rerun ensemble mean:

Background RMSE		Rerun RMSE			
n_{iter}^{max}	Background	$n_{iter} = 1$	$n_{iter} = 2$	$n_{iter} = 3$	$n_{iter} = 4$
1	0.304	0.187			
2	0.304	0.233	0.170		
3	0.304	0.254	0.203	0.163	
4	0.304	0.263	0.227	0.195	0.160

Table : RMSE on the time average of the model state

Lorenz '96 model

Conclusion:

- Modified Lorenz '96 model application.
- Bias correction term is estimated.
- Model rerun exhibits a significant improvement on the non-corrected run.
- Iterative assimilation reduces bias even further.

Lorenz '96 model

Conclusion:

- Modified Lorenz '96 model application.
- Bias correction term is estimated.
- Model rerun exhibits a significant improvement on the non-corrected run.
- Iterative assimilation reduces bias even further.

Time to test the method with a realistic model.

NEMO-LIM2: Bias correction generation

Realistic model: NEMO (Nucleus for European Modelling of the Ocean), coupled to the LIM2 (Louvain-la-Neuve Sea Ice Model) sea ice model.

Bias: Caused by low resolution, currents too weak.

NEMO-LIM2: Bias correction generation

Realistic model: NEMO (Nucleus for European Modelling of the Ocean), coupled to the LIM2 (Louvain-la-Neuve Sea Ice Model) sea ice model.

Bias: Caused by low resolution, currents too weak.

Method implementation: Correct currents by adding a random forcing into the ocean dynamics equations.

Constraints: Model stability, physical restrictions on currents.

NEMO-LIM2: Bias correction generation

Realistic model: NEMO (Nucleus for European Modelling of the Ocean), coupled to the LIM2 (Louvain-la-Neuve Sea Ice Model) sea ice model.

Bias: Caused by low resolution, currents too weak.

Method implementation: Correct currents by adding a random forcing into the ocean dynamics equations.

Constraints: Model stability, physical restrictions on currents.

Assimilation Scheme: Local assimilation with ETKF scheme from OAK (Ocean Assimilation Kit).

NEMO-LIM2: Bias correction generation

Forcing term generation from **random field** $\Psi = \Psi(x, y)$ and from the Hessian of the cost function $J(\Psi)$ by

$$J(\Psi) = \int_{\Omega} L_h^4 (\nabla^2 \Psi)^2 + 2L_h^2 (\nabla \Psi)^2 + \Psi dx. \tag{23}$$

This provides a random **stream function tendency** with constraints on correlation length L_h.

NEMO-LIM2: Bias correction generation

Forcing term generation from **random field** $\Psi = \Psi(x, y)$ and from the Hessian of the cost function $J(\Psi)$ by

$$J(\Psi) = \int_{\Omega} L_h^4(\nabla^2\Psi)^2 + 2L_h^2(\nabla\Psi)^2 + \Psi dx. \tag{23}$$

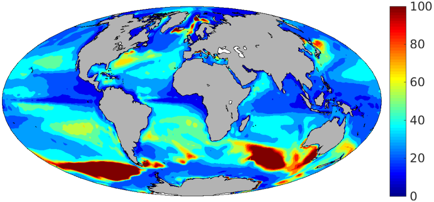
This provides a random **stream function tendency** with constraints on correlation length L_h.

One avoids perpendicular currents with $\nabla\Psi \bullet \mathbf{t} = 0$, and uses spatial filtering to increase model stability.

Average mixed layer depth

Vertical extension with yearly mean mixed layer depth by

$$\Psi'(x,y,z) = \frac{\Psi(x,y)}{1+\exp(\frac{z-T(x,y)}{L_v})}.$$



Yearly average of the mixed layer depth from a NEMO-LIM2 free run, in m.

NEMO-LIM2: Bias correction generation

Zonal and **meridional** forcings from stream function

$$F_u(x, y, z) = -\frac{\partial \Psi'(x, y, z)}{\partial y}, \quad (24) \qquad F_v(x, y, z) = \frac{\partial \Psi'(x, y, z)}{\partial x}. \quad (25)$$

NEMO-LIM2: Bias correction generation

Zonal and **meridional** forcings from stream function

$$F_u(x,y,z) = -\frac{\partial \Psi'(x,y,z)}{\partial y}, \quad (24) \qquad F_v(x,y,z) = \frac{\partial \Psi'(x,y,z)}{\partial x}. \quad (25)$$

Modified ocean dynamics equations:

$$\frac{du}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + fv + \frac{1}{\rho}\frac{\partial \tau_x}{\partial z} + F_u, \quad (26)$$

$$\frac{dv}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - fu + \frac{1}{\rho}\frac{\partial \tau_y}{\partial z} + F_v. \quad (27)$$

NEMO-LIM2: Experiment Set-up

NEMO-LIM2 experiments set-up, both twin and realistic cases, as follow:

NEMO-LIM2: Experiment Set-up

NEMO-LIM2 experiments set-up, both twin and realistic cases, as follow:

Experiment parameters

- Correlation length: 5000km
- SSH spatial variability: 30cm
- Ensemble size: 100 members

NEMO-LIM2: Experiment Set-up

NEMO-LIM2 experiments set-up, both twin and realistic cases, as follow:

Experiment parameters

- Correlation length: 5000km
- SSH spatial variability: 30cm
- Ensemble size: 100 members

Procedure

- Run ensemble with random forcing
- Assimilate SSH observations
- Extract optimal forcing
- Rerun ensemble mean and validate

NEMO-LIM2: Experiment Set-up

NEMO-LIM2 experiments set-up, both twin and realistic cases, as follow:

Experiment parameters

- Correlation length: $5000km$
- SSH spatial variability: $30cm$
- Ensemble size: 100 members

Procedure

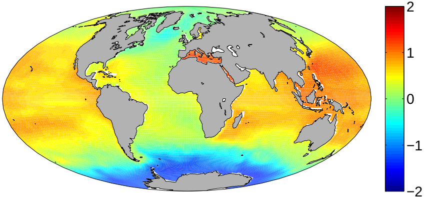
- Run ensemble with random forcing
- Assimilate SSH observations
- Extract optimal forcing
- Rerun ensemble mean and validate

State vector is augmented with the observations: $\mathbf{x}'' = \begin{bmatrix} \overline{SSH} \\ \widehat{F}_u \\ \widehat{F}_v \end{bmatrix}$.

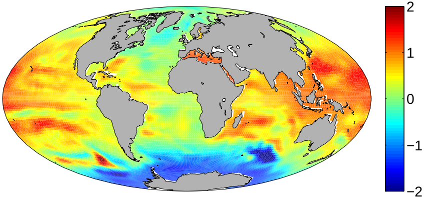
NEMO-LIM2 Twin Experiment: SSH

Ensemble of forcings is run. Observations are taken from the reference run.

Yearly mean sea surface height (in m)



(a) Average of the ensemble.

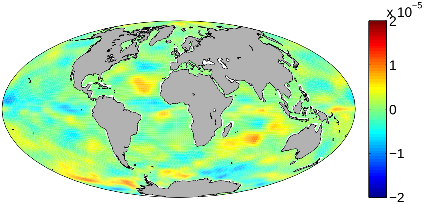


(b) Twin experiment true run.

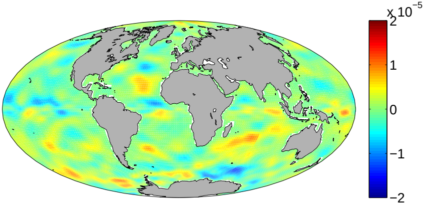
NEMO-LIM2 Twin Experiment: Zonal forcing

After SSH assimilation, the analysis provides estimated bias correction terms.

Zonal forcing (in ms^{-2})



(a) Ensemble mean after analysis.

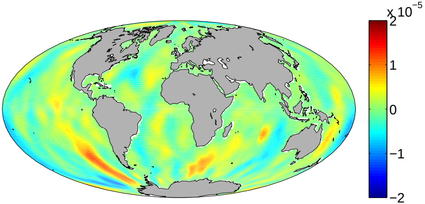


(b) Reference run.

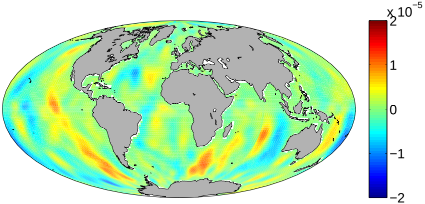
NEMO-LIM2 Twin Experiment: Meridional forcing

After SSH assimilation, the analysis provides estimated bias correction terms.

Meridional forcing (in ms^{-2})



(a) Ensemble mean after analysis.

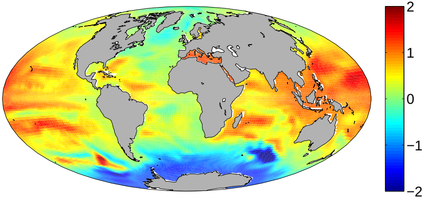


(b) Reference run.

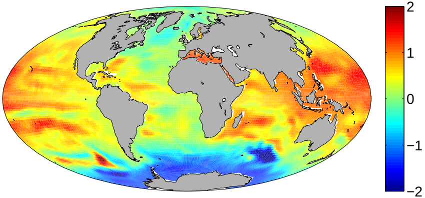
NEMO-LIM2 Twin Experiment: SSH rerun

The optimal forcing is rerun, providing a bias corrected run.

Model rerun SSH (in m)

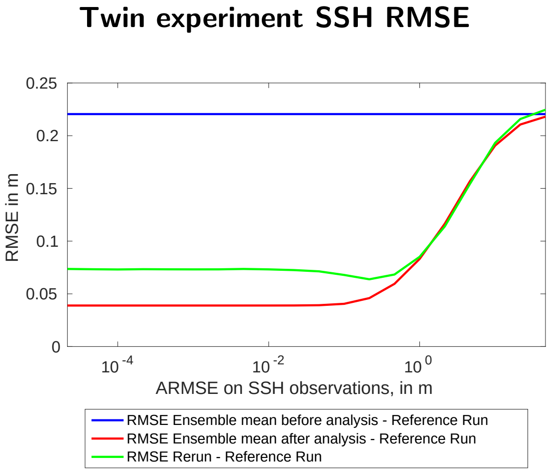


(a) Rerun with optimal forcing.



(b) Reference run.

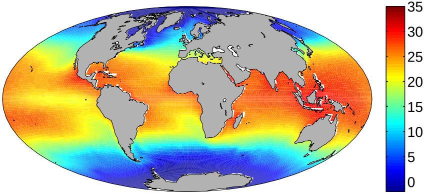
NEMO-LIM2 Twin Experiment: SSH RMSE



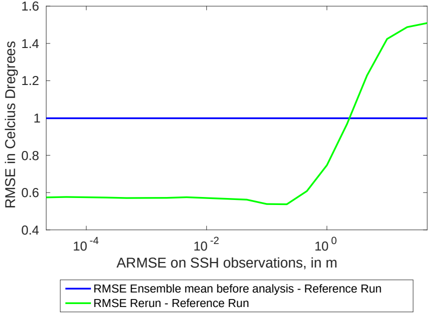
RMSE on SSH from Ensemble Mean before and after analysis, and Rerun, with True Run (in m)

NEMO-LIM2 Twin Experiment: SST Validation

SST validation (in C°)



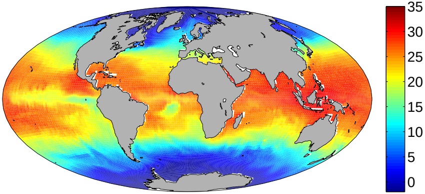
(a) Average of the ensemble.



(b) RMSE on SST from Ensemble Mean after analysis, and Rerun, with True Run (in C°).

NEMO-LIM2 Twin Experiment: SST Validation

SST validation (in C°)



(a) Rerun with optimal forcing.

(b) Reference run.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}					
$\widehat{F}_v$ in ms^{-2}					
SSH in m					
SST in C$^\circ$					
SSS in PSU					

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,\mathrm{C}^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}				
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}				
SSH in m					
SST in C°					
SSS in PSU					

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,C^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}				
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}				
SSH in m	0.220				
SST in C$^\circ$	0.999				
SSS in PSU	0.268				

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,\mathrm{C}^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}	6.27×10^{-7}			
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}	5.81×10^{-7}			
SSH in m	0.220				
SST in C$^\circ$	0.999				
SSS in PSU	0.268				

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,\mathrm{C}^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}	6.27×10^{-7}			
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}	5.81×10^{-7}			
SSH in m	0.220		0.068		
SST in C$^\circ$	0.999		0.539		
SSS in PSU	0.268		0.197		

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,C^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}	6.27×10^{-7}		5.96×10^{-7}	
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}	5.81×10^{-7}		5.45×10^{-7}	
SSH in m	0.220		0.068		
SST in C°	0.999		0.539		
SSS in PSU	0.268		0.197		

Table : RMSE values of the multivariate rerun for a $ARMSE = 1$ C° value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

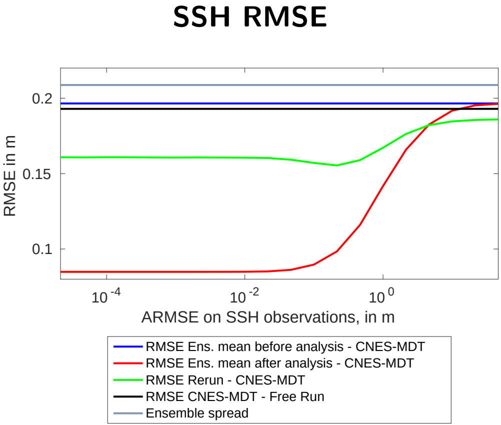
SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}	6.27×10^{-7}		5.96×10^{-7}	
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}	5.81×10^{-7}		5.45×10^{-7}	
SSH in m	0.220		0.068		0.061
SST in C$^\circ$	0.999		0.539		0.509
SSS in PSU	0.268		0.197		0.150

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,$C$^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

NEMO-LIM2 Real Case

SSH observations come from the **mean dynamic topography** (MDT) of CNES-CLS09 (Centre National d'Etudes Spatiales, Collecte Localisation Satellites).

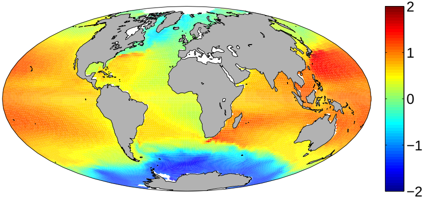


RMSE on SSH with CNES-MDT observations (in m).

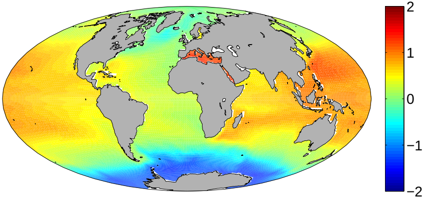
NEMO-LIM2 Real Case: SSH

Strong currents are too weak in NEMO.

Yearly mean SSH (in m)



(a) CNES-MDT observations.

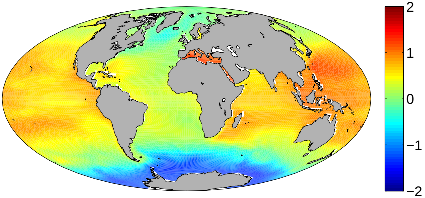


(b) Model free run.

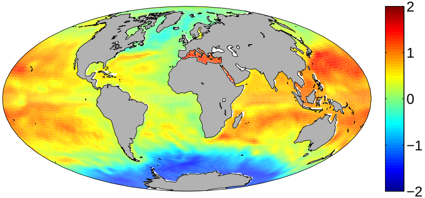
NEMO-LIM2 Real Case: SSH

Corrected rerun shows small scale structure, but also stronger currents.

Yearly mean SSH (in m)



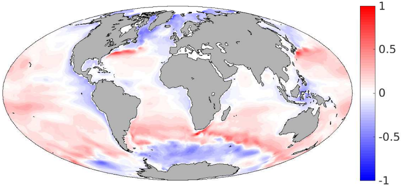
(a) Average of the ensemble.



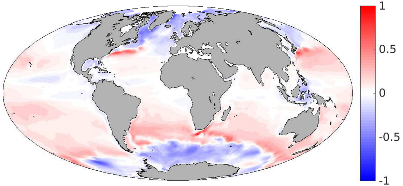
(b) Optimal forcing rerun.

NEMO-LIM2 Real Case: SSH average error

Yearly mean SSH average error with the observations (in m)



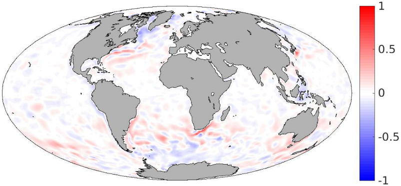
(a) Average of the ensemble.



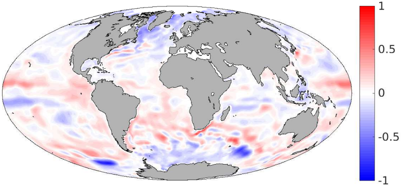
(b) Model free run.

NEMO-LIM2 Real Case: SSH average error

Yearly mean SSH average error with the observations (in m)

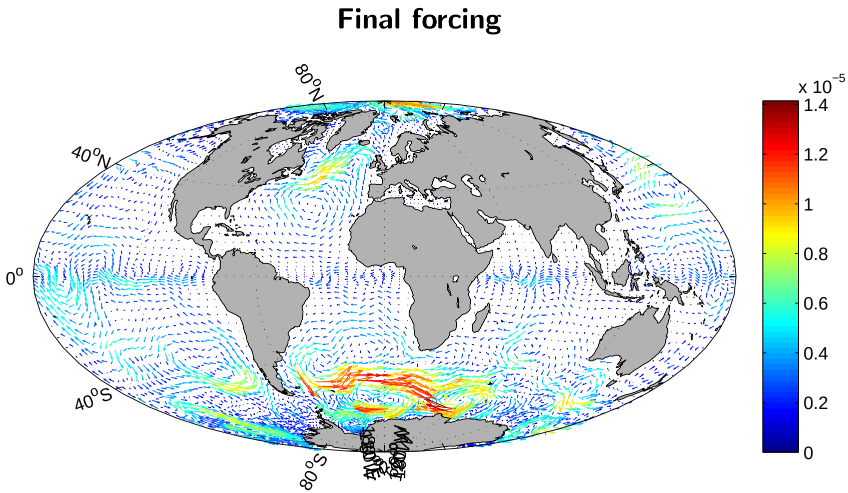


(a) Analysis.



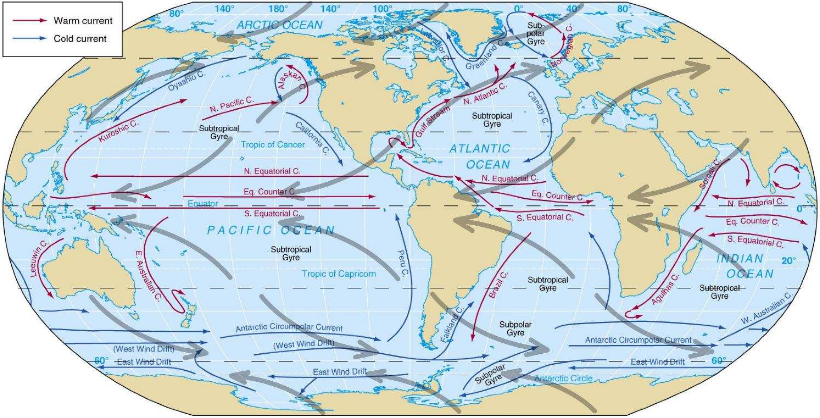
(b) Optimal forcing rerun.

NEMO-LIM2 Real Case: final forcing



Analysed forcing from CNES-MDT observations, used to rerun the model (in ms^{-2}).

Real global current map

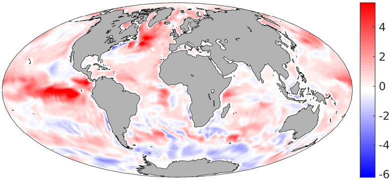


Real global average current map of the oceans. Adapted from http://www.georgemaps.com/

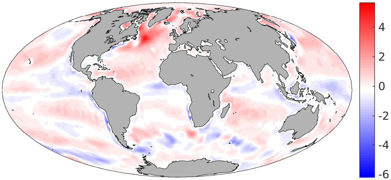
NEMO-LIM2 Real Case: SST Validation

SST climatology from NODC-WOA13V2 data provided by the National Oceanic and Atmospheric Administration (NOAA).

Yearly mean SST average error (in C°)

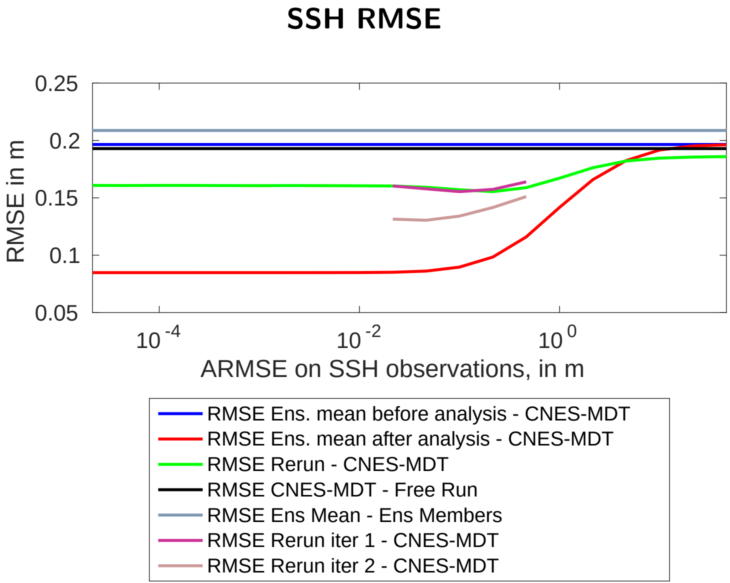


(a) Optimal forcing.



(b) Model free run.

NEMO-LIM2 Real Case: Iterative Analysis



NEMO-LIM2 Real Case: Iterative Analysis

Exact RMSE values on the SSH for **single** and **iterative** experiments.

ARMSE on SSH (in m)		RMSE (in m)			
		Background	Single Assim	Iterative Assim	
R	2**R**			iter 1	iter 2
0.0215	0.0431				
0.0464	0.0928				
0.1000	0.2000				
0.2154	0.4308				
0.4642	0.9284				

Table : RMSE on SSH from the ensemble mean before analysis with CNES-MDT observations, from the forced rerun with the observations, and from the first and second successive iterated assimilations.

NEMO-LIM2 Real Case: Iterative Analysis

Exact RMSE values on the SSH for **single** and **iterative** experiments.

ARMSE on SSH (in m)		RMSE (in m)			
		Background	Single Assim	Iterative Assim	
R	2**R**			iter 1	iter 2
0.0215	0.0431	0.1965			
0.0464	0.0928	0.1965			
0.1000	0.2000	0.1965			
0.2154	0.4308	0.1965			
0.4642	0.9284	0.1965			

Table : RMSE on SSH from the ensemble mean before analysis with CNES-MDT observations, from the forced rerun with the observations, and from the first and second successive iterated assimilations.

NEMO-LIM2 Real Case: Iterative Analysis

Exact RMSE values on the SSH for **single** and **iterative** experiments.

ARMSE on SSH (in m)		RMSE (in m)			
		Background	Single Assim	Iterative Assim	
R	2**R**			iter 1	iter 2
0.0215	0.0431	0.1965	0.1604		
0.0464	0.0928	0.1965	0.1592		
0.1000	0.2000	0.1965	0.1571		
0.2154	0.4308	0.1965	0.1554		
0.4642	0.9284	0.1965	0.1589		

Table : RMSE on SSH from the ensemble mean before analysis with CNES-MDT observations, from the forced rerun with the observations, and from the first and second successive iterated assimilations.

NEMO-LIM2 Real Case: Iterative Analysis

Exact RMSE values on the SSH for **single** and **iterative** experiments.

ARMSE on SSH (in m)		RMSE (in m)			
		Background	Single Assim	Iterative Assim	
R	2**R**			iter 1	iter 2
0.0215	0.0431	0.1965	0.1604	0.1604	
0.0464	0.0928	0.1965	0.1592	0.1579	
0.1000	0.2000	0.1965	0.1571	0.1554	
0.2154	0.4308	0.1965	0.1554	0.1574	
0.4642	0.9284	0.1965	0.1589	0.1640	

Table : RMSE on SSH from the ensemble mean before analysis with CNES-MDT observations, from the forced rerun with the observations, and from the first and second successive iterated assimilations.

NEMO-LIM2 Real Case: Iterative Analysis

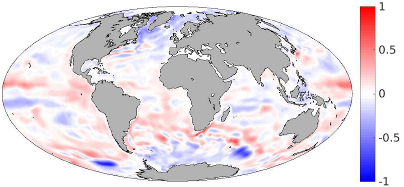
Exact RMSE values on the SSH for **single** and **iterative** experiments.

ARMSE on SSH (in m)		RMSE (in m)			
		Background	Single Assim	Iterative Assim	
R	2**R**			iter 1	iter 2
0.0215	0.0431	0.1965	0.1604	0.1604	0.1315
0.0464	0.0928	0.1965	0.1592	0.1579	0.1305
0.1000	0.2000	0.1965	0.1571	0.1554	0.1341
0.2154	0.4308	0.1965	0.1554	0.1574	0.1416
0.4642	0.9284	0.1965	0.1589	0.1640	0.1511

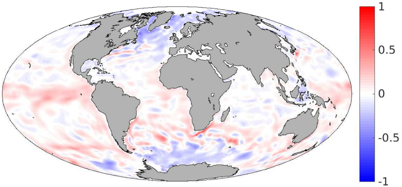
Table : RMSE on SSH from the ensemble mean before analysis with CNES-MDT observations, from the forced rerun with the observations, and from the first and second successive iterated assimilations.

NEMO-LIM2 Real Case: Iterative Analysis SSH average error

Yearly mean SSH average errors (in m)



(a) First iteration rerun.



(b) Second iteration rerun.

NEMO-LIM2 Real Case: Conclusions and perspectives

The twin experiment shows a **significant** reduction on the SSH bias through adequate model forcing. The real case shows that the bias **generation** is important, and the strong **seasonal** cycle in the Antarctic ocean.

NEMO-LIM2 Real Case: Conclusions and perspectives

The twin experiment shows a **significant** reduction on the SSH bias through adequate model forcing. The real case shows that the bias **generation** is important, and the strong **seasonal** cycle in the Antarctic ocean.

Possible development options for the **future**:

- Real 3D forcing
- Time dependent forcing, with seasonal variations
- Validation against other bias correction methods
- Parametrisation of the final forcing
- Extract local optimal forcing

Thesis: Conclusions

Theoretical formulation of the bias correction method.

Thesis: Conclusions

Theoretical formulation of the bias correction method.

The application on the Lorenz '96 shows:

- Estimation of the bias correction term.
- Reduced model bias on the model rerun.

Thesis: Conclusions

Theoretical formulation of the bias correction method.

The application on the Lorenz '96 shows:

- Estimation of the bias correction term.
- Reduced model bias on the model rerun.

NEMO-LIM2 implementation:

- Correction term forcing is stable on the model.
- Bias correction works on complex models.
- Iterative assimilation is more accurate.

Thesis: Conclusions

Theoretical formulation of the bias correction method.

The application on the Lorenz '96 shows:

- Estimation of the bias correction term.
- Reduced model bias on the model rerun.

NEMO-LIM2 implementation:

- Correction term forcing is stable on the model.
- Bias correction works on complex models.
- Iterative assimilation is more accurate.

Thesis objectives:

- Model correction can be used for future projections.
- Applicable to other models.

Thesis: Conclusions

Thank you for your attention.

State of the art: Data Assimilation Notation

Multiple notations through time. Unified notation in Ide et al. (1997).

Variables	Descriptions
$_m$	Time index subscript.
^{f} ^{a}	Forecast and analysis superscripts.
$\mathbf{x}, \mathbf{y}$	State vector, observations.
$\mathbf{M}, \mathbf{H}$	Forward model and observation operators.
$\mathbf{P}$	Covariance matrix of state vector.
$\mathbf{Q}, \mathbf{R}$	Model and observation error covariance matrix.
$\mathbf{K}$	Kalman gain.
^{i}	Ensemble index superscripts.

NEMO-LIM2 Twin Experiment: Multivariate Assimilation

SSH and SST can be assimilated together, for a better estimation of the bias correction forcing.

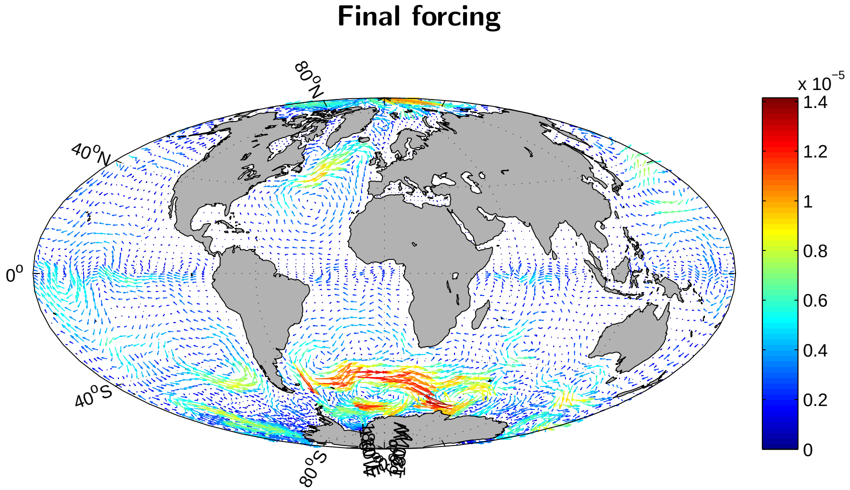
Variable name	Forecast	Monovariate		Multivariate	
		analysis	rerun	analysis	rerun
$\widehat{F}_u$ in ms^{-2}	1.66×10^{-6}	6.27×10^{-7}		5.96×10^{-7}	
$\widehat{F}_v$ in ms^{-2}	1.24×10^{-6}	5.81×10^{-7}		5.45×10^{-7}	
SSH in m	0.220	0.039	0.068	0.0457	0.061
SST in C°	0.999		0.539	0.453	0.509
SSS in PSU	0.268		0.197		0.150

Table : RMSE values of the multivariate rerun for a $ARMSE = 1\,C^\circ$ value, compared to the monovariate assimilation. Empty values are not relevant.

D.4 The program structure

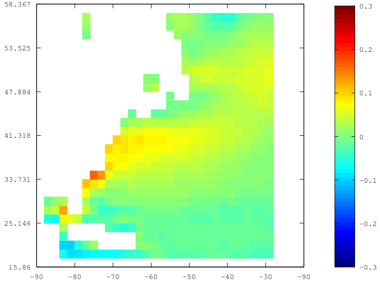
To be done....

NEMO-LIM2 Real Case: final forcing

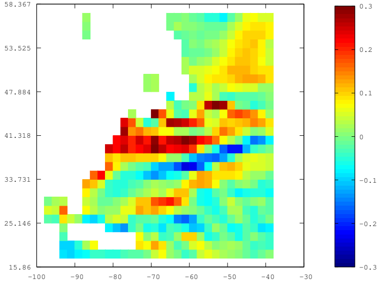


Analysed forcing from CNES-MDT observations, used to rerun the model (in ms^{-2}).

Zonal current Gulf Stream (in ms)



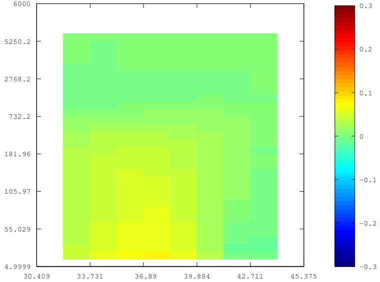
(a) Free model.



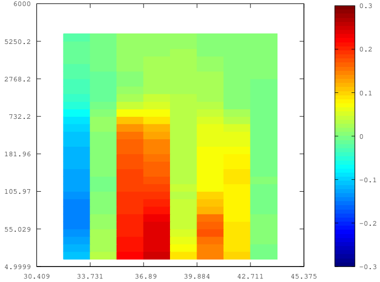
(b) Corrected model.

NEMO-LIM2 Real Case: Gulf Stream

Zonal current Gulf Stream, latitude cut (in m/s)



(a) Free model.



(b) Corrected model.

Free model range: -0.2921 to 0.25035 (m/s). Forced model: -0.57111 to 0.53793 (m/s).