IMPROVING THE PARAMETERIZATION OF ERRORS STATISTICS FOR DATA ASSIMILATION IN A HYCOM BAY OF BISCAY REGIONAL CONFIGURATION

P. BRASSEUR, G. BROQUET, J.M. BRANKART, F. CASTRUCCIO, C. LAUVERNET and J. VERRON

LEGI – Laboratoire des Ecoulements Géophysiques et Industriels – Grenoble, France

Ocean Sciences meeting, 2008, Orlando, 03 – 08 March 2008

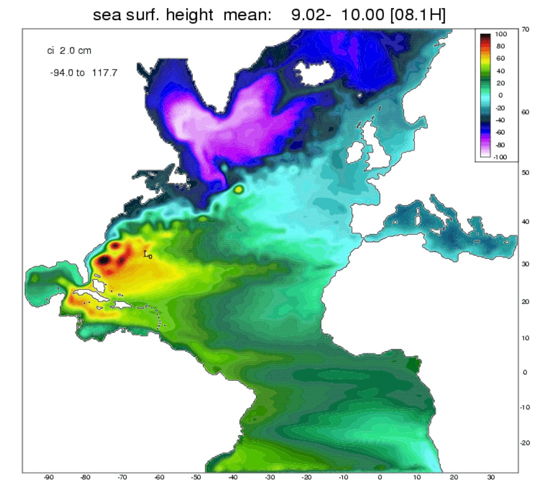
Context

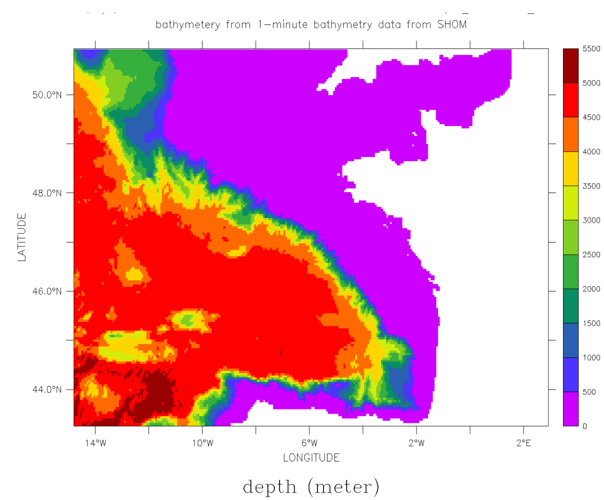
- Development of the **SEEK** (Singular Evolutive Extended **Kalman filter**) Data Assimilation system for both scientific and operational (MERCATOR / GODAE / HYCOM) mesoscale applications : ***Pham et al., 1998; Brasseur and Verron, 2006***

- The SEEK in **MICOM/HYCOM** : experiments in **North Atlantic configurations** within the DIADEM/TOPAZ European project : ***Brankart et al., 2003; Birol et al., 2004; Rozier et al., 2007*** and at the NRL : ***Parent et al., 2007***

- Development of a HYCOM Bay of Biscay configuration at the SHOM

→ application of the SEEK filter to a HYCOM regional configuration in a coastal zone : the **Bay of Biscay model**





Previous issues of the SEEK applications (in OPA / MICOM / HYCOM basin configurations) made more critical because of a higher complexity near the coasts at high resolution : difficulties with the usual crude assumptions of Data Assimilation systems.

Issues on KF/DA errors statistics

KF Forecast step

$$\mathbf{x}_{i+1}^{f} = m\left(\mathbf{x}_{i}^{a}\right)$$

$$\mathbf{P}_{i+1}^{f} = \mathbf{M}\mathbf{P}_{i}^{a}\mathbf{M}^{T} + \mathbf{Q}$$

KF Analysis step

$$\mathbf{K}_{i+1} = \mathbf{P}_{i+1}^{f}\mathbf{H}^{T}\left(\mathbf{H}\mathbf{P}_{i+1}^{f}\mathbf{H}^{T} + \mathbf{R}\right)^{-1}$$

$$\mathbf{x}_{i+1}^{a} = \mathbf{x}_{i+1}^{f} + \mathbf{K}_{i+1}\left(\mathbf{y}_{i+1} - h\left(\mathbf{x}_{i+1}^{f}\right)\right)$$

$$\mathbf{P}_{i+1}^{a} = \left(\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}\right)\mathbf{P}_{i+1}^{f}$$

Errors in the KF (DA in general):

- **Gaussian / not constrained** despite constrained model variables

- $\mathbf{P}_0$, **Q** and **R** a priori described with simple characteristics for calculations

$\mathbf{P}^f$ usually stationary $\mathbf{P}^f=\mathbf{P}_0$; $\mathbf{Q}=0$ (see also in SEEK $\mathbf{Q}=\alpha\mathbf{M}\mathbf{P}_a\mathbf{M}^T$) with $\mathbf{P}_0$ estimated:

• analytically (OI, variational systems) : problem of isotropy/homogeneity, stronger in coastal zones

• with the model variability (+ order reduction: through EOF decomposition in the SEEK): supposes that the model main modes have well reproduced real dynamics

→ **no true description of a model error Q**, whose weight is higher in coastal zones

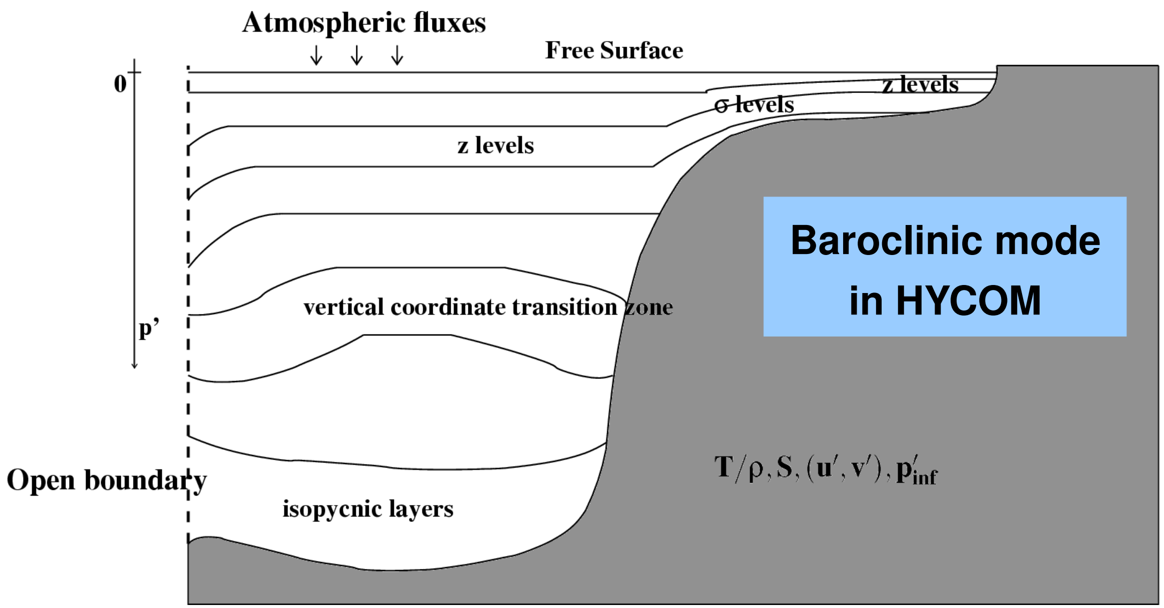
Incorrect representation of error spatial and multivariate covariances:

→ **unbalanced new initial conditions**, violating some laws of the model

→ adjustment in the observed sub-space, but **lack of sensible extrapolation of the obs information**

Applying the KF in HYCOM

Pb of bad representation of error highly sensitive in HYCOM: constrained variables, changes in *T*,*S* and *p* strongly dependent in the hybrid space. Some usual answers:



→ DA in *z* space (strong loss of background and obs information from regridding)

→ **In the SEEK**: extrapolation of data is strongly limited (cancellation of many covariances) + **adjustment operator** (loss of statistical optimality): correction after analysis of densities (*T*/*S*) and of *p* for **hydrostatic stability** and insuring the **minimum layer thicknesses**

Objective : describing more realistic/adapted errors in the HYCOM state space to

- generate more balanced new initial conditions during the KF analysis step

- improve statistical results on both observed and non observed variables

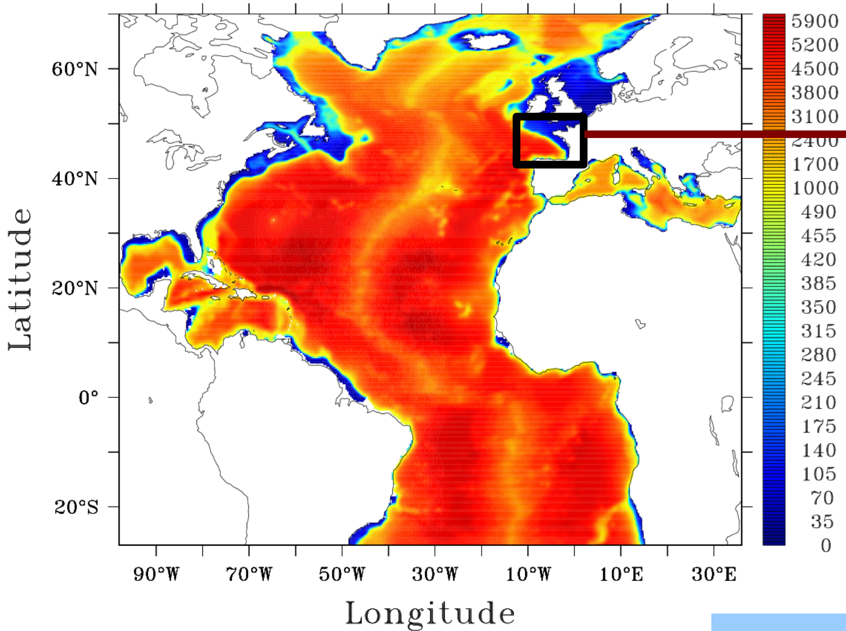
Two approaches

(1) generating/using a better estimation of Q

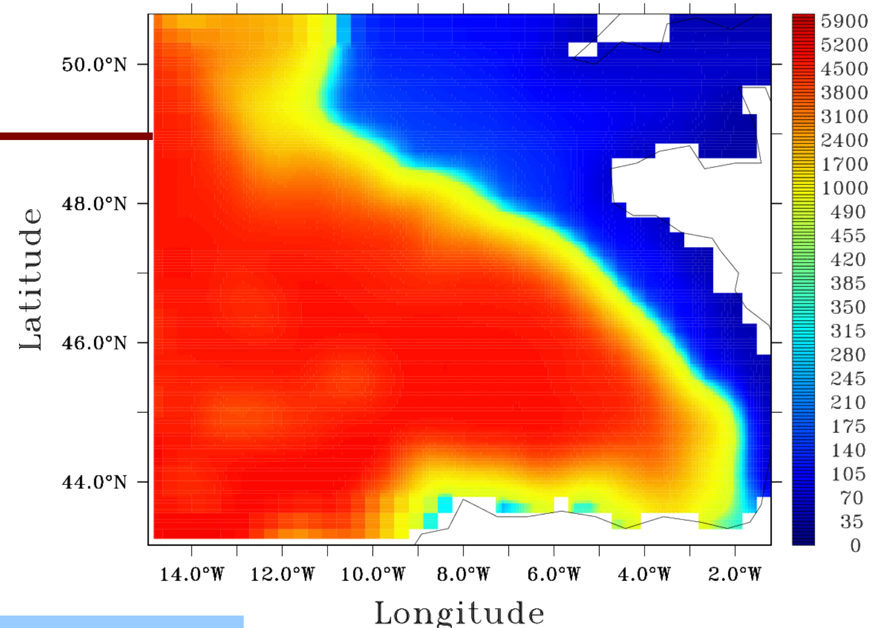
(2) including model inequality constraints on errors/in the KF process

The Bay of Biscay model

NA3, 1/3° resolution, 26 layers



BB15, 1/15° resolution, 26 layers



Bathymetries (from ETOPO5)

Forcings at open boundaries: method of characteristics for barotropic variables: p_b, (u_b, v_b); relaxation for baroclinic variables: S, T, p', (u', v')

Forcings at the surface: Bulk formulation for heat flux, wind stress imposed, no fresh water flux (atmospheric parameters: w, T_a, q_{lw}, q_{shw}, E_v)

An ensemble method for model error statistics estimation

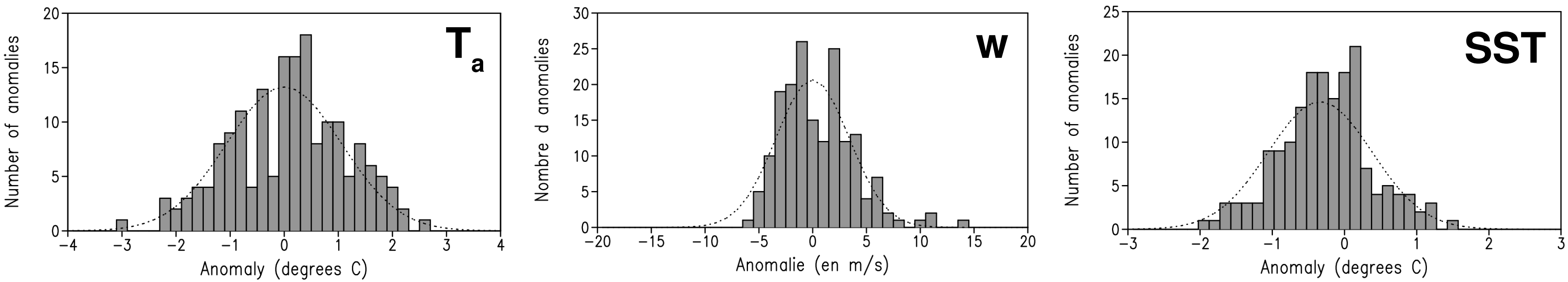
Broquet et al. 2007

Q build from a sample of the model response to perturbations in the main source of error (realism of the estimation and allowance for error non Gaussian/linear propagation)

1) Identification of **main sources of model error**: increased role and weak representation of **surface forcings** in regional case

2) Generation of an ensemble of “perturbed” forcings: use of an **ensemble of realistic forcings**, assuming that the temporal variability (weekly and inter-annual) of data is representative of the forcing error statistics over a given season

3) Generation of the related **ensemble of simulation representative of the model error**



Distribution of 176 anomalies on T_a, w and SST at (11.7°W,45.4°N) and t=16d

Gaussian pdfs generated on not strongly constrained atmospheric parameters/model variables

→ optimal trade off between KF assumptions and realism of error distributions

Imposing inequality constraints in the KF

Lauvernet et al. 2008

Schematic one-dimensional TG pdfs associated to N(0,1) with constraint I : x≤1 or x≤-1

Development of a **Truncated Gaussian filter**: an adaptation of the KF to deal with errors of TG pdfs

1) Description of **the "true state" distribution as the TG distribution** N($\mathbf{x}^f$,$\mathbf{P}^f$,I) associated with an usual "true state" Gaussian distribution N($\mathbf{x}^f$,$\mathbf{P}^f$) and a natural set of inequality constraints I : **the hydrostatic stability and the minimum layer thicknesses condition**

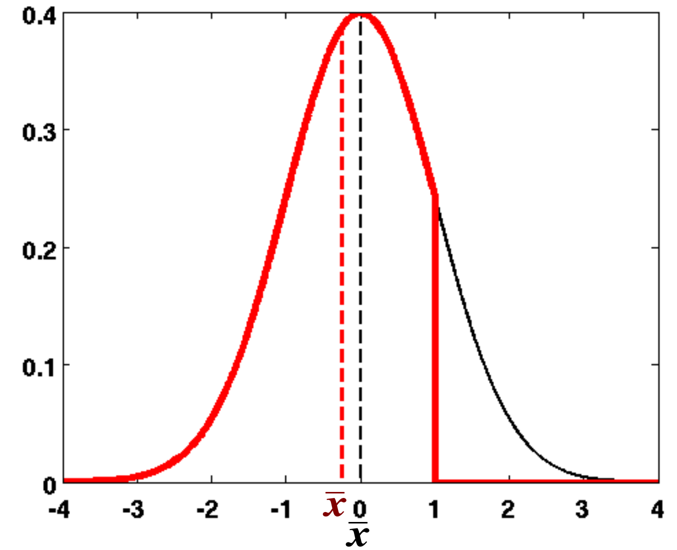
$$\rho_k - \rho_{k+1} \leq \epsilon \qquad p_k \geq p_k^{\min}(H)$$

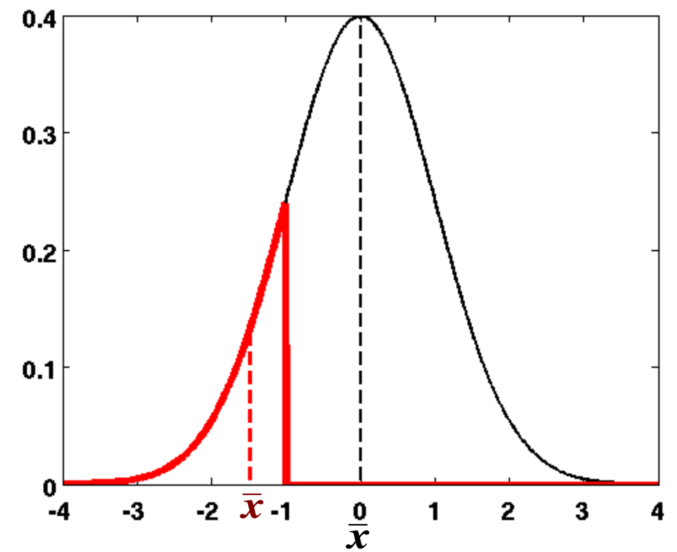
$$p_k - p_{k+1} \leq 0$$

2) Application of **traditional KF update** ($\mathbf{x}^f$,$\mathbf{P}^f$)→ ($\mathbf{x}^a$,$\mathbf{P}^a$) giving the updated optimal error distribution N($\mathbf{x}^a$,$\mathbf{P}^a$,I)

3) Calculation of the **error variance minimizing estimator from** N($\mathbf{x}^a$,$\mathbf{P}^a$,I) to forecast $\mathbf{x}^f$: **use of a Gibbs sampler.**

Quasi-normal approximation: $\mathbf{P}^f$ is obtained with classical KF forecast step from $\mathbf{P}^a$. A true TG ensemble forecasting on the sample of N($\mathbf{x}^a$,$\mathbf{P}^a$,I)→N($\mathbf{x}^f$,$\mathbf{P}^f$,I) has been tested on a z-coordinate ML model for conservation of hydrostatic stability





Representers of error covariances

Bennett, 1992 ; Echevin et al. 2000

KF analysis state update

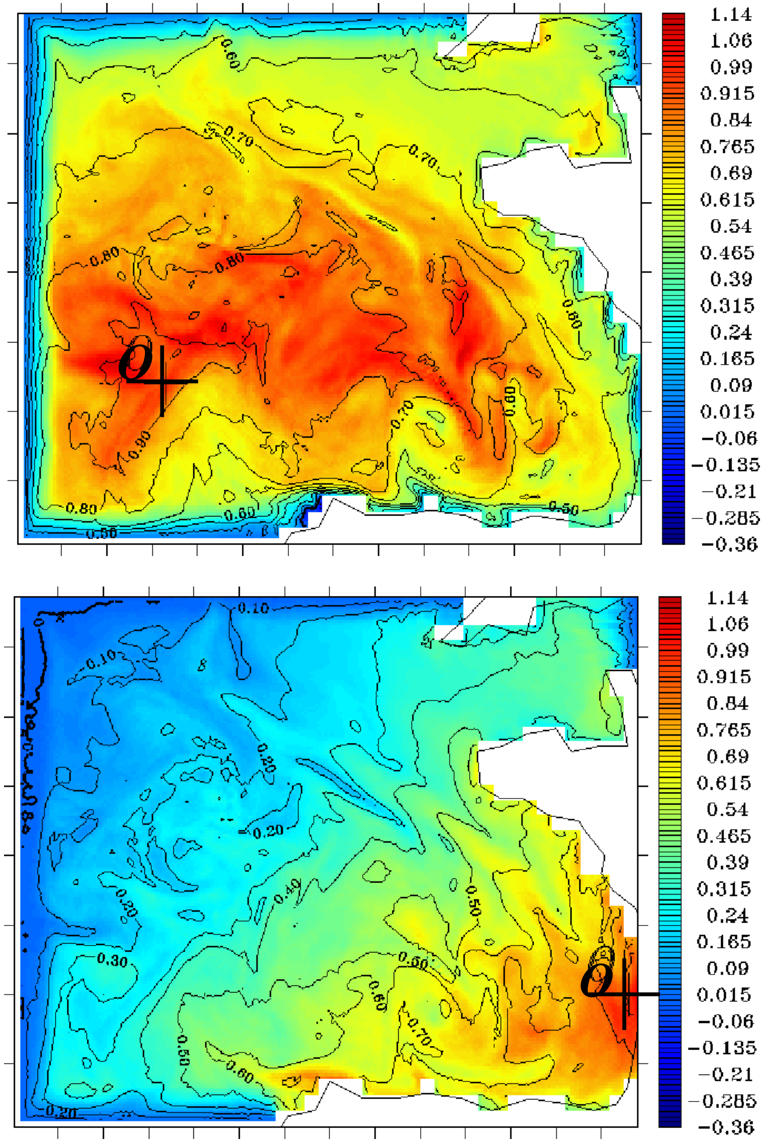
$$K = P^f H^T [H P^f H^T + R]^{-1}$$

$$x^a = x^f + K [y - H x^f]$$

If y scalar (1 obs) : $x^a = x^f + \lambda r$

Representer : $r = (P^f H^T)/(H P^f H^T)$

Representer for Q : $r_Q = (Q H^T)/(H Q H^T)$



SST field of representers of SST obs

- Their convergence with the sample size is used to check the ensemble **Q statistics convergence**
- They confirm the strong **anisotropy** / **inhomogeneity** of the dynamics
- They show through the use of **Q** in the KF a **limited influence of surface observations at depth** and a large spread of their influence in the mixed layer

Impact of the error parameterization for DA

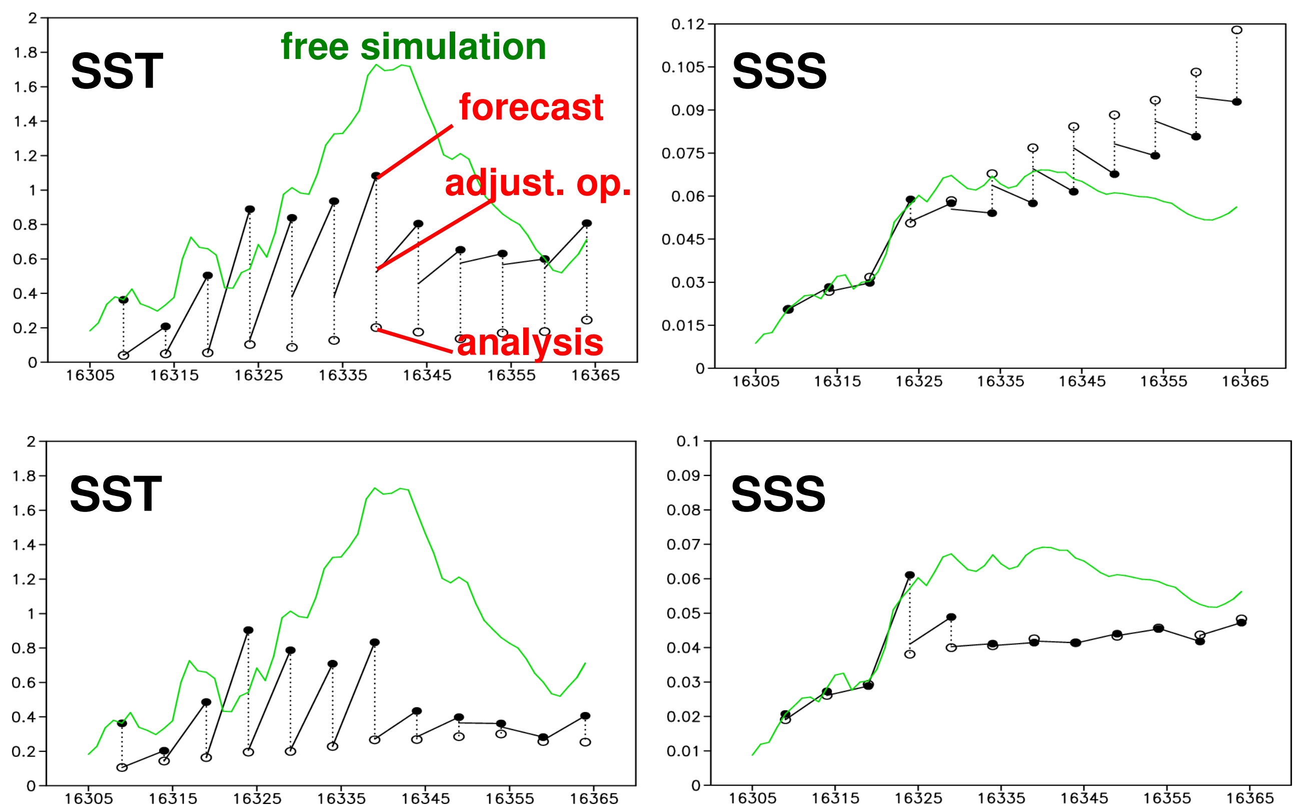
Twin experiments on 60 days, with atmospheric forcing perturbation: the **SEEK filter is applied to assimilate SST** on the perturbed simulation with different static $\mathbf{P}^f$:

- $\mathbf{P}^f$ = the control run (REF) **daily variability** on the 60 days (no order reduction): **EXP1**
- $\mathbf{P}^f$ = the 59 main EOFs of **Q estimated through the ensemble method**: **EXP2**

- In **EXP1**, the state is degraded in the non observed sub space / great impact of adjustment operator

- In **EXP2**, **improvements on the whole state space** / dynamically balanced adjustments: **weak impact of the adjustment operator** (both objectives are reached)

→ **the system demonstrates its sensitivity to the model error parameterization**



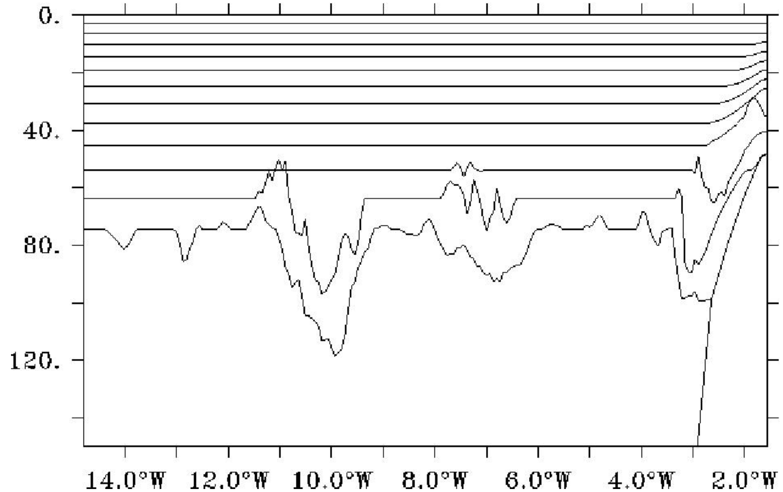
RMS error between REF and EXP1 (top) / EXP2 (bot.)

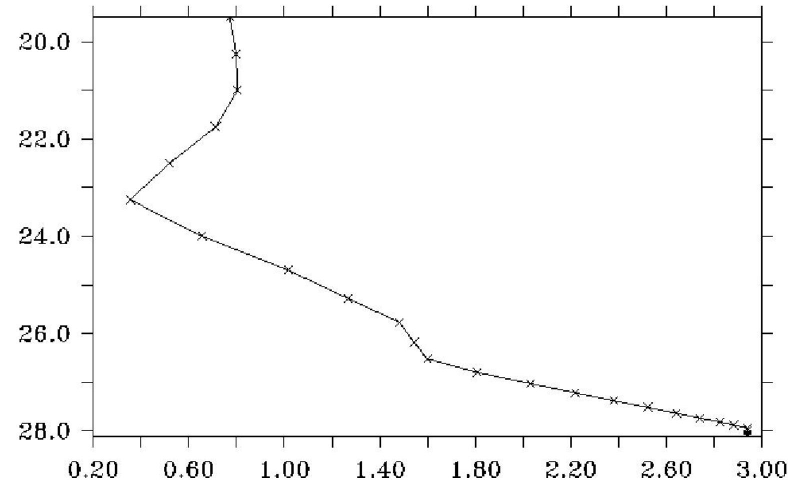
Gaussian vs TG observational update (1)

Identical twin experiments: in **EXP3**, the SEEK is applied to assimilate SST with the same $\mathbf{P}^f$ parameterization as in EXP1 and the **TG observational update** instead of the classical Gaussian update

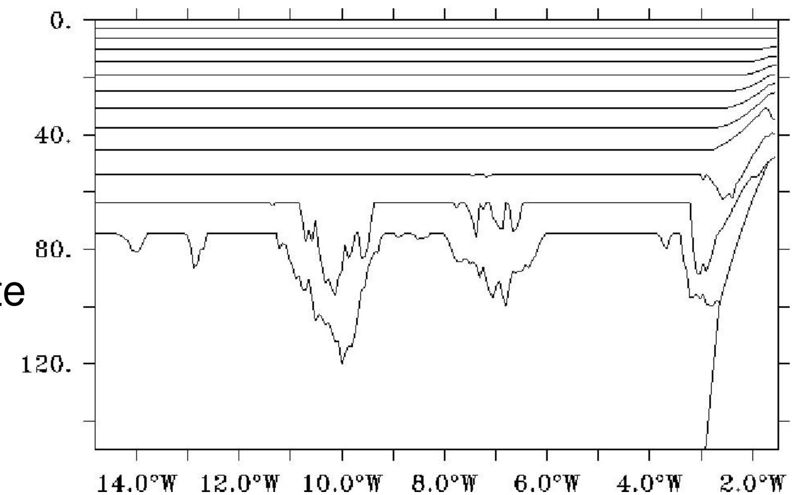
Zonal section (at 45.9N) of interface pressure	**Density profile belonging to that section (at 6.4W)**

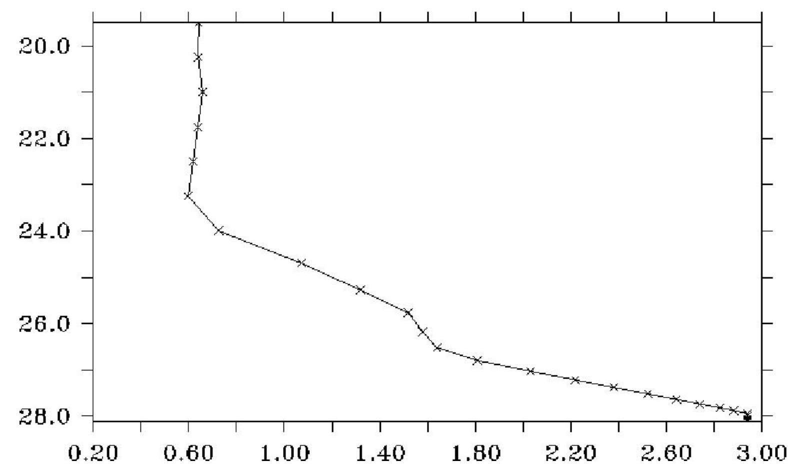
In **EXP1**, all constraints are violated: density inversions, negative layer thicknesses, *p* below minimum value





In **EXP3**, **no more density inversion, all layers have retrieved their minimum depth** (even in the s-coordinate region), and positive thickness.



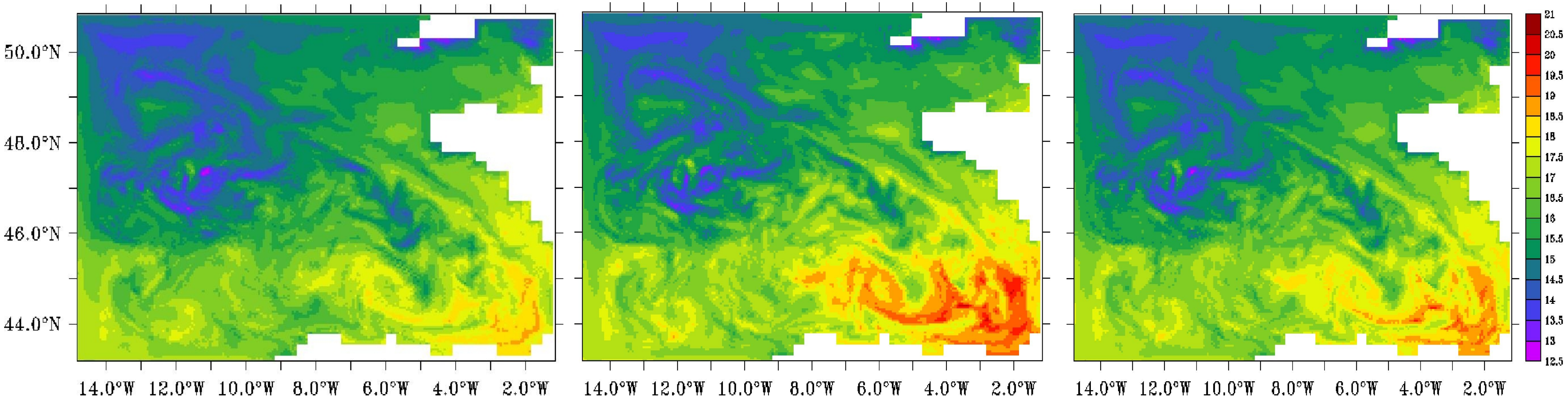


Gaussian vs TG observational update (2)

EXP1: SST field after Gaussian update

EXP1: SST field after Gaussian update and adjustment op.

EXP3: SST field after Trunc. Gauss. update



- The warming relative to the Gaussian update with adjustment operator or with TG update occurs above the first density inversion of the profile in **EXP1**

- For the full 3D temperature field, the RMS error is

  0.72 °C for the Gaussian estimate

  0.83 °C after the adjustment operation

  0.74 °C for the truncated Gaussian estimate.

With the TG observational update T values remain closer to the observations than with the adjustment operator

Conclusion

- Study of two **error parameterization** aspects

→ two different methods for **an improved application of the SEEK filter / DA systems**

- Both ensure :
  - a **statistically more optimal extrapolation** of the observation information
  - a **more physically consistent** solution
- Impact of **Q** realistic parameterization → **fundamental role of the error covariances parameterizations in DA system**
- **TG description** of error is sensible and more adapted

→ These studies have to be expanded :

- application of ensemble methods to estimate the **error model linked to other sources** + linear combination with realistic $\boldsymbol{P}_0$ for **realistic experiments**
- **full TG process** to be used on the real size problem of the Bay of Biscay

→ Combination and generalization of these procedures for other applications / other OGCM

Thank you !