Some recent advances in ocean data assimilation with the SEEK filter

Emmanuel Cosme, Jacques Verron, Frédéric Castruccio, Yann Ourmières,
Céline Robert #, Sergey Skachko, Eric Blayo #, Pierre Brasseur, Jean-Michel Brankart
LEGI, CNRS/INPG/UJF, Grenoble, France, except #: LMC, CNRS/INPG/UJF/INRIA, Grenoble, France.

Ocean data assimilation, Kalman filtering, and the SEEK filter

Ocean data assimilation (DA) aims at obtaining the best possible description of the ocean state using observations along with a ocean circulation model. It is probably the most sophisticated approach for systematic valorization of observational data. DA is used for various issues such as initialization (forecasting issue), reanalysis (for climate studies), models improvement: estimation of parameters, of external forcing, of tracer sources, and a posteriori characterization of the errors of the model, observations, and estimated states.

Kalman filtering (KF) is a four-dimensional DA method based on estimation theory, in contrast with the variational methods, based on control theory. In the KF approach, an estimate of the ocean state is alternately propagated with the dynamical model (equation 1 below) and corrected with available observations (equation 3). Simultaneously, an estimate of the state error statistics is also propagated (eq. 2) and updated (eq. 4). The corrections involve a gain matrix optimally computed in real time (eq. 5).

$$x^f_{i+1} = M_i x^a_i \quad \text{(eq. 1)}$$

$$P^f_{i+1} = M_i P^a_i M_i^T + Q_i \quad \text{(eq. 2)}$$

$$x^a_i = x^f_i + K_i (y_i - H_i x^f_i) \quad \text{(eq. 3)}$$

$$P^a_i = (I - K_i H_i) P^f_i \quad \text{(eq. 4)}$$

$$\text{avec } K_i = P^f_i H_i^T (H_i P^f_i H_i^T + R_i)^{-1} \quad \text{(eq. 5)}$$

The Singular Evolutive Extended Kalman (SEEK) filter is a Kalman filter designed for applications to large and non-linear systems such as an ocean model. Its main characteristics are:
- $\boldsymbol{P}^f$ is considered of low rank and represented in a reduced-size space. Very often, it is initialized with the first EOFs computed from a free model run (eq. 6 and 7);
- Due to model's non-linearity, the state error statistics are propagated using a method of state perturbation (eq. 8);
- The form of the observation error statistics R allows some handling leading to a matrix inversion in the reduced-size space only, instead of the observation space (eq. 9).

a very few dimensions : ~15-20

$$S^f = \overbrace{\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}}^{\text{model's dimension : } \sim 10^7} \quad \text{(eq. 6)}$$

$$P^f = S^f S^{fT} \quad \text{(eq. 7)}$$

$$(S^f_{i+1})_k = M\left(x^a_i + (S^a_i)_k\right) - M(x^a_i),\ k = 1,15-20 \quad \text{(eq. 8)}$$

$$K = S^f [I + (HS^f)^T R^{-1} (HS^f)]^{-1} (HS^f)^T R^{-1} \quad \text{(eq. 9)}$$

Incremental Analysis Updating (IAU)

A significant drawback of sequential methods is the time discontinuity of the solution resulting from intermittent corrections of the model state. To remove this drawback and obtain a smooth assimilation result, the IAU scheme (Bloom et al., 1996) was implemented in the SEEK filtering process. It works as illustrated by Figure 1: After a forecast step, an analysis correction is computed. The forecast step is then run again. In the course of this model integration, the previously computed correction is incrementally incorporated.
Figure 2 shows the time evolution of the temperature profile of a point from the Gulf Stream region over a month. When the IAU is activated (top), the solution is clearly smoother than when it is not (bottom).

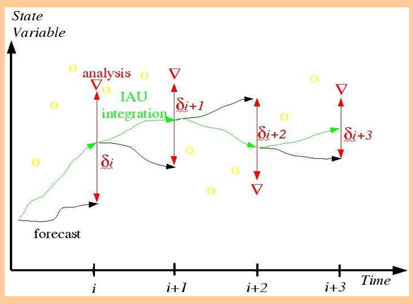


Figure 1 : Operating cycle of the IAU scheme.

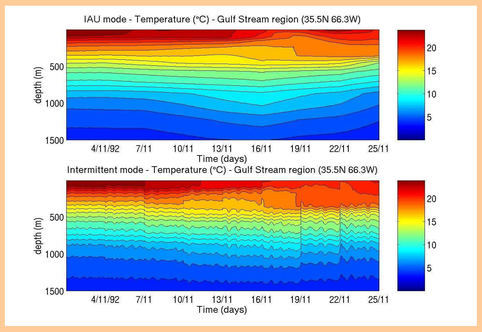


Figure 2 : Time evolution of the temperature profile at 35.5N, 66.3W with the "standard" SEEK filter (bottom) and the SEEK filter with the IAU (top).

Estimation of model parameters

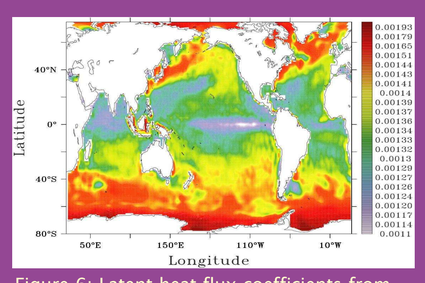


Figure 6: Latent heat flux coefficients from the "true" ocean.

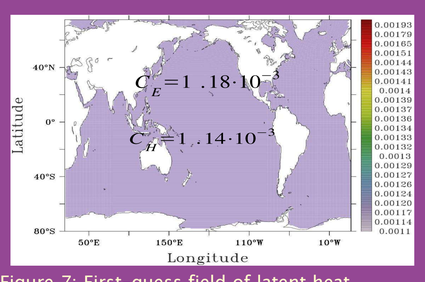


Figure 7: First-guess field of latent heat flux coefficient.

Here the SEEK filter is applied to estimate the time varying bulk coefficients of latent and sensible heat fluxes. The state vector, usually composed of horizontal current velocities, temperature and salinity, is augmented with the two bulk coefficients.
The method has been used in a twin experiments framework. A "true" ocean is provided by an independent model run (Figure 6) and a "working" ocean is initialized with constant bulk coefficients (Figure 7). Temperature and salinity observations are drawn from the former and assimilated in the latter.

Figure 8 depicts the RMS error values on latent heat flux coefficient. These errors rarely exceed 20% of the absolute coefficient values.

Heat fluxes resulting from the corrected heat flux coefficients tightly resemble those of the "true" case (Figure 9).

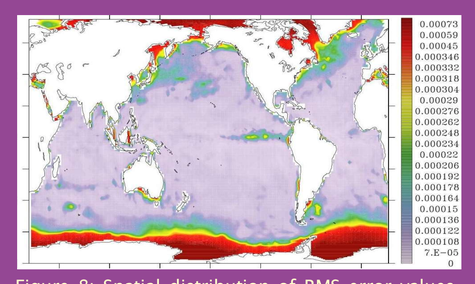


Figure 8: Spatial distribution of RMS error values on latent heat flux coefficient.

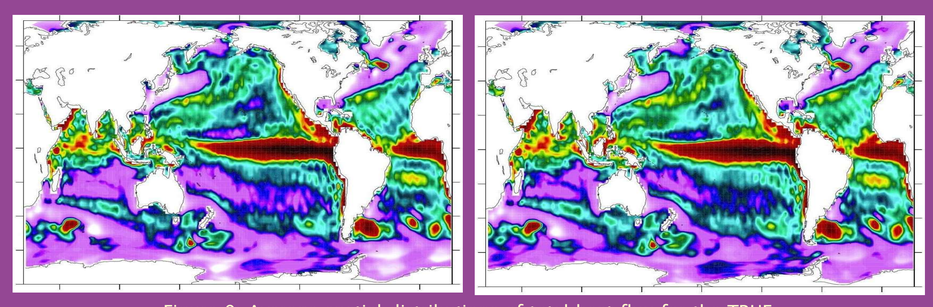
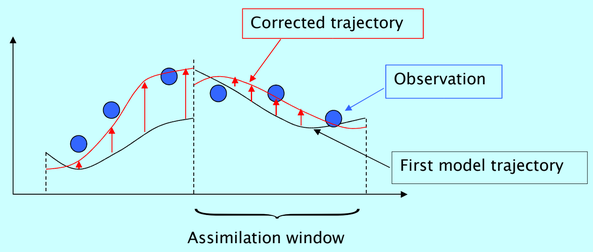


Figure 9: Average spatial distributions of total heat flux for the TRUE ocean (left) and the assimilation experiment (right).

Comparison and hybridization of 4D-VAR and the SEEK filter

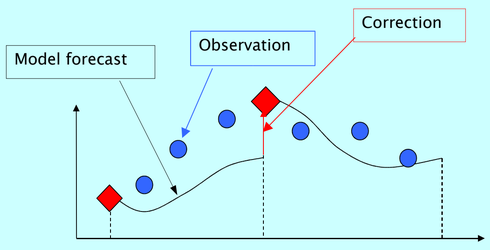
In 4D-VAR, the computation of the solution is based on the minimization of a cost function (below) composed of the "distance" from the solution state to the background state and the "distance" from the solution state to observations. Optimization is performed along a model trajectory over an "assimilation window":
- observations are assimilated at their real existence time;
- $\boldsymbol{P}$ matrix is kept fixed from one assimilation window to the next.

$$J(x) = (x^b - x)^T P^{-1} (x^b - x) + (y - Hx)^T R^{-1} (y - Hx)$$



With the SEEK filter, the assimilation solution is computed directly from the background state and the observations. This calculation is sequential:
- observations are assimilated at fixed times;
- $\boldsymbol{P}$ matrix is dynamically propagated by the model.

$$x = x^b + PH^T (HPH^T + R)^{-1} (y - Hx^b)$$



A reduced-order 4D-VAR (R-4D-VAR) has been developed and compared with the SEEK filter. Both the SEEK and 4D-VAR improve the representation of circulation of the free model.
Over the first 6 months, the SEEK and the reduced-order 4D-Var assimilation provide similar improvements of the free model (Figure 10a). Beyond that, the SEEK filter deviates (Figure 10b). This is due to the triggering of Tropical Instability Waves (TIW) in June. Their propagative nature is well anticipated by R-4D-VAR. (Figure 11).

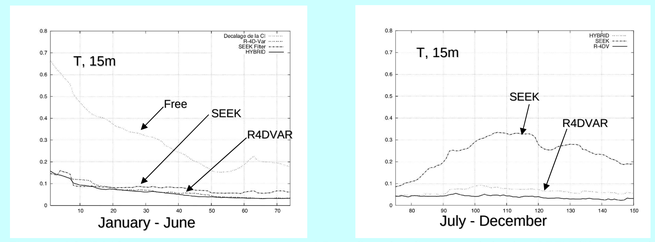


Figure 10: RMS error in temperature at 15 m for the free model, the SEEK filter, and R-4D-VAR at short term (left) and longer term (right).

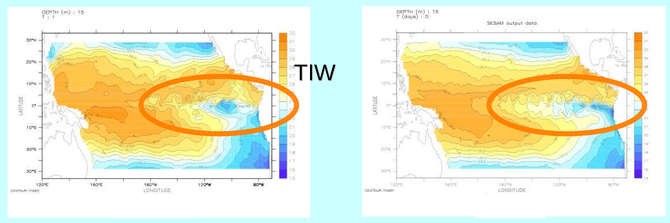


Figure 11: 15 m-temperature field from the assimilation experiments with the SEEK filter and R-4D-VAR, showing the difference in the representation of the TIW.

R-4D-VAR and the SEEK filter have been hybridized (P is propagated now). The hybrid looks generally better than R-4D-VAR and the SEEK filter (Figure 12), but not that much in comparison with R-4D-VAR. The ocean model used here is in a tropical configuration: the dynamics is close to be linear and well-suited to R-4D-VAR.

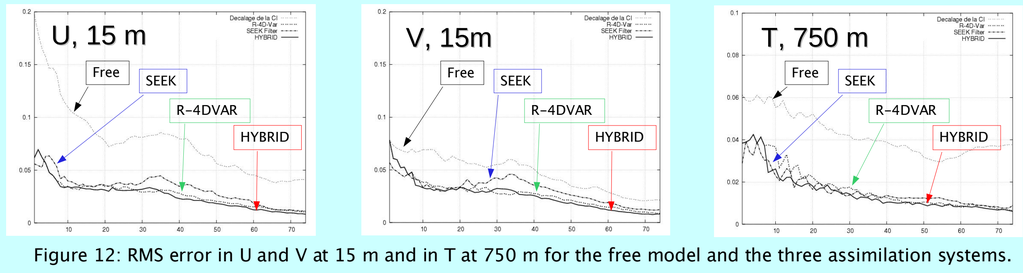


Figure 12: RMS error in U and V at 15 m and in T at 750 m for the free model and the three assimilation systems.

Assimilation of absolute Sea Surface Height and in situ observations

Parameters of satellites navigation are referenced to an ellipsoid and oceanic circulation is physically influenced by the geoid. This is, very briefly, why only the residual component of the measured altimetric signal can be used for assimilation in oceanography. This component, the Sea Level Anomaly (SLA), must be completed with the Mean Dynamic Topography (MDT), i.e. a mean circulation signal referenced to the geoid. So far, a synthetic MDT, typically computed from a free model run, was used. Now, some gravimetric missions (CHAMP, 2000; GRACE, 2002; GOCE, 2006) provide observation-based geoid data and allow the use of absolute altimetric signal.

Simultaneous assimilation of observations of different types can be impossible due to physical inconsistencies between the data. This defect is expected here, when a synthetic MDT is used to reference SLA. Figure 13 shows that the deviation of the MDT estimated from a model run (top) from the mean dynamic height estimated from TAO observations (top) is larger than the deviation of the MDT estimated from GRACE observations from the same data (bottom).

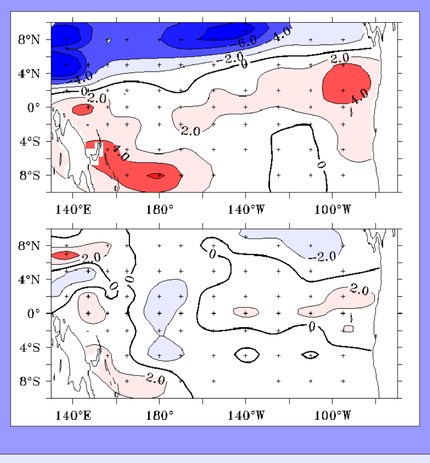


Figure 13: Difference (in cm) between the mean dynamic height estimated from TAO observations and model MDT (top), and MDT computed using GRACE geoid (bottom). The crosses symbolize TAO moorings.

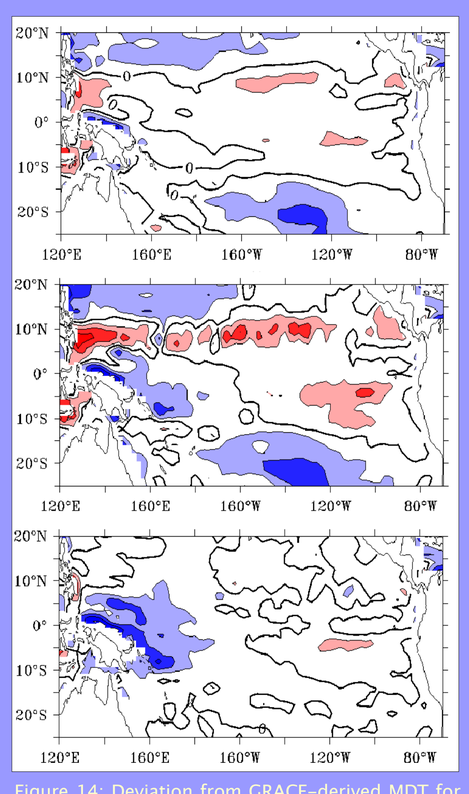


Figure 14: Deviation from GRACE-derived MDT for the free model (top), the model with assimilation of TAO and altimetric ERS data, using a model-based MDT (middle) or GRACE-derived MDT (bottom) as a reference.

Data inconsistency is illustrated on Figure 14. The misfit in MDT in the free model (top) is lower than in the model constrained with assimilation of TAO and altimetric (ERS) observations, using a model-based MDT (middle). When GRACE-derived MDT is used, the misfit is further reduced (bottom). This misfit is not zero because TAO observations brings a MDT signal partially inconsistent with GRACE.

The three runs cited previously are compared with independent temperature observations from XBT profiles acquired over one year in the tropics (Figure 15). The model forced with assimilation of TAO and ERS observations using the model-based MDT (blue) performs better than the free model (green). The use of GRACE-derived MDT further improves model results.

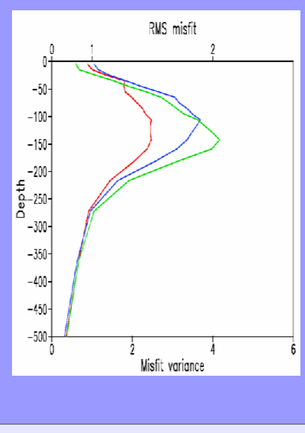


Figure 15: XBT temperature RMS differences for the free model (green), the model with assimilation using the model-based MDT (blue), and using GRACE-derived MDT (red).