

## Data assimilation for marine monitoring and prediction: The MERCATOR operational assimilation systems and the MERSEA developments

By P. BRASSEUR\*<sup>1,2</sup>, P. BAHUREL<sup>2</sup>, L. BERTINO<sup>3</sup>, F. BIROL<sup>2</sup>, J.-M. BRANKART<sup>1</sup>, N. FERRY<sup>2</sup>, S. LOSA<sup>4</sup>, E. REMY<sup>2</sup>, J. SCHRÖTER<sup>4</sup>, S. SKACHKO<sup>1</sup>, C.-E. TESTUT<sup>2</sup>, B. TRANCHANT<sup>2</sup>, P. J. VAN LEEUWEN<sup>5</sup> and J. VERRON<sup>1</sup>

<sup>1</sup>Laboratoire des Ecoulements Géophysiques et Industriels, Grenoble, France

<sup>2</sup>MERCATOR-Océan, Ramonville Saint Agne, France

<sup>3</sup>Mohn-Sverdrup Center/Nansen Environmental and Remote Sensing Center, Bergen, Norway

<sup>4</sup>Alfred Wegener Institute, Bremerhaven, Germany

<sup>5</sup>Instituut voor Marien en Atmosferisch Onderzoek, Utrecht University, the Netherlands

(Received 4 July 2005; revised 3 April 2006)

### SUMMARY

During the past 15 years, a number of initiatives have been undertaken at national level to develop ocean forecasting systems operating at regional and/or global scales. The co-ordination between these efforts has been organized internationally through the Global Ocean Data Assimilation Experiment (GODAE). The French MERCATOR project is one of the leading participants in GODAE. The MERCATOR systems routinely assimilate a variety of observations such as multi-satellite altimeter data, sea-surface temperature and *in situ* temperature and salinity profiles, focusing on high-resolution scales of the ocean dynamics.

The assimilation strategy in MERCATOR is based on a hierarchy of methods of increasing sophistication including optimal interpolation, Kalman filtering and variational methods, which are progressively deployed through the Système d'Assimilation MERCATOR (SAM) series. SAM-1 is based on a reduced-order optimal interpolation which can be operated using 'altimetry-only' or 'multi-data' set-ups; it relies on the concept of separability, assuming that the correlations can be separated into a product of horizontal and vertical contributions. The second release, SAM-2, is being developed to include new features from the singular evolutive extended Kalman (SEEK) filter, such as three-dimensional, multivariate error modes and adaptivity schemes. The third one, SAM-3, considers variational methods such as the incremental four-dimensional variational algorithm.

Most operational forecasting systems evaluated during GODAE are based on least-squares statistical estimation assuming Gaussian errors. In the framework of the EU MERSEA (Marine Environment and Security for the European Area) project, research is being conducted to prepare the next-generation operational ocean monitoring and forecasting systems. The research effort will explore nonlinear assimilation formulations to overcome limitations of the current systems. This paper provides an overview of the developments conducted in MERSEA with the SEEK filter, the Ensemble Kalman filter and the sequential importance re-sampling filter.

KEYWORDS: GODAE Operational ocean forecasting

### 1. INTRODUCTION

Operational oceanography is an emerging field of activities that can be defined as the process of systematic and real-time monitoring and prediction of the state of oceans and coastal seas (including living resources) in a way that will promote and engender wide utility and availability of this information for maximum benefit to the community (Verron and Chassignet 2006). The scientific and technological feasibility of operational ocean forecasting systems results from several factors.

(i) *The maturation of numerical models and computing techniques available to simulate the ocean circulation.* Substantial progress has been achieved especially in terms of: model formulations, discretization techniques, numerical schemes, parametrization of subgrid-scale processes, coupling with the overlying atmosphere and sea ice etc. (Griffies 2006). In addition, the growth of resources for high-performance computing has permitted an increase of the spatial resolution to such an extent that, today, basin-scale models are able to resolve the mesoscale dynamics explicitly.

\* Corresponding author: Laboratoire des Ecoulements Géophysiques et Industriels, BP 53, 38041 Grenoble Cedex 9, France. e-mail: pierre.brasseur@hmg.inpg.fr

© Royal Meteorological Society, 2005.

(ii) *A concerted international effort to establish global ocean observing systems.* During the past decade, major international programs have taken place (e.g. the World Ocean Circulation Experiment) with the objective to better characterize the state of the global ocean. The contribution of satellite altimetry to that effort has been critical, providing for the first time continuous and accurate observation of the surface signature of the ocean dynamics at global scale (Fu and Cazenave 2001). Today, the deployment of ARGO floats covering the major ocean basins provides additional information about the sub-surface ocean properties (Send 2006).

(iii) *The progress achieved in data assimilation techniques.* The first application of data assimilation in oceanography dates back to the 1980s, involving simple ocean models. Since then, the theoretical framework of data assimilation has been progressively adapted to meet the specific requirements of more sophisticated ocean models and observational datasets. Major advances have been achieved to:

- Adapt the assimilation algorithms to ocean systems of very large dimensions;
- Develop sophisticated representations of the error statistics with primitive-equation models;
- Implement multivariate algorithms for assimilating several data types simultaneously; and
- Extend the theoretical framework from linear to nonlinear problems (e.g. Verron 1992; Blayo *et al.* 1997; De Mey 1997; Fukumori 2001; Brusdal *et al.* 2003; Evensen 2003).

As a consequence of this overall progress, ocean forecasting in the meteorological sense is becoming a reality. A number of initiatives have been undertaken at national level to develop ocean forecasting systems operating at regional and/or global scales, e.g. MERCATOR in France, FOAM in the United Kingdom and MFS in Italy (Crosnier and Le Provost 2006); TOPAZ in Norway, HYCOM (Crosnier and Le Provost 2006) and ECCO (Fukumori 2006) in the United States, and BLUElink in Australia (Schiller and Smith 2006). One of the main objectives during the GODAE project (Smith 2006) has been to compare the assimilation systems implemented operationally and evaluate their overall performances in real-time conditions.

The French MERCATOR project (Bahurel 2006) is one of the leading participants in GODAE. The project was launched in 1995 by the major French agencies involved in oceanography. The MERCATOR system is based on several components: the ocean model, the surface forcing fields, the remotely sensed data (e.g. sea surface temperature (SST) and altimetric data) and *in situ* observations (e.g. temperature and salinity profiles). These various components are integrated through an assimilation system, the objective being to provide the best possible description of the real ocean. The assimilation methods are all derived from least-squares estimation principles. In the first part of this paper, the suite of assimilation tools implemented in the MERCATOR system is reviewed and illustrated.

In the future, MERCATOR will be extended to the European scale through the EU MERSEA (Marine EnviRonment and Security for the European Area) project. The overarching objective of MERSEA is to develop a single European system for global monitoring and forecasting of the ocean with a co-ordinated network of regional systems for European waters. This integrated system will form the ocean component of the future Global Monitoring for Environment and Security system. The high-resolution global ocean forecasting system to be shared by the European partners will be inherited mostly from the MERCATOR experience, while FOAM will be focused on the European coastal area, TOPAZ on the Arctic ocean and MFS on the Mediterranean Sea.

In the second part of the paper, we discuss the research effort in data assimilation carried out in MERSEA to achieve this integration. Following the pathway taken by numerical weather prediction (NWP) organizations, it is believed that a strong investment in research and development is necessary to support operational oceanography in the long-term, with strong beneficial feedback on the research side.

An overview of the MERCATOR prototypes in use today is presented in section 2, while the suite of assimilation schemes is presented in section 3. The research directions that are being investigated in the MERSEA project to build the future integrated system are described in section 4, and conclusions follow in section 5.

## 2. THE MERCATOR OPERATIONAL SYSTEM

Operational ocean prediction systems are being developed with a variety of objectives in mind, such as: ocean current hindcasting and short-range forecasting; monitoring and prediction of properties of the surface layer; estimation of the thermodynamic state of the ocean for seasonal and climate predictions; production of retrospective analyses of the changing ocean; and representation of the background physical environment which is critical to the functioning of marine ecosystems.

The representation of the ocean state at eddy-resolving resolution is necessary to meet the requirements of end-users, but this is also fully justified from a scientific point of view. The dominant energetic activity of the mesoscale ocean, its non-deterministic nature and the interactions with the large-scale circulation are now well recognized properties, requiring sophisticated numerical models and assimilation methods that make the best use of sparse observations. To produce reliable forecasts, the models must be initialized with conditions that represent as accurately as possible the actual state of the ocean at eddy-resolving scales. Fortunately, the arrival of satellite observations has played a pivotal role in the development of operational oceanography, providing the observational basis needed to respond appropriately to the 'high-resolution challenge'.

The long-term goal of MERCATOR is, therefore, to provide a three dimensional (3D) description of the global ocean dynamics and thermohaline circulation in terms of temperature, salinity, currents and sea-surface elevation, with the highest possible resolution in space and time permitted by the observing systems and the computational resources. In the future, an additional objective will be to run a coupled physical-biogeochemical model in order to continuously monitor the biogeochemical and ecosystem properties of the ocean.

The MERCATOR prototypes available today consist of an ocean circulation model, an assimilation system and different datasets that are assimilated routinely. The ocean model is based on the rigid-lid version of the OPA-NEMO primitive-equation model developed at the LOCEAN laboratory (Madec *et al.* 1998), and the transition to the free-surface version of the ocean code is underway. Surface forcings consist of daily fields of wind stress, evaporation, precipitation, non-solar and solar heat fluxes provided by analyses and forecasts from the European Centre for Medium-range Weather Forecasts. The surface forcing includes a retroaction term in the net heat flux, based on the difference between the model SST and the weekly Reynolds SST product, in order to reflect the coupling between the ocean and the atmosphere. The main river outflows are represented by an input of fresh water at the river mouth given by the climatological monthly database from UNESCO (Vörösmarty *et al.* 1996). A realistic topography is adapted to the resolution of the model, based on the bathymetric database produced by Smith and Sandwell (1997).

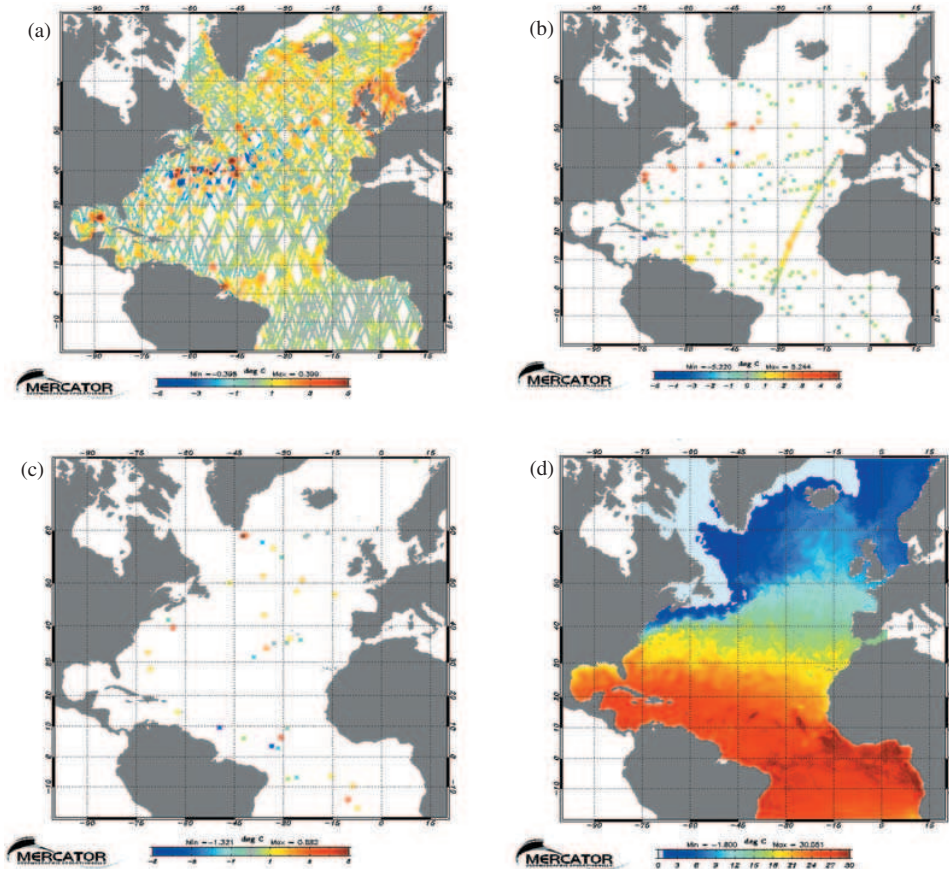


Figure 1. Different kinds of assimilated data for the first week of January 2002, and resulting sea surface temperatures as analysed by the MERCATOR North Atlantic ( $1/3^\circ$ ) prototype: (a) along-track sea-level height anomaly data (from satellites Jason1, ERS2, GFO); (b) *in situ* temperature vertical profiles; (c) salinity vertical profiles; and (d) resulting sea surface temperature analysis.

Input data for the MERCATOR system include *in situ* as well as remotely sensed observations which are used for several applications: forcing, data assimilation, model verification and validation. Here we focus on the data used for assimilation. Figure 1 shows an example of data (observations of sea-level anomalies, temperature and salinity profiles) assimilated over a 7-day window and the resulting model state after assimilation. One can see the high spatial coverage of the altimetric data (three satellites are available, see Fig. 1(a)) and the relative scarcity of *in situ* data, especially salinity profiles (Fig. 1(c)). The continuous line in Fig. 1(b) corresponds to expendable bathythermograph (XBT) temperature profiles collected along shipping tracks. The analysed SST field in Fig. 1(d), which incorporates information coming from both the model integration and the available observations, shows important mesoscale structures.

The system components are assembled into a hierarchy of prototypes; at present three prototypes are being used operationally:

- The first prototype (denoted PSY1) covers the North and equatorial Atlantic with an intermediate resolution of  $1/3^\circ$  and 43 vertical levels. The grid spacing is stretched

in the vertical, with the resolution changing from 12 m at the sea surface to 200 m at the bottom. Every week, this model assimilates vertical temperature/salinity profiles in addition to satellite data from AVHRR\* SST and radar altimetry. The Mediterranean Sea is not explicitly included in the model domain, but its impact on the North Atlantic basin is taken into account through a buffer zone in the Gibraltar Strait and Alboran Sea where a relaxation to climatology is applied.

- The second prototype (PSY2) is a high-resolution model (5 to 7 km horizontal resolution) covering the North Atlantic basin from 9 to 70°N, and including the Mediterranean Sea. There are 43 vertical levels with the resolution changing from 6 m at the surface to 300 m at the bottom of the Atlantic basin. This configuration is more specifically focused on mesoscale processes (Drillet *et al.* 2005). It is intended to provide boundary conditions for coastal modelling in European seas. Until recently only altimetric data were assimilated in PSY2 but the system has been upgraded in late 2005 for incorporation of *in situ* data, still on a weekly basis. The next version of the high-resolution Atlantic model will be developed on the same grid as the future global high-resolution model (1/12°) with: a free surface, a partial-step vertical coordinate, atmospheric bulk formulae, and a sea ice model (Garric and Charpentier 2005; Timmermann *et al.* 2005). In the future, this system is intended to produce ocean analyses every day in order to better meet the requirements of end users.

- The third prototype (PSY3) is a global ocean configuration with 31 levels in the vertical. Today, PSY3 assimilates altimetric data only, but the extension to *in situ* temperature and salinity data is ongoing. There are 21 levels located in the top 1000 m of the water column, and the thickness of the levels varies from 10 m at the surface (within the first 100 m) to 500 m below the 3000 m level. The initial version of the global prototype had a horizontal resolution of 2°, but an upgrade was achieved in late 2005 with a new global version at 1/4° resolution. This prototype will provide up-to-date oceanic initial conditions for seasonal climate predictions every month.

Figure 2 illustrates several results obtained from these models, and the typical oceanic structures present in the numerical model outputs.

### 3. ASSIMILATION SCHEMES IN THE MERCATOR SYSTEM

#### (a) *Incremental approach*

MERCATOR is developing a suite of assimilation tools of increasing complexity (denoted as Système d'Assimilation MERCATOR or SAM) ranging from pragmatic sequential schemes to variational methods. This incremental approach was adopted at the start of the project as a compromise between the operational needs for high-resolution products and the available computational resources. A few years ago, the feasibility of sequential methods based on optimal interpolation (OI) to assimilate altimeter data with mesoscale ocean dynamics was demonstrated (De Mey 1997). In the long-term, high-performance computing resources are expected to increase further, and research and development activities are being conducted in anticipation of the introduction of more advanced assimilation techniques, such as 4D variational assimilation (4D-Var).

The MERCATOR assimilation tools in place today have their roots in the theoretical framework of least-squares statistical estimation. The first release, SAM-1, has been developed from an OI scheme; SAM-1 has been running on an operational real-time basis since early 2001. The second release, SAM-2, is considering a singular extended evolutive Kalman (SEEK) filter analysis method; it has been evaluated and compared to

\* Advanced Very High Resolution Radiometer.

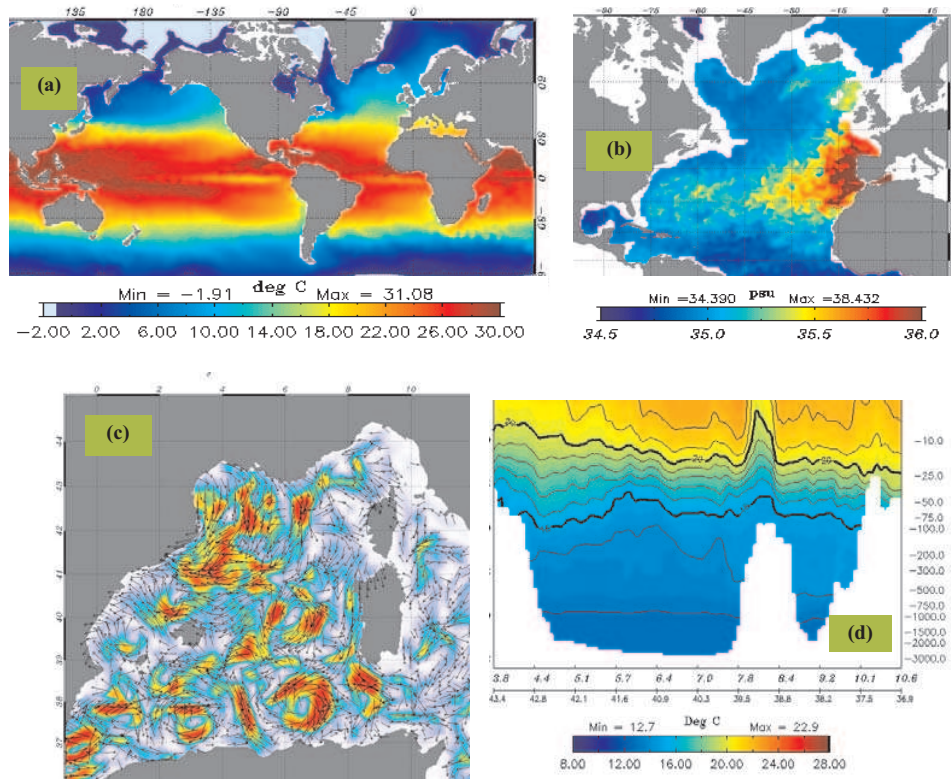


Figure 2. Examples of MERCATOR ocean system outputs from prototypes PSY1, PSY2 and PSY3 (see text) all computed on 1 June 1995: (a) global ocean sea surface temperature field, real-time analysis 1 June 2005—from PSY3; (b) North and Tropical Atlantic ( $1/3^\circ$  model) 1000 m depth salinity, 2-week forecast verifying 15 June 2005—from PSY1; (c) Mediterranean Sea (5–7 km model,  $1/16^\circ$ ) surface currents—from PSY2, and (d) vertical section of sea temperature between Sète and Tunis, 2-week forecast verifying 15 June 2005. See text for further details.

SAM-1 in several hindcast experiments and will be integrated soon into the operational system. The third release, SAM-3, targets more advanced approaches such as 4D-Var, and is still under construction in research and development mode.

#### (b) SAM-1

SAM-1 is based on the reduced order OI method developed by De Mey and Benkiran (2002) and Demirov *et al.* (2003). Two different versions of SAM-1 are implemented in the MERCATOR prototypes. Version 1 (SAM-1v1) employs a vertical extrapolation scheme based on the Cooper and Haines (1996) method to assimilate observations of sea-level anomalies (SLA). The assimilation scheme is described in detail in Ferry *et al.* (2006). The algorithm starts by calculating a SLA increment from ‘innovations’ collected over a one-week cycle (i.e. differences between along-track measurements and the model forecast at the corresponding time within the cycle) using a background-error covariance that is constant in time. The SLA increment is partitioned into a baroclinic and a barotropic contribution, in proportions that are determined according to the behaviour of the model for the period and the location under consideration (Ferry *et al.* 2006). This is achieved by estimating for each water

column how much the forecast dynamic height ( $\eta_{\text{bclinic}}$ ) and the forecast barotropic height ( $\eta_{\text{btropic}}$ ) contribute to the model's forecast sea surface height:  $\eta_{\text{bclinic}}$  is calculated as a vertical density anomaly integration from the bottom to the surface, and  $\eta_{\text{btropic}}$  is calculated away from the shelves (for depths  $>300$  m) by using the approximation  $\eta_{\text{btropic}} = f\psi/(gH)$ , where  $\psi$  is the model barotropic stream function,  $H$  is the ocean depth,  $f$  is the Coriolis parameter and  $g$  is the acceleration due to gravity. The baroclinic (barotropic) partition of the SLA is based on a linear regression between  $\eta_{\text{bclinic}}$  ( $\eta_{\text{btropic}}$ ) and the model SLA, which is achieved using a collection of oceanic states. This collection of oceanic states (13 weekly fields for  $\eta_{\text{bclinic}}$ ,  $\eta_{\text{btropic}}$  and SLA collected during the preceding 3 months) is continually updated during the simulation, which allows the statistics to evolve with time. The linear regression determines the average  $\eta'_{\text{bclinic}}/\text{SLA}$  and  $\eta'_{\text{btropic}}/\text{SLA}$  ratios used to split the SLA increment into its respective baroclinic and barotropic parts. The results indicate that this partition of the SLA increment is time- and space-dependent.

The barotropic component is then converted into an increment of horizontal velocity and barotropic stream function of the model, while the baroclinic part is used to modify the thermohaline structure of the water column by lifting or lowering the isopycnals with the Cooper and Haines (1996) technique. This scheme relies on the idea that, in oceanic regimes where baroclinicity dominates, sea level is closely related to the depth of isopycnals: the deeper the isopycnals, the higher the sea level. Conceptually, if the density changes as a result of temperature increase, the warming water (which is then less dense and takes up more room than cold water) produces a higher SLA and vice versa.

SAM-1v1 has been implemented in the PSY2 and PSY3 prototypes. To illustrate the impact of SLA data on the oceanic state, Fig. 3 represents the sea level and isopycnals along the equatorial Pacific between 180 and 140°W on 22 June 1993, before and after assimilation of altimetric data. The figure actually shows that the displacement of isopycnals is negatively correlated to the sea-level change: the assimilation induces an upward motion of isopycnals in the central and eastern regions, and a downward motion in the western region. In terms of temperature, an upward isopycnal motion is equivalent to raising the isotherms, i.e. to cooling the water at a fixed depth. By decreasing (increasing) the surface elevation, the ocean heat content is therefore decreased (increased).

Although this approach gives satisfactory results for assimilating SLA data in subtropical and midlatitude regions, the method cannot easily be extended to assimilate other observations, such as SST or *in situ* temperature and salinity profiles, with a primitive-equation model. In addition, the modelling error is not necessarily distributed in the vertical as an isopycnal mode, and more complex error structures must be considered. A fully multivariate, multi-data version of the OI scheme was therefore developed to assimilate simultaneously *in situ* temperature and salinity profiles and along-track altimetric data.

Version 2 (SAM-1v2) of the OI assimilation tool is based on the formulation described in De Mey and Benkiran (2002). It has been running operationally since January 2004 in the PSY1 North Atlantic prototype at 1/3°, and was transposed to the PSY2 prototype in late 2005. The background-error covariance is still constant, but the algorithm performs a reduced-order OI by means of multivariate vertical empirical orthogonal functions (EOFs) of barotropic stream function, and vertical temperature and salinity profiles, computed from a priori model simulation. The update of the model state is then expressed as a sum of contributions from each EOF weighted by the product of the innovation and the Kalman gain in the reduced space. The number and shape of the vertical modes permitted by this approach can be adjusted regionally

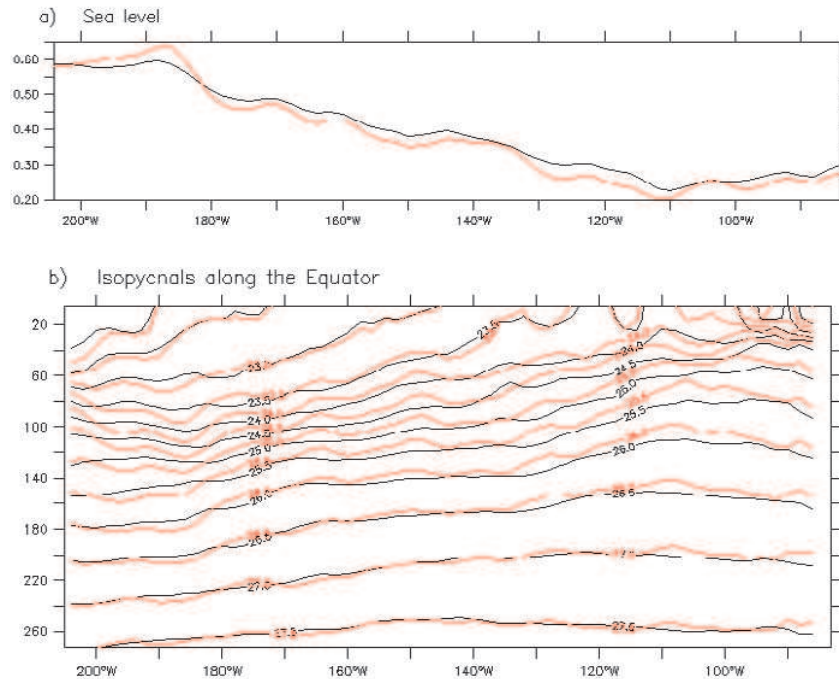


Figure 3. (a) Sea level (m), and (b) isopycnals (water density minus 1000, in  $\text{kg m}^{-3}$ ) at the equator in the Pacific for 23 June 1993, as simulated by the MERCATOR global ocean analysis system. The model state before (after) assimilation is represented by black (red) lines.

and seasonally. This option has been implemented in PSY1 and PSY2, leading to a better representation of the local physics than with the uni-modal scheme implemented in SAM-1v1. A variant of this multivariate scheme has also been implemented in the MFSPP (Mediterranean Forecasting System Pilot Project) model of the Mediterranean Sea (Demirov *et al.* 2003), showing its capacity to accommodate a fairly wide spectrum of surface and sub-surface dynamical regimes.

### (c) SAM-2

By construction, the SAM-1 algorithm assumes that the correlations can be separated into horizontal and vertical correlations. The concept of separability is related to the predominant role of stratification, and the very different scales involved horizontally and vertically in the open ocean. However, different behaviours can be expected near boundaries and coastal regions (e.g. Echevin *et al.* 2000). These limitations motivated the development of a more advanced assimilation scheme, SAM-2, that does not require the hypothesis of horizontal/vertical separability.

The SAM-2 algorithm is inherited from the analysis scheme of the SEEK filter, which is a reduced-order Kalman filter introduced by Pham *et al.* (1998) in the context of ocean circulation models. The error statistics of the SEEK filter are represented in a sub-space spanned by a small number of dominant error directions. The formulation of the assimilation algorithm relies on a low-rank background-error covariance matrix, which makes the calculations tractable even with state vectors of very large dimension.



Several strategies can be adopted to initialize the vectors of the reduced basis. A method involving the computation of EOFs obtained from prior simulations (without assimilation) has been applied in the majority of case-studies; this approach leads to corrections of the model trajectory that are multivariate and consistent with the dynamics of primitive-equation models. The extrapolation of the data from observed to non-observed variables is performed along the directions represented by these error modes which connect all dynamical variables and grid points of the numerical domain. The 3D modal representation for the error statistics is intended to overcome some of the limitations of SAM-1v2 in anisotropic and non-separable regions of the world ocean, such as shallow areas or in the surface layers (Testut *et al.* 2003).

Unlike the original SEEK filter, the variant of SAM-2 considered here does not evolve the error statistics according to the model dynamics (Ballabrera-Poy *et al.* 2001). This would require prohibitive costs given the size of the operational systems. However, some form of evolution of the background error is taken into account by considering different error sub-spaces for the four seasons. The error modes can be computed using different techniques: (i) EOFs of model states extracted from a prior simulation (without assimilation); (ii) EOFs of system states extracted from a prior hindcast experiment (obtained, for instance, with SAM-1); or (iii) EOFs of the system tendencies that occur over weekly cycles. These various approaches are currently being tested by investigating their capacity to control the model trajectory with the PSY1 configuration.

An issue of practical interest is how to suppress spurious correlations with distant variables, that may appear as a result of the truncation of EOFs. In order to prevent the data from exerting an artificial influence at remote distances through large-scale signatures in the EOFs, a simplification of the analysis scheme has been adopted by setting to zero the error covariances between distant variables which are believed to be uncorrelated in the real ocean. Previous experiments with the SEEK filter have shown that the local representation of the error sub-space is particularly effective for capturing the mesoscale features of the turbulent ocean (e.g. Penduff *et al.* 2002; Testut *et al.* 2003; Birol *et al.* 2005). This simplification is implemented in SAM-2 by assuming that distant observations have negligible influence on the analysis. The global system is split into sub-systems, and for each of these the traditional analysis is computed. Only data points located within individual regions, centred on a sub-domain of one or several grid points to be updated, actually contribute to the gain. This approach can be understood as a tuning of the observation operator according to the sub-domain in question. The size of the regions is determined in such a way that the distribution of the observations available on the model domain always provides at least a few data points within each region of influence. The typical size of the regions of influence in the MERCATOR prototypes extends from about 200 to 500 km, i.e. several Rossby radii of deformation.

The analysis step of the conventional Kalman filter is reformulated to take advantage of the low-rank approximation, leading to more efficient inversions of the data in the reduced space than in observation space (Brasseur 2006). To minimize the computational requirements, the analysis kernel in SAM-2 has been massively parallelized and integrated in a generic platform hosting the SAM-1 and SAM-2 kernel families. This platform provides a technological capacity to extend the dimension of the error sub-space up to several hundred modes (typically 200 with the MERCATOR prototype configurations).

Hindcast experiments have been conducted to evaluate the performance of SAM-1 with respect to SAM-2 in a similar set-up; these have demonstrated the overall superiority of SAM-2. Significant improvements have been reported with SAM-2 in most oceanic regions. For instance, coastal regions are now better represented, in particular

across the shelf. Biases in the equatorial regions have been much reduced with SAM-2, and the occurrence of spurious equatorial upwelling patterns, which had been noticed with SAM-1, has been suppressed. As a result, SAM-2 is planned to supersede SAM-1 in all high-resolution prototypes. For example, the multivariate global system at  $1/4^\circ$  resolution now under construction will be based on SAM-2, with an operational start in late 2006 or early 2007.

As an illustration of the spatial properties of the background errors used in SAM-2, Fig. 4 shows the ‘representer’ function as defined by Echevin *et al.* (2000) for the SST in two different regions: in the equatorial Pacific at  $0^\circ\text{N}$ ,  $140^\circ\text{W}$  (PSY3 configuration) and near the eastern Florida coast at  $27^\circ\text{N}$ ,  $80^\circ\text{W}$  in the subtropical Atlantic (PSY1 configuration). The representer field is based on error statistics reflected in the EOFs, and indicates how the model state is influenced by a single observation. In Fig. 4(a), the impact of this virtual SST observation extends eastwards and westwards about  $\pm 10^\circ$  of longitude, but also  $1^\circ$  northwards and southwards. The shape of the representer is slightly anisotropic zonally, reflecting the behaviour of equatorial dynamics. In Fig. 4(b), the representer function exhibits a structure influenced by the Gulf Stream current, which flows eastwards out of the Caribbean Sea and circulates northwards following the American coast. The largest values of the representers are located near  $27^\circ\text{N}$ ,  $80^\circ\text{W}$  with a meridional structure. Values greater than  $0.5\text{ degC}$  can be also found west of Florida, which may indicate that SST changes could also be due to surface heat fluxes at larger scale in this particular area.

The SAM-2 scheme has been tested in the PSY1 eddy-permitting North Atlantic configuration, assimilating a multivariate set of observations (along track altimetry, *in situ* temperature and salinity profile data, and SSTs). The estimation state vector includes temperature, salinity and barotropic stream function, and a geostrophic adjustment is performed after each analysis step to extend the correction to the whole model state. Several hindcast experiments were conducted during 2003 to validate the method with independent (non-assimilated) *in situ* temperature data profiles. The distribution of the profiles covers all regions of the North Atlantic basin, but the best data coverage is achieved in the zonal band between  $20$  and  $50^\circ\text{N}$ .

The skill of the assimilation is illustrated in Fig. 5, which shows the vertical distribution of the misfit variance computed between the validation profiles and the climatology, the run without assimilation and the hindcast experiment. The model simulation without assimilation competes reasonably well with the climatology, and improves the fit to the data in the top 70 m only. The assimilation improves the temperature field at all depths, with a significant reduction of the error in the thermocline.

The performance of the assimilation in correcting non-observed variables such as the currents has also been verified. Figure 6 shows the annual mean surface current computed from the hindcast experiment in the Gulf of Mexico and Caribbean Sea; this displays a much more realistic structure of the mean flow (especially off Cape Hatteras) than typical representations obtained from simulations without assimilation (e.g. Chassignet *et al.* 1996).

It is intended to upgrade the future MERCATOR prototypes to include the SAM-2 assimilation tool, and to pursue the development of the algorithm by improved temporal strategies, such as the Incremental Analysis Updating method (Bloom *et al.* 1996), and new statistical parametrizations such as adaptive schemes (e.g. Brankart *et al.* 2003; Testut *et al.* 2003). An extension to a wider variety of assimilation data types is also foreseen, in the perspective of new observing systems such as sea surface salinity measurements from satellites with the forthcoming SMOS (Soil Moisture and Ocean Salinity) mission.

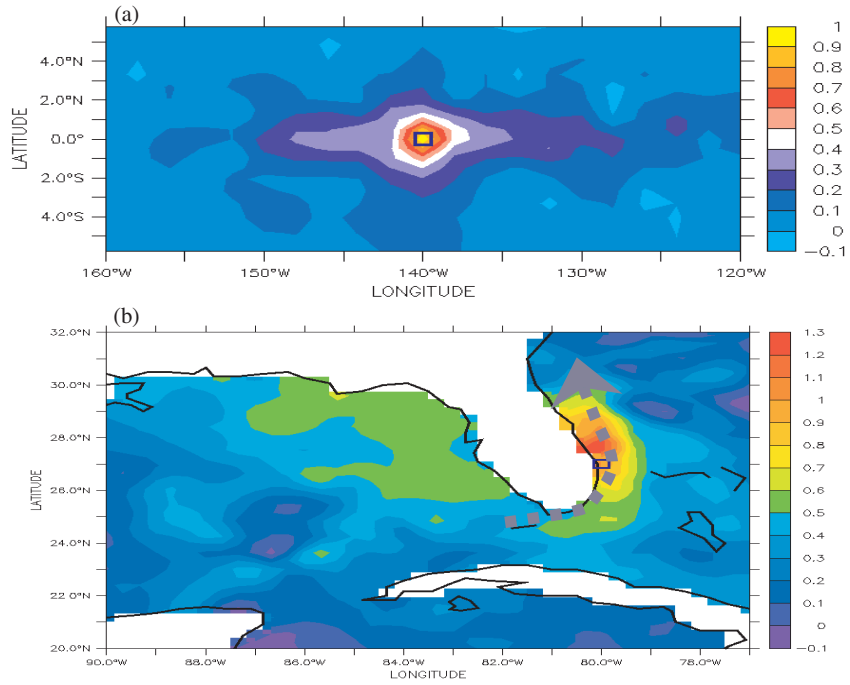


Figure 4. Representer function associated with a SST observation ( $^{\circ}\text{C}$ ): (a) the representer is computed with respect to a point located at  $0^{\circ}\text{N}$ ,  $140^{\circ}\text{W}$  (blue square), in the central Pacific; (b) the representer is relative to a point located at  $27^{\circ}\text{N}$ ,  $80^{\circ}\text{W}$  (blue square), in the Atlantic. The arrow indicates the Gulf Stream path. See text for details.

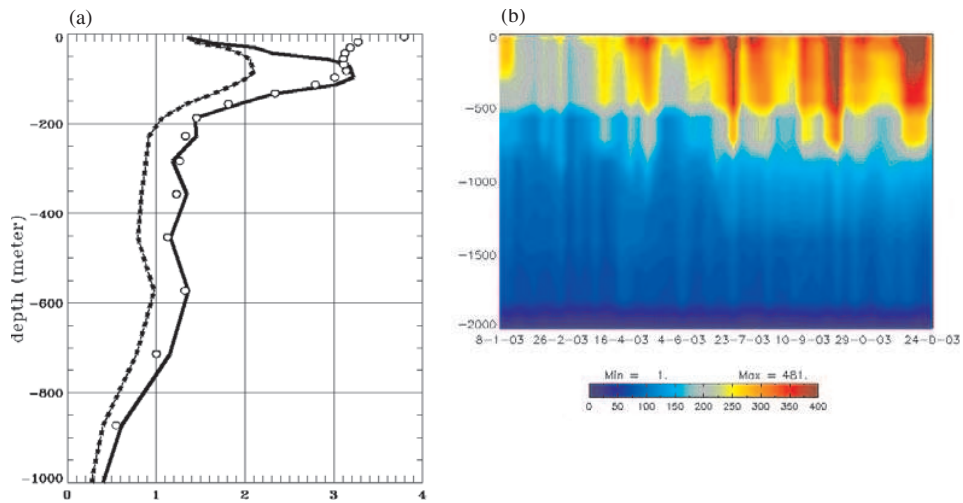


Figure 5. Vertical distribution in the North Atlantic Ocean domain of: (a) variance of the temperature misfit (in  $\text{degC}^2$ ) between independent temperature data, and: (i) climatology (small circles), (ii) the control run (solid line) and (iii) the assimilation simulation (dashed line) during 2003; (b) number of measurements available for each assimilation cycle during 1993, as a function of depth.

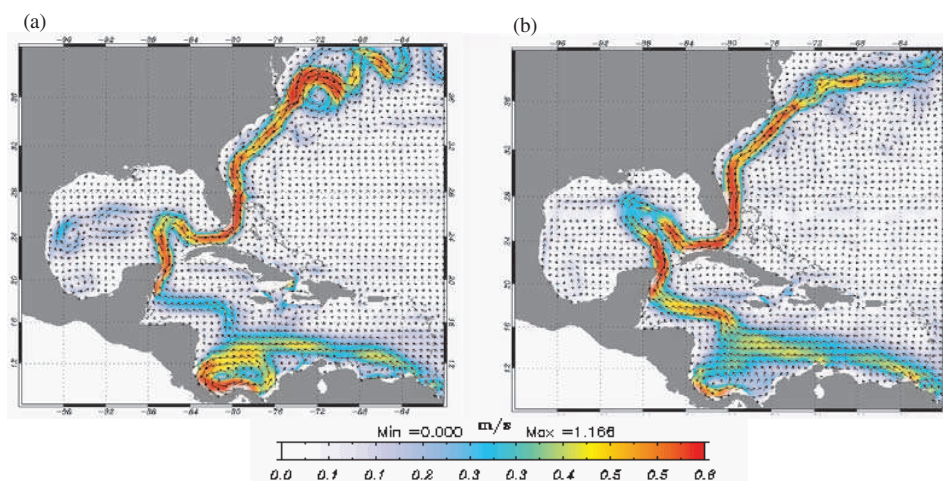


Figure 6. Annual mean surface current ( $\text{m s}^{-1}$ ) in the region of the Gulf of Mexico (Florida strait, Cape Hatteras, Yucatan strait), computed from the model: (a) without assimilation, and (b) from a hindcast experiment performed during 2003 with SAM-2. See text for details.

#### (d) SAM-3

A major simplification made in the SAM-1 and SAM-2 schemes is to perform the analysis at regular time intervals that do not necessarily correspond with the measurement times. This was identified as a serious weakness of operational NWP systems, and lead many operational centres to initiate major research efforts to develop 4D assimilation algorithms such as 4D-Var (Courtier *et al.* 1994). Variational assimilation methods were then developed in the oceanographic context (e.g. Luong *et al.* 1998; Weaver *et al.* 2003). The 4D-Var algorithm takes rigorous account of the temporal dimension by including the prognostic model equations as constraints in the assimilation problem. The 4D-Var formulation is based on the minimization of a cost function that measures the weighted model departures from observations and from the model background state. The minimizing solution is the closest model trajectory that simultaneously fits both sources of information (observations and background) within error bars set by estimates of their respective error covariance matrices (the inverse of these matrices are used to weight the observation and background terms in the cost function). At a given time, this solution is constrained by both past and future observations available in the assimilation window. As a result, the variational approach is particularly well suited to control the dominant processes taking place in tropical regions, where equatorial waves can propagate over large distances within a typical assimilation window width of several days.

Hence the implementation of a third assimilation strategy, SAM-3, has been initiated in MERCATOR, which is based on the incremental variational assimilation system developed at CERFACS (Centre Européen de Recherche et Formation Avancée en Calcul Scientifique) for the OPA model (Weaver *et al.* 2003). A preliminary assessment of the first SAM-3 version (based on a 3D-Var algorithm) has been carried out, using hindcast experiments to assimilate altimetric data or temperature/salinity profiles in the global model configuration at  $2^\circ$  resolution. Further evaluations of the variational system will be conducted by comparing re-analyses obtained with SAM-2 and SAM-3 in

the tropical regions. The outcome of these evaluations will provide some guidance in defining the MERCATOR assimilation strategy concerning SAM-3 in coming years.

In order to meet the operational requirements of the MERCATOR applications, several technical and scientific features of the algorithm need to be investigated. First of all, SAM-3 must be able to assimilate *in situ* and altimetric data simultaneously in a global ocean configuration. Most global applications to date (e.g. those conducted within the framework of the EU ENACT\* project) have involved the assimilation of each of these data types separately. A key point for a successful assimilation is how the multivariate background-error covariance is specified. A promising technique to model these covariances implicitly has recently been developed by Weaver *et al.* (2006); this uses a balance operator to transform from model space, where variables are highly correlated, to a control space where variables are approximately uncorrelated.

A second question that requires further investigation is the feasibility of running variational assimilation algorithms with eddy-resolving ocean models and, most importantly, the required computational resources for doing so. So far, only a few 4D-Var studies have been conducted in the context of eddy-resolving ocean models, and these have shown the necessity of reducing the length of the assimilation window to preserve the validity of the tangent-linear approximation (Luong *et al.* 1998). The exploration of these issues with realistic basin-scale models at eddy-permitting resolution is in progress.

A third issue that is worthy of investigation is the possibility of combining the advantages of 4D-Var assimilation with the propagation of the error statistics from one assimilation cycle to the next, as permitted by Kalman filters. Veersé *et al.* (2000) proposed a hybrid algorithm that combines the variational analysis with the state error propagation of the SEEK filter. In future, such approaches could be explored to provide a hybrid version of the SAM-2 and SAM-3 systems.

#### 4. DEVELOPMENT OF AN ADVANCED ASSIMILATION CAPACITY IN MERSEA

##### (a) *The assimilation research and development effort in MERSEA*

Most operational forecasting systems in evaluation today, not only in MERCATOR but also in the FOAM, HYCOM and MFS projects, are based on simplified assimilation schemes such as reduced-order OI. These methods are fairly robust and require relatively few computer resources, however, they rely on quite severe assumptions about the error statistics which are rarely verified in reality, and therefore cannot make optimal use of all available observations to estimate the state of the ocean and predict its evolution. The assimilation scheme of TOPAZ, however, is inherited from the more advanced Ensemble Kalman filter (EnKF) method, which for some applications is reduced to an ensemble OI scheme (Evensen 2003).

One of the goals of MERSEA is to improve the existing assimilation methods implemented in European forecasting system by: (i) addressing new issues which are particularly relevant for ocean forecasting at high resolution, such as the existence of nonlinear processes and non-Gaussian statistical behaviours; and (ii) extending the assimilation to new data types such as sea-ice parameters or biogeochemical properties in coupled circulation/ecosystem models. The focus in MERSEA is on fully multivariate assimilation methodologies with proven capabilities, that have been developed and applied extensively in previous European operational oceanography projects (Brusdal *et al.* 2003).

\* European Union funded project: ENhanced ocean data Assimilation and ClimaTe prediction.

Three classes of assimilation methods are being considered here:

- Very computationally efficient multivariate statistical schemes (e.g. SAM-2 or the ensemble-based OI schemes) which use time-invariant error statistics. This effort will mainly serve the consolidation phase of the MERCATOR and MFS systems.
- Ensemble-based methods (EnKF and SEEK) which can evolve the error statistics in time using nonlinear models, and thereby provide more realistic predictions of error statistics to be used in the analysis scheme. Earlier versions of these methods have been tested during the DIADEM experiment (Brusdal *et al.* 2003), and their further development in MERSEA will serve the consolidation phase of the TOPAZ assimilation system.
- Fully nonlinear filters such as the Sequential Importance Resampling (SIR) filter, which is a property-conserving Monte Carlo method designed to handle non-Gaussian error statistics. However, the research on nonlinear filters is exploratory, and the transition to a particular operational system will depend on the progress achieved during MERSEA.

In addition to these methods, research and development efforts on variational assimilation techniques will be pursued by MERCATOR and collaborators to develop the SAM-3 variant of the MERCATOR assimilation system.

(b) *MERSEA developments with the SEEK filter*

Turbulent momentum, heat and fresh water fluxes at the air–sea interface (usually computed using bulk formulations) are one of the main sources of error in ocean models (Large 2006), which strongly reduce the operational capacity to provide realistic forecasts of thermohaline characteristics of the mixed layer and of the surface ocean currents. This problem is explored in the framework of MERSEA, in order to better understand the nature of these errors and improve our knowledge of the ocean–atmosphere fluxes by assimilation of oceanic observations.

The starting point of this investigation is a reduced order OI scheme (like the SAM-2 scheme or the SEEK filter with a pre-determined background error covariance) in which the background-error covariance matrix is represented by means of a set of 3D error modes (e.g. EOFs of the model variability) in the state space of the ocean model. The idea is to augment the control space of the filter to include, in addition to the state variables, information about the air–sea fluxes. This information could be the fluxes themselves, or the atmospheric parameters from which they are computed. An approach is developed here that includes a selection of key parameters of the bulk formulae in the control vector, because: (i) these parameters are likely to persist in time (the aim is to improve the forecast); (ii) they are expected to be controllable by ocean observations (provided that the chosen parameters are linearly linked to the value of the flux); and (iii) they are assumed to be the real source of error (in spite of a possible risk of compensating for errors in the atmospheric parameters by correcting the bulk coefficients).

In the example discussed here, only the sensible-heat flux coefficient ( $C_H$ ) and the latent-heat flux coefficient ( $C_E$ ) are included in the control vector. The procedure is tested using twin assimilation experiments with a similar model, as in the MERCATOR global prototype at 2° resolution. The reference simulation (the true ocean) is a standard interannual simulation for the year 1993, with the original bulk formula:

$$Q_S = \rho_a C C_H W (T_w - T_a),$$

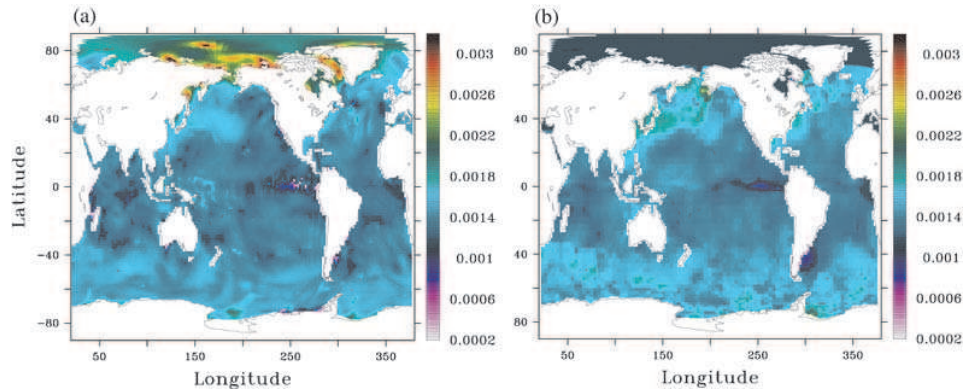


Figure 7. Spatial distribution of the latent-heat flux coefficient: (a) as obtained in the reference simulation, and (b) as reconstructed by the statistical analysis with the augmented state vector. See text for details.

where  $\rho_a$  is the air density,  $C$  is the air specific heat,  $W$  is the wind speed,  $T_w$  and  $T_a$  are the sea-surface and air temperatures, respectively, and:

$$Q_L = \rho_a L C_E W \max(0, q_s - q_a),$$

where  $L$  is the latent heat of vaporization, and  $q_s$  and  $q_a$  are the surface and atmospheric specific humidities, respectively.  $C_E$  and  $C_H$  receive complex parametrizations depending, in particular, on the stability of the air column close to the sea surface. Figure 7(a) shows the value of  $C_H$  in the reference simulation for 31 January 1993. Synthetic observations of temperature and salinity profiles are then sampled from this reference simulation, to be assimilated in a modified simulation in which the values of  $C_E$  and  $C_H$  are kept constant ( $C_E = 1.12 \times 10^{-3}$  and  $C_H = 10^{-3}$ ). Hence, the experiment is built in such a way that the only source of error in the model is due to  $C_E$  and  $C_H$ . The sampling has been designed to approximately mimic the Argo network of profiling floats, i.e. one profile every 10 days and every three grid points in both directions.

To perform the assimilation experiment with sequential corrections of the bulk coefficients, a background-error covariance matrix in the augmented control space was generated as follows. From a series of six initial conditions extracted every 10 days from the reference simulation in 1993, an ensemble of 10-day forecasts was computed using the ocean model with 10 different values of  $C_E$  and  $C_H$ . The dispersion of these values reflects the a priori uncertainty of the corresponding parameters. The 10-day forecast period was chosen to correspond to the assimilation window of the sequential assimilation scheme. In that way, an ensemble of  $6 \times 9$  10-day forecast anomalies (augmented by the  $C_E$  and  $C_H$  anomalies) was obtained, and the 20 dominant EOFs were used to parametrize the background-error covariance matrix for the assimilation experiment. However, since in each member of the ensemble the coefficients are constant horizontally, the resulting error covariance matrix is unable to represent correctly the horizontal correlation structure between errors in  $C_E$  and  $C_H$ , and the ocean state error. Hence, this ensemble will only be used to estimate the local multivariate correlation between the ocean state and the bulk coefficients.

Figure 7(b) illustrates the statistical analysis obtained for the augmented state vector after a forecast from perfect initial conditions, using temperature and salinity observations sampled from the reference simulation. The analysed field for the latent-heat flux coefficient ( $C_E$ ) is compared to the same field in the reference simulation.

It shows that significant values of the bulk coefficients can be inferred from only temperature and salinity observations. This experiment is ideal in the sense that the only source of error is in  $C_E$  and  $C_H$ , and the atmosphere and ocean initial conditions are both perfectly known, but it illustrates that it is theoretically possible to estimate the turbulent air–sea flux bulk coefficients by inverting oceanic observations. To go further two questions need to be addressed. How many bulk parameters can be controlled using only oceanic observations? What is the impact of initial errors in the ocean state? In other words, is it possible to control the full system, i.e. the ocean state and the flux parameters, with the available ocean observation system?

(c) *MERSEA developments with the EnKF*

The EnKF was originally introduced by Evensen (1994) as a more stable alternative to the extended Kalman filter. Monte Carlo (ensemble) model forecasts have the advantage of not requiring a linearization of the numerical model, and are therefore better suited for strongly nonlinear systems. A review of the EnKF applications can be found in Evensen (2003). The analysis scheme is based on linear statistical estimation theory, which is conceived for Gaussian variables. The analysis scheme can, however, be adapted to nonlinear estimation by geo-statistical extensions (Bertino *et al.* 2003). The evolutive SEEK filter shares many common features with the EnKF, as pointed out by Brusdal *et al.* (2003). The main differences are in the Monte Carlo propagation step and in the details of the solution algorithm of the analysis scheme.

In the TOPAZ operational system, which is the Arctic component of MERSEA, the EnKF is applied to the HYCOM model. Lisæter *et al.* (2003) have evaluated the assimilation of satellite-measured ice concentrations by the EnKF in a coupled ice–ocean model. The TOPAZ model used here is identical, but it has double the horizontal resolution in order to represent mesoscale processes. Two examples of forecast-error covariances in winter situations are given in Figs. 8(a) and (b), showing the predicted ensemble correlation between ice concentration and sea surface salinity at two instants separated by three weeks. The EnKF forecast-error statistics are evolving dynamically with model error prescribed in the forcing fields, so that every member runs with slightly different wind stress and heat fluxes. The resulting multivariate ensemble correlation is a combination of different model processes. In Figs. 8(a) and (b) the correlation intensifies along the ice edge (presented as the ensemble average 15% ice concentration isoline in green) and near land. The negative correlation in the interior of the ice pack can be due to wind-induced ice convergence. Parts of the marginal ice zone with negative correlations can be areas where ice is blown into the ice pack by southerly winds (wind fields not shown here) and open water is mixed with the fresher water below. The same applies along coasts from which winds are blowing the ice away. Other parts with positive correlations are areas where new ice is freezing and brine is released into the surface waters.

Since the forecast multivariate statistics are strongly related to winds, the use of an evolutive data assimilation method where wind fields are used as a source of model error is fully justified (K. A. Lisæter 2006, personal communication). Further improvements can be achieved by applying non-Gaussian transformations when the ice–ocean relationships are strongly nonlinear.

(d) *MERSEA developments with the SIR filter*

The SIR filter is an ensemble-based Monte Carlo method similar to the EnKF. However, the SIR filter updates probabilities of the ensemble members and not the ensemble states themselves. This difference makes the SIR filter a truly variance-minimizing



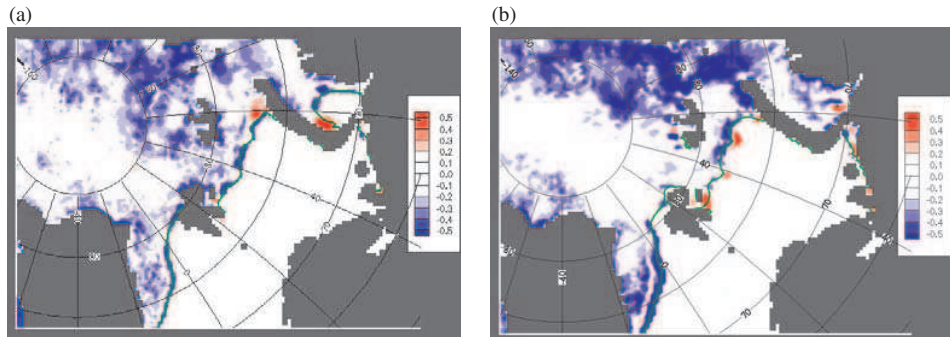


Figure 8. Correlations between ice concentration and sea surface salinity on: (a) 21 December 2005, and (b) 11 January 2006.

scheme, which can easily be applied for nonlinear systems with any probability density function (pdf) that is not necessarily Gaussian.

Such a method is well suited to the problem of assimilating data into strongly non-Gaussian ecosystem models that MERSEA needs to investigate. Indeed, monitoring the ocean environment requires not only physical but also ecosystem models. Although significant advances have been made in recent years, understanding and modelling the complex processes in ecosystems requires an integrated approach based on observed data. MERSEA will contribute to progress by developing an advanced biogeochemical model to be coupled with the global ocean circulation model.

The SIR filter has been used successfully for simultaneous state and parameter estimation in a relatively simple ecosystem model with 15 poorly known model parameters (e.g. Losa *et al.* 2003). The observations came from the BATS (Bermuda Atlantic Time-series Study) dataset, i.e. real observations were used. The obtained model solution agreed reasonably well with the data, even with independent values which had not been assimilated. The model parameters, however, revealed strong seasonal variations, which may point to possible uncertainties in the parametrization of biological processes. This motivated us to also try to estimate the model noise level in the data assimilation scheme.

In general, the magnitude of the noise level is very difficult to determine; usually it is based on ‘educated guesses’. On the one hand this is satisfactory, because it is here where scientific intuition comes in; on the other hand, a more objective way of determining it is desirable. Furthermore, there is no a priori reason why the model noise should not be time-dependent.

In the SIR filter, the pdf of the ecosystem model is represented by a finite number of ensemble members. By integrating the ensemble forward in time, subject to model noise, the evolution of the pdf with time is simulated. As soon as observations are present, each ensemble member is weighted by the ‘distance’ to these observations. To be more precise, the weight  $w_i$  for ensemble member  $\psi_i$  is calculated from:

$$w_i = \frac{p(d | \psi_i)}{\sum_i p(d | \psi_i)},$$

which follows directly from Bayes theorem (Kivman 2003; Losa *et al.* 2003; Van Leeuwen 2003). Here,  $p(d | \psi_i)$  is the pdf of the observations,  $d$ , given the ensemble member,  $i$ . The specific form for this pdf will be presented below. This leads to a

weighted ensemble, and the statistical moments, like the mean of the pdf or its variance, are calculated using the weighted members.

After several updates of the ensemble by observations, some ensemble members get relatively high weight, while the weight of others becomes negligible. This leads to a reduction in the effective size of the ensemble. To avoid this, the ensemble is re-sampled after each update to give each member equal weight again. This re-sampling can be done in several ways. Here we used the weights as probabilities of the members, and drew a new ensemble from the resulting probability, with replacement\*.

The result of the re-sampling is an ensemble of the same size, in which some ensemble members are identical and others have disappeared. This new ensemble is then integrated forward in time to the next observation time, where the weighting and re-sampling is repeated. To increase the spread in the re-sampled ensemble, the identical members are slightly perturbed. No general rule exists on the size and form of the perturbations. In the experiments described here the perturbations are applied to the parameter values of identical members.

The SIR filter has been implemented for estimating poorly known biological parameters of an ecosystem model developed by Drange (1996), which describes the dynamics of phytoplankton, zooplankton, bacteria, dissolved inorganic nitrogen—represented by nitrate and ammonium—and particular and dissolved organic matter within the upper mixed layer. The ecosystem model was constrained by BATS data, in particular by nitrate, chlorophyll, dissolved organic nitrogen and carbon concentrations, observed for the period from December 1988 to January 1994. Various biological parameters have been adjusted.

It is worth noting that the experiment is very similar to those designed and discussed in detail in the study by Losa *et al.* (2003). Thus, an initial ensemble of 1000 particles is drawn randomly from an exponential distribution:

$$p(\psi) = \bar{\psi} \exp\left(-\frac{\psi}{\bar{\psi}}\right),$$

where  $\bar{\psi}$  is equal to a first guess. Then, each ensemble member evolves according to the model equations, with some random model (system) noise added to the model at every time step of the integration. In the study by Losa *et al.* (2003) model noise was estimated simply by trial and error. Here, the level of the system noise is implemented as nine additional parameters added to the 15 biological parameters to be optimized. At the analysis step when filtering the ensemble the observation errors are assumed to be Gaussian distributed, which leads to weights of the following form:

$$w_i \approx \exp\{-0.5(d - \psi_i)^2/s^2\},$$

where  $s^2$  is the variance of the observation.

As mentioned above, all the optimized parameters of the re-sampled ensemble are perturbed to avoid ensemble collapse. The procedure is as follows: if, at an analysis step, some parameter values  $a$  are re-sampled many times, a uniform pdf is created in the interval ( $a - \text{nearest smaller value}$ , to  $a + \text{nearest higher value}$ ), and new parameter values are drawn from this uniform density.

Figure 9(a) depicts the temporal evolution of one of 15 optimized model parameters (red solid curve) obtained from the experiment with the variable level of model noise, against previous estimates of the same parameters (black solid curves) when the model

\* It has been shown that this procedure introduces extra variance in the ensemble (Liu and Chen 1998), which can be minimized by other even more efficient re-sampling procedures.

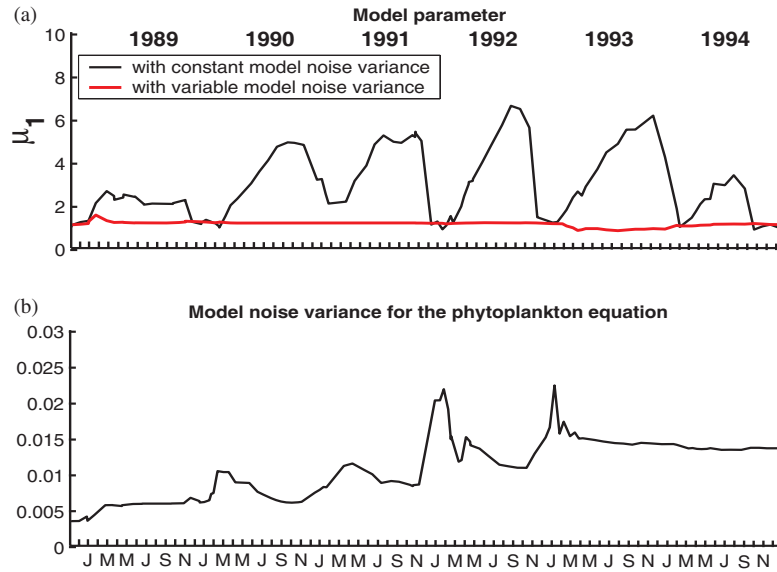


Figure 9. The evolution of the estimates for: (a) one of the optimized biological model parameters, and (b) model-noise variance. All the parameters are normalized with respect to their initial values. See text for details.

noise-level is constant in time. One can see that the previous estimates revealed quite strong seasonal variations and trends, which are difficult to understand, while optimizing the model noise has allowed one to get model parameter values almost constant in time. The temporal variations are present on Fig. 9(b), which depicts the model-noise variance over the period 1989–94.

The results obtained with the SIR filter show that strong seasonal variations found in estimated model parameters can be linked to errors in the model equations and forcing fields. In reality some of the biological parameters may vary in time as they may depend on a number of environmental conditions (water temperature, for instance). Such dependencies however are not yet parametrized properly.

In principle, one would like to run the SIR filter scheme with parameter and noise estimation along with the biogeochemical model. Unfortunately, this is too expensive in operational mode. A practical implementation is to estimate from a long hindcast the seasonal dependence of the model parameters and the noise level. In operational mode, the biogeochemistry model is then run with these seasonally varying parameters, and a realization of the model noise from the appropriate noise level at that time. More research is needed to find the optimal strategy.

## 5. CONCLUSIONS

In this paper, an overview has been given of the progress made in the field of ocean data assimilation, in the perspective of operational monitoring and forecasting. In the MERCATOR context, advantage is taken from the whole spectrum of complexity that assimilation methods can offer to provide the best possible description of the ocean in space and time. Through the MERSEA project, a transition is underway from OI-type statistical schemes to more advanced methods which can address nonlinear processes more rigorously.

The GODAE initiative has contributed to the establishment of a feedback loop within the oceanographic community, in order to guarantee coherent progress between science, research and development activities, operational applications, and the user requirements. This is also a main goal of the MERSEA project. The expected benefit will be improved quality standards of operational products and services, and thereby strengthened credibility in, and sustainability of, operational oceanography through the satisfaction of users.

## ACKNOWLEDGEMENTS

This work has been partly supported by the MERSEA project of the European Commission under Contract SIP3-CT-2003-502885.

## REFERENCES

- Bahurel, P. 2006 MERCATOR OCEAN global to regional ocean monitoring and forecasting. Pp. 381–395 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Ballabrera-Poy, J., Brasseur, P. and Verron, J. 2001 Dynamical evolution of the error statistics with the SEEK filter to assimilate altimetric data in eddy-resolving ocean models. *Q. J. R. Meteorol. Soc.*, **127**, 233–253
- Bertino, L., Evensen, G. and Wackernagel, H. 2003 Sequential data assimilation techniques in oceanography. *Int. Stat. Rev.*, **71**(2), 223–241
- Birol, F., Brankart, J.-M., Lemoine, J. M., Brasseur, P. and Verron, J. 2005 Assimilation of satellite altimetry referenced to the new GRACE geoid estimate. *Geophys. Res. Lett.*, **32**, L06601, doi: 10.1029/2004GL02329
- Blayo, E., Mailly, T., Barnier, B., Brasseur, P., Le Provost, C., Molines, J. M. and Verron, J. 1997 Complementarity of ERS-1 and TOPEX/Poseidon altimeter data in estimating the ocean circulation: Assimilation into a model of the North Atlantic. *J. Geophys. Res.*, **102**(C8), 18573–18584
- Bloom, S. C., Takacs, L. L., Da Silva, A. M. and Levina, D. 1996 Data assimilation using incremental analysis updates. *Mon. Weather Rev.*, **124**, 1256–1271
- Brankart, J.-M., Testut, C.-E., Brasseur, P. and Verron, J. 2003 Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: Application to a 1993–96 hindcast of the North Atlantic Ocean circulation. *J. Geophys. Res.*, **108**, 3074, doi: 10.1029/2001JC001198
- Brasseur P. 2006 Ocean data assimilation using sequential methods based on the Kalman Filter. Pp. 271–316 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Brusdal, K., Brankart, J. M., Halberstadt, G., Evensen, G., Brasseur, P., van Leeuwen, P. J., Dombrowsky, E. and Verron, J. 2003 A demonstration of ensemble-based assimilation methods with a layered OGCM from the perspective of operational ocean forecasting systems. *J. Mar. Systems*, **40–41**, 253–289
- Chassignet, E., Smith, L. T., Bleck, R. and Bryan, F. O. 1996 A model comparison: Numerical simulations of the North Atlantic ocean circulation in depth and isopycnic coordinates. *J. Phys. Oceanogr.*, **26**, 1849–1867
- Cooper, M. and Haines, K. 1996 Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **101**(C1), 1059–1077
- Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994 A strategy for operational implementation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**, 1367–1388
- Crosnier, L. and Le Provost, Ch. 2006 Internal metrics definition for operational forecast systems intercomparison: Example in the North Atlantic and Mediterranean Sea. Pp. 455–465 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. E. Chassignet and J. Verron. Springer, Dordrecht, the Netherlands
- De Mey, P. 1997 Data assimilation at the oceanic mesoscale: A review. *J. Meteorol. Soc. Jpn*, **75**, 415–425

- De Mey, P. and Benkiran, M. 2002 A multivariate reduced-order optimal interpolation method and its application to the Mediterranean basin-scale circulation. In *Ocean forecasting: Conceptual basis and applications*. Eds. N. Pinardi and J. D. Woods. Springer Verlag, Berlin, Germany
- Demirov, E., Pinardi, N., Fratiani, C., Tonani, M., Giacomelli, L. and De Mey, P. 2003 Assimilation scheme of the Mediterranean Forecasting System: Operational implementation. *Ann. Geophysicae*, **21**, 189–204
- Drange, H. 1996 An isopycnic coordinate model of the seasonal cycling of carbon and nitrogen in the Atlantic Ocean. *Phys. Chem. Earth*, **25**, 503–509
- Drillet, Y., Bourdalle-Badie, R., Siefriid, L. and Le Provost, C. 2005 The MEDDIES in the Mercator North Atlantic and Mediterranean Sea eddy-resolving model. *J. Geophys. Res.*, **110**, C03016, doi: 10.1029/2003JC002170
- Echevin, V., De Mey, P. and Evensen, G. 2000 Horizontal and vertical structure of the representer functions for sea surface measurements in a coastal circulation model. *J. Phys. Oceanogr.*, **30**, 2727–2635
- Evensen, G. 1994 Sequential data assimilation with a non-linear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162
- 2003 The Ensemble Kalman Filter theory and practical implementation. *Ocean Dyn.*, **118**, 1–23
- Ferry, N., Remy, E., Brasseur, P. and Maes, C. 2006 The Mercator global ocean operational analysis/forecast system: Assessment and validation of an 11-year reanalysis. *J. Mar. Systems*, in press
- Fu, L.-L. and Cazenave A. 2001 *Satellite altimetry and earth sciences: A handbook of techniques and applications*. International Geophysics Series, Vol. 69. Academic Press, London, UK
- Fukumori, I. 2001 Data assimilation by models. Pp. 237–265 In: *Satellite altimetry and earth sciences, a handbook of techniques and applications*. Eds. L.-L. Fu and A. Cazenave. Academic Press, San Diego, USA
- 2006 What is data assimilation really solving, and how is the calculation actually done? Pp. 317–342 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. E. Chassignet and J. Verron. Springer, Dordrecht, the Netherlands
- Garric, G. and Charpentier, P. 2005 ‘Sea ice: A new component of the Mercator project’. Pp. 4–11 in MERCATOR Newsletter No 16. Available from MERCATOR-OCEAN, Ramonville-St-Agne, France (or <http://www.mercator.com.fr/>)
- Griffies, S. 2006 Some ocean models fundamentals. Pp. 19–73 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Kivman, G. A. 2003 Sequential parameter estimation for stochastic systems. *Nonlinear Processes Geophys.* **10**, 253–256
- Large, W. 2006 Surface fluxes for practitioners of global ocean data assimilation. Pp. 229–270 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Lisæter, K. A., Rosanova, J. and Evensen, G. 2003 Assimilation of ice concentration in a coupled ice–ocean model, using the Ensemble Kalman filter. *Ocean Dyn.*, **53**, 368–388
- Liu, J. S. and Chen, R. 1998 Sequential Monte-Carlo methods for dynamic systems. *J. Am. Stat. Ass.*, **93**, 1032–1044
- Losa, S. N., Kivman, G. A., Schroeter, J. and Wenzel, M. 2003 Sequential weak constraint parameter estimation in an ecosystem model. *J. Mar. Systems*, **43**, 31–49
- Luong B., Blum, J. and Verron, J. 1998 A variational method for the resolution of data assimilation problems in oceanography. *Inverse Problems*, **14**, 979–997
- Madec, G., Delecluse, P., Imbard, M. and Lévy, C. 1998 ‘OPA 8.1 ocean general circulation model reference manual’. Notes du pôle de modélisation IPSL. Available from <http://www.lodyc.jussieu.fr/NEMO/>
- Penduff, Th., Brasseur, P., Testut, C.-E., Barnier, B. and Verron, J. 2002 Assimilation of sea-surface temperature and altimetric data in the South Atlantic Ocean: Impact on basin-scale properties. *J. Mar. Res.*, **60**, 805–833
- Pham, D. T., Verron, J. and Roubaud, M. C. 1998 A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Systems*, **16**, 323–340

- Schiller, A. and Smith, N. 2006 BLUElink: Large-to-coastal scale operational oceanography in the southern hemisphere. Pp. 427–439 in: *Ocean weather forecasting: An integrated view of oceanography*. Eds. E. Chassignet and J. Verron. Springer, Dordrecht, the Netherlands
- Send, U. 2006 *In situ* observations: Platforms and techniques. Pp. 191–206 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Smith, N. 2006 Perspectives from the Global Ocean Data Assimilation Experiment. Pp. 1–17 in *Ocean weather forecasting: An integrated view of oceanography*. Eds. J. Verron and E. Chassignet. Springer, Dordrecht, the Netherlands
- Smith, W. H. F. and Sandwell, D. T. 1997 Global sea floor topography from satellite altimetry and ship depth soundings. *Science*, **277**, 1956–1962
- Testut, C.-E., Brasseur, P., Brankart, J. M. and Verron, J. 2003 Assimilation of sea-surface temperature and altimetric observations during 1992–1993 into an eddy-permitting primitive equation model of the North Atlantic Ocean. *J. Mar. Systems*, **40–41**, 291–316
- Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C. and Duliere, V. 2005 On the representation of high latitude processes in the ORCA-LIM global coupled sea ice–ocean model. *Ocean Modelling*, **8**, 175–201
- Van Leeuwen, P. J. 2003 A truly variance minimizing filter for large-scale applications. *Mon. Weather Rev.*, **131**, 2071–2084
- Veersé, F., Pham, D.-T. and Verron, J. 2000 ‘4D-Var/SEEK: A consistent hybrid variational-smoothing data assimilation method’. Rapport de recherche INRIA No 3902, Domaine de Voluceau, Rocquencourt, Le Chesnay, France (available from <http://www.inria.fr/rrrt/>)
- Verron, J. 1992 Nudging altimeter data into quasi-geostrophic ocean models. *J. Geophys. Res.*, **97**(C5), 7497–7491
- Verron, J. and Chassignet, E. (Eds) 2006 *Ocean weather forecasting: An integrated view of oceanography*. Springer, Dordrecht, the Netherlands
- Vörösmarty, C. J., Fekete, B. and Tucker, B. A. 1996 ‘A contribution to IHP-V Theme discharge database, version 1’. River 1.0 (RivDIS v1.0), Volumes 0 through 6. Technical Documents in Hydrology Series. UNESCO, Paris
- Weaver, A., Vialard, J. and Anderson, D. L. T. 2003 Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. I: Formulation, internal diagnostics, and consistency checks. *Mon. Weather Rev.*, **131**, 1360–1378
- Weaver, A., Deltel, C., Machu, E., Ricci, S. and Daget, N. 2006 A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 000–999