Data assimilation in operational ocean forecasting systems: the MERCATOR and MERSEA developments

By P. BRASSEUR[1,2], P. BAHUREL[2], L. BERTINO[3], F. BIROL[2], J.-M. BRANKART[1], N. FERRY[2], S. LOSA[4], E. REMY[2], J. SCHRÖTER[4], S. SKACHKO[1], C.-E. TESTUT[2], B. TRANCHANT[2], P.J. VAN LEEUWEN[5] , J. VERRON[1],

(1) Laboratoire des Ecoulements Géophysiques et Industriels, BP 53, 38041 Grenoble Cedex 9, France.
(2) MERCATOR-Océan, 8-10 rue Hermès, Parc technologique du canal, 31526 Ramonville Saint Agne, France.
(3) Mohn-Sverdrup Center/Nansen Environmental and Remote Sensing Center, Thormølensgate 47, N-5006 Bergen, Norway.
(4) AWI, P.O Box 120161, 27515 Bremerhaven, Germany.
(5) IMAU, Utrecht University, P.O Box 80005, 3508 TA Utrecht - The Netherlands.

SUMMARY

During the past fifteen years, a number of initiatives were undertaken at national level to develop ocean forecasting systems operating at regional and/or global scales. The coordination 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 to GODAE. The MERCATOR systems assimilate a variety of observations in routine 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 Variationnal methods, which are progressively deployed through the SAM (Système d'Assimilation MERCATOR) series. SAM-1 is based on the reduced-order optimal interpolation SOFA scheme which can be operated using "altimetry-only" or "multivariate" setup; 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 SEEK filter (such as three-dimensional, multivariate error modes and adaptivity schemes). The third one, SAM-3, considers variational methods such as the incremental 4D-VAR algorithm.

Most operational forecasting systems evaluated during GODAE are based on least-square statistical estimation concepts, assuming gaussian errors. In the framework of the E.U. MERSEA project, R&D activities are conducted to prepare the operational ocean monitoring and forecasting systems of the next generation. The research effort aim at the exploration of non-linear assimilation formulations to overcome the limitations of the current systems. This paper provides an overview of the developments conducted in MERSEA with the SEEK filter, the EnKF filter and the SIR filter.

KEYWORDS: Data assimilation, operational ocean forecasting, GODAE.

1. INTRODUCTION

Operational Oceanography is an emerging field of activites 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, 2005*]. The scientific and technological feasability of operational ocean forecasting systems results from several factors:

- 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, parameterization of sub-grid scale processes, coupling with the overlying atmosphere and sea ice etc. [Griffies *et al*., 2005]; in addition, the growth of resources for high performance computing have permitted an increase of the spatial resolution to such an extent that, today, basin-scale models are able to resolve the mesoscale dynamics explicitly.

- the concerted effort of the nations to establish global ocean observing systems: during the past decade, major international programs took place in the World Ocean (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 a continuous and accurate observation of the surface signature of the ocean dynamics at global scale [Fu and Cazenave, 2001].
- the progress achieved in data assimilation techniques: the first application of data assimilation in oceanography dates back to the eighties, involving simple ocean models; then, the theoretical framework of data assimilation has been progressively adapted to meet the specific requirements of more sophisticated ocean models and observational data sets : major advances have been achieved to (i) adapt the assimilation algorithms to ocean systems of very large dimensions, (ii) develop sophisticated representations of the error statistics with primitive equation models, (iii) implement multivariate algorithms for assimilating several data types simultaneously, and (iv) extend the theoretical framework from linear to non-linear problems [e.g. Verron, 1992 ; Blayo *et al*., 1997 ; De Mey 1997 ; Fukumori 2001 ; Evensen 2003; Brusdal *et al*., 2003].

As a consequence of this overall progress, ocean forecasting in the meteorological sense is becoming a reality although this was not identified as a central issue in the early nineties. A number of initiatives were 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, MFS in Italy, TOPAZ in Norway, HYCOM and ECCO in the United States and BLUElink in Australia), and the coordination between these efforts started to be organized internationally through the GODAE (Global Ocean Data Assimilation Experiment) project [Smith, 2005]. One of the main GODAE objectives will be to compare the assimilation systems implemented operationally and evaluate their overall performances in real-time conditions.

The French MERCATOR project [Bahurel 2005] is one of the leading participants to GODAE. The project was launched in 1995 by the major French agencies involved in oceanography. The MERCATOR system is based on two components, the ocean model and the remotely sensed (e.g. SST, altimetric data) and *in situ* (e.g. temperature and salinity profiles) observations, that are integrated through an assimilation system with the objective to provide the best possible description of the real ocean. The assimilation methods are all derived from least-square estimation principles. In the first part of this paper, the suite of assimilation tools implemented in the MERCATOR system will be reviewed and illustrated, focusing on the scientific developments that are conducted in R&D mode.

The transition between the demonstration stage achieved so far by MERCATOR and the consolidation phase is being conducted through the european MERSEA project which represents the oceanic contribution to GMES (Global Monitoring for Environment and Security). In the second part of the paper, the research effort in data assimilation carried out in MERSEA will be discussed and illustrated. Following the pathway taken by NWP organizations, it is believed that a strong investment in R&D is necessary to support operational oceanography in the long-term, with strong beneficial feedback on the research side. The research directions described in Section 4 will be targetted to the exploration of new approaches focused on some non-linear aspects of data assimilation.

2. THE MERCATOR OPERATIONAL MONITORING AND FORECASTING 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 the surface layer's properties, 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 backgroud 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 recognised 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 resolution. 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 MERCATOR system provides a full 3D depiction of the ocean dynamics and thermohaline circulation (temperature, salinity, currents, sea surface elevation,…), with a priority given to high resolution (eddy resolving) scales. Information is available on a near-real-time and routine basis consisting of weekly analyses and 2-week forecasts, in addition to retrospective analyses performed in hindcast mode.

The MERCATOR prototypes consist of an ocean model, different datasets that are assimilated and an assimilation system. The ocean model used today in MERCATOR is based on the rigid-lid version of the OPA-NEMO primitive equations 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 forcing consist of daily fields of wind stress (i.e. the friction of the wind on the ocean surface), evaporation, precipitation, non-solar and solar heat fluxes provided by the European Center for Medium-range Weather Forecast (ECMWF) analyses and forecasts. The surface forcing includes a retroaction term in the net heat flux, based on the difference between the model Sea Surface Temperature (SST) and the weekly Reynolds SST product, in order to describe 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 data base from UNESCO [Vörösmarty *et al.*, 1996].

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 will 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 notice the high spatial coverage of the altimetric data (3 satellites are available, see Fig. 1a) and the relative scarcity of *in situ* data, especially salinity profiles (Fig. 1c). The continuous line in Figure 1b corresponds to XBT temperature profiles collected along shipping tracks. The analysed SST field which gathers information coming from both the model integration and the available observations is displayed in Figure 1d and shows important mesoscale structures.

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

• The first prototype (noted PSY1) covers the North and Equatorial Atlantic with an intermediate resolution of 1/3° and 43 vertical levels distributed from 12 m at the surface to 200 m at the bottom [Benkiran *et al.*, 2005]; this model simultaneously assimilates "surface" data from SST and satellite altimetry and vertical temperature/salinity profiles ; the Mediterranean Sea is not explicitely included in the model domain, but its impact on the

North Atlantic basin is taken into account through a buffer zone covering the Gibraltar Strait and Alboran Sea.

- The second prototype (noted PSY2) is a high resolution model (5 to 7 km horizontal resolution, 43 vertical levels from 6 m at the surface to respectively 200 m and 300 m at the bottom of the Mediterranean Sea and the Atlantic) covering North Atlantic basin from 9°N to 70°N and Mediterranean Sea; 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. Today, only altimetric data are assimilated in PSY2 but an upgrade is underway for incorporation of *in situ* data. The next version of the high resolution Atlantic model will design the future global high resolution model (1/12°) with a free surface, a partial step vertical coordinate, the atmospheric bulk formulae and a sea ice model.
- The third prototype (noted PSY3) is a global ocean configuration with an horizontal resolution of 2°x2°cos(latitude) and 31 levels on the vertical, that assimilates altimetric data only. There are 21 levels located in the top 1,000 meters 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 3,000 m level ; a realistic topography based on the ETOPO5 bathymetric data base is used [Ferry *et al.*, 2005]. Every month, this prototype provides up-to-date estimates of oceanic initial conditions for seasonal climate predictions experienced in NWP centers. An upgrade in the resolution will be achieved soon for the global configuration from 2° to 1/4°.

Figure 2 illustrates the spatial coverage of these models and the typical oceanic structures present in the numerical model outputs.

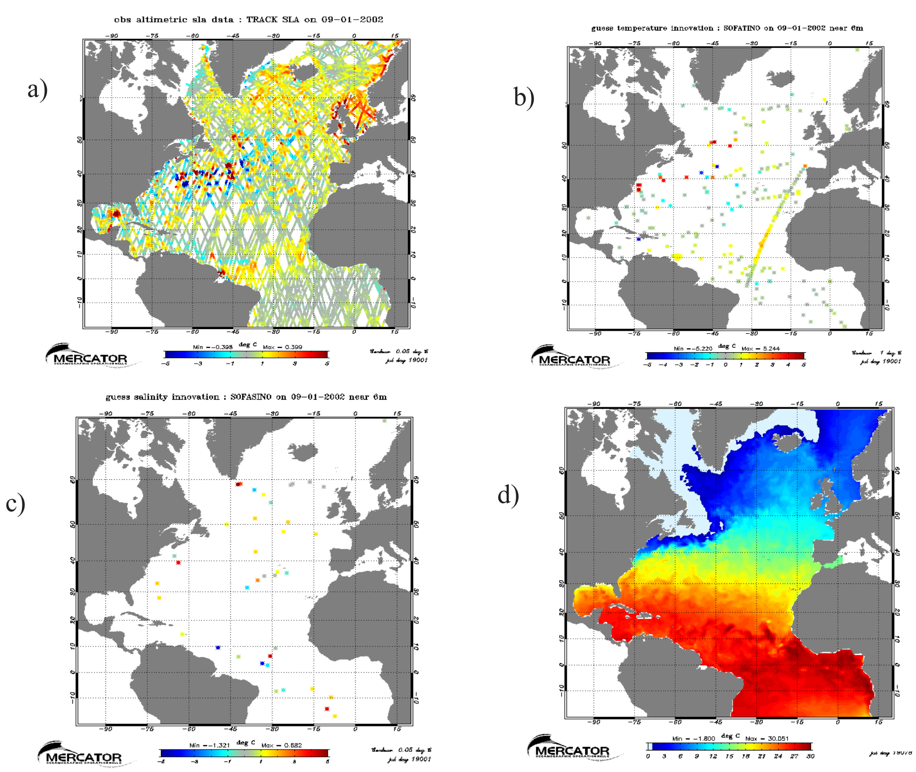

Figure 1: Different kind of assimilated data for a particular week (a, b, c) and resulting sea surface temperature analysed by the MERCATOR North Atlantic (1/3°) prototype. Along track SLA data (Jason1, ERS2, GFO) (a), available temperature (b) and salinity (c) vertical profiles and analysed sea surface temperature (d).

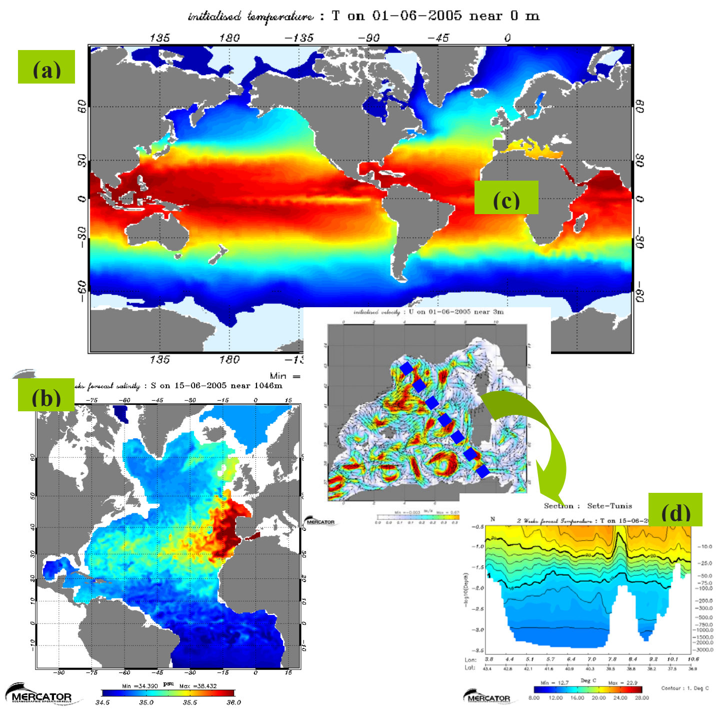

Figure 2: Examples of Mercator Ocean system outputs. (a) Global Ocean (~2° model) Sea Surface Temperature field, Real-Time Analysis 1st June 2005 computed 1st June 2005; (b) North & Tropical Atlantic (1/3° model) 1000 m depth Salinity, 2-week forecast 15: June 2005 computed 1st June 2005; (c) Gibraltar strait (5-7 km model, ~1/15°) surface currents (note the presence of eddies); Real-Time Analysis 1st June 2005; (d) Mediterranean sea (5-7 km model, 1/16°) Temperature vertical section between Sète and Tunis (and surface current map) ; 2-week forecast 15 June 2005 computed 1st June 2005.

3. ASSIMILATION SCHEMES IN THE MERCATOR SYSTEM

(a) Incremental approach

MERCATOR is developing a suite of assimilation tools (called "SAM" for Système d'Assimilation MERCATOR) of increasing complexity, from sub-optimal sequential schemes to variational methods. This incremental approach was adopted at the start of the project to conciliate the operational needs for high-resolution, the available computing resources and the feasability of sequential methods to assimilate altimeter data with mesoscale ocean dynamics [De Mey 1997]. In the long-term, high-performance computing resources are expected to increase, and R&D activities are conducted to anticipate the introduction of more elaborated techniques such as the variational method.

The MERCATOR assimilation tools in place today find their roots in the theoretical framework of least-square statistical estimation. The first release, SAM-1 has been elaborated from an optimal interpolation scheme; SAM-1 is running on an operational real-time basis since early days of year 2001. The second release, SAM-2, is considering a Singular Extended Evolutive Kalman (SEEK) filter analysis method; it has been evaluated and intercompared to SAM-1 in several hindcast experiments and will be integrated in the operational chain soon. The third one, SAM-3, is targetting more advanced approaches such as the 4D variational method and is still under construction in R&D mode.

(b) SAM-1

The SAM-1 assimilation tool is based on the reduced order optimal interpolation method developed by De Mey and Benkiran [2002) and Demirov *et al*. [2003]. Two different versions of this tool are implemented in the MERCATOR prototypes.

Version 1 (noted SAM-1v1) considers 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 details in Ferry *et al*. [2005]. The algorithm starts by calculating an SLA increment from innovations collected over a one-week cycle (i.e. the difference between the along-track measurements and the model forecast at the corresponding time within the cycle), using the SOFA interpolation scheme and 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. The barotropic component is then converted into an increment of horizontal velocity and barotropic streamfunction of the model, while the baroclinic part is used to modify the thermohaline structure of the water column by lifting-lowering of isopycnals with the Cooper and Haines technique. This scheme relies on the idea that, in oceanic regimes where baroclinicity dominates, the 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 higher SLA and conversely.

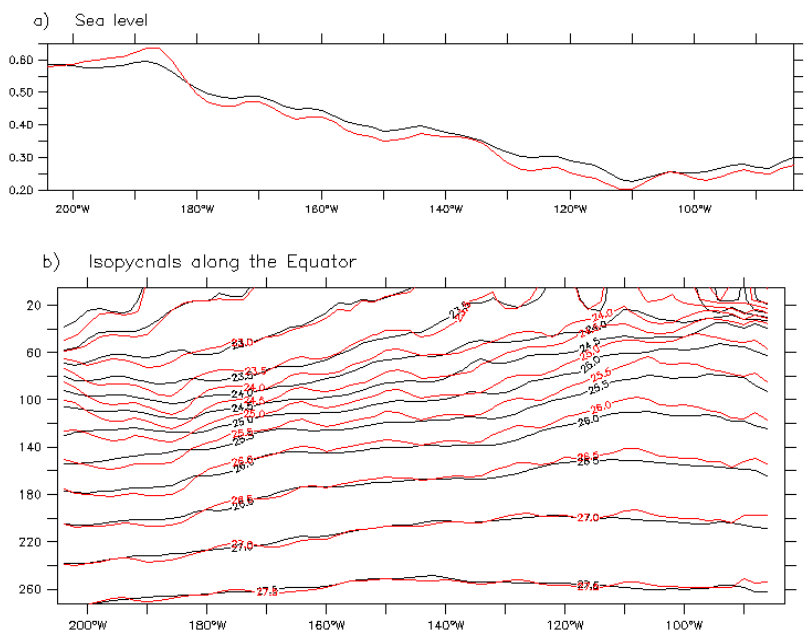

Figure 3: Sea level (a) and isopycnals (b) at the Equator in the Pacific for 23 June 1993 as simulated by the MERCATOR coarse resolution global ocean analysis system. The model state before (respectively after) assimilation is represented by black (respectively red) lines. Sea level is in meter, isopycnals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

SAM1v1 has been implemented in the PSY2 and PSY3 prototypes. To illustrate the impact of SLA data on the oceanic state, Figure 3 represents the sea level and the isopycnals along the Equatorial Pacific between 180°W and 140°W, before and after assimilation of altimetric data for 22 June 1993. The figure shows that the assimilation induces an upward motion of the 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 raise the isotherms or equivalently to cool the water at a fixed depth. The impact in terms of sea level is shown in Figure 3a: by decreasing the ocean heat content, the surface elevation is decreased in such a way as to better fit the observations.

Although this approach gives satisfactory results to assimilate SLA data in sub-tropical and mid-latitude regions, the Cooper and Haines method cannot be easily extended to assimilate other observations like SST, or *in situ* temperature and salinity profiles with a primitive-equation model. In addition, the modelling error is not necessarily distributed on the vertical as a Cooper and Haines 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 T/S profiles and along-track altimetric data.

Version 2 (noted 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 PSY-1 MERCATOR North Atlantic prototype at 1/3° and it should be transitioned to the PSY2 prototype very soon. The background error covariance is still constant, but the algorithm performs a reduced order optimal interpolation on the vertical by means of multivariate EOFs (Empirical Orthogonal Functions of barotropic stream function and vertical temperature and salinity profiles) computed from a prior 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 (~ a few tens) and shape of the vertical modes permitted by this approach can be adjusted regionally and seasonally, leading to a better representation of the local physics than with the uni-modal scheme implemented in SAM-1v1 [Benkiran *et al*., 2005]. A variant of this multivariate scheme has been implemented in the MFSPP model of the Mediterranean Sea [Demirov *et al*., 2003], showing its capacity to accomodate 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 a product of 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 behaviors can be expected near boundaries and coastal regions [e.g. Echevin *et al*., 2000]. These limitations motivated the development of a more advanced assimilation system, noted SAM-2, that doesn't 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 mesoscale ocean models. The error statistics of the SEEK filter is 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 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 empirical orthogonal functions 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 dynamically consistent. 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 SAM1v2 in anisotropic and non separable regions of the world ocean such as shallow areas or in the surface layers.

Unlike the original SEEK filter [Ballabrera-Poy *et al.*, 2001], the variant of SAM-2 which is considered here doesn't evolve the error statistics according to the model dynamics [Testut *et al.*, 2003]. This would require prohibitive costs given the size of the operational systems. However, some form of evolutivity 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 model 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 have been tested recently, investigating their respective capacity to control the model trajectory with the PSY1 configuration.

An issue of practical interest is the estimation of small correlations associated with distant variables. In order to prevent the data from exerting a spurious influence at remote distances through large-scale signatures in the EOFs, a simplification of the analysis scheme has been adopted by enforcing 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, centered 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 extend 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 the observation space [Brasseur 2005]. To minimise the computational requirements, the analysis kernel in SAM2 has been massively parallelized and integrated in a generic platform hosting the SAM-1 and SAM-2 kernel families. This platform is driven by the PALM software which makes the coupling between the model codes and the assimilation schemes more effective. 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).

As an illustration of the spatial properties of the background errors used in SAM-2, Figure 4 shows the representer function as defined by Echevin *et al.* [2000] for the sea surface temperature in two different regions: in the equatorial Pacific at 0°N, 140°W (PSY3 configuration) and near the eastern Florida coast at 27°N, 80°W, in the sub-tropical Atlantic (PSY1 configuration). Based on model statistics reflected in the EOFs, the representer indicates how the model SST is modified by the assimilation of one single observation 1°C warmer than the model SST in the absence of observation error. In Figure 4a, the impact of this virtual observation extends eastwards and westwards (about ±10° of longitude) but also 1° north and southward. The shape of the representer is slightly anisotropic, a nice feature due to the EOF analysis of a "good" prior series of model states. In Figure 4b, 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 representer highest values are located near 27°N, 80°W with a meridional structure. Values greater than 0,5°C can be also found east of Florida, which may indicate that SST changes could be also due to surface heat fluxes at larger scale in this particular area.

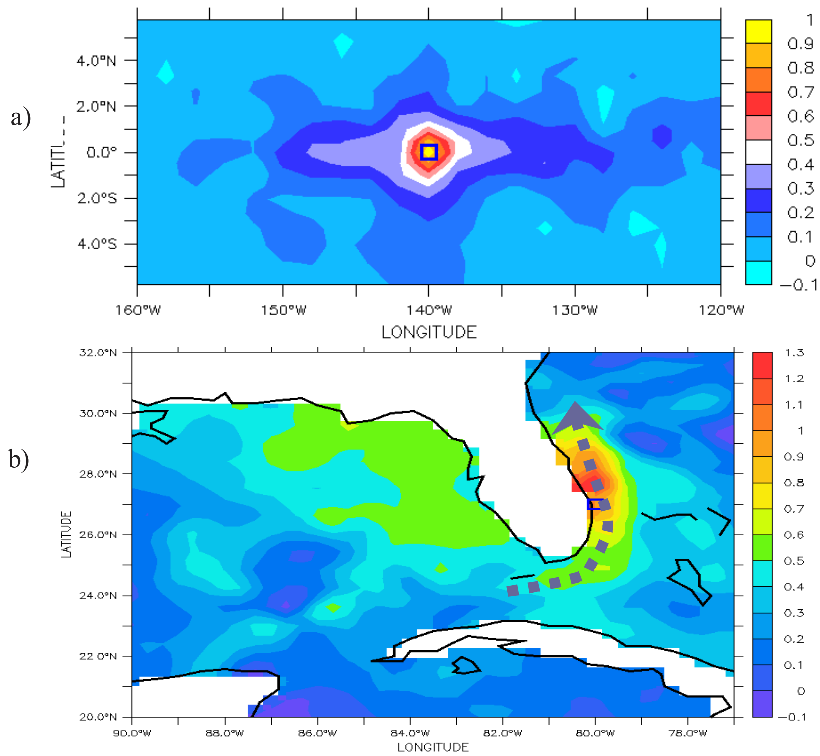

Figure 4: SST representer function of a +1°C sea surface temperature observation anomaly: (a) the representer is computed with respect to a SST point located at 0°N, 140°W (blue square), in the Central Pacific; (b) the representer is relative to a SST point located at 27°N, 80°W (blue square), in the Atlantic. The arrow indicates the Gulf Stream path.

The SAM2 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 data, sea surface temperature). The estimation state vector includes the temperature, salinity and barotropic height model variables, and a geostrophic adjustment is performed after each analysis step to extend the correction to the whole model state. Several hindcast experiments have been conducted during the year 2003 to validate the method with independent *in situ* temperature data. The skill of the assimilation is illustrated in figure 5, showing the vertical distribution of the misfit variance computed between independant (non-assimilated) temperature profiles and the climatology, the control run and the hindcast experiment. The model simulation without assimilation competes reasonably well with the climatology and improves the fit to the data in the sub-surface only. The assimilation improves the temperature field at all depths, with a significant reduction of the error in the thermocline.

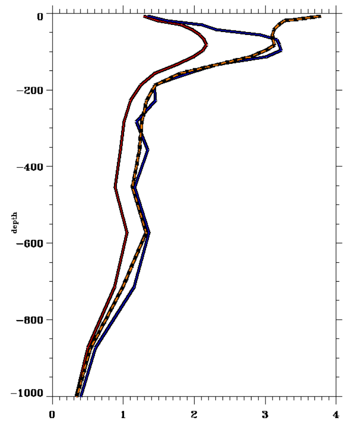

Figure 5: Variance of the temperature misfit (in °C^2) between in-situ data and: (i) the climatology (orange - dashed line), (ii) the control run (blue line) and (iii) the assimilation simulation (red line) until 1000 meter depth during 2003 over the North Atlantic.

The performance of the assimilation to correct non-observed variables such as the surface currents has been verified too. Figure 6 shows the annual mean surface current computed from the hindcast experiment in the region of the Gulf of Mexico, displaying a much more realistic structure of the mean flow by comparison with typical representations obtained from simulations without assimilation.

It is intended to elaborate the future MERCATOR prototypes (global ¼° and regional North Atlantic 1/12° resolution models) on the basis of the SAM-2 assimilation tool, and to pursue the development of the algorithm by improved temporal strategies such as the Incremental Analysis Updating method [Cosme *et al.*, this issue] and new statistical parameterizations 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 mission.

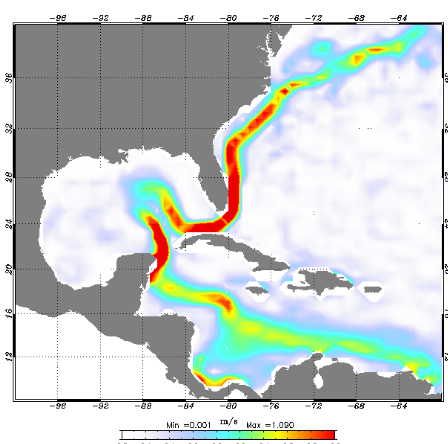

Figure 6. Annual mean surface current (m.s^{-1}) in the region of Gulf of Mexico (Florida strait, Cap Hatteras, Yuccatan strait), computed from a hindcast experiment performed during 2003 with SAM2.

(d) SAM-3

As for the temporal dimension of the assimilation problem, a major simplification is considered in the SAM-1 and SAM-2 schemes since the analysis is performed at regular times intervals which do not necessarily correspond to the measurement times. This was identified as one of the major limitations of operational NWP systems, impacting sometimes severely the forecast performances. The 4D-VAR variational formulation developed with a global ocean model by Weaver *et al*. [2003] takes a rigorous account of the temporal dimension and has the additional advantage of computing increments that are consistent with the linearized model dynamics. The four-dimensional variational formulation is based on the minimization of a cost function that measures the weighted squared difference between observations (including a background state) and the model counterpart. The given solution is the closest model trajectory to the observations and is dynamically consistent with the linearized model equations and the background state. At a given time, this solution is constrained by both past and future observations available in the assimilation window.

The development of a third type of assimilation tool (noted SAM-3) has been initiated in MERCATOR which relies on the OPAVAR system. The technical and scientific features of the algorithm are under investigation to assimilate *in situ* and altimetric data simultaneously with the PSY3 global ocean configuration. The variational approach is particularly well suited to control of the dominant processes taking place in the tropical region where equatorial waves can propagate over several grid points during an assimilation window.

As a first evaluation of the sensitivity of the variational system to the prescribed error statistics, different hindcast experiments have been performed using SAM-3 in a 3D-VAR setup, assimilating temperature and salinity profiles in the PSY3 configuration with different distributions of observation error variance on the vertical. A comparison has been made between a control run, a hindcast experiment perfomed period during the european ENACT project over the 1987-2001 using a constant observation error, and a new hindcast experiment with a depth-dependent observation error. The RMS misfit between the three experiments and the temperature and salinity profiles are shown in Figure 7. A further evaluation of the variational system will be conducted soon by comparing hindcast experiments performed with SAM-2 and SAM-3 in the PSY3 configuration. Some form of hybridation between the 4D-VAR and the SEEK filter is being considered, based on the approach proposed by Cosme *et al.* [this issue].

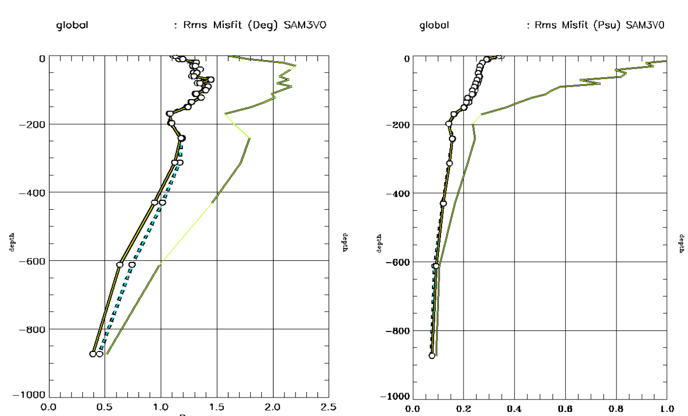

Figure 7: RMS misfit to observations in temperature (left) and salinity (right) for the control run (thin line), with a constant observation error profile (dashed line) compared to a depth dependent error (thick line).

4. Development of an advanced assimilation capacity in MERSEA

(a) The assimilation R&D 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 optimal interpolation (ROOI). 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; therefore, they cannot make optimal usage of all available observations to estimate the state of the ocean and predict its evolution.

The ambition for the MERSEA will be to improve the existing assimilation tools implemented in the european forecasting system, with the aim to (i) address new issues which are particularly relevant for ocean forecasting at high resolution, such as the existance of non-linear processes and non-gaussian statistical behaviours, (ii) extend the assimilation to new data types such as sea-ice parameters or biogeochemical properties in coupled circulation/ecosystem models. The focus in MERSEA is set 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]. Three classes of assimilation methods are being considered:

- very computationally efficient multivariate statistical schemes (e.g. SAM-2 or the ensemble-based OI schemes) which use time invariant approximate error statistics;
- advanced ensemble based methods (EnKF and SEEK) which can evolve the error statistics in time using non-linear models, and thereby provide realistic predictions of error statistics to be used in the analysis scheme. These have been demonstrated during the DIADEM experiment [Brusdal *et al*. 2003] and the EnKF is presently used in the TOPAZ system;
- fully non-linear filters such as the SIR filter, which is a property conserving Monte-Carlo method designed to handle non-Gaussian error statistics.

The progress achieved in these different categories of schemes are illustrated here below in a number of case studies.

(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 2005], which strongly penalize the operational capacity to provide realistic forecasts of the 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 Kalman filter or an optimal interpolation scheme (like the SAM-2 scheme or the SEEK filter with a pre-determined background error) 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 then 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. Instead, 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 [Large 2005] (in spite of a possible risk to compensate errors on 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 configuration as in the MERCATOR PSY3 prototype. The reference simulation (the true ocean) is a standard interannual simulation for the year 1993, with the original bulk formula:

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

where ρ_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 temperature, and

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

where L is the vaporisation latent heat, q_s and q_a are the surface and atmospheric specific humidities. C_E and C_H receive complex parameterizations depending, in particular, on the stability of the air column close to the sea surface. Figure 8 shows the value of C_H in the reference simulation for January 31st, 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 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.

To perform the assimilation experiment with sequential corrections of the bulk coefficients, a relevant background error covariance matrix in the augmented control space has to be provided. For that purpose, an ensemble of ocean models was built using different values of C_E and C_H. From this ensemble of ocean models, an ensemble of 5-day forecasts was computed with a series of initial conditions distributed in time in 1993. The 5-day forecast period was chosen to correspond to the assimilation window of the sequential assimilation scheme. In that way, an ensemble of 5-day forecast anomalies (augmented with the C_E and C_H anomalies) was obtained, and the 20 dominant EOFs were used to parameterize 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 C_E and C_H errors and the ocean state error. If it is used as it is, the above covariance matrix can only generate corrections of C_E and C_H that are constant over the world ocean. Hence, this ensemble will only be used to estimate the local multivariate correlation between the ocean state and the bulk coefficients while we will rely on the local SEEK algorithm [Brankart *et al*., 2003] to eliminate the long range correlation coefficients. In that way, local corrections of C_E and C_H will be obtained.

Figure 8 illustrates the statistical analysis obtained for the augmented state vector after a 5-day forecast from perfect initial condition, using temperature and salinity observations sampled from the reference simulation. The analyzed 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 are in C_E and C_H and the atmosphere and ocean initial conditions are both perfectly known, but it illustrates the theoretical possibility to estimate the turbulent air-sea flux bulk coefficients by inverting oceanic observations. To go further, several 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 ?

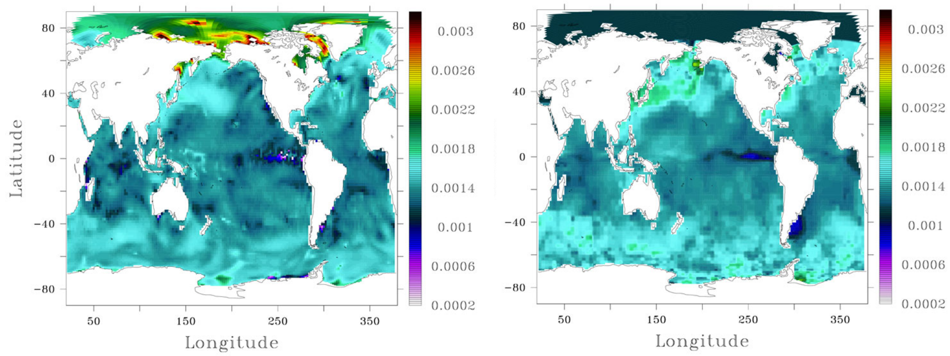

Figure 8. Spatial distribution of latent heat flux coefficient : as obtained in the reference simulation (left pannel), and as reconstructed by the statistical analysis with the augmented state vector (right pannel).

(c) MERSEA developments with the EnKF

Statistical estimation techniques are based on a theoretical framework in the linear case while in the non-linear case one generally resorts to Monte-Carlo methods. In the two-step approach of sequential data assimilation, the Ensemble Kalman filter [Evensen 2003] performs a Monte-Carlo propagation step while the analysis step is based on the linear estimation. This linear analysis step is therefore optimal for multi-Gaussian distributed model state variables in all model grid cells. When this assumption is not respected, the linear analysis can result in a bias that is neither due to observation nor model bias but inherent to the estimation process. In Bertino *et al.* [2003], a common tool of geostatistics, the Gaussian anamorphosis, is suggested to extend the analysis step to a wider class of variables, those that can be transformed to multi-Gaussian by an individual non-linear transformation of each state variable. The resulting analysis of an ensemble of forecast states then allows non-linear estimation, and the method is not restricted to the sole EnKF but applicable to any sequential data assimilation method. Even if this approach assumes that the joint density function of transformed Gaussian variables will remain multi-Gaussian, the ability to reduce estimation biases related to non-Gaussian variables was illustrated on a simple 1-dimensional ecosystem model, on the ground that a non-linear transformation that excludes negative concentrations of ecosystem variables avoids the post-processing of negative estimates obtained by linear estimators, and cancels a large source of estimation bias [Bertino *et al*. 2003].

In 3-dimensional physical ocean models the state variables are less characterized by the accumulation of zero values than the ecosystem variables, but some exceptions to the multi-Gaussian assumption are clear: in variable vertical coordinate models (isopycnal or hybrid such as HYCOM), the layer thickness is a state variable with positive values and obviously does not fit in the Gaussian framework. Thus the TOPAZ system (Arctic component of the MERSEA system, based on HYCOM and the EnKF) has applied a lognormal law to the layer thickness error fields instead of a Gaussian law at the initialization of the EnKF. This avoids the need to post-process negative initial layer thicknesses and the introduction of nondifferenciable fields. The lognormal law has however not proved satisfactory in the analysis step because of local particular cases. The very thin model layers in regions of strong gradients or of outcropping create spurious ensemble correlations. Further, the layer thickness distribution presents some diracs in the regions transitioning form isopycnal layers to fixed z-levels. There the lognormal law does not seem more appropriate than the Gaussian and it seems that an anamorphosis of hybrid layer thicknesses should be adapted by regions rather than applied globally. The use of a linear multivariate update in fixed coordinate models appears more satisfactory since all state variables (velocities, temperature and salinity) seem reasonably Gaussian. However, the counterpart is that a multivariate linear analysis cannot conserve non-linear relationships between these variables (i.e., density), so that non-linear transformations might be considered even for Gaussian state variables. A natural extension of this idea would be the assimilation of categorical variables by analogy with plurigaussian conditional simulations [Armstrong *et al*. 2001] assuming underlying Gaussian variables which classes determine different categories.

(d) MERSEA developments with the SIR filter

The Sequential Importance Resampling (SIR) filter is an ensemble-based Monte Carlo method similar to the Ensemble Kalman Filter (EnKF). However, the SIR updates probabilities of the ensemble members and not the ensemble states themselves. This difference makes the SIR filter to be a truly variance minimizing scheme, which can be easily applied for non-linear systems with any probability density function (pdf) that is not necessarily Gaussian.

Such a method just perfectly suits the problem of data assimilation into strongly non-Gaussian ecosystem models that MERSEA needs to investigate. Monitoring the ocean environment indeed 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 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 Bermuda Atlantic Time-series Study data set, i.e. real observations were used. The obtained model solution agreed reasonably well with the data, even with independent (not assimilated) ones. The model parameters however revealed strong seasonal variations, which may point to possible uncertainties in parameterization of biological processes. This motivated us to try to estimate the model noise level in the data-assimilation scheme also.

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 Sequential Importance Resampling (SIR) 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 ψ_i is calculated from

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

which follows directly from Bayes theorem (see e.g. Kivman, 2003; Losa *et al.*, 2003; Van Leeuwen, 2003). $p(d \mid \psi_i)$ is the pdf of the observations given this ensemble member i. The specific form for this pdf will be presented below. This leads to a weighted ensemble, and each moment, like the mean of the pdf, or its variance, is 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 of the effective size of the ensemble. To avoid this, the ensemble is resampled after each update, to give each member equal weight again. This resampling can be done in several ways. Here we used the weights as probabilities on the members, and draw a new ensemble from the resulting probability, with replacement[1].

The result of the resampling is an ensemble of the same size, in which some ensemble members are identical, and other have disappeared from the ensemble. This new ensemble is then integrated forward in time to the next observation time, where the weighting and resampling is repeated. To increase the spread in the resampled ensemble jittering is sometimes applied, in which identical members are slightly perturbed. No general rule exists on the size and form of the perturbations. In the experiments described here some jitter is 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 - presented 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. A number of 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

$$p(\boldsymbol{a}) = \bar{\boldsymbol{a}} \, \mathbf{exp}(-\frac{\boldsymbol{a}}{\bar{\boldsymbol{a}}})$$

where $\bar{\boldsymbol{a}}$ 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 9 additional parameters added to 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

$$w_i = C \, \mathbf{exp}\left(-0.5(d-\psi_i)^2 / s^2\right),$$

where s^2 is the variance of the observation.

As mentioned above, jitter is applied to all the optimized parameters of the resampled ensemble to avoid ensemble collapse. The jittering procedure is as follows. If, at an analysis step, some parameter values *a* are resampled many times, a uniform pdf is created in the interval (*a*- nearest smaller value, *a*+ nearest higher value), and new parameter values are drawn from this uniform density.

Figure 9 (upper panel) depicts the temporal evolution of one of 15 optimized model parameters (red solid curves), obtained from the experiment with the variable level of the

[1] 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 resampling procedures.

model noise, against previous estimates of the same parameters (black solid curves) when the model-noise level is constant in time Losa *et al.* [2003]. 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 the bottom panel of figure 9, which depicts the model noise variance over the period 1989-1994.

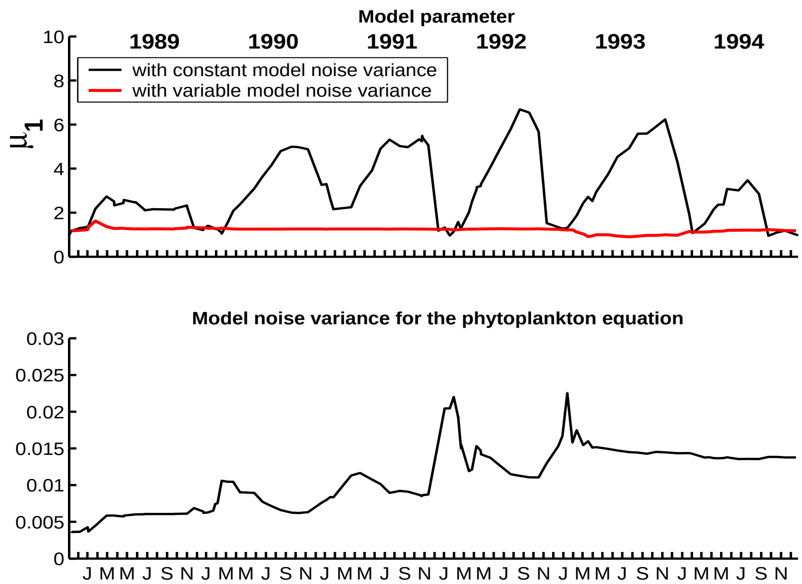

Figure 9: The evolution of the estimates for one of the optimized biological model parameters (upper panel) and model noise variance (bottom panel). All the parameters are normalized with respect to their initial values.

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

In principle one would like to run the SIR-scheme with parameter and noise estimation along with the biogeochemical model. Unfortunately, that 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 instant. 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 systems. In the MERCATOR context, advantage is taken from the whole spectum of complexity that assimilation methods can offer to provide the best possible depiction 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 non-linear processes more rigorously.

The GODAE initiative has contributed to establish a feedback loop in the oceanographic community, which guarantees a coherent progress between upstream research, R&D activities, operational applications, and the users. This is also pivotal ambition of the MERSEA project. The expected benefit will be to improve the quality standards of operational products and services, and thereby strengthen the credibility in, and sustainability of operational oceangraphy through the users'satisfation.

ACKNOWLEDGEMENTS

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

REFERENCES

Armstrong, M., Galli, A., Le Loch, G., Geoffroy, F. and Eschard, R. 2001. *Plurigaussian simulations*, Kluwer, Dordrecht.

Bahurel P., 2005 : MERCATOR OCEAN global to regional ocean monitoring and forecasting. In : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century (J. Verron and E. Chassignet Eds.)*, Kluwer Academic Press, in press.

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. Met. Soc*., **127**, 233-253

Benkiran, M., Greiner, E., Dombrowsky, E., 2005. The Multi variate multi data assimilation in the Mercator project. *J. Marine System* special issue, in revision.

Bertino, L., Evensen, G., 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. Letters,*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.

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., 2005 : Ocean Data Assimilation using Sequential Methods based on the Kalman Filter. In : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century (J. Verron and E. Chassignet Eds.)*, Kluwer Academic Press, in press.

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.

Cooper M. and K. Haines, 1996: Altimetric assimilation with water property conservation, *J. Geophys. Res.*, **101** (C1), 1059-1077.

Cosme E., F. Castruccio, Y. Ourmières, C. Robert, J. Verron and P. Brasseur, 2005 : Recent advances in ocean data assimilation with the SEEK filter, *Q. J. R. Met. Soc., submitted.*

De Mey, 1997. Data assimilation at the oceanic mesoscale: a review. *J. Met. Soc. Japan,* Special issue on "Data assimilation in meteorology and Oceanography: Theory and practice", **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, N. Pinardi and J.D. Woods, Eds, Springer Verlag, Berlin, Heidelberg, New York, 472 pp.

Demirov E., N. Pinardi, C. Fratianni, M. Tonani, L. Giacomelli, and P. De Mey, 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. *Physics and Chemistry of the Earth,* **25**, 503-509.

Drillet, Y., R. Bourdalle-Badie, L. Siefridt, and C. Le Provost. 2005. The MEDDIES in the Mercator North Atlantic and Mediterranean Sea eddy-resolving model., *Journal of Geophysical Research*, Vol. **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., 2003: The Ensemble Kalman Filter theory and practical implementation*, Ocean Dyn*., **118**, 1-23.

Ferry, N., E. Remy, P. Brasseur, C. Maes, 2005: The Mercator global ocean operational analysis / forecast system: assessment and validation of an 11-year reanalysis. J. of Marine Systems, in press.

Fu L.-L. and Cazenave A., 2001 : Satellite Altimetry and Earth Sciences, a Handbook of Techniques and Applications, *Academic Press, International Geophysics Series,* Vol. **69**, 463 pp.

Fukumori I., 2001 : Data assimilation by models. In : *Satellite Altimetry and Earth Sciences, a Handbook of Techniques and Applications (Fu L.-L. and A. Cazenave Eds.), Academic Press, 237-265.*

Griffies S., 2005 : Some ocean models fundamentals. In : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century (J. Verron and E. Chassignet Eds.)*, Kluwer Academic Press, in press.

Kivman, G.A., 2003. Sequential parameter estimation for stochastic systems. *Nonlinear Process. Geophys*. **10**, 253-256.

Large W., 2005 : Surface fluxes for practitioners of global ocean data assimilation. In : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century (J. Verron and E. Chassignet Eds.)*, Kluwer Academic Press, in press.

Liu J.S. and Chen R., 1998 : Sequential Monte-Carlo methods for dynamic systems, *J. Amer. Stat. Ass*., **93**, 1032-1044.

Losa, S.N, Kivman, G.A., Schroeter, J., Wenzel, M., 2003. Sequential weak constraint parameter estimation in an ecosystem model. *Journal of Marine Systems*, **43**, 31-49.

Madec G., Delecluse P., Imbard M., Lévy C. 1998. OPA 8.1 ocean general circulation model reference manual, *Notes du pôle de modélisation IPSL*, 91 pp.

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., J. Verron and M. C. Roubaud, 1998 : A singular evolutive extended Kalman filter for data assimilation in oceanography, *J. Mar.Syst*., **16**, 323-340.

Smith N., 2005 : Perspectives from the Global Ocean Data Assimilation Experiment. In : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century (J. Verron and E. Chassignet Eds.)*, Kluwer Academic Press, in press.

Testut C.-E., P. Brasseur, J.M. Brankart and J. Verron, 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. Syst*., **40-41**, 291-316.

Van Leeuwen, P.J., 2003. A truly variance minimizing filter for large-scale applications, *Mon. Weather Rev*. **131**, 2071-2084.

Verron J., 1992 : Nudging altimeter data into quasi-geostrophic ocean models, *J. Geophys. Res,* **97**(C5), 7497-7491.

Verron J. and Chassignet E., 2005 : *GODAE, an Integrated View of Oceanography : Ocean Weather Forecasting in the 21st Century*, Kluwer Academic Press, in press.

Vörösmarty, C.J., B. Fekete, and B.A. Tucker. 1996. River Discharge Database, Version 1.0 (RivDIS v1.0), Volumes 0 through 6. A contribution to IHP-V Theme 1. *Technical Documents in Hydrology Series*. UNESCO, Paris.

Weaver A., J. Vialard and D. L.T. Anderson, 2003: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks, *Mon. Wea. Rev*., **131**, 1360-1378.