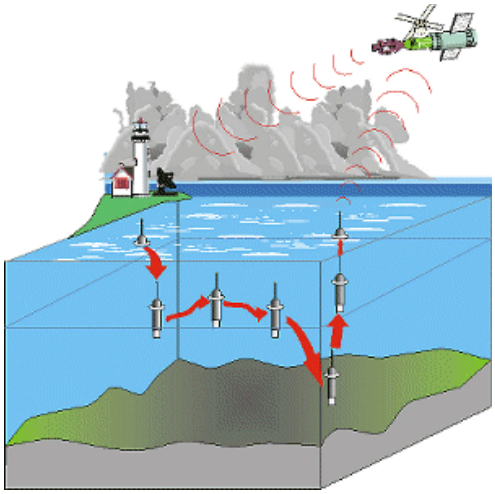
MERCATOR

Quarterly Newsletter

Editorial – July 2009



Argo floats Temperature and Salinity profiles (left panel), as well as satellite altimeter (right panel) Sea Surface Height Anomaly constitute the 2 main observations sources which are assimilated into the Ocean Forecast systems. Left Panel: Measurement cycle for an Argo float. Credit: Ifremer. Right panel: Jason-2, French-American satellite (Cnes/Nasa) measuring sea surface height. Credit: AVISO

Greetings all,

This month's newsletter is devoted to Data Assimilation and its techniques and progress for operational oceanography.

Gary Brassington is first introducing this newsletter with a paper telling us about the international summer school for "observing, assimilating and forecasting the ocean" which will be held in Perth, Western Australia in 11-22 January 2010 (http://www.bom.gov.au/bluelink/summerschool/). The course curriculum will include topics covering the leading edge science in ocean observing systems, as well as the latest methods and techniques for analysis, data assimilation and ocean modeling.

Scientific articles about Data Assimilation are then displayed as follows: The first article by Broquet et al. is dealing with Ocean state and surface forcing correction using the ROMS-IS4DVAR Data Assimilation System. Then, Cosme et al. are describing the SEEK smoother as a Data Assimilation scheme for oceanic reanalyses. The next article by Brankart et al. is displaying a synthetic literature review on the following subject: Is there a simple way of controlling the forcing function of the Ocean? Then Ferry et al. are telling us about Ocean-Atmosphere flux correction by Ocean Data Assimilation. The last article by Oke et al. is dealing with Data Assimilation in the Australian BlueLink System.

The next October 2009 newsletter will review the current work on ocean biology and biogeochemistry.

We wish you a pleasant reading!

Contents

The International Summer School for Observing, Assimilating and Forecasting the Ocean: 11-22 January 2010 in Perth Australia

By Gary B. Brassington

Bureau of Meteorology, BMRC, GPO Box 1289K, Melbourne, Victoria 3001, Australia

Over the past decade the international ocean community has largely succeeded in the implementation of the Global Ocean Observing System as was planned in the OceanObs'99 meeting. Today, we now enjoy a sophisticated in situ observing systems of an Argo array in excess of 3000 floats that spans the worlds ocean basins, a surface drifting buoy network on the order of 1250 floats as well as repeat XBT lines and an expanding mooring network in the equatorial waveguides through the TAO/TRITON, PIRATA and now RAMA mooring arrays. The introduction of gliders, OceanSITES and HF radars continues to show an ever increasing range of instrumentation to complement and extend GOOS over the next decade, the subject of the forthcoming OceanObs'09.

The past decade has similarly seen the introduction of a sophisticated range of satellites that remotely observe the ocean's surface properties including sea level, SST and ocean colour. Satellite SST is the most sophisticated network with IR and Microwave instrumentation, sun synchronous and geostationary orbits offering both high accuracy and high data coverage. Satellite altimetry has successfully maintained the high quality altimetry missions of Topex-Poseidon, Jason and now Jason-2. Similarly the ERS, ERS2 and Envisat have provided a continuum of high latitude observation of altimetry. Geosat and GFO have similarly contributed to the altimetry observations where multiple narrow-swath tracks are required to observe the oceans mesoscale. Plans for wide-swath and constellations offer exciting new advances for oceanography and ocean prediction. New platforms to observe the sea surface salinity are nearing their launch with SMOS and Aquarius which will provide the first global measurements of this important property.

Collectively, GOOS represents an explosion in the volume of high quality information about our oceans state and circulation. This does not mean that the job is done, there remain many mysteries in our understanding of the ocean circulation particularly in respect to the role of eddies both in terms of transports and abyssal circulation on short and longer timescales. Similarly the integration of GOOS with coastal observing systems will also lead to new advances in our understanding of ocean dynamics. None the less the observing system we currently enjoy, if maintained at present levels will continue to provide new understanding about the world's ocean for years to come.

In addition to objectives of improvements to coverage and quality, the ocean scientific community has been successful in adhering to the new paradigm of open access policies for data (and their processing) and real-time reporting. Both of these changes have been fundamental to the rapid progress that continues to be made in this field and the corresponding expansion in capability to many new countries. To date these goals have largely been achieved for GOOS, however the community will need to remain vigilant as to the benefits ocean and climate science and services in order to sustain this paradigm into the future. Encouraging all members to adhere to this new paradigm is an important ongoing activity. Demonstrating the benefit of this new capability to short- to medium-range ocean prediction was central to the Global Ocean Data Assimilation Experiment (GODAE). Throughout the period of GODAE the ocean state and circulation has been demonstrated to be predictable to a measurable level of skill by ocean forecasting systems. The success story of GODAE (and GOOS), which concluded in 2009, has been the implementation of operational ocean prediction systems and services by international centres spanning Europe, North America, Asia and Oceania with activities emerging in both Africa and South America. Some of these systems are now second generation and support a growing range of public and commercial services and applications.

Ocean forecasting continues to be a rapidly developing area of scientific research and technological development. Ocean service providers will require continuous improvement in performance and quality of services if they are to meet the demands of an ever increasing and diverse user community. Continued progress in this field requires an expanding supply of talented scientists and technicians to continue to innovate and develop the next generation systems and downstream and coupled services. It is a serious challenge for university curriculum to track the leading edge of any science. This is particularly the case for ocean forecasting where both the technological infrastructure and research scientists reside largely within research institutions and agencies that are outside the university system. One forum that serves to fill this gap is summer schools. Summer schools are short courses composed of subjects delivered by leading scientists targeted for students that have the aptitude and aspiration to pursue a given specialised field. GODAE supported a first summer school that was held in Lalonde les Maures, France in 2004 (http://www-meom.hmg.inpg.fr/Web/GODAE_school/) which was very successful with many new scientists emerging and the lecturers contributing to a book publication (Ocean Weather Forecasting: An Integrated View of Oceanography, Chassignet, Eric P.; Verron, Jacques, Eds.).

GODAE and JCOMM are coordinating the second in this series called the International Summer School for Observing, Assimilating and Forecasting the Ocean to be held in 11th-22nd January 2010 in Perth Australia (http://www.bom.gov.au/bluelink/summerschool/). The objectives for this summer school remain the same as that stated for the

first summer school, "to train the next generation of young scientists and professionals who will be the developers and users of operational oceanographic outputs". However since 2004, there have been numerous advances: the operational systems and their performance; introduction of new observing system technologies; advances in data portal technologies; developments of downstream products and services; and the emergence of ocean forecasting support services; all of which require an updated curriculum. In addition, there is an on-going training need for existing systems to maintain and advance their systems as well as support the capacity emerging in many new countries particularly India, China, Brazil and South Africa. This two week course will provide an up to date curriculum of lectures and tutorials on ocean modelling, ocean data assimilation, ocean observing systems, operational oceanography and their applications. The course material will be delivered by approximately 30 leading experts including: Pierre Brasseur; Harley Hurlburt; Pierre-Yves Le Traon; Eric Chassignet; Eric Dombrowsky; Andy Moore; and many more. The summer school is aiming to provide approximately 60 places for early career researchers and early career professionals that intend to pursue a career in this field. Later career professionals will also be considered subject to availability of places. Students should register their interest in attending the summer school through the online form on the website before the 31st July 2009. Students that are currently registered can be found online (http://www.bom.gov.au/fwo/IDZ00033.html). Student places will be announced by September. A limited number of travels, accommodation scholarships will be made available and additional funds are being sort particularly for students from regions that have less capacity to support students. It is important to note that summer schools can not occur without significant sponsorship to assist students to attend. The summer school organisers gratefully acknowledge the Australian Bureau of Meteorology, CSIRO, University of Western Australia, NOAA, the IOC Perth Office and the many national funders for their support.

Ocean State and Surface Forcing Correction using the ROMS-IS4DVAR Data Assimilation System

By Grégoire Broquet[1], Andrew M. Moore[1], Hernan G. Arango[2], Christopher A. Edwards[1] and Brian S. Powell[3]

[1]University of California, Santa Cruz, CA, USA
[2]Rutgers University, New Brunswick, NJ, USA
[3]University of Hawaii, Manoa, HI, USA

Abstract

The Incremental Strong constraint 4D-Variational (IS4DVAR) system of the Regional Ocean Model System (ROMS) for data assimilation is presented. It has been applied to the California Current System (CCS) to assimilate remotely-sensed and in situ oceanic observations. Results from both twin and realistic experiments are presented. The IS4DVAR control vector is comprised of the model initial conditions and surface forcing, and a focus is given on the ability of ROMS-IS4DVAR to correct the latter. ROMS-IS4DVAR always reduces misfits between the model and the observations that are assimilated. However, without corrections to the surface forcing, the assimilation of surface data can degrade the temperature structure at depth. This behavior is prevented by using surface forcing adjustment in ROMS-IS4DVAR which can reduce errors between the model and surface observations through corrections to surface forcing rather than to temperature at depth. Wind stress corrections generate abnormal spatial and temporal variability in its divergence and curl. However, twin experiments indicate that corrections to wind stress and surface heat flux tend to reduce errors in these forcing fields.

Introduction

Three 4D-Variational (4DVAR) data assimilation systems have been developed for the Regional Ocean Model System (ROMS): the Incremental Strong constraint 4DVAR (IS4DVAR), the Physical-state Statistical Analysis System (4DPSAS) and the representer based 4DVAR (R4DVAR), which will be described in detail by Moore et al. (2009) (in preparation). This ensemble of systems is unique for a community ocean general circulation model.

All three 4DVAR systems adjust the model initial condition at the beginning of assimilation windows. While both 4DPSAS and R4DVAR allow for weak constraint data assimilation (with the adjustment of some increments added to the model state vector at each time step, to account for model error), IS4DVAR is technically limited to strong constraint data assimilation (assuming the model is perfect), which is the main reason for the development of the different methods. However, options for adjusting the surface forcing and open boundary conditions at each time step have also been developed for all three systems. Options for preconditioning the minimizations inherent to each 4DVAR technique (Tshimanga et al., 2008), and for modeling the multivariate error covariances of the background control state (Weaver et al 2005) are also available for all ROMS-4DVAR systems.

The purpose of this paper is to illustrate the capabilities of the ROMS-IS4DVAR. Powell et al. (2008) and Broquet et al. (2009) present some realistic applications of ROMS-IS4DVAR respectively in the Intra-America Seas and in the California Current System (CCS), but only the initial conditions were adjusted in those cases. The impact of adjusting the surface forcing is described here. Corrections on the initial condition allow for controlling the mean position and structure of the currents in the CCS (Broquet et al., 2009). But in this region where upwelling phenomena dominate the variability in the coastal dynamics, a major potential source of error is local wind forcing (Veneziani et al., 2009a and 2009b). On the other hand, errors in heat and fresh water fluxes are known to be important (Doyle et al., 2009). The use of surface forcing adjustment therefore provides important information about the nature of errors in either the wind and thermal forcing, or in the way they are imposed in ROMS. Experiments with correction of surface forcing using ocean data assimilation can already be found for example in Stammer et al. (2002). Vossepoel et al. (2004) reveal difficulties when attempting to reconstruct known wind stress variations from oceanic data, despite clear improvements in the ocean circulation from the wind stress adjustments themselves.

Section 2 summarizes the principles of ROMS-IS4DVAR and of the surface forcing adjustment in ROMS-IS4DVAR. Section 3 presents the ROMS/ROMS-IS4DVAR CCS configurations. Section 4 shows some results from the use of ROMS-IS4DVAR in the CCS, illustrating the impact of the surface forcing adjustment. Conclusions are given in section 5.

Sequential adjustment of initial condition and forcing in ROMS-IS4DVAR

ROMS-IS4DVAR is derived from the incremental 4DVAR system of ECMWF (Courtier et al. 1994; Weaver et al. 2003). The theory underlying this method and an extensive bibliography can be found in these papers.

ROMS-IS4DVAR can be applied sequentially on cycles than span a desire time period. For each given cycle the model control variables are corrected to ensure that the model is closer to the observations $\mathbf{y}^o$ in a least-squares sense. The control space vector is denoted **s**, the observation space vector **y**, and the model space vector **x** (each space being dependent on location, time, and the physical nature of each comprised variable). In most data assimilation systems **s** is composed of the assimilation cycle initial condition $\mathbf{x}(t_0)$, but it can be enlarged, as is the case for ROMS-IS4DVAR, to surface and open boundary forcing fields at every time step.

Consider the operator *G*, composed of the model *M* and operators that transform the model **x**-space into the observation **y**-space at a given time. *G* projects **s**-space into the **y**-space. Linearizing *G* along a model trajectory generated from a given control vector **s** yields the tangent linear operator **G** and its adjoint $\mathbf{G}^T$. **G** is used to evaluate changes in the model solution projected in the **y**-space from small corrections on **s**, while $\mathbf{G}^T$ is used to evaluate the sensitivity of the difference between the observations and the model solution in **y**-space to changes in **s**-space. In practice, $\mathbf{G}^T$ and **G** are based on the code for the tangent linear and adjoint model of the model *M* that were developed for ROMS by Moore et al. (2004).

Combining these linear operators, an estimate of the optimal correction $\delta\mathbf{s}^a$ in **s**-space during a given cycle can be obtained, that minimizes, in a least square sense, the difference between the model solution and both the observations $\mathbf{y}^o$ and a background estimate $\mathbf{s}^b$ of **s**, typically from a previous assimilation cycle. When using ROMS-IS4DVAR, these differences are minimized in the **s**-space by minimizing the following cost function:

$$J(\delta s) = \frac{1}{2}(\delta s)^T B^{-1}\delta s + \frac{1}{2}(G\delta\delta - d)^T R^{-1}(G\delta\delta - d)$$

where $\mathbf{d}=\mathbf{y}^o-G(\mathbf{s}^b)$; differences to the observations and to $\mathbf{s}^b$ are weighted with the inverse covariance matrices, respectively **R** and **B**, of the error on these sources of information that are assumed Gaussian distributed and independent. The determination a priori of **R** and **B** is a major issue for data assimilation systems. The cost function *J* is quadratic and the minimizations are performed using a conjugate gradient algorithm.

Here $\mathbf{s}^b$ is comprised of the initial condition from the data assimilation analysis from the preceding cycle and the surface forcing for the current cycle. The model trajectory generated with $\mathbf{s}^b$ will be called the forecast. After data assimilation and estimation of $\delta\mathbf{s}^a$, the model *M* is run with the adjusted control vector (the sum of $\mathbf{s}^b$ and the estimate of $\delta\mathbf{s}^a$) to yield the analysis for the present cycle, and a background initial condition for the following cycle.

ROMS/ROMS-IS4DVAR CCS configurations and observations

ROMS (Shchepetkin and McWilliams, 2005; Haidvogel et al., 2008) denoted *M*, is a terrain following vertical coordinate model with **x** composed of the 2D sea surface elevation (SSH), and the 3D temperature (T), salinity (S) and zonal and meridional velocities (u,v). The ROMS CCS configurations cover the domain 134°W - 115.5°W and 30°N - 48°N with Mercator grids, and they use realistic bathymetry from ETOPO2 (cf Figure 1). The 1/3° horizontal resolution configuration (WC3) and 1/10° horizontal resolution configuration (WC10) use respectively 30 and 42 vertical levels. WC10 is described in detail in Veneziani et al. (2009a and 2009b). In both cases, GLS k-omega vertical mixing scheme, harmonic horizontal viscosity (mixing coefficient: 4 $m^2\ s^{-1}$), free-slip boundary conditions, a quadratic bottom friction (bottom drag coefficient: 2.5×10^{-3}) and a sponge layer with higher viscosity near the open boundaries are used. Open boundary conditions are imposed using monthly averaged data from the Estimating the Circulation and Climate of the Ocean - Global Ocean Data Assimilation Experiment (ECCO-GODAE) model, Flather condition for the barotropic variables, and clamped conditions for the baroclinic variables. The surface forcing data of meridional and zonal wind stress $\tau=(\tau_u,\tau_v)$, heat and fresh water flux (Q_T and Q_W) are derived from the daily data of the multi-grid atmospheric component of the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS) model (Doyle et al., 2009). In WC10, Bulk fluxes computations are used with atmospheric parameters from COAMPS, while in WC3 τ, Q_T and Q_W are directly imposed (to test their adjustment with data assimilation).

WC3 was used to test the efficacy of various ROMS-4DVAR configurations, while WC10 was used to evaluate the improvements in the CCS circulation dynamics from data assimilation in a more realistic case. *M* was run for the period 1999-2004, initialized from a 7 years spin-up (with climatological forcing) in both CCS configurations. This so called free run will be denoted FREE10 (FREE3) in WC10 (WC3). Data assimilation was conducted with 14 days cycles, covering in WC10 the period 1999-2004 and in WC3 the period 2000-2004 (using the state of respectively FREE10 and FREE3 at the beginning of those periods as the first background initial condition).

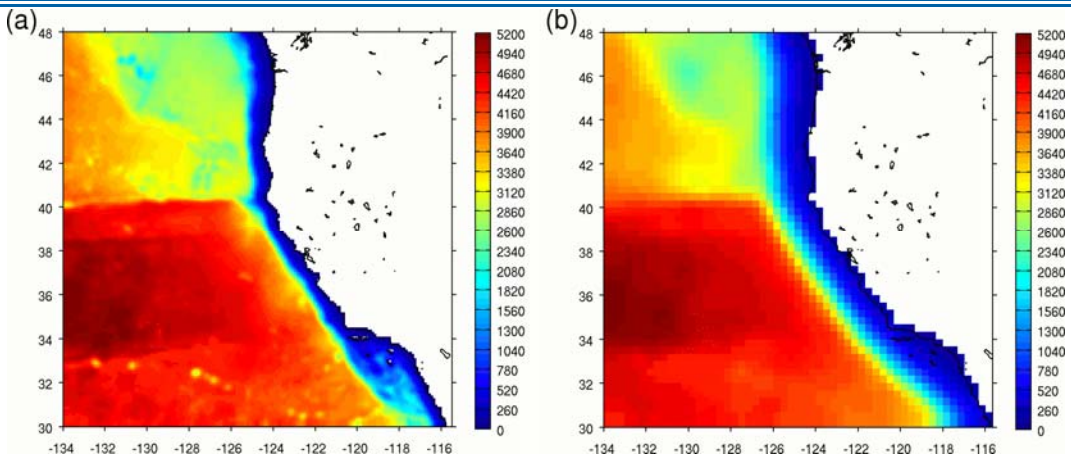


Figure 1

Bathymetry (in m depth) of (a) WC10 (b) WC3

Data assimilated for realistic experiments in the CCS are gridded Aviso SSH anomalies (available every 7 days with an horizontal resolution ~1/3°), daily gridded sea surface temperature (SST) from COAMPS (of higher resolution than 1/10° everywhere, but interpolated at each model resolution before assimilation), T and S profiles from CTD measurements of California Cooperative Ocean Fisheries Investigation (CalCOFI) and GLOBal ocean ECosystems dynamics (GLOBEC), Northeast Pacific (NEP)/ CCS mesoscale survey cruises (see Figure 2).

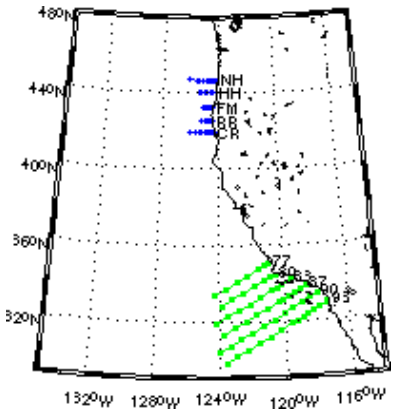


Figure 2

Typical location of measurements used: GLOBEC (blue) and CalCOFI (green)

The choice of data assimilation parameters for ROMS-IS4DVAR in the CCS is described in detail by Broquet et al. (2009) and summarized here: **R** is assumed diagonal with observational errors assigned to either the measurement error or the error of representativeness, whichever is higher. Measurements error were chosen with the following standard deviations: 0.2 cm for SSH, 0.4° for SST, 0.1° for in situ T and 0.01 psu for in situ S. The standard deviation of the observations values within a grid cell is used for the error of representativeness. **B** was modeled with spatial univariate correlations, from a diffusion operator (Derber and Bouttier, 1999) and assumed to be temporally uncorrelated. Correlations between errors in τ_u and τ_v or between errors in surface forcing from different time are not included, although $\mathbf{G}^T$ provides some multivariate correlations and correlations in time via the model dynamics. Decorrelation length scales used here, for all control variables of the initial condition are as follows: horizontally 50 km, and vertically (for 3D variables) 30 m; for the surface forcing: 100 km horizontally. The surface forcing is corrected every time-step by interpolation of daily adjustments in the control space, since COAMPS forcing data are daily averages. This provides continuity in the corrections applied to surface forcing. Standard deviations in **B** are estimated monthly with the model temporal variability (from daily outputs of FREE10/FREE3) for control variable of the initial condition and with the COAMPS temporal variability (from daily data) for the surface forcing.

The minimization of J was interrupted after 20 loops of minimization on WC3 and 10 loops on WC10 because it would be practically impossible to reach the point when $\nabla J \sim 0$. However it has been shown (Broquet et al., 2009) that the solution after 10/20 loops accounts for most of the decrease in J toward the true minimum of J. It is also possible with ROMS-4DVAR methods to update regularly the model trajectory used to yield **G** and $\mathbf{G}^T$ from the last estimate of $\delta\mathbf{s}^a$, with the so called outer-loops (at each outer loop a minimization of *J* is conducted, yielding a new estimate of $\delta\mathbf{s}^a$, and thus a new model trajectory is used in the following outer-loop). This option does not yield significant impact on the results in WC10 and WC3, and it is not used here (we use only 1 outer-loop).

Results from ROMS-IS4DVAR in WC10 and WC3

Realistic experiments

Broquet et al. (2009) describe experiments of sequential initial condition adjustments from assimilation of Aviso SSH, COAMPS SST, CalCOFI and GLOBEC data with ROMS-IS4DVAR in the CCS, that are denoted ALL10 in WC10 and ALL3 in WC3 (cf Table 1). Some statistical comparisons with the COAMPS SST and CalCOFI winter S data for FREE10 and ALL10 are shown in Figure 3. Difference between the model and the observations will hereafter be referred to as "error" (even though the true errors are the differences between the model and the "true circulation"). Data assimilation improves the circulation after analysis during each cycle, and when forecasting during the following cycle, as shown on the SST (Figure 3a). This improvement is apparent both at the surface (Figure 3a) and at depth (Figure 3b). Both the model bias, and the standard deviation (STD) of the error are reduced, leading to an overall decrease in the Root Mean Square (RMS) error. The decrease in the error STD during data assimilation is not only due to a better fit of the model spatial and temporal variability to the observed variability, but also to an increase of spatial and temporal correlation between the model and the observations (not shown). The system, when adjusting the initial conditions, is able to anticipate and correct the main sources of error within a range of more than 14 days. However, the error computed from observations that are not assimilated shows that such improvements apply far more to the mean circulation than to the circulation variability or to the correlation between the model and the observations. Another issue is that most of the model corrections to fit the observed SST and SSH occur in the mixed layer and the thermocline, while **B** has a tendency to smooth these corrections, resulting in a diffuse thermocline, and a warm bias between 50 m and 100 m for most of the year.

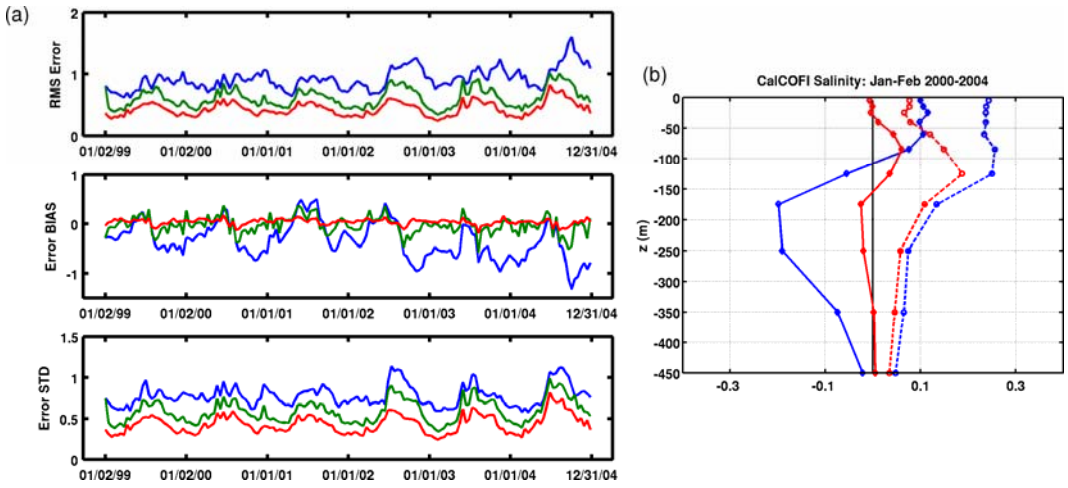


Figure 3

(a) Error statistics (in °C) for each assimilation cycle in WC10 relative to all SST COAMPS data available during each cycle; (b) Error bias (continuous lines) and Error STD (dashed lines) in salinity (in psu) computed from CalCOFI measurements in June-July during the period 2000-2004 as a function of depth. Blue: FREE10; green: forecast in ALL10; red: analysis in ALL10 (source: Broquet et al. 2009)

Corrections applied regularly by ROMS-IS4DVAR in ALL10 and ALL3 exert an influence on the circulation far from observations and on unobserved variables by virtue of the adjoint model and of the model *M* during both analysis and forecast cycles. This results in significant modifications to the whole circulation in the CCS. This is illustrated by the vertical section of summer mean meridional velocities at the latitude of Monterey Bay, shown in Figure 4. The broad equatorward California Current and the poleward California Under Current have a very different core position and intensity in FREE10 (Figure 4a) and ALL10 (Figure 4b). The alongshore currents from the regularly sampled CalCOFI lines 80 and 90 in ALL10 agree more favorably with published geostrophic estimates (e.g. Bograd and Lynn, 2003) than those of FREE10 (not shown). At 37°N, however, there are few

observations, but the currents there are similar to those along lines 80 and 90 in both ALL10 and FREE10, which would suggest that data assimilation has had an equally positive impact on the circulation in this poorly observed region.

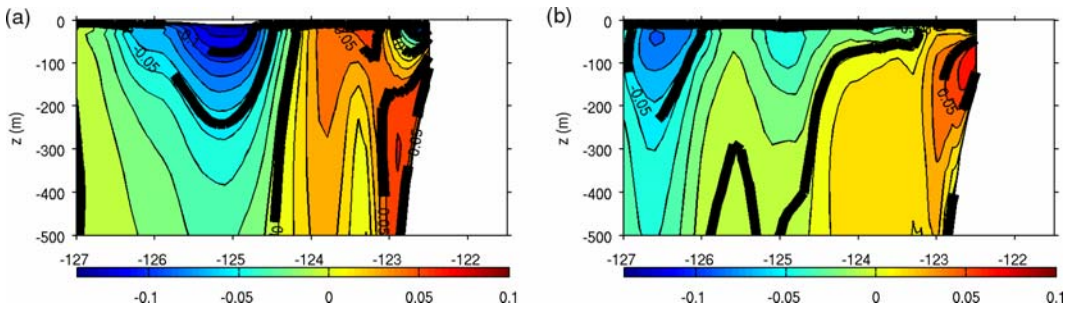


Figure 4

Summer (July–September) mean for 2000–2004 of meridional velocity in ms^{-1}: vertical cross-shore section at latitude 37°N in (a) FREE10 and (b) ALL10 (source: Broquet et al. 2009)

Twin experiments with forcing adjustment

Adjustment of wind stress and surface heat fluxes was initially tested in WC3 using identical twin experiments. During these experiments, the model is run from January 1 2000 with the same parameters as in FREE3 except that it uses COAMPS data for τ or for Q_T lagged by 1 year. The new free simulations obtained are denoted TW-FREE3 and TQ-FREE3 respectively. FREE3 is used as a surrogate for the true circulation, and observations were extracted from this simulation to replicate the distribution of observations used in ALL3: weekly SSH, daily SST, T and S at CalCOFI and GLOBEC locations and time of measurement. Perfect observations are assimilated in simulations conducted with the same surface forcing as TW-FREE3 (TQ-FREE3), denoted TW-ALL3 (TQ-ALL3) when only the cycle initial conditions are adjusted, and TW-ALL3adW (TQ-ALL3adQ) when both the cycle initial conditions and τ (Q_T) are adjusted (cf Table 1).

Reference	Configuration	Type of experiment	Forcing lagged in time	Observations of SSH,SST and in situ T,S assimilated	Forcing adjusted by data assimilation
FREE10	WC10	realistic	none	none	none
ALL10	WC10	realistic	none	real	none
FREE3	WC3	realistic	none	none	none
ALL3	WC3	realistic	none	real	none
ALL3adW	WC3	realistic	none	real	τ
ALL3adQ	WC3	realistic	none	real	Q_T
ALL3adWQ	WC3	realistic	none	real	τ,Q_T
TW-FREE3	WC3	twin	τ	none	none
TW-ALL3	WC3	twin	τ	extracted from FREE3	none
TW-ALL3adW	WC3	twin	τ	extracted from FREE3	τ
TQ-FREE3	WC3	twin	Q_T	none	none
TQ-ALL3	WC3	twin	Q_T	extracted from FREE3	none
TQ-ALL3adQ	WC3	twin	Q_T	extracted from FREE3	Q_T

Table 1

Main experiments conducted in the CCS.

The twin experiments reveal that data assimilation (in TW-ALL3, TW-ALL3adW, TQ-ALL3 and TQ-ALL3adQ) can decrease the error to the assimilated observations, and also generally reduce the errors in the entire 3D circulation. This means that a great part of the error from surface forcing can be corrected through the cycle initial conditions. However, it appears that when compared with TW-FREE3, TW-ALL3 increases the error in temperature below 100 m, close to the coast, in summer (not shown), where variations in the wind stress most impact the ocean circulation. When compared to TQ-FREE3, TQ-ALL3 often increases the error in T below 100 m depth, far from the coast, when important mean increase or decrease occurs in the perturbation of Q_T. This is because the present ROMS-IS4DVAR CCS configuration excessively increases (decreases) the mixed layer temperature to compensate for a decrease (increase) in Q_T (not shown).

The increase of the error in T below 100 m depth with data assimilation does not occur in TW-ALL3adW and TQ-ALL3adQ. During the analysis and forecast cycles, a decrease in error in observed and unobserved variables occurs when comparing TW-ALL3adW (TQ-ALL3adQ) to TW-FREE3 or TW-ALL3 (TQ-FREE3 or TQ-ALL3). Enlarging the control vector to include surface boundary conditions gives more freedom to the system to minimize the cost function, which naturally yields smaller error. The decrease of error in unobserved variables confirms that the surface forcing adjustments influence in a positive way the entire ocean circulation.

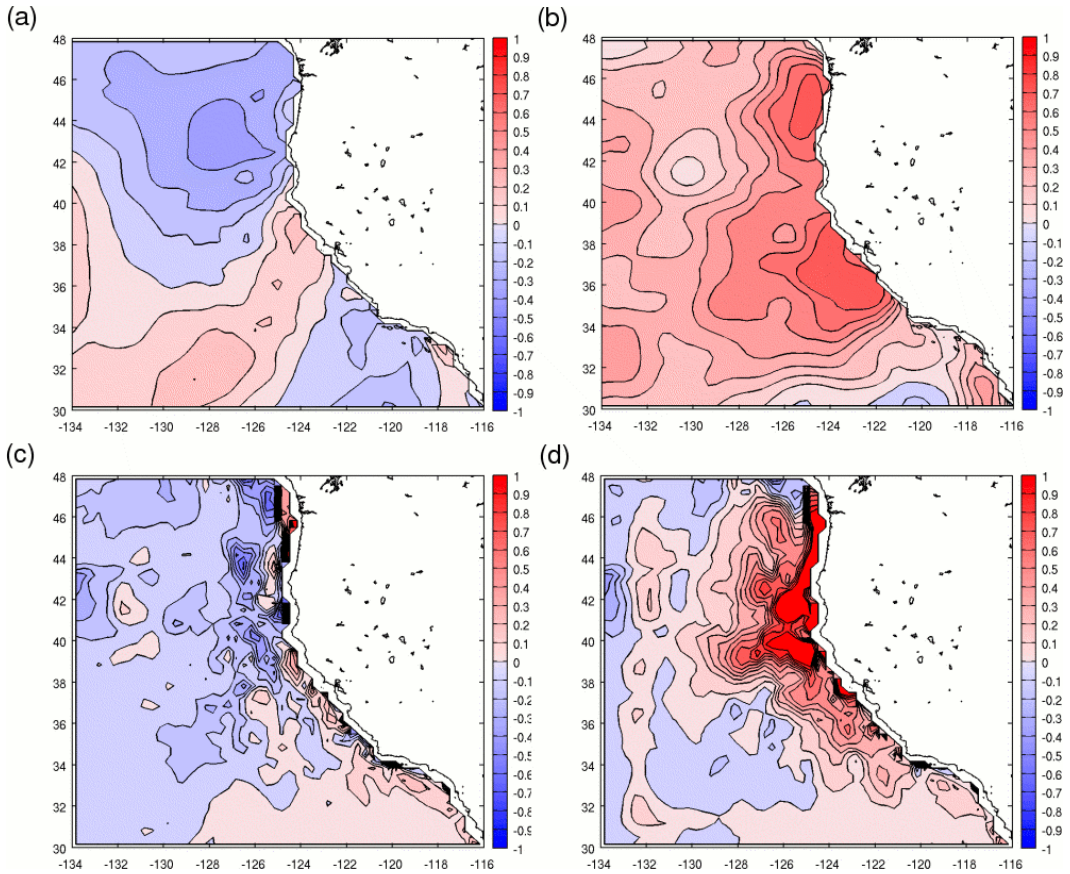


Figure 5

Statistics of the error for the wind stress in TW-ALL3 (a,c) and TW-ALL3adW (b,d) during summer 2002: (a,b) temporal correlation with true data for zonal wind, (c,d) error in the temporal variability of the wind stress curl in 10^{-8} Nm^{-3}.

On seasonal time-scales, the error introduced by the lag in τ decorrelates the true and perturbed data. Both seasonal mean and variability of τ in COAMPS data are similar from year to year. Figure 5a illustrates the low correlations between τ_v data used in TW-FREE3 and TW-ALL3, and the true τ_v used in FREE3, during summer 2002. Figure 5b characterizes the ability of the adjustment of τ in TW-ALL3adW to increase this correlation, especially close to the coast where upwelling occurs and the response time of the ocean to change in wind is short. Errors in both components of τ are thus reduced in TW-ALL3adW. However, abnormal structures may appear in the curl and divergence of τ, mainly because corrections to τ_u and τ_v are decorrelated in **B**. Those structures are mainly local and short-lived too, because there is no temporal correlation in **B** to constraint the daily forcing adjustments. This increases the error in the variability of the curl and divergence of the wind, as illustrated by Figures 5c and 5d. In the present ROMS-IS4DVAR system, the extent to which ocean data can be used to correct surface forcing is limited by this aspect and requires further work.

The adjustment of Q_T strongly decreases all components of the error introduced in this flux. Even on seasonal time scales, there may be important differences from year to year. In TQ-ALL3adQ biases and random errors in heat flux variability are very well corrected when compared to TQ-FREE3, and the correlation between adjusted and true Q_T significantly increases (not shown). The use of daily SST observations facilitates the appropriate correction of the heat fluxes.

Realistic experiments with forcing adjustment

Realistic experiments, denoted ALL3-adW, ALL3-adQ and ALL3-adWQ (cf Table 1) were also conducted with the adjustment of cycle initial conditions, τ and/or Q_T and Q_W from the assimilation of Aviso SSH, COAMPS SST, CalCOFI and GLOBEC T and S data in WC13. As during the twin experiments, the difference between the model and assimilated observations decreases during ALL3-adW, ALL3-adQ and ALL3-adWQ (the latter giving the best results) after analysis when compared to ALL3 and FREE3, as illustrated for the SST in Figure 6a. However, during the forecast cycle, the errors in ALL3-adWQ and ALL3 are similar.

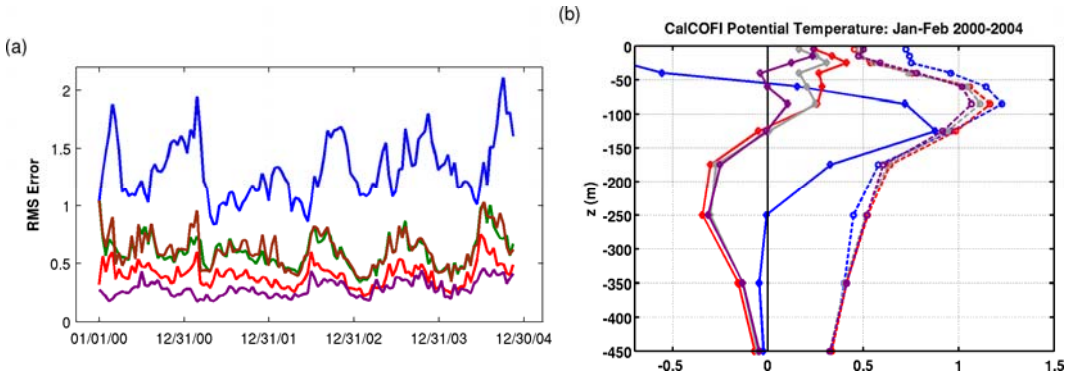


Figure 6

a) RMS Error in °C for each assimilation cycle in WC3 relative to all SST-COAMPS data available during each cycle. b) Error bias (continuous lines) and Error STD (dashed lines) in temperature (in °C) computed from CalCOFI measurements in January-February during the period 2000-2004 as a function of depth. Blue: FREE3; green: forecast in ALL3; red: analysis in ALL3; grey: analysis in ALL3-adW; brown: forecast in ALL3-adWQ; purple: analysis in ALL3-adWQ.

The problem of the weakening of the thermocline gradient by data assimilation, especially strong in winter, is still present when adjusting the surface forcing. However the effect is smaller when adjusting the surface forcing, a part of the correction applied close to the thermocline during ALL3 being converted into corrections on surface forcing during ALL3-adW, ALL3-adQ and ALL3-adWQ. Figure 6b shows that the bias and Error STD relative to the CalCOFI winter T data are smaller when using adjustment of either τ or Q_T and Q_W in addition to adjustment of the initial condition. The most sensitive part of this decrease is on the bias on T below 100 m, which ALL3 usually increases by weakening the thermocline.

Similar to the twin experiments, corrections in τ may generate abnormal structures of curl and divergence in the wind, that varies especially over the continental shelf (not shown). The main trends in the surface forcing corrections during ALL3-adWQ are illustrated on Figure 7. In spring and summer, data assimilation tends to weaken the alongshore north and north-west winds, considerably reducing the strength of coastal upwelling. During winter, mean corrections on the wind are weak. Because the confidence on the COAMPS wind data yields uncertainties smaller than those mean corrections, and because the upwelling cross-shore transport (not shown) in ALL3-adW is more consistent with that of FREE10 and ALL10 than with that of FREE3, it is supposed that those mean corrections account for the lack of horizontal resolution and the poor representation of capes in WC3 more than for an error in COAMPS data. The heat fluxes are mainly increased during fall-winter, while in spring-summer, this increase only occurs in northern part of the CCS, and the heat fluxes are reduced in the south east portion of the CCS.

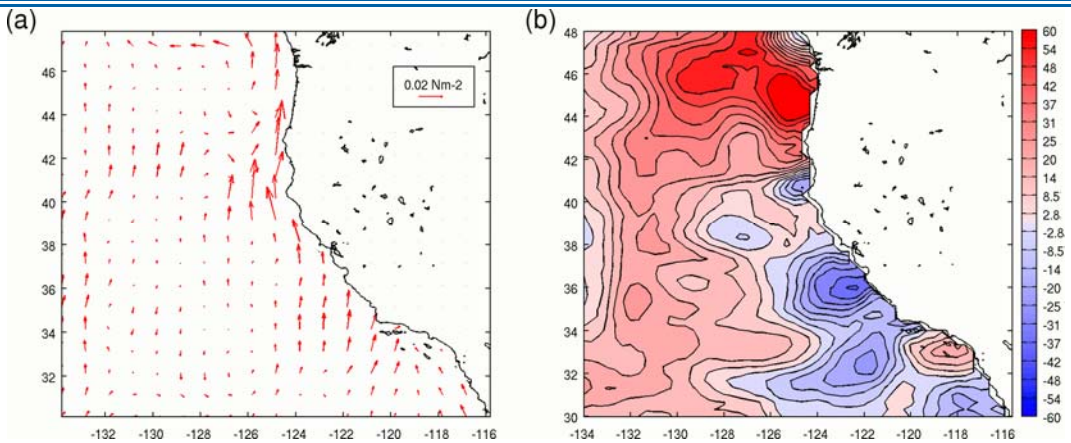


Figure 7

Summer mean adjustment on wind stress (a) and surface heat flux in Wm^{-2} (b) during ALL3-adWQ.

Conclusion

The applications of ROMS-IS4DVAR to adjust both the initial condition and the surface forcing indicate that data assimilation produces a sensible inversion of the information from assimilated data to the ocean circulation and atmospheric forcing. During twin experiments, adjustment of initial condition and surface forcing in ROMS-IS4DVAR reduces the error on non-observed variables of the ocean state and surface forcing. However, imperfect definition and description of error sources, and, thus, imperfect parameter settings in ROMS-IS4DVAR can lead to the system generating errors in the model or in the forcing. This point is particularly evident in the twin experiments presented here where surface forcing is the only source of error; when data assimilation is used to adjust only initial conditions (assuming an absence of error in surface forcing) it can produce errors on temperature at depth. The adjustment of surface forcing increases errors in the curl and divergence of the wind stress because correlations between the errors on the wind stress components at different time or between the zonal and meridional component are not represented in **B**. To improve adjustments to surface forcing, correlations in time and between zonal and meridional wind components may be useful in **B**. Options now developed in ROMS-4DVAR methods for multivariate correlations in **B** (based upon empirical T-S relationship, geostrophy, etc) only concern correlations between variables in the initial condition. Experiments on WC3 also lack realism and the model error in this configuration has to be interpreted by the data assimilation system as an error in the initial condition or surface forcing. Experiments with surface forcing adjustment in WC10, which is more realistic, should provide understanding about the influence of surface forcing, the quality of the data used and of the error that is generated with the methods used to derive ocean surface forcing from these data. A better representation of the different error sources with the use of a larger control vector and application of weak constraint data assimilation (using ROMS-4DPSAS and ROMS-R4DVAR) is also expected to improve results.

Acknowledgements

This work is supported by the National Oceanographic Partnership Program (NOPP) project

NA05NOS4731242. We would like to thank James D. Doyle (NRL), David Foley (NOAA), Stephen Ralston (NOAA) and Patrick Heimbach (MIT) for useful scientific discussions and for providing data used in those studies. The ECCO-GODAE data was provided by the ECCO Consortium for Estimating the Circulation and Climate of the Ocean funded by NOPP.

References

Bograd, S.J., and Lynn, R.J. 2003: Long-term variability in the Southern California Current System. Deep-Sea Research, 50, 2355-2370, doi:10.1016/S0967-0645(03)00131-0.

Broquet, G., Edwards, C.A., Moore, A.M., Powell, B.S., Veneziani, M., and Doyle J.D. 2009: Application of 4D-Variational data assimilation to the California Current System. Dyn. Atmos. Oceans, in press, doi:10.1016/j.dynatmoce.2009.03.001.

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.

Derber, J.D., and Bouttier, F. 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus A 51 (2), 195-221 doi:10.1034/j1600-0870.1999.t01-2-00003.x.

Doyle, J.D., Jiang, Q., Chao, Y., Farrara, J. 2009: High-resolution atmospheric modeling over the Monterey Bay during AOSN II. Deep Sea Research II doi10.1016/j.dsr2.2008.08.009.

Haidvogel, D.B., Arango, H., Budgell, W.P., Cornuelle, B.D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W.R., Hermann, A.J., Lanerolle, L., Levin, J., McWilliams, J.C., Miller, A.J., Moore, A.M., Powell, T.M., Shchepetkin, A.F., Sherwood, C.R., Signell, R.P., Warner, J.C., and Wilkin, J. 2008: Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. Journal of Computational Physics 227, 3595-3624.

Moore, A. M., Arango, H.G., Di Lorenzo, E., Cornuelle, B.D., Miller, A.J., and Neilson, D.J. 2004: A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. Ocean Modelling. 7 227-258.

Moore, A.M., Arango, H.G., Broquet, G., Powell, B.S., Zavala-Garay, J., and Weaver, A.T. 2009: The Regional Ocean Modeling System (ROMS) 4-Dimensional Variational Data Assimilation Systems. In preparation.

Powell, B.S., Arango, H.G., Moore, A.M., Di Lorenzo, E., Milliff, R.F., and Foley, D. 2008: 4DVAR Data Assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). Ocean Modelling 25, 173–188.

Shchepetkin, A.F., and McWilliams, J.C. 2005: The Regional Oceanic Modeling System: A split-explicit, free-surface, topography-following-coordinate ocean model. Ocean Modeling. 9, 347-404.

Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heinbach, P., Marotzke, J., Adcraft, A., Hill, C.N. and Marshall, J. 2002: Global ocean circulation during 1992-1997 estimation from ocean observations and a general circulation model, J. Geophys. Res. 107, C9, 3118, doi:10.1029/2001JC000888.

Tshimanga, J., Gratton, S., Weaver, A.T., and Sartenaer, A. 2008: Limited-memory preconditioners with application to incremental variational data assimilation. Q. J. R. Meteorol. Soc., 134, 751-769.

Veneziani, M., Edwards, C. A., Doyle, J. D., and Foley D. 2009a: A central California coastal ocean modeling study: 1. Forward model and the influence of realistic versus climatological forcing. J. Geophys. Res., 114, C04015, doi:10.1029/2008JC004774.

Veneziani, M., Edwards, C. A., and Moore, A. M. 2009b: A central California coastal ocean modeling study: 2. Adjoint sensitivities to local and remote forcing mechanisms. J. Geophys. Res., 114, C04020, doi:10.1029/2008JC004775.

Vossepoel, F.C., Weaver, A.T., Vialard, J., and Delecluse, P. 2004: Adjustment of Near-Equatorial Wind Stress with Four-Dimensional Variational Data Assimilation in a Model of the Pacific Ocean. Monthly Weather Review 132(8): 2070.

Weaver, A.T., Deltel, C., Machu, E., Ricci, S., and Daget, N. 2005: A multivariate balance operator for variational ocean data assimilation. Q. J. R. Meteorol. Soc., 131, 3605-3625.

Weaver, A.T., Vialard, J., and Anderson, D.L.T. 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. Monthly Weather Review. 131, 1360-1378.

A Data Assimilation Scheme for Oceanic Reanalyses: the Seek Smoother

By Emmanuel Cosme, Jean-Michel Brankart, Pierre Brasseur and Jacques Verron

LEGI, Grenoble, France

Abstract

This paper provides a short description of a reduced rank square root smoother well designed for reanalyses of the ocean circulation. The smoother algorithm naturally comes as a complement to the associated filter, resulting in a very limited extra numerical cost. It is implemented with the Singular Evolutive Extended Kalman (SEEK) filter and the Nucleus for European Modelling of the Ocean (NEMO) model in an idealized configuration. Although the *model* configuration is idealized, the *assimilation* configuration is not: an error model is introduced, by playing on the model resolution in the experimental design. A realistic observation density is also used. We show the benefits of the smoother in comparison with the filter. The smoother is of particular value for observation steps at which observations are sparse. Finally, we discuss two important difficulties which, we believe, will be met when the smoother is implemented in a present-day operational system: the propagation of state error statistics and the parameterization of model errors.

Introduction

Reanalyses of the ocean circulation are needed to examine the variability and tendencies of climate. By essence, reanalyses are constructed retrospectively, over a continuous, limited period in the past. Their construction involves an ocean circulation model, observations of the real ocean, and a data assimilation tool to combine both sources of information and get the most accurate representation of the circulation. The Kalman filter and its approximation, optimal interpolation, are of common use for such exercise. The Global Ocean ReanalYsis and Simulation (GLORYS) project, lead in MERCATOR-Océan with the SAM-2 (*Système d'Assimilation Mercator*, Brasseur et al, 2005), is such an example. And yet, estimation theory says that the Kalman filter, and *a fortiori* optimal interpolation, is not the optimal tool for this purpose (Cohn et al, 1994).

The Kalman filter, under some assumptions on the dynamical system's properties, solves the problem of estimating a state, and associated error covariances, given past and present observations. If subsequent observations are available, they are not used by the filter to estimate the present state (they will be used later to estimate states to come). In the retrospective construction of a reanalysis, subsequent observations are obviously available. In place of the Kalman filter, a smoother is designed to make use of subsequent observations in the estimation process (e.g. Gelb, 1974; Simon, 2006).

This paper synthesizes the recent development of a smoother algorithm based on the Singular Evolutive Extended Kalman (SEEK, Pham et al, 1998) filter, and its implementation with an idealized, high resolution configuration of NEMO model. Rather than describing the algorithm in detail (see Cosme et al, 2009), we emphasize the gain due to the smoother with respect to the filter, and discuss two important issues concerning the forthcoming implementation in an operational system.

The SEEK smoother

The SEEK filter

The SEEK filter (Pham et al, 1998; Brasseur and Verron, 2006) is a square root Kalman filter, the formulation of which is particularly adapted to high dimensional systems as ocean circulation models. As a square root filter, the state error statistics are described with an ensemble of error modes, from which the covariance matrix of the standard Kalman filter is straightforward to compute (leaving aside the dimension problem here). The reduced rank strategy consists in limiting the number of these modes to only a few; generally less than 100 in ocean circulation problems. This drastic sampling enables the dynamical propagation of the error modes with the model. Let us denote S^f_k the matrix of which columns are formed by the forecast error modes at a time index k; the state correction due to the filter analysis is written:

$$x^a_k - x^f_k = S^f_k z_k \qquad (1)$$

where x^a_k and x^f_k denotes the analysis and forecast states at time index k. z_k is a vector, easily obtained from the Kalman filter equations. The filter correction of the forecast state simply appears as a linear combination of the forecast error modes. The forecast error modes themselves are updated at the filter analysis step, to provide the filter analysis error modes:

$$S^a_k = S^f_k (I+\Gamma_k)^{-1/2} \qquad (2)$$

where Γ_k is a low dimensional matrix computed from the forecast error modes and the observation error covariance matrix. We refer to Brasseur and Verron (2006) for more detail.

The SEEK smoother

Let us suppose we have an estimate of a state at a time index $i<k$, that contains the past observational information till time index $k-1$ (in the case $i=k-1$, it is simply the filter analysis at $k-1$). We note this state vector $x^a_{i|k-1}$ and $S^a_{i|k-1}$ the (also known) associated matrix of error modes. It can be shown (Cosme et al, 2009) that this estimate is corrected using the observations at time index k, with the following formulas:

$$x^a_{i|k}-x^a_{i|k-1}=S^a_{i|k-1}z_k \qquad (3)$$

$$S^a_{i|k}=S^a_{i|k-1}(I+\Gamma_k)^{-1/2} \qquad (4)$$

These formulas involve the same vector z_k and matrix Γ_k as the filter. This results in a almost cost-free retrospective estimation of the past state with the present observations.

The model error issue

Dealing with the model error in a reduced rank square root filter is a recurrent and not fully resolved problem. A widespread approach, especially because it is a good background for adaptive schemes, consists in inflating the forecast error modes by multiplying them by a factor larger than 1. This is sometimes called covariance inflation, or forgetting factor strategy. Let α be the inflation factor that multiplies the forecast error modes after their dynamical propagation. A short calculation shows that Equation (4) must be modified as:

$$S^a_{i|k}=\alpha^{-1}S^a_{i|k-1}(I+\Gamma_k)^{-1/2} \qquad (5)$$

This is straightforward to implement, but one must be aware that $S^a_{i|k}$ does not hold anymore the error modes associated to $x^a_{i|k}$ strictly speaking (see Cosme et al, 2009, for detail).

Numerical experiment

Model and observations

The NEMO system (Madec, 2008) is used to simulate a double gyre circulation confined between closed boundaries at 25 and 45°N and over 30° in longitude. The domain is discretized with a 120 (x) by 94 (y) by 11 (z) grid of the Mercator projection type. The resulting resolution is close to 1/4°. The circulation is forced by a zonal wind with a sinusoidal variation with latitude. The lateral boundaries are frictionless but the circulation is subject to a linear friction at the bottom (C_D=2.65 10^{-4} s^{-1}) and a biharmonic lateral dissipation (A_h=-8 10^{10} m^4 s^{-1}). The time integration is performed with a step of 15 minutes and a leap frog scheme. After a 50-year spin up, 10 years of simulation are used to initialize the assimilation experiments: the mean is used for the initial state, the first 20 EOF modes are extracted to form the initial state error covariance.

We perform twin experiments, what means that observations are synthetic, drawn from a reference model simulation. This approach enables a perfect monitoring of errors, and then offers an appropriate framework for methodological developments. The assimilation problem can be gradually complexified by considering observations less and less connected to the dynamical model used for the experiment. For our problem, the reference simulation is run using our double gyre model at a 1/6° resolution (instead of 1/4° as in the nominal configuration), with friction and dissipation coefficients adjusted according to the resolution change. The output fields of Sea Surface Height (SSH) and temperature are noised to simulate observation errors. Then, observations are extracted according to networks of realistic density: SSH is observed similarly to TOPEX/POSEIDON, temperature is observed on a regular grid that roughly mimics the ARGO network density. Figure 1 depicts the SSH tracks in a 10 day interval and temperature profile location in an 18 day interval. These pseudo observations are assimilated in the 1/4° model every 2 days over a 360 day experiment. An inflation factor α=1.054 is used.

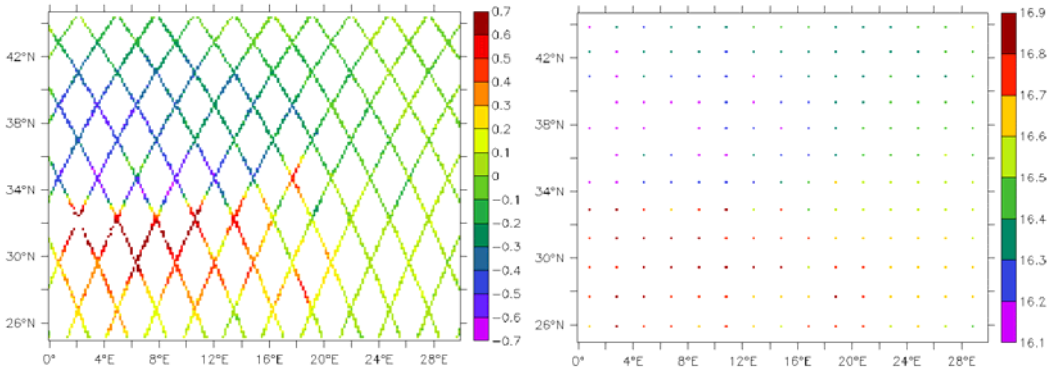


Figure 1

Observation networks used in the assimilation experiment. left: ten days of along-track altimetric observations; right: 18 days of vertical profiles of temperature. One profile every 2° and every 18 days provides a network density (1/36 obs/°/day) coarsely similar to the ARGO network density of one profile every 3° and every 10 days (1/30 obs/°/day).

Results

At any time of the experiment interval, the filter analysis is expected to reduce the error in the forecast state. By 'error', we mean here the departure of the estimate from 'reality' (the 1/6° simulation). We measure this error by the root mean square of this departure, and talk about Root Mean Square Error (RMSE). The filter analysis must then reduce the forecast RMSE. When subsequent observations are used to further correct this analysis estimate through the smoother (Equation 3), the RMSE is still expected to reduce. Figure 2 illustrates this error reduction for 7 days in the first 6 months of the experiment. It can be noticed that (i) the impact of the smoother is significant in comparison to the filter's (in some cases, most of the correction is due to the smoother); and (ii) subsequent observations beyond 8 days have no more impact on the state estimates. We say that 8 days is the best *smoother lag* for the present system. The RMSE reduction is particularly important in the first days, of which the state estimates have not (or to a limited extent) inherited of observational information from the past (before Day 0).

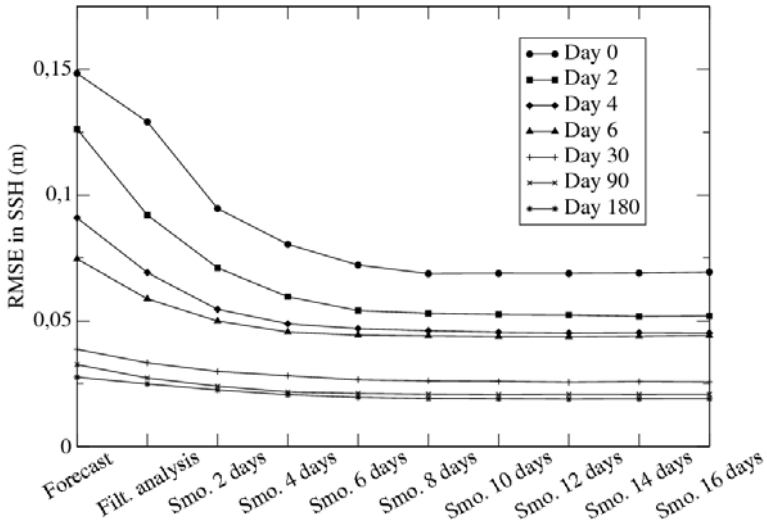


Figure 2

Root Mean Square Error in SSH (meters) for 7 days in the experiment. The RMSE is plotted as a function of the filter/smoother analysis step.

The smoother is most effective with non stationary observation networks. This is illustrated by Figure 3, depicting the fields of error in SSH at initial time, before the filter analysis, after the filter analysis (performed with the observations at the same time), and after four retrospective smoother analyses involving the observations of days 2 to 8. The error reduction due to the filter alone is rather poor, because the SSH observations used for the filter analysis (shown on the central graph) are along tracks that do not cross all the regions where errors are large. These tracks gather two days of satellite passes, consistently with the assimilation cycle. After 4 assimilation cycles, 10 days of SSH observations (8 in addition to the observations at the filter analysis time), that is, the full network shown on Figure 1 (left), has been used. The error in SSH is efficiently reduced (Figure 3, right).

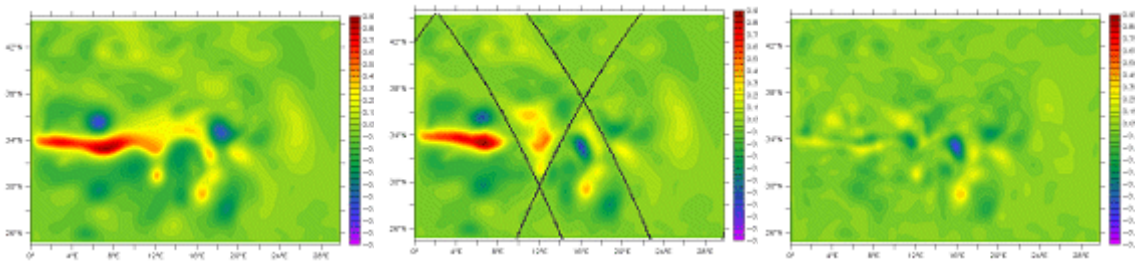


Figure 3

Error in SSH at the first day of the experiment, in meters. Left: initial state (considered as a forecast); center: filter analysis; right: 8-day smoother analysis. SSH observations available for the filter analysis are distributed along the tracks displayed on the central graph.

Towards operational implementation

Optimal interpolation: an obstacle to optimal smoothing?

Operational data assimilation systems based on the Kalman filter are often run with prescribed error modes, i.e. use optimal interpolation (e.g. SAM-2, Brasseur et al, 2005). The implementation of a smoother algorithm based on an optimal interpolation scheme is theoretically nonsensical, and in practice would probably have a poor effect, or even deteriorate the previous states estimates.

It is nonsensical because the smoother retrospective corrections are based on the existence of significant covariances between the state errors at the time of correction (indexed with *i* previously) and the time of observation (*k*). However, the stationarity of the error modes, in the optimal interpolation approach, is valid for long assimilation cycles, the length of which is typically the decorrelation time scale of the dynamical model. Thus, two consecutive states (an analyzed one and its subsequent forecast) should have uncorrelated errors.

Running a smoother with optimal interpolation may also deteriorate the previous states estimation. When the Kalman filter propagates the error modes with the nonlinear model, it tends to progressively attenuate the correlations with errors of the past estimates. The smoother analyses reduce the smoother errors $(S^{a}_{i|k})$ themselves. Both effects result in the convergence to zero of the smoother corrections, when *k* runs away from *i*. This can be seen on Figure 2: beyond 8 days, no correction occurs. Prescribed error modes, depending on how they have been constructed, may reflect significant covariances between errors at different times and generate unjustified corrections by the smoother. Note that some recipes may be used, such as the Incremental Analysis Updating scheme (Bloom et al, 1996), to distribute a filter analysis correction all along the previous assimilation cycle. They work in the spirit of a smoother, but do not take into account the decorrelation errors throughout the cycle and obviously result in suboptimal states estimation.

Impact of the model error parameterization

It is well known that in the Kalman filter, the model error must be appropriately parameterized to avoid filter's divergence. In the smoother approach, the model error contributes to decorrelate the errors that affect the states at time indices *i* and *k* when *k* increases, leading to the fading of smoother corrections at long time distances. This process may not occur correctly if the model error parameterization is erroneous. In the most unfavorable case, smoother corrections can even deteriorate the state estimates. This is illustrated in Figure 4. This is the counterpart of Figure 2, for an experiment similar to the one presented previously, except that no inflation factor is used. That means that the model error term is neglected. The smoother retrospective analyses are extended to 16 assimilation cycles (32 days). In this experiment, corrections due to observations beyond 10 ten days increase the RMSE in SSH.

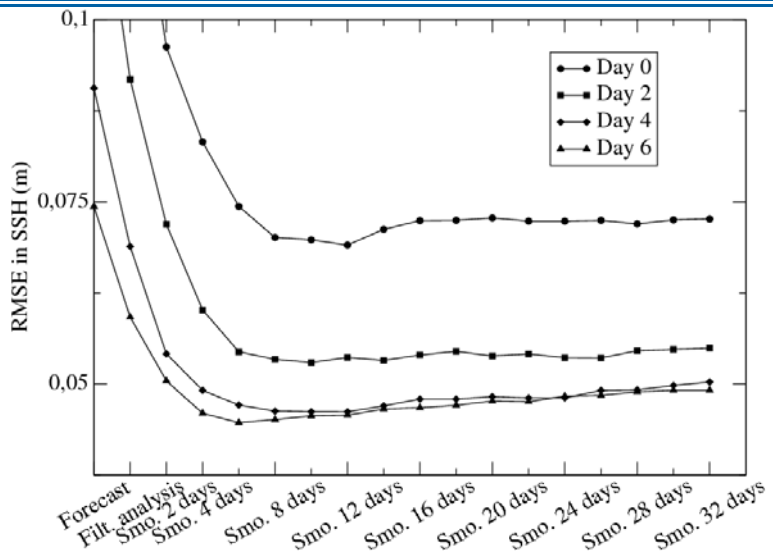


Figure 4

Same as Figure 2 in the perfect model experiment. Only days 0 to 6 are shown, and the smoother works over 32 days.

Conclusion

The SEEK smoother algorithm has been developed and implemented with the NEMO system. It is particularly designed to make reanalyses of the ocean circulation, for it is not constrained to work over a short time interval. Moreover, the smoother is not numerically more expensive than the filter. It produces significant improvements in regard to the filter, especially when the space and time distribution of the observations is not regular.

We emphasize two aspects that may constitute serious obstacles to the operational implementation of the smoother. These aspects will be the topics of specific research in the near future.

Acknowledgements

This work is supported by CNRS/INSU through the LEFE/ASSIM program and by the ANR program. Calculations are performed at the IDRIS computation facility center. The team is supported by the French National Space Center, CNES.

References

Bloom, S.C., Takacs, L., Da Silva, A.M. and Ledvina, D. 1996: Data assimilation using incremental analysis updates. Mon. Wea. Rev. 124, 1256-1271.

Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.-M., Ferry, N., Losa, S., Remy, E., Schröter, J., Skachko, S., Testut, C.-E., Tranchant, B., van Leeuwen, P.J., and Verron, J. (2005). Data assimilation for marine monitoring and prediction : the MERCATOR operational assimilation systems and the MERSEA developments. Quarterly Journal of the Royal Meteorological Society, 131, 3561-3582.

Brasseur, P., and Verron, J. 2006: The SEEK filter method for data assimilation in oceanography: a synthesis, Ocean Dynamics, DOI:10.1007/s10236- 006-0080-3.

Cohn, S.E., Sivakumaran, N.S. and Todling, R. 1994: A fixed-lag Kalman smoother for retrospective data assimilation, Mon. Wea. Rev., 122 , 2838-2867.

Cosme, E.,Brankart, J.M., Verron, J., Brasseur, P. and Krysta, M. 2009: Implementation of a Reduced-rank, square-root smoother for high resolution ocean data assimilation, submitted to Ocean Modelling.

Gelb, A. 1974: Applied optimal estimation, M.I.T. press, 374 pp.

Madec, G. 2008: Nemo reference manual, ocean dynamics component : NEMO-OPA. preliminary version, Tech. rep., Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No 1288-1619, 2008, 91 pp, available at: http://www.locean-ipsl.upmc.fr/NEMO/general/manual/index.html.

Pham, D. T., Verron, J. and Roubaud, M.C. 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Marine. Sys., 16 , 323-340.

Simon, D. 2006: Optimal state estimation, Wiley & sons, 530 pp.

Is there a simple way of controlling the Forcing Function of the Ocean?

By Jean-Michel Brankart, Bernard Barnier, David Béal, Pierre Brasseur, Laurent Brodeau, Grégoire Broquet, Frédéric Castruccio, Emmanuel Cosme, Claire Lauvernet, Pierre Mathiot, Marion Meinvielle, Jean-Marc Molines, Yann Ourmières, Thierry Penduff, Sergey Skachko, Chafih Skandrani, Clément Ubelmann and Jacques Verron

LEGI-MEOM, Grenoble, France

Abstract

The imperfect accuracy of the ocean forcing function has been an important issue in several PhD thesis or postdoctoral researches that were conducted by the MEOM group in the recent years. The purpose of this paper is to propose a synthesis of the respective contributions of these studies to the problem of controlling atmospheric forcing parameters using oceanic observations. Particular emphasis is given to (i) idealized experiments in which MERCATOR reanalysis data are used to estimate parameter corrections, and (ii) ensemble simulations investigating the non-Gaussian character of forcing errors.

Introduction

During the last few years, the MEOM group conducted a series of modelling or assimilation studies, which were most often not directly connected with the problem of controlling the forcing function of the ocean. It happened however that, in many of these studies, interesting questions were raised about the specification or the influence of the ocean atmospheric forcing. By considering these studies together, we can thus get a useful integrated view of the subject and propose synthetic lessons about the control problem, which it is the purpose of this paper to summarize.

The forcing parameters that were considered in this set of studies include wind speed, air temperature or humidity, precipitations, turbulent exchange coefficients,... In view of the decisive influence that these parameters can have on key aspects of the model simulations, there is undoubtedly an important benefit to expect if the available ocean observations could be used to help improving their accuracy. Operational forecasts in particular would directly benefit from a substantial upgrade of the forcing parameters. However, before a solution to this inverse problem can be obtained, it is necessary to examine several important questions, which may be formulated as follows: “What kind of error can be expected for the forcing parameters? How do they propagate to the state of the system and to observed quantities? Can this dependence be inverted to identify relevant parameter corrections?”

In order to propose answers, the first step (section 1) is to describe our individual studies one by one and see what specific light they shed on the above three questions. In that section, we mainly focus on analysing the sensitivity of model solutions to forcing errors. In the second step (section 2), we turn to examining in more detail the outcome of experiments in which the control of atmospheric parameters is effectively attempted, with a special emphasis on the approximations that were introduced. In the third step (section 3), we further describe the studies exploring possible methods to obviate difficulties arising from the non-Gaussian behaviour of forcing errors. And finally, the conclusion sums up the outcome of the various studies and propose an answer to the general question that makes the title of this paper.

To what extent does model accuracy depend on forcing errors?

As a first approach to the problem, we start by giving a list of the individual studies that we have performed, describing for each of them its original purpose or motivation and explaining how it is connected to the subject of this paper.

Deterministic studies

Brodeau (2007) studied the problem of designing an improved set of atmospheric surface variables and fluxes to force the DRAKKAR hierarchy of global ocean/sea-ice general circulation models (The DRAKKAR group, 2007). This work led to the development of several tools dedicated to the elaboration and validation of the atmospheric forcing function and to the construction of the 3rd and 4th version (DFS3 and DFS4) of the DRAKKAR forcing system, in view of carrying out ocean hindcasts and reanalyses for the last five decades (Brodeau et al., 2009). Many long-term sensitivity experiments were performed with the ORCA2 NEMO configuration, and helped understand the oceanic impact of certain key changes in the forcing function. Such explicit model simulations experiments were also found very useful to improve the consistency of the forcing itself, as an improvement to the usual flux calculation from a prescribed sea surface temperature.

Mathiot (2009) studied ways of improving the modelling of key factors (sea ice, ice shelves, air temperature and humidity, surface wind) influencing the formation and modification of dense waters on the Antarctic continental shelf. Using the 1/2° and 1/4°

circumantarctic DRAKKAR configuration of the NEMO model, his sensitivity experiments show that, among the several parameterizations that have been improved, the fine-tuning of the atmospheric forcing (by an appropriate downscaling of the ERA40 reanalysis) has produced the dominant contribution in the improvement of the Antarctic-born water masses (Mathiot et al., 2009a, b), which is in turn a key factor for a correct modelling of the global thermohaline circulation.

Stochastic studies

Castruccio (2006) studied the problem of jointly assimilating altimetric and *in situ* observations in a Tropical Pacific ocean model. His experiments were performed using a reduced order Kalman filter (a version of the SEEK filter) to assimilate these data into the ORCA2 configuration of the NEMO model, and were chiefly intended to explore the role of an accurate geoid (such as that provided by the GRACE mission) for consistently assimilating these two datasets (Castruccio et al., 2006; 2007; 2008). As a subsidiary result, his experiments also indicate that, in the tropical Pacific, wind forcing error is the key factor explaining the difference that is raised between simulated and observed mean dynamic topography. Simulating realistic random wind perturbations is thus found to be one possible solution for producing consistent model error statistics.

Ubelmann (2009) compared the ability of various altimetric observation scenarios for controlling the Tropical Atlantic circulation, as simulated by a 1/4° DRAKKAR configuration of the NEMO model. In his experiments, he makes a clear distinction between the North Brazil current region where most of the error is due to initial conditions, and the region dominated by the signal of the Tropical Instability Waves (TIW), which is mainly driven by the wind forcing (Ubelmann et al., 2009). In the latter case, his observation system simulation experiments were conducted on the basis of ensemble forecasts with random wind perturbations (with appropriate structure). With this system, he was able to simulate realistic TIW error structures, and to assess their controllability by satellite altimetry.

Broquet (2006) performed similar Monte Carlo experiments to characterize the model error that is produced by the atmospheric forcing, in a 1/15° resolution HYCOM configuration of the Bay of Biscay, which he nested in a 1/3° North Atlantic model. The objective was mainly to study the spatial structure of the errors that are induced on the model state with the idea of using them as a parameterization of the error statistics that are required for data assimilation experiments (Broquet et al., 2007). But the study also provides interesting insights into the relation between forcing errors and forecast errors, showing for instance a significant divergence between the central forecast (i.e. with the unperturbed mean forcing) and the mean forecast of the Monte Carlo simulation as a result of the nonlinearity of the model response to a forcing difference.

Estimation problems

Skachko et al. (2007, 2009) studied the problem of estimating the turbulent exchange coefficients that control evaporation, latent and sensible heat fluxes at the air-sea interface using ocean temperature and salinity observations. Twin experiments, carried out with the ORCA2 configuration of the NEMO model, have shown that a joint control of the unknown parameters together with the state of the system (by augmenting the control vector of the reduced order Kalman filter) is best appropriate for eliminating the model biases that are due to parameter errors, thus improving the short term forecast of the mixed layer properties.

Skandrani et al. (2009) extended this study by considering the more realistic problem of using MERCATOR reanalysis data to control an increased number of uncertain atmospheric parameters. More detail about the method and results obtained in this study is given in the next section.

Lauvernet et al. (2009) performed Monte Carlo simulations by introducing random errors on atmospheric parameters driving a one-dimensional model of the ocean mixed layer. The main purpose of these experiments was to assess the behaviour of a new filtering approach based on the truncated Gaussian assumption but they also provided an original information about the statistical relationship between forcing errors and the shape of the mixed layer (as shown in section 3 below).

Béal et al; (2009) also performed Monte Carlo experiments by introducing wind perturbations in a coupled circulation-ecosystem model (NEMO/LOBSTER) of the North Atlantic (using the 1/4° DRAKKAR configuration NATL4). The main purpose of this study was to generate a realistic type of ecosystem errors and investigate ways of controlling them using ocean colour observations. But, as the previous study, this also gives a consistent statistical view on the ocean response to wind errors and thus on the level of complexity of the inference problem (see section 3).

This set of studies encompasses a wide range of possible ways of looking at the forcing problem, from a deterministic understanding of forcing corrections, to a stochastic exploration of the forcing influence, and the statistical estimation of forcing parameters. All of them point towards the ubiquitous sensitivity of the ocean behaviour to forcing particularities and to the intricate character of this dependence. In the perspective of the estimation problem, two difficulties stand out in this general picture: (i) The atmospheric forcing exerts its influence over a wide range of timescales, and the forcing parameters are subject to both instantaneous and time-averaged physical constraints. For instance, whereas on the short term, evaporation and precipitation need only to be inside reasonable error bars, on the long term, the ocean cannot possibly receive more rain than it evaporates. Inferring parameter corrections from short term information only may thus produce inconsistent long term behaviour, to the prejudice of the climatic relevance of the DRAKKAR hindcasts or MERCATOR reanalyses. (ii) The influence of the forcing is intimately nonlinear

and produces uncertainty in the ocean simulation that is far from Gaussian. For instance, a model simulation performed with the average (or median, or most probable) forcing function may be very different from the average (or median, or most probable) model response to a representative sample of possible forcings. Hence, as long as forcing uncertainty does not vanish, using the best forcing function does not necessarily produce the best model simulation. The only way out of this apparent paradox is to keep track of the dependence of model results upon forcing uncertainty (e.g. by means of ensemble simulations). This explicit knowledge can then be used to compute best estimates of any nonlinear diagnostic (e.g. by averaging over the ensemble) and to specify an optimal criterion for the inference problem.

Can atmospheric forcing parameters be controlled using sea surface observations?

In the context of the MERSEA project, Skandrani et al. (2009) investigated the problem of controlling atmospheric forcing parameters using idealized experiments in which MERCATOR ocean reanalysis data are used as the reference simulation (i.e. the truth of the problem). Synthetic observation datasets (for sea surface temperature and sea surface salinity) are extracted from the reanalysis to be assimilated into a low resolution global ocean model (ORCA2). With respect to Skachko et al. (2009) who investigated the same problem using twin assimilation experiments, this new study is thus more realistic, since the difference between model and reanalysis is now very similar in nature to the real error. However, solving this more realistic problem required a better specification of the prior information about the parameters and their associated uncertainty. It turned out indeed that making appropriate assumptions on that respect is increasingly important as the estimation problem is becoming more realistic, because it is more and more difficult to make the distinction between forcing errors and the other potential sources of error in the system. An additional important objective is thus to find means of identifying properly the part of the model-observation misfit that can be interpreted as resulting from inaccurate atmospheric parameters.

For that purpose, our plan is to apply sequentially a Bayesian inference method to compute piecewise constant optimal parameter corrections. This is done using a Monte Carlo algorithm to simulate the ocean response to parameter uncertainty, and using the resulting ensemble representation of the prior probability distribution to infer optimal parameter corrections from the ocean surface observations. Compared to Skachko et al. (2009), we improved the definition of this prior probability distribution in two ways. First, the error statistics are estimated locally in time for each assimilation cycle, by performing a sequence of ensemble forecasts around the current state of the system (thus no more constant). And second, it is assumed to be a truncated Gaussian distribution (as proposed by Lauvernet et al., 2009, as an improvement to the classical Gaussian hypothesis), in order to avoid the most extreme and non-physical parameter corrections (more than 3 times the prior standard deviation). See Skandrani et al. (2009) for a detailed presentation of the parameterization.

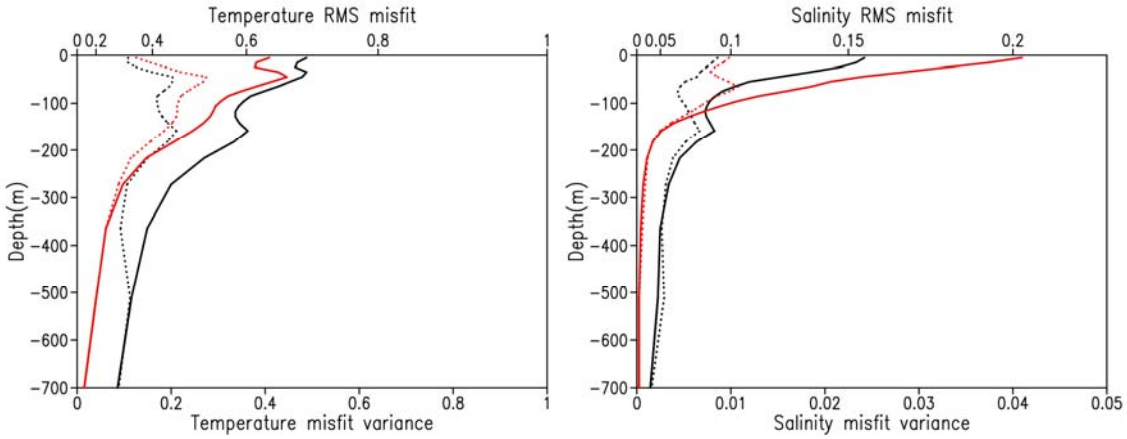


Figure 1

Temperature (left panel, in °C) and salinity (right panel, in psu) root mean square misfit (over a one-year diagnostic period, from June 30, 1993 to June, 29, 1994) with respect to reanalysis data in the North Atlantic (black lines) and North Pacific (red lines) oceans, as obtained with (dotted line) and without (solid line) parameter optimization. (The latter simulation is thus free of any kind of data assimilation.)

Figure 1 shows the resulting temperature and salinity misfit for the two simulations, with and without parameter optimization (for the North Atlantic and North Pacific oceans). This comparison illustrates the beneficial impact of parameter corrections over the whole depth of the mixed layer. The large error patterns close to the surface, corresponding to mixed layer errors that are present in the reference model simulation (without parameter optimization) and that are clearly due to forcing errors, are indeed very substantially

reduced. Overall, the results of the experiments show that it is possible to compute piecewise constant parameter corrections, with predefined amplitude limitations, so that long-term model simulations become much closer to the reanalysis data. The misfit variance is typically divided by 3 over the global ocean. However, the model that is used to perform these experiments is a low resolution model that does not represent correctly important resolution-dependent circulation features (like Western boundary currents). The consequence is that, without parameter optimization, part of this model error is still incorrectly ascribed by the scheme to the parameters and the prescribed amplitude limitations always saturate in these regions, which indicates that the parameter corrections are systematically unrealistic. Such problems can only be circumvented either by improving the model or by correcting these errors by data assimilation.

Yet, despite the deficiencies just mentioned, the experiments reported in Skandrani et al. (2009) already represent a significant step in the right direction: by constructing the prior distributions locally in time, and by imposing strict limitations to the amplitude of the correction, we can be sure at least that the parameter estimates always remain in a realistic range of values (i.e. inside their local range of variation in the input atmospheric data). To go further, a key element is certainly to continue improving the prior parameter probability distribution, which we have just shown to be a key issue in the computation of more realistic parameter estimates.

Can non-Gaussian methods be made efficient enough?

In their study, Lauvernet et al. (2009) performed Monte Carlo experiments with a 1D model of the ocean mixed layer, assuming that the only source of error in the system comes from the atmospheric data (wind and air temperature only), with known probability distribution. From this distribution, they sample an ensemble of atmospheric data and generate the corresponding ensemble of model simulations. To build a Gaussian parameterization of the forecast error distribution, they use a reduced-order representation of the covariance of this 1000-member ensemble. The resulting Gaussian distribution is illustrated by a 30-member sample in Figure 2 (left panel). From this Figure, it is apparent that many profiles of the sample are hydrostatically unstable, thus physically unacceptable. This already shows that the Gaussian assumption is not a correct way of representing the forecast ensemble (in which all profiles are hydrostatically stable). Even in this very simple example, forcing induced errors are thus very far from being well captured by a Gaussian model. Figure 2 (middle panel) shows a sample of the updated Gaussian distribution using an accurate observation of sea surface temperature, which typically leads to unstable density profiles. This is a classical problem with this model if the inequality constraints are not taken into account.

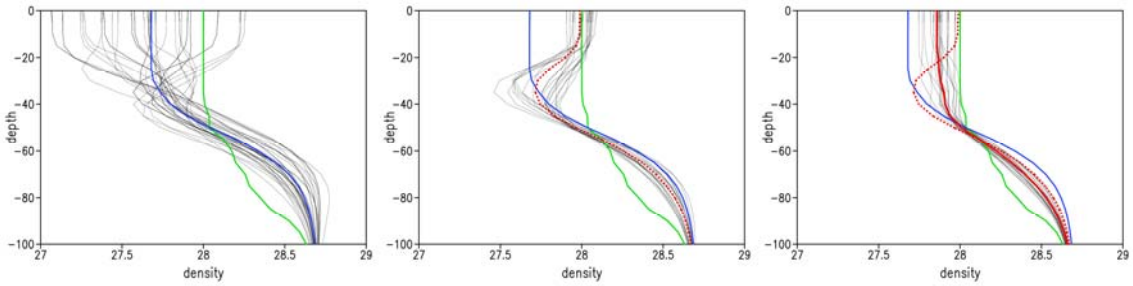


Figure 2

Samples (30 members) of density profiles (in kg/m³) as a function of depth (in m) from the prior Gaussian (left panel), updated Gaussian (middle panel) and updated truncated Gaussian probability distribution (right panel). The Figure also shows the true state (green line), the background (blue line), the mean of the updated Gaussian distribution (dotted red line), and the mean of the updated truncated Gaussian distribution (solid red line).

The originality of this work is then to introduce a truncated Gaussian assumption for the prior probability distribution, setting to zero the prior probability on all regions of the state space violating a set of inequality constraints (including here all unstable profiles). Figure 2 (right panel) shows a sample of the updated truncated Gaussian distribution (this character is preserved by the observational update) using the same observation. Such a sample can be computed using a Gibbs sampler, as explained in Lauvernet et al. (2009), thus providing an updated ensemble with the required hydrostatic stability. More generally, this study shows that it is possible to build an optimal nonlinear filter based on the assumption that error probability distributions are truncated Gaussian distributions. The justification for this assumption is that the resulting filter is numerically much more efficient than more general methods that do not make any assumption about the shape of the probability distributions (such as particle filters). It is especially more efficient because the observational update can be performed using the same formula as the Kalman filter, and because truncated Gaussian distributions can be sampled very efficiently (with a Gibbs sampler).

In a distinct study, Béal et al. (2009) also performed Monte Carlo experiments to study the effect of wind errors in a coupled physical-biogeochemical model of the North Atlantic (set up by Ourmières et al., 2009), with a specific focus on the analysis of the ecosystem response to these errors. Figure 3 (left panel) illustrates for instance the direct impact of wind uncertainties on the simulated mixed layer depth. In general, the stronger the wind, the deeper the mixed layer, even if there can be specific situations in

which the relationship is no more monotonous. Then, as a direct consequence of a deepening of the mixed layer, the sea surface temperature is decreased by mixing with deeper waters (Figure 3, middle panel). This enhanced mixing generally imports nitrates and exports phytoplankton from the surface layer, thus also modifying the ecosystem behaviour (Figure 3, right panel). As a general rule, the response is observed to be rather complex, depending in particular on the local stratification, in such a way that even the general features of the probability distributions can change radically in space and time (e.g. sign and strength of the correlation, shape of the regression lines, asymmetry between positive and negative anomalies, presence of thresholds,...). Nonlinearities in the model lead indeed to many kinds of non-Gaussian behaviours that cannot be properly handled by classical linear estimation methods.

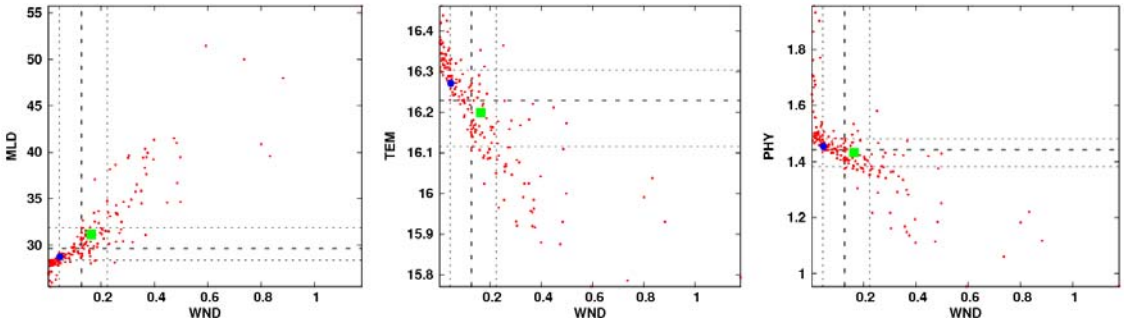


Figure 3

200-member, 1-day ensemble forecast at 47°W, 40°N. The scatterplots connect wind magnitude (in m/s) to mixed layer (left panel, in m), sea surface temperature (middle panel, in °C) and phytoplankton concentration (right panel, in mmol N/m³). The Figure ure also shows the reference simulation without wind perturbation (blue dot), the mean of the ensemble (green square), and the medians (dashed lines) and quartiles (dotted lines) of the marginal sample distributions.

In order to tackle this problem at moderate cost (i.e. in a way that is compatible with large size data assimilation problems), a simplified approximate nonlinear scheme has been introduced in this study. The idea is to perform a nonlinear change of variables (anamorphosis) separately for each variable in the control vector, by remapping the ensemble percentiles of their marginal distribution to Gaussian percentiles. In that way, the additional cost required to make the observational update nonlinear is negligible: the main cost is in the computation of an ensemble forecast that is sufficient in size to identify properly the transformation functions. Test experiments show that this simplified scheme is often sufficient to detect and to exploit the nonlinear relationships between observations and estimated variables, thus restoring the control of the system in situations for which linear estimation fails (because of weak or inexistent linear correlation). In many regions of the North Atlantic, a very substantial reduction of error variance has been obtained for the most important ecosystem variables.

Conclusion

The general lesson that can be drawn from the various studies enumerated in section 1 is that the realism of ocean model simulations closely depend on the quality of the forcing function, and that there is much to be gained from a direct correction of forcing parameters to fit ocean observations. Furthermore, in the quest for better operational forecasts, the endless improvement of the initial conditions is bound to become useless as soon as forcing errors start prevailing. However, a distinctive feature of this inverse problem is that parameter corrections can only rely on indirect observations, so that they closely depend on the parameterization of the statistical dependence between forcing errors and forecast errors. In addition, we have observed in section 3 that this dependence cannot usually be accurately described by a Gaussian model, so that nonlinear estimation cannot be avoided. This nonlinearity of the forcing influence together with the sensitivity of the indirect estimation to statistical approximations make the inverse problem ill-conditioned by nature and a specific care must be taken to avoid unstable behaviours.

In terms of method, our studies largely rely on Monte Carlo experiments for exploring the effect of forcing errors on the model solution. In principle, the ensemble forecast can also provide a sufficient statistical information to infer a correction to the forcing function from the available ocean observations. With this information in hand, it is indeed possible either to follow a simple Gaussian assumption (relying on the mere covariance of the ensemble forecast) or to turn to more sophisticated schemes (as those presented in section 3) if needed. As an alternative, the control of the forcing function of the ocean can also be performed using a variational approach (e.g. Stammer et al., 2004). But the particular algorithm that is used to compute the solution does not change the nature of the problem. And if it is certainly true that a variational method is in some sense more natural to embody the nonlinear deterministic relationship between forcing and forecast errors, it may sometimes be excessively difficult to find the global minimum of a complex cost function or to take a specific action for avoiding non-physical estimates. The most appropriate method thus generally depends on the particular circumstances and, whatever the computational approach, the key criterion in selecting a suitable parameterization

is always to look for an acceptable tradeoff between numerical efficiency and a sufficient generality to deal with the problem at stake.

In our attempt to control a few atmospheric parameters using MERCATOR reanalysis data (as described in section 2), it has been found that such a tradeoff could be achieved using a linear update of the parameter, provided that the error covariances are derived from a 100-member ensemble forecast (performed from the current initial condition) and if the amplitude of the correction is forced inside the most probable region of the prior Gaussian (less than 3 times the standard deviation). More physical bounds on the value of the parameters could also be imposed using a more sophisticated truncated Gaussian assumption (as shown in section 3). However, this introduction of *ad hoc* safeguards to compensate for crude statistical assumptions can only be done at the expense of a redefinition of the inverse problem, which is here reduced to finding a reasonable forcing function (e.g. within predefined error bars) to fit the available observations, with a lesser concern for the intrinsic accuracy of the updated parameters. On the other hand, even if these statistical assumptions may look rather crude in view of the non-Gaussian character of the problem, they are also quite sophisticated since a large size ensemble forecast is systematically performed to obtain a description of the error covariance that is accurate enough to update the forcing parameters. This level of sophistication is already unreachable for large size ocean models, for which parameter estimates could only be obtained by performing the ensemble forecast at much lower resolution. Such a complicated suite of shortcomings and approximations cannot be evaded because a simple method that is, at the same time, general enough to grasp the full complexity of the inverse problem and efficient enough to cope with full-size MERCATOR systems does not exist yet.

Acknowledgements

The authors wish to acknowledge the various kinds of support (including doctoral grants) that they received for the several studies described in this paper: CNES, CNRS and SHOM, the Groupe Mission Mercator-Coriolis, the MERSEA project (funded by the E.U. under contract No. AIP3-CT-2003-502885) and the U.S. Office of Naval Research (under Grant n° N00014-05-1-0731) as part of the NOPP HYCOM project. Calculations were conducted with the support of IDRIS/CNRS.

References

Béal, D., Brasseur, P., Brankart, J.-M., Ourmières, Y. and Verron, J. 2009: Controllability of mixing erros in a coupled plysical biogeochemical model of the North Atlantic: a nonlinear study using ensemble anamorphosis. Ocean Science, submitted.

Brodeau, L., 2007: Contribution à l'amélioration de la fonction de forçage des modèles de circulation générale océanique. PhD thesis, Université Joseph Fourier, Grenoble.

Brodeau, L., Barnier, B., Penduff, T., Treguier, A.M. and Gulev, S. 2009: An ERA-40 based atmospheric forcing for global ocean circulation models. *Ocean Modelling*, in revision.

Broquet, G. 2007: Caractérisation des erreurs de modélisation pour l'assimilation de données dans un modèle océanique régional du Golfe de Gascogne. PhD thesis, Université Jospeh Fourier, Grenoble.

Broquet, G., Brasseur, P., Rozier, D., Brankart, J.-M. and Verron, J. 2008: Estimation of model errors generated by atmospheric forcings for ocean data assimilation: experiments in a regional model of the Bay of Biscay . *Ocean dynamics*, 58(1):1-17.

Castruccio, F. 2006: Apports des données gravimétriques GRACE pour l'assimilation de données altimétriques et in-situ dans un modèle de l'Océan Pacifique Tropical. PhD thesis, Université Joseph Fourier, Grenoble.

Castruccio, F., Verron, J., Gourdeau, L., Brankart, J.-M. and Brasseur, P. 2006: On the role of the GRACE mission in the joint assimilation of altimetry and TAO data in a tropical Pacific Ocean model. *Geophys. Res. Letters*, 33:L14616.

Castruccio, F., Verron, J., Gourdeau, L., Brankart, J.-M. and Brasseur, P. 2007: GRACE: an improved MDT reference for altimetric data assimilation. Mercator Newsletter, 25:20-31.

Castruccio, F., Verron, J., Gourdeau, L., Brankart, J.-M. and Brasseur, P. 2008: Joint altimetric and in-situ data assimilation using the GRACE mean dynamic topography: a 1993-1998 hindcast experiment in the Tropical Pacific Ocean. *Ocean dynamics*, 58(1):43-63.

Lauvernet, C., Brankart, J.M., Castruccio, F., Broquet, G., Brasseur, P. and Verron, J. 2009: A truncated Gaussian filter for data assimilation with inequality constraints: application to the hydrostatic stability condition in ocean models. *Ocean Modelling*, 27:1-17.

Mathiot, P. 2009: Influence du forçage atmosphérique sur la représentation de la glace de mer et des eaux de plateau en Antarctique dans une étude de modélisation numérique. PhD thesis, Université Joseph Fourier, Grenoble.

Mathiot, P., Barnier, B., Gallée, H., Molines, J.-M., Le Sommer, J., Juza, M. and Penduff, T. 2009: Introducing katabatic winds in global ERA40 fields to simulate their impact on the Southern Ocean and sea-ice. *Ocean Modelling , submitted*.

Mathiot, P., Jourdain, N., Barnier, B., Gallée, H., Molines, J.-M., Le Sommer, J. and Penduff, T. 2009: Sensitivity of model of Ross Sea polynyas to different atmospheric forcing sets, Ocean Dynamics, submitted.

Ourmières, Y., Brasseur, P., Levy, M., Brankart, J.M. and Verron, J. 2009: On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. *Journal of Marine Systems*, 75(1-2):100-115.

Skachko, S., Brankart, J.-M., Castruccio, F., Brasseur, P. and Verron, J. 2006: Air-sea fluxes correction by sequential data assimilation. Mercator Newsletter, 22:24-28.

Skachko, S., Brankart, J.-M., Castruccio, F., Brasseur, P. and Verron, J. 2009: Improved turbulent air-sea flux bulk parameters for the control of the ocean mixed layer: a sequential data assimilation approach. *Journal of Atmospheric and Oceanic Technologies*, 26(3):538-555.

Skandrani, C., Brankart, J.-M., Ferry, N., Verron, J., Brasseur, P. and Barnier, B. 2009: Controlling atmospheirc forcing parameters of global ocean models by sequential assimilation of sea surface Mercaror-Ocean reanalysis data, Ocean Science, submitted.

Stammer, D., Ueyoshi, K., Köhl, A., Large, W.G., Josey, S.A. and Wunsch, C. 2004: Estimating air-sea fluxes of heat, freshwater, and momentum through global data assimilation, Journal of Geophysical Research, 109(C5):C05023.

The Drakkar Group 2007: Eddy-permitting Ocean Circulation Hindcasts of past decades. CLIVAR Exchanges 42, 12(3):8-10.

Ubelmann, C. 2009: Etude de scénarios d'altimétrie satellitaire pour le contrôle de la circulation océanique dans l'océan Atlantique tropical par assimilation de données. PhD thesis, Université Joseph Fourier, Grenoble.

Ubelmann, C., Verron, J., Brankart, J.M., Cosme, E. and Brasseur, P. 2009: Impact of upcoming altimetric missions on the control of the three-dimensional circulation in the tropical Atlantic ocean. *Journal of Operational Oceanography*, 2(1):3-14.

Ocean-Atmosphere Flux correction by Ocean Data Assimilation

By Nicolas Ferry and Mahé Perrette

Mercator Ocean, Ramonville St Agne, France

Abstract

Part of the ocean surface forecasting error is due to errors in the atmospheric surface forcing. Only few ocean data assimilation methods propose formulations with background covariances "exiting" from the ocean to go into the atmosphere thereby analyzing ocean-atmosphere fluxes. This study describes and tests a method of sequential ocean data assimilation to produce a correction of ocean-atmosphere fluxes consistent with observations assimilated and therefore with the analyzed ocean state. A new variable, defined as the time integral of the surface flux is introduced in the assimilation control vector. A preliminary study carried out with representers shows that this quantity is the best choice to correlate ocean upper layer temperature anomalies (respectively salinity) anomalies with the ocean-atmosphere heat (respectively freshwater) flux. A realistic case is then considered with an ocean general circulation model assimilating real observations (sea surface temperature and profiles of temperature and salinity). This allows building a 1-year long analysed ocean surface flux time series consistent with the analysed ocean state. These corrections vary geographically and over time (seasonal) with an average value of about 10 $W.m^{-2}$, which corresponds to the magnitude of the uncertainties generally accepted for the heat flux. A one year free run forced with the analysed surface heat flux turns out to be more realistic than the control simulation. The SST biases are strongly reduced as well as the rms of the model innovation (observation minus observation). A possible implementation in an operational framework is finally discussed.

Introduction

The analysis and forecast of the ocean state have made considerable progress over the last decade, particularly by improving the ocean-atmosphere fluxes provided by numerical weather prediction centres but also through progresses in the field of data assimilation and ocean numerical modelling. Generally speaking, the ocean forecast skill is closely related to (i) uncertainties in the ocean initial condition (obtained by data assimilation), (ii) limitations of physical parameterizations used in ocean numerical models and (iii) uncertainties associated to ocean-atmosphere fluxes used as boundary condition (at the surface). This study proposes to improve ocean-atmosphere fluxes using a sequential ocean data assimilation method. The idea is to produce a correction of ocean-atmosphere flux consistent with the assimilated ocean observations and therefore analyzed ocean state.

Few studies have investigated the issue of analysed surface fluxes consistent with an analysed oceanic state. The studies of Bonekamp et al. (2001) and Vossepoel et al. (2004) are a first attempt to propose wind stress corrections consistent with the ocean state. Only Bonekamp et al. (2001) implement their method in a realistic framework. However, the 4D-VAR approach used in these studies gives satisfactory results only in the tropics for long assimilation windows (2 months) and is expensive. Stammer et al. (2004) also propose a 4D-VAR approach and build wind stress, heat and freshwater fluxes corrections consistent with the analyzed global oceanic state. The method is objectionable because of the authors entirely attribute the model forecasting errors to errors in the surface fluxes.

We present here a method based on a reduced rank Kalman filter (SEEK formulation, Pham et al. 1998), focusing on the surface heat flux (the method is also applicable to the freshwater flux). The approach is similar to Skachko et al. (2006) or Skandrani et al. (2009) and Brankart et al. (2009, this issue) and consist to augment the control vector with *ad hoc* atmospheric variables. The choice of the control variables is crucial since it determines the features of the new error covariances and how the surface flux will be corrected by ocean data assimilation. Skachko et al. (2006) assume that most of the surface heat flux error comes from uncertainties in the C_E and C_H bulk coefficients for latent and sensible heat. In our study, the approach is different: we do not want to attribute the heat flux error to a particular parameter of the bulk formulation, but rather to the surface total heat flux itself. From a data assimilation point of view, the inverse mathematical problem we want to solve should be less under determinated than with the other approaches.

The paper is organised as follows. We first present the methodology (Section 2), and we test it in a simple framework with representers (Section 3). Then, the method is tested in an ocean general circulation model assimilating real observations (Section 4). Finally, a possible implementation in a pre-operational framework is discussed (Section 5).

Heat flux correction methodology

The basic idea of the heat flux correction method is that a fraction of the forecast error of the surface layer temperature is related to errors in the ocean-atmosphere heat flux. Surface forcing errors therefore result in mixed layer and / or thermocline

temperature and / or salinity errors. The assimilation of ocean observations aims to minimize the forecast errors, in particular in the surface layers. Through the error covariances between the observed quantities and an appropriate new variable in the control vector, it becomes possible to diagnose the corresponding flux correction. Assuming that δT is the temperature forecast error, we affirm that a fraction δT^Q of this forecast error is due to a heat flux error δQ. Figure 1 illustrates the concept.

Figure 1: Diagram showing the error covariances between the heat flux and the temperature forecast error in the surface layers (see text for details).

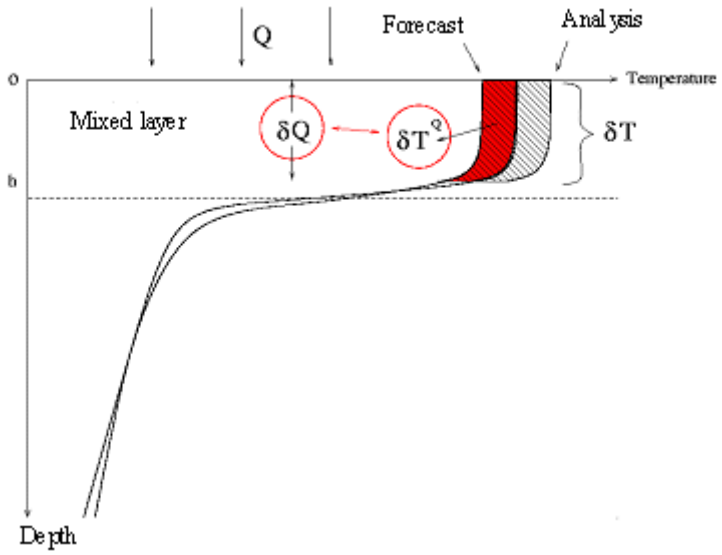


Figure 1

Diagram showing the error covariances between the heat flux and the temperature forecast error in the surface layers (see text for details).

In this study we use Mercator Assimilation System (SAM) version 2 based on a reduced rank Kalman filter (SEEK filter, Pham et al., 1998, Testut et al. 2003, 2005). The assimilation system uses 5 control variables (namely the barotropic height, temperature, salinity, zonal and meridional velocities). The covariance of these 5 variables is modelled by a collection (~ 250) of seasonally variable error modes calculated from a long free OGCM simulation. SAM2 scheme is modified in this study by introducing a new control variable that has the following properties: (i) it depends on the net heat flux and (ii) it is strongly correlated with changes in ocean temperature. A study of the mixed layer heat budget reveals that the time integral of surface heat flux is the quantity that maximizes the correlation with temperature changes observed in the upper ocean which is the part of the ocean that directly interacts with the atmosphere. This leads to augment the control vector with the new variable T^Q defined by:

$$T^Q = \int 1/(\rho\ Cp\ h_o) \times Q\ dt \qquad (1)$$

with Q the net surface heat flux, ρ the sea water density, Cp the seawater specific heat capacity and h_o a mean value of the mixed layer depth. It can be shown that this new control variable maximizes the covariance between changes the mixed layer temperature and the one induced by the surface heat flux. A data assimilation system including this new parameter in the control vector will produce, thanks to the multivariate covariances, an increment δT^Q. It reflects the error of the surface flux in terms of upper layer temperature change. It is then possible to retrieve the heat flux increment δQ^a associated to δT^Q by using the following inversion relationship:

$$\delta Q^a = h_o\ \rho\ Cp \times \delta T^Q / \Delta t \qquad (2)$$

With Δt the length of the assimilation window. In fact one assumes that the temperature change δT^Q is mostly due to a heat flux change δQ^a, acting all over the mixed layer. We can note that the mixed layer depth h_o formally disappears when one combines (1) and (2). The choice of this quantity has therefore little impact on the covariance between heat flux and temperature, mainly because the error covariance is seasonally variable. In other word we think that the SST forecast error is more related to errors in the surface fluxes that to an erroneous mixed layer depth. The results obtained in the following show that this assumption is reasonable.

Preliminary study with representers

The use of representers is an inexpensive and powerful way to study the error covariance structure. In our case, we are interested in the cross representer p(δSST, δ Q^a) where δQ^a is the analyzed surface flux given by Eq. (2). This allows analyzing the covariance between an SST error and the associated heat flux error. In our case, the representer p(x_i, x_j) between the control variables x_i and x_j is the quantity cov(x'_i, x'_j)/var(x'_i), where cov () and var () operators are the covariance et variance of an appropriate series of error modes.

In order to investigate the dependency of the new control variable δQ^a to an observed quantity like SST (which is *a prori* the most correlated ocean parameter with δQ^a), we focus on a region where the surface fluxes are known to be the main forcing for the upper ocean heat budget. We choose the north-eastern Atlantic, off Portugal (36°W-8°W/24°N-55°N), where the advection plays a minor role and where errors in the surface fluxes may be a good candidate to explain errors in the SST. First, we check that SST changes significantly co-vary with the surface heat flux. Figure 2 shows that in that region, an innovation of +1ºC in surface temperature is associated to a surface heat flux error of the order of ~ 70 W/m^2, with a maximum located near the +1°C SST pulse. The length scale associated to the representer is about several hundreds of km, the typical correlation scale for atmospheric cyclone depressions. The anisotropic shape of the representer reflects the large zonal component of the atmospheric circulation. Figure 3 shows the seasonal cycle of the same representer at 22°W/39°N: The amplitude of the seasonal cycle is large, with a minimum in summer, when the mixed layer is thin (i.e. the heat capacity of the ocean is weak). In winter it is the opposite: a large heat flux is necessary to produce a 1ºC SST change because of the large depth of the mixed layer.

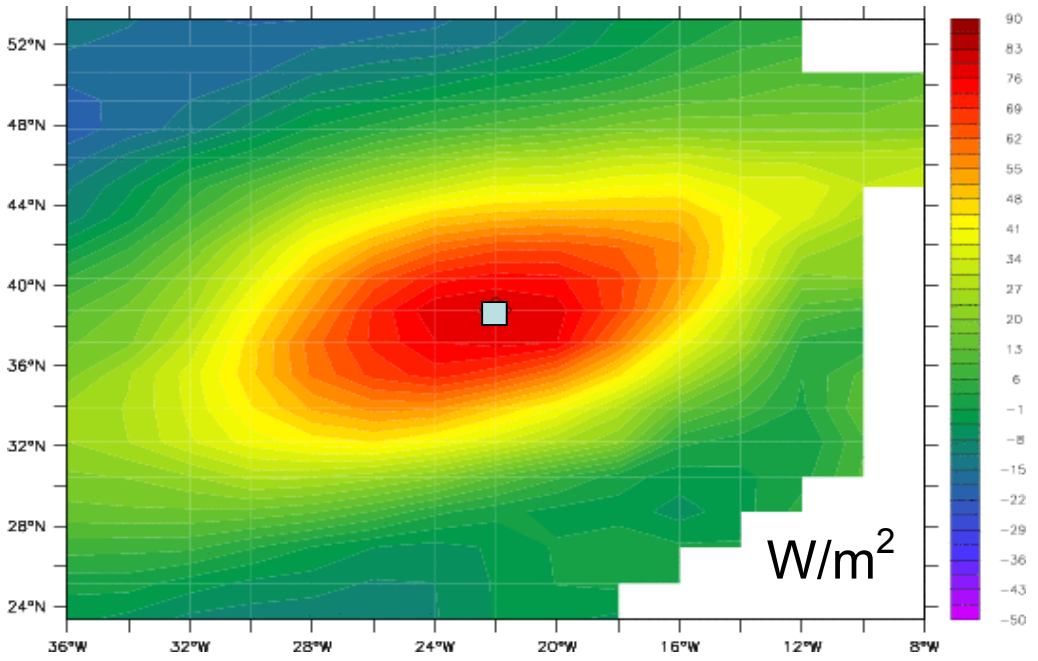


Figure 2

Heat flux δQ^a representer p(δSST, δQ^a) of a +1°C SST impulse at 22°W, 39°N (grey square) (contours, min=-50, max=90), in June.

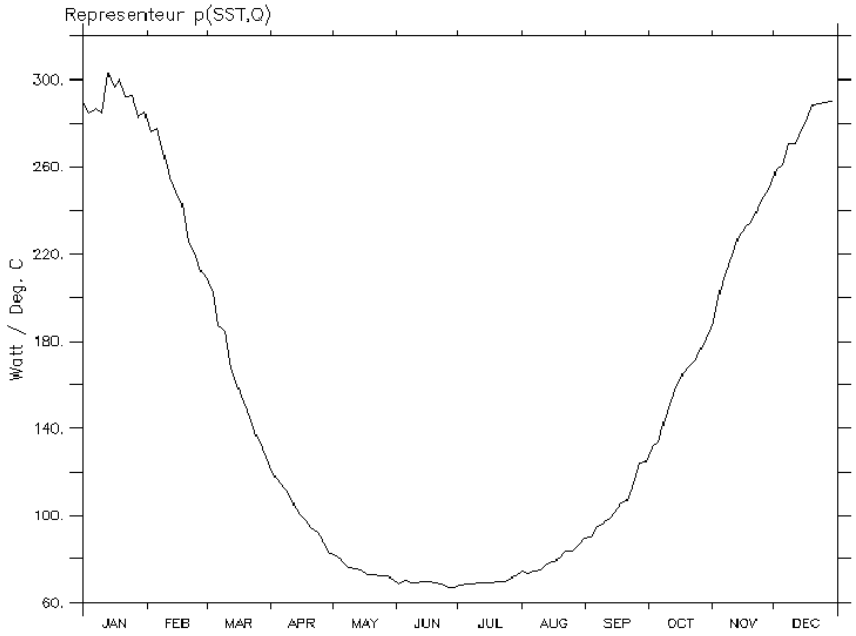


Figure 3

Seasonal variability of the covariance at the surface between the heat flux control variable δQ^a and a +1°C SST impulse located at (22°W, 39°N). Note the large surface heat flux (~ 300W/m2) in winter, consistent with a deep mixed layer (ie high heat capacity of the ocean).

In general, covariances are used to project the observation information onto the control variable. In Figure 4 we show the "reverse" representer $p(\delta Q^a, \delta SST)$ (i.e. control variable projected onto observation) of a +100 W/m^2 heat flux pulse during 7 days at 22°W/39°N. This means that if surface heat flux observation were available, the impact of a single perfect (no observation error) observation would generate a temperature correction as illustrated in Figure 4. The temperature response shows a strong seasonal cycle which results in summer in an intense correction in temperature localized within the first 10~20 meters of the ocean, which indicates the strong summer stratification. In winter, the same impulse of heat flux is associated to a much weaker temperature correction, but spreads over a thicker ocean layer (~90 m depth). We can note that the representer decorrelates as distance increases, without any “noisy” signals at distance. This is due to the choice of a large number of error modes (~250) to build the model covariance.

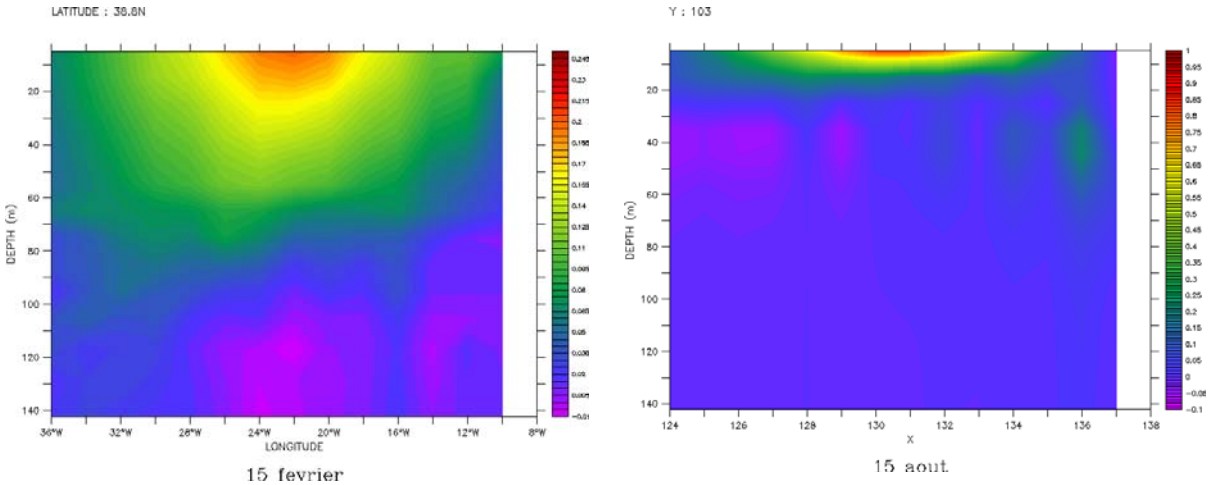


Figure 4

Zonal vertical section of the representer $p(\delta Q^a, \delta SST)$. A +100$W/m^2$ impulse (during 7 days) located at (22°W, 39°N) is associated to changes in the temperature field along the latitude 22°W. Scale: [-0.01, 0.24] °C for (a) and [-0.1, 1] °C for (b). Note the strong seasonal variability of the vertical extension associated to the mixed layer seasonal cycle in summer and winter.

We studied the geographical structure of several representers between surface temperature and surface heat flux at mid latitude in the North-eastern Atlantic, a place where significant correlation is expected between heat flux errors and upper ocean

temperature errors. The new implemented control variable T^Q introduced correlates well with SST changes and demonstrate the efficiency of the approach in the region considered. Other regions have been studied in a similar manner (not shown), with different dynamical regimes (western boundary current, uppwelling, western / eastern equatorial, Antarctic circumpolar regions). Most of the results indicate that the covariances between the control variables are consistent with our knowledge of the thermo-dynamics of theses regions. These preliminary results indicate that the approach seems realistic and provides satisfactory results in terms of the physical ocean processes description. These results are encouraging enough to let us think that the method developed shows some good realism and could be implemented for further tests in a data assimilation system.

Data assimilation of real observations

A series of experiments with and without data assimilation is carried out to test the effectiveness of surface heat flux correction by ocean data assimilation. Mercator coarse resolution global ocean data assimilation system, based on NEMO ORCA2 model configuration and SAM2 multivariate data assimilation scheme (Testut et al., 2005) is used. It is a system very similar to that used in operational at Mercator to produce initial conditions for coupled atmosphere ocean seasonal forecasting at Météo France (Balmaseda et al., 2008). However, some differences with the operational system exist: (i) the present model configuration uses the CLIO bulk formulation for ocean-atmosphere fluxes instead of the flux formulation, (ii) we only assimilate SST and temperature and salinity in situ profiles (no altimetric data) (iii) and we implement the surface heat flux correction method previously described. Three 1-year experiments (year 2005) are performed (Table 1). CTL is the control simulation, with no data assimilation, no heat flux correction. In ASSIM experiment we assimilate SST data and temperature and salinity profiles to produce surface heat flux corrections (a series of 7-day average heat flux corrections). The heat flux corrections ($\delta Q^a(k)$, k=1,52, k is the assimilation cycle number) are diagnosed but not used in that experiment. Finally, the heat flux corrections diagnosed in ASSIM ($\delta Q^a(k)$, k=1,52) are used in QCOR, an experiment with no data assimilation, to test whether the analysed heat flux help improving the model SST or not.

	DATA ASSIMILATION	An analysed surface heat flux Q^a is produced	Analysed surface heat flux Q^a is used
CTL	No	No	No
ASSIM	**Yes** SST, T&S profiles	**Yes** correction $\delta\mathbf{Q}^a(k)$ is produced	No
QCOR	No	No	Yes $\delta\mathbf{Q}^a(k)$ is applied (from **ASSIM** exp.)

Table 1

Main features of CTL, ASSIM and QCOR experiments.

The globally averaged bias and standard deviation for the SST innovations in the 3 experiments are presented in Figure 5. We are interested in that variable because it is the most sensitive oceanic parameter to surface heat flux. In the control simulation (CTL), the SST innovation bias and standard deviation are important, much higher than in the two other experiments. The cold SST bias is a known flaw of the CLIO bulk formulation. The simulation with data assimilation (ASSIM) significantly reduces the cold bias but also the rms of the SST misfit. The surface heat flux corrections produced in ASSIM experiment are then used in QCOR. The results show a great improvement in terms of SST innovation bias and rms compared to CTL. The corrections ($\delta Q^a(k)$) are entirely prescribed during the 1 year simulation and allow a significant improvement in QCOR. This experiment is even better than ASSIM in terms of SST bias. It should be noted that QCOR experiment, which benefits from an analysed surface heat flux forcing, is almost as good as ASSIM, and is largely better than CTL.

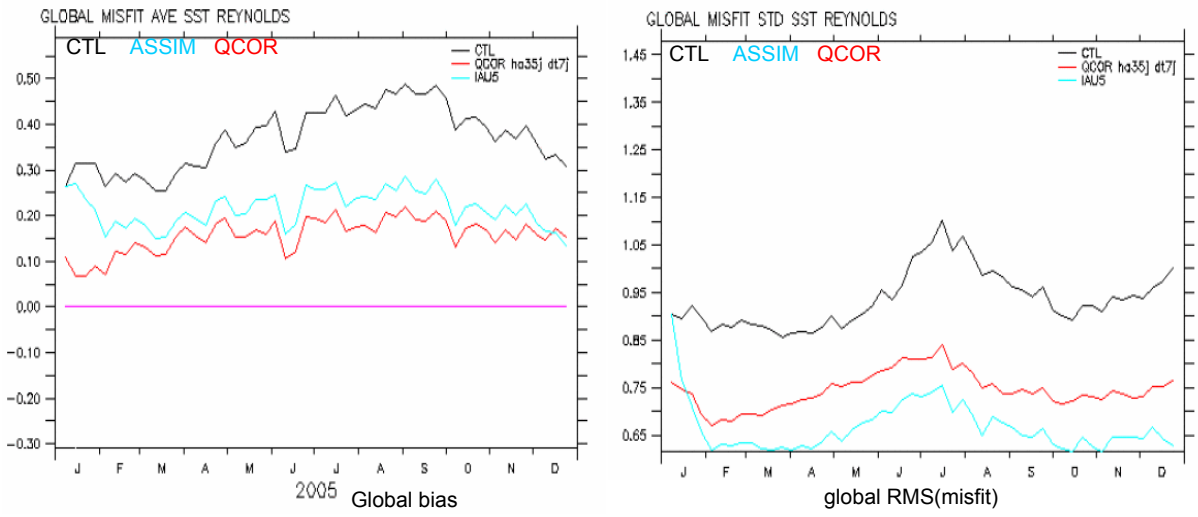


Figure 5

Globally averaged SST innovations bias and rms for the three experiments CTL, ASSIM and QCOR (in °C).

In order to have a closer look at the geographical patterns of the heat flux corrections diagnosed in ASSIM, Figure 6 shows different quantities diagnosed in this experiment. Figure 6a represents the annual average of the mixed layer temperature increment δT in ASSIM. One can notice that between 50°N/50°S the mean increment is positive, of the order of a few tenths of degree Celsius per 7 days and tend to counteract the cold bias of the model.

Figure 6b shows the annual average of the weekly heat flux corrections δQ^a diagnosed in ASSIM and used in QCOR. This quantity is positive, of the order of 15W.m^{-2}. This corresponds to the average heat flux bias present associated to the CLIO bulk formulation which overestimates the ocean-atmosphere surface cooling. Differences in the spatial patterns can be noted between the eastern / western side of the basins where the dynamics are very different (weak / strong currents). It may also be noted that in the region of the Antarctic Circumpolar Current (ACC), the heat flux corrections can be much stronger (~ 100W.m^{-2}). As we will see later, this surface heat flux correction is certainly overestimated in that area.

Figure 6c represents the annual average of δT^Q, the equivalent mixed layer temperature change associated to the heat flux correction δQ^a (see Eq. (2)). We note that the structures and intensity are similar to those of the temperature increment δT in ASSIM simulation (Figure 6a), except in the tropics (20°N/20°S) and in the western boundary current regions. This means that at middle and high latitudes, far from the western side of the gyres, much of the forecast error of temperature within the mixed layer is due to ocean-atmosphere flux. This is coherent with the thermodynamics of these regions, where a large part of the surface heat budget is due to air-sea exchanges (e.g. Gill and Niiler, 1972, Ferry et al., 2000). This is however not the case in the tropical oceans where it is mostly ocean dynamics that controls the variability of the surface layers (i.e. most of the forecast error is due to ocean dynamics and not air-sea heat fluxes).

By comparing the average ratio $R = < \delta T^Q/\delta T >$ of temperature corrections induced by the heat flux to the increments of temperature averaged over the mixed layer, it is possible to know whether surface heat flux (R ~ 1) or ocean dynamics (advection, diffusion, ie R ~ 0) is most probably responsible for the mixed layer temperature forecast error. Figure 6d shows this ratio. It appears that away from the tropics (20°N/20°S), heat flux corrections induce mixed layer temperature changes close to that induced by the unmodified assimilation system (R ~ 1, Figure. 6d). Near the equator, where the thermal structure is mostly constrained by the dynamics, heat flux errors do not explain the error on the SST (the ratio R is close to 0). We can see some noisy values of R in the western boundary current regions and in the ACC. This reflects mainly the areas where the ocean model has strong SST biases due to inaccurate front's location. In these regions, the data assimilation system awkwardly tries to reduce these biases by the way of (unrealistic) strong surface heat fluxes.

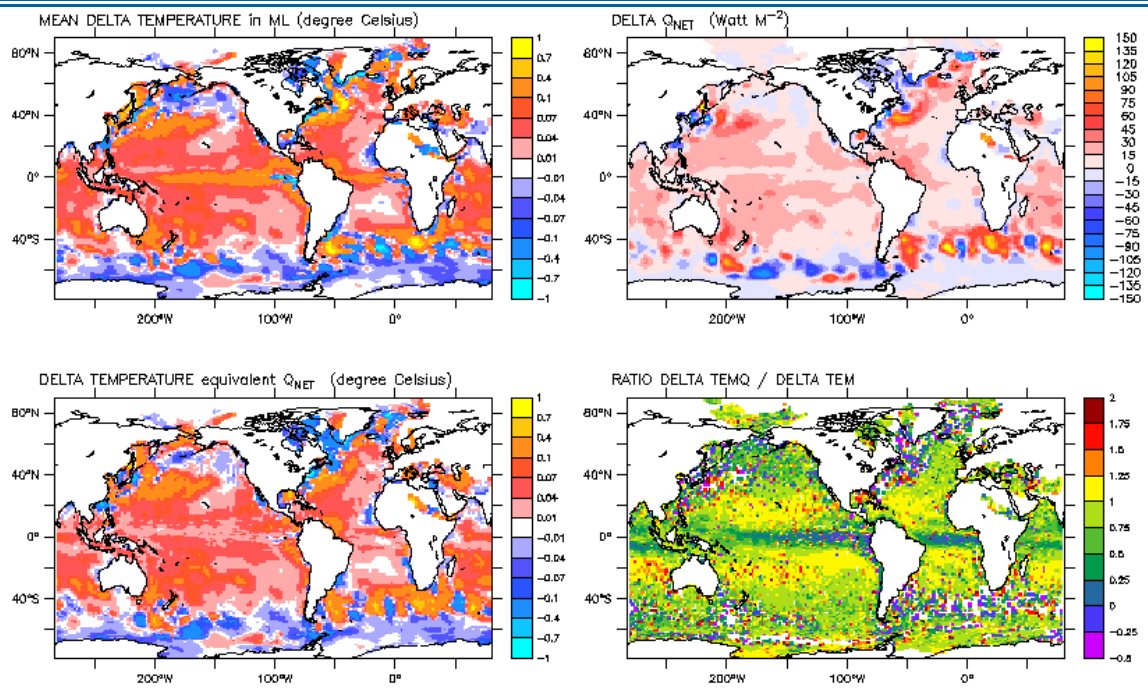


Figure 6

(a) annual mean temperature increment δT averaged within the mixed layer for ASSIM (in °C). (b) Annual mean of the heat flux correction δQ^a diagnosed in ASSIM and used in QCOR (in $W.m^{-2}$). (c) Annual mean of δT^Q, the equivalent temperature change associated to the heat flux correction δQ^a (see Eq. (2)) (in °C). (d) Ratio $R = <\delta T^Q>/<\delta T>$ of the mean equivalent temperature change associated to the heat flux correction (i.e. Figure 6c) over the mean increment averaged within the mixed layer (i.e. Figure 6c). A ratio close to 1 indicates that the heat flux is fully responsible for the forecast error in temperature; a ratio close to 0 indicates that other processes than surface heat flux may be related to the SST forecast error.

These results are however very encouraging with respect to that method of heat flux correction with ocean data assimilation. Away from the strong current regions the diagnosed surface heat flux corrections seem realistic and physically consistent with our knowledge of the ocean thermo-dynamics and of the model deficiencies.

Conclusions and perspectives

A method for correcting the air-sea heat flux by ocean data assimilation is presented. A new variable has been introduced in the control vector which is the integral of the net surface heat flux received by the ocean. The heat flux correction is then deduced from this new variable analysis increment, assuming that the increment is the integral of the error of heat flux on the window of assimilation. A preliminary study of some representers shows a satisfactory behaviour of this new control variable in terms of seasonal variations and spatial correlation lengths (horizontal and vertical).

This method is then implemented in a realistic framework. A data assimilation experiment with Mercator assimilation system version 2 (SAM2) and coarse resolution model ORCA2 is used to obtain a series of net heat flux corrections for the year 2005. These corrections have an order of magnitude similar to the error bars generally admitted on air-sea heat fluxes, except in areas of strong meso-scale activity (i.e. frontal zones) where the corrections can be very large, because of SST biases in the model. The temperature increment associated to the heat flux correction (average within the mixed layer), is compared to the mixed layer temperature increment. The average ratio of these quantities ($R = \delta T^Q/\delta T$) is close to 1 at middle and high latitudes, while it is almost zero at the equator. This indicates that except in the tropical region and near western boundary currents, the error on the temperature in the upper ocean layer is largely explained by the heat flux. This is consistent with our knowledge of the ocean thermo-dynamics.

A twin experiment is then carried out with: (i) a control run and (ii) a model integration forced with the analysed surface heat fluxes obtained previously. The results show that the SST misfit average and rms are dramatically reduced. In regions where the SST bias (wrong position of the fronts for example) is large, the data assimilation system tries to correct this bias by adjusting the surface heat flux. In these areas, the heat flux correction analysed is sometimes too strong and unrealistic, and this may lead to too strong stratification / convection in the ocean upper layers. A limitation of the heat flux increment appears to be necessary in these areas. A way to achieve this would be to limit the heat flux increment, i.e. to impose the ratio R to be in the range [0;1].

It is interesting to notice that the air-sea flux correction method described here is easily usable in diagnostic mode, to provide for example surface fluxes consistent with the analyzed state oceanic. This possible application is considered in the context of GLORYS ¼ ° global ocean reanalysis supported by Groupe Mission Mercator-Coriolis. The direct use of the analysed heat flux corrections in the data assimilation system (using the heat flux increment in the equations of the dynamic with an Incremental Analysis Update (IAU) method, Bloom et al., 1996) has not been addressed here. The main difficulty is not to correct twice the temperature field. Indeed, the temperature increment provided by the analysis includes errors in the mixed layer processes, in the ocean dynamics... but also errors in the surface fluxes. Thus, one should remove the part due to surface fluxes to the temperature increment, in order to apply (with an appropriate IAU method) both the modified temperature increment and the analysed heat flux correction. Finally, the approach presented here may also be implemented to correct the precipitation from salinity observations provided by Argo profiling floats array and future SMOS mission (to be launched this year) and also the wind stress at the surface (Chelton and Song, 2009).

References

Balmaseda M., O. Alves, A. Arribas, T. Awaji, D. Behringer, N. Ferry, Y. Fujii, T. Lee, M. Rienecker , T. Rosati, D. Stammer, 2008 : Ocean Initialization for Seasonal Forecasting. Final Symposium of the Global Ocean Data Assimilation Experiment, invited paper, 12-15 November, 2008, Nice, France.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. Mon. Wea. Rev., 124:1256–1271.

Bonekamp, H., G. van Oldenborgh, and G. Burges (2001), Variational Assimilation of Tropical Atmosphere-Ocean and expendable bathythermograph data in the Hamburg Ocean Primitive Equation ocean general circulation model, adjusting the surface fluxes in the tropical ocean, J. Geophys. Res., 21:16693–16709.

Chelton D. B., Q. Song, 2009: Observations and Modeling of SST Influence on Surface Winds, proc. ECMWF Workshop on Ocean-Atmosphere Interactions, 10-12 November 2008, http://www.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2008/Atmos_Ocean_Interaction/Chelton.pdf

Ferry, N., G. Reverdin, and A. Oschlies, 2000: Seasonal sea surface height variability in the North Atlantic Ocean, J. Geophys. Res., 105:6307–6326.

Gill, A.E., and P.P. Niiler, 1973: The theory of the seasonal variability in the ocean, Deep Sea Res., 20:141-177.

Pham D.T., J. Verron and M. C. Roubauda, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. J. of Mar. Sys., 16:323-340.

Vossepoel F. C., A. T. Weaver, J. Vialard and P. Delecluse, 2004 : Adjustment of Near-Equatorial Wind Stress with Four-Dimensional Variational Data Assimilation in a Model of the Pacific Ocean. Mon. Weather Rev, 132. 2070–2083. DOI: 10.1175/1520-0493(2004)132.

Skachko S., J.-M. Brankart, F. Castruccio, P. Brasseur and J. Verron, 2006 : Air-sea fluxes correction by sequential data assimilation. Mercator Ocean Quarterly Newsletter #22 – July 2006 – 24-28. http://www.mercator-ocean.fr/documents/lettre/lettre_22_en.pdf

Skandrani C., J-M.Brankart, N. Ferry, J. Verron, P. Brasseur, and B. Barnier, 2009: Controlling atmospheric forcing parameters of global ocean models: sequential assimilation of sea surface Mercator-Ocean reanalysis data, Ocean Sci. Discuss., 6, 1129–1171.

Stammer D., K. Ueyoshi, A. Köhl, W. G. Large, S. A. Josey, and C. Wunsch, 2004: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation, J. Geophys. Res., 109, C05023doi:10.1029/2003JC002082

Testut C.E., B. Tranchant, F. Birol, N. Ferry and P. Brasseur, 2005 : SAM2: The Second Generation of MERCATOR Assimilation System, Geophysical Research Abstracts, Vol. 7, 06816, 2005, SRef-ID: 1607-7962/gra/EGU05-A-06816.

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. Marine Systems, 40-41:291-316. doi:10.1016/S0924-7963(03)00022-8

Data Assimilation in the Australian Bluelink System

By Peter R. Oke[1,2] , Gary B. Brassington[1,3], David A. Griffin[1,2], Andreas Schiller[1,2]

[1]Centre for Australian Weather and Climate Research, Australia
[2]Commonwealth Scientific and Industrial Research Organisation Wealth from Oceans National Research Flagship, Hobart, Australia
[3]Bureau of Meteorology, Melbourne, Australia

Abstract

The Bluelink Ocean data assimilation system (BODAS) is an ensemble-based system that underpins Australia's operational short-range ocean forecast system. The primary test-bed for the Bluelink system is the series of Bluelink ReANalysis (BRAN) experiments. Over the life of Bluelink, BRAN experiments have been used to assess the performance of the system, and to test new developments prior to integration into the forecast system and operational trials. BRAN experiments have helped identify problems with the model, assimilation system, data processing, and model initialisation. In this paper, the recent improvements of the Bluelink system are highlighted, along with some preliminary results from the application of BODAS to a relocatable coastal ocean model, also developed under Bluelink.

Introduction

Bluelink is a partnership between the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the Bureau of Meteorology (BoM) and Royal Australian Navy (RAN). The primary objective of Bluelink is to develop and improve Australia's capabilities in short-range ocean forecasting and reanalysis. The Bluelink forecast system (Brassington et al. 2007) first became operational at the BoM in August 2007, and has since produced two 7-day forecasts each week. The main components of the Bluelink system are the Ocean Forecasting Australia Model (OFAM) and the Bluelink Ocean Data Assimilation System (BODAS). The primary test-bed for the Bluelink system is the series of Bluelink ReANalysis (BRAN) experiments – multi-year data assimilating model runs.

The purpose of this paper is to describe the Bluelink system, particularly BODAS and its recent enhancements, and to review some of the lessons learnt from a series of BRAN experiments. This paper is organized as follows: a short description of the Bluelink model is presented, followed by a description of BODAS. A summary of a series of BRAN experiments are described, followed by results from a recent BRAN experiment, and a demonstration of the application of BODAS to a coastal ocean forecast system.

Bluelink Ocean Model - OFAM

The global model used here is based on the Modular Ocean Model (Griffies et al., 2004) and is called the Ocean Forecasting Australia Model (OFAM). The first version of OFAM, OFAM1, used version 4p0d. OFAM2, which is still being developed, uses version 4p1. The horizontal resolution of OFAM varies from 2^{o} in the North Atlantic to $1/10^{o}$ in the 90^{o}-sector centred on Australia and south of 16^{o}N. OFAM1(2) has 47(51) levels in the vertical, with 20(24) levels in the top 200 m, and 35 levels in the top 1000 m, with a minimum of 10(5) m resolution near the surface. The horizontal grid has 1191 and 968 (1191 and 997) points in the zonal and meridional directions, respectively. The bottom topography for OFAM1 was a composite of a range of different sources, including dbdb2 (provided by the United States Naval Research Laboratory) and the General Bathymetric Charts of the Ocean (GEBCO) and AGSO2002. The OFAM2 topography is based on the Smith and Sandwell (1997) v11.1 bathymetry. The model uses the third-order quicker scheme for tracer advection (Leonard, 1979). Horizontal viscosity is resolution and state-dependent based on the Smagorinsky-scheme (Griffies and Hallberg, 2000). The turbulence closure model used by OFAM is the hybrid mixed-layer scheme described by Chen et al. (1994).

For long model runs, such as free spin-up runs and BRAN experiments, OFAM is forced by 6-hourly atmospheric fluxes from ECMWF, using fields from ERA-40 (Kallberg et al., 2004), for the period prior to August 2002, and 6-hour operational forecasts thereafter. The operational Bluelink forecast system uses 6-hourly forcing from the BoM Global Atmospheric Prediction System (GASP, e.g., Schulz et al. 2007; soon to be replaced with a version of the Unified Model, Rawlins et al. 2007).

Bluelink Ocean Data Assimilation System - BODAS

The Bluelink Ocean Data Assimilation System (BODAS; Oke et al. 2008) was initially developed for data assimilation into a global ocean forecast system. The requirements of such a system are to facilitate the assimilation of different observation types, in all possible dynamical regimes, including those of the open ocean, shelf zones and marginal seas. The assimilation of

multiple observation types makes a multivariate assimilation preferable, whereby observations of one type (e.g., sea-level) influence the increments to model fields of all types (e.g., sea-level, temperature, salinity, velocity). The requirement to assimilate in a variety of different regions and dynamical regimes encourages the adoption of inhomogeneous and anisotropic background error covariance estimates, since background errors in different regions are expected to be characterised by different length-scales, and with different orientations. Multivariate, inhomogeneous, and anisotropic covariance estimates are readily obtained using ensemble data assimilation methods. It is for this reason that the Bluelink team opted to develop an ensemble-based data assimilation system. The salient aspects of BODAS are as follows:

BODAS employs an ensemble optimal interpolation (EnOI) scheme that uses a stationary ensemble of intraseasonal model anomalies, or modes, to approximate the system's background error covariance. Because we expect the background field errors of a short-range forecast system to be dominated by eddy-scale features, the ensemble is comprised of ensemble members that contain eddy-scale variability. In practice, this is achieved by computing each ensemble member by high-passing a long model run. At present, Bluelink applications use (up to) 120-ensemble members, computed from the last 10-years of a 15-year free run of OFAM. Each ensemble member is a 3-day mean minus the 3-month mean centered at the same time. The current operational system uses a 72-member ensemble.

An important feature of BODAS is covariance localisation. Using ensemble data assimilation, the influence of an observation on the model state is determined by the ensemble-based covariance between the observed state element and all other state elements. Because the ensemble is small compared to the dimension of the model subspace, the ensemble is rank-deficient and suffers from sampling error (Houtekamer and Mitchell 2001; Mitchell et al. 2002; Oke et al. 2006). The rank-deficiency means that the ensemble does not have enough degrees of freedom to adequately fit the model-data misfits (background innovations) during an assimilation step. The sampling error means that the ensemble-based covariances are noisy – particularly for long-distance covariances that are really expected to be zero. For example, sea-level errors in the Tasman Sea are not expected to be correlated with sea-level errors in the Gulf of Mexico. However, for a small ensemble, the ensemble-based covariance may be non-zero. These artificial long-distance covariances are eliminated in practice by multiplying the ensemble-based covariance by a localising correlation function (Houtekamer and Mitchell, 2001). Here, the localising function is a homogeneous, isotropic, quasi-Gaussian function with an e-folding length-scale of about 2-3 degrees. As a result, the influence of an individual observation on the model state depends on both the ensemble-based covariances and the distance between the observed location and the location of each model state element. For the covariances over short distances (less than a few hundred metres), the details of the ensemble-based covariance - including the length-scales, inhomogeneity, and the anisotropy – are retained when localisation is used (Oke et al. 2005). But the long-distance covariances are eliminated.

At present, BODAS routinely assimilates along-track sea-level anomaly (atSLA) data from all available satellite altimeters and coastal tide gauges from around Australia, plus Sea Surface Temperature (SST) observations from the Pathfinder database and AMSR-E passive microwave radiometer. Recent developments permit the assimilation of GHRSST L2P data (e.g., NAVOCEANO L2P AVHRR). In-situ temperature and salinity observations from Argo floats, the Tropical Atmosphere–Ocean (TAO) array, CTD and XBT (temperature only) surveys from a variety of different field surveys, including WOCE, Indian Ocean Thermal Archive (IOTA) and others, are also routinely assimilated. Explicit observation error estimates are assigned to each observation according to their expected instrument error, their “age” relative to the analysis time, and an estimate of their representation error – see Oke et al. (2008) for details. Representation error estimates are obtained using the method described by Oke and Sakov (2008). This method yields error estimates that depend on the model grid. For example, where the model grid spacing is $1/10^{o}$ the representation error is small – because the model and observations can “represent” variability of comparable scales. However, where the model is coarse, say 2^{o} in the North Atlantic, the representation error is large – because the model cannot represent all of the features and variability represented by the observations. Although somewhat counter-intuitive, this difference in representativeness is ascribed as an error to the observation, so that the model doesn’t over-fit the data by “trying” to reproduce scales that are not resolvable on the model grid. That is, so the analysis step doesn’t try to fit what the model regards as noise.

Despite the fact that the ocean is under-sampled, the number of discrete satellite observations is too large to be efficiently assimilated directly by BODAS. This is addressed by assimilating super-observations for SLA and SST, and by selecting only a sub-set of in situ temperature and salinity profiles to assimilate. The calculation of super-observations simply refers to the spatial averaging of SLA and SST data prior to assimilation. Super-observations are ascribed a smaller error, depending on the distribution and number of observations that are averaged. The amount of averaging and sub-sampling done by BODAS is flexible, and can readily be modified for different scenarios. Because of the spatially varying resolution of OFAM, a typical application of BODAS involves super-obing the SLA and SST data to a nominal resolution of $4/10^{o}$-$6/10^{o}$ around Australia (i.e., every 4^{th} or 6^{th} model grid point) and coarser elsewhere. Similarly, it is typical to select one temperature and salinity profile every degree around Australia, and coarser elsewhere. For short experiments that are focused on a particular region, or event, the resolution of the super-obing and sub-sampling can easily be modified to retain more observations in regions of particular interest.

BODAS calculates a global analysis of the model state by performing many (approximately 500) analyses on sub-domains of the model grid. For each sub-domain, observations from a halo around that sub-domain are used to influence the analysis. Provided the extent of the halo is chosen to match the distance over which the localizing function goes to zero, adjacent sub-domains produce analyses that are seamless at their point of intersection (i.e., spatially continuous), and the analysis of the full model state is equivalent to a global inversion (Figure 1). This approach differs from many ensemble-based systems (e.g., Houtekamer and Mitchel 2001; Brasseur et al. 2006; Bertino et al. 2008), who compute analyses, one grid point at a time, using observations only in the vicinity of each grid point.

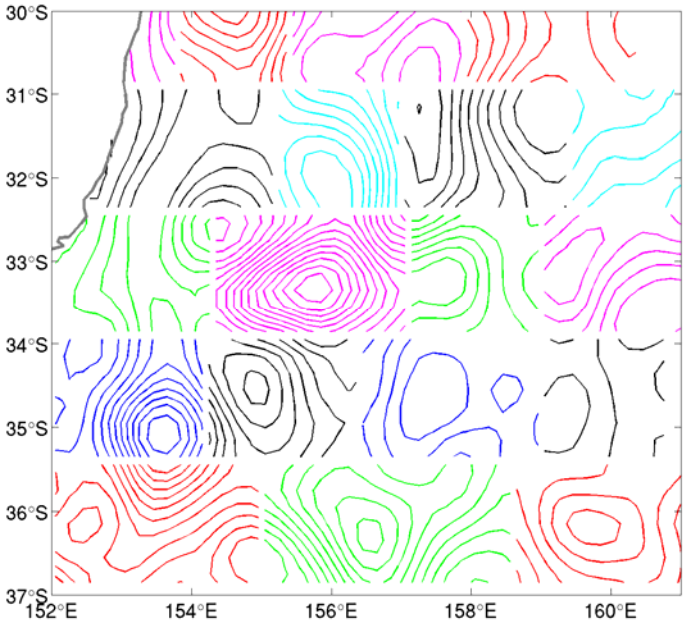


Figure 1

An example of the increments for sea-level (the contour interval is 5 cm) in the Tasman Sea produced by BODAS. Different colours represent increments computed independently. The meridional extent of each sub-domain is pre-determined, but the zonal extent of each sub-domain is adaptive, and depends on the density of the observations. Note the continuity of the increments in adjacent sub-domains.

Since it's development, BODAS has been used for many different applications, including global reanalyses (Schiller et al. 2008), operational global ocean forecasting (Brassington et al. 2007), seasonal prediction (http://poama.bom.gov.au/research/assim/index.htm), observing system evaluation (Oke and Schiller 2007), observing system design (e.g., Oke et al. 2009), and more recently, regional (Sandery and Brassington 2008) and coastal data assimilation (see below). Some examples of these applications are described below.

Reanalysis Experiments

BRAN experiments are typically multi-year data-assimilating model runs. The purposes of BRAN experiments are twofold. Firstly, BRAN experiments are intended to facilitate testing and development of new versions of the Bluelink System prior to operational trials. Secondly, BRAN experiments are intended to provide a service to the research community for understanding ocean variability and dynamics. In this section, a review of BRAN activities is presented, along with some scientific results on ocean variability around Australia.

To date, two long (>12 years) BRAN experiments have been performed (BRAN1 and BRAN2p1), two intermediate-length (1-4 years) experiments have been performed (BRAN1p5 and BRAN2p2), and several short (3-6 month) experiments have been performed. These experiments differ in the time periods simulated, the data that is assimilated, frequency of assimilation, forcing fields, ensemble size, and the method of initialisation. A summary of the configuration of each of the main BRAN experiments is given in Table 1.

	BRAN1	BRAN1p5	BRAN2p1	BRAN2p2
Time period	10/1992-12/2004	1/2003-6/2006	10/1992-12/2006	4/2006-4/2008
Data assimilated	atSLA, T/S	atSLA, T/S, SST	atSLA, T/S, SST	atSLA, T/S, SST
Ensemble size	72	72	120	120
Assimilation interval (d)	3	7	7	7
Surface forcing	ECMWF fluxes	ECMWF fluxes	ECMWF heat/PmE fluxes & 10 m winds	ECMWF heat/PmE fluxes & 10 m winds
Rivers	none	none	seasonal	seasonal
SST (SSS) restoring	30-d (30-d)	none (none)	none (30-d)	none (30-d)
Initialisation	Updates to U, V, T, S and η in single step	Nudging to T, S & η with 1-d time-scale	Nudging to T, S & η with max(1-d,T_{in}) time-scale.	IAU to T, S, U, V, and η over 12-hours.
Known problems	Error in surface heat fluxes & bugs in BODAS	Some in situ T profiles processed incorrectly	Topographic errors in some shallow Straits	Topographic errors in some shallow Straits

Table 1

Summary of the configuration for BRAN experiments (atSLA is along-track sea-level anomaly; T/S refers to in situ temperature and salinity observations, including vertical profiles and mooring observations; SST is sea-surface temperature - to date the only SST data assimilated by BRAN is from the Pathfinder data base and from the AMSR-E mission; PmE is precipitation minus evaporation; SSS is sea-surface salinity; η is model sea-level; T_{in} is the local inertial period; IAU is Incremental Analysis Updating; ECMWF fluxes refers to ERA-40 prior to 10/2002 and ECMWF 6-hour forecasts thereafter).

Results from BRAN1 are described by Oke et al. (2005). This study demonstrates that BRAN can produce realistic mesoscale variability around Australia. However, this study also identified some problems with the Bluelink system. An error was identified in the way the surface heat flux was applied that resulted in the development of a warm bias. Some bugs were found in BODAS that meant that the salinity updates were incorrect for the first 4 years of the run. Initialization shocks, resulting from the model being updated in a single time-step, sometimes seriously degraded the reanalysis, and made the reanalyzed fields quite noisy in both space and time. All of these problems were addressed prior to operational trials of the first Bluelink forecast system at the BoM, and prior to the performance of the following BRAN experiments.

Results from BRAN1p5 are described by Oke et al. (2008). This study includes a more comprehensive assessment of BRAN, including comparisons with with-held observations. Quantitatively, it was shown that reanalyzed fields in the region around Australia in BRAN1p5 are typically within 6–12 cm of withheld atSLA observations, within 0.5–0.9°C of observed SST, and within 4–7 cm of observed coastal sea-level. Comparisons with Argo profiles and surface drifting buoys show that BRAN1p5 fields are within 1°C of observed sub-surface temperature, within 0.15 psu of observed sub-surface salinity, and within 0.2 m/s of near-surface currents. The fields produced by BRAN1p5 are smooth and look realistic. But it is clear from the model-data comparisons that most of the observations are under-fitted. Based on this study, initialization was identified as a key area in which the Bluelink system could be improved. Analysis of the time-mean and root-mean squared increments to sea-level also identified some biases in sea-level (Oke et al. 2008). The largest of these biases tend to be along the path of the Antarctic Circumpolar Current (ACC), indicating that perhaps the mean sea-level (MSL) field used for BRAN1p5 was inadequate. The MSL field used for all completed BRAN experiments is the time-mean of a 15-year non-assimilating run of OFAM1. A revised MSL field has recently been generated by constraining a multi-year run tightly to climatological temperature and salinity in a so-called diagnostic run using OFAM2. Research on this aspect of the Bluelink system is ongoing.

Results from BRAN2p1 are described by Schiller et al. (2008). This study includes a description of the salient features of the reanalysed circulation in the Australasian region. It included comparisons with observed and reanalyzed transport estimates for the key regions around Australia. The total (top-to-bottom) annual mean transport through the Indonesian straits, and its standard deviation, are 9.7 ± 4.4 Sv from the Pacific to the Indian Ocean with a minimum in January (6.6 Sv) and a maximum in April (12.3 Sv). The circulation of the Leeuwin Current, along the west coast of Australia, is dominated by eddy variability with a mean southward transport of 4.1 ± 2.0 Sv at 34°S. Off southern Australia, the eastward South Australian Current advects 4.5 ±

2.6 Sv at 130°E. At 32°S the East Australian Current transports 36.8 ± 18.5 Sv southward. The Coral Sea exhibits a quasi-permanent gyre between north-eastern Australia and Papua-New Guinea that is associated with the Hiri Current, which flows along the south coast of Papua-New Guinea and advects 8.2 ± 19.1 Sv into the Western Pacific Ocean. The results from BRAN2p1 are much better than BRAN1, and are very similar to those in BRAN1p5 – and like BRAN1p5, BRAN2p1 fields tend to under-fit the assimilated observations.

The latest BRAN experiment is BRAN2p2. The main difference between BRAN2p2 and BRAN2p1 and 1p5 is the initialization. The Incremental Analysis Updating (IAU) method described by Bloom et al. (1996) was adopted for BRAN2p2. Recall that BRAN2p1 and 1p5 both used nudging (Table 1). For this experiment, the increments were applied over 12 hours, with a constant weight. Examples of velocity fields at 145 m depth off south-west Western Australia, in May 2006, from BRAN1p5, 2p1, 2p2, and observations are presented in Figure 2. The observed velocity maps presented are each based on 9-days of ship-board Acoustic Doppler Current Profiler (ADCP) measurements, collected during an R. V. *Southern Surveyor* cruise in May 2006 (data provided courtesy of M. Feng) and follows a similar cruise in 2003 (Feng et al. 2007). The BRAN fields, presented for comparison, are time-averages over the sampling periods. In these example, a pair of counter-rotating eddies are evident in the observations. The eddies have a radius of only about 0.5° – approximately 50 km. This means that they are only just resolvable by a 1/10° resolution model – approximately 10 km resolution – with only about 5 grid points from the eddy center to its outer boundary. Similarly, these features are only just resolvable by the observing system. The standard GDR altimeter data, for example, comprise one estimate every 7 km along-track, and track separations are typically over 100 km. Despite these limitations, there is some evidence of these eddies in BRAN1p5 and BRAN2p1. However, they are not well reproduced. By contrast, the reproduction of these eddies in BRAN2p2 is very good, with even some of the asymmetric shapes of these eddies reproduced. There remains some errors in BRAN2p2 in the reproduction of the position of these eddies, however BRAN2p2 clearly represents an improvement in the Bluelink system in this region at this time, compared to earlier versions. The main difference between BRAN2p1 and BRAN2p2 is the initialization – nudging versus IAU.

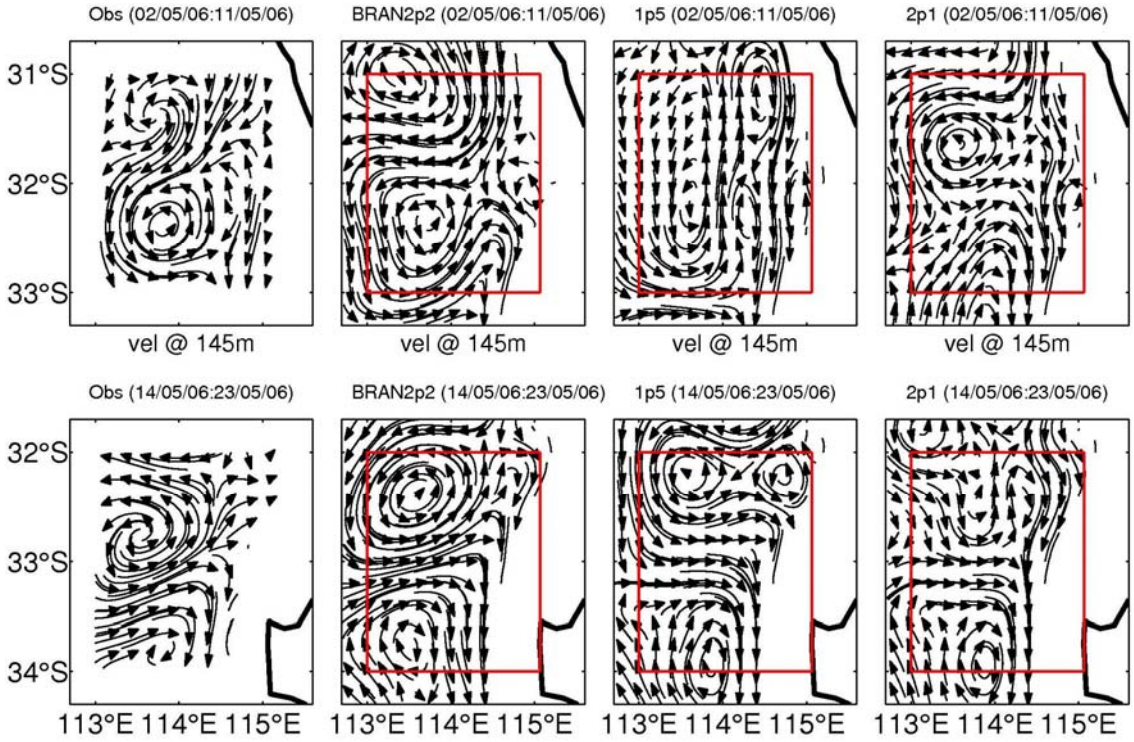


Figure 2

An example of velocities at 145 m depth from observations (left; courtesy of M. Feng) and from different versions of the BRAN for two different periods (top and bottom) off south-west Western Australia. The observed fields are mapped from ship-board ADCP measurements collected over a 9-day period. The BRAN fields are time-averages over the same period. The extent of the observation region is denoted by the red box over the BRAN fields.

Global Data Assimilation

An example of sea-level fields from the BRAN2p2 with drifter-derived velocities and trajectories overlaid is presented in Figure 3. The BRAN fields are monthly means and include the MSL. The drifter data are from the entire month. The drifter data represents the time-varying ocean circulation and is a measure of the time-integrated circulation. This is not necessarily well represented by the monthly mean sea-level fields of BRAN, but provided the variability of the circulation over each month is not too large, this comparison provides an independent assessment of the reanalyzed circulation. Note that data from the surface drifting buoys are not assimilated into BRAN. In general there is good agreement between the drifter trajectories and the sea-level contours, indicating that there is independent agreement between the reanalyzed and observed circulation. The examples in Figure 3 include situations where the drifter trajectories cross the sea-level contours. This is due on occasions to the effects of wind, or may be because a mean field (sea-level) is being compared to a Lagrangian description of the circulation (drifters). It may also be because the mesoscale features reproduced in BRAN are not precisely in the correct positions, or with the correct structures.

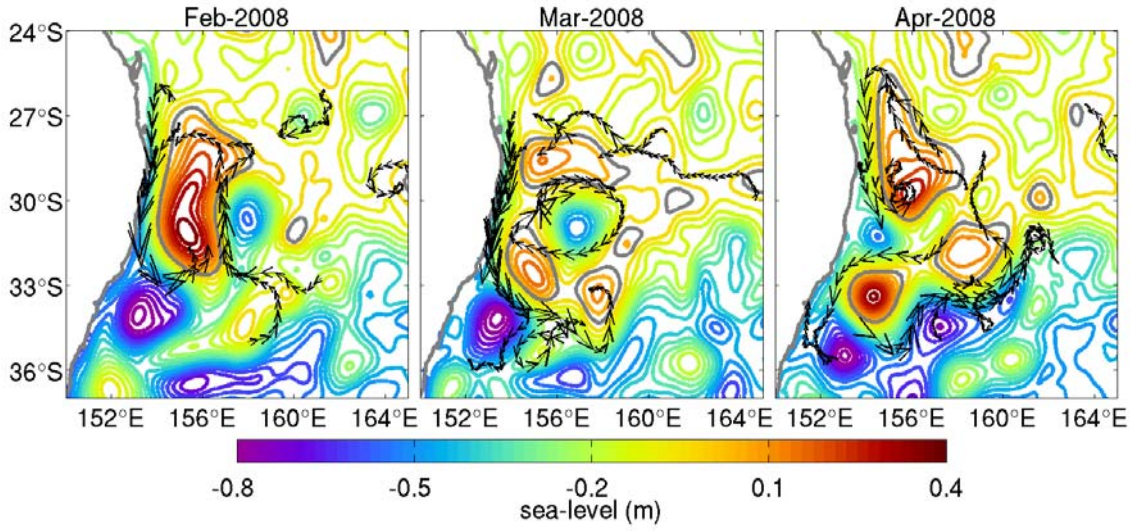


Figure 3

Monthly mean sea-level from BRAN (version 2p2), with surface drifter velocities and trajectories overlaid

Coastal Data Assimilation

In addition to the development and application of the Bluelink global forecast and reanalysis system, the Bluelink team has developed a relocatable ocean atmosphere model (ROAM). The ROAM system is controlled by a graphical user interface that enables a non-expert user to quickly define the extent of a model domain, a forecast period (e.g., 1-7 days), and the key model components (i.e., ocean, atmosphere, waves), and execute a forecast independently in near-real-time. The intension is for ROAM to be applied by an operator for domains of around 100 to 500 km in extent. The resolution of the ocean component of ROAM is typically 1-10 km, and the model is nested within either the Bluelink ReANalysis (BRAN) system for delayed-mode applications, or the operational Bluelink forecast system for near-real-time forecasts. A recent development under Bluelink is the incorporation of BODAS into the ROAM control system and the addition of ocean data assimilation to the user's choice of specifications. The benefits of the addition of ocean data assimilation to ROAM is demonstrated here through an example to the Bonney coast, a region of frequent wind-driven upwellings, off South Australia.

The ocean model used in ROAM is the Sparse Hydrodynamic Ocean Code (SHOC; Herzfeld 2009). SHOC is a z-level primitive equation model that has been developed at CSIRO over many years. For this application, the horizontal resolution of SHOC is 5 km – twice the resolution of the Bluelink model. The surface wind stress is the same for BRAN and SHOC, and is from ERA40.

Within ROAM, SHOC is typically integrated for up to a 7-day forecast. For the examples presented here, daily mean fields of velocity, temperature, salinity, and sea-level from BRAN2p1 are used to construct the initial and boundary fields for SHOC. Each integration of SHOC includes a 4-day spin-up period, followed by a 7-day forecast.

The data assimilation in BRAN is sequential, and is performed on a 7-day update cycle. BRAN can therefore be considered to be a series of 7-day forecasts. For the examples described here, the SHOC forecasts are synchronised with the BRAN update cycle so that BRAN forecasts can be directly compared to SHOC forecasts, with and without data assimilation - hereafter denoted as SHOC and SHOC+DA, respectively. A sequence of 8 forecasts cycles are reported here, including a 4-day spin-up

and 7-day forecast for each cycle. The period chosen for this comparison is the 2-month period spanning February/March 1995. This period corresponds to a series of wind-driven upwelling events, and is the focus of a detailed study by Griffin et al. (1997), who sought an explanation for a massive Pilchard die-off that occurred off southern Australia at this time.

The version of BODAS that is applied to SHOC is the same as that used for both the reanalysis and operational Bluelink systems. For the SHOC+DA runs, all of the assimilation calculations are performed on a sub-domain of the global model, with $1/10^{o}$ resolution. For each day of the 4-day spin-up period, the BRAN fields are modified by BODAS. For each day, BODAS treats the daily mean BRAN fields as the background field, and combines these fields with SST observations from 3-day composite AVHRR fields produced by CSIRO. An analysis field is generated for each day of the 4-day spin-up ($t<0$), and the increments from the last day of the spin-up ($t=0$) are used to update the "analysis" fields for the forecast period ($t>0$). This is intended to reduce the discontinuities in time that may occur in the transition from the spin-up period to the forecast period. In practice, on day 1 of the forecast, 80% of the day 0 increment is applied to the BRAN fields. On day 2, 60% of the day 0 increment is applied, and so on. This aspect of the assimilation has not yet been tuned properly.

Note that the assimilation performed here is all done on a sub-domain of the global model grid – not on the grid of the coastal model. One advantage of this is that the ensemble from the global model can readily be used for coastal data assimilation. Of course, this assumes that the statistics of both the global and coastal models are comparable. Another advantage is that the coastal model, SHOC, is integrated in almost the same manner for both the free run without data assimilation, and the run with data assimilation.

An example of the SST field from independent (un-assimilated) observations and from 5-day forecasts from BRAN, SHOC and SHOC+DA is presented in Figure 4. This Figure shows a strong signature of wind-driven upwelling in the observations, with very cold waters upwelled to the surface and becoming advected offshore. The wind stress prior to this period is moderate and upwelling favourable (Figure 4f). Despite the coarse resolution of the ERA40 forcing fields used here ($2x2^{o}$), BRAN produces an upwelling, but it is weaker than the observed event. Similarly, SHOC produces an upwelling, but is also too weak. The SHOC+DA run produces a stronger upwelling that is in better agreement with the observations.

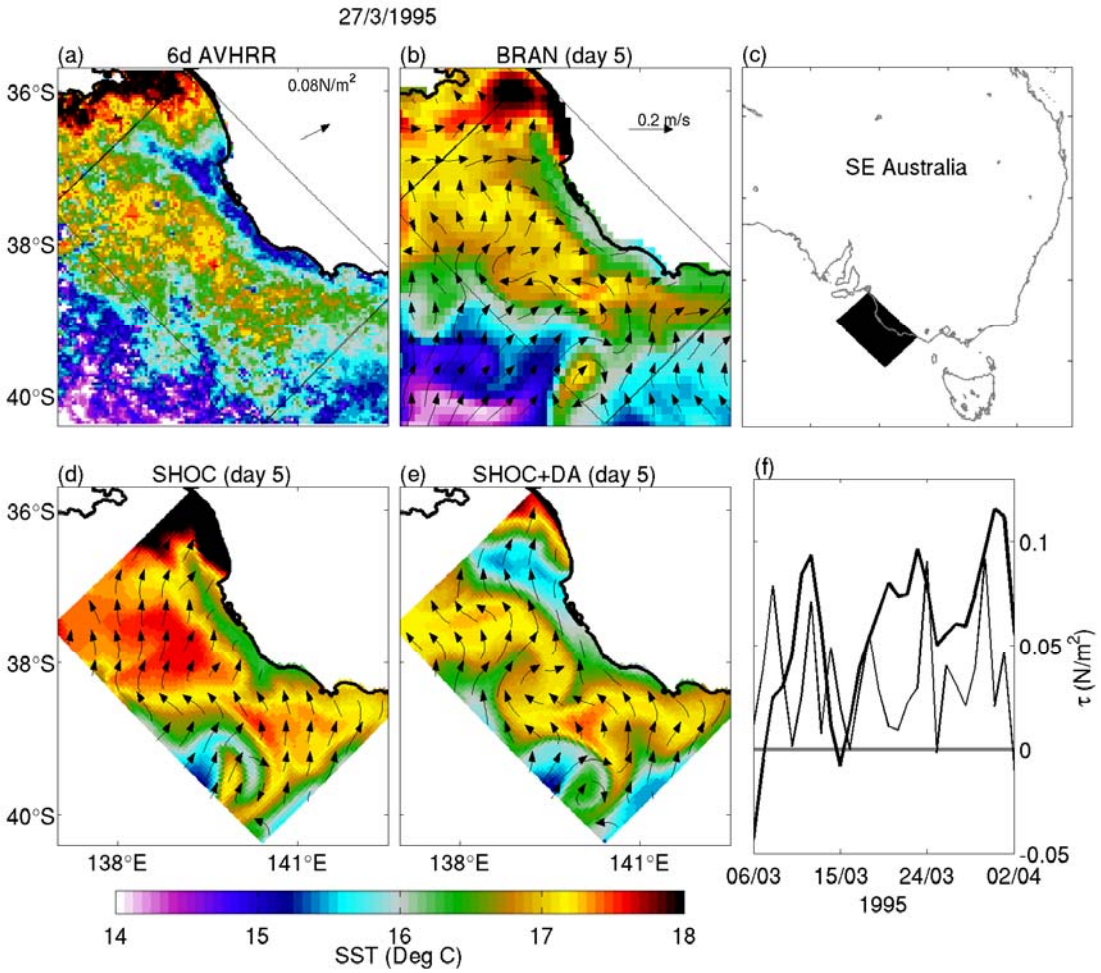


Figure 4

An example of SST from (a) 6-d composite AVHRR, (b) daily mean BRAN (version 2p1), (d) daily mean SHOC, and (e) daily mean SHOC plus data assimilation. The model fields are valid 5 days after initialisation. The arrow in panel (a) shows the daily mean wind stress along with the magnitude. The region of the SHOC domain is shown in panel (c) and the time series of zonal (bold) and meridional (thin) wind stress is plotted in panel (f). The arrows in panels (b, d, and e) show the daily mean surface velocities

The model fields are compared to 1-day composite AVHRR SST observations across 8 consecutive 7-day forecasts. These statistics are summarized in Figure 5, showing the root-mean-squared difference (RMSD) fields, presented as a function of the forecast lead time. The forecast lead time is negative during the spin up period and positive during the forecast period. Figure 5 shows that the RMSD is greatest for SHOC without assimilation, and is smallest for SHOC+DA. The difference between these runs is greatest during the spin-up period, when SST data are assimilated, but remains significant out to 7-day forecasts.

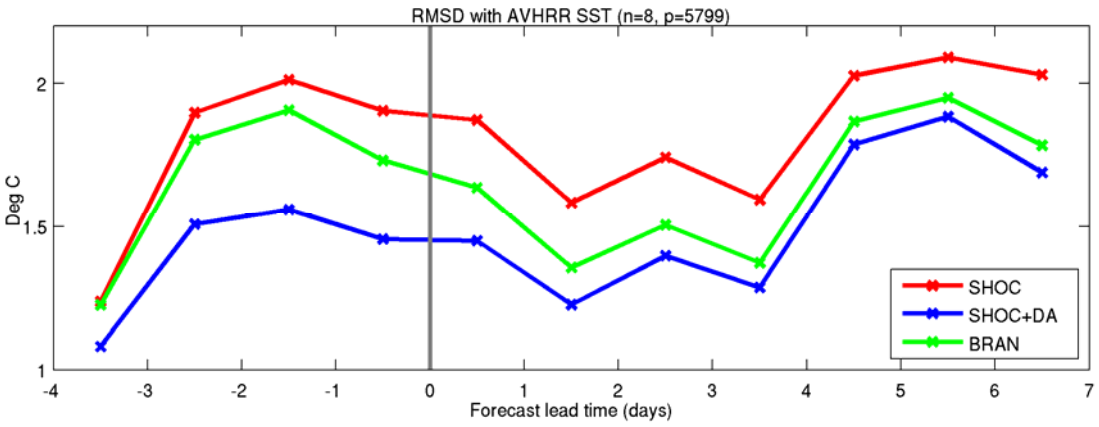


Figure 5

Root-mean-squared difference between 1-day composite AVHRR SST and daily mean SST from SHOC (no assimilation), SHOC+DA (with assimilation of AMSR-E SST), and BRAN (version 2p1)

Because SHOC has higher resolution than BRAN, one might expect SHOC, even without data assimilation, to out-perform BRAN. However, this is not the case here. This is probably because the difference in resolution is only small (5 km compared to 10 km). SHOC also has an additional source of error through the open boundaries. The fields are well behaved at the boundaries (Figure 4) with incoming features retaining their structure and out-going features leaving the domain with no obvious artifacts. However, the boundaries certainly remain a source of error.

Despite the main event considered here, a wind-driven upwelling, being due to surface forcing, rather than initialization, there is still a significant benefit of updating the initial conditions and boundary fields to better match reality. This is one demonstration of the benefit of data assimilation in coastal models.

Conclusion

BODAS was initially developed under Bluelink for global ocean data assimilation. BODAS was initially developed for short-range forecasting of the mesoscale circulation in the open ocean. But since its development, BODAS has also been incorporated into the operational Bluelink forecast system, run at the BoM, and has routinely been used for reanalysis experiments. Many aspects of the Bluelink system have been improved as a result of the BRAN experiments and the system has demonstrated measureable improvements over the lifetime of Bluelink. In addition to global data assimilation, BODAS has also been used for observing system evaluation, observing system design, and coastal data assimilation. Development of BODAS is ongoing. Specific challenges ahead include the application of BODAS to a global 1/10^{o} model that is planned for Bluelink. Better use of observations is also an important ongoing challenge and the problem of model initialisation remains an issue. Research in these areas continues under Bluelink

Acknowledgements

Financial support for this research is provided by CSIRO, the Bureau of Meteorology, and the Royal Australian Navy as part of the Bluelink project, and the US Office of Naval Research (Grant No. N00014-07-1-0422). The author also acknowledges the many contributions of the Bluelink science team. Satellite altimetry is provided by NASA, NOAA, ESA and CNES. Drifter data are provided by NOAA-AOML and SST observations are provided by NASA, NOAA and Remote Sensing Systems.

References

Bertino, L., and Lisaeter, K.A. 2008: The TOPAZ monitoring and prediction system for the Atlantic and Arctic oceans, *Journal of Operational Oceanography,* **1**, 15–19, 2008.

Bloom, S.C., Takas, L.L., Da Silva, A.M., Ledvina, D. 1996: Data assimilation using incremental analysis updates. *Monthly Weather Review,* **124**, 1256–1271.

Brassington, G.B., Pugh, T., Spillman, C., Schulz, E., Beggs, H., Schiller, A., Oke, P.R. 2007: BLUElink> development of operational oceanography and servicing in Australia. *Journal of Research Practice Information Technology,* **39**, 151–164.

Brasseur, P., Bahurel, P., Bertino, L., Birol, F., Brankart, J.-M., Ferry, N., Losa, S., Remy, E., Schrter, J., Skachko, S., Testut, C.-E., Tranchant, van Leeuwen, P.-J., Verron, J. 2006: Data assimilation in operational ocean forecasting systems: the MERCATOR and MERSEA developments. *Quarterly Journal of the Royal Meterological Society,* **131**, 3561-3582.

Chen, D., Rothstein, L.M., Busalacchi, A.J. 1994: A hybrid vertical mixing scheme and its application to tropical ocean models. Journal of Physical Oceanography, **24**, 2156-2179.

Feng, M., Majewski, L.J., Fandry, C.B., and Waite, A.M. 2007: Characteristics of two counter-rotating eddies in the Leeuwin Current system off the Western Australian coast. *Deep-Sea Research, **54***, 961–980.

Griffies, S.M., Hallberg, R.W. 2000: Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models. *Monthly Weather Review*, **128**, 2935-2946.

Griffies, S.M., Pacanowski, R.C., Rosati, A. 2004: A technical guide to MOM4. GFDL Ocean Group Technical Report No. 5. NOAA/Geophysical Fluid Dynamics Laboratory, 371pp.

Griffin, D.A., Thompson, P.A., Bax, N.J., Bradford, R.W., Hallegraff, G.M. 1997: The 1995 mass mortality of pilchard: no role found for physical or biological oceanographic factors in Australia, *Marine and Freshwater Research*, **48**, 27-42.

Herzfeld, M. 2009: Improving stability of regional numerical ocean models. Ocean Dynamics **59**, 21-46.

Houtekamer, P.L.,Mitchell, H.L. 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review,* **129**, 123–137.

Kallberg, P., Simmons, A., Uppala, S., Fuentes, M., 2004: The ERA-40 archive, European Centre for Medium-range Weather Forecasts (ECMWF), ECMWF Re-Analysis Project (ERA), ERA-40 Project Report Series 17, 31 pp.

Leonard, B.P. 1979: A stable and accurate convective modelling procedure based on quadratic upstream interpolation. *Computational Methods of Applied Mechanical Engineering,* **19**, 59–98.

Mitchell, H.L., Houtekamer, P.L., and Pellerin, G. 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Monthly Weather Review,* **130**, 2791–2808.

Oke, P.R., Schiller, A., Griffin, D.A., Brassington, G.B. 2005: Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. *Quarterly Journal of the Royal Meteorological Society,* **131**, 3301–3311.

Oke, P. R., P. Sakov, and S. P. Corney 2006: Impacts of localisation in the EnKF and EnOI: Experiments with a small model. *Ocean Dynamics,* **57**, 32-45.

Oke, P.R., and Schiller, A. 2007: Impact of Argo, SST and altimeter data on an eddy-resolving ocean reanalysis. *Geophysical Research Letters,* **34**, L19601, doi:10.1029/2007GL031549.

Oke, P.R., Brassington, G.B., Griffin, D.A. and Schiller, A. 2008: The Bluelink Ocean Data Assimilation System (BODAS), *Ocean Modelling,* **21**, 46-70, doi:10.1016/j.ocemod.2007. 11.002.

Oke, P. R., and P. Sakov 2008: Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology,* **25**, 1004-1017.

Oke, P. R., P. Sakov and E. Schulz, 2009: A comparison of shelf observation platforms for assimilation into an eddy-resolving ocean model. Dynamics of Atmospheres and Oceans, doi:10.1016/j.dynatmoce.2009.04.002, in press.

Rawlins, F., Ballard, S.P., Bovis, K.J., Clayton, A.M., Li, D., Inverarity, G.W., Lorenc, A.C., and Payne T.J. 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society,* **133**, 347–362.

Sandery, P.A., and Brassington, G.B. 2008: Preliminary evaluation of a coupled ocean-atmosphere prediction system. *CAWCR Research Letters,* **1**, 19-23.

Schiller, A., Oke, P.R., Brassington, G.B., Entel, M., Fiedler, R., Griffin, D.A., and Mansbridge, J.V. 2008: Eddy-resolving ocean circulation in the Asian-Australian region inferred from an ocean reanalysis effort. *Progress in Oceanography,* **76**, 334-365.

Schulz, E.W., Kepert, J.D. and Greenslade, D.J.M. 2007: An Assessment of Marine Surface Winds from the Australian Bureau of Meteorology Numerical Weather Prediction Systems, *Weather and Forecasting*, **22**, 613-636.

Smith, W.H.F., and Sandwell, D.T. 1997: Global seafloor topography from satellite altimetry and ship depth soundings, *Science*, **277**, 1957-1962.

Notebook

Editorial Board:

Laurence Crosnier

Secretary:

Monique Gasc

Articles:

Contact :

Please send us your comments to the following e-mail address: webmaster@mercator-ocean.fr

Next issue: October 2009