Data assimilation for coastal zone monitoring and forecasting

Geir Evensen and Helge Drange [a]

[a]Nansen Environmental and Remote Sensing Center,
Edvard Griegsvei 3a
N-5037 Solheimsviken
Norway

In this note we attempt to identify the main developments in data assimilation and ecosystem modelling that must be made in the next few years to build an efficient coastal zone monitoring and prediction system. By the coastal zone we mean the oceans along and sometimes outside the continental margins which may be of particular interest for comercial utilization, say, within the fishing and/or oil industry. As an example this includes most of the Nordic Seas and the Mediterranean but excludes the major part of the Atlantic basin. Various candidates for the data assimilation methodologies are discussed in addition to presenting the status of current data assimilation systems for Ocean General Circulation Models (OGCMs). The use of data assimilation methods for models of the marine ecosystem is far less developed and a significant effort needs to be invested to implement and examine various assimilation techniques with such models. In addition, there is an urgent need for making observations of ecosystem variables available on a regular basis. The high spatial and temporal coverage which is needed in data assimilation suggest that remotely sensed observations will be crucial, e.g. from ocean color sensors.

1. Introduction

The need for better monitoring and modelling of the marine environment has increased dramatically in recent years, especially along coastal boundaries and shelf regions where human activities are extensive and pollution has a significant impact. Prediction of natural hazards, preservation of marine life and commercial utilization of resources like oil, gas, minerals, hydrothermal energy and marine food would benefit from an operational coastal zone monitoring and prediction system. This has been clearly demonstrated by a number of unpredicted events over the last few years, including storm surges, harmful algal blooms and oil spills.

A future operational coastal ocean and environmental monitoring and forecasting system will provide estimates of variables of both the physical and the biogeochemical marine environment. The system will enable early warning and execution of cost effective precautions in the case of potential harmful events. In addition, a good understanding of the processes in the marine ecosystem is of great importance for resource management. The potential for marine monitoring systems has been pointed out in several publications

[17,18,15], and is identified as an area of great importance within EuroGOOS.

Some of the most important variables to predict are those which are related to the coastal zone and open ocean ecosystems, including the temporal and spatial distribution of planktonic biomass and plant nutrients. Knowledge of these variables is needed in order to assess the response of the marine ecosystem to various anthropogenic activities, to predict the water quality, to estimate new and regenerated marine production, and to understand the coupling between the physics and ecosystem dynamics in the marine cycling of nutrients and carbon.

A coastal ocean monitoring system will have to be built on methodologies for efficient integration of observations and numerical models. Oceanic *in situ* observations are sparse in space and time and, thus, the huge amount of remote sensing data provided by the satellites observing the environment will play a key role in an operational system even though they only provide information from the ocean surface. On the other hand, ocean circulation models do hold information about the physical processes which govern the general ocean circulation, although they must be used together with information from observations to give a realistic description of the real world. Such integrated use of observations and models is best done using data assimilation methods, which, in an optimal way, merges the information about the dynamics contained in a model with the information about the current state of a system contained in a set of measurements.

Recent developments in ecosystem modelling and data assimilation methodologies will together with the new satellite observation systems provide a basis for the implementation of an operational ocean monitoring and forecasting system which focuses on the coastal zone ocean and ecosystem dynamics. The time frame of such a system is expected to be about four to five years, allowing time for the implementation and validation of data assimilation systems for coupled primitive equation and ecosystem models and also for building the framework of an operational system.

2. Data assimilation in OGCMs

The currently available data assimilation applications for OGCMs are based on rather simplistic assimilation schemes. By that is meant that none of these takes proper error statistics into account and *ad hoc* approaches are used for the assimilation. Thus, even if one now has a good understanding of how to formulate and solve the inverse or data assimilation problem, only a few simplistic approaches exist for realistic primitive equation models. This is due to the strong nonlinearities of the mesoscale ocean dynamics and the huge numerical load associated with such systems.

Examples of existing data assimilation implementations with OGCMs are given by Derber and Rosati [4], where an objective analysis technique was used to update the model temperature in a version of the Cox model [24]; in Ezer and Mellor [13] where a univariate optimal interpolation algorithm was used with vertical projection of surface information in the Blumberg and Mellor model [1]; by Cooper and Haines [2] who used the the vertical projection method based on water property conservation in the Cox model [3]; and by Malanotte-Rizzoli and Young [20], where a nudging technique where used in experiments for the Gulf Stream in a semi-spectral model.

The fundamental problems related to formulation and solution of the general assimilation problem are now well understood and significant progress has been made during the last two or three years in developing advanced data assimilation systems which handle strong nonlinearities at an affordable numerical cost, and where proper error statistics are used in the analysis step. These methods have been implemented and validated with less complicated models and there is now a significant ongoing effort, worldwide, in implementing and validating these more advanced data assimilation systems for OGCMs.

It is expected that so-called sequential methods will be the most efficient for operational data assimilation in OGCMs. Relatively simple versions of such methods, i.e. various versions of Optimal Interpolation (OI), have been used in the atmospheric community for decades [14]. The method is based on an assumption of known error statistics for the model forecast and the measurements at each particular time where measurements are available. Given the model forecast and the measurements with specified error covariances, a variance minimizing analysis estimate is calculated and used to reinitialize the model for the further integration until the next time when observations are available.

OI is a rather simple assimilation scheme and the efficiency and accuracy of the results depend crucially on the quality of the specified error covariances which determine the influence an observation will have on the model state. A data assimilation system based on OI is probably the simplest methodology that can provide reliable results when used with OGCMs.

More advanced data assimilation techniques apply time dependent and dynamically consistent error statistics. This requires the forward integration of an error covariance equation for the error statistics, e.g. by using an Extended Kalman Filter (EKF) [6,7], or, as a better alternative, one can integrate an ensemble of ocean states as is done in the recently proposed Ensemble Kalman Filter (EnKF) [8,9,12].

The recent developments related to so-called advanced methods like the EnKF, and the significant improvement of available computer resources, now suggest that such advanced methods should be implemented also with OGCMs. These methodologies have proven very successful when used with less complicated, but still highly nonlinear, dynamical models and there is a significant ongoing effort in implementing such advanced data assimilation methods with OGCMs.

The EnKF is essentially a Monte Carlo method for predicting error statistics where an ensemble of ocean states is integrated forward in time, and the error statistics which are needed to perform a variance minimizing analysis can be calculated from the ensemble. A clever analysis scheme provides both an analyzed estimate and a reinitialized ensemble with the correct analyzed covariance after measurements have been assimilated. In the limit of an infinite ensemble size this method can be characterized as the optimal variance minimizing sequential method for nonlinear dynamics. The method provides statistical error estimates for the analysis without additional computations. The method completely overcomes the major problems reported for the Kalman filter when used with nonlinear dynamics. That is, there are no closure problems associated with the forward integration of error statistics, and if the ocean model can apply open boundaries this is also true for the EnKF. The numerical load has also been significantly reduced compared to the standard Kalman filter, and the method can now be applied for realistic domains and resolution on extant computer resources. The numerical cost corresponds to 100–500 forward model integrations.

The method has recently been applied with a multilayer quasi-geostrophic model for the Agulhas retroflection area, where Geosat altimetry was assimilated in a study of the Agulhas eddy shedding process [12].

3. Observations to be assimilated in OGCMs

The most important observations to be used in an operational system will be Radar Altimeter data, e.g. from TOPEX/POSEIDON and ERS–2, and sea surface temperatures, e.g. from the ERS–2 ATSR. These data are already available for use in preoperational data assimilation systems, however in a fully operational system the access time for the most recent observations becomes important. Probably such observations should be distributed on a daily basis. Further, real time analyses and predictions from the European weather services must be used for ensuring a proper forcing of the model and for making it possible to generate realistic predictions of the marine system.

4. Status of present ecosystem models

Several prognostic ecosystem models have been developed over the last decade in order to describe the cycling of nutrients and carbon in the marine environment. The model state variables typically consist of 1–3 groups of phytoplankton organisms, 1–2 groups of zooplankton organisms, bacteria, 2–3 nutrients, total dissolved inorganic carbon and total alkalinity, and dissolved and particulate organic matter. Although the marine ecosystem is very complex, including a high degree of temporal and spatial variability, modelling experiments show that the major features of the marine ecosystem dynamics are reasonably well understood [25,5]. State-of-art ecosystem models, combined with information about the physical-biogeochemical state of the ocean, can therefore be used in a monitoring and forecasting system.

5. Data assimilation in ecosystem models

There are only a few publications available on data assimilation in ecosystem models (e.g. [16,21,23,22]). Thus it is natural to start with examining data assimilation methods for zero dimensional models (where variables are integrated in the vertical) to see how the assimilation methods are capable of retrieving the observed variability. The next steps are then first an extension to a 1–dimensional model where the vertical is resolved, and finally to the full 3–dimensional model. Initially, a relatively simple ecosystem model should be used in the development of the data assimilation systems. One argument for working with relatively simple models in the development phase is that there are currently not enough observations available to constrain all of the variables in a multi-compartment ecosystem model. The data assimilation methods should be developed in a rather general context to be easily adapted to new and more advanced ecosystem models as such models develop in the future.

Other data assimilation techniques than those used with OGCMs should be examined for the ecosystem dynamics because of the vastly different mathematical properties of an ecosystem model compared to OGCMs (e.g. no nonlinear advection term). Thus, in addition to the EnKF discussed for the OGCMs, one should also consider so-called variational methods. One such candidate is the weak constraint gradient descent solver [10],[11]. In a weak constraint variational formulation one allows the model dynamics to contain errors and attempts to find a solution which is close to the observations and at the same time “almost” satisfies the model dynamics. “Close” is defined in some sense, normally by minimizing the squares of the residuals between the estimate and the observations and the model dynamics. The method has proven very successful with strongly nonlinear dynamics, and has proven superior to other advanced methods in one particular example with the Lorenz model [19] since the method seeks the maximum likelihood solution independent of the nonlinearities of the model. Another strength is that the method does not require any integration of the model equations since a model solution in space and time is substituted in each iteration.

6. A coastal ocean monitoring and forecasting system

An accurate prediction of the coastal ecosystem will rely strongly on the quality of available estimates of ocean currents and mixing processes. Thus, reliable primitive equations ocean circulation models must be used in combination with data assimilation systems to provide the physical fields that advect and mix the ecosystem and nutrient variables. Further, inverse calculations or data assimilation systems must be used also with the ecosystem models to take advantage of information from observations about the state of the coupled physical-biogeochemical system.

The observing systems will have to consist of both *in situ* and remotely sensed information. Information about geostrophic velocities are available from altimeter data, the sea surface temperature may be estimated from IR images, and information about chlorophyll *a* concentration can be determined from ocean color sensors. *In situ* information is required for calibration of the satellite sensors, to add more model variables to the data assimilation schemes, and for extracting information on sub-surface variables.

An important property of a coupled physical and ecosystem model is that there exists to the lowest order only a one way coupling from physical variables to ecosystem variables. Thus one can first solve the data assimilation problem for the OGCM and then use the analyzed advective velocities, mixed layer parameters and thermodynamic variables as input to the data assimilation system for the ecosystem model.

The major issue for the ecosystem assimilation problem is the lack of useful observations of biogeochemical variables. The only satellite sensors available yet that may provide useful information are the Coastal Zone Color Scanner (CZCS) and also the recently launched Ocean Color Temperature Scanner (OCTS) onboard the ADEOS satellite. Future planned remotely sensed ocean color observations will be those collected by SEAWIFS and MERIS In addition to ocean color sensors, regular surface and sub-surface *in situ* nutrient and biomass observations are needed in order to properly constrain the ecosystem model in a data assimilation context.

REFERENCES

1. A. F. Blumberg and G. L. Mellor, A description of a three-dimensional coastal ocean circulation model, in *Three-Dimensional Coastal Ocean Models*, edited by N. Heaps, pp. 1–16, American Geophysical Union, Washington, DC, 1987.

2. M. Cooper and K. Haines, Altimetric assimilation with property conservation, *J. Geophys. Res.*, *101*, 1059–1077, 1996.
3. M. D. Cox, An eddy-resolving numerical model of the ventilated thermocline: Time dependence, *J. Phys. Oceanogr.*, *17*, 1044–1056, 1987.
4. J. Derber and A. Rosati, A global oceanic data assimilation system, *J. Phys. Oceanogr.*, *19*, 1333–1347, 1989.
5. H. Drange, An isopycnic coordinate model of the seasonal cycling of carbon and nitrogen in the Atlantic Ocean, *Physics and Chemistry of the Earth*, 1996, Submitted.
6. G. Evensen, Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.*, *97*(C11), 17,905–17,924, 1992.
7. G. Evensen, Open boundary conditions for the extended Kalman filter with a quasi-geostrophic model, *J. Geophys. Res.*, *98*(C9), 16,529–16,546, 1993.
8. G. Evensen, Inverse methods and data assimilation in nonlinear ocean models, *Physica D*, *77*, 108–129, 1994.
9. G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162, 1994.
10. G. Evensen, Advanced data assimilation for strongly nonlinear dynamics, *Mon. Weather Rev.*, *125*, 1342–1354, 1997.
11. G. Evensen and N. Fario, A weak constraint variational inverse for the Lorenz equations using substitution methods, *J. Meteor. Soc. Japan*, 1996, In print.
12. G. Evensen and P. J. van Leeuwen, Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model, *Mon. Weather Rev.*, *124*, 85–96, 1996.
13. T. Ezer and G. L. Mellor, Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf stream model, *J. Phys. Oceanogr.*, *24*, 832–847, 1994.
14. M. Ghil and P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.*, *33*, 141–266, 1991.
15. P. M. Haugan, G. Evensen, J. A. Johannessen, O. M. Johannessen, and L. Pettersson, Modeled and observed mesoscale circulation and wave-current refraction during the 1988 Norwegian continental shelf experiment, *J. Geophys. Res.*, *96*(C6), 10,487–10,506, 1991.
16. J. Ishizaka, Data assimilation for biogeochemical models, in *Towards a Model of Ocean Biochemical Processes*, edited by G. T. Evans and M. J. R. Fasham, pp. 295–316, Berlin, 1993, Springer-Verlag.
17. J. A. Johannessen, L. P. Røed, O. M. Johannessen, G. Evensen, B. Hackett, L. H. Petterson, P. M. Haugan, S. Sandven, and R. Shuchman, Monitoring and modeling of the marine coastal environment, *Photogrametric Eng. and Remote Sensing*, *59*(3), 351–361, 1993.
18. J. A. Johannessen, P. W. Vachon, and O. M. Johannessen, ERS–1 SAR imaging of marine boundary layer processes, *Earth Observation Quarterly, ESA*, (46), 1–5, 1994.
19. E. N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.*, *20*, 130–141, 1963.
20. P. Malanotte-Rizzoli and R. E. Young, Assimilation of global versus local data sets into a regional model of the Gulf Stream system: 1. Data effectiveness, *J. Geophys.*

Res., *100*, 24,773–24,796, 1995.

21. R. J. Matear, Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P, *J. Marine. Res.*, *53*, 571–607, 1995.
22. P. Prunet, J.-F. Minster, V. Echevin, and I. Dadou, Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean (2). Adjusting a simple trophic model to chlorophyll, temperature, nitrate and pCO_2 data, *Global Biogeochemical Cycles*, *10*(1), 139–158, 1996.
23. P. Prunet, J.-F. Minster, D. Ruiz-Pino, and I. Dadou, Assimilation of surface data in a one-dimensional physical-biogeochemical model of the surface ocean (1). Method and preliminary results, *Global Biogeochemical Cycles*, *10*(1), 111–138, 1996.
24. A. Rosati and K. Miyakoda, A GCM for upper ocean simulation, *J. Phys. Oceanogr.*, *18*, 1601–1626, 1988.
25. J. L. Sarmiento, R. D. Slater, M. J. R. Fasham, H. W. Ducklow, J. R. Toggweiler, and G. T. Evans, A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone, *Global Biogeoch. Cycles*, *7*, 417–450, 1993.