Development of Operational Data-Assimilating Water Quality Modelling System for South-East Tasmania

Margvelashvili, N.[1], J.S. Parslow[1], M. Herzfeld[1], K. Wild-Allen[1], J. Andrewartha[1], F. Rizwi[1], E. Jones[1]

[1] *CSIRO Marine and Atmospheric Research, GPO Box 1538, HOBART TAS 7001, Australia*
Email: nugzar.margvelashvili@csiro.au

***Abstract*-With the rapid advances in on-line observing system applications, the paradigm in environmental modelling is shifting from one-off models for specific purposes, to operational models, sequentially assimilating data streams from in situ and remote sensors. Such models can provide products and services to support a wide range of applications, from short-term forecasting to long-term scenarios, and are expected to deliver superior performance much more cost-effectively. In the marine field, this is most advanced for circulation models at large ocean scales. The potential benefit from these advances is even greater in the coastal zone, where human uses, impacts and ecosystem services are concentrated. However, there are substantial challenges to be overcome. Coastal applications typically require biogeochemical, ecological, and ultimately socioeconomic models. These additional models are more complex, with higher uncertainty, and require different approaches to data assimilation and uncertainty analysis. The uncertainties arise from a number of sources including poorly known parameters, structural errors and stochastic forcing. When model realisations are sufficiently fast, Monte Carlo techniques can be used to improve the model performance and assess its quality, otherwise alternative estimation techniques are required.**

This paper describes the development of an operational, data-assimilating coastal model for SE Tasmania, integrating across hydrodynamics, sediment dynamics and biogeochemistry. Inputs and outputs from the model are expected to be integrated into the regional information system (INFORMD), and to be used directly in multiple management applications, and as input into ecosystem models. A hydrodynamic model, nested inside an operational global model, will be assimilating data from the coastal sensor network and other sources, including remote sensing. The model is based on an operational modelling platform developed by CSIRO through the BlueLink project (ROAM), and will be used to implement and test data-assimilation techniques for coastal models under development in BlueLink. Operational sediment dynamic and biogeochemical models, will be coupled to the hydrodynamic model, either directly or through intermediate transport models. Data-assimilating techniques for these models currently are under development in Computational and Simulation Sciences theme, CSIRO. This paper outlines preliminary results from these developments. A number of candidate techniques including Kalman Filter, Particle Filter and MCMC are discussed. The utility of fast and cheap statistical surrogates of complex models (emulators) for sequential data assimilation is illustrated through the trial application of emulators to one-dimensional sediment/pollutant and 3-d sediment transport models.

INTRODUCTION

The South-East Tasmanian (SETAS) coast is a rapidly developing region, containing some of the fastest growing municipalities in Australia. The coastal ecosystems directly support a wide range of human activities including aquaculture, fisheries, recreation and tourism. They also provide indirect support for numerous industries, urban centres and rural land-use practices through the provision of 'ecosystem services'. The region encompasses the Derwent-Huon estuaries and coastal ecosystems, and can be considered as a microcosm of the issues facing coastal development globally and elsewhere in Australia.

Management of the increasing range and intensity of the human use of coastal ecosystems is a recognised problem in SETAS. Major bottlenecks come from limitations in the kind and quantity of observations, and from the lack of reliable interpretations of the data. Coastal managers require cost effective, versatile decision support tools integrating observations with models. To meet these requirements, CSIRO Wealth from Oceans (WfO) project "Linking models and sensors in the Derwent region" will develop and implement operational data-assimilating modelling system for SE Tasmania, integrating across hydrodynamics, sediment dynamics, and biogeochemistry, and linked to sensor networks and other real time data sources. The project is part of INFORMD umbrella program, a collaborative program between the Tasmanian government, University of Tasmania and CSIRO to support sustainable development of the Derwent-Huon coastal ecosystem. The program is anticipated to deliver a new 'state of the art' observation network, information management system, new models for data interpretation and prediction, and support tools to deliver knowledge to industry, planners, policy developers and managers.

978-1-4244-5222-4/10/$26.00 ©2010 IEEE

This paper outlines a strategy for the operational modeling system for SE Tasmania comprising (a) the development and implementation of the Environmental Modelling Suite (EMS), a backbone of the operational system, (b) set-up and acquisition of on-line data-streams and (c) development of capabilities for sequential assimilation of these data into the integrated modeling system. Preliminary results from operational runs of the hydrodynamic model are illustrated and strategies for the development of operational data-assimilating biogeochemical models are discussed.

I. Operational Models

The operational model is based on EMS developed by CSIRO over twenty years of research and applications to Australian coastal waters. EMS consists of a set of component models which can be used separately or together. The hydrodynamic model provides a versatile tool for predicting 3-d advection, mixing and tracer transport in coastal ecosystems. The sediment model adds a layered sediment bed, and simulates the physical exchange of particulate and dissolved tracers between the water column and the bed. The biogeochemical module provides a sophisticated representation of the cycling of carbon and nutrients through coupled pelagic-benthic ecosystems.

Hydrodynamics

SHOC (Sparse Hydrodynamic Ocean Code) is a general purpose 3 dimensional, finite-difference hydrodynamic model applicable on spatial scales ranging from estuaries to regional ocean domains [1]. It is based on the equations of momentum, continuity and conservation of heat and salt, employing the hydrostatic and Boussinesq assumptions. The model has a free surface and uses mode splitting to separate the 2d mode from the 3d mode. Several turbulence closure schemes including k-e, and Mellor-Yamada 2.0, 2.5, can be invoked to simulate vertical mixing in a turbulent flow and a variety of advection schemes can be applied to tracers.

An operational SETAS model is implemented on a curvilinear grid with the resolution ranging from 250m in the estuaries to 2.5 km at offshore boundaries. The model is nested into the global ocean operational model OceanMAPS: (www.bom.gov.au/bluelink/products/prod_oceanmaps.html), which resolves the Australasian region at 10 km, and supplies sea level, temperature and salinity on the open sea boundaries. At the river boundaries the model is forced with river flow from the Huon and Derwent Rivers, and it uses MesoLAPS atmospheric products for the surface flux specification. A prototype operational model has been successfully implemented in near real-time (www.emg.cmar.csiro.au). A higher resolution model, capable of supplying output at ~100 m resolution within the estuaries and increasing to ~2 km offshore, has been nested within this larger scale model.

Wave model

Operational wave forecast is obtained from the Spectral Wave Model Simulations (SWAN). SWAN solves equations for the directional wave spectrum on a regular grid based on local wind input, wave dissipation, nonlinear wave-wave interactions and propagation of waves from non-local sources. The model predicts significant wave height, period and direction of the surface gravity waves (locally generated wind-waves and swell with a period of 1-20 seconds). The wind fields for the wave model are derived from the Bureau of Meteorology (BOM) MesoLAPS model predictions. The open ocean boundary conditions are obtained from the JFREMER global wave model Wavewatch3 (WW3).

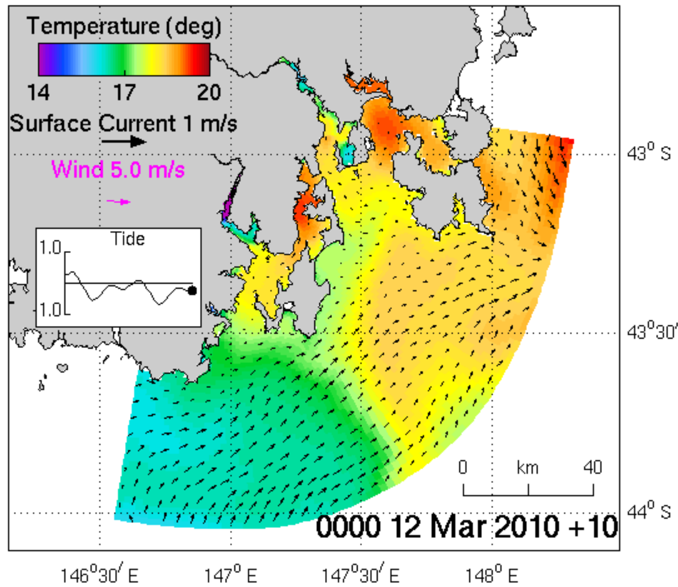

Figure 1. A snapshot of simulated SST and currents in SETAS

Sediment model

One of the key objectives of the operational hydrodynamic/ wave models is to provide near-real time forcing for the sediment and biogeochemical models. The sediment model of EMS adds a multilayer sediment bed to the EMS grid and simulates sinking, deposition and resuspension of multiply size-classes of suspended sediment [2]. The model solves advection-diffusion equations of the mass conservation of suspended and bottom sediments and is particularly suitable for representing fine sediment dynamics, including resuspension and transport of biogeochemical particles.

Operationalization of the sediment model requires near-real-time access to data representing sediment loads from catchments, discharge from point sources (eg storm-water outfalls) and the sediment concentrations at the open ocean boundaries. Field data, remote sensing, and estimates based on the scaling relationships between river flow and sediment concentrations will be used to specify such data sets.

Biogeochemistry

The biogeochemical module of EMS has evolved through a series of case studies around Australian coast including Port Phillip Bay Environmental Study [3], Gippsland Lakes Environmental Study [4], Ord-Bonaparte study [5]. Each study addressed specific environment and ecological questions resulting in the development, implementation and testing of a diverse range of model components. The model domain is organized in 3 'zones': pelagic, epibenthis and sediments. Dissolved nutrients are advected and diffused throughout the model domain in an identical fashion to temperature and salinity while particulate substances sink and are resuspended in the same way as sediment particles. For each layer and grid location the time evolution of each model substance is the sum of conservative advection, diffusion and sinking processes and non-conservative bio-geochemical rate processes. Non-conservative biogeochemical processes are organized into pelagic processes of phytoplankton and zooplankton growth and mortality, detritus remineralisation, and fluxes of dissolved oxygen, nitrogen, and phosphorus; epibenthic processes of growth and mortality of macroalgae and seagrass, and sediment based processes of phytoplankton mortality, microphytobenthos growth, detrital remineralisation, and fluxes of dissolved substances. Further details of the biogeochemical model, including model equations and parameter values, are included in [6].

Ground observations, remote sensing and climatology data will be used to specify boundary conditions and point sources for the operational biogeochemical model.

Transport solutions

Both sediment and biogeochemical models can run in a mode fully coupled to the hydrodynamic model and in a transport mode when the sediment and biogeochemical transport is simulated off-line, using precalculated velocities and diffusion coefficients. The transport model uses semi-Lagrangian techniques to allow long time steps, and has been shown to run up to 100 times faster than the fully coupled Eulerian model. While some accuracy is being forfeited because of non-conservative nature of the Lagrangian schemes, numerical experiments show generally good agreement between fully coupled Eulerian and off-line Lagrangian models.

II. DATA-STREAMS

Access to high quality on-line data streams is critical for the successful operation of the near-real-time modeling system. The data sets are required to drive models, and test and improve their predictions. Conventional monitoring programs based on monthly samples provide valuable insights into the long-term trends of the ecosystem dynamics but typically have poor spatial and temporal resolution hindering their interpretation over shorter time/space scales. Remote sensing products tend to have high spatial resolution of surface layers but may have limited resolution in time. Over the last years novel observation techniques, integrating a relatively high spatial coverage with the high temporal resolution, such as AUVs and sensor networks, have been gaining momentum. Within the INFORMD program a number of such novel techniques are being explored, including deployment of the Slocium Glider in Storm Bay, the development of a network of a relatively cheap sensors (thermistor strings) for the SETAS, and a trial deployment and evaluation of novel sensors for in-situ measurement of the nutrient concentrations (ISUS). These experimental developments are complemented with a number of more conventional deployments of stationary mooring equipped with the reliable high-quality sensors that will provide a test bed for the experimental facilities.

III. DATA-MODEL FUSION

Aforementioned observational platforms have the potential to deliver data sets with the resolution and quality unprecedented earlier in SETAS. Interpretation of such data is not trivial because of high dimensionality of observations representing wide range of spatial and temporal scales of the ecosystem dynamics. Numerical data-assimilating models are irreplaceable for evaluating and interpreting such data streams. The challenge is to devise and implement data-assimilating schemes that would scale with the data dimensionality and cope with the computational constraints and complexity the numerical models itself.

Ensemble Kalman Filters (EnKF) have been very successful in recovering state variables of complex space-time resolving ocean models using limited number of model runs [7]. The reduced version of EnKF based on optimal interpolation has been developed by CSIRO in partnership with BoM through the BlueLink project and implemented to global ocean model having 10km resolution around Australian coast [8]. The SETAS hydrodynamic model, nested inside this operational global model, will be assimilating data from the coastal sensor network and other sources, including remote sensing. The model will be based on operational modelling platform developed by CSIRO through the BlueLink project (ROAM), and will be used to implement and test data-assimilation techniques for coastal models under development in BlueLink.

While the focus of this study is on near-real-time products, sediment and biogeochemical models are often required to deliver evaluations of long-term management scenarios rather than short-term forecasts. Such long-term simulations tend to be more sensitive to the specification of the model parameters and processes rather than initial conditions, and are more analogous to climate modeling than ocean forecasting. Applications of EnKF to the problem of the combined parameter-state estimation are much less advanced and documented compared to the state evaluation problem. Another candidate technique, particle filtering, while theoretically more general and appealing, suffers from the sample impoverishment and ensemble collapse particularly with high-dimensional problems.

In what follows we consider novel approach to the data assimilation aiming at the recovery of both state and parameter estimates for fine-resolution ocean models. The method is based on dimension reduction and emulation of the model to allow Bayesian inferences to be applied to computationally-expensive models. The dimension reduction is carried out via the Singular Value Decomposition (SVD) of the model ensemble; the emulation is based on Gaussian Process Models (GPM). The technique can also be considered an approximate Particle Filter with two classes of particles: model-particles representing an ensemble of computationally expensive model solutions, and emulator-particles representing an ensemble of fast and cheap model approximations.

An outline of the assimilation strategy is as follows: assume that at time t0 posterior samples of initial states and parameters are known. An ensemble of models is initialized with these samples and propagated forward in time until new data sets become available (say at time t1). Having known ensemble predictions, the model state variables are decomposed into a set of basis functions at time t0 and time t1. Decomposition coefficients at time t0 are augmented with the model parameters and mapped to decomposition coefficients at time t1 using Gaussian process models [9]. Each GPM predicts decomposition coefficient at time t1 as a function of the model parameters and decomposition coefficients at time t0. Combined with the basis-function decomposition, GPMs provide fast and cheap approximation of the complex model, called emulator. The emulator is used as a model substitute during MCMC sampling; the sampling strategy in general follows that of Dowd [10]. A more detailed description of the technique can be found in [11].

The data-assimilation algorithm has been tested through the assimilation of synthetic data sets into a one-dimensional model of sediment/pollutant transport in an idealised vertically-resolved benthic-pelagic system. The resuspension and deposition of sediments is driven by tidal velocities superimposed by two moderate flood events. The pollutant is represented by the particulate and dissolved forms of the sediment reactive material. The model has 4 unknown parameters: settling velocity, bottom roughness, and two sorption coefficients regulating pollutant exchange between solid and liquid phases. Synthetic data are given by 2 hourly sediment and contaminant concentrations at the reference level above the seabed. The data are produced by the "twin" model run contaminated with the measurement error.

Figure 2 illustrates estimated state variables vs true values. The ensemble degenerates after day 1 but then recovers from this state and successfully tracks true values within the 95% confidence interval. Numerical experiments, carried out with the different assimilation scenarios (varying size of the ensemble, number of basis functions, observation error etc), show robust performance of the assimilation procedures for a wide range of the assimilation settings.

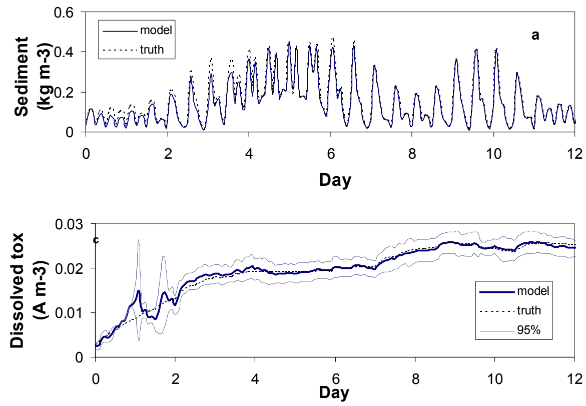

Figure 2. Estimated state variables of 1d sediment/pollutant model vs "truth".

Preliminary evaluations have also been carried out with a 3-d model of sediment transport in SE Tasmania. The model simulates two size-classes of the suspended sediment and assimilates 12 hourly sediment concentrations at the surface produced by the "twin" model run. The assimilation procedure updates 3 model parameters (2 settling velocities and the bottom roughness) and 4 state variables (concentration of 2 sediment fractions in benthic and pelagic layers). The model ensemble consists of 16 members and SVD decompositions are truncated after the first 3 modes. As illustrated in figure 3, the assimilation technique gives reliable estimates of the unknown parameters and concentrations, encouraging further development of the technique.

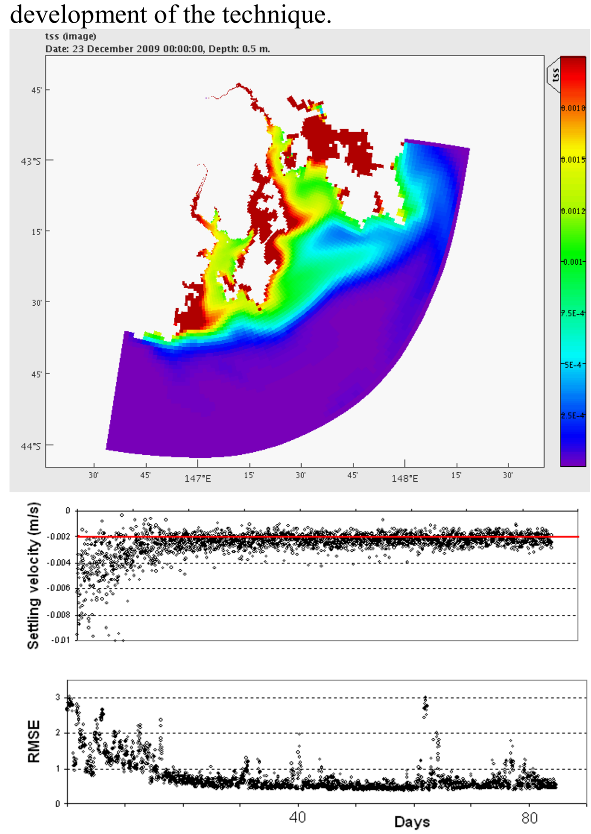

Figure 3. Surface TSS (top) and estimated settling velocity and RMSE (bottom) of the predicted sediment concentrations. TSS is given in kg m-3, and the red line shows "true" settling velocity.

IV. DISCUSSION AND CONCLUSION

One of the key objectives of the operational modeling system is to deliver near-real-time simulation products into the information system that would provide decision support for the coastal management operations. The quality of such services will be largely influenced by the reliability of the developed system, including robustness of the model operations and quality of the simulated data and measurements. The modeling system comprises a number of the numerical models driven by various on-line data streams. Failure to access one of such data-streams would result in a failure of the whole system. The problem is further exacerbated by the large number of boundary conditions and forcing data required by the model. To ensure robust performance of the system, statistically validated estimates of the input data based on current and past observations are likely to be used. High uncertainty and stochastic nature of such data begs for a more general question about uncertainty of the modeling system predictions.

The modelling system is anticipated to provide a quantitative assessment of its own quality. The challenge is to evaluate and estimate uncertainty of the integrated system as a whole; distinct from the uncertainty of the individual components. The evaluation procedures must also take into account the key sources of the uncertainty including stochastic boundary conditions and forcing, poorly known parameters, structural errors, initial conditions, data error etc. While conceptually such evaluation can be carried out via Monte Carlo sampling, computational constraints make direct application of this method not feasible. The strategy based on the dimension reduction and emulation of complex models nested into a hierarchical framework we believe is one of the potential avenues to address this problem. Fast and cheap statistical surrogates of complex models have a potential to underpin statistical evaluation of complex computationally expensive models. Preliminary evaluations illustrated in this paper show promising results, however, at this stage, more research is required to better understand the capabilities and limitations of this approach.

The developments described in this paper are focused on the evaluation of water quality parameters, instrumental to material cycling at higher trophic levels. In a management context it is often important also to understand and predict ecological processes on ecosystem levels nested into a broader socio-economic framework. Within the INFORMD program the development of ecosystem models is underway ("Atlantis") that gives some promise for a future data-assimilating model integrating across various ecosystem dimensions.

The SETAS operational modeling platform highlights a strategy for the delivery of science to practice distinct from the conventional research studies. Instead of the one-off scientific research within the project life-time addressing particular science questions in the region and then moving to a new case-study area, the strategy is to build an infrastructure for a long-terms ongoing support of the research and applications in SETAS to ensure sustainable development of the region. Such infrastructure would require on-line monitoring capabilities, operational data-assimilating models, and information system integrated into the decision support framework. Scientific and technical challenges at the early development stage are likely to be exacerbated with time with the challenges of long-term maintenance of such system. The delivery partners and clients must be identified and communication channels established at early stages of the development to ensure transition of the decision support capabilities beyond the project life-time.

The tools and technologies developed in SE TAS are anticipated to be transferrable to other regions and modeling systems. More generally the developed system is anticipated to have an educational value for various community groups including students and environmental managers. Open access to observations and simulation products via the information system would facilitate communication between involved parties and make more transparent decision procedures through the awareness of stakeholders to the quality of data sets underpinning the management decisions.

ACKNOWLEDGMENT

This research was supported by CSIRO Wealth from Oceans, and Computational and Simulation Science themes.

REFERENCES

[1] M. Herzfeld, "An alternative coordinate system for solving finite difference ocean models", *Ocean Modelling*, vol. 14, pp. 174-196, 2006.

[2] N. Margvelashvili, "Stretched Eulerian coordinate model for coastal sediment transport", *Computers and Geosciences*, vol. 35, pp. 1167-1176, 2009.

[3] A.G. Murray and J.S. Parslow, "Modelling nutrients impact in Port Phillip Bay – a semi-enclosed marine Australian ecosystem", *J. Marine Freshwater Res.*, vol. 50, pp. 597-612, 1999.

[4] J.S. Parslow, P. Sakov, and J. Andrewartha, "Integrated Model Implementation and Calibration – Final Report",. Gippsland Lakes Environmental Study Technical Report, CSIRO, 2001.

[5] J.S. Parslow, J. Volkman, I. Webster, B. Robson, M. Herzfeld, N. Margvelashvili, A. Revill, J. Andrewartha, P. Sakov, **"**Response To Catchment Flows and Loads in Australian Tropical and Temperate Estuaries", Paper delivered at Coast-to-Coast Conference, Hobart, 2004.

[6] K.Wild-Allen, P.A. Thompson, J.K. Volkman, and J.S. Parslow, "Use of a coastal biogeochemical model to select environmental monitoring sites", *J. Marine Systems, special edition 'Proceedings of the 41st Liege Colloquium on Ocean Dynamics'*., in press.

[7] G.Evenson, "The Ensemble Kalman Filter: theoretical formulation and practical implementation"*, Ocean Dynamics,* vol. 53, pp. 343-365, 2003.

[8] P. Oke, R., G. B. Brassington, D. A. Griffin, A. Schiller, "The Bluelink Ocean Data Assimilation System (BODAS)". *Ocean Modelling,* vol. 20, pp. 46-70, 2008.

[9] N.Margvelashvili and E.Campbell, "Gaussian Orthogonal Function Emulation of the suspended sediment model", unpublished.

[10] M. Dowd, "Bayesian statistical data assimilation for ecosystem models using Marcov Chain Monte Carlo", *J. Marine Systems*, vol. 68, pp. 439-456, 2007.

[11] N.Margvelashvili, "Sequential data-assimilation in fine-resolution models using error-subspace emulators: theory and preliminary evaluation", unpublished.