Bayesian Hierarchical Models to Augment the Mediterranean Forecast System

L. Mark Berliner
Department of Statistics, The Ohio State University
1958 Neil Ave.
Columbus, OH 43210
phone: (614) 292-0291 fax: (614) 292-2096 email: mb@stat.osu.edu

Ralph F. Milliff
Colorado Research Associates Division, NWRA
3380 Mitchell Lane
Boulder, CO 80301
phone: (303) 415-9701 fax: (303) 415-9702 email: milliff@cora.nwra.com

Christopher K. Wikle
Department of Statistics, University of Missouri
146 Middlebush Hall
Columbia, MO 65211
phone: (573) 882-9659 fax: (573) 884-5524 email: wikle@stat.missouri.edu

Emanuele Di Lorenzo
School of Earth and Atmospheric Sciences, Georgia Institute of Technology
311 Ferst Drive
Atlanta, GA, 30332
phone: (404) 894-3994 fax: (404) 894-5638 email: edl@eas.gatech.edu

Award Number: N00014-07-1-0512
http://www.cora.nwra.com/MedBhm/

LONG-TERM GOALS

The overall project goal has been to test the feasibility and practicality of Bayesian Hierarchical Model (BHM) methods in aspects of the Mediterranean Forecast System (MFS); an operational ocean data assimilation and forecast system.

OBJECTIVES

Three main objectives have been pursued in support of the project goal. They are:

1. a surface wind BHM (MFS-Wind-BHM) to drive ensemble ocean data assimilation and forecasts in MFS;

2. a time- and depth-dependent background error covariance BHM (MFS-Error-BHM) to evolve the background error covariance matrix in 13 sub-regions of the MFS forecast domain; and
3. a BHM to demonstrate super-ensemble forecast capabilities (MFS-SuperEnsemble-BHM) for ocean applications.

APPROACH

MFS-Wind-BHM

The wind-driven component of the circulation comprises the major part of the total circulation in the Mediterranean Sea on the 10-day timescales of the MFS operational forecasts. MFS-Wind-BHM is designed to perturb the surface wind forcing for data assimilation and forecast steps of the MFS ocean forecast cycle as a means of generating an ocean ensemble forecast. The surface wind perturbations are constrained to be within the uncertainties prescribed in data stage and process model distributions. Data stage distributions are constructed from measurement error models for the QuikSCAT surface wind retrievals and from analysis-forecast error statistics for the ECMWF surface wind fields. The process model distribution is based on the leading order terms of the Rayleigh Friction Equations (RFE; Stevens et al., 2002). The purturbations are drawn from the posterior distribution of the MFS-Wind-BHM model for the 14-day sequential data assimilation step leading to ensemble ocean initial conditions, and for the 10-day forecast step generating the ensemble forecast. Ten member ensembles are used in the pilot experiments. Initial condition and forecast posterior mean fields have been described in previous reports. Initial condition and forecast uncertainty are quantified by the spread in the ensemble fields for the ocean state.

MFS-Error-BHM

The approach, recent work completed, and results for the latest implementation of MFS-Error-BHM are presented in the FY2010 annual report for the ONR Physical Oceanography Program BRC project, “Bayesian Hierarchical Model Characterization of Model Error in Ocean Data Assimilation and Forecasts”, (Berliner, Wikle, Milliff; co-PIs). We review the salient points in abbreviated sections here and refer the reader to that report for more detailed discussion and figures.

In general, the dimension of the full state ($n \times n$) error covariance matrix Σ_t is reduced by expanding in a set of $n \times p$ basis functions Φ (e.g. EOFs, where $n >> p$) as

$$\Sigma_t = \Phi B_t \Phi'$$

where the time dependent background error covariance matrix of interest is B_t of dimension $p \times p$. This matrix is not, in general, diagonal since the amplitudes of the basis functions are allowed to change with time.

We define an error process model as $e_t = Z\beta_t + \eta_t$ where Z are vertical basis functions based on multivariate (i.e. T and S) profiles, the β_t are time-dependent amplitudes and η_t is a zero-mean Gaussian error term to account for additional uncertainty that is introduced e.g. in the dimension reduction.

The key assumption is that β_t is Gaussian, with zero mean and variance given by B_t from above. We use a modified Cholesky decomposition (Chen and Dunson, 2003) to write

$$B_t = \Lambda_t \Gamma_t \Gamma'_t \Lambda_t$$

where Λ_t is a diagonal matrix with elements proportional to the standard deviations of the elements of β_t, and Γ_t is a lower triangular matrix associated with correlations among the β_t.

At the next level of the Bayesian hierarchy, we model the Λ_t and Γ_t as autoregressive time series, where the autoregression parameter distributions are specified at still lower levels of the BHM.

Data stage distributions are built from *misfit* d_t and *anomaly* q_t datasets from MFS forecasts. The model misfits are forecast differences with respect to *in-situ* observations. The anomalies are departures from the model "year minus day" climatologies. These vectors can be written

$$\begin{aligned} d_t &= H_t(X_{t|t-1}) - x_t^{obs} \\ q_t &= X_{t|t-1} - \bar{x} \end{aligned} \tag{1}$$

where H_t is the operator that moves the forecast $X_{t|t-1}$ to the observation x_t^{obs} locations for comparison, and $\bar{X}_t$ is the climatology value for the model state variable X.

MFS-SuperEnsemble-BHM

Super-ensembles of ocean forecasts in the Mediterranean Sea can be constructed to quantify a multi-model posterior distribution and model biases for target ocean processes. Combining ensembles from several models is a challenging problem since the models may have different biases, variabilities and skills. Adjusting for these characteristics and managing uncertainty are keys. The Bayesian strategy suggested in Berliner and Kim (2008) offers a method that can treat these issues. In this approach ensembles are processed much like observational data and combined with a prior probability distribution for the state variables of interest.

MFS-SuperEnsemble-BHM is our implementation of the Berliner and Kim (2008) methodology for temperature and salinity profile evolution (e.g. $T(z,t)$ and $S(z,t)$) at two locations in the Rhodes Gyre region of the eastern Mediterranean. The periods of interest span February-March for the years 2006-2009.

We have assembled ensembles from three models: NEMO, OPA, and MedROMS. Each model has been forced by 10 realizations from MFS-Wind-BHM, as well as by the ECMWF analysis winds for the periods of interest. A total of 33 realizations serve as our data stage inputs. The MFS analysis fields, version SYS3a2, are used to formulate the MFS-SuperEnsemble-BHM prior (or process model stage distribution).

Two approaches are used to analyze the SYS3a2 output to form a prior:

(1) simple, data analytic summaries of state variable means and covariances serve as estimates of prior means and covariances; and

(2) a stochastic time series model (specifically, a multivariate autoregressive model including parameters for two temporal lags) is fit to the MedROMS output and then used as the prior.

WORK COMPLETED

MFS-Wind-BHM

Two manuscripts (Milliff et al., 2010; and Bonazzi et al., 2010) have been submitted, reviewed, and revised for publication in the *Quarterly Journal of the Royal Meteorological Society*. In part 1 (Milliff et al., 2010) the theory and implementation of MFS-Wind-BHM is described in detail. An appendix includes the term-by-term definitions of the probability distributions that arise at each level of the model hierarchy. In part 2 (Bonazzi et al., 2010), the impact of the ensemble forecast methodology based on MFS-Wind-BHM perturbations is documented. Forecast uncertainty is concentrated at the ocean mesoscale; it is shown to be time and location dependent.

MFS-Error-BHM

Wikle has been computing B_t based on misfit and anomaly data stages for the period January-May 2007, for the MFS sub-region in the Gulf of Lions. Wikle supplies vertical EOFs (Z) and B_t to Dr. Srdjan Dobricic at MFS, who then runs twin hindcast experiments with the operational error covariance and the B_t from MFS-Error-BHM for this period. Until now, there have been "order-of-magnitude" differences in the computed error covariances between the two methods.

During the summer visit by Prof. Nadia Pinardi to NWRA/CoRA it was finally determined that the regional seasonal cycle removal procedure at MFS is specific and should be implemented in MFS-Error-BHM as well. Preliminary experiments using only anomaly (q_t) data stage inputs demonstrate that following the MFS pre-processing removes the discrepancy. Figures in the ONR annual report for "Characterizing Model Error" (see Related Projects) demonstrate that, even in the absence of d_t data stage inputs, the forecast impact of MFS-Error-BHM is neutral. Experiments are underway now to introduce d_t back into the MFS-Error-BHM and quantify forecast impacts at MFS.

MFS-SuperEnsemble-BHM

We have assembled all needed datasets and completed algorithmic development. Archetypal numerical codes have been developed, run, and assessed for effectiveness. Currently, those codes are being scaled up to the multivariate setting of our $S(z,t)$ and $T(z,t)$ datasets. We have designed experiments to exemplify and assess the method, including the role of the prior and processing the Bayesian output for forecasting purposes as well as model assessment (i.e., estimates of biases).

A major paper presenting the results will be completed and submitted in December, 2010.

While the code development for MFS-SuperEnsemble has been underway, further MedROMS development and validation have been undertaken by Di Lorenzo and colleagues at Georgia Tech. A MedROMS ensemble integration (whole basin, 3 members so far) spanning the period 1963-2009 has been produced using monthly ECMWF forcing.

With Dr. Hazem Nagy, Di Lorenzo has produced an early draft of a paper on the climate variability of the Mediterranean eastern basin circulation. The ensemble mean reproduces quite well the satellite data in the period after 1990. The climate simulations also capture the upper circulation inversion that occurred in the Ionian sea in the 1990s. Di Lorenzo et al. are able to show that these inversions have occurred in the past on multi-decadal timescales. These mutli-decadal fluctuations associated with the upper ocean inversion of the Ionian circulation are captured in the first mode of sea-surface height anomaly variability in the individual realizations and in the ensemble mean. Di Lorenzo et al have also isolated an atmospheric bridge from the Pacific to the Mediterranean Sea in their analyses. It appears that the second mode of variability in the eastern basin circulation integrates the effects of interannual tropical Pacific variability; so much so that the principal component of the eastern basin circulation mode reproduces indices of Pacific climate (i.e. ENSO and PDO) with correlation $R = 0.86$. The Georgia Tech group is exploring mechanisms and implications for the deeper circulation and on deep water formation in the Mediterranean. Interestingly, the effect of the NAO does not appear to dominate in the eastern basin.

While the MedROMS climate-scale research is not a task directly associated with any of the MFS BHM objectives, it does demonstrate model stability and utility in this newest component of the MFS-SuperEnsemble-BHM data stage inputs. As noted, a climate-scale manuscript is in preparation and ONR funding will be acknowledged.

Relevant Presentations

(Berliner, Milliff, Pinardi, Wikle) Informal presentations and discussions at the annual “All-Hands” project meeting at NWRA/CoRA, August, 2010.

(Milliff, Wikle; session co-conveners) Probabilistic Models in Ocean Sciences: Applications in Data Assimilation, Coupled Ecosystem Models and Air-Sea Interaction Studies, American Geophysical Union, Ocean Sciences Meeting, Portland, OR, February, 2010.

(Berliner) Combining Models and Data: The Bayesian Approach to Modeling and Prediction, Invited Talk, AGU Ocean Sciences Meeting, Portland, OR, February, 2010.

(Pinardi) A new method for ocean ensemble forecasting with quantification of wind uncertainties. Invited talk. AGU Ocean Sciences Meeting, Portland, OR, February, 2010.

(Milliff, Pinardi, Wikle, Berliner, Bonazzi) Process model considerations for a surface wind Bayesian hierarchical model. Poster, AGU Ocean Sciences Meeting, Portland, OR, February 2010.

(Milliff) Estimating semivariograms to build covariance matrices for J. Workshop on the ROMS 4D-Var Data Assimilation Systems for Advanced ROMS Users, University of California, Santa Cruz, July 2010.

(Wikle) A hierarchical approach to motivate spatio-temporal statistical models. Institute for Pure and Applied Mathematics (IPAM), UCLA, Los Angeles, CA, May 25, 2010.

(Wikle) Bayesian hierarchical models to augment the Mediterranean forecast system. Invited talk. Iowa State University. Ames, IA, October 15, 2009.

(Wikle) Don’t forget the process! Using scientific process knowledge to motivate spatio-temporal models. Invited talk. SAMSI Program on Space-Time Analysis for Environmental Mapping, Epidemiology and Climate Change, Opening Workshop, RTP, North Carolina, September 14, 2009.

(Wikle) A class of nonlinear spatio-temporal dynamic models. Invited Talk, Joint Statistics Meetings, Washington, DC, August 4, 2009.

RESULTS

MFS-Wind-BHM

We are awaiting word from *QJRMS* on the status of the revised manuscripts; Milliff et al., 2010 (part I) and Bonazzi et al., 2010 (part II).

MFS-Error-BHM

Embedded scales in the error covariance estimations of ocean forecast systems act to rescale the error covariance magnitudes. Anomaly data stage inputs are probably not sufficient to represent abrupt regime shifts in the ocean state and provide added value over operational methods using a fixed, seasonal, background error covariance. Experiments adding misfit data stage inputs will be useful in modelling error covariance response to ocean regime shifts in the Mediterranean on sub-seasonal timescales.

Under seperate ONR funding, the MFS-Error-BHM methodology is being adapted to the California Current System and the Regional Ocean Model System, four-dimensional variational data assimilation and forecast tools developed by Prof. A. Moore and colleagues.

IMPACT/APPLICATIONS

MFS-Wind-BHM

BHM methods are proving feasible and practical in an array of ocean data assimilation and forecast applications. The papers, Milliff et al. (2010) and Bonazzi et al. (2010) will serve as detailed generic references for this methodology in ocean and atmospheric sciences.

MFS-Error-BHM

Refining estimates of the time-dependent changes in ocean states and forecast uncertainty across regime shifts adds value to ocean forecast system output.

MFS-SuperEnsemble-BHM

First steps in developing super-ensemble ocean forecast techniques for targeted fields and/or processes will provide for feasibility and practicality assesments of this popular, but poorly understood, methodology. Model bias distributions will demonstrate a new means of comparing forecast tools that are being developed in the community (i.e. ROMS, OPA, NEMO).

TRANSITIONS

Informal communications continue with scientists in the Ocean Modelling branch of the Naval Research Laboratory, Bay St. Louis, MI.

RELATED PROJECTS

“Estimating Ecosystem Model Uncertainties in Pan-Regional Syntheses and Climate Change Impacts on Coastal Domains of the North Pacific Ocean”, NSF US Globec Program, October 2008 - October 2011.

“Bayesian Hierarchical Model Characterization of Model Error in Ocean Data Assimilation and Forecasts”, ONR Physical Oceanography Program; Milliff, Berliner, Wikle, co-PIs.

“Quantifying the Amplitude, Structure and Influence of Model Error during Ocean Analysis and Forecast Cycles”, ONR Physical Oceanography Program, A. Moore (PI).

REFERENCES

Berliner, L.M. and Y. Kim, 2008: “Bayesian design and analysis for super ensemble based climate forecasting”, *J. Climate*, **21**, 1891-1910.

Chen, Z. and Dunson, D. B., 2003: “Random effects selection in linear mixed models”, *Biometrics*, **59**, 762–769.

Stevens, B., J. Duan, J.C. McWilliams, M. Munnich and J.D. Neelin, 2002: “Entrainment, Rayleigh friction, and boundary layer winds over the tropical Pacific”, *J. Climate*, **15**, 30-44.

PUBLICATIONS

Bonazzi, A., N. Pinardi, S. Dobricic, R.F. Milliff, C.K. Wikle and L.M. Berliner, 2010: “Ocean Ensemble Forecasting, Part II: Mediterranean Forecast System Response”, revised for *Quarterly Journal of the Royal Meteorological Society*.

Milliff, R.F., A. Bonazzi, C.K. Wikle, N. Pinardi and L.M. Berliner, 2010: “Ocean Ensemble Forecasting, Part I: Ensemble Mediterranean Winds from a Bayesian Hierarchical Model”, revised for *Quarterly Journal of the Royal Meteorological Society*.

Wikle, C.K. and M.B. Hooten, 2010: A general science-based framework for spatio-temporal dynamical models. Invited discussion paper for TEST, Official Journal of the Spanish Society of Statistics and Operations Research, In press.