Predictive Skill, Predictive Capability and Predictability in Ocean Forecasting

A.R. Robinson, P.J. Haley, Jr., P.F.J. Lermusiaux, W.G. Leslie

Harvard University
29 Oxford Street
Cambridge, MA 02138
robinson@pacific.deas.harvard.edu

Abstract - **We discuss the concepts involved in the evaluation and quantitative verification of ocean forecasts and present two predictive skill experiments to develop and research these concepts, carried out in the North Atlantic and Mediterranean Sea in 2001 and 2002. Ocean forecasting involves complex ocean observing and prediction systems for ocean regions with multi-scale interdisciplinary dynamical processes and strong, intermittent events. Now that ocean forecasting is becoming more common, it is critically important to interpret and evaluate regional forecasts in order to establish their usefulness to the scientific and applied communities.**

The Assessment of Skill for Coastal Ocean Transients (ASCOT) project is a series of real-time Coastal Predictive Skill (CPSE) and Rapid Environmental Assessment (REA) experiments and simulations focused on quantitative skill evaluation, carried out by the Harvard Ocean Prediction System group in collaboration with the NATO SACLANT Undersea Research Centre. ASCOT-01 was carried out in Massachusetts Bay and the Gulf of Maine in June 2001. ASCOT-02 took place in May 2002 in the Corsican Channel near the island of Elba in the Mediterranean Sea. Results from the ASCOT exercises highlight the dual use of data for skill evaluation and assimilation, real-time adaptive sampling and skill optimization and present both real-time and *a posteriori* evaluations of predictive skill and predictive capability.

I. INTRODUCTION

Ocean states that are initially close often separate rapidly over time. Predictability studies for regional forecast systems and for interdisciplinary nonlinearities are important. Quantitative studies and methodologies are needed for quantifying the theoretical predictability, the inherent ability to determine the future state of ocean phenomena from the current state. The predictability limit is the theoretical time necessary for two slightly different true ocean states to become undistinguishable from two arbitrarily chosen states. It is inherent to the growth rate of errors for a perfect prediction model. It depends on the true ocean processes under consideration (days to months for mesoscale ocean processes) and is a function of the initial uncertainty.

The ability of a system to predict certain ocean phenomena is the predictive capability of the system for those phenomena. It considers all sources and reductions of errors (initial and boundary conditions, model, data, etc.) and their evolution. It needs to be quantified as a function of the observation network, models and assimilation criterion used. The system predictive capability is ultimately limited by predictability. Before the predictability limit is reached, other sources of error (quality and quantity of data, forcings, model structures and parameters, initialization and assimilation schemes, etc.) limit the predictive capability.

In general, one can decompose the predictive capability error or total error covariance into its estimated predictability and model error components. These two components can also be further divided into phase errors, the right feature at the wrong place and time, and structural errors, an approximate feature at the right place, e.g. wrong meander wavelengths or eddy characteristics. With data assimilation, observation model errors also influence these components. Around data times, the error covariance is then reduced in accord with the assimilation criterion used (Robinson *et al*., 1998). Studying errors and assessing the relative importance of their components is made challenging by a lack of (multi-scale) data. The multivariate and multi-scale properties of ocean dynamics as well as the regional peculiarities, both in space (coastal vs. deep ocean) and time (external forcings vs. internal mesoscale turbulence) also need to be considered. Future research issues will include the areas of: sensitivity to the selection of algorithm (streamfunction, surface pressure), numerical parameters, grid resolution, subgrid scale parameterization, data assimilation methodology, data coverage, etc.

For the evaluation and verification of ocean forecasting, significant guidance from meteorology exists, but much of the methodology must be ocean-specific and devised for both generic and regional-specific qualitative and quantitative skill metrics. Forecasts are ultimately limited by loss of predictability related to the non-linear transfer of data errors from smaller to larger scales. As a regional forecast system is developed, the predictive capability is less than the predictability limit. Any quantitative measure of forecasting ability or skill can be termed Predictive Skill (e.g. Root-Mean Square Error (RMSE) and Pattern Correlation Coefficient (PCC)).

II. THE HARVARD OCEAN PREDICTION SYSTEM

The Harvard Ocean Prediction System (HOPS) (Fig. 1) is an integrated system of data analysis and assimilation schemes, and a suite of coupled interdisciplinary (physical, acoustical, optical, biogeochemical-ecosystem) dynamical models (Robinson *et al*., 1998; Robinson, 1999). This
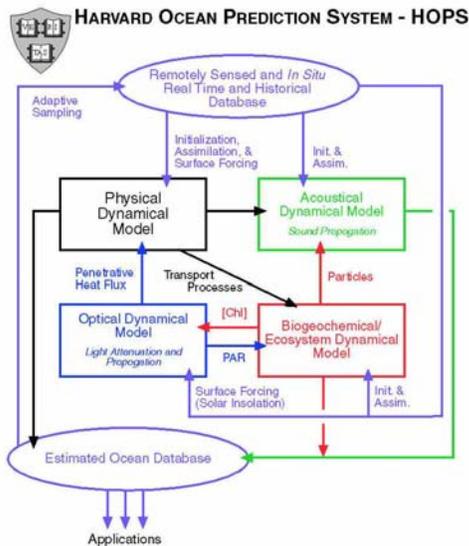
Fig. 1. A schematic of the components of the Harvard Ocean Prediction System.

system was developed for producing interdisciplinary oceanic field estimates that include effective and efficient data assimilation, dynamically consistent model initialization, multi-scale nesting, and model-driven adaptive sampling with feedbacks. HOPS employs a primitive equation (PE) physical dynamical circulation model. Boundary layers (top and bottom) and isopycnal and diapycnal turbulence are modeled through process parameterization and scale-dependent filters. Multiple sigma vertical coordinates are calibrated for accurate modeling of steep topography. Multiple two-way nests are an existing option for the horizontal grids.

The ability to measure the performance and skill of ocean nowcast/forecast systems such as HOPS is a growing area of interest. Quantitative demonstrations of ocean mesoscale forecast skill are in an early stage of development. In operational practice, the simple common sense measures, like root-mean-square error or performance with respect to persistence (e.g. Tapp, 1986; Carton, 1987) are most often employed. The sparseness of oceanic data, compared to atmospheric data, makes the verification problem generally different. Often the lack of sufficient data requires one to devise clever measures of useful skill, such as mean frontal axis position error (Willems *et al.*, 1994) or eddy-spawning event statistics (Robinson *et al.*, 1989).

A robust set of metrics to quantitatively measure the performance of coupled model-observation systems needs to be developed. Such metrics should be carefully designed to objectively reveal and differentiate the behaviors of the numerics, physics, and parameterizations and their relative impact in the system's ability to estimate reality. Papers can be found in the literature (e.g., for atmospheric models: Williamson *et al.*, 1992, and for ocean models: Haidvogel and Beckmann, 1998) that discuss objective, idealized tests.

Lynch (1995) provides an overview of the problem from a coastal ocean viewpoint. Recent research on the measures of skill of forecasts using real data (Murphy and Epstein, 1989; Krzysztofowicz, 1992; Stamus *et al.*, 1992; Murphy, 1996; Buizza and Palmer, 1998), and on the verification of prediction systems and study of predictive capabilities (Murphy, 1991; Briggs and Levine, 1997; Ward and Navarra, 1997; Majewski, 1997; Xue, *et al.*, 1997; Ghil and Jiang, 1998), are also useful.

III. COASTAL PREDICTIVE SKILL EXPERIMENTS

A Coastal Predictive Skill Experiment (CPSE) measures the ability of a forecast system to combine model results and observations in coastal domains or regimes and to accurately define the present state and predict the future state. Rapid Environmental Assessment (REA) is defined in the naval environment as "the acquisition, compilation and release of tactically relevant environmental information in a tactically relevant time frame". An REA CPSE is designed to determine forecast skill on the basis of minimal and covertly attainable observations.

Considerations for the design and analysis of CPSEs relate to the quantifiable skill of the CPSEs and to the accuracy of error estimates. The skills relate to the ability of the system to: i) attain specific accuracy; ii) track specific processes; and, iii) provide sufficiently accurate information for practical needs or for decision making processes; etc. Important applications for the measure of skill involve the objective intercomparison between: i) forecasting systems or particular configurations of observational network; ii) measurement models; and, iii) dynamical models of an ocean prediction system. The measurements of skill are usually statistical measures; but other measurements are required: i) phase errors, in order to account for the fact that a particular structure might evolve at a different rate in the model than in reality; and, ii) structural errors, in order to account for the fact that a particular feature may evolve with the wrong characteristics. After this has been done for a number of regions, synthesis and abstraction can be expected to lead to some simple overall concepts. A comprehensive set of standard metrics needs to be defined for CPSEs. This requires substantial additional research and community effort in order to define skill measurements and validation procedures in the coastal ocean for a variety of processes, regions, and applications. Additionally, each CPSE will need its own set of metrics, assembled from standard metrics, which reflect the unique regional dynamics.

IV. ASSESSMENT OF SKILL FOR COASTAL OCEAN
TRANSIENTS (ASCOT)

The ASCOT project is a series of real-time CPSE/REA experiments and simulations focused on quantitative skill evaluation and cost-effective forecast system development. The overall goal of the ASCOT project is to enhance the efficiency, improve the accuracy and extend the scope of

nowcasting and forecasting of oceanic fields for CPSE and for REA in the coastal ocean and to quantify such CPSE and REA capabilities.

ASCOT-01

ASCOT-01 was carried out in Massachusetts Bay and the Gulf of Maine in June 2001. The specific objectives of ASCOT-01 were to: 1) carry out and quantitatively evaluate in Massachusetts Bay and the Gulf of Maine a coupled multi-scale interdisciplinary real-time forecast experiment; 2) obtain a data set adequate to define coupled dynamical processes (submeso-, meso-, bay-, gulf- scales) that govern the formation and evolution of structures and events, including generic processes and the coupling of wind-forced events and buoyancy currents; and, 3) obtain an intensive data set adequate for definitive quantitative skill assessment and suitable for the design of minimal data requirements for both REA and for an efficient regional monitoring and prediction system.

As a predictive skill experiment, ASCOT-01 included over sampling, in order that sources of error could be tracked. During the verification survey a significant fraction of the initialization survey was repeated. Adaptive sampling survey patterns were designed to address: 1) the interactions of Massachusetts Bay and the Gulf of Maine (inflow updates, exchanges, etc.); 2) response to storms or air-sea exchanges (upwelling, structures of currents and gyres, bifurcation structures in the Gulf of Maine, etc.); coupling of wind-response and buoyancy currents; reduction of multi-variate forecast errors; and, update of information for feature model parameters
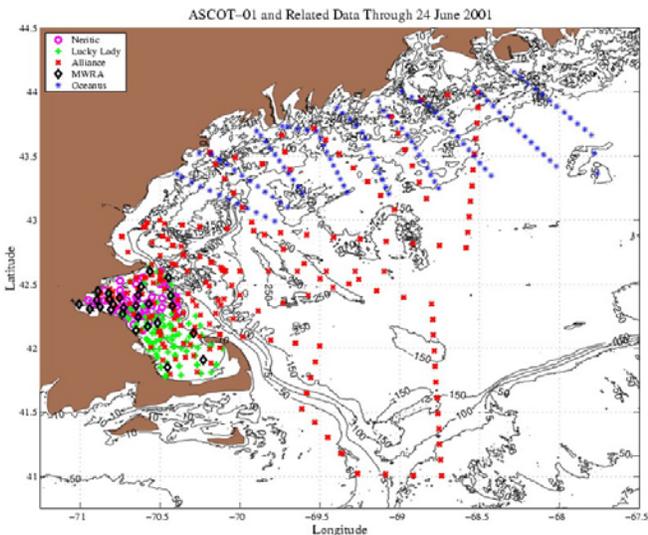

Fig. 2. CTD station positions for ASCOT-01

The ASCOT-01 experiment utilized three dedicated research vessels: the NRV Alliance (NATO SACLANT Undersea Research Centre), the RV Lucky Lady (University of Massachusetts – Dartmouth), and the RV Neritic (University of Massachusetts – Boston). In addition to dedicated research vessels, other hydrographic data were available during this time. As part of the ECOHAB project

(http://www.whoi.edu/science/B/ecohab/), the RV Oceanus made a CTD survey on a dense station grid in the northern Gulf of Maine (OC366) during early June 2001. Chief scientist Jim Churchill made the data available to the ASCOT program in real time. The Massachusetts Water Resources Authority (MWRA) conducts routine surveys of Massachusetts Bay through an ongoing monitoring program (http://www.mwra.state.ma.us/harbor/html/mbmon.htm). Data from survey WF017 was provided to the ASCOT program in real time.

Atmospheric forcing fields were acquired from the US Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC). The FNMOC fields included surface pressure, air temperature, 12-hour forecast precipitation, surface winds, relative humidity, cloud cover, sea surface temperature and mixed layer depth. The model and analysis fields, including 00Z and 12Z nowcasts and forecasts for up to 144 hours on a 1-degree or 2-degree resolution grid, were downloaded via the Internet. These gridded fields were interpolated in space and time onto the HOPS model grids and used to compute fluxes that drove the HOPS models at the surface. Flux analyses were used whenever possible and forecast fluxes were replaced by the analyses as those analyses became available

Data analysis, data assimilation and numerical simulations were carried out on a daily basis in real-time throughout the duration of the exercise. Data was analyzed, quality controlled and processed as it was received and made available for assimilation into HOPS. Model simulations were carried out both aboard the Alliance and at Harvard. The operational physical forecasts were performed aboard the Alliance, while the biological forecasts were completed at Harvard (Besiktepe, *et al*., 2992; Lermusiaux, *et al*., 2002). Here the focus is on the physical forecasts.
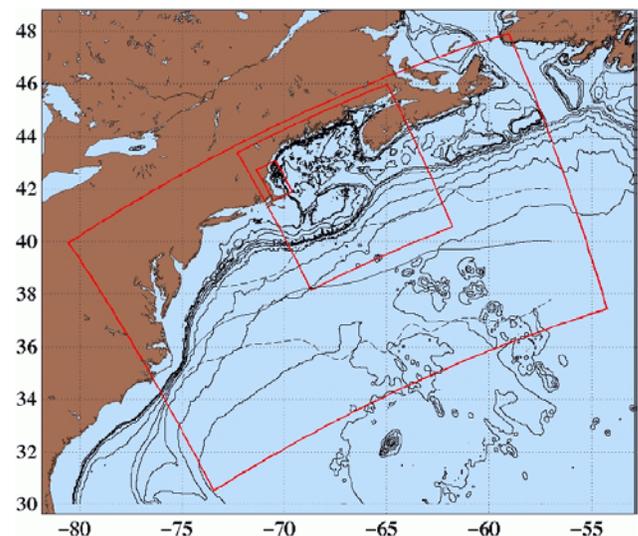

Fig. 3. Nested modeling domains for ASCOT-01

The physical ASCOT-01 simulation and operational system consisted of a set of three two-way nested domains: the

Northwest Atlantic (NWA), the Gulf of Maine (GOM) and Massachusetts Bay (MB) (Fig. 3). A two-way nested domain pair consists of a dynamical model defined in two domains, one with coarser resolution containing the other with finer resolution. Currently, technical considerations require a 3:1 ratio of domain resolutions. Information from the finer resolution domain is used to replace information in the coarser resolution domain areas that intersects with the finer resolution domain (up-scale). Information from the coarser resolution domain around the boundaries of the finer resolution domain is interpolated to improve boundary information in the finer resolution domain (down-scale). In the operational context, there was two-way nesting between the NWA and GOM (NWA/GOM) domains and the GOM and MB (GOM/MB) domains. The NWA/GOM nested run provided boundary conditions for the GOM during the GOM/MB nested run. The coupled physical/biological simulations were run in a stand-alone Massachusetts Bay domain.

Typical model forecasts lasted for seven model days. This duration allows for the assimilation of hydrographic and remotely sensed data, a nowcast and short (four day) forecast. Lengthier forecasts are possible but were not considered necessary for this operation. Forecasts were available on a daily basis after the initialization survey in order to provide adaptive sampling patterns for the subsequent day's sampling. Products were available from the NRV Alliance via the experiment web site (http://www.deas.harvard.edu/~leslie/ASCOT01). Products continue to be available at this site.

ASCOT-02

ASCOT-02 took place in May 2002 in the Corsican Channel near the island of Elba in the Mediterranean Sea. The ASCOT-02 objectives were to: 1) carry out and quantitatively evaluate in the Eastern Ligurian Sea in the region of the island of Elba a multi-scale real-time forecast experiment; 2) obtain a data set adequate to define coupled dynamical processes (submeso-, meso-, bay-, gulf- scales) that govern the formation and evolution of structures and events, including generic processes and the coupling of wind-forced events and buoyancy currents; 3) obtain an intensive data set adequate for definitive quantitative skill assessment and suitable for the design of minimal data requirements for both REA and for an efficient regional monitoring and prediction system; 4) support AUV exercises; and 5) provide a rigorous real time test of a distributed ocean prediction system technology.

The ASCOT-02 simulation system consisted of a pair of two-way nested domains: the Channel and Elba domains (Fig. 4). The Elba domain was designed to provide high resolution forecasts for the area to the north of Elba. Operationally, only results from the Channel domain were issued. As in ASCOT-01, data analysis, data assimilation and numerical simulations were carried out on a daily basis in real-time throughout the duration of the exercise. Atmospheric forcing was utilized as in ASCOT-0. All

simulations were carried out at Harvard. Products were (and remain) available via the experiment web site (http://www.deas.harvard.edu/~leslie/ASCOT02).
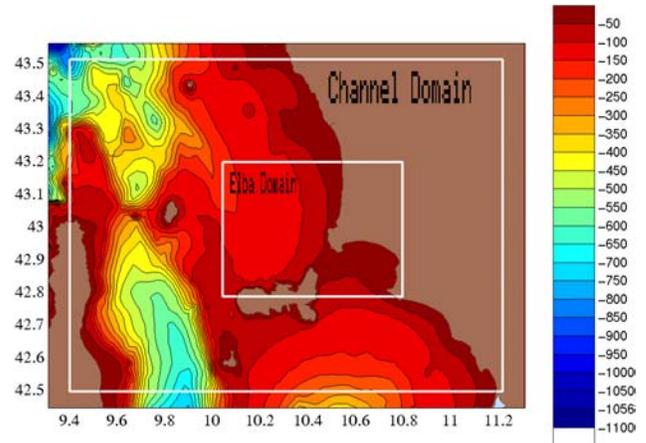

Fig. 4. ASCOT-02 modeling domains.

The hydrographic data for ASCOT-02 was collected by the NRV Alliance. A total of 269 successful CTD casts were carried out from 7-17 May 2002. The station positions from the 7-10 May initialization survey are shown in Fig. 5. A verification survey from 14-17 May repeated these positions. Adaptive sampling and AUV missions were carried out from 10-14 May.
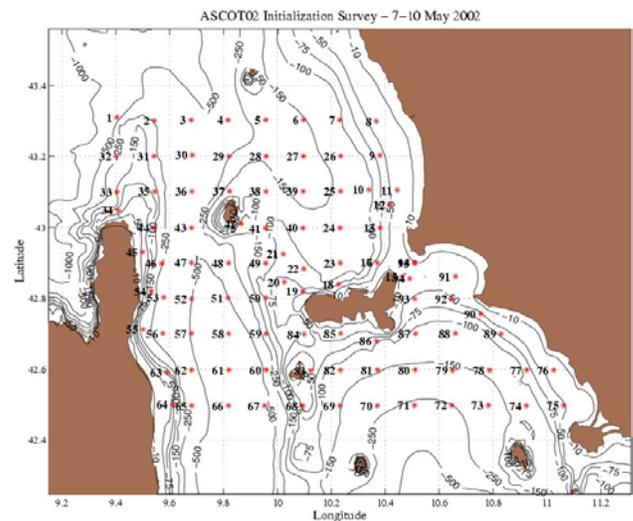

Fig. 5. CTD station positions for ASCOT-02 initialization survey

V. OPERATIONAL PREDICTIVE SKILL METRICS

Skill of the operational forecasts is measured using the metrics, Root-Mean-Square Error (RMSE) and Pattern Correlation Coefficient (PCC). These numbers are computed model level by model level or on depth surfaces, and as a volume average. Perfect values of the RMSE and PCC are, respectively, zero and one.

The metrics RSME and PCC are defined by:

$$\text{RMSE} \equiv \sqrt{(T^f - \hat{T})^T (T^f - \hat{T})} \doteq \left\| T^f - \hat{T} \right\|_2 \text{, and}$$

$$\text{PCC} \doteq \frac{(T^f - T^b)^T (\hat{T} - T^b)}{\left\| T^f - T^b \right\|_2 \left\| \hat{T} - T^b \right\|_2} \text{,}$$

where $\hat{T}$ denotes the true ocean, T^f its forecast, T^b a background field vector (e.g. large-scale field, climatological field, etc.), and $\| . \|_2$ the vector ℓ_2 norm.

A classic measure of skill is to compare the RMS and PCC of the forecast with those of the initial conditions (persistence). If the RMSE of the forecast is smaller than that of the IC, the forecast has RMS-skill or beats persistence. Similarly, if the PCC of the forecast is larger than that of the IC, the forecast has PCC-skill or has better patterns than persistence. The units of the RMSE are those of the quantity being evaluated. For temperature, the units are ^{o}C, and for salinity, PSU. PCC is a non-dimensional quantity.

VI. ASCOT PREDICTIVE SKILL RESULTS

ASCOT-01

The ASCOT-01 skill metrics have been determined by defining the initialization survey (6-8 June 2001) fields as "persistence", the forecast from the end of the assimilation of the initialization survey data as the "forecast", and the verification survey fields (15-20 June) as the "verification". The RMS and PCC statistics are calculated only in regions where the non-dimensional observation errors are estimated to be less than or equal to 0.3. The salinity observation errors for the verification survey are shown in Fig. 6. The statistics are only calculated, therefore, in the regions of Fig. 6 that are medium to dark blue. The statistics are computed at constant depths. The depths (2m, 10m, 40m) have been determined by important features in the vertical structure of the verification fields, e.g. depth of the mixed layer, center of the thermocline and deep (quiescent) conditions. Two additional intermediate depths (5m, 20m) have been added for completeness.

The comparison of forecast and persistence to verification for ASCOT-01 is shown in Fig. 7 (RMS Error) and Fig. 8 (PCC) for both temperature and salinity. The statistics of the forecast fields are shown as solid lines while persistence is shown as dotted. The bottom point in Fig. 7 and Fig 8 indicates the mean value of the points above. For both RMS Error and PCC, the forecast is clearly better than persistence for temperature in the upper water column. Below the center of the thermocline (10m), the values are essentially identical. For salinity, forecast results are better than persistence down to 20m. In the mean, forecast RMS errors are lower by approximately 1°C for temperature and

0.05 PSU for salinity. Temperature mean PCC values are approximately the same for the forecast and persistence, while for salinity, the mean PCC is significantly greater for the forecast versus persistence. These objective measures indicate a skillful forecast for ASCOT-01.
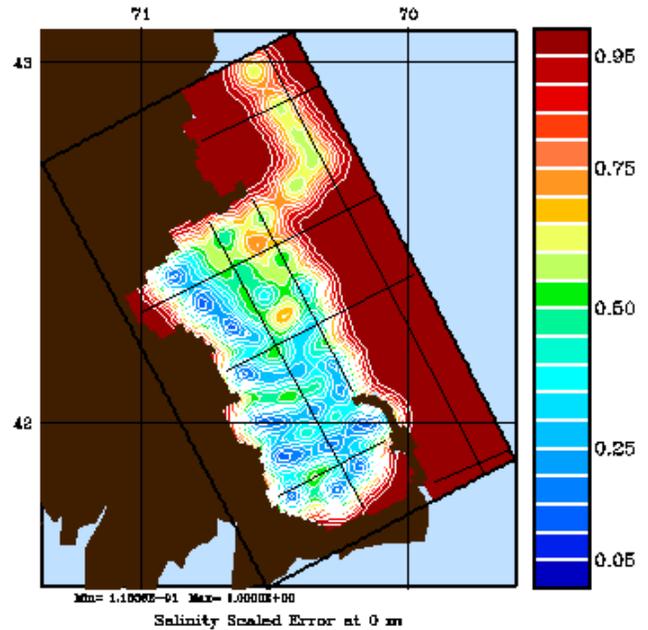

Fig. 6. Salinity observation errors for ASCOT-01 verification fields.
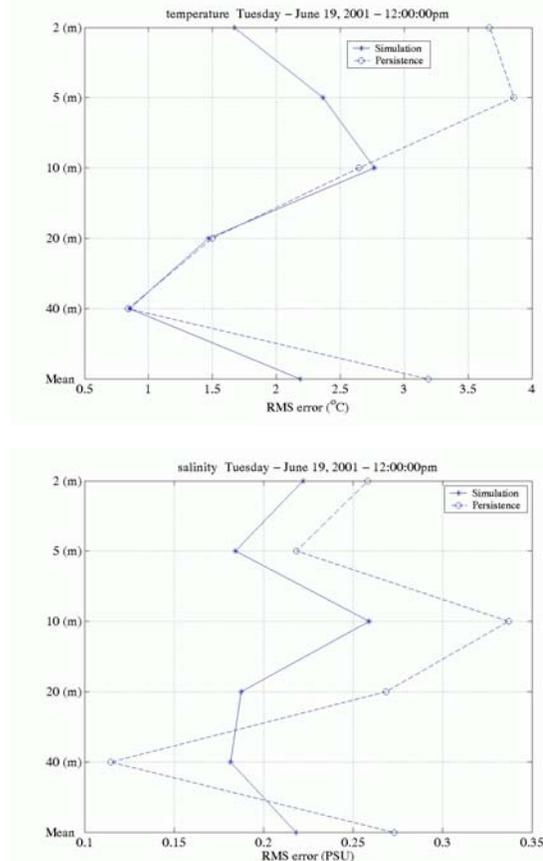

Fig. 7. ASCOT-01 RMS Error - Temperature (top), Salinity (bottom)

temperature Tuesday – June 19, 2001 – 12:00:00pm
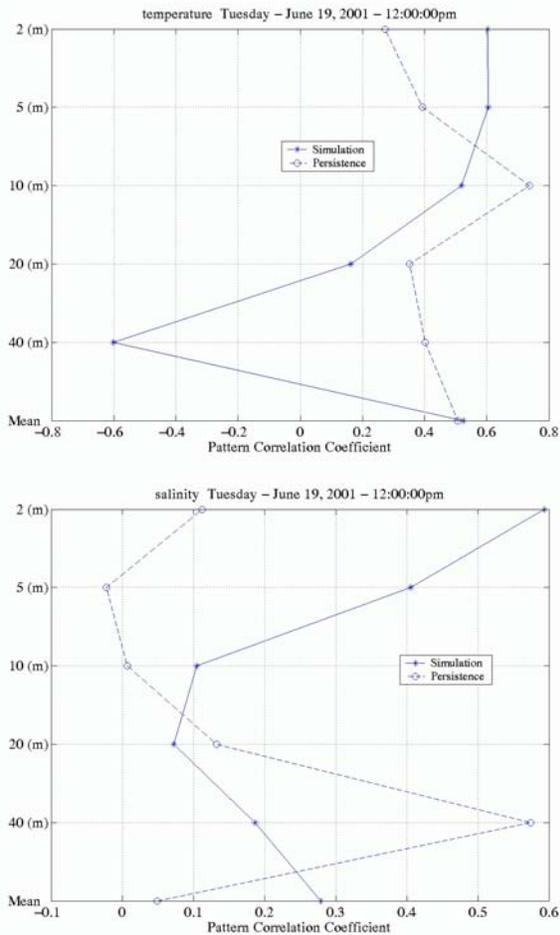

salinity Tuesday – June 19, 2001 – 12:00:00pm

Fig. 8. ASCOT-01 PCC - Temperature (top), Salinity (bottom)

ASCOT02

For ASCOT-02, there are two sets of verification fields, 15 May and 16 May 2002. The forecast, which includes data assimilation, is then 8.5 days (15 May 1200Z) or 9.5 days (16 May 1200Z) from the beginning of the initialization survey (7 May 000Z). Persistence is defined as the analysis of the initialization survey. RMS and PCC statistics are again only calculated in regions where the observation error is less than or equal to 0.3. Fig. 9 shows the observation errors for 15 May (top) and 16 May (bottom). Due to the logistics of the initialization survey (sailed from north to south) and adaptive sampling patterns (not shown) which focused more extensively on the area to the north of the island of Elba, the verification fields for 16 May have a smaller time difference from the last assimilation update than do the verification fields for 15 May. The statistics are computed at constant depths. The depths (2m, 20m, 40m, 200, and 400m) have been determined by important features in the vertical structure of the verification fields, e.g. depth of the mixed layer, center of the thermocline, center of the pycnocline, center of the halocline and deep (quiescent) conditions. Two additional intermediate depths (10m, 100m) have been added for completeness.

The ASCOT-02 comparison of forecast and persistence to verification for 15 and 16 May is shown in Fig. 10 (RMS Error) and Fig. 11 (PCC) for both temperature and salinity. Blue lines in each figure denote 15 May, while green lines indicate 16 May. The statistics of the forecast fields are shown as solid lines while persistence is shown as dotted. The bottom point in Fig. 10 and Fig 11 indicates the mean value of the points above. Note that the statistics are not calculated at 400m for 16 May, as this region does not reach this depth. For both temperature and salinity RMS error, the forecasts for both 15 and 16 May are clearly superior to persistence above the pycnocline (40m). Below that depth, the results are nearly identical. The mean temperature forecast RMS error is approximately 0.45^{o}C lower than that of persistence for 15 May and 0.85^{o}C lower for 16 May. The mean salinity forecast RMS error is approximately 0.005 PSU lower than that of persistence for 15 May and 0.055 PSU lower for 16 May. The temperature forecast PCC is larger than the persistence PCC above the center of the thermocline (20m) for both dates, but is lower below. The mean PCC for temperature indicates skill in the forecast, especially for 15 May. For salinity, results are similar. The mean May 16 salinity PCC indicates skill in the forecast, while for May 15, results are similar.
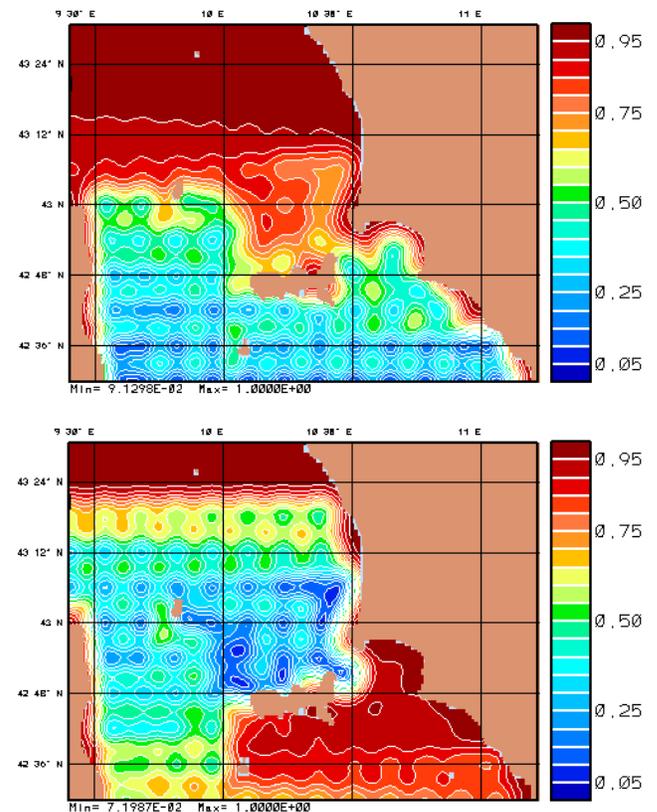

Fig. 9. Salinity observation errors for ASCOT-02 verification fields. 15 May (top), 16 May (bottom).

VII. CONCLUSIONS

Two intensive, coastal ocean surveys have been described. These experiments provided the basis for evaluating the skill of ocean prediction systems and for designing the minimal requirements for a rapid environmental assessment. An initial set of objective skill metrics, root-mean-square error and pattern correlation coefficient, have been used to assess the real--time skill of the Harvard Ocean Prediction System. The usefulness of such systems has been quantitatively demonstrated as a result of that assessment.

A couple of outstanding research issues remain in the design of skill metrics. First, no attempts have been made to account for phase or structural errors. Second, the overall results of these skill comparisons is that HOPS has better skill above the thermocline than below. This could be the result of relatively consistent fields below or it could be an indicator of a need for a better bottom layer representation. Overall, the RMS error and PCC provide meaningful quantitative measures of forecast reliability. The dual use of data for assimilation and skill evaluation should become standard for every forecasting operation.
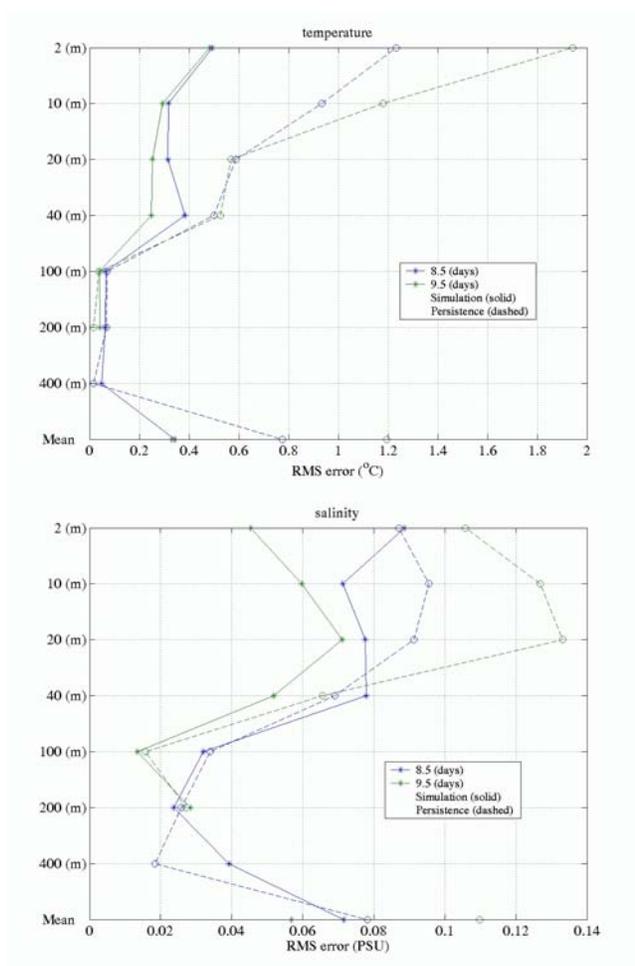

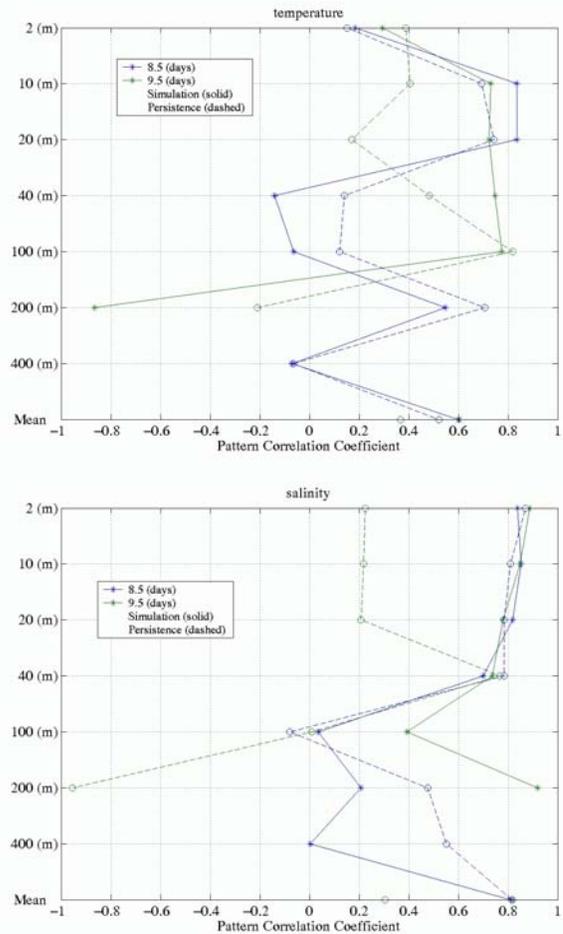Fig. 10. ASCOT-02 RMS Error – Temperature (top), Salinity (bottom)

Fig. 11. ASCOT-02 PCC - Temperature (top), Salinity (bottom)

REFERENCES

Besiktepe, S.T., P.F.J. Lermusiaux and A.R. Robinson, 2002. Coupled physical and biochemical data driven simulations of Massachusetts Bay in late summer: real-time and post-cruise data assimilation, Journal of Marine Systems, M. Gregoire, P.F.J. Lermusiaux and P. Brasseur (Eds.), submitted.

Briggs, W.M. and R.A. Levine, 1997. Wavelets and field forecast verification. Monthly Weather Review, 125, 6, 1329-1341.

Buizza, R. and T.N. Palmer, 1998. Impact of ensemble size on ensemble prediction. Monthly Weather Review, 126, 9, 2503-2518.

Carton, J.A., 1987. How Predictable are the Geostrophic Currents in the Recirculation Zone? Journal of Physical Oceanography, 17(6): 751–762.

Ghil, M. and N. Jiang, 1998. Recent forecast skill for the El Nino/Southern Oscillation. Geophysical Research Letters, 25, 2, 171-174.

Glenn, S.M. and A.R. Robinson, 1995. Verification of an Operational Gulf Stream Forecasting Model, in Quantitative Skill Assessment for Coastal Ocean Models, Coastal and Estuarial Studies, American Geophysical Union, 47, 469--499.

Haidvogel, D.B. and A. Beckmann, 1998. Numerical models of the coastal ocean. The Sea, Vol. 10, 466-478.

Krzysztofowicz, R., 1992. Bayesian correlation score: a utilitarian measure of forecast skill. Monthly Weather Review, 120, 1, 208-219.

Lermusiaux P.F.J., A.R. Robinson, P.J. Haley and W.G. Leslie, 2002. Advanced interdisciplinary data assimilation: Filtering and smoothing via Error Subspace Statistical Estimation. Proceedings of the "The OCEANS 2002 MTS/IEEE" conference, in press.

Lozano, C.J., A.R. Robinson, H.G. Arango, A. Gangopadhyay, N.Q. Sloan, P.J. Haley, L.A. Anderson and W.G. Leslie, 1996. An Interdisciplinary Ocean Prediction System: Assimilation Strategies and Structured Data Models, in Modern Approaches to Data Assimilation in Ocean Modeling (P. Malanotte-Rizzoli, editor), Elsevier Oceanography Series, 413--452, Elsevier Science, The Netherlands.

Lynch, D.R. (ed.), 1995. Quantitative Skill Assessment for Coastal Ocean Models. Coastal and Estuarine Studies, 47, American Geophysical Union.

Majewski, D., 1997. Operational regional prediction. Meteorology and Atmospheric Physics, 63, 1-2, 89-104.

Miller, A.J., P.-M. Poulain, A.R. Robinson, H.G. Arango, W.G. Leslie and A. Warn-Varnas, 1995. Quantitative Skill of Quasigeostrophic Forecasts of a Baroclinically Unstable Iceland Faeroe Front, Journal of Geophysical Research, 100(C6), 10833--10849.

Murphy, A.H., 1991. Forecast verification: its complexity and dimensionality. Monthly Weather Review, 119, 7, 1590-1601.

Murphy, A.H., 1996. General decompositions of MSE-based skill scores: measures of some basic aspects of forecast quality. Monthly Weather Review, 124, 10, 2353-2369.

Murphy, A.H. and E.S. Epstein, 1989. Skill scores and correlation coefficients in model verification. Monthly Weather Review, 117, 3, 572-581.

Robinson, A.R., 1996. Physical Processes, Field Estimation and Interdisciplinary Ocean Modeling, Earth-Science Reviews, 40(1/2), 3--54.

Robinson, A.R., S.M. Glenn, M.A. Spall, L.J. Walstad, G.M. Gardner, and W.G. Leslie, 1989. Forecasting Gulf Stream meanders and rings. EOS, 70(45), 1464-1473.

Robinson, A.R., H.G. Arango, A.J. Miller, A. Warn-Varnas, P.-M. Poulain, and W.G. Leslie, 1996a. Real-time operational forecasting on shipboard of the Iceland-Faeroe Frontal variability. Bulletin of the American Meteorological Society, 77(2), 243-259.

Robinson, A.R., H.G. Arango, A. Warn-Varnas, W.G. Leslie, A.J. Miller, P.J. Haley and C.J. Lozano, 1996b. Real-Time Regional Forecasting, Modern Approaches to Data Assimilation in Ocean Modeling (P. Malanotte-Rizzoli, editor), Elsevier Oceanography Series, 377--410, Elsevier Science, The Netherlands.

Robinson, A.R., P.F.J. Lermusiaux, and N.Q. Sloan, 1998. Data Assimilation. In THE SEA: The Global Coastal Ocean, Volume 10: Processes and Methods (K.H. Brink and A.R. Robinson, eds.), John Wiley and Sons, NY, 541-594, 1998.

Stamus, P.A., Carr, F.H., Baumhefner, D.P., 1992. Application of a scale-separation verification technique to regional forecast models. Monthly Weather Review, 120, 1, 149-163.

Tapp, M.C., 1986. A new simple measure of forecast accuracy. Weather, 41, 4, 135-136.

Ward, M.N. and A. Navarra, 1997. Pattern analysis of SST-forced variability in ensemble GCM simulations: Examples over Europe and the tropical Pacific. Journal of Climate, 10, 9, 2210-2220.

Willems, R.C., S.M. Glenn, M.F. Crowley, P. Malanotte-Rizzoli, R.E. Young, T. Ezer, G.L. Mellor, H.R. Arango, A.R. Robinson, and C-C. Lai, 1994. Experiment evaluates ocean models and data assimilation in the Gulf Stream. EOS, 75, 34.

Williamson, D.L., J.B. Drake, J.J. Hack, R. Jakob, and P.N. Swartztrauber, 1992. A standard test for numerical approximations to the shallow water equations in spherical geometry. J. Comp. Phys., 102, 211-224.

Xue, Y., Cane, M.A., Zebiak, S.E., Palmer, T.N., 1997. Predictability of a coupled model of ENSO using singular vector analysis. II. Optimal growth and forecast skill. Monthly Weather Review, 125, 9, 2057-2073.