Automation of Ocean Model Performance Metrics

James D. Dykes[1], Jay F. Shriver[1] and Sean Ziegeler[2]

[1]Naval Research Laboratory
Code 7320 Bldg 1009
Stennis Space Center, MS 39529 USA

[2]High Performance Technologies Inc. / HPCMP PETTT Program
NRL Code 7320, Stennis Space Center, MS 39529 USA

***Abstract* - A system to rapidly and automatically assess the performance of numerical ocean modeling systems was developed by the U.S. Naval Research Laboratory (NRL). This includes the calculation of quantitative, objective metrics of the accuracy of ocean forecasts. We will present results from this system, including metrics of surface and subsurface analysis and forecast fields. This work supports the U.S. Naval Oceanographic Office (NAVOCEANO), which provides oceanographic products in response to requests for environmental support for Navy operations. The development of a comprehensive automated system that provides model performance information is expected to increase the consistency of results, reduce errors, and reduce time required to generate oceanographic products.**

I. Introduction

A continual requirement exists to quickly and frequently evaluate the validity and accuracy of oceanographic data, models, algorithms, and products with performance metrics that are meaningful and applicable to the supported mission. Results from these evaluations will help make performance improvements to the model and products, better assess the ocean environment, and provide decision makers with an improved perspective on the ocean environment and the product. In addition to meeting operational needs, this work supports research, development, and evaluation of new analysis and forecast systems intended for operational use. The numerical models being assessed by this system have applications other than for Navy support, including providing high resolution boundary conditions for even higher resolution coastal models; tracking pollutants; managing fisheries and other marine resources; assessing ocean impacts on oil rigs and other structures; predicting storm surge resulting from hurricanes; and providing inputs to water quality assessment.

NRL has developed new core operational components that include the required algorithms, methodology, software, and guidance as follows: a) An automated system that creates, and stores the metrics of present and future ocean modeling analysis and forecast systems, in real-time and over longer space and time scales, b) A subset of specifically acoustic metrics for the evaluation of oceanographic data and models for mission support, and c) An automated system that facilitates data collection and provides metrics of user forecasts and the operational impacts of those forecasts. This paper will focus on the first of these three.

II. Methodology

Since environmental analyses and forecasts are highly dependent on numerical ocean models (e.g., Navy Coastal Ocean Model (NCOM)[1][2], and the HYbrid Coordinate Ocean Model (HYCOM) [3][4]), the ocean forecasters are interested in their accuracy. Some standard metrics are already produced in various capacities and are now being produced automaticly. Examples include time series comparisons, vertical profile comparisons, axis error of ocean features, anomaly correlation, RMS error, and skill score. Parameters or state variables of interest include temperature, salinity, currents, sonic layer depth, and sound velocity gradients. As a component of the Navy Coupled Ocean Data Assimilation (NCODA)[5] analysis and data quality control software (OCNQC), a regular feed of quality-controlled in-situ observations (e.g., XBTs, CTDs, profiling floats, glider data, and surface ship observations) is used.

Data structures and formats have been defined to facilitate database queries and analysis of model-observation and model-model comparisons. Observation files come in the OCNQC format and is publicly available on the Global Ocean Data Assimilation Experiment (GODAE)[6] server where the data and software is provided and maintained by NRL Monterey. The model output at NAVOCEANO is processed into netCDF using a standard convention based on COARDS as published by University Corporation of Atmospheric Research (UCAR). A convention in netCDF that can handle atmospheric, ocean and wave model output makes the processing of model output highly flexible. Software has been designed to support frequent data processing (multiple data cuts per day) and multiple-nested models. Routines for generating automated evaluations of model forecast statistics have been developed and pre-existing tools have been collected to create a generalized tool set, which included user-interface tools to the metrics data.

An automated system was installed on the DoD High Performance Computing (HPC) machines of the Navy DoD Supercomputer Resource Center (DSRC) where the models whose performance is to be monitored resides. Once the system is set

0-933957-38-1/09/$20.00

up for a list of models to be processed, it runs fully automatically without human intervention. Here the routines that compile model-observations and model-model comparisons are run in real-time, using software compiled in C and tested on multiple platforms. As the database of comparisons accumulate, the latest files are transferred to the user spaces at NAVOCEANO where ocean forecasters use the data to aid in interpreting ocean model nowcasts and forecasts. A schematic of the autometrics system is shown in Figure 1.

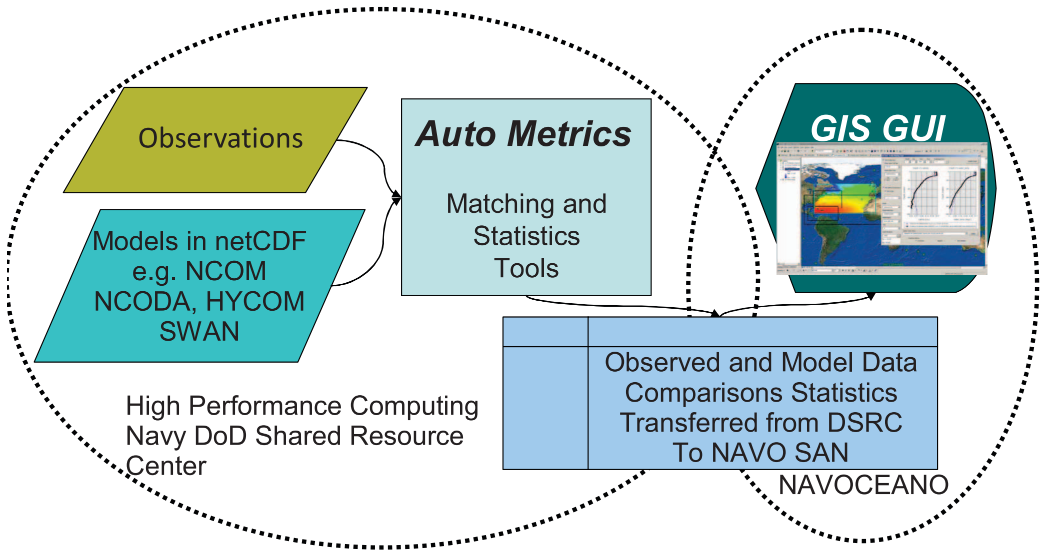

Figure 1: Schematic of the automated metrics system.

The model-observations comparison database consist of files of matches, a.k.a. matchups, between modelled and observed data. For every observation point in geographic location and forecast cycle time there corresponds modelled values linearly interpolated in time and space. For observation profiles e.g. CTD and XBT, the observed values are interpolated to the modelled levels using a piece-wise hermite polynomial scheme, although many times the vertical density of observed data is so high compared to the model level density, that the resulting observed interpolated values to model levels amounts to subsampling. For now glider observations are reported as profiles, but later the format of glider data will be considered the most general case of finding an individual point in the four-dimensional space, x, y, z, and t, corresponding to longitude, latitude, level and time. Usually, the profiled observations are assimilated into the model analysis but they can easily be excluded without making a huge impact on model performance while providing a source of independent data.

The measured parameters from observation profiles includes temperature, salinity and bottom depth. The files we use were processed in the OCNQC routines and thus include data quality control flags to allow us to decide the criteria for excluding data. The routine that builds the matchups database also computes sound speed based on the observed and modelled profiles results of which are added to the database of matched up data with which to compute statistics. In addition, selected acoustics characteristics of the ocean such as sonic layer depth (SLD) are computed based on the sound speed profile and added to the matchup database. These acoustic parameters are computed based on each of the original, uninterpreted observation profiles and on the modelled profiles using the model levels.

Generally, the coverage of observation profiles is not as complete as that of remotely sensed data, which includes a systematic report in swathes of brightness temperature and regular tracks of highly accurate altimeter height measurements. Matchups could be accomplished with these data, biut the database would quickly become unmanageably large for timely and practical processing. Another disadvantage is that only surface data is available. In addition, these data almost always assimilated into the model analysis where synthetic profiles are derived based on these data.

Having processed remotely sensed data into the analysis, i.e. the intialization of a model forecast run, one can use the analysis as a basis for model performance by see what the forecasts have done to look like the analysis valid for the same time. Model-model comparisons are important since they help the oceanographer forecaster get a sense of the model performance as well as the rate of change of the environmental conditions. The simple difference between two fields, say, the current temperature analysis and the 24-hour forecast, provide a sense that the model is “behaving” well. Further, a collection of these over time contributes to statistics that reveal model tendencies. The mean differences and RMS differences for each of the grid points over a long enough time may provide a statistical significant result distinguishing the model performance spatially over the domain, i. e. the spatial information would reveal where in a certain domain does the model better handle the physics.

Whether it be model-model or observation-model comparison-based statistics, a reference to the level of model skill is needed to help give the model evaluator a sense of the level of performance. Typically, statistics of forecasts based on model skill are compared to the statistics of using persistence as a predictor. Persistence takes current conditions and predicts that this will be the same in the future without any other consideration. Arguably, where the environment clearly changes very slowly a forecast for conditions in 72 hours can be a fairly reasonable predictor, in which case we may be actuallly regarding a climatological feature. However, in regimes of rapidly changing conditions, persistence is expected to be a very bad predictor. The hope is that in either case a forecast with added skill beyond simply maintaining the same value should "beat" persistence, i.e. the statistics of the comparisons between ground truth and forecasts should be better than comparisons of persistent condition to reality. If this were not the case, than this implies that the subject forecasting method, the model, has little skill. The matchup system utilized in this automated system is also implemented to compare all the ground truth for a certain forecast period to the initial state variables, producing a database of the same size as the matchups based on the interpolating in space and time for model output forecasts. The statistics for both are then compared.

A one-for-one comparison of gridded data does not tell the whole story. Besides determining the error at a fixed position, it is also important to determine the displacement error (i.e., how far is a forecasted feature from its nowcasted location). Automated displacement error algorithms (both magnitude and direction) have been developed and implemented to assess forecasted feature placement accuracy and is explained in the following way. The displacement vector field is generated using a deformable registration method[7] . A two-dimensional cubic B-spline mesh is imposed over the forecast data set. Each control point of the mesh can be adjusted in the x or y direction, and each adjustment produces a smooth distortion. An advanced gradient-descent optimization routine iteratively chooses the adjustments to improve the squared-errors between the forecast and analysis data sets. The B-spline mesh can be transformed into a displacement field with a vector at each data point.

As a further improvement, the error displacement are constructed using the gradients of the scalar fields. The gradient reduces the influence of regional biases. For example, if an eddy feature were warmer in the analysis but remained in the same location, the gradient would not change as much. This approach utilizes a Gaussian-smoothed gradient, which widens the high-gradient features making it easier to track them from one data set to the next.

III. Results

Processing of data and model comparisons are run at the Navy DoD Supercomputing Resource Center (DSRC) and the results of the comparisons and statistics calculation are visualized using GIS GUIs and tools within the client/viewer for ocean product performance metrics system depicted in Figure 2. Also, model performance metrics in the form of statistics of model vs. observations can be displayed in a window of the client/viewer as depicted in Figure 3. From the database of comparisons between model output and observation profiles the paired match-up profiles are grouped in 24-hour segments. Their mean differences, RMS errors, and correlations are computed for each model run for each model level. Several model runs are displayed in a series providing an indicator of model performance trends. For example, the RMS error of the 0-24 hour forecasts are usually less than the 24-48 hour RMS error, which is expected. Also, from these graphics comparisons between model and observed values that are very different can be easily identifed and even rooted out if warranted.

An additional feature within the GUI tool where model performance statistics are displayed is the ability to display the profile matchups as they are selected in the scatterplot (Figure 4). This is particularly useful when the investigation of extreme observations are warranted. This allows the visual determination of valid observations to catch those rogue values that were not caught by OCNQC. This also allows for the elimination of selected observations from the loaded feature set and the recomputation of statistics.

Figure 5 shows an example of the results from the automated displacement error algorithms (both magnitude and direction) which assess forecasted feature placement accuracy. This display compares forecasted with analysis sea surface temperature output from NCOM. The vectors in Figure 5(all identical) represent the relative movement of data points from a 24-hour forecast (a,c) to an analysis the next day (b,d). This provides complimentary information to that of a simple point-wise difference between the two. The latter results in error values in the unit space of the scalar (e.g., temperature). This novel method results in error quantities in spatial units (e.g., degrees of latitude/longitude). Figures 5(c,d) show the smoothed gradient of (a,b), respectively. Visually, the changes in feature positions are more apparent when looking at (c,d). The next step is to assess model feature placement in comparison to observations, and this work is ongoing.

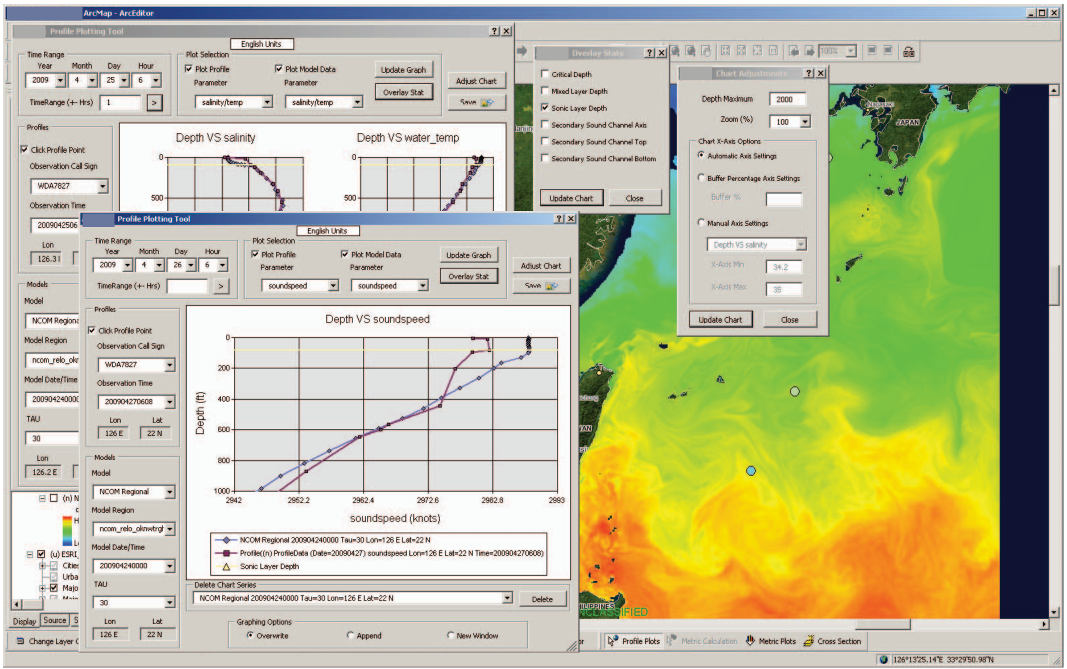

Figure 2: Example of display in the client/viewer used by oceanographers at NAVOCEANO. A point within that same domain was selected to present the profiles of models and observed temperature, salinity and soundspeed at that point.

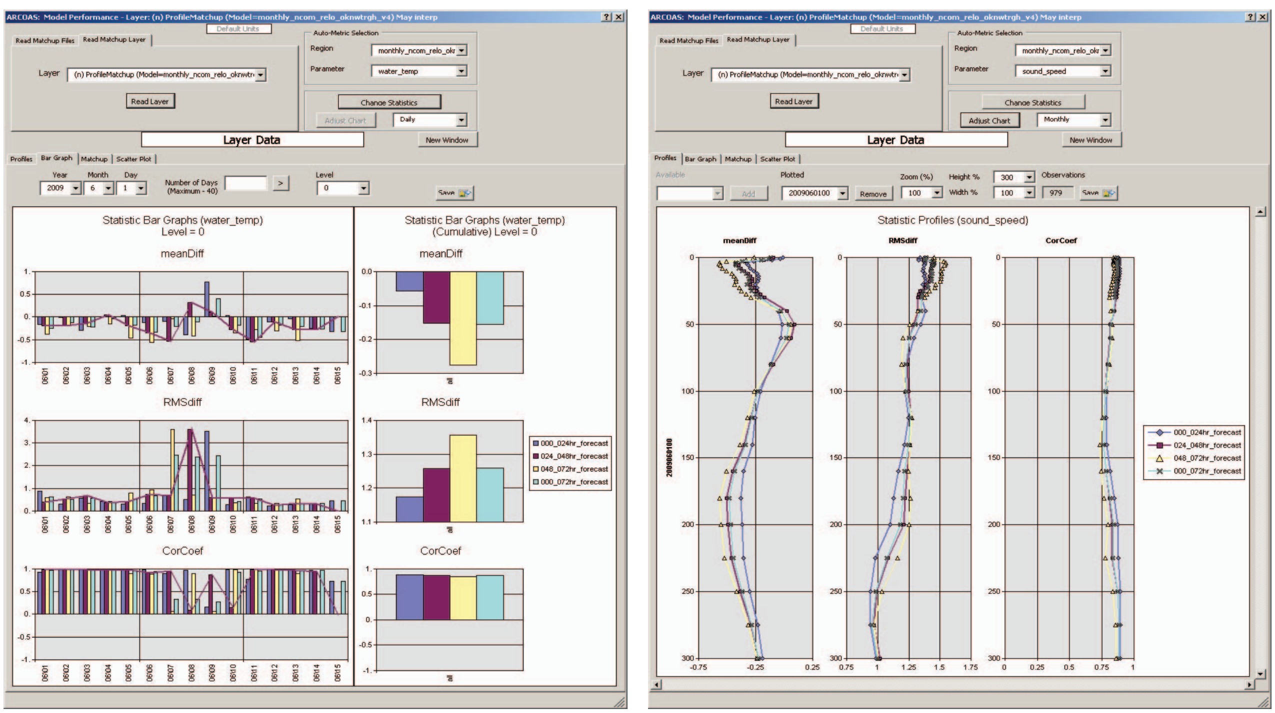

Figure 3: Display of statistics of comparisons between model and observation profiles as summary bar graphs and profile plots within the metrics client/viewer used by oceanographers at NAVOCEANO. In this case the bar graphs revealed abarrent model-obervation comparisons for a few model runs during the month. The observations were discovered to be bad.

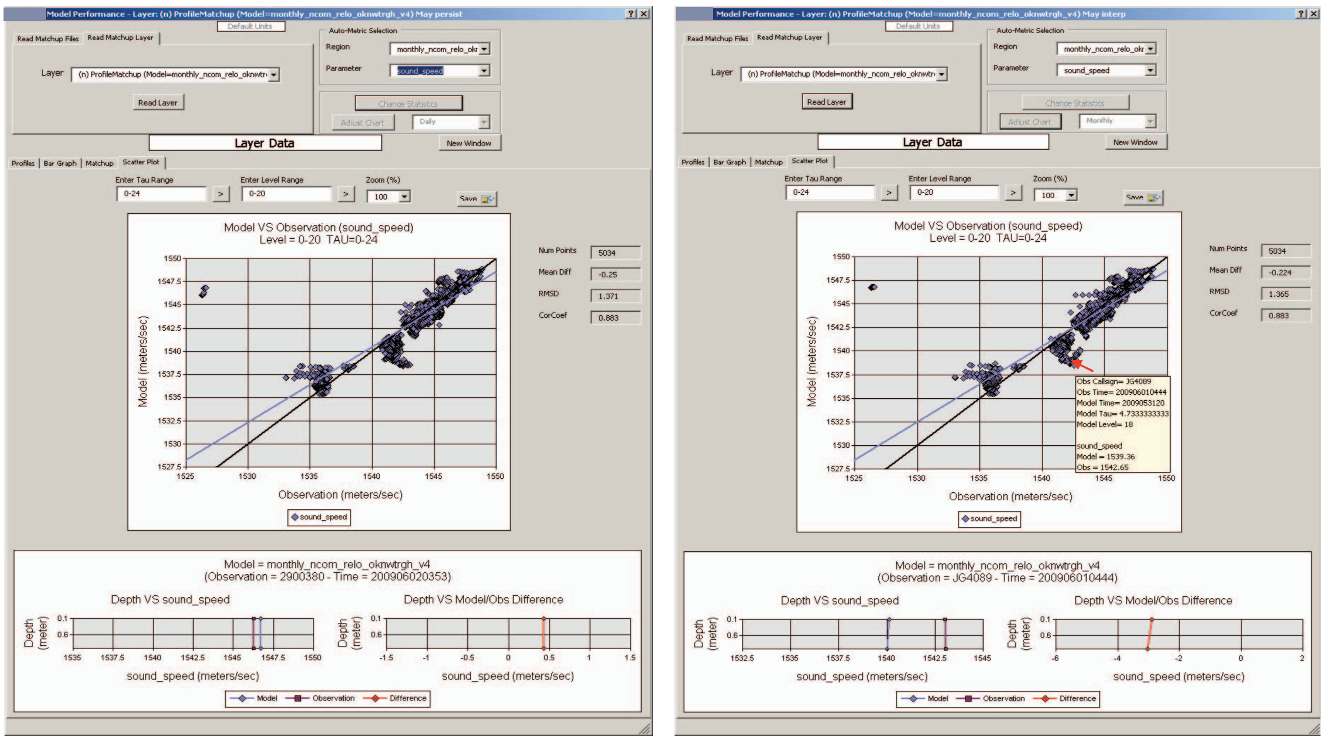

Figure 4: Display of scatter plots of model observation comparisons, on the left for persistence matchups and on the right for interpolated model forecasts matchups. These two are displayed side-by-side to determine with the model has the skill to predict better than persistence.

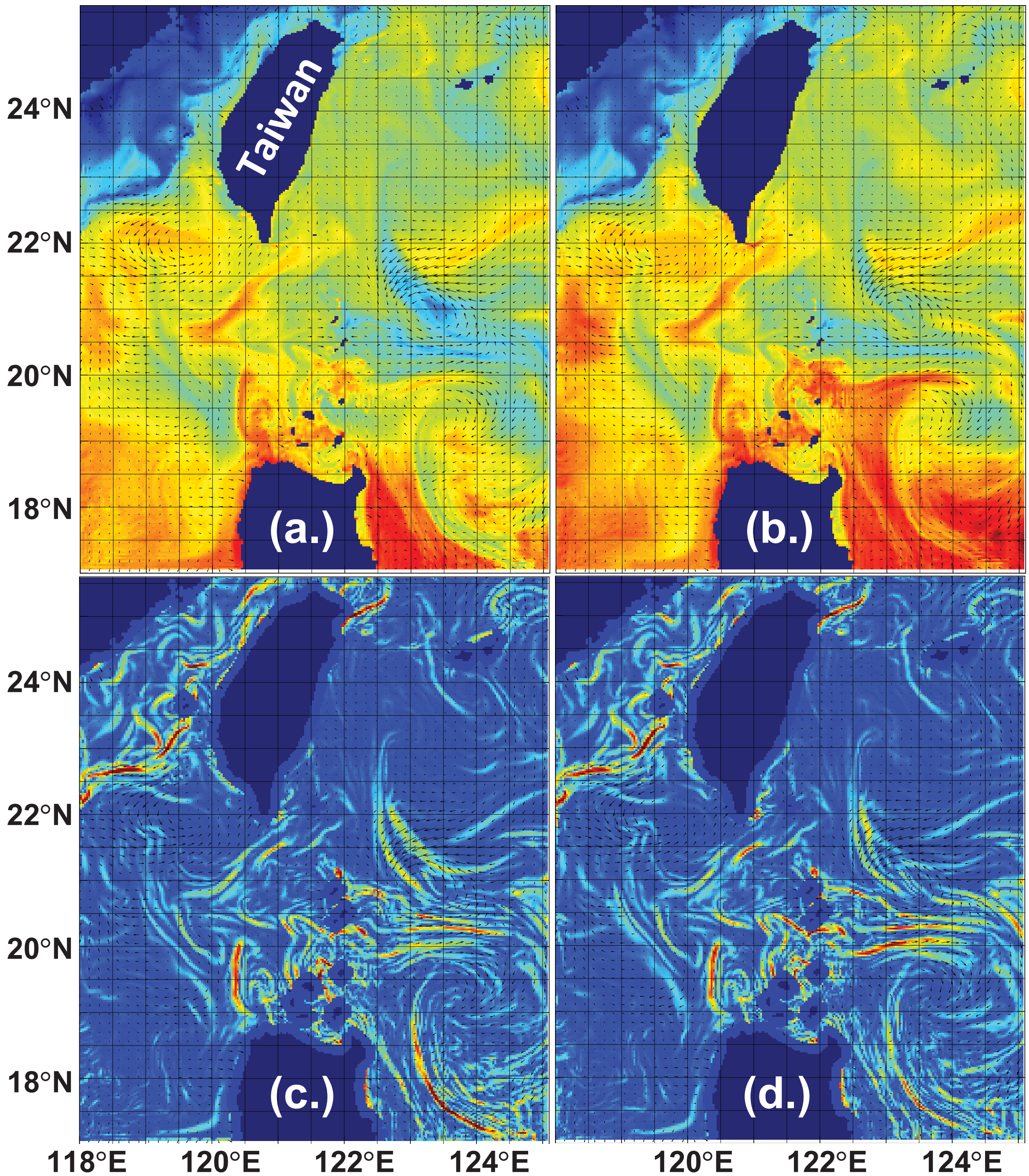

Figure 5: The vectors in (a-d) are identical and represent spatial displacement to scale. (a) 24-hour forecast of temperature in °C, (b) Model analysis the next day, (c) Smooth gradient of the 24-hour temperature forecast, (d) Smooth gradient of model analysis with all panels ranging from blue (low) to red (high).

Acknowledgments

This work was sponsored by the Office of Naval Research (program element 0602435N) as part of the project "Automation of Ocean Product Performance Metrics". The computations in this paper were completed utilizing Defense Department High Performance Computing time at the Navy DSRC.

References

[1] Halliwell G. R., 2004: Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid-Coordinate Ocean Model (HYCOM), Ocean Modelling, Volume 7, Issues 3-4, 285-322.
[2] Chassignet, E.P., L.T. Smith, G.R. Halliwell, and R.Bleck. 2003: North Atlantic simulation with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference density, and thermobaricity. Journal of Physical Oceanography 33:2,504–2,526.
[3] Martin, P.J., 2000: Description of the Navy Coastal Ocean Model Version 1.0. NRL/FR/7322—00-9962, Naval Research Laboratory, Stennis Space Center, MS
[4] Barron, C.N., A.B. Kara, P.F. Martin, R.C. Rhodes, L.F. Smedstad, 2006: Formulation, implementation and examination of vertical coordinate choices in the Global Navy Coastal Ocean Model (NCOM). J. Coastal Eng. 11, 347 – 375.
[5] Cummings, J.A. 2005. Operational multivariate ocean data assimilation. Quarterly Journal of the Royal Meteorological Society 131:3,583–3,604.
[6] Smith, N. and M. Lefèbvre, 1997: The global ocean data assimilation experiment (GODAE), in Proc. Int. Symp. Monitoring Oceans in the 2000s: An Integrated Approach Biarritz, France.
[7] Rueckert, D., et al, 1999: Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Transactions on Medical Imaging, 18:8, 712-721.