THE OFFICIAL MAGAZINE OF THE OCEANOGRAPHY SOCIETY

Oceanography

DEVELOPMENT OF A HINDCAST/FORECAST MODEL FOR THE PHILIPPINE ARCHIPELAGO

BY HERNAN G. ARANGO, JULIA C. LEVIN, ENRIQUE N. CURCHITSER, BIN ZHANG, ANDREW M. MOORE, WEIQING HAN, ARNOLD L. GORDON, CRAIG M. LEE, AND JAMES B. GIRTON

ABSTRACT. This article discusses the challenges of developing a regional ocean prediction model for the Philippine Archipelago, a complex area in terms of geometry, bathymetry-dominated dynamics and variability, and strong local and remote wind forcing, where there are limited temporal and spatial ocean measurements. We used the Regional Ocean Modeling System (ROMS) for real-time forecasting during the Philippine Straits Dynamics Experiment (2007–2009) observational program. The article focuses on the prediction experiments before and during the exploratory cruise period, June 6–July 3, 2007. The gathered observations were not available in real time, so the 4-Dimensional Variational (4D-Var) data assimilation experiments were carried out in hindcast mode. The best estimate of ocean state (nowcast) is determined by combining satellite-derived products for sea surface temperature and height, and subsurface temperature and salinity measurements from several hydrographic assets over a sequential five-day data assimilation window. The largest source of forecast uncertainty is from the prescribed lateral boundary conditions in the nearby Pacific Ocean, especially excessive salt flux. This result suggests that remote forcing and inflows from the Pacific are crucial for predicting ocean circulation in the Philippine Archipelago region. The lateral boundary conditions are derived from 1/12° global HYbrid Coordinate Ocean Model (HYCOM) daily snapshots. The incremental, strong-constraint 4D-Var data assimilation successfully decreased temperature and salinity errors of the real-time, nonassimilative control forecast by 38% and 49%, respectively.

INTRODUCTION

Some of the greatest challenges in real-time ocean forecasting are posed by the lack of continuous surface and subsurface observations and of reliable predictive atmospheric forcing. Usually, ocean prediction is needed next to coastal areas for various scientific and operational purposes, requiring the use of regional, fine spatial- and temporal-resolution models. These models, in turn, require lateral boundary conditions from larger-scale basin or global forecasting models. Because the main goal in forecasting is to predict circulations and events that have not yet been observed, it is essential to have the best estimate of the atmospheric/ocean state (nowcasts) to initialize such models. The best state circulation estimate is achieved with advanced data assimilation techniques that combine models (the *prior*) and observations, taking into account *a priori* hypotheses about errors in initial conditions, boundary conditions, surface forcing, and imperfect model physics (Moore et al., 2011a). There are extensive, in-depth scholarly reviews of data assimilation in meteorology and oceanography, including Bengtsson et al. (1981), Tarantola (1987), Daley (1991), Ghil and Malanotte-Rizzoli (1991), Bennett (1992, 2002), and Wunch (1996).

The length of ocean forecasts in uncoupled prediction systems is limited by the availability, frequency, and delivery of real-time atmospheric products. Regional, fine temporal- and spatial-resolution atmospheric models exist in few places globally and are usually available every 6, 12, or 24 hours and confined to two- to three-day forecasts. Alternatively, medium-range atmospheric forecasts are available for longer periods (7–15 days), but at coarser temporal and spatial resolution. For example, the National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS) products are available every 12 hours at 2.5° spatial resolution and for a 15-day forecast period (Kleist et al., 2009). Additionally in 4-Dimensional Variational (4D-Var) data assimilation, the length of the forecast cycle is limited by the validity of the tangent linear assumption. The growth of linearly unstable modes can lead to spurious circulations (Moore et al., 2004, 2009, 2011b). In coastal ocean applications, the tangent linear assumption depends on the grid resolution and nonlinear dynamics and is usually valid from three to 30 days.

In this paper we present our first attempt at building a regional ocean prediction model for the Philippine Archipelago based on the Regional Ocean Modeling System (ROMS) and its comprehensive 4D-Var data assimilation algorithms (Moore et al., 2011a,b,c). This work was part of the Philippine Straits Dynamics Experiment (PhilEx), sponsored by the Office of Naval Research, with the goal of improving our capability to predict the inherent spatial and temporal variability near the Philippine straits, and thus contribute to the development of reliable prediction systems. The PhilEx program involved scientists from various institutions and included both field observations and numerical modeling. The gathered observations were not available in real time due to logistics and the exploratory nature of the program. Therefore, only real-time forecasts without data assimilation were carried out in support of the exploratory cruise (June 6–July 3, 2007), the Joint US/Philippines Cruise (November 22–December 30, 2007), and the two regional Intensive Observational Period cruises (IOP-08 from January 9 to February 1, 2008, and IOP-09 from February 27 to March 21, 2009). Each forecasting cycle, updated daily, was run for nine days (four-day hindcast and five-day forecast). The model was initialized four days prior to the forecast cycle starting day to use reanalyzed atmospheric and boundary forcing. These real-time forecasts were issued and posted at http://www.myroms.org/philex. The issued forecasts were analyzed later when the collected data were processed and made available. Systematic errors and biases were found in the forecasted temperature and salinity fields. Here, we present the corrections to such errors and biases by combining observations and models using a primal form of the incremental strong constraint 4D-Var (I4D-Var) in hindcasting mode. In particular, we concentrate on the summer 2007 exploratory cruise period.

PHILIPPINE ARCHIPELAGO

The Philippine Archipelago region is characterized by complex topography and a collection of seas connected by many straits and passages (see Figure 1). The dynamics in the archipelago are also complex due to interactions among circulation (mean and strong tidal flows), strong local and remote seasonal forcing, passage constrictions, overflow across topographic sills, and ventilation of internal seas (Han et al., 2009; Gordon et al., 2011; Hurlburt et al., 2011; May et al., 2011). At the surface, the archipelago is subject to strong, seasonally reversing monsoon wind forcing. Through the lateral boundaries, it is influenced by the fresher water inflow from the South China Sea (SCS) via Mindoro Strait (sill depth ~ 420 m) and Balabac Strait (south of Palawan and Balabec islands; sill depth < 50 m) and by the neighboring saltier tropical Pacific Ocean to the east (Han et al., 2009; Gordon et al., 2011).

Han et al. (2009) used a ~ 5 km ROMS model to estimate the relative importance of local, remote, and tidal forcing on the dynamics of the region. They found that local winds play a crucial role in archipelago surface circulation during the peak monsoon periods. However, the simulations show that the remote forcing dominates the seasonal variations of transports above 40 m across all the major straits and passages of the region, including Mindoro, Balabac, Dipolog, and Tablas straits.

MODEL AND CONFIGURATION

Several separate ROMS grids of varying geographical extent and resolution (~ 15 km, ~ 10 km, ~ 5 km, and ~ 2 km) have been developed for the Philippine Archipelago for annual, seasonal, and forecasting simulations. Only the 5-km and 2-km grids have been one-way nested. ROMS is a mature numerical framework used routinely by both the scientific and operational communities to study ocean dynamics over a wide range of spatial (estuaries to basin) and temporal (days to seasons, years to decades) scales. It is unique because it is the only ocean community framework to include the adjoint-based analysis and prediction tools that are available in Numerical Weather Prediction (NWP), such as 4D-Var data assimilation (Moore et al., 2011a), adjoint sensitivity analysis (Moore et al., 2009), ensemble prediction (Powell et al., 2008; Javier Zavala-Garay, Rutgers University, *pers. comm.*, 2010), observation impact and sensitivity (Moore et al., 2011c), adaptive sampling (Zhang et al., 2010b), and generalized linear stability analysis of the circulation (Moore et al., 2004).

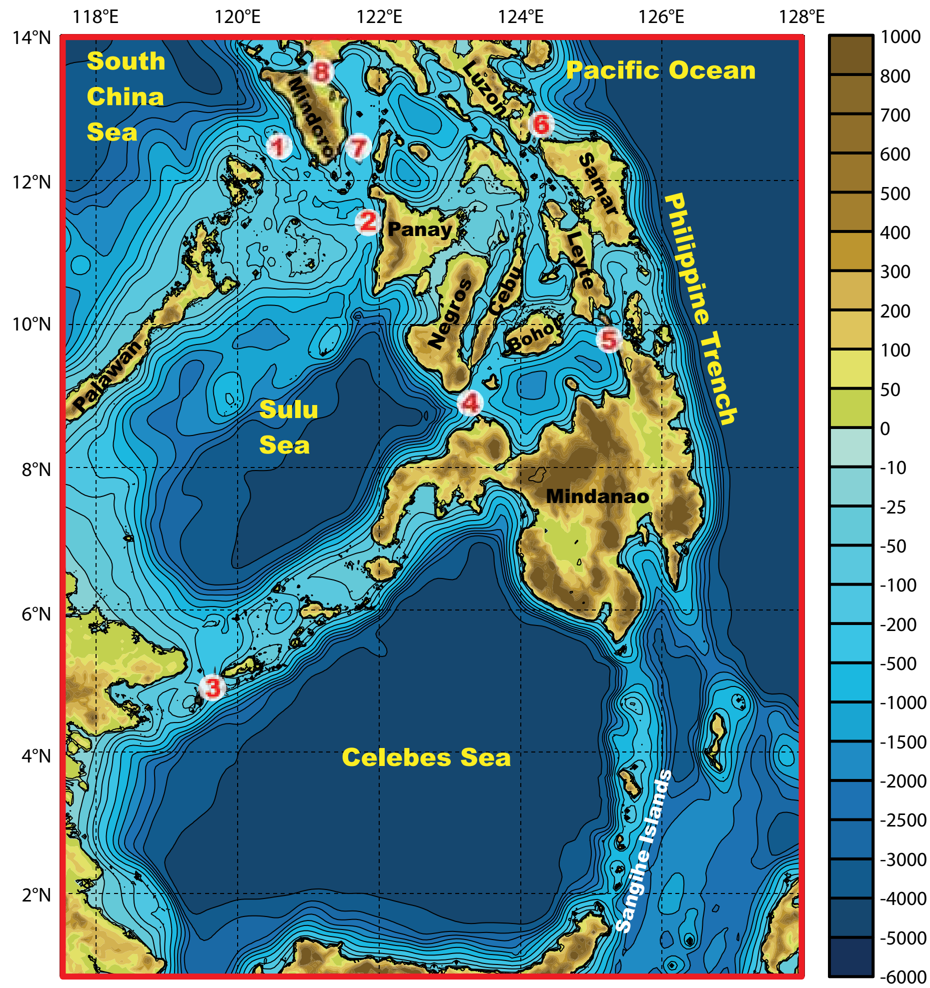

Figure 1. Philippine Archipelago bathymetry (m) and elevation (m). The model grid is indicated by the red box and has ~ 5.5-km average resolution. The major straits and passages with approximated actual sill depth are also shown in a counterclockwise order: (1) Mindoro Strait ~ 420 m, (2) Panay Strait ~ 570 m, (3) Sibutu Passage ~ 320 m, (4) Dipolog Strait ~ 504 m, (5) Surigao Strait ~ 60 m, (6) San Bernardino Strait ~ 80 m, (7) Tablas Strait ~ 565 m, and (8) Verde Island Passage ~ 70 m.

The dynamical kernel of ROMS solves the three-dimensional, free surface, Reynolds-averaged, Navier-Stokes primitive equations using the hydrostatic vertical momentum balance and Boussinesq approximation (Haidvogel et al., 2000, 2008; Shchepetkin and McWilliams, 2005, 2009). The governing dynamical equations are discretized on a vertical terrain-following coordinate system. The horizontal coordinates are orthogonal and curvilinear, allowing Cartesian, spherical, and polar spatial discretization on an Arakawa C-grid. Its numerical kernel includes accurate and efficient algorithms for time-stepping, advection, pressure gradient (Shchepetkin and McWilliams 2003, 2005, 2009), several subgridscale parameterizations (Durski et al., 2004; Warner et al., 2005), and various bottom boundary layer formulations to determine the stress exerted on the flow by the bottom.

Figure 1 shows the ROMS grid, indicated by the red box, used in the results presented here. It extends from 117°30'E to 128°E and 0°48'N to 14°N and has an average resolution of 5.5 km. The model bathymetry is interpolated from the Smith and Sandwell (1997) data set, which has a one-minute longitude/latitude resolution, and smoothed to suppress the systematic computational errors in the discretization of the horizontal pressure gradient force (Shchepetkin and McWilliams, 2003). The vertical coordinate is discretized into 42 levels; level thicknesses depend on local water depth. In the deepest parts of the Sulu and Celebes basins, around 10 levels make up the top 200 m of the water column, which is sufficient to model surface processes. The vertical resolution is much finer across the important straits and passages and adequate to resolve the overflows.

The initial and lateral open-boundary conditions for free-surface, temperature, salinity, and momentum are derived from the 1/12° global HYbrid Coordinate Ocean Model (HYCOM; Bleck, 2002), which includes assimilation of satellite sea surface temperature, along-track altimetry data, and available in situ temperature and salinity observations via the Navy Coupled Ocean Data Assimilation (NCODA) system (Cummings, 2005). Atmospheric forcing is from the Navy's Operational Global Atmospheric Prediction System (NOGAPS) three-hour, half-degree resolution. It includes near-surface wind, air temperature, air pressure, humidity, shortwave and longwave radiation, and precipitation.

The model is configured with four open boundary conditions. All the boundary conditions for temperature, salinity, and three-dimensional momentum are prescribed and interpolated in space and time at every time-step from the HYCOM daily snapshots. Additionally, the HYCOM interpolated values for free-surface and vertically integrated momentum plus tides are used as inflow in the Flather (1976) radiation boundary conditions for these fields. Tidal amplitudes and currents are derived from the Oregon State University Tidal Prediction Software (OTPS) from Egbert et al. (1994) and Egbert and Erofeeva (2002). It includes four semidiurnal (K_2, S_2, M_2, N_2) and four diurnal (K_1, P_1, O_1, Q_1) tidal components.

Ocean surface turbulent fluxes for momentum, heat, and moisture are computed using an air-sea boundary layer formulation based on the bulk parameterization Fairall et al. (2003) adapted from the Coupled Ocean-Atmosphere Response Experiment (COARE) algorithm for the computation of surface wind stress, sensible heat, latent heat, and evaporation minus precipitation. This algorithm uses, as input, atmospheric products from NOGAPS interpolated at every time-step from snapshots of appropriate fields. The generic length scale turbulence closure (Warner et al., 2005), configured as *k-kl*, is used to parameterize vertical mixing processes. A quadratic drag formulation is used to parameterize bottom momentum stress. The quadratic drag coefficient is 0.003.

OBSERVATIONS

The Philippine Archipelago is probably one of the least-observed regions of the world ocean. Frequent cloud cover in the archipelago restricts the use of remote-sensing products derived from satellite data, such as sea surface temperature (SST) and ocean color. Synoptic maps of SST are usually a blend of high-resolution thermal infrared measurements over several days, coarser-resolution microwave data (which is

Hernan G. Arango *(arango@marine.rutgers.edu) is Research Associate, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA.* **Julia C. Levin** *is Assistant Research Professor, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA.* **Enrique N. Curchitser** *is Associate Research Professor, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ, USA.* **Bin Zhang** *is Postdoctoral Fellow, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.* **Andrew M. Moore** *is Professor, Ocean Sciences Department, University of California, Santa Cruz, CA, USA.* **Weiqing Han** *is Associate Professor, Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA.* **Arnold L. Gordon** *is Professor, Earth and Environmental Sciences and Associate Director, Ocean and Climate Physics, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA.* **Craig M. Lee** *is Principal Oceanographer and Associate Professor, Applied Physics Laboratory, University of Washington, Seattle, WA, USA.* **James B. Girton** *is Affiliate Assistant Professor, Applied Physics Laboratory, University of Washington, Seattle, WA, USA.*

unaffected by clouds), and first-guess climatology. The sea surface height (SSH) anomaly derived from satellite altimetry measurements is contaminated by the archipelago geometry and is only reliable in the Sulu and Celebes seas. Finally, historical hydrographic data within the archipelago are limited, and global climatology databases are biased due to the presence of multiple connected islands during gridding, often causing seawater properties from different provinces to merge across landforms. Therefore, building a continuous forecasting system in the Philippine Archipelago is extremely challenging due to the scarcity of ocean observations and their limitations. The extensive data collected within the archipelago during the PhilEx field program provide a unique opportunity to investigate the observational requirements needed to maintain such a forecasting system.

The observations that are available from various sources are shown in Figure 2 and include:

a. Daily gridded maps of sea level anomaly at 1/3° horizontal resolution from altimetry products produced by Ssalto/Duacs (Segment Sol multimissions d'ALTimétrie, d'Orbitographie et de localisation précise/Data Unification and Altimeter Combination System), and distributed by Aviso (Archiving, Validation and Interpretation of Satellite Oceanographic data), with support from the Centre National d'Études Spatiales (CNES). It is a merged product using all altimetry (Jason-1, Envisat, and Geosat Follow-on [GFO]) measurements relative to the mean dynamic topography (MDT) data estimated by Rio et al. (2005). The thermosteric signal was removed using the method of Willis et al. (2004). The map shown in Figure 2a is for June 6, 2007. Notice that data are only available in the Sulu, Celebes, and South China seas, and the Pacific Ocean to the east (see Figure 2a). In the sequential 4D-Var experiments considered here, the mean surface height is obtained from a multiyear integration of ROMS and the model-derived tidal signals are added to the satellite anomalies. Consequently, the high-frequency tidal signal, which is not constrained by the altimetry product, is not contaminated during assimilation. An observational/representativeness error of 4 cm is assigned to SSH in the 4D-Var estimates.
b. Satellite-derived SST products from the Open-source Project for a Network Data Access Protocol (OPeNDAP) catalog maintained by the National Oceanic and Atmospheric Administration (NOAA) CoastWatch Program. Figure 2b,c shows a couple of SST estimates for June 6, 2007. Figure 2b is an optimally interpolated (OI) SST at 10-km horizontal resolution that blends the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) aboard the National Aeronautics and Space Administration (NASA) Aqua spacecraft, the TRMM Microwave Imager (TMI) aboard NASA's Tropical Rainfall Measuring Mission (TRMM) satellite, and the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the NASA Aqua and Terra spacecrafts. A data comparison of this SST product against temperature data from the exploratory cruise gives a root mean square error (RMSE) of 0.5°C. Figure 2c shows an experimental five-day composite SST product at 0.1° horizontal resolution derived from microwave and infrared sensors aboard multiple platforms, including AMSR-E and MODIS carried by the Aqua satellite, Advanced Very-High Resolution Radiometers (AVHRR) aboard the NOAA Polar-orbiting Operational Environmental Satellites (POES), and imagers aboard the NOAA Geostationary Operational Environmental Satellites (GOES). No attempt is made to objectively analyze the blended SST product as in Figure 2b. Notice the lack of data in the Sulu Sea and warmer SST along Mindoro. These observations indicate that the first guess for the OI in Figure 2b is older than five days and the values have been persisted to create a complete map without gaps. Data from single-satellite sensors like MODIS are also available but are obscured by the frequent clouds over the Philippine Archipelago. The SST estimates from MODIS (not shown) are much warmer than those from the other products, yielding a 0.75°C RMSE when compared with the cruise data. The OI-blended SST product is used in the 4D-Var experiments shown here with representativeness standard deviation of 0.8°C.
c. Hydrographic measurements from the PhilEx exploratory cruise (June 6–July 3, 2007) are used for assimilation and verification. Not all the data are assimilated. Figure 2d shows the station locations (red symbols) from R/V *Melville* (Gordon et al., 2011), ElectroMagnetic-Autonomous

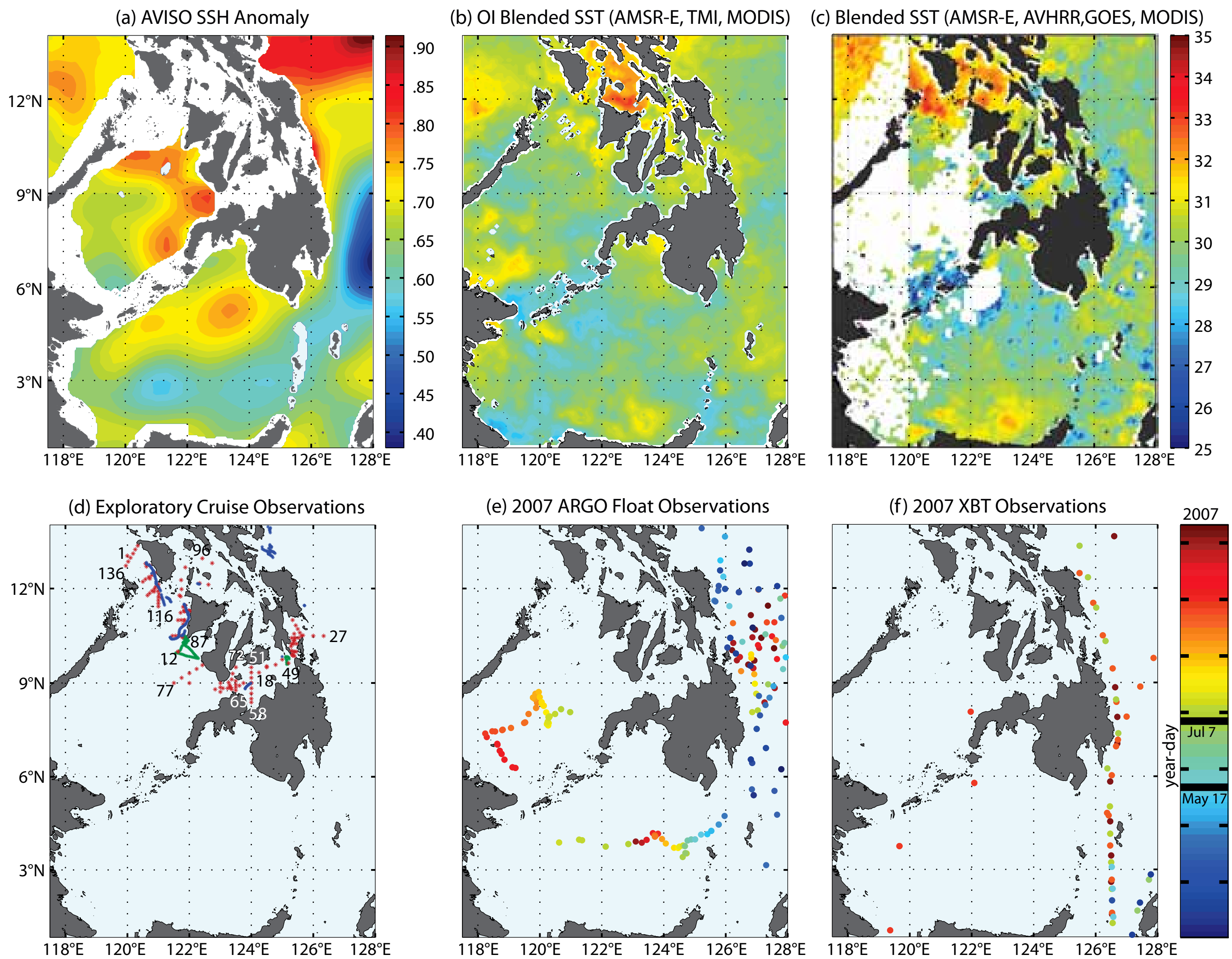

Figure 2. Observation data sets available for assimilation. (a) Sample of gridded, satellite-derived sea surface height (SSH) anomaly from Aviso (Archiving, Validation and Interpretation of Satellite Oceanographic data) for June 6, 2007. (b) Sample of gridded optimally integrated (OI) sea surface temperature (SST) blend from microwave (AMSR-E and TMI) and infrared (MODIS) satellite sensors for June 6, 2007. (c) Sample of experimental five-day SST composite using data from microwave (AMSR-E) and infrared (MODIS, AVHRR, POES, GOES) sensors for June 6, 2007. (d) Exploratory cruise track, June 6–July 3, 2007, showing conductivity/temperature/depth (CTD) stations (red), EM-APEX profiling float trajectories (blue), and two glider tracks (green). (e) Argo float profiles colored by 2007 year-day. (f) Expendable bathythermograph (XBT) temperature profiles colored by 2007 year-day. AMSR-E = Advanced Microwave Scanning Radiometer for EOS. TMI = NASA Tropical Rainfall Measuring Mission Microwave Imager. MODIS = Moderate Resolution Imaging Spectroradiometer. AVHRR = Advanced Very-High Resolution Radiometer. POES = NOAA Polar-orbiting Operational Environmental Satellites. GOES = Geostationary Operational Environmental Satellites. EM-APEX = ElectroMagnetic-Autonomous Profiling Explorer.

Profiling EXplorer (EM-APEX) float trajectories (blue symbols) near Panay and Mindoro islands (Girton et al., 2011), and glider measurements (green symbols) between Panay and Negros islands and near Surigao Strait. Only temperature and salinity observations are assimilated with a representativeness error of 0.1°C and 0.02, respectively. These values were scaled with depth to take into account the variations of the observations error covariance in the water column.

d. The few temperature and salinity profiles available from Argo floats. Figure 2e shows 2007 Argo float observations obtained from the quality-controlled UK Met Office EN3 data set (Ingleby and Huddleston, 2007). The symbols are colored according to 2007 year-day. There are only a few observations in the Sulu and Celebes seas. Most observations are confined next to the Philippines Trench in the Pacific Ocean. For the hindcast period considered here, there are four and 11 profiles (light-blue and greenish circles) in the Sulu and Celebes seas, respectively.

e. Expendable bathythermograph (XBT) profiles obtained from the ship-of-opportunity program. Figure 2f shows 2007 XBT temperature profile locations from the UK Met Office EN3 data set. Again, the time of the profiles is color coded by the year-day. Nearly all the profiles are located in the western Pacific Ocean.

The climatology data set used in our hindcast experiments was remapped using a Fast Marching Method (FMM; Agarwal, 2009) that takes into account archipelago geometry during field gridding and avoids correlations across landforms. The World Ocean Atlas (WOA) data set (Boyer et al., 2005) is still too coarse (0.25° x 0.25°) to be used inside the Philippine Archipelago; it is based on objective mapping correlations with straight Euclidean distances that violate coastline constraints. Improved monthly climatology was computed for our application grid (Figure 1) using the FMM method (Agarwal, 2009). It included similar historical observations to those in the WOA data set.

The climatological temperature and salinity data are also weakly assimilated with larger standard deviation values (0.8°C and 0.18, respectively) to better constrain the posterior analysis in areas lacking subsurface observations. This method also controlled spurious forcing, in deep waters, from open boundary conditions for temperature and salinity from HYCOM daily snapshots.

RESULTS AND ANALYSIS

In the real-time forecast experiments, ROMS was initialized on March 2, 2007, and the initial and lateral boundary conditions were derived from the 1/12° horizontal resolution global HYCOM with NCODA. Atmospheric forcing was computed from NOGAPS, and OTPS tides were imposed on all four open boundaries of the grid shown in Figure 1. The model was run for more than two months to allow enough time for the density field to adjust to the bathymetry and suppress the excitation of internal waves arising from the interpolation between ROMS and HYCOM grids. Then, the real-time forecasts were run daily from May 17 to July 7, 2007. Each run period was for nine days and included a four-day hindcast and five-day forecast. The strategy was to initialize ROMS four days prior to each sequential forecast cycle to allow reanalyzed and more accurate atmospheric and boundary forcing. Several horizontal and vertical cross sections, at six-hour intervals, of potential temperature, salinity, and momentum were issued for each forecast via the Internet. A total of 50 forecasts were distributed prior to and during the exploratory cruise.

These real-time forecasts were validated several months after the cruise ended when the data were processed and became available for the 4D-Var data assimilation experiments. Moreover, quality-controlled ocean data are not available in real time. The Aviso sea surface anomaly product (Figure 2a) is usually delayed one month from the current day for reanalysis products, and up to six days for near-real-time products. The gridded OI SST reanalysis product (Figure 2b) that combines microwave and infrared measurements may be delayed over a month, whereas the near-real-time blended product (Figure 2c) is late by between one and two days. Similarly, hydrographic observations from the UK Met Office EN3 quality-controlled data set may be delayed from several days up to a month. Therefore, the 4D-Var data assimilation experiments were performed in hindcast mode and focused on the exploratory cruise period of June 6 to July 3, 2007, for brevity.

The data assimilation procedure used for combining observations and model, in time and space, is the ROMS I4D-Var algorithm. This algorithm has been used successfully for the US East Coast (Zhang et al., 2010a), Gulf of Mexico and Caribbean Sea (Powell et al., 2008, 2009), California Current System (Broquet et al., 2009, 2010; Moore et al., 2011b,c), and East Australia Current (Javier Zavala-Garay, Rutgers University, *pers. comm.*, 2010). The model *prior* error statistics for the background ocean state are derived from a multiyear, de-tided run of the Philippine Archipelago described in Han et al. (2009). The model error is assumed to be Gaussian and uncorrelated in time, with a horizontal and vertical decorrelation scale of 40 km and 10 m, respectively. We selected a five-day, sequential data assimilation cycle starting on June 6, 2007. This time window is well within the valid range of the tangent linear assumption for this area of seven to 15 days. The observations assimilated include satellite-derived SSH and SST, and subsurface temperature and salinity profiles from various platforms collected during the exploratory cruise. Only the initial conditions are adjusted by using the observations available from the previous five days. In the sequential daily experiments, the model initial conditions (*prior*) are the best ocean state estimate (nowcast) from the previous five-day data assimilation window. Then, the nonlinear model is run freely in verification mode (no data assimilation) for five days, and the RMSE between model and new observations, not yet assimilated, is computed to evaluate the forecasting skill or improvements over the real-time control forecast.

Figure 3 shows the temperature and salinity RMSE between model and observations along the cruise track from June 6–24, 2007. The blue curve shows the values for the nonassimilative real-time control forecast, while the red curve shows the values of the daily nowcasts resulting from the I4D-Var minimization between model and observations for the previous five days. The black curve shows the values of the nonlinear model run in hindcast mode and initialized from the nowcast during the verification step without data assimilation. It should be noted that the error in the nowcast is based on the observations that have been used during I4D-Var estimation. It is a measure of how close the model is getting to the observations. In contrast, the error in the hindcast run is computed with observations that have not yet been assimilated. It can be seen in Figure 3a that the temperature correction is between 1.0° and 2.0°C. The correction for salinity (Figure 3b) is also significant, at nearly 0.2. Most noticeably, the forecast initialized from the nowcast (black curve) has better skill than the nonassimilative, real-time control run (blue curve), especially for salinity. The larger salinity correction is primarily due to correcting the excessive salt flux through the lateral boundaries and for uncertainties in the surface freshwater flux (evaporation minus precipitation) in the control forecast.

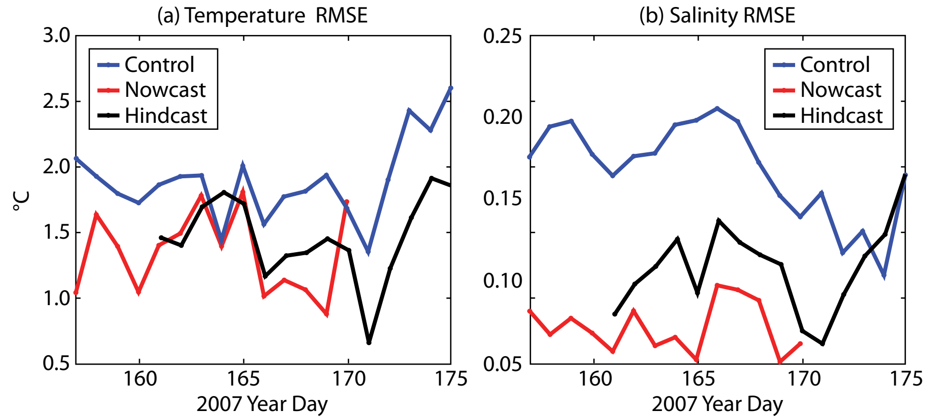

Figure 3. Root mean square error between model and observation for the real-time control forecast without data assimilation (blue curve), I4D-Var best estimate (nowcast) using observations from the previous five days (red curve), and nonlinear model verification run in hindcast mode and initialized from the I4D-Var nowcast (black curve): (a) temperature (°C) and (b) salinity.

Figure 4 shows the time-averaged surface temperature and salinity for the forecast cycle from June 26 to July 1, 2007. It exemplifies the time-averaged spatial structure correction from I4D-Var data assimilation (Figure 4b,e), compared against the time-averaged control run without data assimilation (panels a and d) and the time-averaged observations and climatology (panels c and f). The observations for surface temperature are based on the time-averaged OI SST product (Figure 4c), whereas the FMM June climatology is used for surface salinity (Figure 4f). It is clear that the assimilation of SST is able to restore mesoscale features in the Sulu and Celebes seas, as well as reduce the surface-water cooling inside the archipelago due to excessive vertical mixing. The salinity correction from assimilation is most noteworthy. Figure 4d shows that the real-time control forecast has excessive surface salinity almost everywhere, especially in the Pacific Ocean and inside the archipelago. The surface salinity observations near Mindoro, Panay, and Dipolog straits, shown by colored circles in Figure 4f, are much closer to the FMM June climatology than the real-time control forecast. Evidently, the excessive salt is coming into the archipelago from the lateral open boundaries in the Pacific Ocean, shown in Figure 4d as salty plumes entering the domain. It is similar to the buoyancy flows and impulses observed at the mouth of rivers and estuaries, but saltier instead of fresher. On the other hand, water from the SCS that flows into the Sulu Sea via Mindoro Strait appears to be too fresh in the control run (Figure 4d). The weak data assimilation of climatology next to the Pacific Ocean boundary has removed most of the excessive salinity entering the domain and mitigated the low-salinity inflow from the SCS into the Sulu Sea after nearly a month of sequential I4D-Var. Notice that the anomalous salinity is still present at the Pacific Ocean boundaries; it is due to the prescribed open boundary conditions from HYCOM daily snapshots. Still, some patches of excessive salinity remain near San Bernardino and Surigao straits, Leyte Gulf, and in the eastern Bohol Sea in the data assimilation estimate. There is a need for more observations in these areas to help constrain the ocean state estimate. Unfortunately, Argos float profiles are lacking in this area during this period. Perhaps this Pacific Ocean boundary will benefit, in the future, from salinity measurements from the Aquarius satellite, despite its coarse resolution and surface range. Also, correcting the open boundary conditions using 4D-Var may improve the salinity boundary conditions. We will try this approach in the future.

Figure 5 shows the impact of I4D-Var data assimilation on the exploratory cruise track for temperature (upper panels) and salinity (lower panels). The section's abscissa is in terms of cruise track CTD station numbers shown in Figure 2c. These panels display the observation profiles (Figure 5a,d), the real-time control forecast minus observations (Figure 5b,e), and the I4D-Var data assimilation hindcast minus observations (Figure 5c,f). The temperature correction is mainly due to the model's over-diffused thermocline. The model subsurface temperature in the control forecast is colder than the observations, especially across Mindoro

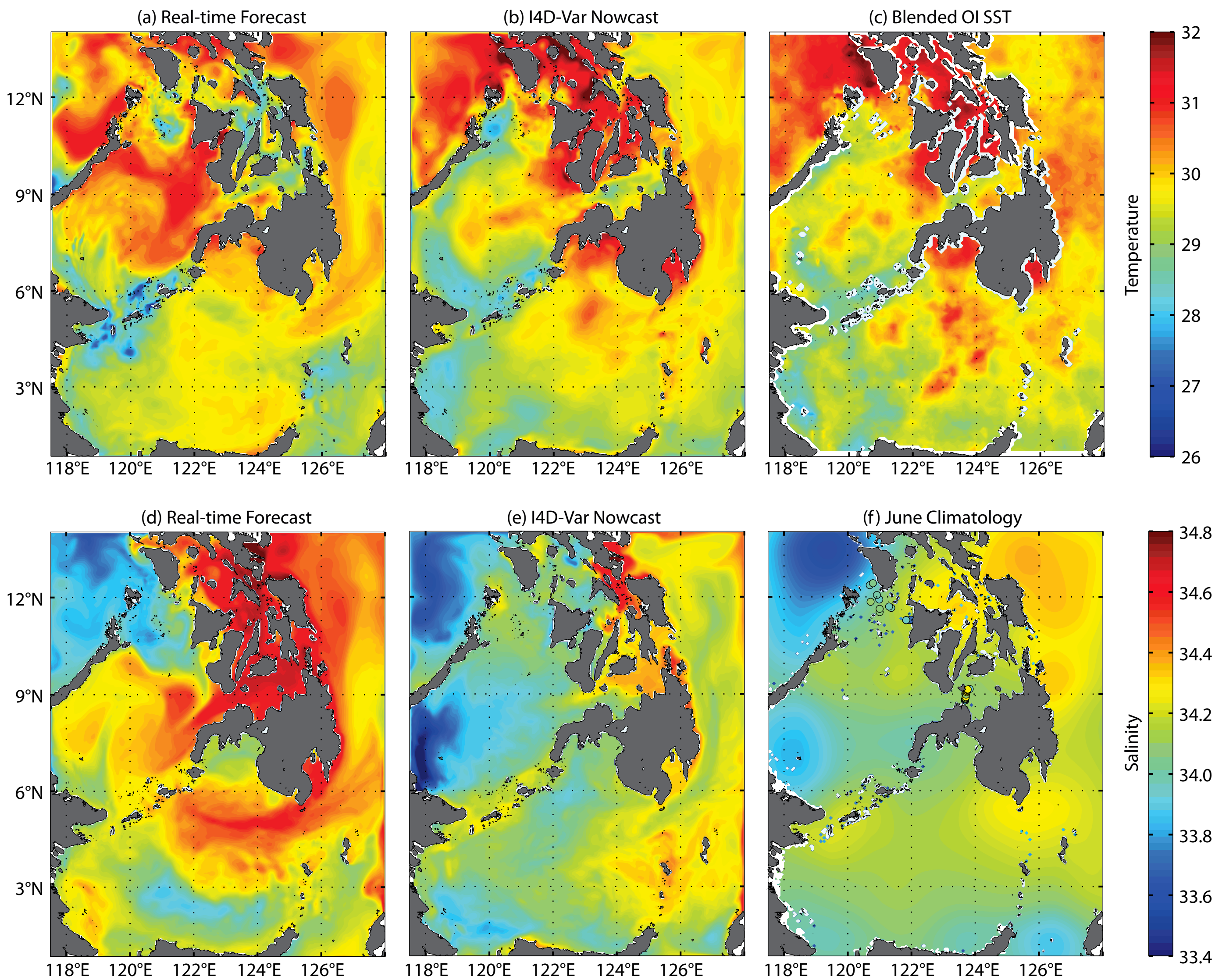

Figure 4. Sea surface temperature (upper panels, °C) and salinity (lower panels) time average from June 26–July 22, 2007. (a) Real-time temperature forecast without data assimilation. (b) Temperature hindcast with data assimilation. (c) Gridded OI SST blend from microwave (AMSR-E and TMI) and infrared (MODIS) sensors. (d) Real-time salinity forecast without data assimilation. (e) Salinity hindcast with data assimilation. (f) Objectively analyzed June salinity climatology using the FMM (Fast Marching Method) algorithm and CTD sea surface salinity (colored circles).

and Panay straits (stations 80–140), indicating over-mixing in these areas. Data assimilation renders a more reliable temperature profile along the cruise track when comparing the departures between model and observations in Figure 5b,c. The RMSE in temperature is decreased 38% from 2.13° to 1.32°C. Similar corrections are noticeable in the salinity estimate. Figure 5e illustrates the excessive salt propagating from the Pacific Ocean open boundary, in the upper 300 m, into Leyte Gulf, across Surigao Strait, the Bohol Sea, and portions of the Sulu Sea (stations 28–80), as shown in the surface map (Figure 4d). The I4D-Var data assimilation successfully removes the anomalous high salinity, as indicated in Figure 5f. Analysis of stations 26 and 27 in the Pacific Ocean shows that the assimilation removes bogus salinity from the surface and creates a more pronounced subsurface salinity layer. Improvement is also noticeable in the Sibuyan Sea (stations 90–100). Overall, the salinity RMSE is decreased by data assimilation from 0.176 to 0.091, a 49% improvement.

CONCLUSION

In this paper, we describe some of the challenges for implementing a hindcast/nowcast/forecast system for the Philippine Archipelago. Usually, regional ocean numerical models are imperfect and observations are temporally and

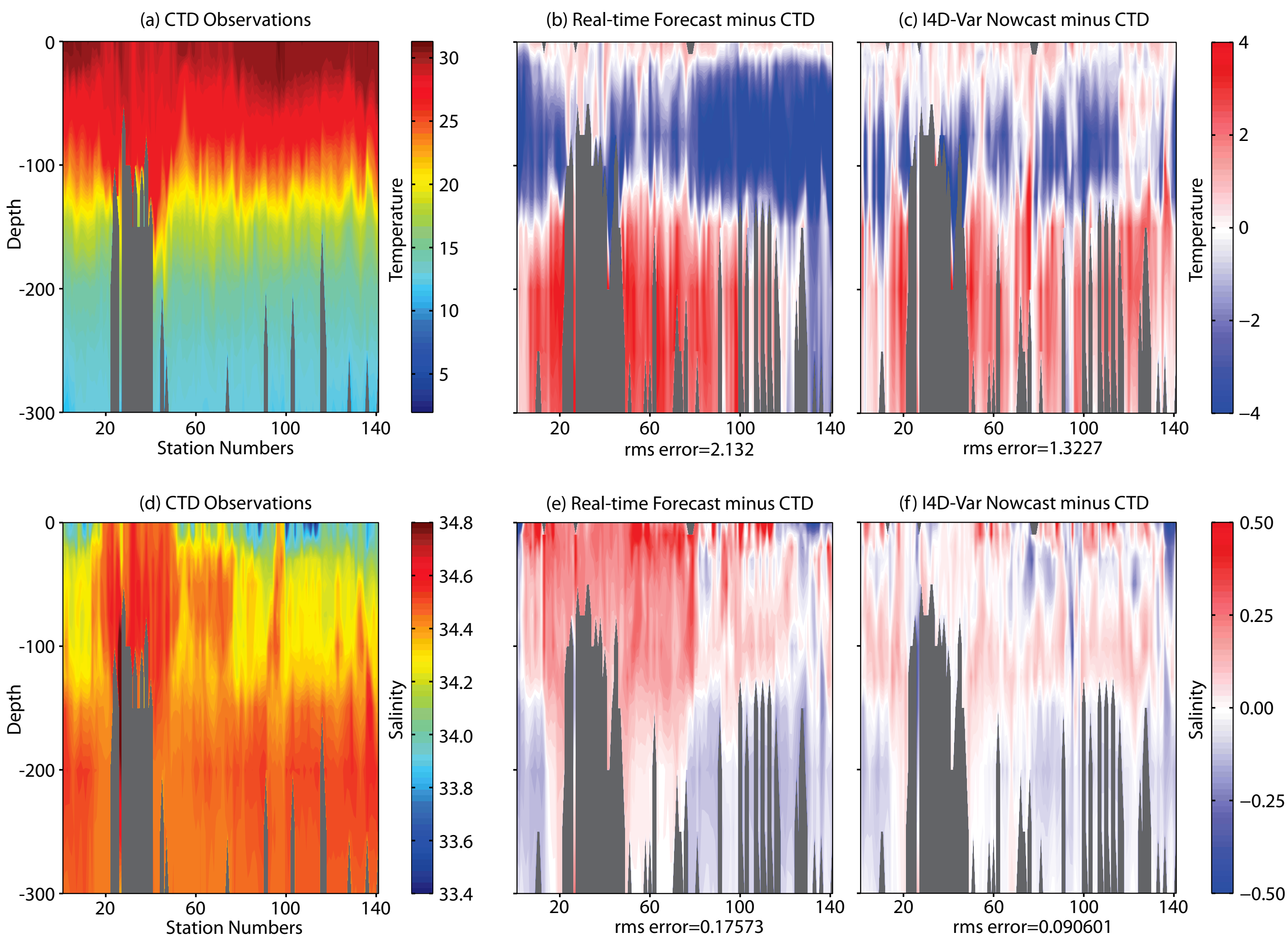

Figure 5. Data assimilation impact for temperature (upper panels, °C) and salinity (lower panels) along the exploratory cruise track. (a) CTD temperature observations, (b) real-time control temperature forecast minus CTD observations, (c) I4D-Var data assimilation temperature hindcast minus CTD observations, (d) CTD salinity observations, (e) real-time control salinity forecast minus CTD observations, and (f) I4D-Var data assimilation salinity hindcast minus CTD observations. Figure 2d shows the numbered cruise track CTD stations.

spatially sparse and incomplete. This limitation could not be more true in the Philippine Archipelago with its complex geometry, bathymetry-dominated flows, and strong local and remote wind forcing. The bathymetry of some of the straits and passages is not well known, and this area is good for using a downscaling modeling approach with grids of increasing resolution. However, it also presents a challenge when specifying lateral boundary conditions for large-scale mean flows, mesoscale and submesoscale variability, and strong external and internal tidal forcing. Obviously, regional, fine-resolution atmospheric models are necessary. The satellite-derived data that are essential in any regional ocean prediction system are not that reliable in the Philippine Archipelago due to the frequency of clouds when deriving SST or the island geometry in the altimetry-derived SSH. Scientific, high-quality, blended multi-sensor (microwave and infrared, altimetry) products are not available for real-time forecasting because they are usually delayed by at least several days. Lower-quality, near-real-time products are available with a one- to two-day delay. Even with the intense measurement program of the ONR PhilEx research initiative, there are still large areas in need of sampling in time and space, and these data may be important to support and evaluate an ocean prediction system for the Philippine Archipelago.

It is clear from the results presented here that the open boundaries in the Pacific Ocean are the source of the larger uncertainty in the forecast, especially salinity. This observation demonstrates the importance of remote forcing and inflows from the Pacific Ocean and South China Sea in affecting the Philippine Archipelago region. The I4D-Var successfully mitigated the over-diffused thermocline, colder subsurface temperatures, and excessive salt flux from the prescribed lateral boundary conditions for salinity. Climatology data can be assimilated with larger uncertainty in areas close to the open boundaries to constrain the estimation and preserve the water masses properties. The data assimilation estimates for temperature and salinity are improved by 38 and 49%, respectively. There is still a lot of modeling work ahead with ROMS adjoint-based algorithms and the PhilEx data sets. The next logical step is to use 4D-Var to correct the surface forcing, open boundary conditions, and model error (i.e., weak constraint). Much of that work is already underway and reported elsewhere. Our results suggest that to achieve accurate depiction and prediction of Philippine Archipelago circulation, data assimilation together with improved surface and lateral boundary forcings and improved model physics are necessary.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Office of Naval Research (N00014-07-1-0417). We thank Joe Metzger, Harley Hurlburt, and Ole Martin Smedstad (NRL) for kindly providing us with the global HYCOM data for initial and lateral boundary conditions and the NOGAPS atmospheric forcing in a timely manner. We are indebted to Arpit Agarwal and Pierre Lermusiaux for objectively analyzing the monthly temperature and salinity climatology for our grid using the FMM algorithm. The authors wish to thank all R/V *Melville* scientists and crew who helped to collect all the ocean measurements and assisted with the logistics. The tidal forcing data were obtained from http://www.coas.oregonstate.edu/research/po/research/tide/index.htm. The satellite-derived SST products were obtained from http://coastwatch.pfel.noaa.gov/data.html server. We thank David Foley for his work on the satellite-derived SST products. The quality-controlled Argo and XBT data were obtained from the http://www.metoffice.gov.uk/hadobs server.

REFERENCES

Agarwal, A. 2009. *Statistical Field Estimation and Scale Estimation for Complex Coastal Regions and Archipelagos*. MS Thesis, Massachusetts Institute of Technology, Department of Mechanical Engineering, May 2009.

Bengtsson, L., M. Ghil, and E. Källen. 1981. *Dynamic Meteorology: Data Assimilation Methods*. Springer-Verlag, 330 pp.

Bennett, A.F. 1992. *Inverse Methods in Physical Oceanography*. Cambridge University Press, 347 pp.

Bennett, A.F. 2002. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press, 234 pp.

Bleck, R. 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modeling* 4:55–88.

Boyer, T., S. Levitus, H. Garcia, R.A. Locarnini, C. Stephens, and J. Antonov. 2005. Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25° grid. *International Journal of Climatology* 25(7):931–945.

Broquet, G., C.A. Edwards, A.M. Moore, B.S. Powell, M. Veneziani, and J.D. Doyle. 2009. Application of 4D-Variational data assimilation to the California Current System. *Dynamics of Atmospheres and Oceans* 48:69–92, doi:10.1016/j.dynatmoce.2009.03.001.

Broquet, G., A.M. Moore, H.G. Arango, and C.A. Edwards. 2010. Corrections to ocean surface forcing in the California Current System using 4D-variational data assimilation. Ocean Modelling 36:116–132, doi:10.1016/j.ocemod.2010.10.005.

Cummings, J.A. 2005. Operational multivariate ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society* 131(613):3,583–3,604.

Daley, R. 1991. *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
Durski, S.M., S.M. Glenn, and D.B. Haidvogel. 2004. Vertical mixing schemes in the coastal ocean: Comparison of the level 2.5 Mellor-Yamada scheme with an enhanced version of the K profile parameterization. *Journal of Geophysical Research* 109, C01015, doi:10.1029/2002JC001702.
Egbert, G.D., A.F. Bennett, and M.G.G. Foreman. 1994. TOPEX/POSEIDON tides estimated using a global inverse model. *Journal of Geophysical Research* 99:24,821–24,852.
Egbert, G.D., and S.Y. Erofeeva. 2002. Efficient inverse modeling of barotropic ocean tides. *Journal of Atmospheric and Oceanic Technology* 19:183–204.
Fairall, C.W., E.F. Bradley, J.E. Hare, A.A. Grachev, and J.B. Edson. 2003. Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *Journal of Climate* 16:571–591.
Flather, R.A. 1976. A tidal model of the northwest European continental shelf. *Memoires de la Société Royale des Sciences de Liège* 10:141–164.
Ghil, M., and P. Malanotte-Rizzoli. 1991. Data assimilation in meteorology and oceanography. *Advances in Geophysics* 33:141–266.
Girton, J.B., B.S. Chinn, and M.H. Alford. 2011. Internal wave climates of the Philippine seas. *Oceanography* 24(1):100–111.
Gordon, A.L., J. Sprintall, and A. Ffield. 2011. Regional oceanography of the Philippine Archipelago. *Oceanography* 24(1):14–27.
Haidvogel, D.B., H. Arango, W.P. Budgell, B.D. Cornuelle, E. Curchitser, E. Di Lorenzo, K. Fennel, W.R. Geyer, A.J. Hermann, L. Lanerolle, and others. 2008. Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. *Journal of Computational Physics* 227(7):3,595–3,624.
Haidvogel, D.B., H.G. Arango, K. Hedstrom, A. Beckmann, P. Malanotte-Rizzoli, and A.F. Shchepetkin. 2000. Model evaluation experiments in the North Atlantic Basin: Simulations in nonlinear terrain-following coordinates. *Dynamics of Atmospheres and Oceans* 32:239–281.
Han, W., A.M. Moore, J. Levin, B. Zhang, H.G. Arango, E. Curchitser, E. Di Lorenzo, A.L. Gordon, and J. Lin. 2009. Seasonal surface ocean circulation and dynamics in the Philippine Archipelago region during 2004–2008. *Dynamics of Atmospheres and Oceans* 47:114–137.
Hurlburt, H.E., E.J. Metzger, J. Sprintall, S.N. Riedlinger, R.A. Arnone, T. Shinoda, and X. Xu. 2011. Circulation in the Philippine Archipelago simulated by 1/12° and 1/25° global HYCOM and EAS NCOM. *Oceanography* 24(1):28–47.
Ingleby, B., and M. Huddleston. 2007. Quality control of ocean temperature and salinity profiles: Historical and real-time data. *Journal of Marine Systems* 65:158–175.
Kleist, D.T., D.F. Parrish, J.C. Derber, R. Treadon, W.-S. Wu, and S. Lord. 2009. Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather and Forecasting* 24:1,691–1,705.
May, P.W., J.D. Doyle, J.D. Pullen, and L.T. David. 2011. Two-way coupled atmosphere-ocean modeling of the PhilEx Intensive Observational Periods. *Oceanography* 24(1):48–57.
Moore, A.M., H.G. Arango, E. Di Lorenzo, B.D. Cornuelle, A.J. Miller, and D.J. Neilson. 2004. A comprehensive ocean prediction and analysis system based on the tangent linear and adjoint of a regional ocean model. *Ocean Modelling* 7:227–258.
Moore, A.M., H.G. Arango, E. Di Lorenzo, A.J. Miller, and B.D. Cornuelle. 2009. An adjoint sensitivity analysis of the Southern California Current circulation and ecosystem. *Journal of Physical Oceanography* 39:702–720.
Moore, A.M., H.G. Arango, G. Broquet, B.S. Powell, J. Zavala-Garay, and A.T. Weaver. 2011a. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part I: Formulation and overview. *Progress in Oceanography*, submitted.
Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B.S. Powell, D. Foley, J. Doyle, D. Costa, and P. Robinson. 2011b. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part II: Performance and applications to the California Current System. *Progress in Oceanography*, submitted.
Moore, A.M., H.G. Arango, G. Broquet, C. Edwards, M. Veneziani, B.S. Powell, D. Foley, J. Doyle, D. Costa, and P. Robinson. 2011c. The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems, Part III: Observation impact and observation sensitivity in the California Current System. *Progress in Oceanography*, submitted.
Powell, B.S., H.G. Arango, A.M. Moore, E. Di Lorenzo, R.F. Milliff, and D. Foley. 2008. 4DVAR data assimilation in the Intra-Americas Sea with the Regional Ocean Modeling System (ROMS). *Ocean Modelling* 25:173–188.
Powell, B.S., A.M. Moore, H.G. Arango, E. Di Lorenzo, R.F. Milliff, and R.R. Leben. 2009. Near real-time ocean circulation assimilation and prediction in the Intra-Americas Sea with ROMS. *Dynamics of Atmospheres and Oceans* 48:46–68.
Rio, M.-H., P. Schaeffer, J.-M. Lemoine, and F. Hernandez. 2005. Estimation of the ocean mean dynamic topography through the combination of altimetric data, in-situ measurements, and GRACE geoid: From global to regional studies. *Proceedings of the GOCINA [Geoid and Ocean Circulation in the North Atlantic] International Workshop*. April 13–15, 2005, Luxembourg.
Shchepetkin, A.F., and J.C. McWilliams. 2003. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. *Journal of Geophysical Research* 108(C3), 3090, doi:10.1029/2001JC001047.
Shchepetkin, A.F., and J.C. McWilliams. 2005. The Regional Ocean Modeling System: A split-explicit, free-surface, topography following coordinates ocean model. *Ocean Modelling* 9:347–404.
Shchepetkin, A.F., and J.C. McWilliams. 2009. Computational kernel algorithms for fine-scale, multiprocess, longtime oceanic simulations. Pp. 119–182 in *Handbook of Numerical Analysis: Computational Methods for the Atmosphere and Oceans.* R.M. Teman and J.J. Tribbia, eds, Elsevier Science, doi:10.1016/S1570-8659(08)01202-0.
Smith, W.H.F., and D.T. Sandwell. 1997. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277:1,956–1,962.
Tarantola, A. 1987. *Inverse Problem Theory: Methods for Data Filtering and Model Parameter Estimation*. Elsevier, 613 pp.
Wang, X., and Y. Chao. 2004. Simulated sea surface salinity variability in the tropical Pacific. *Geophysical Research Letters* 31, L02302, doi:10.1029/2003GL018146.
Warner, J.C., C.R. Sherwood, H.G. Arango, and R.P. Signell. 2005. Performance of four turbulence closure methods implemented using a generic length scale method. *Ocean Modelling* 8:81–113.
Willis, J.K., D. Roemmich, and B.D. Cornuelle. 2004. Interannual variability in upper ocean heat content, temperature, and thermosteric expansion on global scales. *Journal of Geophysical Research* 109, C12036, doi:10.1029/2003JC002260.
Wunsch, C. 1996. *The Ocean Circulation Inverse Problem*. Cambridge University Press, 442 pp.
Zhang, W.G., J.L. Wilkin, and H.G. Arango. 2010a. Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation. *Ocean Modelling* 35:119–133.
Zhang, W.G., J.L. Wilkin, and J.C. Levin. 2010b. Towards an integrated observation and modeling system in the New York Bight using variational methods. Part II: Repressenter-based observing strategy evaluation. *Ocean Modelling* 35:134–145.