



NRL/MR/7320--08-9148

# Validation Test Report for the Global Ocean Prediction System V3.0 – 1/12° HYCOM/NCODA: Phase I

E.J. METZGER  
H.E. HURLBURT  
A.J. WALLCRAFT  
J.F. SHRIVER  
L.F. SMEDSTAD

*Ocean Dynamics and Prediction Branch  
Oceanography Division*

O.M. SMEDSTAD  
P. THOPPIL  
D.S. FRANKLIN

*QinetIQ North America  
Planning Systems, Inc.  
Slidell, LA*

November 26, 2008

Approved for public release; distribution is unlimited.

# REPORT DOCUMENTATION PAGE

*Form Approved*  
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

|   |                                    |  |   |   |  |
|---|------------------------------------|--|---|---|--|
| <b>1. REPORT DATE (DD-MM-YYYY)</b><br>26-11-2008  |                                    | <b>2. REPORT TYPE</b><br>Memorandum Report |   | <b>3. DATES COVERED (From - To)</b>   |  |
| <b>4. TITLE AND SUBTITLE</b><br><br>Validation Test Report for the Global Ocean Prediction System V3.0 – 1/12° HYCOM/NCODA: Phase I   |                                    |  |   | <b>5a. CONTRACT NUMBER</b>  |  |
|   |                                    |  |   | <b>5b. GRANT NUMBER</b>   |  |
|   |                                    |  |   | <b>5c. PROGRAM ELEMENT NUMBER</b><br>0603207N                               |  |
| <b>6. AUTHOR(S)</b><br><br>E.J. Metzger, O.M. Smedstad,* P. Thoppil,* H.E. Hurlburt, A.J. Wallcraft, D.S. Franklin,* J.F. Shriver, and L.F. Smedstad  |                                    |  |   | <b>5d. PROJECT NUMBER</b>   |  |
|   |                                    |  |   | <b>5e. TASK NUMBER</b>  |  |
|   |                                    |  |   | <b>5f. WORK UNIT NUMBER</b><br>73-5094-18-5                                 |  |
| <b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b><br><br>Naval Research Laboratory<br>Oceanography Division<br>Stennis Space Center, MS 39529-5004  |                                    |  |   | <b>8. PERFORMING ORGANIZATION REPORT NUMBER</b><br><br>NRL/MR/7320--08-9148 |  |
| <b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b><br><br>Space & Naval Warfare Systems Command<br>2451 Crystal Dr.<br>Arlington, VA 22245-5200   |                                    |  |   | <b>10. SPONSOR / MONITOR'S ACRONYM(S)</b><br><br>SPAWAR                     |  |
|   |                                    |  |   | <b>11. SPONSOR / MONITOR'S REPORT NUMBER(S)</b>                             |  |
| <b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b><br><br>Approved for public release; distribution is unlimited.   |                                    |  |   |   |  |
| <b>13. SUPPLEMENTARY NOTES</b><br><br>*QinetIQ North America, Planning Systems, Inc., Slidell, LA   |                                    |  |   |   |  |
| <b>14. ABSTRACT</b><br><br>Global Ocean Prediction System Version 3.0 (V3.0) is comprised of the 1/12° global HYbrid Coordinate Ocean Model (HYCOM) and the Navy Coupled Ocean Data Assimilation (NCODA) system. It is a next-generation system capable of nowcasting and forecasting the oceanic “weather,” which includes the three-dimensional ocean temperature, salinity and current structure, the surface mixed layer, and the location of mesoscale features such as eddies, meandering currents, and fronts. It supports the current Navy requirements for Numerical Modeling (METOC 98-01), Hi-Resolution Surface and Sub-Surface Currents Model (OCEAN 93-08), Tactical Environmental/Acoustic Routing (OCEAN 91-11), and Air/Sea Drift Prediction (OCEAN 91-15). V3.0 is scheduled to replace the existing nowcast/forecast system (V2.6) based on the 1/8° Navy Coastal Ocean Model (NCOM), 1/32° Navy Layered Ocean Model (NLOM), 1/8° Modular Ocean Data Analysis System (MODAS) and NCODA. This report describes the validation testing performed on one-year hindcasts of V3.0 and V2.5+, an interim version of V2.6 that does not include in situ profile assimilation and referred to throughout this report as V2.5. The phase I validation tasks include evaluation of a) the large scale circulation features, b) sea surface height variability and eddy kinetic energy, c) vertical temperature and salinity structure, d) quantities that describe the underwater acoustical environment such as mixed layer depth, sonic layer depth, deep sound channel axis and below layer gradient, e) sea surface temperature, and f) coastal/island sea level variability. Analyses are performed on both hindcast and forecast simulations. For these quantities in areas of Navy interest, V3.0 is generally performing equal to or slightly better than V2.5, with the exception of sonic layer depth. |                                    |  |   |   |  |
| <b>15. SUBJECT TERMS</b><br>HYCOM            Global ocean nowcast/forecast system<br>NCODA           Model validation   |                                    |  |   |   |  |
| <b>16. SECURITY CLASSIFICATION OF:</b>  |                                    |  | <b>17. LIMITATION OF ABSTRACT</b><br><br>UL | <b>18. NUMBER OF PAGES</b><br><br>85  | <b>19a. NAME OF RESPONSIBLE PERSON</b><br>E. Joseph Metzger        |
| <b>a. REPORT</b><br>Unclassified  | <b>b. ABSTRACT</b><br>Unclassified | <b>c. THIS PAGE</b><br>Unclassified        |   |   | <b>19b. TELEPHONE NUMBER (include area code)</b><br>(228) 688-4762 |



## CONTENTS

|   |    |
|---|----|
| 1.0 INTRODUCTION .....                                  | 1  |
| 2.0 V3.0 SYSTEM COMPONENTS.....                         | 4  |
| 2.1 Global HYCOM.....                                   | 4  |
| 2.2 NCODA.....  | 6  |
| 2.3 PIPS.....   | 8  |
| 2.4 The HYCOM/NCODA Runstream .....                     | 9  |
| 3.0 TESTING RESULTS .....                               | 9  |
| 3.1 Validation Experiments .....                        | 9  |
| 3.2 Large Scale Mean and Variability .....              | 10 |
| 3.3 Temperature/Salinity vs. Depth Error Analysis ..... | 14 |
| 3.4 MLD/SLD/DSC/BLG Error Analysis.....                 | 21 |
| 3.5 Sea Surface Temperature Error Analysis .....        | 24 |
| 3.6 Coastal/Island Sea Level Error Analysis.....        | 27 |
| 4.0 SUMMARY AND RECOMMENDATIONS.....                    | 28 |
| 5.0 ACKNOWLEDGEMENTS .....                              | 31 |
| 6.0 REFERENCES .....                                    | 33 |
| 7.0 TABLE OF ACRONYMS .....                             | 37 |



## 1.0 INTRODUCTION

Development of an advanced global ocean nowcasting/forecasting system has been of long-time Navy interest. Such a system will provide the capability to depict (nowcast) and predict (forecast) the oceanic “weather”, some components of which include the three dimensional (3-D) ocean temperature, salinity and current structure, the surface mixed layer and the location of mesoscale features such as eddies, meandering currents and fronts. The space scales of these eddies and meandering currents are typically about 100 km and currents speeds can easily exceed  $1 \text{ ms}^{-1}$  in the western boundary current regions of the Kuroshio, Gulf Stream and Somali Current. So, relatively high horizontal and vertical resolution numerical ocean models are needed to depict the 3-D ocean structure with accuracy superior to climatology and/or persistence (i.e. a forecast of no change). Knowledge of the oceanic mesoscale has many naval applications, including tactical planning, optimum track ship routing, search and rescue operations, long-range weather prediction, inputs to coastal models, and knowledge of high current shear zones.

The previous operational global ocean nowcast/forecast system run at the Naval Oceanographic Office (NAVOCEANO) consisted of the  $1/8^\circ$  Navy Coastal Ocean Model (NCOM),  $1/32^\circ$  Navy Layered Ocean Model (NLOM) and  $1/8^\circ$  Modular Ocean Data Analysis System (MODAS) and was known as Global Ocean Prediction System (GOPS) Version 2.5 (V2.5). The motivation for the NLOM-NCOM-MODAS approach is described in Rhodes et al. (2002) and should be viewed as a multi-model system with each component having a different function. MODAS provides an optimum interpolation ocean analysis with the ability to generate synthetic temperature and salinity profiles based on sea surface height (SSH) and sea surface temperature (SST) (Fox et al., 2002). NLOM employs

high horizontal resolution and low vertical resolution (7 layers including the mixed layer); this near-global model assimilates SSH data along altimeter tracks and acts as a dynamical interpolator to accurately depict the oceanic mesoscale features. The initial operational horizontal resolution was  $1/16^\circ$  (Smedstad et al., 2003) but this was increased to  $1/32^\circ$  (Shriver et al., 2007) on 6 March 2006. NCOM extended the model domain to cover the global ocean (including the Arctic Ocean), into shallow waters (5 m depth) and it provided higher vertical resolution in the mixed layer (Barron et al., 2006; Kara et al., 2006). It assimilated MODAS-derived synthetic temperature and salinity profiles based on climatological relationships and using  $1/32^\circ$  NLOM SSH anomalies and  $1/8^\circ$  MODAS SST. Validation and evaluation of this system can be found in Barron et al. (2007a) and Barron et al. (2007b).

As of 24 June 2008, the operational NCOM-based nowcast/forecast system began assimilating *in-situ* profile observations via the Navy Coupled Ocean Data Assimilation (NCODA) system (Cummings, 2005). Additional improvements incorporated into the system at that time included: a) the use of mixed layer depth from  $1/32^\circ$  near-global NLOM to improve the MODAS-derived synthetic temperature and salinity profiles resulting in better representation of the surface sonic layer and reduced transmission loss errors, b) a new version of the NCOM code with version control, standardized input/output, Earth System Model Framework (ESMF) compliance (Hill et al., 2004) and a nine-point buoyancy gradient filter and c) a 4-day (96 hour) forecast. On 17 September 2008, the transition of GOPS V2.6 was completed with the addition of an ice modeling component based on the Los Alamos National Laboratory Community Ice Code (CICE) (Bitz and Lipscomb, 1999; Lipscomb and Hunke, 2004; Posey et al., 2008). This system

provides medium-range resolution (mid-latitude nominally at 15 km at 40°N) and currently produces a nowcast/forecast capability out to four days (96 hours). V2.6 has been declared an operational product in providing boundary conditions for nested high-resolution fixed and relocatable coastal forecasting systems and for producing sound speed profiles. Evaluations against this version are not detailed in this report because a sufficiently long hindcast experiment was not available in the time frame needed.

A next generation system based on the HYbrid Coordinate Ocean Model (HYCOM) has been under development at the Naval Research Laboratory (NRL) since 2000 and will be described in more detail in the next section. HYCOM is unique in that it allows a truly general vertical coordinate and is designed to provide a major advance over the existing operational global ocean prediction systems, since it overcomes design limitations of the present systems as well as limitations in vertical and horizontal resolution. The result should be a more streamlined system with improved performance and an extended range of applicability. This new system will be able to nowcast and forecast the Class 2 (and some Class 1 and 4) ocean responses to atmospheric forcing as discussed in Hurlburt et al. (2008a).

This report will discuss the various components of this new GOPS (Version 3.0 (V3.0)) and the validation efforts in relation to V2.5. The validation is designed to occur in phases because of the considerable resources (both manpower and computational) required. Table 1 lists the Phase I and II validation tasks and this report will focus on the former set.

## **2.0 V3.0 SYSTEM COMPONENTS**

### **2.1 Global HYCOM**

As configured within GOPS V3.0, HYCOM has a horizontal equatorial resolution of  $.08^\circ$  or  $\sim 1/12^\circ$  ( $\sim 9$  km). The previously stated  $1/8^\circ$  resolution of global NCOM applies to mid-latitudes whereas its equatorial resolution is  $45/256^\circ$ . Thus, the resolution difference between the two models is 2.2x finer grid spacing in HYCOM. This makes HYCOM eddy-resolving while NCOM is only eddy-permitting. Eddy-resolving models can more accurately simulate western boundary currents and the associated mesoscale variability and they better maintain more accurate and sharper ocean fronts. In particular, an eddy-resolving ocean model allows upper ocean – topographic coupling via flow instabilities, while an eddy-permitting model does not because fine resolution of the flow instabilities is required to obtain sufficient coupling (Hurlburt et al., 2008b). The coupling occurs when flow instabilities drive abyssal currents that in turn steer the pathways of upper ocean currents (Hurlburt et al., 1996 in the Kuroshio; Hogan and Hurlburt, 2000 in the Japan/East Sea; and Hurlburt and Hogan, 2008 in the Gulf Stream). In ocean prediction this coupling is important for ocean model dynamical interpolation skill in data assimilation/nowcasting and in ocean forecasting, which is feasible on time scales up to about a month (Hurlburt et al., 2008a).

The HYCOM grid is on a Mercator projection from  $78.64^\circ\text{S}$  to  $47^\circ\text{N}$  and north of this it employs an Arctic dipole patch where the poles are shifted over land to avoid a singularity at the North Pole. This gives a mid-latitude (polar) horizontal resolution of approximately 7 km (3.5 km). This version employs 32 hybrid vertical coordinate surfaces with potential density referenced to 2000 m and it includes the effects of thermobaricity

(Chassignet et al., 2003). Vertical coordinates can be isopycnals (density tracking), often best in the deep stratified ocean, levels of equal pressure (nearly fixed depths), best used in the mixed layer and unstratified ocean and sigma-levels (terrain-following), often the best choice in shallow water. HYCOM combines all three approaches by choosing the optimal distribution at every time step. The model makes a dynamically smooth transition between coordinate types by using the layered continuity equation. The hybrid coordinate extends the geographic range of applicability of traditional isopycnic coordinate circulation models toward shallow coastal seas and unstratified parts of the world ocean. It maintains the significant advantages of an isopycnal model in stratified regions while allowing more vertical resolution near the surface and in shallow coastal areas, hence providing a better representation of the upper ocean physics. HYCOM is configured with options for a variety of mixed layer submodels (Halliwell, 2004) and this version uses the K-Profile Parameterization (KPP) of Large et al. (1994). A more complete description of HYCOM physics can be found in Bleck (2002).

The ocean model uses the Fleet Numerical Meteorology and Oceanography Center (FNMOC) 3-hourly 0.5° Navy Operational Global Atmospheric Prediction System (NOGAPS) forcing that includes these fields: air temperature at 2 m, surface specific humidity, net surface shortwave and longwave radiation, total (large scale plus convective) precipitation, ground/sea temperature, zonal and meridional wind velocities at 10 m, mean sea level pressure and dewpoint temperature at 2 m. The first six fields are input directly into the ocean model or used in calculating components of the heat and buoyancy fluxes while the last four are used to compute surface wind stress with temperature and humidity based stability dependence. Currently the system uses the 0.5°

application grid NOGAPS products (i.e. already interpolated by FNMOC to a constant 0.5° latitude/longitude grid); however HYCOM can also (and preferably) use the 0.5° computational grid (i.e. a Gaussian grid – constant in longitude, nearly constant in latitude) products. Typically atmospheric forcing forecast fields extend out to 120 hours (i.e. the length of the HYCOM/NCODA forecast). On those instances when atmospheric forecasts are shorter than 120 hours, an extension is created based on climatological products. The last available NOGAPS forecast field is then gradually blended toward climatology to provide forcing over the entire forecast period.

## **2.2 NCODA**

NCODA is a fully three-dimensional multivariate optimum interpolation (MVOI) scheme (Cummings, 2005). The three-dimensional ocean analysis variables include temperature, salinity, geopotential and the vector velocity components which are all analyzed simultaneously. In support of HYCOM, a new analysis variable was added to NCODA that corrects the model layer pressure of the hybrid vertical coordinates. It can be run in stand-alone mode but here is cycled with HYCOM to provide updated initial conditions for the next model forecast in a sequential incremental update cycle. Corrections to the HYCOM forecast are based on all observations that have become available since the last analysis. These include surface observations from satellites, including altimeter SSH anomalies, SST, and sea ice concentration, plus *in-situ* SST observations from ships and buoys as well as T & S profile data from XBTs, CTDs and Argo floats. See Table 1 in Cummings (2005) for a complete list. All observations must be quality controlled and this is done via NCODA\_QC which is operational at

NAVOCEANO. By combining these various observational data types via data assimilation and using the dynamical interpolation skill of the model, the 3-D ocean environment can be more accurately nowcast and forecast.

Section 4c in Cummings (2005) provides a detailed description of the two NCODA approaches for projecting surface observations downward to perform the 3-D ocean analysis. Here we present a brief discussion on the implementation in the system being transitioned. In FY08, significant effort went into improving the overall assimilation methodology, e.g. improvements were made to the vertical remapping between the HYCOM first-guess in hybrid space to NCODA z-levels (the piecewise parabolic method was adopted). The technique based on Cooper and Haines (CH) (1996) for direct assimilation of observed SSH change was also modified. Examples include a new temperature-based definition of the HYCOM mixed layer, below which CH becomes active and improved remapping from NCODA analysis z-levels back to HYCOM hybrid space. Substantial progress was made and that showed the assimilated observational inputs were more accurately ingested into the ocean model than before these modifications. However, error analyses based on comparison to non-assimilated T & S profile observations indicated relatively large bias and root mean square error (RMSE). This was traced back to a subsurface warm temperature bias in the non-assimilative HYCOM simulation used to initialize the assimilative system. Cummings (2005) notes a disadvantage of CH is that it cannot correct for model bias or long-term drift of water mass characteristics. Thus the second approach for downward projection was tested – assimilation of synthetic T & S profiles computed from MODAS similar to what is done in GOPS V2.5. These profiles are only created where the satellite based SSH anomalies with

respect to the previous day's ocean analysis exceed a user-defined value. Error analyses of the same non-assimilated T & S profile observations using the MODAS approach yielded much smaller bias and RMSE than the CH approach and thus MODAS synthetics were chosen for the downward projection methodology. One disadvantage to the MODAS approach is that the system does not cycle on the HYCOM mean SSH, but rather utilizes the MODAS mean SSH. Attempts were made to modify NCODA to use the model mean SSH, but progress was slow and because of the need to complete a year-long hindcast for validation, the validation experiment had to be integrated without this modification.

### **2.3 PIPS**

Initial plans also called for HYCOM to be fully two-way coupled with PIPS via ESMF and both were to have run on the same computational grid. Regional (1/12° Bering Sea) versions of two-way coupled HYCOM/PIPS systems have been tested and demonstrated to NAVOCEANO, but use as a global coupled system has been delayed due to unforeseen problems by the Los Alamos development team in implementing the tripole grid configuration. As of this writing, these issues appear to have been resolved and testing and evaluation are underway, but the working version came too late to be included in this transition. Global HYCOM does include the built-in energy loan, thermodynamic ice model. In this non-rheological system, ice grows or melts as a function of SST and heat fluxes. In addition, Special Sensor Microwave Imager (SSM/I) ice concentration observations from NCODA are directly inserted into the model.

## **2.4 The HYCOM/NCODA Runstream**

A depiction of the HYCOM/NCODA runstream is shown in Figure 1. The first NCODA ocean analysis is performed at  $\tau = -126$  hours with the analysis window for altimeter (all other) data spanning  $\pm 36$  ( $\pm 12$ ) hours. (The first hindcast goes back 5+ days from the nowcast because of late arriving satellite altimeter data. An examination of the timeliness of the historical altimeter data determined an additional data gain of 18% between four and five days; orbits also improve with the age of the data.) After the NCODA analysis, HYCOM is run for 24 model hours with the NCODA analysis incrementally updating the ocean model over the first six hours, thus at 00Z HYCOM has fully ingested the observational data. The NCODA analysis and HYCOM hindcast cycle repeats itself daily up to the nowcast time and HYCOM continues to run in forecast mode out to 120 hours.

## **3.0 TESTING RESULTS**

### **3.1 Validation Experiments**

The V3.0 hindcast began integrating on model day 16 May 2007 and was initialized from a non-assimilative HYCOM experiment using a climatology built on the 1978-2002 European Centre for Medium-Range Weather Forecasts (ECMWF) forcing. The wind forcing in that spin-up simulation was scaled relative to the 1999-2002 monthly NASA Quick Scatterometer (QuikSCAT) wind speeds. The overall impact of this QuikSCAT scaling is to increase the magnitude of the surface wind stress, but nothing was done to change the direction. The hindcast used analysis quality 3-hourly  $0.5^\circ$  NOGAPS forcing with the QuikSCAT scaling also applied to the winds. The validation

error analyses were performed over the year-long period 1 June 2007 – 31 May 2008. This hindcast was brought up to real-time and is the spin-up to the system being transitioned.

In order to examine model error as a function of forecast length, a series of forecasts were integrated and all were initialized from the hindcast described above. On the 1<sup>st</sup>, 8<sup>th</sup>, 15<sup>th</sup> and 22<sup>nd</sup> of each month, five-day HYCOM forecasts were run for a total of 48 forecast integrations. Forecast quality 3-hourly 0.5° NOGAPS atmospheric forcing was used and no oceanic data were assimilated into the ocean model.

GOPS V2.5 ran at NAVOCEANO over the same June 2007-May 2008 time frame. It used analysis quality 3-hourly 0.5° NOGAPS atmospheric forcing up to the nowcast time and forecast quality forcing thereafter. Each day it created a 3-day forecast. The nowcasts (forecasts) from V2.5 have thus been compared against the V3.0 hindcasts (forecasts), although the forecast comparison ends at three days.

### **3.2 Large Scale Mean and Variability**

A first order requirement of any nowcast/forecast system is the accurate representation of the large scale ocean circulation and mesoscale variability. Both V3.0 and V2.5 correctly simulate the basin-wide gyre systems as depicted in Figure 2. Panel a) shows the mean dynamic ocean topography (MDOT) from Maximenko and Niiler (2005) that spans the period 1992-2002. This improved global mean sea level is obtained by combining two datasets: 1) the large-scale mean sea level that is based on measurements from the twin-satellite Gravity Recovery and Climate Experiment (GRACE) mission and 2) the mesoscale sea level tilt that is derived from the momentum balance as seen in near-

surface drifting buoys, satellite altimeter data and the National Centers for Environmental Prediction (NCEP) wind data. The hybrid product reveals complex structures of the main currents even after averaging over 10 years and depicts their location in the large-scale near-surface circulation. The simulated mean SSH over the hindcast period is shown in Figures 2b,c. Qualitatively, the large scale gyre systems and major current systems are well represented in both V3.0 and V2.5. There is more small scale structure embedded within the simulated means, but this is due to the sampling period, one year for the model results versus 11 years for MDOT. Quantitatively, the standard deviation of the difference between V3.0 and MDOT is 7.3 cm and that between V2.5 and MDOT is slightly higher at 7.6 cm. The standard deviation of the difference between the V3.0 and V2.5 is 4.3 cm. The simulated major western boundary currents also agree well with the observed pathways as shown for the Gulf Stream (Figure 3) and the Kuroshio (Figure 4); the currents separate from the coast at the proper latitude and the simulated mean meanders compare well with the observations. Note that in Figure 3 the overlaid observed pathway is the north wall position and thus the core of the simulated current is south of the mean.

While a more thorough evaluation of the velocity structure will be performed in the Phase II validation, a few examples are provided to highlight differences between the two systems in an area of key Navy interest. The velocity structure in the top 300 m of the water column is shown for Luzon Strait (Figure 5) and east of Taiwan (Figure 6). Observations from Liang et al. (2003) are based on a 10-year composite of shipboard acoustic Doppler current profiler (ADCP) data. There are two branches of the Kuroshio in Luzon Strait, one which penetrates into the South China Sea basin and a second branch

that bypasses to the east of the Babuyan Islands (Metzger and Hurlburt, 2001). The observations indicate the inner and outer branches have a subsurface core speed of .6 m/s and .4 m/s, respectively with a weak reversal in the lee of the island chain. Both systems simulate this basic pattern but the subsurface core speed of the inner branch is better represented in V3.0 as is the longitude of the weak counter flow. The core speed of the by-passing branch in V2.5 agrees more closely with the observations, but the maximum is surface trapped (similar to the inner branch), rather than being subsurface. Following the Kuroshio northward, the observations indicate a two core system at 22°N and 23°N on the east side of Taiwan that eventually merges into a single core system at 24°N and 25°N (Figure 6). This structure is depicted (missing) in V3.0 (V2.5), the core speed of the Kuroshio is similar to (weaker than) the observations and the maximum is subsurface (surface trapped). V3.0 continues to have a stronger Kuroshio in the upper 1000 m between the continental shelf of the East China Sea and the Ryukyu Island chain as well as in Tokara Strait south of Kyushu, Japan (not shown).

The SSH variability can be used as a measure of oceanic mesoscale eddy activity and Figure 7 compares observations versus simulated results. The observations were obtained from Collecte Localisation Satellites (CLS) and span October 1992 through May 2007. They are based on available altimeter data from several satellites (ERS-1 and 2, TOPEX/POSEIDON, Geosat Follow-On, Jason-1 and Envisat). Areas of high eddy activity are seen in the western boundary current regions, Agulhas Retroflexion region, Gulf of Mexico, Brazil-Malvinas region, along the Antarctic Circumpolar Current and east of Australia. Both V3.0 and V2.5 reproduce the higher SSH variability in these regions, although it is weaker than indicated by the observations. Insufficient model SSH

variability is a long-standing problem in the ocean modeling community. Inadequate horizontal model resolution is one reason for the difference and this is highlighted by noting that V3.0 generally has more variability than the lower horizontal resolution V2.5. Some differences are also attributed to the different averaging periods, e.g. the higher observed variability in the equatorial Pacific Ocean is associated with El Niño events. These were absent during the time frame of the hindcasts and thus the equatorial Pacific Ocean has relatively low variability in the simulated results. The global areal average of the SSH variability outside the ice mask is 7.8, 7.3 and 6.0 cm for Figures 7a-c, respectively. Figures 8 and 9 zoom in on the Gulf Stream and Kuroshio regions and they indicate that V3.0 simulates the meandering currents and eddy field in these regions better than V2.5.

Eddy kinetic energy (EKE) is a measure of the mesoscale variability with respect to ocean current velocities. It is higher in regions of mesoscale flow instabilities, such as the meandering and eddy shedding regions of the western boundary currents, or regions of active eddy shedding such as the Gulf of Mexico or Agulhas Retroflexion. A comparison of observed versus simulated EKE for the North Atlantic Ocean is shown in Figure 10. Panel a) is taken from Fratantoni (2001) and is based on 1500 drifting buoys drogued at 15 m depth that span the time frame 1990-1999 whereas the lower two panels are from the V3.0 and V2.5 hindcasts. Qualitatively, surface EKE in V3.0 agrees better with the climatological observations than V2.5 and this is seen in more detail by examining the magnitude and areal extent of the EKE associated with the Gulf Stream (zoom of this region in Figure 11). V3.0 simulated EKE along the core of the current reaches values comparable to the observations ( $2790 \text{ cm}^2/\text{s}^2$ ), although the location of the

simulated maxima does not exactly agree with the observed maxima. This is due to the shorter averaging period of the hindcast and greater smoothing of the observational data analysis. East of 50°W the high EKE associated with the northeastward North Atlantic Current (an extension of the Gulf Stream) is also better represented in V3.0 than V2.5 as is the slightly elevated EKE along 35°N associated with the Azores Current. In the equatorial Atlantic, both models represent the North Brazil Current eddy shedding but V3.0 has higher EKE associated with the North Equatorial Counter Current.

Schmitz (1996) provides a climatological depiction of EKE at 700 m depth in the Gulf Stream region and it is compared against simulated results in Figure 12. The observations are based on the float data from Owens (1984), Owens (1991) and Richardson (1993). Maximum observed values of EKE reach  $\sim 700 \text{ cm}^2/\text{s}^2$  along the core of the Gulf Stream and in V3.0, but V2.5 EKE values are significantly weaker. This implies a more realistic velocity field at depth along with the corresponding eddy activity in V3.0 than V2.5.

### **3.3 Temperature/Salinity vs. Depth Error Analysis**

The analysis in the previous section focused mostly on the surface, but accurate nowcasting and forecasting of the subsurface temperature, salinity and velocity structure is equally important and pertinent to naval operations. The vertical distributions of temperature and to a lesser extent salinity determine the sound speed properties. The near-surface stratification, surface mixed layer and thermocline gradient all play important roles in the sound propagation. Furthermore, the sound speed transmission can be complicated by the occurrence of temperature inversions and subsurface salinity extrema. Therefore, the

predictability of T and S is vital to accurate simulation of the underwater acoustical environment, as discussed in section 3.4.

A temperature and salinity versus depth error analysis was performed on both year-long V3.0 and V2.5 hindcasts using profile data from the Global Ocean Data Assimilation Experiment (GODAE) server. The database was separated into assimilated and unassimilated profiles. As expected, V3.0 is performing much better at assimilated profile locations and this is only briefly discussed in this report. For a given observation, the models were sampled at the nearest gridpoint and interpolated in the vertical to the observation depths.

The world ocean was broken into eight regions as shown in Figure 13. Region MER4d contains the major western boundary currents and additional analysis is performed over the western Pacific Ocean (120-170°E, 20-50°N) and Arabian Sea (45-80°E, 0-24°N). A near-global region defined as MERall comprises the four MERxx regions. Very few profiles existed for validation in the four polar regions (ANTarc, ARCPac, ARCatl and ARCOcn). Thus, no analysis from these regions is discussed in this report because of the paucity of observations that can lead to non-robust statistics and the challenging nature of modeling these nearly isothermal regions. The analysis was broken into boreal seasons defined as summer (June-July-August [JJA]), fall (September-October-November [SON]), winter (December-January-February [DJF]) and spring (March-April-May [MAM]).

The statistical metrics used in this report are mean error (ME), root mean square error (RMSE) and non-dimensional skill score (SS). They are defined as:

$$ME = \bar{Y} - \bar{X},$$

$$RMSE = \left[ \frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2 \right]^{1/2},$$

$$R = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) / (\sigma_x \sigma_y),$$

$$SS = R^2 - [R - (\sigma_y / \sigma_x)]^2 - [(\bar{Y} - \bar{X}) / \sigma_x]^2,$$

where  $\bar{X}(\bar{Y})$  and  $\sigma_x(\sigma_y)$  are the mean and standard deviation of the observed (simulated) data and R is the correlation. Additional information on these statistical measures can be found in Murphy (1995). ME is the annual bias and RMSE the absolute error between the model and data. SS is a non-dimensional quantity based on correlation squared, conditional bias (middle term on the right-hand-side) and unconditional bias (right term on the right-hand-side). The conditional bias is that associated with differences in the standard deviations of the two data sets and unconditional bias is a measure of the difference in the means. A SS of 1.0 is perfect and negative skill score indicates poor performance.

### 3.3.1 Hindcast Experiments: Temperature

A level of uncertainty is associated with the NCODA analysis of the observations and an example of this is highlighted in Figure 14 that shows the ME and RMSE over the near-global region MERall for September 2007. The red lines indicate the error of the analysis (relative to the observations) on the standard NCODA z-levels; it is the lowest error that could be expected from the system. There is essentially no bias and RMSE is less than 0.4°C at all depths; these values are typical for the other months of the hindcast. A very small amount of additional error is introduced in the vertical remapping from NCODA z-levels to NCODA hybrid space (green lines), i.e. onto HYCOM hybrid vertical coordinates. The black curves indicate the error between the assimilated profiles and the

V3.0 analysis at 00Z, i.e. at the end of the six hour period after the NCODA analysis has been incrementally inserted into the model. The blue curves represent the error of the 24-hour HYCOM forecast and they have similar magnitudes to the error analysis against unassimilated profiles that follow later in this section.

As indicated above, over the past year significant progress was made within NCODA to increase accuracy regarding assimilation of the profile observations. This is shown season-by-season in Figure 15 for the near-global region MERall. Similar to Figure 14, the mean error is very small and RMSE peaks just above  $0.5^{\circ}\text{C}$  between  $\sim 50\text{-}200$  m, the depth range of high variability in the mixed layer and thermocline; SS is also very high. As would be expected, V3.0 outperforms V2.5 at these assimilated profile locations.

A comparison of the two systems against unassimilated profiles is shown in Figure 16. Across most of the world ocean, both systems are performing very similarly. The ME is largest in the upper 200 m of the water column, but it is generally small and slightly lower in V2.5 (approximately  $-0.1^{\circ}\text{C}$  over the top 500 m) compared to V3.0 ( $-0.2^{\circ}\text{C}$ ). RMSE reaches a maximum between 50-200 m depending upon the season and V3.0 has less error below 50 m in spring. SS is again high in all seasons for both systems. Focusing on the western boundary current region (MER4d) both systems have a stronger cold bias that is greater than  $0.5^{\circ}\text{C}$  at most depths in spring and summer (Figure 17). V2.5 has higher bias in summer and fall while it is higher in V3.0 in winter and spring. RMSE is generally higher in most seasons compared to the near-global statistics and the maximum error reduces in summer and fall. In spring, RMSE is consistently high down to 500 m and this is most likely associated with the transition between a deep winter-time to shallower spring-time mixed layer and thermocline. Lower RMSE and higher SS in V3.0 suggest that it

responds with greater accuracy to the seasonal change. Zooming in even farther on the western Pacific region (Figure 18) the cold bias increases slightly but the maximum RMSE grows by nearly 0.5°C in some seasons. Note also that where V3.0 shows lower error than V2.5, the spread is larger in this region where the Kuroshio is dominant than in the region that comprises both western boundary currents. Lastly, the temperature vs. depth error analysis is shown for the Arabian Sea in Figure 19. Here the systematic cold bias is replaced by a cool surface bias and warm bias at mid-depths in some seasons. This asymmetry is most likely the cause of the greatly reduced SS in this region.

Overall, this error analysis using unassimilated profiles indicates the two systems are performing with similar accuracy across much of the globe. However, there are both seasons and regions where V3.0 has lower RMSE and higher skill than V2.5, namely the springtime across the globe (and especially in the western Pacific) and the summer and fall in the western boundary current region and again in the western Pacific Ocean.

### **3.3.2 Hindcast Experiments: Salinity**

Both V3.0 and V2.5 use sea surface salinity (SSS) relaxation to keep the upper ocean salinity balance on track and free of long-term drift that may result from inadequate evaporation and precipitation forcing. In V3.0, the e-folding time scale is 30 days times the mixed layer depth (MLD) / 30 m, hence it is stronger when the MLD is shallow and weaker when it is deep. In V2.5, the e-folding time scale also varies as a function of MLD and is twice the value used in V3.0; hence SSS relaxation is only half as strong. The salinity vs. depth error analysis against unassimilated profiles is shown for the near-global region MERall in Figure 20. In the upper 100 m of the water column, V3.0 shows reduced ME

compared to V2.5; this is most likely related to the stronger SSS relaxation in the former system. Below approximately 200 m the bias is essentially zero for both systems in all seasons. The RMSE curves are similar between the two systems in all seasons with slightly lower near surface error in V3.0, but V2.5 holds a slight advantage between 50-200 m in some seasons. For completeness, the salinity vs. depth error analysis is shown for the same regions as the temperature analysis, i.e. Figure 21: the MER4d region, Figure 22: the western Pacific Ocean and Figure 23: the Arabian Sea. The summer-time negative salinity bias in the top 100 m is more prevalent in the two western boundary current regions, otherwise these profiles are similar to the near-global ME statistics. RMSE errors between the whole domain and the western boundary current regions are also similar, but as with the temperature analysis, V3.0 has reduced RMSE with depth in the spring compared to V2.5. The SS for the western Pacific region is much different than that for MER4d. This metric is much more sensitive to the number of profiles used in the calculation and the total number of observations at depth is much smaller than that depicted in the top row of Figure 22. The salinity structure in the Arabian Sea is dominated by salinity maxima at the near-surface (approximately upper 100 m) and a mid-depth (~200-300 m). The former is associated with the high evaporation in the region and the latter by advection of high salinity waters from the Persian Gulf and Red Sea. Figure 23 indicates that both systems do not adequately simulate salinity in this region. Relatively large seasonal biases exist, especially in Fall and near-surface RMSE reaches large values.

### 3.3.3 Forecast Experiments: Salinity

The focus shifts next on the system's ability to forecast temperature and salinity as a function of depth. This is illustrated in Figure 24 that shows temperature vs. depth for 1-day to 5-day V3.0 forecasts over the near-global MERall region for spring 2008. Each forecast is initialized from the hindcast and integrated with forecast quality forcing as described in section 3.1. The profiles used for validation are those that would eventually be assimilated into the system. The results are similar for all regions and seasons, and they indicate the error growth rate over the forecast period is small for both the bias and RMSE. Thus on these time scales, V3.0 is not drifting away from the analysis state nor back to having a subsurface warm bias as was seen in the non-assimilative spin-up. The error growth rate is small but does not monotonically increase because a different set of profiles are used in each forecast analysis. These same results hold for V2.5, but those forecasts only go out to 3 days.

Figure 25 is the same analysis but with V3.0 and V2.5 plotted together for the 1-day to 3-day forecasts; two transition seasons are displayed, spring and Fall. Similar to the hindcasts, the same tendencies related to bias and RMSE exist for the spring (compare the top two rows of Figure 25 with the right-most column in Figure 16). However, for the Fall season, the ME is essentially the same and V3.0 has reduced RMSE below ~50 m in the forecasts (compare the bottom two rows of Figure 25 with the second column in Figure 16). The maximum RMSE in the V3.0 forecasts is essentially the same as in the hindcast results, but it is ~0.2°C higher in V2.5.

### 3.4 MLD/SLD/DSC/BLG Error Analysis

Ocean acoustic models require accurate sound speed fields to characterize the acoustical ducts within the water column. Reliable sound speed depends upon accurate 3-D temperature and salinity fields, including knowledge of the upper ocean mixed layer, i.e. the surface layer with nearly constant temperature and density vs. depth. Along with the MLD, naval operations are also interested in the sonic layer depth (SLD), deep sound channel axis (DSC) and below layer gradient (BLG) (defined below) as these characterize the underwater acoustical environment. In this report, these quantities are derived from Naval Oceanographic Office Reference Publication 33 (RP 33, 1992) with the exception that the sound speed equation by Chen and Millero (1977) and later correction by Millero and Li (1994) is used rather than that by Wilson (1960). SLD can be very sensitive to sound speed and the former equation was found to be more accurate than the latter (R. Helber, 2008, personal communication).

Figure 26 is a schematic representation of the acoustics related quantities validated in this section. MLD is defined by a  $0.1^{\circ}\text{C}$  change from the surface to a given depth. The SLD is the vertical distance from the surface to the depth of the sound speed maximum, often but not always at the base of the mixed layer. Helber et al. (2008) provides a good analysis relating MLD and SLD and the implications on upper ocean acoustical trapping. Below the SLD, sound speed typically decreases with decreasing temperature until the effects of increasing pressure cause sound speed to increase. The depth of the relative minimum below the SLD is the deep sound channel (DSC). Within the deep duct, low frequency sound energy can potentially be carried very long distances, thousands of kilometers in some cases. Lastly, below layer gradient (BLG) is defined as

the sound speed rate of change with depth per 100 feet in the first 300 feet below the SLD or below the surface if the SLD is absent.

To validate these quantities, unassimilated profile observations were used and profiles were sampled at the nearest model gridpoint over the year-long hindcast. The hindcast results were then vertically interpolated to the observation depths and the four quantities were derived. Because of the skewed nature of MLD and SLD, i.e. summertime values can be very shallow and wintertime values much deeper, we define these metrics:

$$\text{Median Bias Error (MdBE)} = \text{median}(\text{model}) - \text{median}(\text{observation})$$

$$\text{Median Absolute Error (MdAE)} = \text{median}(|\text{model} - \text{observation}|)$$

$$\text{Relative MdAE} = \text{MdAE}(\text{V2.5} - \text{observation}) - \text{MdAE}(\text{V3.0} - \text{observation})$$

The relative MdAE is used to determine which system is performing comparatively better, e.g. if the observed MLD = 40 m, V3.0 MLD = 42 m and V2.5 MLD = 44 m, then MdAE (V3.0 – observed) = 2 m and MdAE (V2.5 – observed) = 4 m. The relative MdAE = 2 m, so positive values indicate V3.0 has less absolute error while negative values indicate V2.5 has less absolute error.

In the temperature vs. depth analysis section, the verification against assimilated profiles (Figure 15) indicated the observations are being accurately ingested into V3.0. Similarly, Figure 27 shows relative MdAE for MLD and SLD using assimilated profiles. As expected, V3.0 is performing significantly better than V2.5 over the entire globe and where profile data have been assimilated, these quantities are accurately nowcast.

With regard to unassimilated profiles, the horizontal distribution of MdBE is shown in Figures 28, 30, 32 and 34 while the corresponding relative MdAE is in Figures 29, 31, 33 and 35 for MLD, SLD, DSC and BLG, respectively. The analysis is limited to the  $\pm 50^\circ$

latitude band because poleward of this range the near isothermal nature of the water column can lead to large biases in both systems, especially for MLD and SLD. With regard to MLD (Figure 28), both systems have a shallow bias, especially within the tropical Pacific Ocean. A weaker shallow bias also exists in the Atlantic and Indian Oceans in V3.0 whereas V2.5 exhibits a deep bias in the southern Indian Ocean. This more pervasive shallow bias in V3.0 is probably related to the manner in which MODAS synthetics are used in the system. On their own, the climatological relationships within MODAS relating SSH and SST to subsurface T and S tend to produce shallower than observed mixed layers. These synthetic profiles are then used “as is” in V3.0. However, V2.5 compares mixed layer depth from  $1/32^\circ$  NLOM with that derived from MODAS synthetic profiles and chooses the deeper of the two, making the upper ocean isothermal to that depth. The NLOM MLDs also exhibit a shallow bias, but it is smaller than the bias using straight MODAS synthetics. Similar attempts to modify the MLD calculated from MODAS synthetics in V3.0 were not undertaken for this validation effort and the overall statistics in V2.5 are marginally better than V3.0 (see Tables 2-4). However, the existing 6.4 Ocean Data Assimilation project (a companion to this project) has plans in early FY09 to test similar and more advanced procedures to improve MLD performance in V3.0. Relative MdAE is shown in Figure 29 for the whole domain with zooms on the western Pacific and Arabian Sea. There is limited validation data in the marginal seas of the western Pacific, but V3.0 has a slight advantage in the South China Sea and Japan/East Sea as well as within the Kuroshio south of Japan. The Arabian Sea is one of the few regions where relative MdAE is higher overall in V3.0 than V2.5.

The overall patterns of MdBE in SLD (Figure 30) are very similar to that for MLD, but the bias is generally more negative. The spread of the MdBE and RMSE between V3.0 and V2.5 is also larger (Tables 2-4) than for MLD. Within the western Pacific and Arabian Sea regions, V2.5 has lower absolute error than V3.0. While the patterns and statistics are very similar, both systems exhibit a very strong negative bias and high RMSE for representing the DSC (Figure 32 and Tables 2-4). This bias tends to be largest in the more poleward latitudes of the Pacific Ocean. The relative MdAE of the DSC (Figure 33) is rather patch-like, but V3.0 has less absolute error than V2.5 across the globe and within the western Pacific and Arabian Sea. The patterns of BLG (Figure 34) are also very similar between the two systems with an overall negative bias over most of the globe, indicating a weaker than observed slope. The MdBE is highest in tropical latitudes and this tends to create lower overall statistics (Table 2); however, focusing on the western Pacific, MdBE is reduced and the relative MdAE is comparable in the two systems.

### **3.5 Sea Surface Temperature Error Analysis**

V3.0 assimilates available satellite and *in-situ* SST observations from the daily 5-day hindcast up to the nowcast time; in forecast mode there is no relaxation of SST. V2.5 relaxes to a 2-D MODAS SST analysis up to the nowcast time; in forecast mode, it relaxes to this persisted MODAS SST at each time step by modifying the surface heat fluxes. Because of this strong surface constraint, it is expected that both systems would have relatively small bias and RMSE at the nowcast time and this is born out by an SST error analysis at each of the 48 analysis times and the one through three (or five) day forecasts. The error analysis has been broken down by the data type of the observations: MCSSTs,

drifting buoys, fixed buoys and ship reports (bucket, hull or intake). Each of these types has their own sources of observational error and clearly the ship reports are the least reliable. Generally the fixed buoys provide relatively accurate SST, though the spatial sampling across the globe is rather limited, thus this report will utilize the satellite and drifting buoy observations in the error analysis. These last two observation types are also the most voluminous, especially the satellite data. As described in Cummings (2005), the observations are reduced in number by the formation of “super-observations”, i.e. averaged into larger bins.

The error statistics for ME, RMSE, SS and correlation against the MCSST observations are shown in Tables 5a, 6a and 7a for the globe limited to  $\pm 45^\circ$  latitude, the Northwest Pacific (120-170°E, 20-50°N) and the Arabian Sea (45-80°E, 0-24°N), respectively. The same analysis over the same regions but against the drifting buoy observations is provided in Tables 5b, 6b and 7b. Across most of the globe (Table 5a and 5b), both systems show a cool bias at the analysis time and throughout the forecast period. In V3.0 the ME is very small at the analysis time but probably not exactly zero because of the “super-observation” approach. The bias gradually increased and did not reach values comparable to V2.5 until sometime in between the 3 and 4-day forecasts. The bias in V2.5 at the analysis time may be higher than that in V3.0 because of errors associated with the formation of the 2-D MODAS SST field to which V2.5 is relaxed. Note the bias (and RMSE) does not change significantly as a function of forecast length in V2.5 because the MODAS SST is persisted in time. RMSE in V3.0 is also lower at the analysis time and it again reaches values comparable to V2.5 by about the 3-day forecast. SS and correlation values are generally always very high and are listed for completeness. Focusing on the

northwest Pacific Ocean (Tables 6a and 6b) and comparing similar data types to the larger region, both the ME and RMSE are higher in V3.0 and the error growth increases more rapidly. This increase is likely associated with small misplacement of the simulated Kuroshio and eddy field. The error growth rate in V2.5 is also higher in this region compared to the larger region because use of a persisted SST is not optimal near this rapidly changing western boundary current. In the Arabian Sea (Tables 7a and 7b), the difference in the ME between V3.0 and V2.5 at the analysis time is the largest of the three regions examined. In addition, the ME out to a 5-day V3.0 forecast is smaller than that at the analysis time (or throughout the forecast) in V2.5. For the MCSST comparison, the RMSE at the 4-day V3.0 forecast is similar to that at the analysis time for V2.5. Shriver et al. (2007) did a quantitative comparison of the eddy field in this region and compared simulated SSH from  $1/16^\circ$  and  $1/32^\circ$  NLOM with ocean color observations. One of their conclusions was that increased horizontal resolution had a large impact on the ability of the assimilative system to accurately depict the eddy regime in this region. This SST error analysis also suggests the finer horizontal resolution in V3.0 is leading to a more accurate SST analysis and forecast.

The spatial distribution of the ME relative to the MCSST satellite observation is shown in Figure 36 for the nowcast and clearly there is less error in V3.0 than V2.5. At the 3-day forecast (Figure 37), the patterns of ME are similar in the two systems with relatively high bias in the eastern equatorial Pacific, along the west coast of the U.S. and in the northern Japan/East Sea.

### 3.6 Coastal/Island Sea Level Error Analysis

An error analysis has been performed against simulated vs. observed daily sea level obtained from the Joint Archive for Sea Level Center at the University of Hawaii Sea Level Center (Caldwell and Merrifield, 1992). The location of the 147 stations used in the analysis is depicted in Figure 38 and it includes both open ocean island stations and coastal stations. The observations have been de-tided and atmospheric pressure loading effects have been removed similar to the methodology described in Barron et al. (2004). The sea level change can be a deterministic response to the atmospheric forcing or non-deterministic and associated with mesoscale flow instabilities. Since relatively fine horizontal resolution atmospheric forcing is used and both ocean prediction systems employ data assimilation, the simulated sea level should be accurately represented. Shown in Figure 39 are histograms of correlation and RMSE for the year-long hindcasts. The shorter “tails” on the histogram plots indicate overall higher correlation and lower RMSE in V3.0 than V2.5 which is quantified by median correlation of 0.80 (0.73) and median RMSE of 5.8 cm (6.3 cm) in V3.0 (V2.5). The percentage of points in V3.0 (V2.5) with correlation higher than or equal to the bar centered on .8 is 68% (31%) and the percentage of points in V3.0 (V2.5) with RMSE lower than or equal to the bar centered on 6 cm is 68% (49%).

The error growth rate as a function of forecast length is relatively small as shown in Figure 40, which depicts correlation and RMSE averaged for all 147 stations at the analysis time and for the 1-day through 3-day forecasts. Forecast sea level was compared against the observed sea level at the proper date and averaged across all stations. V2.5 exhibits a weak trend of decreasing correlation and increasing RMSE as a function of forecast length,

whereas V3.0 actually shows slightly higher correlation and lower RMSE for the one and two day forecasts. The sample size (48 nowcasts or forecasts) may be too small to meaningfully interpret these trends, but the overall median values are similar to the year-long hindcast statistics that use 366 records.

Lastly, Figure 41 shows an example of simulated and observed sea level along the west coast of the U.S. at Neah Bay, WA. Note the short time scale, yet often large amplitude change in sea level, in particular the event near the beginning of December 2007 where sea level rises nearly 50 cm. While the phase of the sea level signal is good in both systems, V3.0 more accurately simulates the amplitudes. While not present in this hindcast, short time scale sea level increases greater than 1 m have been seen along the west coast of the U.S. in other HYCOM hindcasts. Such a large sea level change could have an impact on naval operations and V3.0 more accurately simulates these events. These large events can be associated with coastally trapped waves generated from tropical cyclones (Zamudio et al., 2002) or coastally trapped Kelvin waves generated during El Niño events (Kessler et al., 1995).

#### **4.0 SUMMARY AND RECOMMENDATIONS**

Global Ocean Prediction System V3.0 is comprised of the 1/12° global HYbrid Coordinate Ocean Model and the Navy Coupled Ocean Data Assimilation system. It is a next-generation system capable of nowcasting and forecasting the oceanic “weather”, some components of which include the three dimensional ocean temperature, salinity and current structure, the surface mixed layer and the location of mesoscale features such as eddies, meandering currents and fronts. The system uses 3-hourly forcing from the Navy

Operational Global Atmospheric Prediction System. This report described the validation testing performed on V3.0 and the nowcast/forecast system based on 1/8° Navy Coastal Ocean Model, 1/32° Navy Layered Ocean Model and 1/8° Modular Ocean Data Analysis System (V2.5).

The validation testing is designed to occur in phases because of the considerable resources (both manpower and computational) required. The next validation test phase will include comparison to V2.6. The phase I validation tasks include evaluation of a) the large scale circulation features, b) sea surface height variability and eddy kinetic energy, c) vertical temperature and salinity structure, d) quantities that describe the underwater acoustical environment such as mixed layer depth, sonic layer depth, deep sound channel axis and below layer gradient, e) sea surface temperature and f) coastal/island sea level variability. Error analyses have been performed on both hindcasts and forecasts, with the latter providing estimates of the system's predictive skill.

With regard to the validation tasks defined above, the large scale circulation features are simulated marginally better in V3.0 than V2.5. A quantitative comparison of simulated vs. observed mean sea level from Maximenko and Niiler (2005) indicate lower standard deviation of the difference between the two datasets for V3.0. The SSH variability and eddy kinetic energy fields provide a good measure of the oceanic mesoscale eddy field. V3.0 is the more energetic system and its SSH variability compares quantitatively closer to satellite based observations than that from V2.5. At the surface and at depth (700 m), the eddy kinetic energy fields in V3.0 qualitatively shows better agreement with observations based on float data in the North Atlantic. Thus it simulates a more realistic basin-wide eddy environment.

A seasonal error analysis of temperature/salinity vs. depth indicates the two systems are performing similarly. Across most of the globe, the subsurface temperature bias is small and RMSE reaches a maximum between ~50-200 m. In one region of high Navy interest, the western Pacific Ocean, V3.0 outperforms V2.5 with regard to RMSE in the spring and summer seasons. In the more challenging Arabian Sea, both systems perform very similarly with somewhat high, nonsystematic bias and RMSE approaching 2°C near 100 m depth in some seasons. The salinity vs. depth error analysis indicates a smaller near-surface (top 100 m) salinity bias for V3.0 than V2.5. Neither system has a large error growth rate as a function of forecast length for temperature or salinity vs. depth.

The underwater acoustical environment can be defined by the mixed layer depth, sonic layer depth, deep sound channel and below layer gradient. A negative bias exists in the simulated depths and gradients for both systems. For MLD, median bias error is slightly larger in V3.0 across the globe, but it has less absolute error in the Arabian Sea and in the marginal seas and Kuroshio region of the western Pacific. The difference between the two systems is largest for the SLD and here V2.5 outperforms V3.0 across most of the globe and within the areas of high Navy interest. This more pervasive shallow bias in V3.0 may be related to the manner in which MODAS synthetics are used in the system. On their own, the climatological relationships within MODAS relating SSH and SST to subsurface T and S tend to produce shallower than observed mixed layers. These synthetic profiles are then used “as is” in V3.0. However, V2.5 compares mixed layer depth from 1/32° NLOM with that derived from MODAS synthetic profiles and chooses the deeper of the two, making the upper ocean isothermal to that depth. Efforts will be

undertaken in FY09 to make similar mixed layer adjustments in V3.0. Analysis of the deep sound channel and below layer gradient indicates similar performance by both systems.

Sea surface temperature is assimilated up to the nowcast time. An error analysis indicates the bias and RMSE are relatively small, especially in V3.0. Across most of the globe, the bias and RMSE of a 3-day to 4-day V3.0 forecast is similar in magnitude to that from the V2.5 nowcast. The V3.0 improvement over V2.5 is greatest in the Arabian Sea. Regarding the simulation of coastal/island sea level, V3.0 substantially outperforms V2.5 in terms of correlation and RMSE statistics during the year-long hindcast.

Overall, this report has determined that V3.0 is performing equal to or marginally to notably better than V2.5 for all validation tasks with the exception of the sonic layer depth. The companion Ocean Data Assimilation project has plans in FY09 to improve the representation of the MLD in V3.0, and thus it seems appropriate to re-evaluate the SLD analysis after those improvements are in place.

## **5.0 ACKNOWLEDGEMENTS**

This work was funded as part of the NRL 6.4 Large Scale Prediction and 6.4 Ocean Data Assimilation projects, managed by the Space and Naval Warfare Systems Command under program element 0603207N. The numerical simulations were performed on the NAVOCEANO IBM-Power 4+ and Power 5+ at Stennis Space Center, Mississippi using grants of computer time from the Department of Defense High Performance Computing Modernization Program. Critical evaluation and advice was provided by the members of the review panel: Frank Bub, Chris DeHaan, Andrea Mask and Bruce Lunde (NAVOCEANO), Joe Metzger and Harley Hurlburt (NRL) and Mike Carron (Northern Gulf Institute). The authors would like to thank the various NRL contributors for their support in global

HYCOM/NCODA development and validation. Among them are James Cummings for continued development of NCODA, Birol Kara for his work on the mixed layer analysis, Bob Helber for the software and expert advice with the MLD/SLD/DSC/BLG calculations, Charlie Barron for technical discussions related to NCOM and Tammy Townsend for clarifying the time frame of the NCOM system developments.

## 6.0 REFERENCES

- Barron, C.N., A.B. Kara, H.E. Hurlburt, C. Rowley and L.F. Smedstad, 2004: Sea surface height predictions from the Global Navy Coastal Ocean Model during 1998-2001. *J. Atmos. Oceanic Technol.*, **21**, 1876-1893.
- Barron, C.N., A.B. Kara, P.J. Martin, R.C. Rhodes and L.F. Smedstad, 2006: Formulation, implementation and examination of vertical coordinate choices in the Global Navy Coastal Ocean Model (NCOM). *Ocean Modelling*, **11**, 347-375.
- Barron, C.N., A. B. Kara, R. C. Rhodes, C. Rowley and L. F. Smedstad, 2007a: Validation Test Report for the 1/8° Global Navy Coastal Ocean Model Nowcast/Forecast System. *NRL Tech Report* NRL/MR/7320--07-9019.
- Barron, C.N., L. F. Smedstad, J. M. Dastugue and O. M. Smedstad, 2007b: Evaluation of ocean models using observed and simulated drifter trajectories: Impact of sea surface height on synthetic profiles for data assimilation. *J. Geophys. Res.*, **112**, C07019, doi:10.1029/2006JC002982.
- Bitz, C.M. and W.H. Lipscomb, 1999: An energy-conserving thermodynamic model of sea ice. *J. Geophys. Res.*, **104**, 15669-15677.
- Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Modelling*, **4**, 55-88.
- Caldwell, P. and M. Merrifield, 1992: Building an archive of tropical sea level. *Earth System Monitor*, **3**, 3-6.
- Chassignet, E.P., L.T. Smith, G.R. Halliwell and R. Bleck, 2003: North Atlantic simulations with the HYbrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference pressure, and thermobaricity. *J. Phys. Oceanogr.*, **33**(12), 2504-2526.
- Chen, C.T. and F.J. Millero, 1977: Speed of sound in seawater at high pressures. *J. Acoust. Soc. Am.*, **62**, 1129-1135.
- Cooper, M. and K.A. Haines, 1996: Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **24**, 1059-1077.
- Cummings, J.A., 2005. Operational multivariate ocean data assimilation. *Quart. J. Royal Met. Soc.*, **131**, 3583-3604.
- Fox, D.N., W.J. Teague, C.N. Barron, M.R. Games and C.M. Lee, 2002: The Modular Ocean Data Analysis System (MODAS). *J. Atmos. Oceanic Technol.*, **19**, 240-252.

- Fratantoni, D.M., 2001: North Atlantic surface circulation during the 1990's observed with satellite-tracked drifters. *J. Geophys. Res.*, **106**, 22067-22093.
- Halliwel, G. R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid Coordinate Ocean Model (HYCOM), *Ocean. Model.*, **7**(3–4), 285–322.
- Helber, R.W., C.N. Barron, M.R. Carnes and R.A. Zingarelli, 2008: Evaluating the sonic layer depth relative to the mixed layer, *J. Geophys. Res.*, **113**, C07033, doi:10.1029/2007JC004595.
- Hill C., C. DeLuca, V. Balaji, M. Suarez, A. da Silva, 2004: The Architecture of the Earth System Modeling Framework. *Computing in Science and Engineering*, **6**, 18-28.
- Hogan, P.J. and H.E. Hurlburt, 2000: Impact of upper ocean – topographic coupling and isopycnal outcropping in Japan/East Sea models with 1/8° to 1/64° resolution. *J. Phys. Oceanogr.* **30**, 2535–2561.
- Hurlburt, H.E., E.P. Chassignet, J.A. Cummings, A.B. Kara, E.J. Metzger, J.F. Shriver, O.M. Smedstad, A.J. Wallcraft and C.N. Barron, 2008a: Eddy-resolving global ocean prediction. In: Hecht, M., Hasumi, H. (Eds.), *Ocean Modeling in an Eddy Regime*, Geophysical Monograph 177. American Geophysical Union, Washington, DC, pp. 353-381.
- Hurlburt, H.E. and P.J. Hogan, 2008: The Gulf Stream pathway and the impacts of the eddy-driven abyssal circulation and the Deep Western Boundary Current. *Dyn. Atmos. Oceans*, **45**, 71–101.
- Hurlburt, H.E., E.J. Metzger, P.J. Hogan, C.E. Tilburg and J.F. Shriver, 2008b: Steering of upper ocean currents and fronts by the topographically constrained abyssal circulation. *Dyn. Atmos. Oceans*, **45**, 102-134, doi:10.1016/j.dynatmoce.2008.06.003.
- Hurlburt, H.E., R.C. Rhodes, C.N. Barron, E.J. Metzger, O.M. Smedstad, and J.-F. Cayula, 2000: A feasibility demonstration of ocean model eddy-resolving nowcast/forecast skill using satellite altimeter data, Naval Research Laboratory Tech. Report NRL /MR/7320--00-8235, 23 pp., Naval Research Laboratory, Stennis Space Center, MS.
- Hurlburt, H.E., A.J. Wallcraft, W.J. Schmitz Jr., P.J. Hogan and E.J. Metzger, 1996. Dynamics of the Kuroshio/Oyashio current system using eddy-resolving models of the North Pacific Ocean. *J. Geophys. Res.*, **101 (C1)**, 941–976.
- Kara, A.B., C.N. Barron, P.J. Martin, L.F. Smedstad and R.C. Rhodes, 2006: Validation of interannual simulations from the 1/8° global Navy Coastal Ocean Model (NCOM). *Ocean Modelling*, **11**, 376-398.
- Kessler, W. S., M. J. McPhaden, and K. M. Weickmann (1995), Forcing of intraseasonal Kelvin waves in the equatorial Pacific, *J. Geophys. Res.*, **100**, 10,613–10,631.

- Large, W. G., J. C. Mc Williams, and S. C. Doney, 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* **32**, 363-403.
- Lee, T. and P. Cornillon, 1996: Propagation and growth of Gulf Stream meanders between 75°W and 45°W. *J. Phys. Oceanogr.*, **26**, 225-241.
- Liang, W.D., T.Y. Yang, M.T. Ko and W.S. Chuang, 2003: Upper-ocean currents around Taiwan, *Deep Sea Res. II*, **6-7**, 1085-1105.
- Lipscomb, W.H. and E. Hunke, 2004: Modeling sea ice transport using incremental remapping. *Mon. Weather Rev.*, **132**, 1341-1354.
- Metzger, E.J. and H.E. Hurlburt, 2001: The nondeterministic nature of Kuroshio penetration and eddy shedding in the South China Sea. *J. Phys. Oceanogr.*, **31**, 1712-1732.
- Millero, F.J. and X. Li, 1994: Comments on “On equations for the speed of sound in seawater”. *J. Acoust. Soc. Am.*, **95**, No. 5, 2757-2759.
- Maximenko, N. A. and P.P. Niiler, 2005: Hybrid decade-mean global sea level with mesoscale resolution, in *Recent Advances in Marine Science and Technology, 2004*, edited by N. Saxena, pp. 55–59, Honolulu: PACON International.
- Murphy, A.H., 1995: The coefficients of correlation and determination as measures of performance in forecast verification. *Wea. Forecasting*, **10**, 681-688.
- Owens, W.B., 1991: A statistical description of the mean circulation and eddy variability in the northwestern Atlantic using SOFAR floats. *Prog. Oceanog.*, **28**, 257-303.
- Owens, W.B., 1984: A Synoptic and statistical description of the Gulf Stream and subtropical gyre using SOFAR floats. *J. Phys. Oceanogr.*, **14**, 104-113.
- Posey, P.G., L.F. Smedstad, R.H. Preller, E.J. Metzger and S.N. Carroll, 2008: User’s Manual for the Polar Ice Prediction System (PIPS) Version 3.0. *NRL Tech. Report*, NRL/MR/7320—08-9154.
- Rhodes, R.C., H.E. Hurlburt, A.J. Wallcraft, C.N. Barron, P.J. Martin, O.M. Smedstad, S.L. Cross, E.J. Metzger, J.F. Shriver, A.B. Kara and D.S. Ko, 2002: Navy real-time global modeling systems, *Oceanography*, **15(1)**, 29-43.
- Richardson, P.L., 1993: A census of eddies observed in North Atlantic SOFAR float data. *Prog. Oceanog.*, **31**, 1-50.

- RP 33: Fleet Oceanographic and Acoustic Reference Manual, 1992: Naval Oceanographic Office Reference Publication 33, Naval Oceanographic Office, Stennis Space Center, MS.
- Schmitz, W.J., 1996: *On the World Ocean Circulation: Volume I. Some Global Features/North Atlantic Circulation*. Woods Hole Oceanog. Inst. Tech. Rept. WHOI-96-03, 150 pp.
- Shriver, J.F., H.E. Hurlburt, O.M. Smedstad, A.J. Wallcraft and R.C. Rhodes, 2007: 1/32° real-time global ocean prediction and value-added over 1/16° resolution. *J. Mar. Sys.*, **65**, 3-26.
- Smedstad, O.M., H.E. Hurlburt, E.J. Metzger, R.C. Rhodes, J.F. Shriver, A.J. Wallcraft, and A.B. Kara, 2003: An operational eddy-resolving 1/16° global ocean nowcast/forecast system. *J. Mar. Sys.*, **40-41**, 341-361.
- Wilson, W.D., 1960: Equation for the speed of sound in sea water. *J. Acoust. Soc. Amer.*, **32**, 1357.
- Zamudio, L., H.E. Hurlburt, E.J. Metzger and O.M. Smedstad, 2002: On the evolution of coastally trapped waves generated by Hurricane Juliette along the Mexican West Coast. *Geophys. Res. Lett.*, **29(23)**, 2141, doi:10.1029/2002GL014769.

## 7.0 TABLE OF ACRONYMS

|           |   |
|-----------|---|
| ADCP      | Acoustic Doppler Current Profiler                     |
| BLG       | Below Layer Gradient                                  |
| CH        | Cooper-Haines   |
| CLS       | Collecte Localisation Satellites                      |
| DSC       | Deep Sound Channel                                    |
| ECMWF     | European Centre for Medium-Range Weather Forecasts    |
| EKE       | Eddy Kinetic Energy                                   |
| ESMF      | Earth System Modeling Framework                       |
| FNMOC     | Fleet Numerical Meteorology and Oceanography Center   |
| GODAE     | Global Ocean Data Assimilation Experiment             |
| GOPS      | Global Ocean Prediction System                        |
| GRACE     | Gravity Recovery and Climate Experiment               |
| HYCOM     | HYbrid Coordinate Ocean Model                         |
| KPP       | K-Profile Parameterization                            |
| MdAE      | Median Absolute Error                                 |
| MdBE      | Median Bias Error                                     |
| MDOT      | Mean Dynamic Ocean Topography                         |
| ME        | Mean Error  |
| MLD       | Mixed Layer Depth                                     |
| MODAS     | Modular Ocean Data Analysis System                    |
| MVOI      | MultiVariate Optimal Interpolation                    |
| NAVOCEANO | Naval Oceanographic Office                            |
| NCEP      | National Centers for Environmental Prediction         |
| NCODA     | Navy Coupled Ocean Data Assimilation                  |
| NCOM      | Navy Coastal Ocean Model                              |
| NLOM      | NRL Layered Ocean Model                               |
| NOGAPS    | Navy Operational Global Atmospheric Prediction System |
| NRL       | Naval Research Laboratory                             |
| PIPS      | Polar Ice Prediction System                           |
| QuikSCAT  | Quick Scatterometer                                   |
| RMSE      | Root Mean Square Error                                |
| S         | Salinity  |
| SLD       | Sonic Layer Depth                                     |
| SS        | Skill Score   |
| SSH       | Sea Surface Height                                    |
| SSMI      | Special Sensor Microwave Imager                       |
| SSS       | Sea Surface Salinity                                  |
| SST       | Sea Surface Temperature                               |
| T         | Temperature   |

| <b>Table 1: Validation tasks for GOPS V3.0</b>   |   |
|--|---|
| <b>Phase I</b>   |   |
| 1. Large scale circulation features  | Determine the correct placement of large-scale features such as gyres, strong fronts, and currents.   |
| 2. SSH variability/Eddy Kinetic Energy (EKE)   | Determine if the system has a realistic level and distribution of energy at depths where observations are available.                                      |
| 3. Mixed layer depth (MLD) / sonic layer depth (SLD) / deep sound channel (DSC) / below layer gradient (BLG) | Compare simulated quantities with that from non-assimilated profile data (e.g. XBTs, Argo floats, CTDs and moored buoys) for both nowcasts and forecasts. |
| 4. Vertical profiles of temperature (T) and salinity (S)   | Perform a quantitative comparison of simulated vs. observed non-assimilated profiles for both nowcast and forecasts.                                      |
| 5. Sea surface temperature   | Evaluate whether the models are producing acceptable nowcasts and forecasts of sea surface temperature.   |
| 6. Coastal sea level   | Assess the model's ability to represent observed sea surface heights for nowcast and forecasts.   |
| <b>Phase II</b>  |   |
| 7. Provide boundary conditions to nested models  | Nest Relocatable NCOM within V3.0 and compare inner model with the solution forced by V2.5/V2.6.  |
| 8. Comparison with drifting buoys  | Evaluate the system's ability to produce ocean currents that yield drifter and Argo float trajectories similar to observations.                           |
| 9. Current cross-sections  | Evaluate model velocity cross-sections through qualitative and quantitative comparisons.  |
| 10. Eddy tracking  | Evaluate the system's ability to track mesoscale eddies.  |
| 11. Long-term forecast skill (up to 30 days)   | Evaluate forecasts of T&S vs. depth relative to persistence and climatology and vs. nowcasts of SSH mapping of ocean fronts and eddies.                   |

|     | number of unassim profiles | MdBE               |                    | RMSE              |                   | Relative MdAE (V3.0 $\leq$ V2.5) |
|-----|----------------------------|--------------------|--------------------|-------------------|-------------------|----------------------------------|
|     |                            | V3.0               | V2.5               | V3.0              | V2.5              |                                  |
| MLD | 66387                      | -7 m               | -3 m               | 40 m              | 39 m              | 44%                              |
| SLD | 50681                      | -16 m              | -10 m              | 67 m              | 60 m              | 42%                              |
| DSC | 40187                      | -49 m              | -49 m              | 200 m             | 187 m             | 57%                              |
| BLG | 66495                      | -0.4<br>m/s/100 ft | -0.4<br>m/s/100 ft | 2.2<br>m/s/100 ft | 2.0<br>m/s/100 ft | 47%                              |

|     | number of unassim profiles | MdBE               |                    | RMSE              |                   | Relative MdAE (V3.0 $\leq$ V2.5) |
|-----|----------------------------|--------------------|--------------------|-------------------|-------------------|----------------------------------|
|     |                            | V3.0               | V2.5               | V3.0              | V2.5              |                                  |
| MLD | 9700                       | -7 m               | -1 m               | 40 m              | 41 m              | 37%                              |
| SLD | 6792                       | -20 m              | -10 m              | 60 m              | 54 m              | 34%                              |
| DSC | 6597                       | -60 m              | -60 m              | 201 m             | 203 m             | 58%                              |
| BLG | 9699                       | -0.3<br>m/s/100 ft | -0.1<br>m/s/100 ft | 2.3<br>m/s/100 ft | 2.2<br>m/s/100 ft | 46%                              |

|     | number of unassim profiles | MdBE               |                    | RMSE              |                   | Relative MdAE (V3.0 $\leq$ V2.5) |
|-----|----------------------------|--------------------|--------------------|-------------------|-------------------|----------------------------------|
|     |                            | V3.0               | V2.5               | V3.0              | V2.5              |                                  |
| MLD | 1537                       | -1 m               | 4 m                | 21 m              | 21 m              | 59%                              |
| SLD | 1082                       | -20 m              | -10 m              | 30 m              | 22 m              | 30%                              |
| DSC | 387                        | 0 m                | 0 m                | 120 m             | 114 m             | 57%                              |
| BLG | 1536                       | -0.8<br>m/s/100 ft | -0.8<br>m/s/100 ft | 2.3<br>m/s/100 ft | 2.2<br>m/s/100 ft | 48%                              |

| <b>Table 5a: SST error statistics vs. ~33,000,000 MCSST observations<br/>Limited between 45°S – 45°N</b> |            |      |      |      |             |      |             |      |
|--|------------|------|------|------|-------------|------|-------------|------|
|  | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|  | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis   | -.02       | -.22 | .36  | .61  | .99         | .98  | 1.0         | .99  |
| 1-d fcst   | -.09       | -.19 | .44  | .61  | .99         | .98  | 1.0         | .99  |
| 2-d fcst   | -.14       | -.19 | .52  | .62  | .99         | .98  | .99         | .99  |
| 3-d fcst   | -.18       | -.20 | .60  | .63  | .98         | .98  | .99         | .99  |
| 4-d fcst   | -.22       | -    | .67  | -    | .98         | -    | .99         | -    |
| 5-d fcst   | -.26       | -    | .72  | -    | .98         | -    | .99         | -    |

| <b>Table 5b: SST error statistics vs. ~885,000 drifting buoy observations<br/>Limited between 45°S – 45°N</b> |            |      |      |      |             |      |             |      |
|---|------------|------|------|------|-------------|------|-------------|------|
|   | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|   | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis  | -.03       | -.22 | .27  | .64  | 1.0         | .99  | 1.0         | .99  |
| 1-d fcst  | -.10       | -.21 | .43  | .65  | .99         | .98  | 1.0         | .99  |
| 2-d fcst  | -.15       | -.21 | .55  | .67  | .99         | .98  | .99         | .99  |
| 3-d fcst  | -.19       | -.21 | .64  | .68  | .98         | .98  | .99         | .99  |
| 4-d fcst  | -.23       | -    | .71  | -    | .98         | -    | .99         | -    |
| 5-d fcst  | -.25       | -    | .77  | -    | .98         | -    | .99         | -    |

| <b>Table 6a: SST error statistics vs. ~1,170,000 MCSST observations<br/>Northwest Pacific: Limited between 120-170°E and 20-50°N</b> |            |      |      |      |             |      |             |      |
|--|------------|------|------|------|-------------|------|-------------|------|
|  | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|  | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis   | -.15       | -.33 | .49  | .71  | 1.0         | .99  | 1.0         | 1.0  |
| 1-d fcst   | -.27       | -.32 | .63  | .72  | .99         | .99  | 1.0         | 1.0  |
| 2-d fcst   | -.36       | -.34 | .76  | .75  | .99         | .99  | 1.0         | 1.0  |
| 3-d fcst   | -.45       | -.36 | .88  | .80  | .99         | .99  | 1.0         | 1.0  |
| 4-d fcst   | -.53       | -    | .99  | -    | .98         | -    | .99         | -    |
| 5-d fcst   | -.59       | -    | 1.08 | -    | .98         | -    | .99         | -    |

| <b>Table 6b: SST error statistics vs. ~35,700 drifting buoy observations<br/>Northwest Pacific: Limited between 120-170°E and 20-50°N</b> |            |      |      |      |             |      |             |      |
|---|------------|------|------|------|-------------|------|-------------|------|
|   | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|   | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis  | -.07       | -.19 | .39  | .91  | 1.0         | .98  | 1.0         | .99  |
| 1-d fcst  | -.22       | -.19 | .63  | .93  | .99         | .98  | 1.0         | .99  |
| 2-d fcst  | -.32       | -.20 | .84  | .96  | .98         | .98  | .99         | .99  |
| 3-d fcst  | -.42       | -.21 | 1.02 | 1.00 | .97         | .97  | .99         | .99  |
| 4-d fcst  | -.48       | -    | 1.15 | -    | .97         | -    | .99         | -    |
| 5-d fcst  | -.53       | -    | 1.25 | -    | .96         | -    | .99         | -    |

| <b>Table 7a: SST error statistics vs. ~1,100,000 MCSST observations</b> |            |      |      |      |             |      |             |      |
|---|------------|------|------|------|-------------|------|-------------|------|
| <b>Arabian Sea: Limited between 45-80°E and 0-24°N</b>                  |            |      |      |      |             |      |             |      |
|   | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|   | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis  | -.08       | -.42 | .40  | .68  | .94         | .83  | .97         | .95  |
| 1-d fcst  | -.18       | -.41 | .47  | .66  | .92         | .84  | .97         | .95  |
| 2-d fcst  | -.24       | -.40 | .54  | .65  | .89         | .84  | .96         | .95  |
| 3-d fcst  | -.29       | -.40 | .61  | .65  | .86         | .84  | .95         | .95  |
| 4-d fcst  | -.34       | -    | .67  | -    | .84         | -    | .94         | -    |
| 5-d fcst  | -.38       | -    | .73  | -    | .81         | -    | .93         | -    |

| <b>Table 7b: SST error statistics vs. ~15,800 drifting buoy observations</b> |            |      |      |      |             |      |             |      |
|--|------------|------|------|------|-------------|------|-------------|------|
| <b>Arabian Sea: Limited between 45-80°E and 0-24°N</b>                       |            |      |      |      |             |      |             |      |
|  | Mean Error |      | RMSE |      | Skill Score |      | Correlation |      |
|  | V3.0       | V2.5 | V3.0 | V2.5 | V3.0        | V2.5 | V3.0        | V2.5 |
| Analysis   | -.09       | -.45 | .32  | .75  | .94         | .68  | .97         | .91  |
| 1-d fcst   | -.18       | -.44 | .43  | .73  | .89         | .70  | .96         | .92  |
| 2-d fcst   | -.25       | -.42 | .53  | .71  | .84         | .72  | .94         | .92  |
| 3-d fcst   | -.29       | -.41 | .61  | .69  | .79         | .73  | .93         | .92  |
| 4-d fcst   | -.33       | -    | .67  | -    | .74         | -    | .91         | -    |
| 5-d fcst   | -.35       | -    | .71  | -    | .71         | -    | .91         | -    |



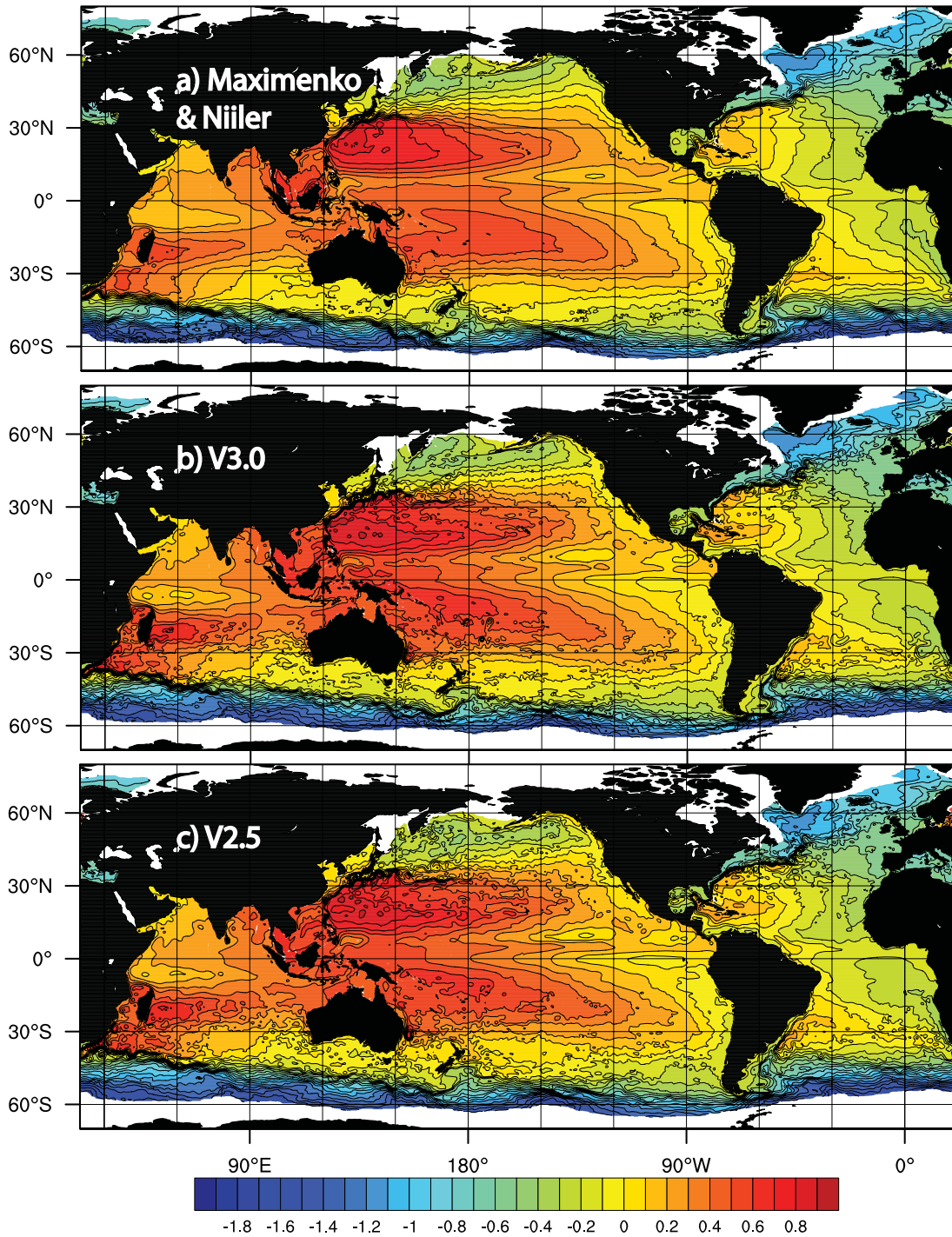


Figure 2: Mean sea surface height (SSH) in meters from a) Maximenko and Niiler (2005) based on satellite altimeter, drifting buoy and wind data spanning 1992–2002 and an improved geoid, GRACE (Gravity Recovery and Climate Experiment) Gravity Model 01 (GGM01), b) Global Ocean Prediction System (GOPS) Version 3.0 (V3.0) and c) GOPS V2.5. The simulated means span the hindcast time period 1 June 2007 – 31 May 2008. In all panels the white area near the poles is a 1982–2007 annual mean sea ice coverage mask from the Climate Diagnostics Center optimum interpolation SST analysis. The non-ice areal mean has been subtracted from each panel before plotting and is 12.3/12.6/13.0 cm respectively for the three panels.

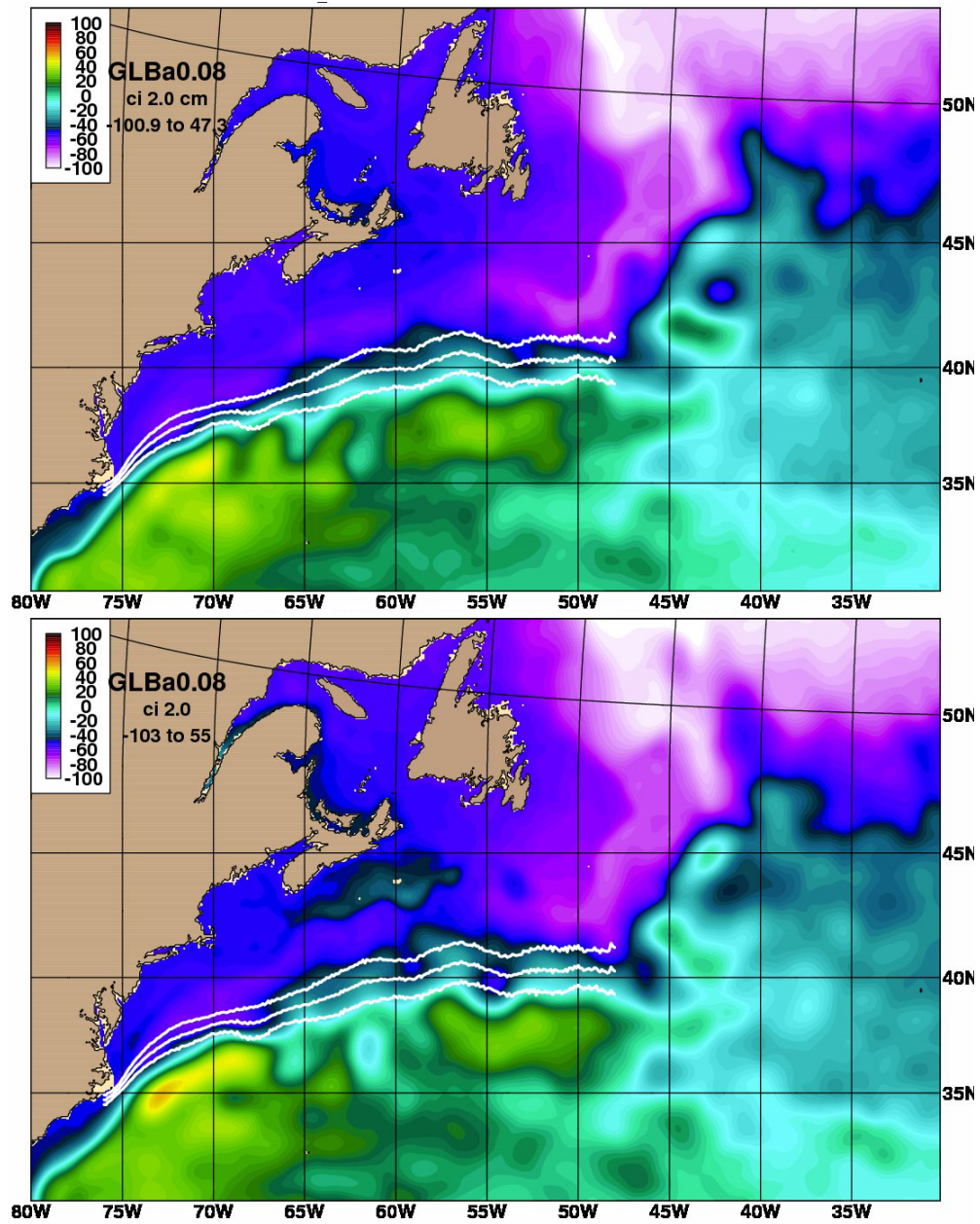


Figure 3: Mean SSH in the Gulf Stream region from V3.0 (top) and V2.5 (bottom) over the hindcast period 1 June 2007 – 31 May 2008. Overlaid is the observed mean north wall position of the Gulf Stream ( $\pm 1$  standard deviation) based on satellite SST.

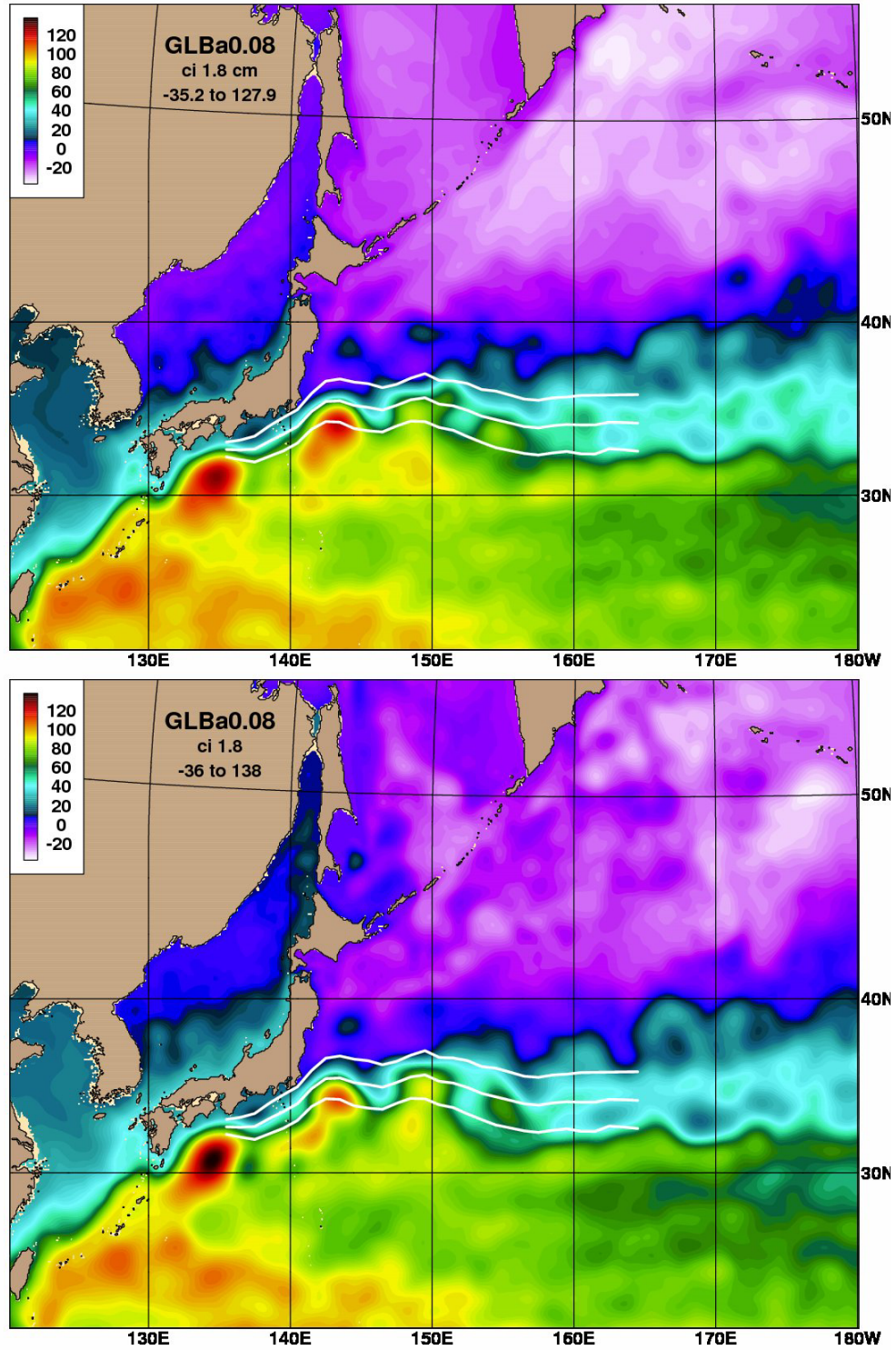


Figure 4: Mean SSH in the Kuroshio region from V3.0 (top) and V2.5 bottom over the hindcast period 1 June 2007 – 31 May 2008. Overlaid is the observed 1992-2004 mean axis of the Kuroshio ( $\pm$  1 standard deviation) based on satellite altimeter data.

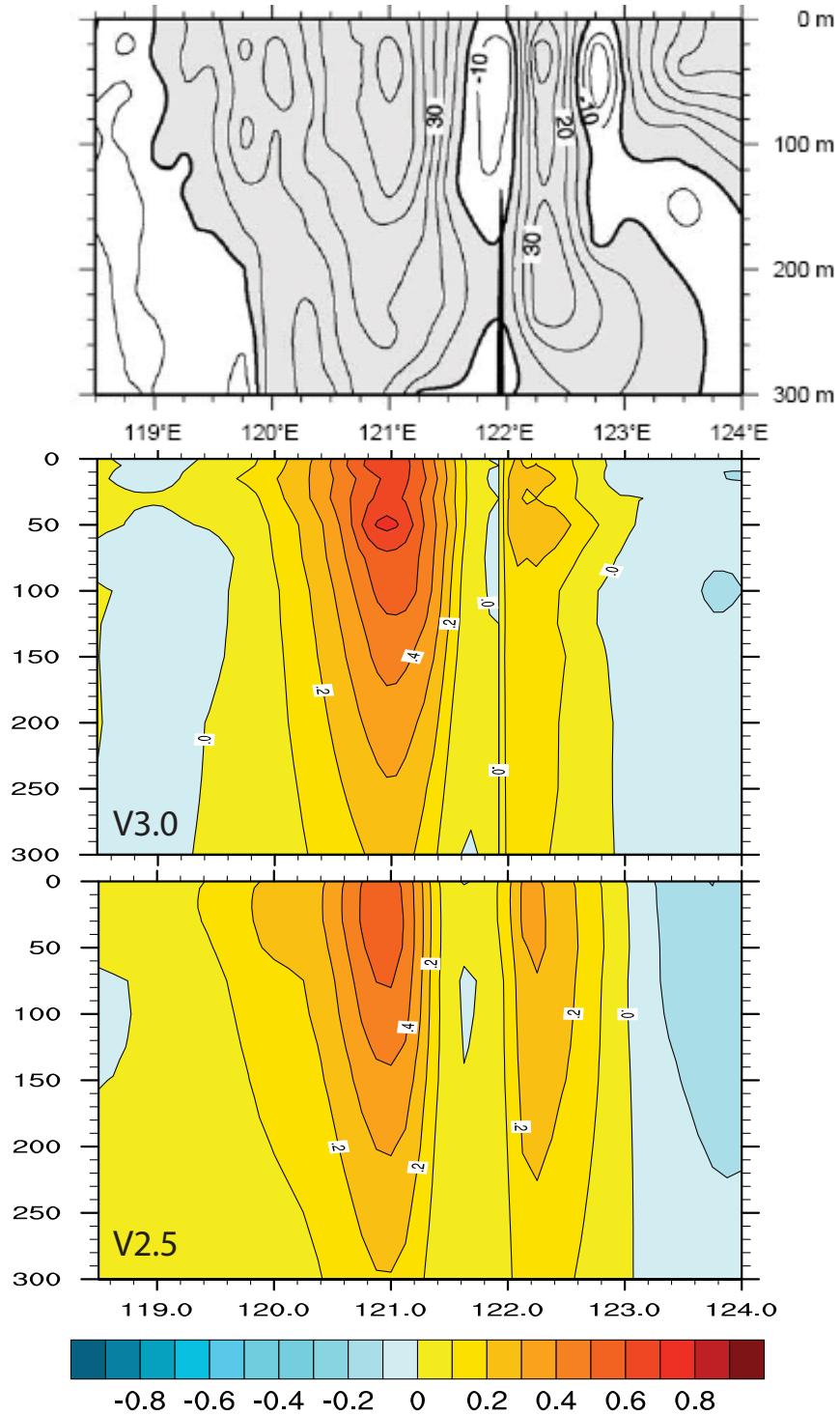
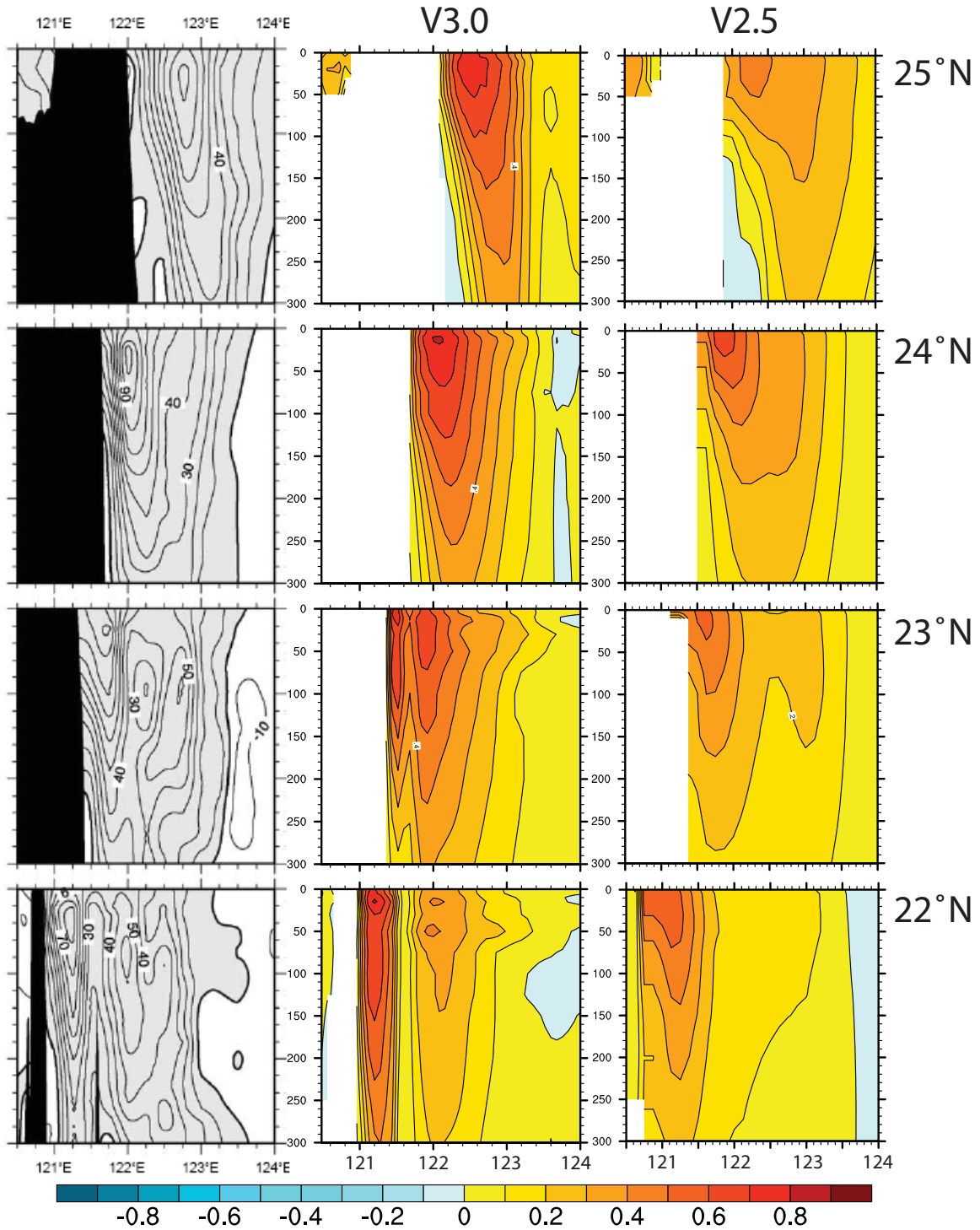
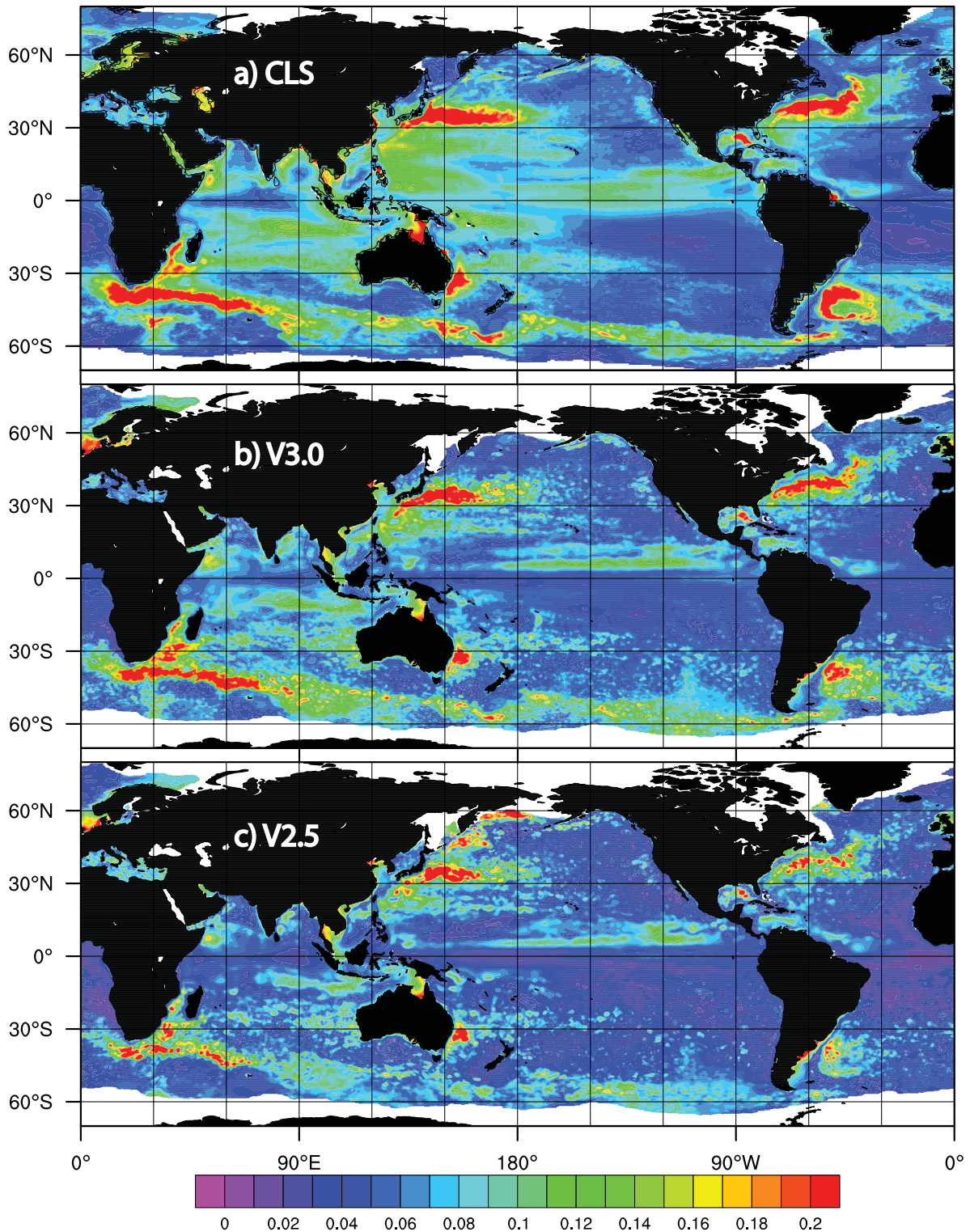


Figure 5: Meridional velocity (m/s) in the top 300 m of the water column in Luzon Strait at 21°N from the observations of Liang et al. (2003) (top), V3.0 (middle) and V2.5 (bottom).



**Figure 6: Meridional velocity (m/s) in the top 300 m of the water column east of Taiwan at 22°N, 23°N, 24°N and 25°N (bottom through top rows, respectively) from the observations of Liang et al. (2003) (left column), V3.0 (middle column) and V2.5 (right column).**



**Figure 7:** SSH variability (m) from a) Collecte Localisation Satellites (CLS) spanning October 1992 through May 2007 and based on available altimeter data from several satellites, b) GOPS V3.0 and c) GOPS V2.5. The simulated results span the hindcast time period 1 June 2007 – 31 May 2008. The CLS data has its own ice mask which differs slightly from the mask used on the simulated results, which is the same as in Figure 2. The non-ice areal average of the variability is 7.8/7.3/6.0 cm respectively, for the three panels.

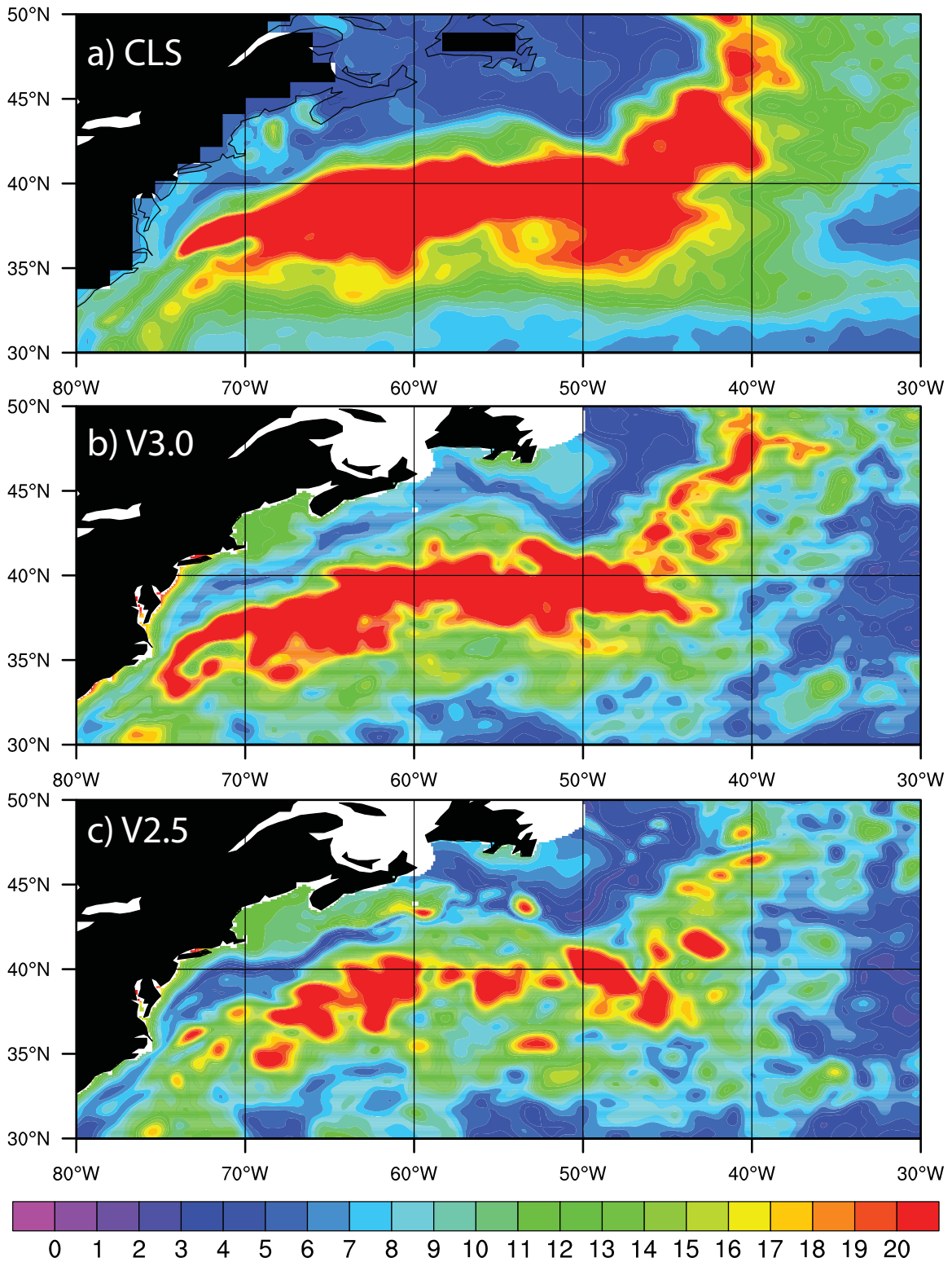


Figure 8: As in Figure 7 except for the Gulf Stream region.

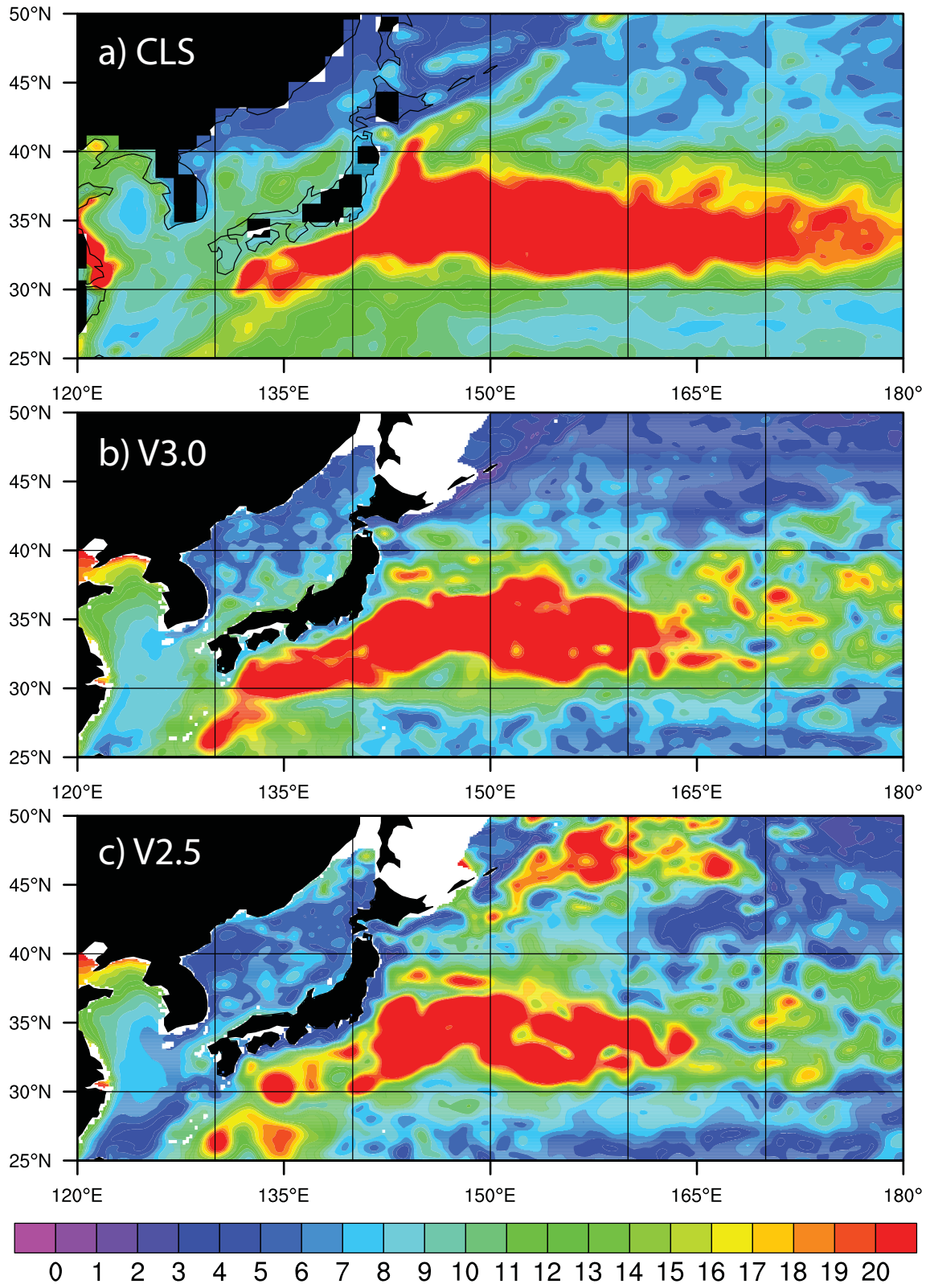


Figure 9: As in Figure 7 except for the Kuroshio region.

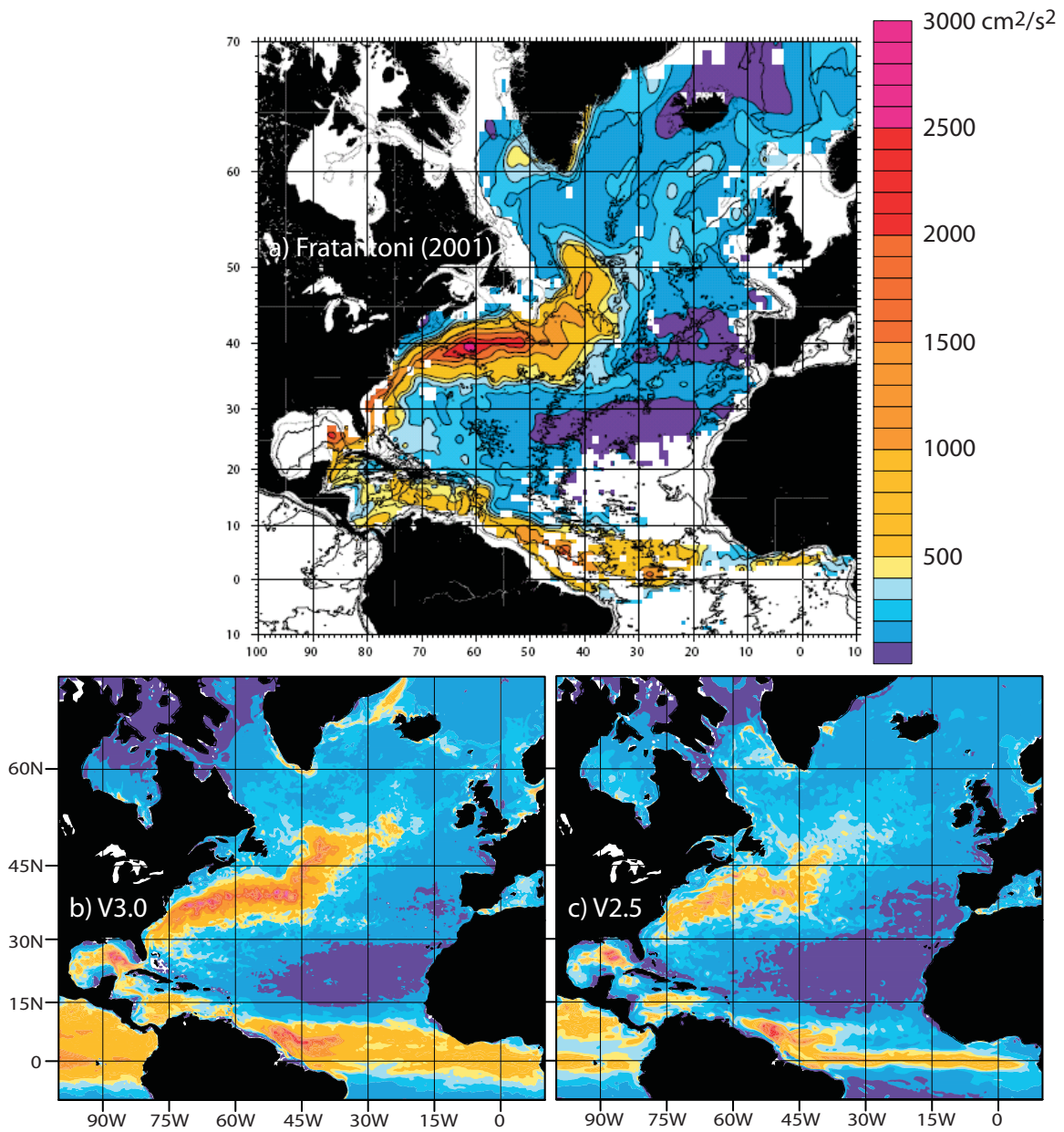


Figure 10: North Atlantic surface eddy kinetic energy (EKE) ( $\text{cm}^2/\text{s}^2$ ) from a) Fratantoni (2001), b) GOPS V3.0 and c) GOPS V2.5. The observations are an average over the time frame 1990-1999 and are based on drifting buoys drogued at 15 m. The simulated EKE fields span the hindcast time period 1 June 2007 – 31 May 2008.

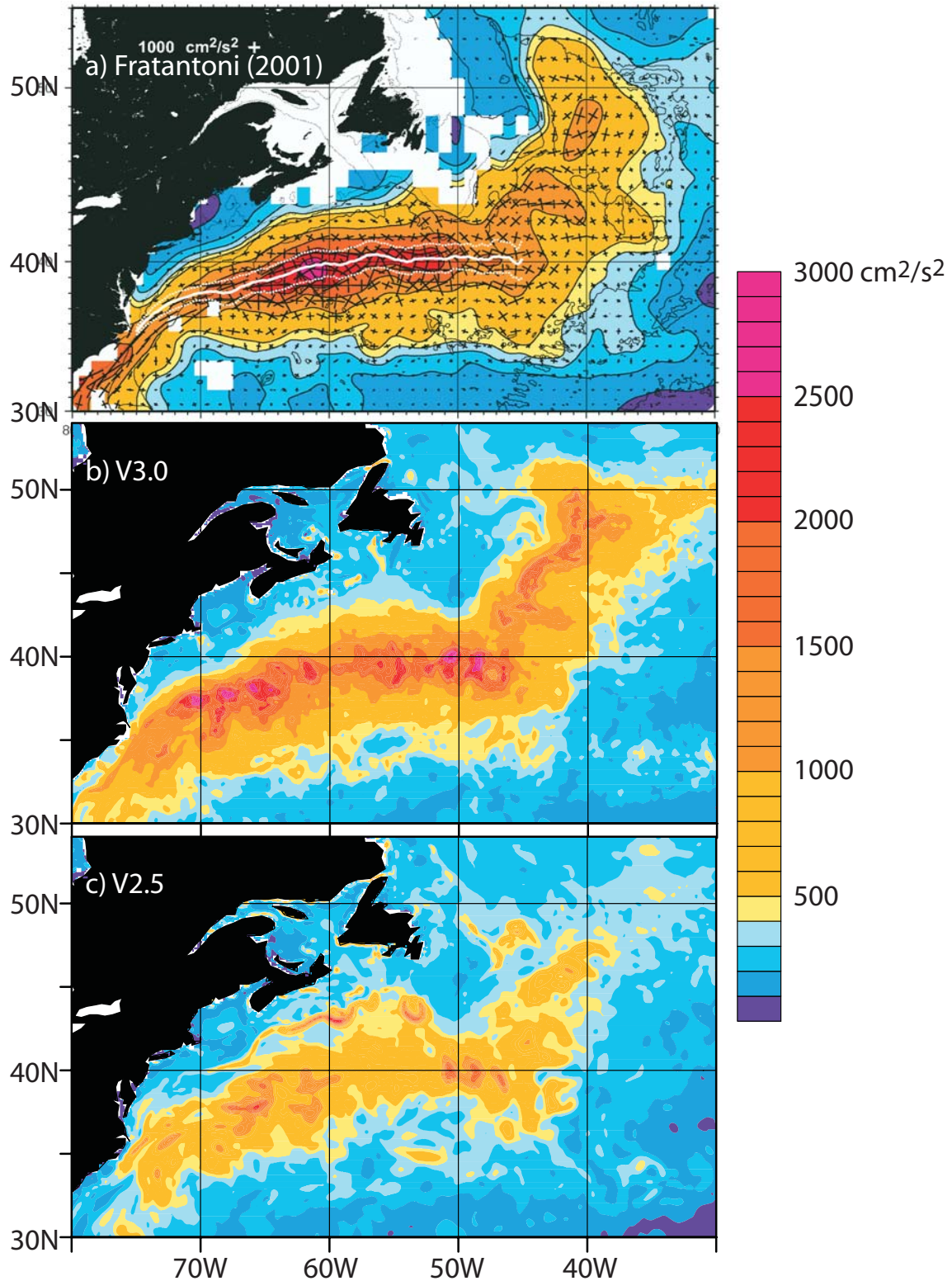


Figure 11: As in Figure 10 but a zoom on the Gulf Stream region. The white lines in panel a) are the mean Gulf Stream pathway  $\pm$  one standard deviation based on satellite SST observations from Lee and Cornillon (1996).

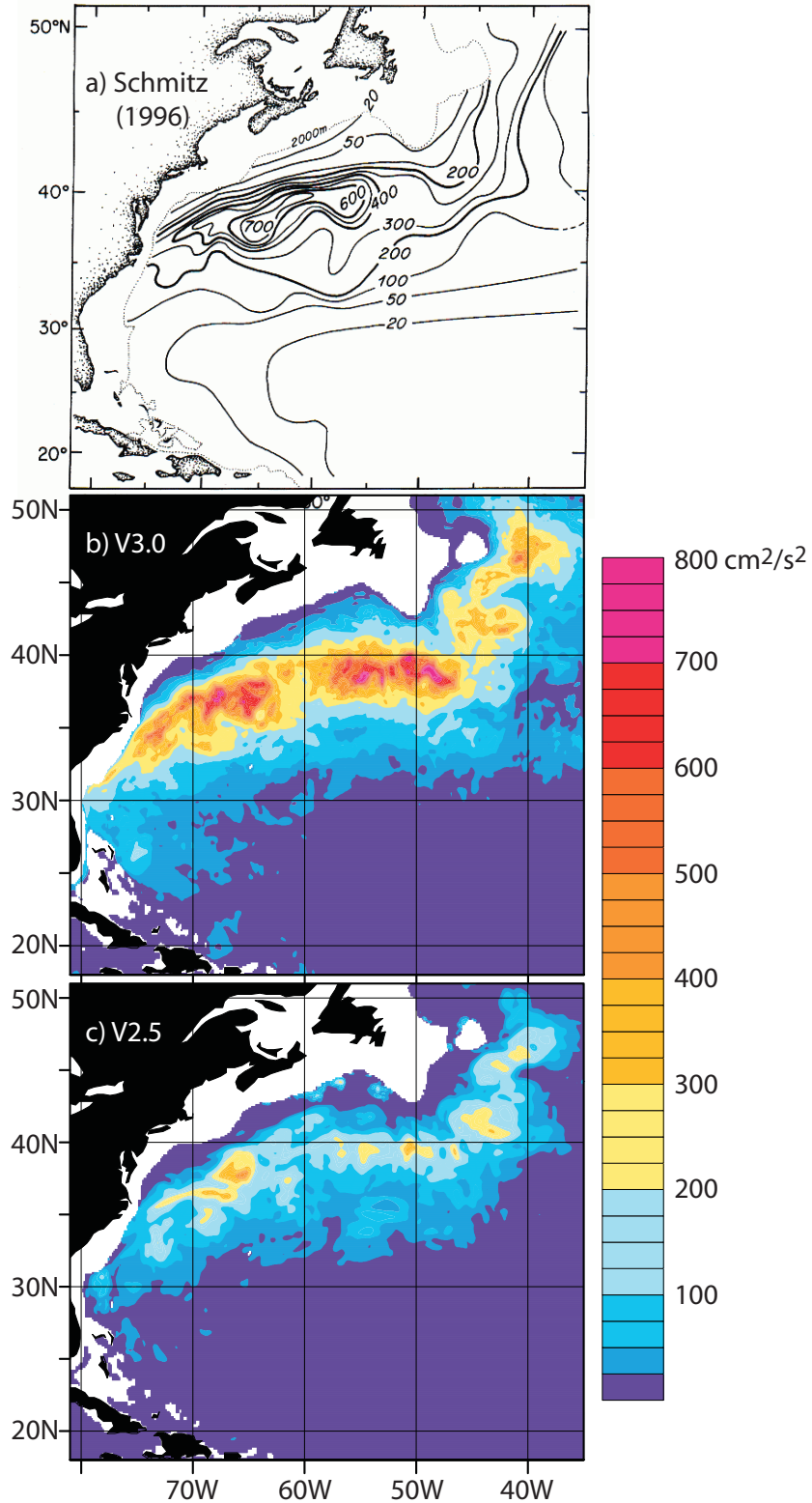


Figure 12: EKE (cm<sup>2</sup>/s<sup>2</sup>) for the Gulf Stream region at 700 m depth from a) Schmitz (1996), b) GOPS V3.0 and c) GOPS V2.5. The observations are compiled from SOFAR float observations. The simulated results span the hindcast time frame, 1 June 2007 – 31 May 2008.

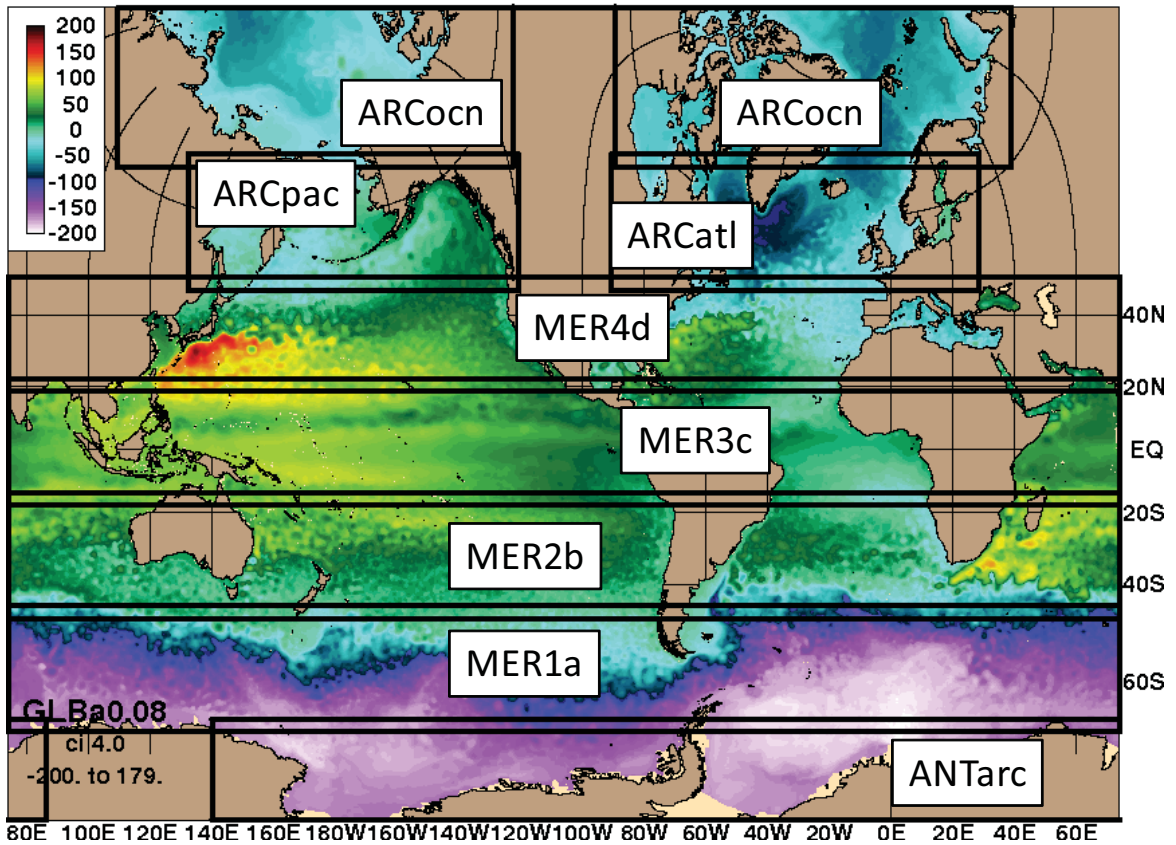
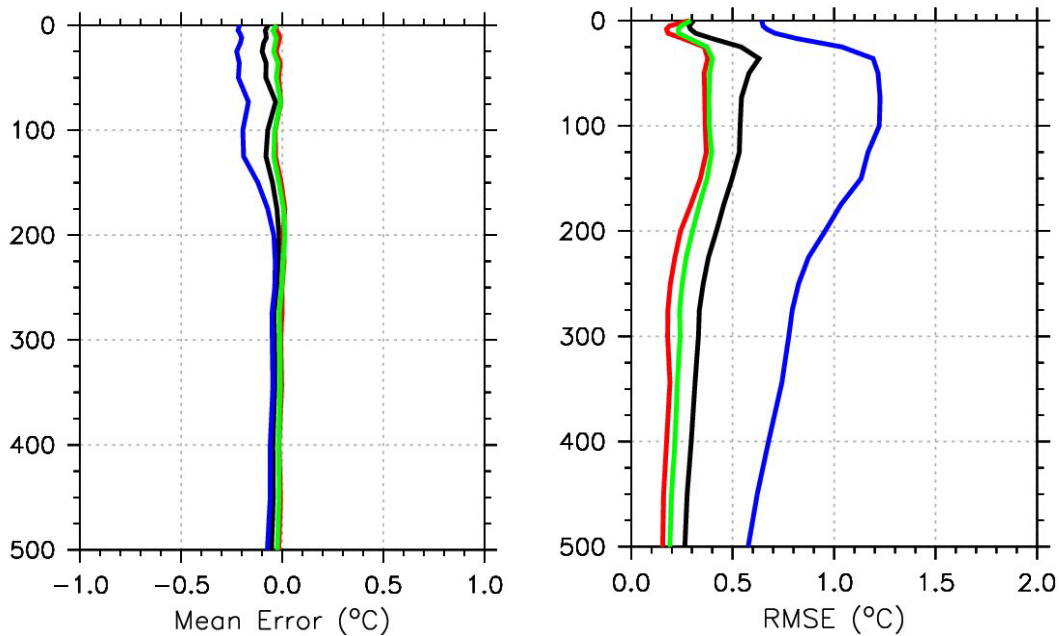
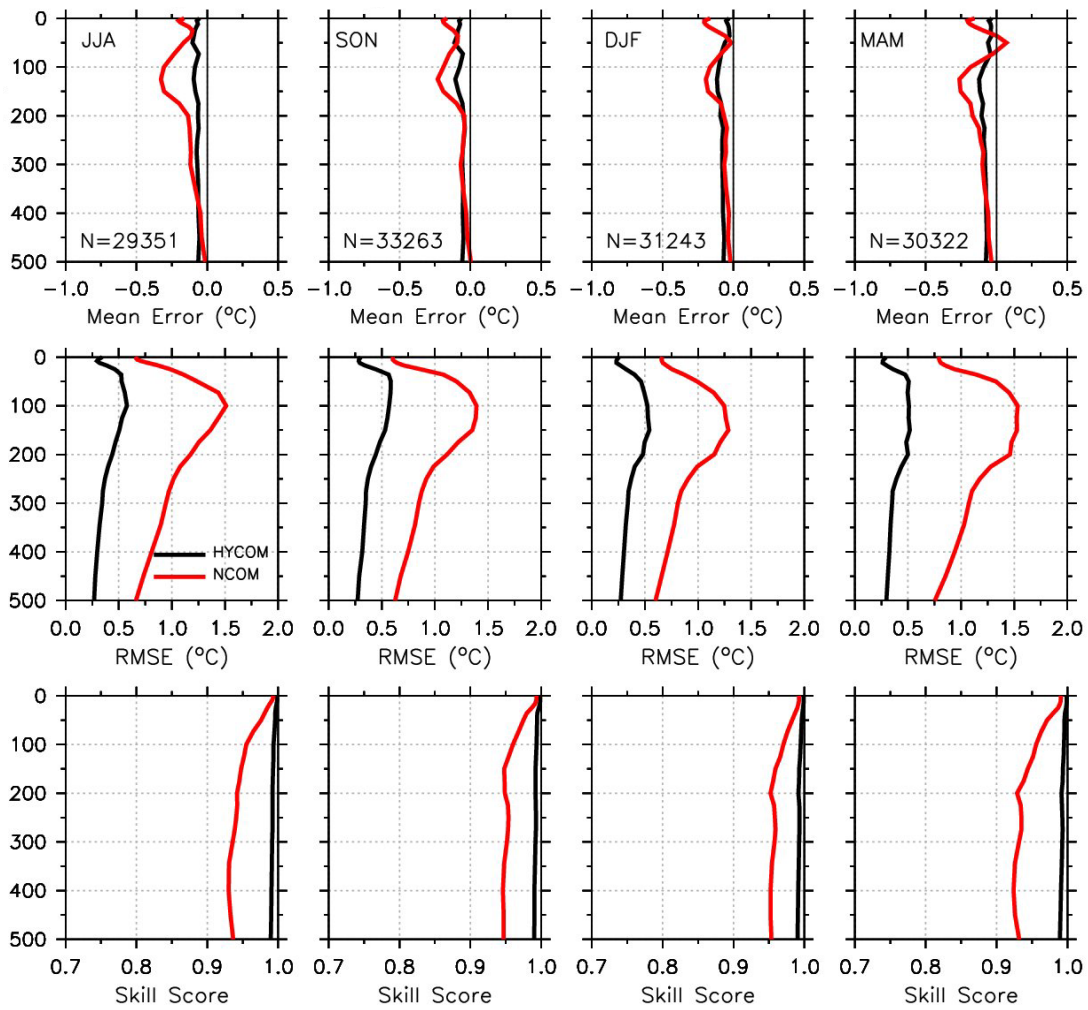


Figure 13: Region names used in the temperature/salinity vs. depth error analysis. Additionally, MERall spans the four MERxx regions show above. The Western Pacific Ocean (WPAC) subregion is defined by 120-170°E, 20-50°N and an Arabian Sea (ASEA) subregion spans 45-80°E, 0-24°N.



**Figure 14: Temperature (°C) vs. depth error analysis in the upper 500 m of the near-global MERall region for September 2007 using 10719 assimilated profiles. The left panel is mean error and the right panel is RMSE. The red curve is the error between the observations and the NCODA analysis in z-space, the green curve between the observations and the NCODA analysis in hybrid space, the black curve between the observations and the V3.0 analysis at 00Z and the blue curve between the observations and the 24-hour V3.0 forecast.**



**Figure 15: Temperature (°C) vs. depth error analysis in the upper 500 m against *assimilated* profiles of the near-global MERall region for the four seasons – first column = summer (JJA), second column = fall (SON), etc. The top, middle and bottom rows are mean error, RMSE and skill score, respectively. The black curves represent V3.0 and the red curves represent V2.5. The number of assimilated profiles used in each season is indicated by N = xxxxx in the top row. This is the number of near-surface observations used and this value decreases with depth since not all profiles exist down to 500 m depth.**

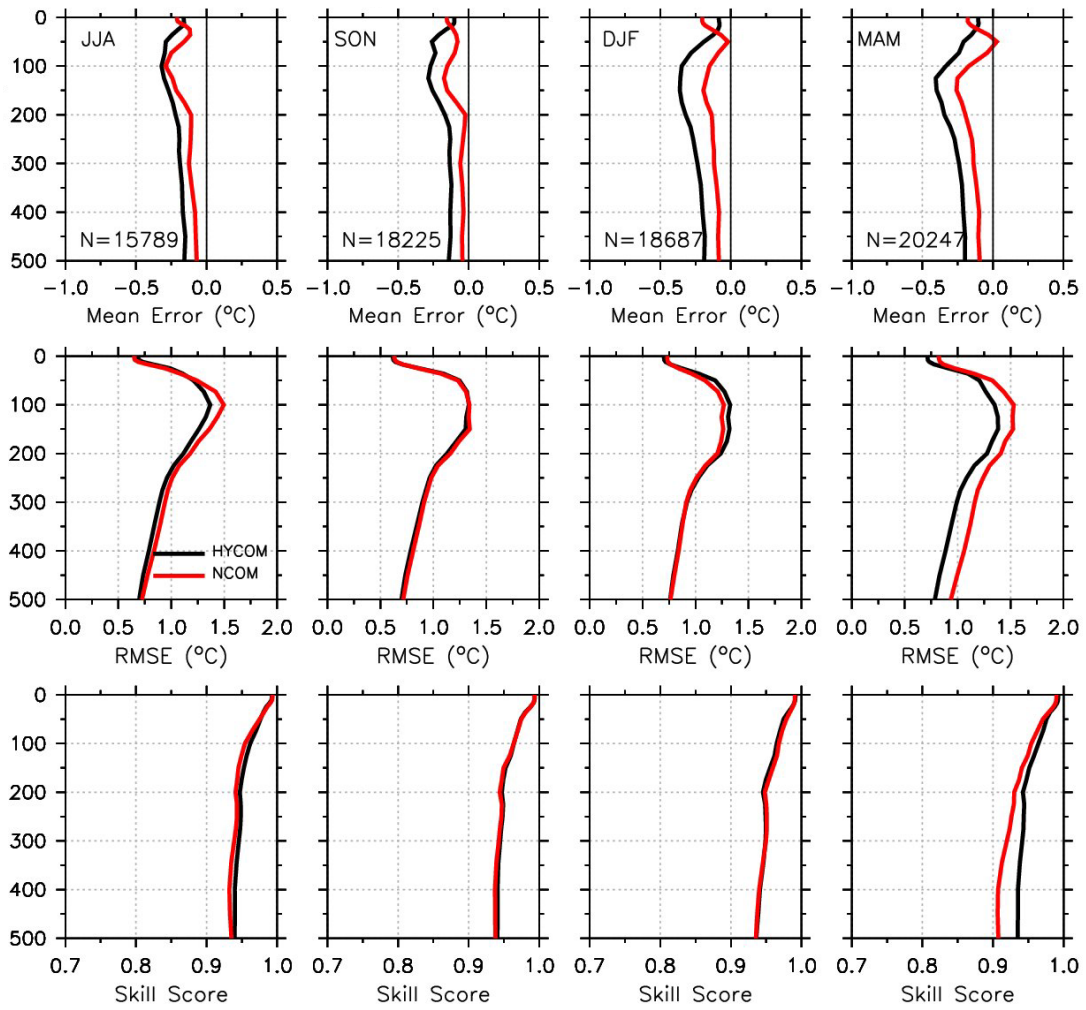
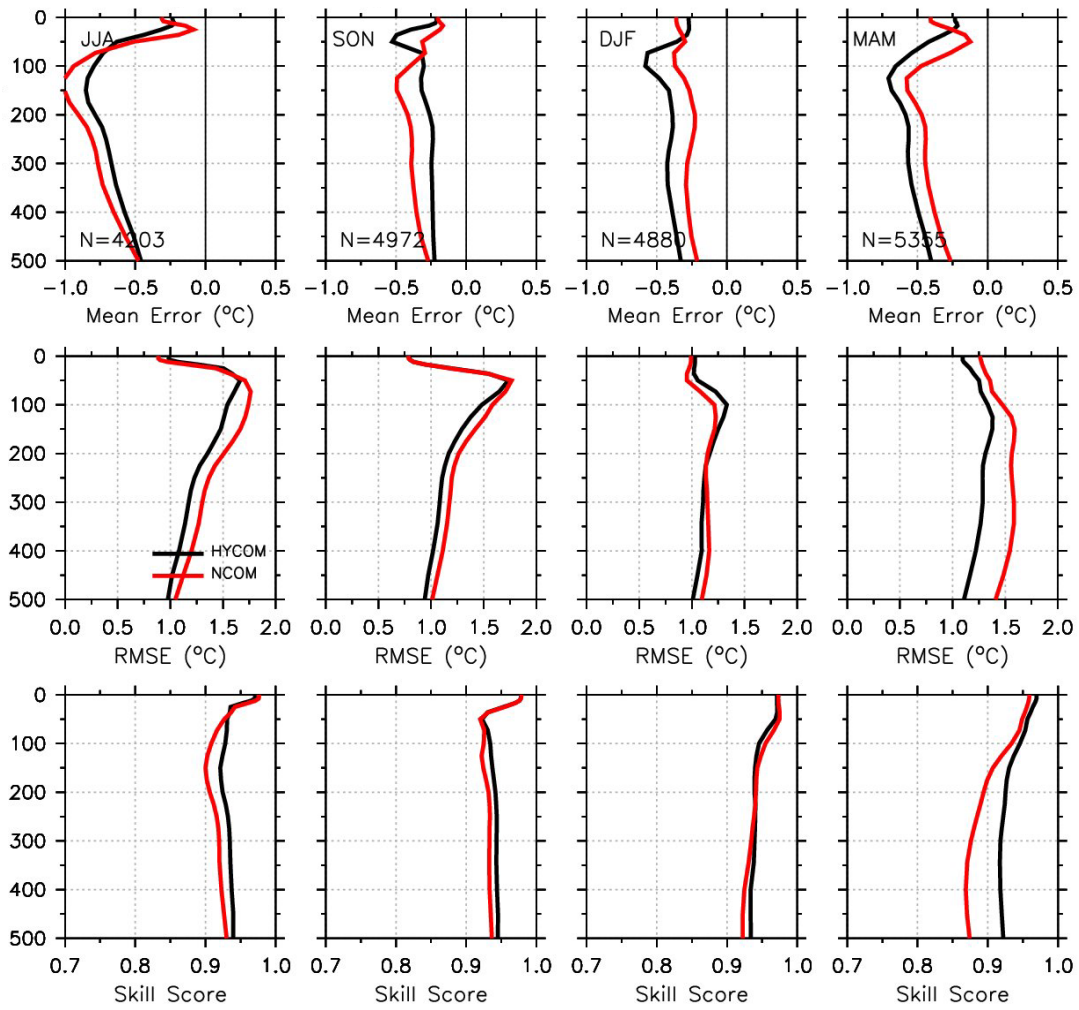
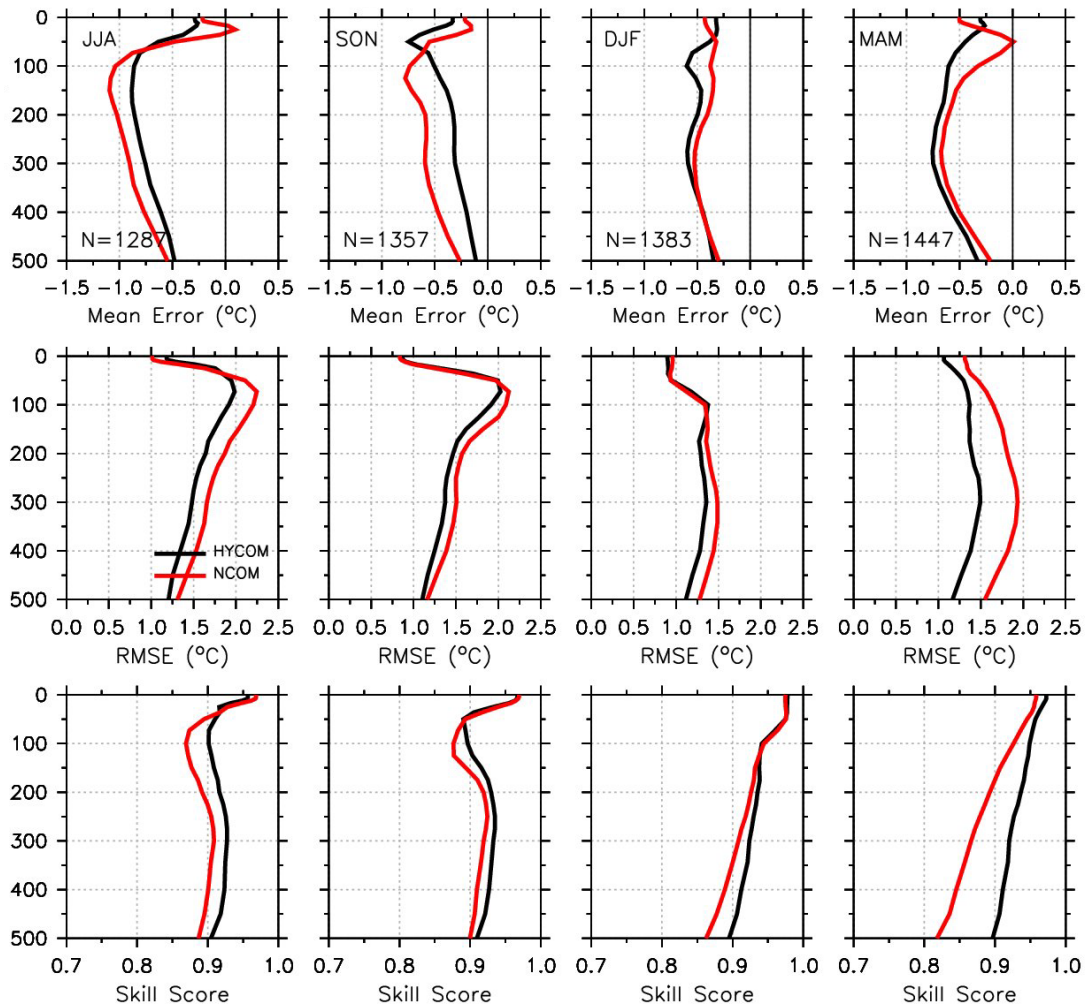


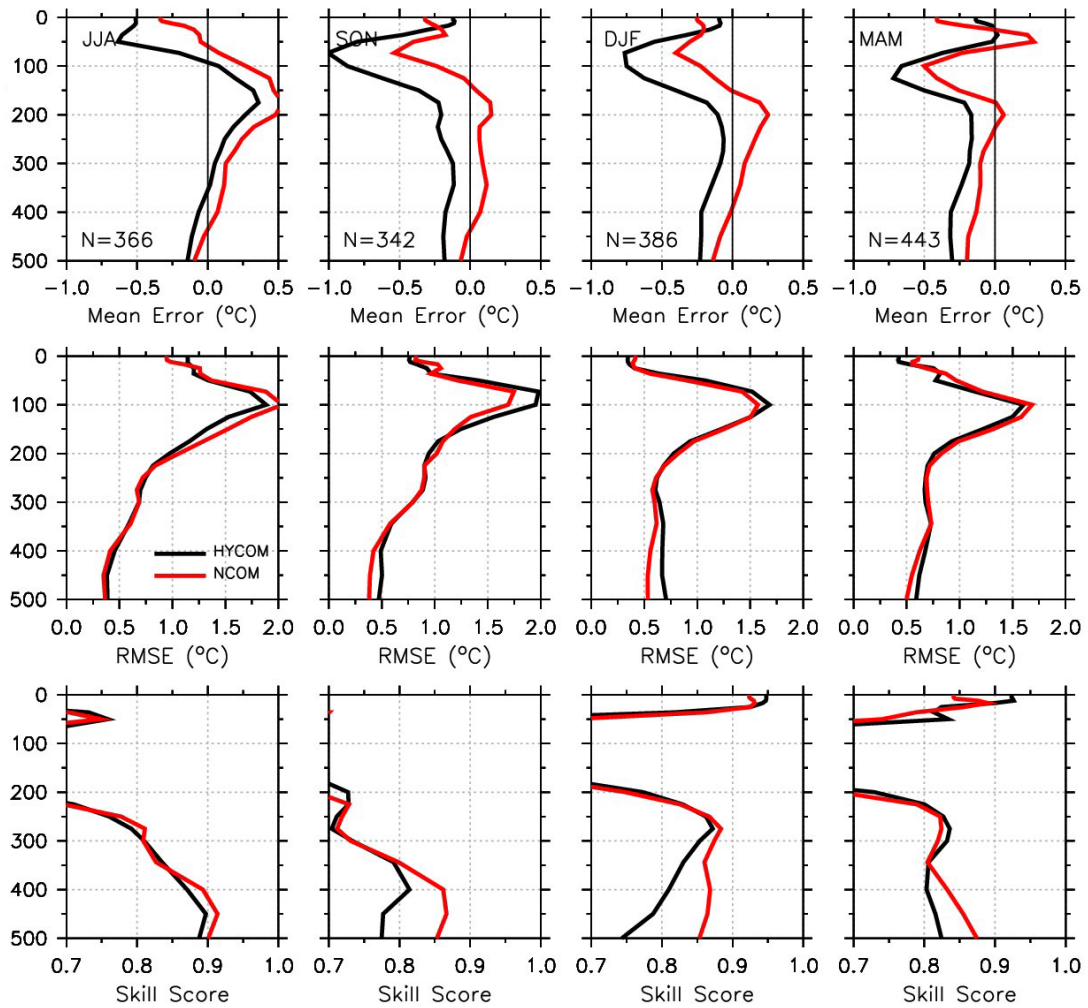
Figure 16: As in Figure 15 except for *unassimilated* profiles.



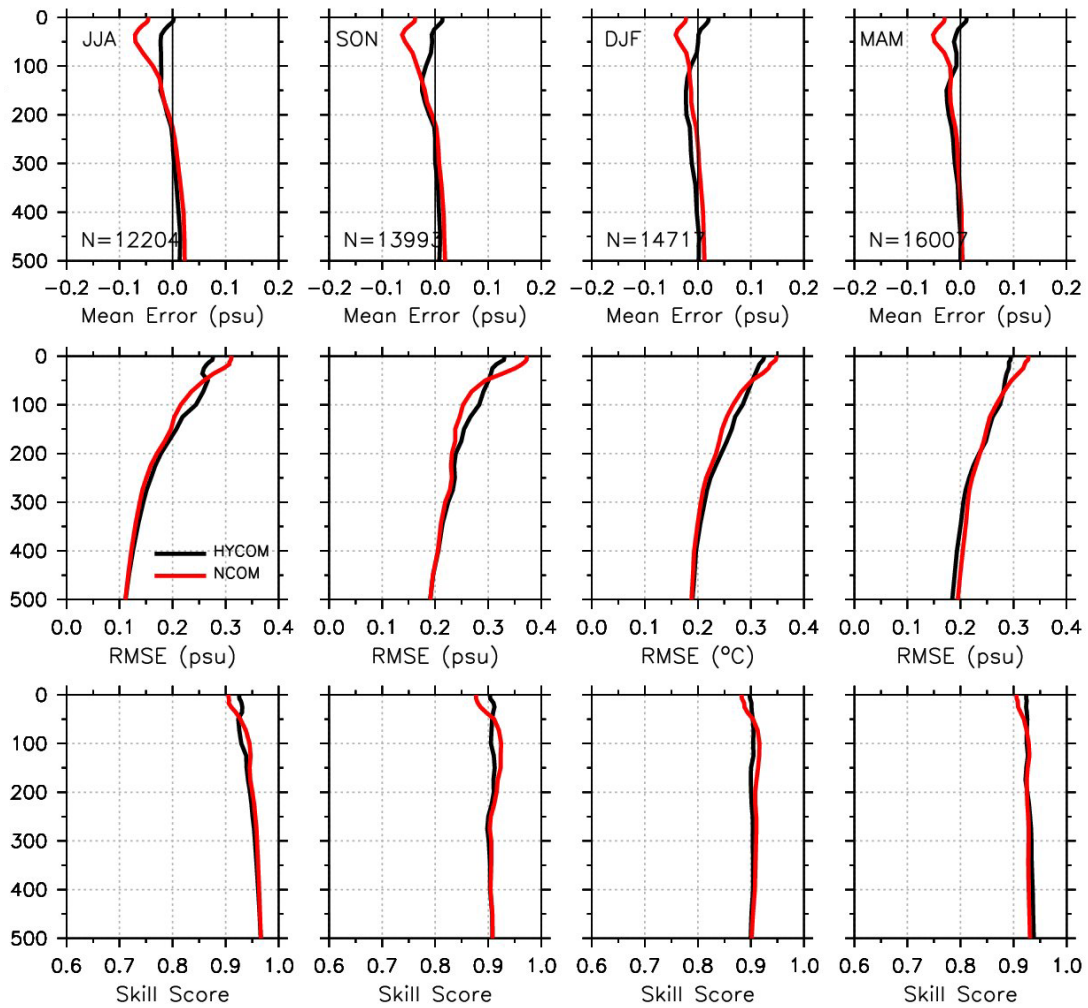
**Figure 17: As in Figure 16 except for region MER4d.**



**Figure 18:** As in Figure 16 except for the western Pacific region (120-170°E, 20-50°N). Note the scales on the x-axes of the ME and RMSE panels are different than those on Figure 16.



**Figure 19: As in Figure 16 except for the Arabian Sea region (45-80°E, 0-24°N).**



**Figure 20: Salinity (psu) vs. depth error analysis in the upper 500 m against *unassimilated* profiles of the near-global MERall region for the four seasons – first column = summer (JJA), second column = fall (SON), etc. The top, middle and bottom rows are mean error, RMSE and skill score, respectively. The black curves represent V3.0 and the red curves represent V2.5. The number of assimilated profiles used in each season is indicated by N = xxxxx in the top row. This is the number of near-surface observations used and this value decreases with depth since not all profiles exist down to 500 m depth.**

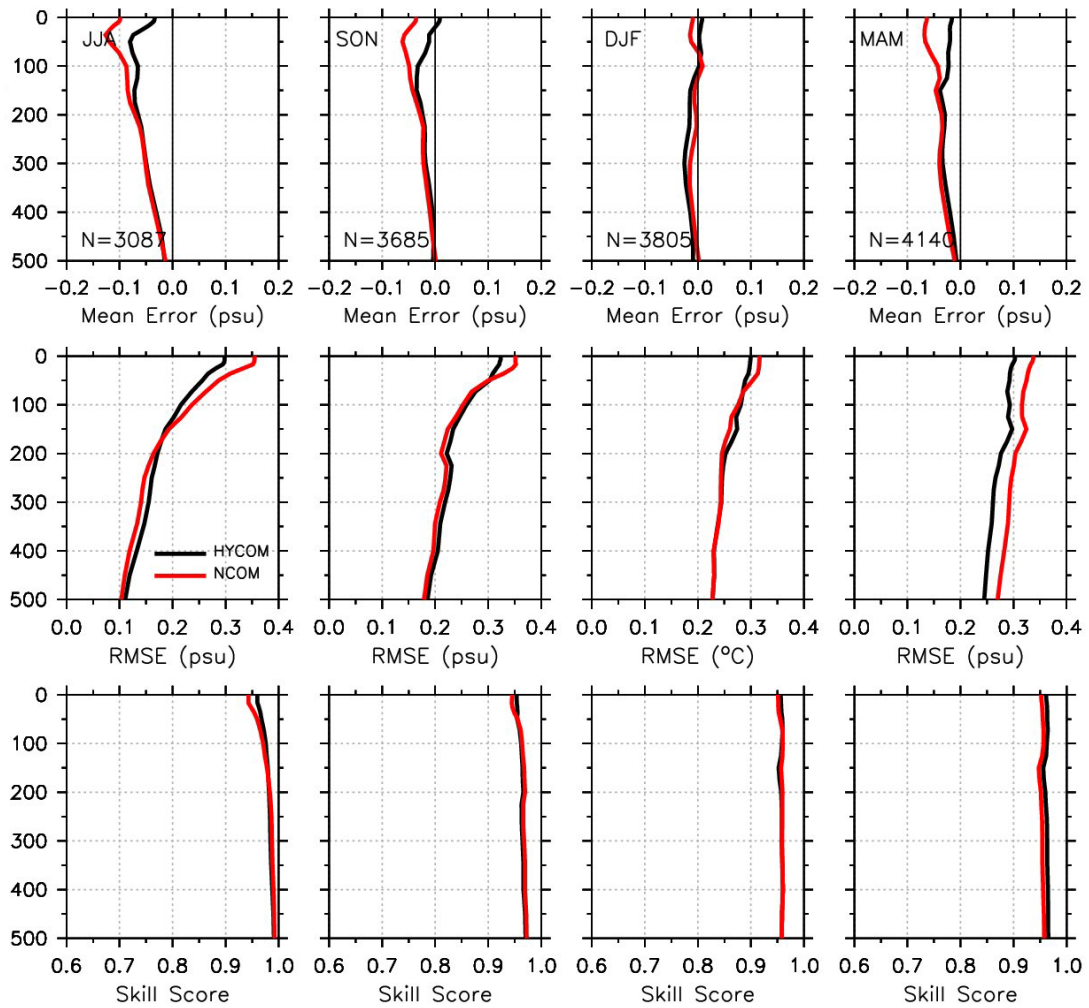
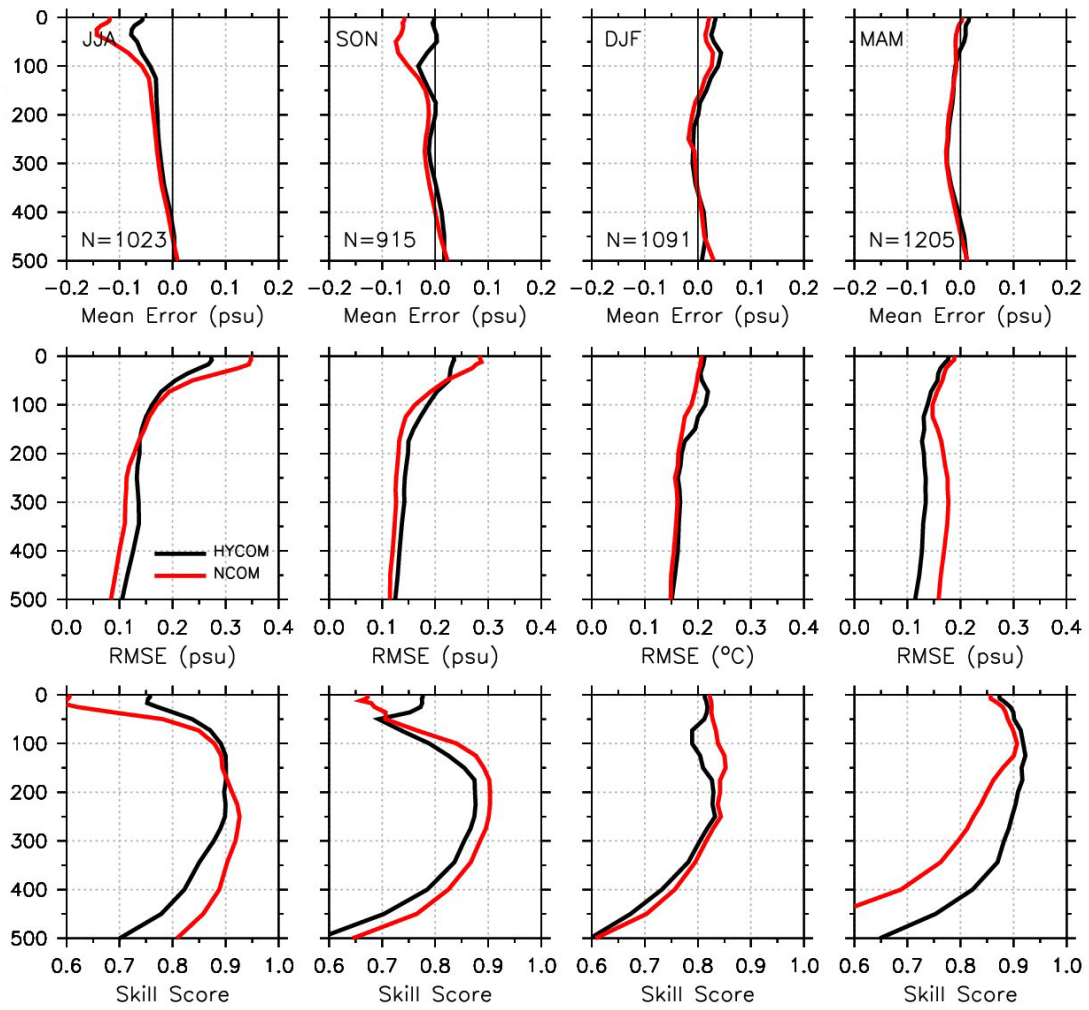
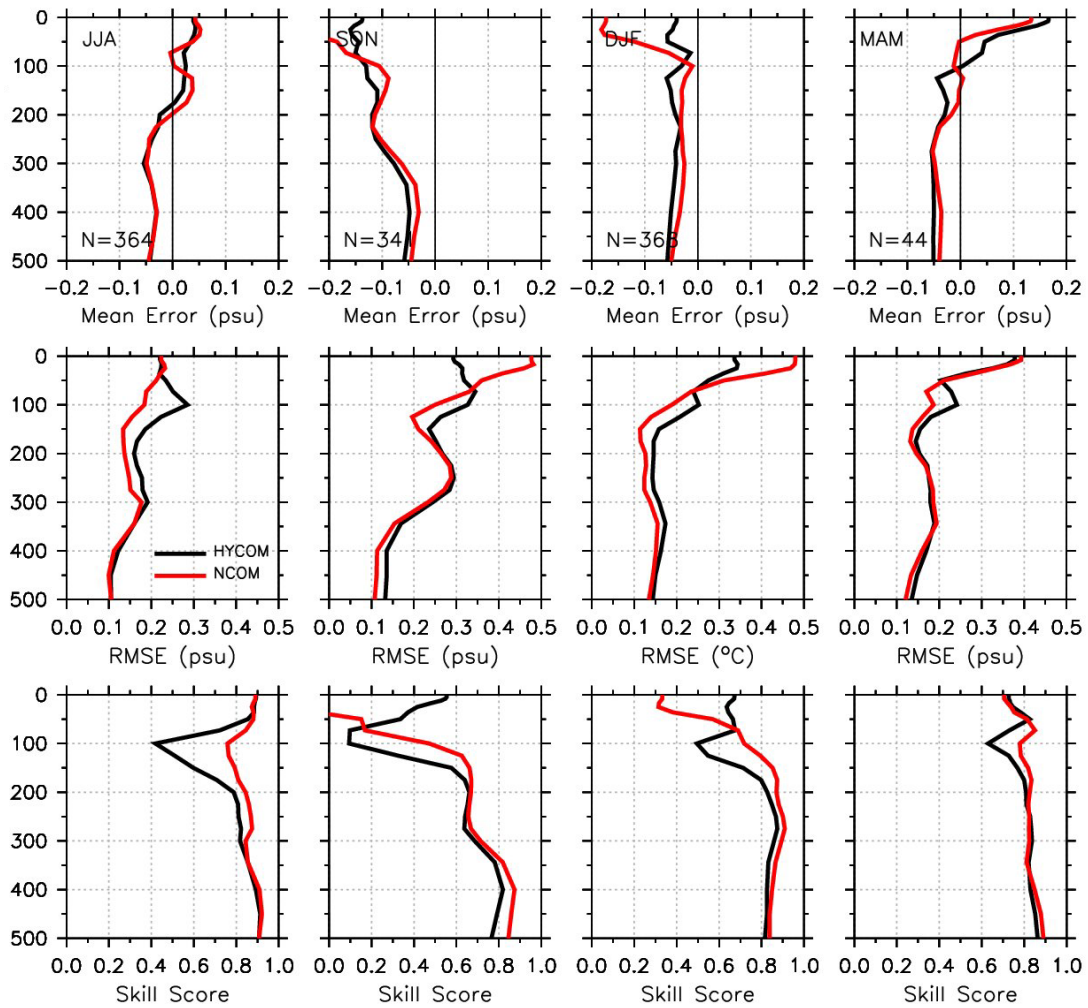


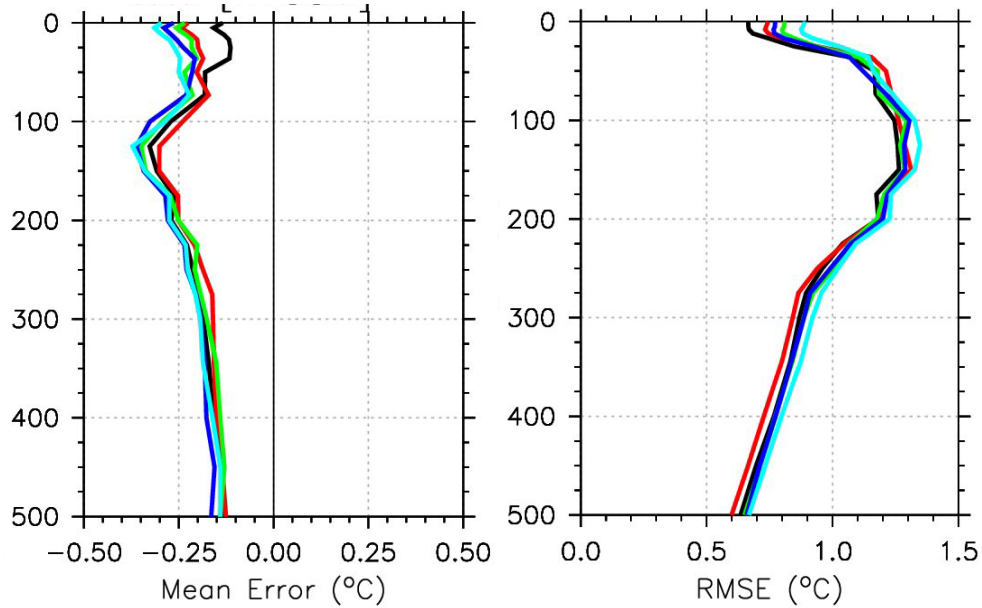
Figure 21: As in Figure 20 except for the MER4d region.



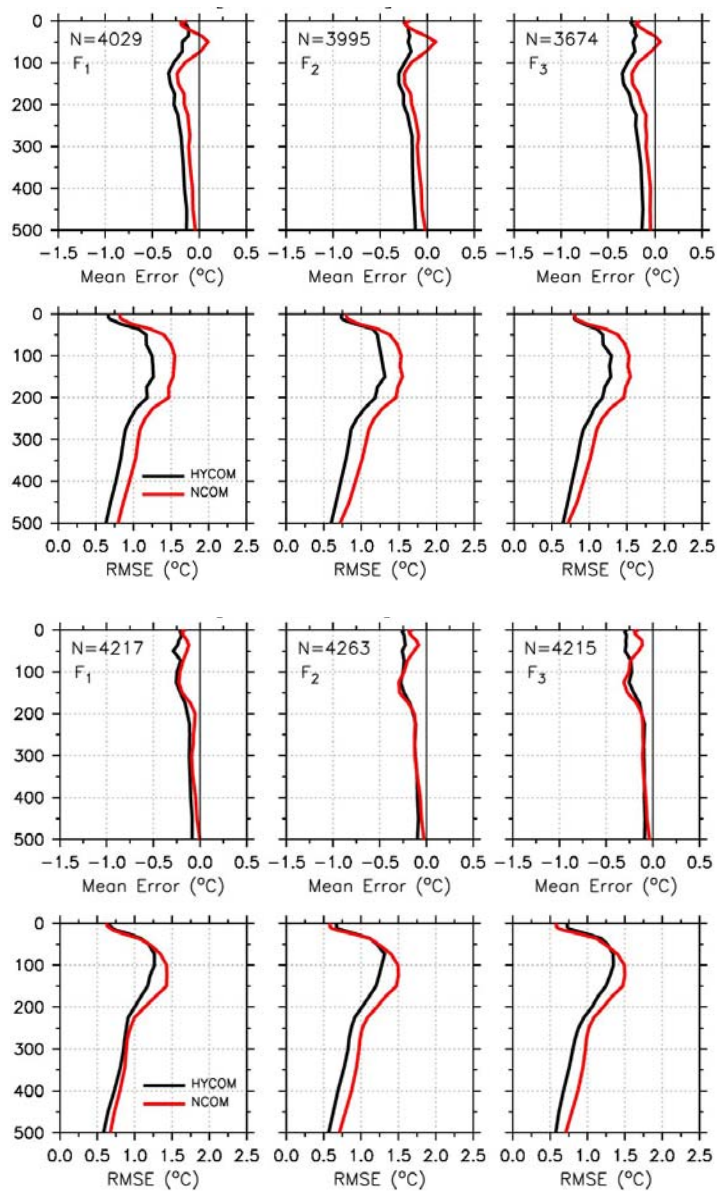
**Figure 22: As in Figure 20 except for the western Pacific Ocean region (120-170°E, 20-50°N).**



**Figure 23:** As in Figure 20 except for the Arabian Sea region (45-80°E, 0-24°N). Note that the range of the x-axes for RMSE and SS is different than that for Figure 20.



**Figure 24: Temperature (°C) vs. depth error analysis in the upper 500 m of the near-global MERall region for Spring (MAM) 2008 using 3874 unassimilated profiles. The left panel is mean error and the right panel is RMSE. The lines represent the error associated with an n-day forecast from the initial state of the hindcast: black = 1-day forecast, red = 2-day forecast, green = 3-day forecast, blue = 4-day forecast and cyan = 5-day forecast.**



**Figure 25: Temperature (°C) vs. depth error analysis in the upper 500 m of the near-global MERall region for Spring (MAM) 2008 (top two rows) and Fall (SON) 2007 (bottom two rows). The first and third rows are mean error and the second and fourth rows are RMSE. The left column represents a 1-day forecast ( $F_1$ ), the middle column a 2-day forecast ( $F_2$ ) and the right column a 3-day forecast ( $F_3$ ). The black curves are V3.0 and the red curves are V2.5. The number of unassimilated profiles used for each forecast is shown as  $N = \text{xxxx}$  in the mean error panels. This is the number of near-surface observations used and this value decreases with depth since not all profiles exist down to 500 m depth.**

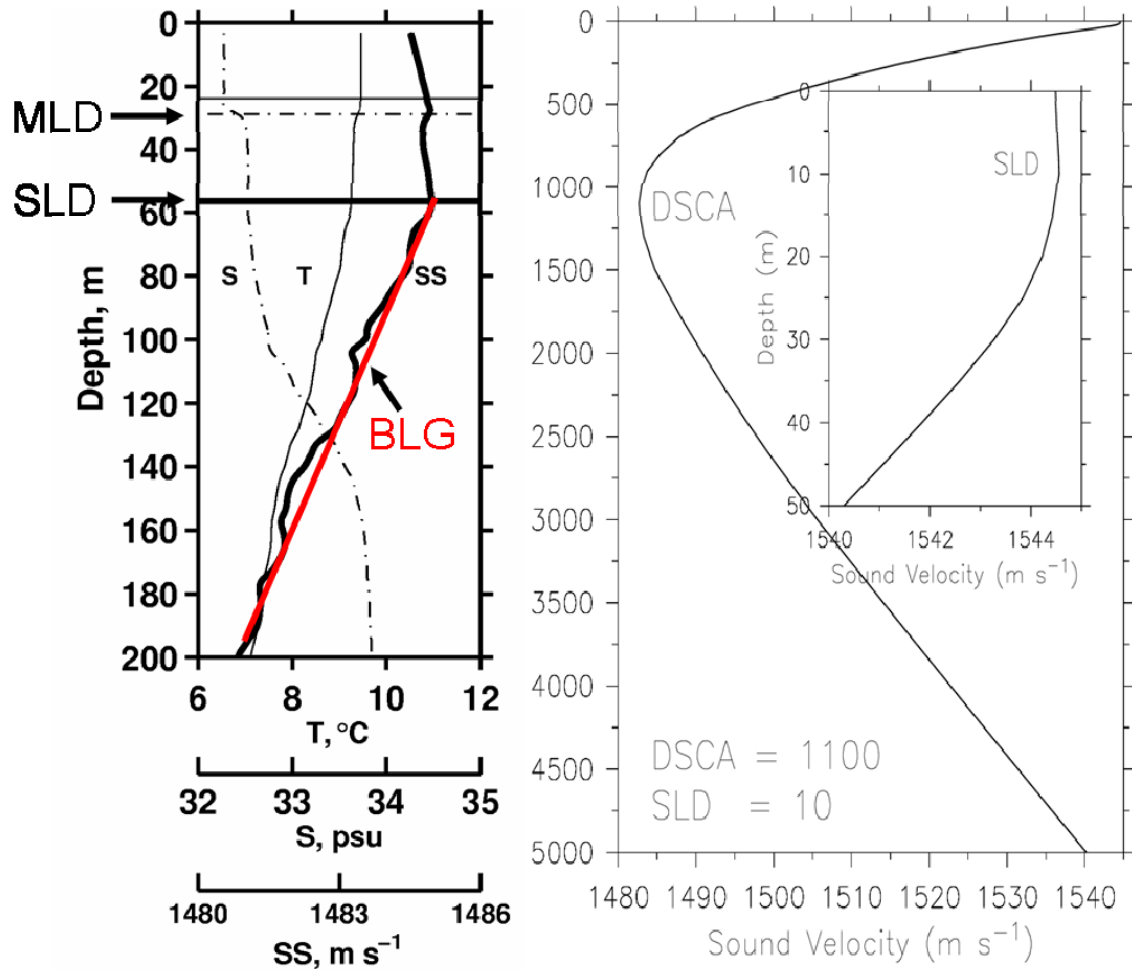


Figure 26: Typical temperature, salinity and sound speed (SS) profiles with representative mixed layer depth (MLD), sonic layer depth (SLD), below layer gradient (BLG) and deep sound channel (DSC) axis overlaid. See the text for a more detailed description of each variable.

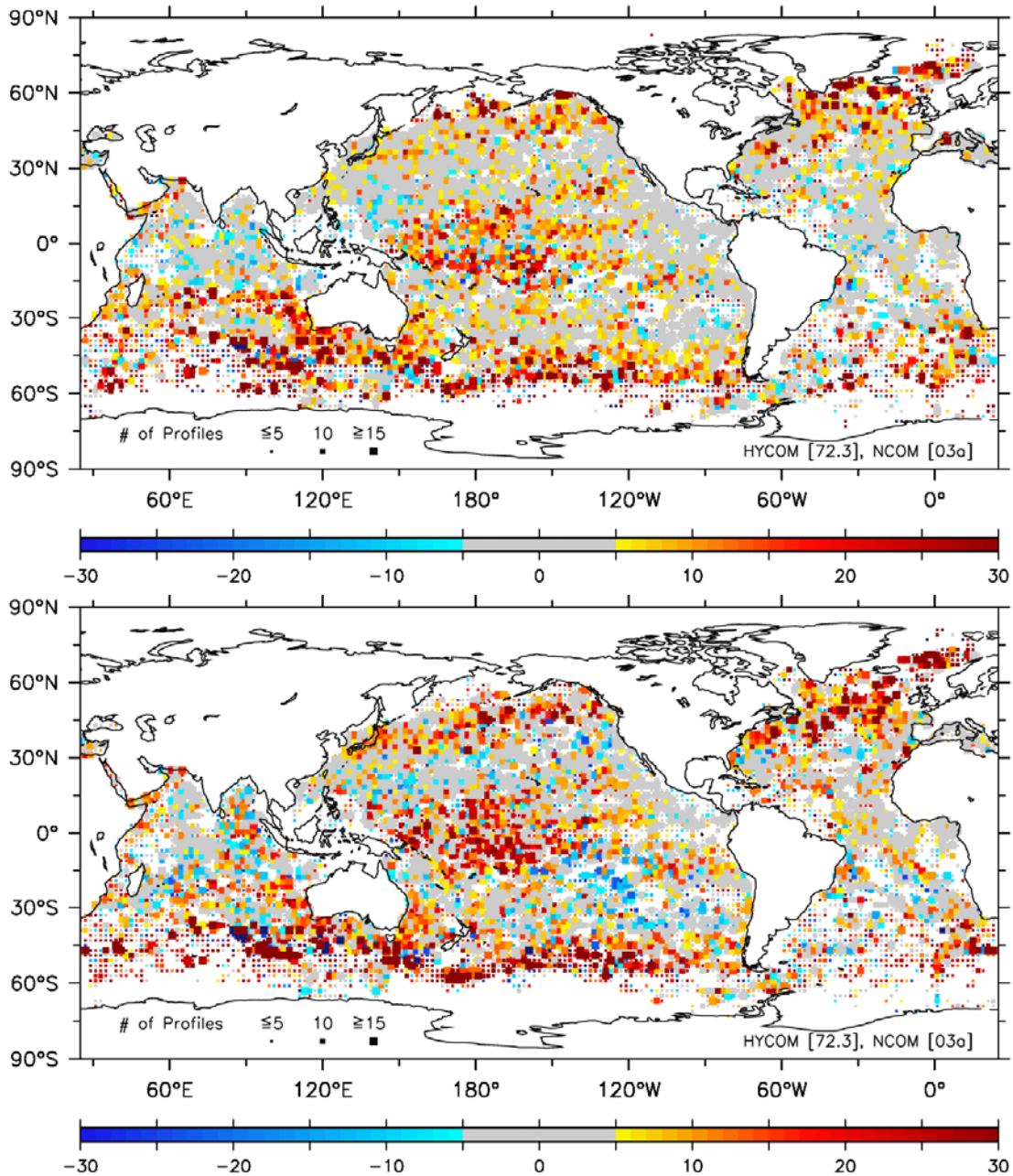
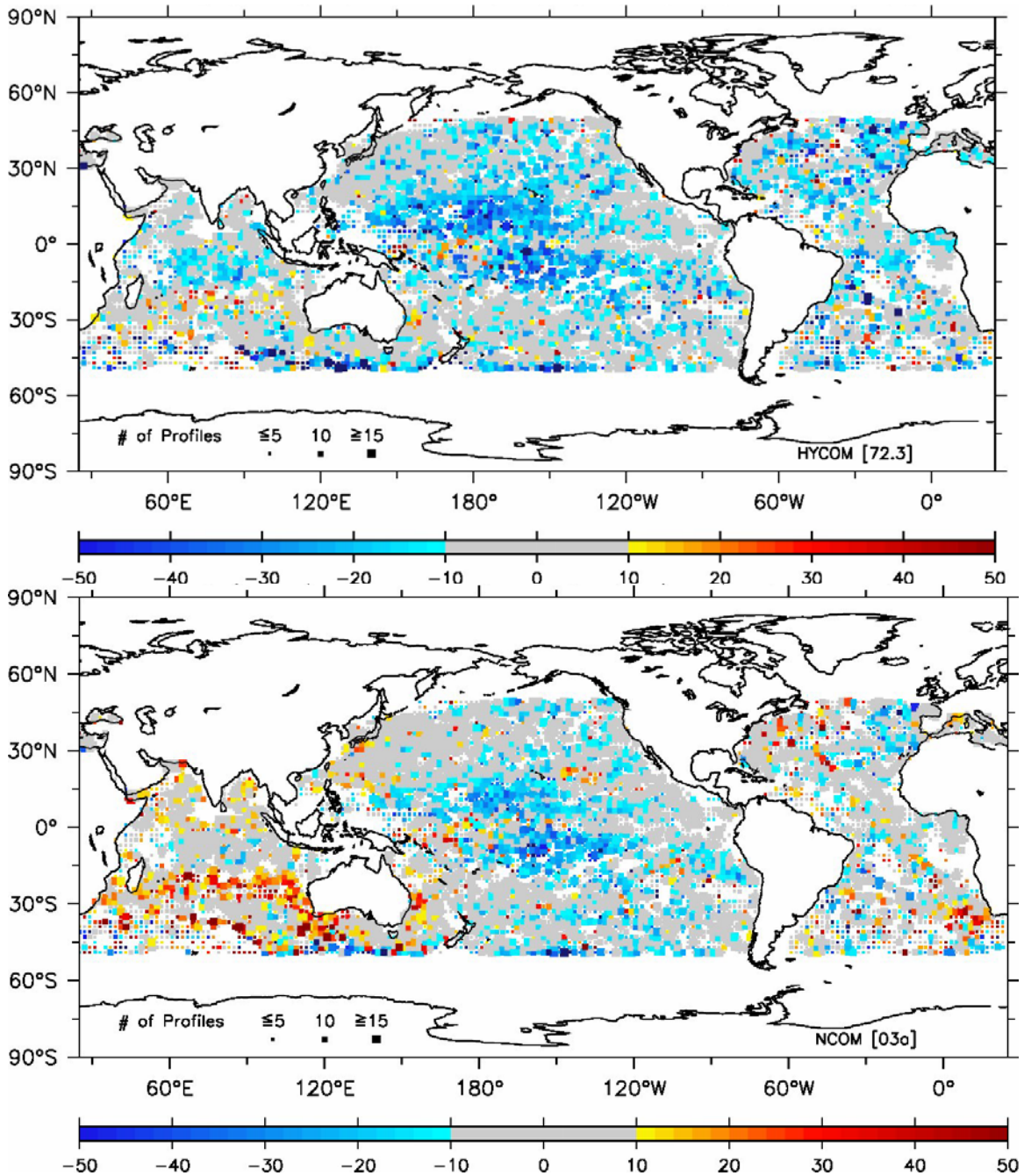
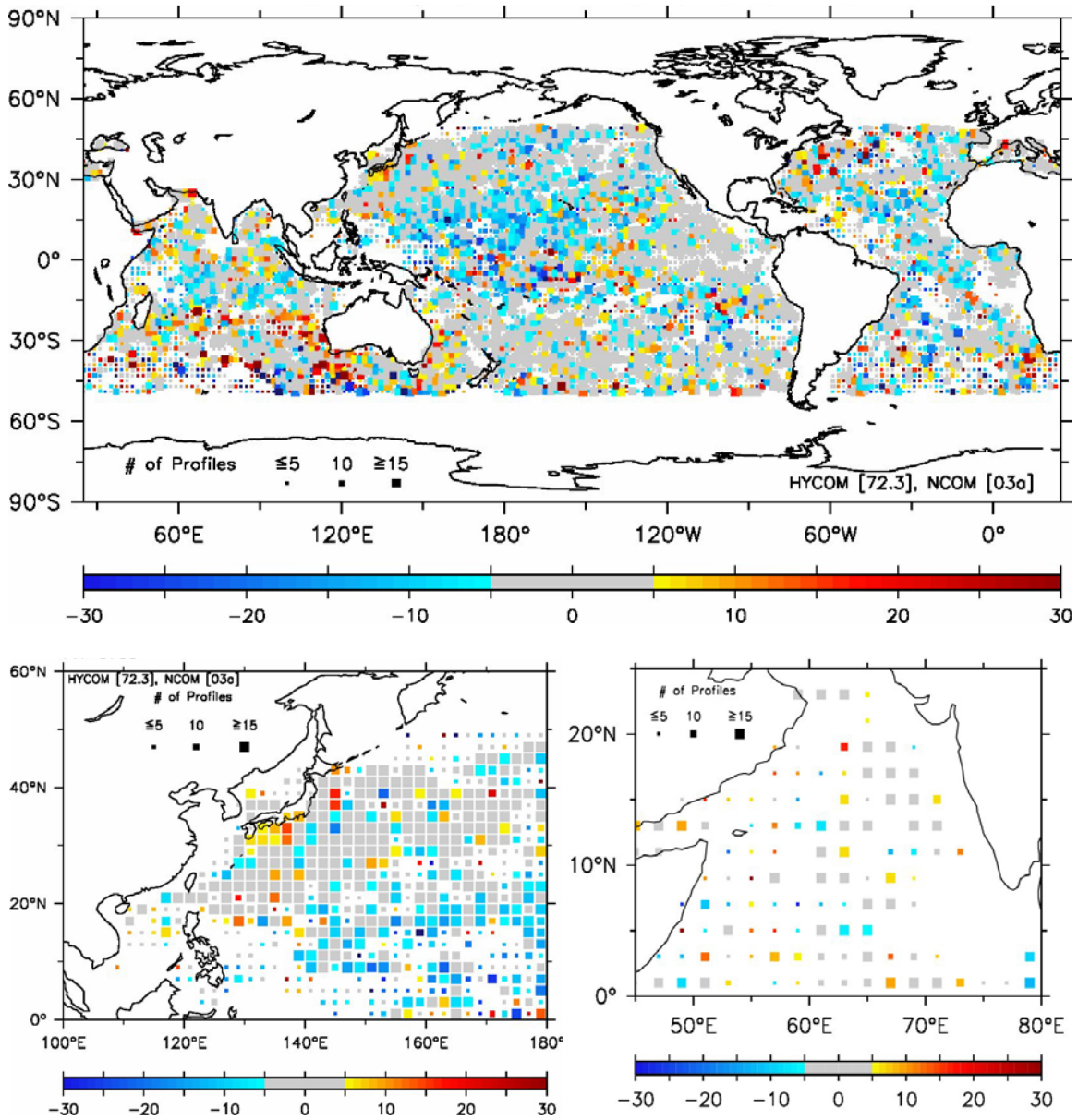


Figure 27: Relative Median Absolute Error (MdAE) for MLD (top) and SLD (bottom) for the whole domain using 131182 assimilated profiles. Positive (negative) values indicate V3.0 (V2.5) has lower absolute error. MdAE is less than 5 m in those boxes colored gray. The data are averaged over  $2^\circ$  bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend within Antarctica. The percentage of points where V3.0 has equal or less absolute error than V2.5 is 73% for MLD and 76% for SLD.



**Figure 28: MLD Median Bias Error (MdBE) for V3.0 (top) and V2.5 (bottom) against 66387 unassimilated profiles over the hindcasts spanning 1 June 2007-31 May 2008. For V3.0, the basin average MdBE = -7 m, RMSE = 40 m and 53% of the points are within 10 m of the observation. For V2.5, the basin average MdBE = -3 m, RMSE = 39 m and 57% of the points are within 10 m of the observation. Red values indicate the simulated MLD is deeper than observed while blue values indicate it is shallower. The analysis is limited to  $\pm 50^\circ$  latitude. The data are averaged over  $2^\circ$  bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend within Antarctica.**



**Figure 29: MLD relative Median Absolute Error (MdaE) for the whole domain (top) using 66387 unassimilated profiles, western Pacific Ocean (bottom left) using 9700 unassimilated profiles and the Arabian Sea (bottom right) using 1537 unassimilated profiles. Positive (negative) values indicate V3.0 (V2.5) has lower absolute error. The analysis in the top panel is limited to  $\pm 50^\circ$  latitude. MdaE is less than 5 m in those boxes colored gray. The data are averaged over  $2^\circ$  bins and the number of profiles within each bin is indicated by the size of each individual square as denoted by the legend within Antarctica. The percentage of points where V3.0 has equal or less absolute error than V2.5 is 44% for the whole domain, 37% in the western Pacific and 59% in the Arabian Sea.**

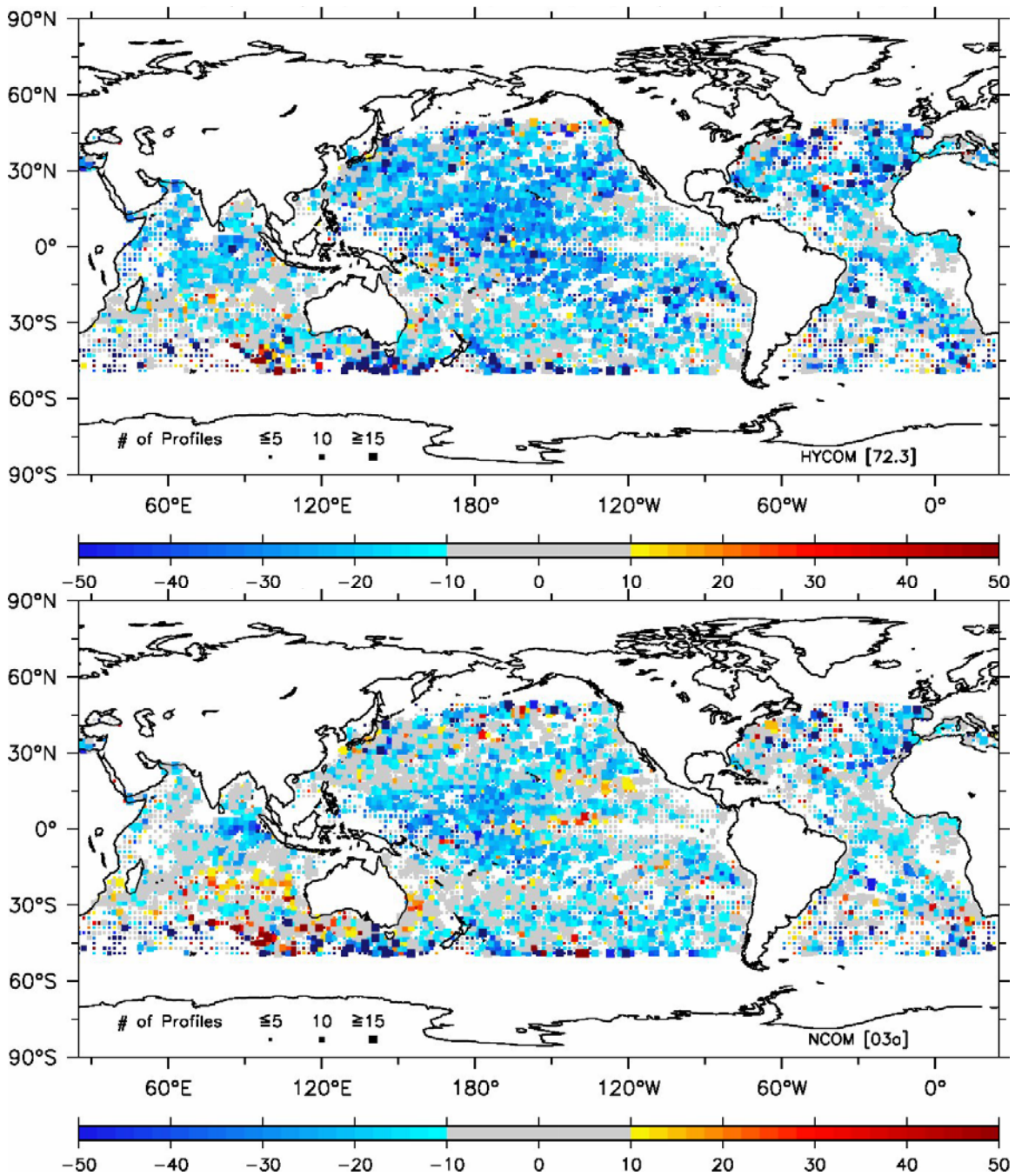


Figure 30: As in Figure 28 except for SLD and the number of unassimilated profiles is 50681. For V3.0, the basin average MdBE = -16 m, RMSE = 67 m and 25% of the points are within 10 m of the observation. For V2.5, the basin average MdBE = -10 m, RMSE = 60 m and 34% of the points are within 10 m of the observation.

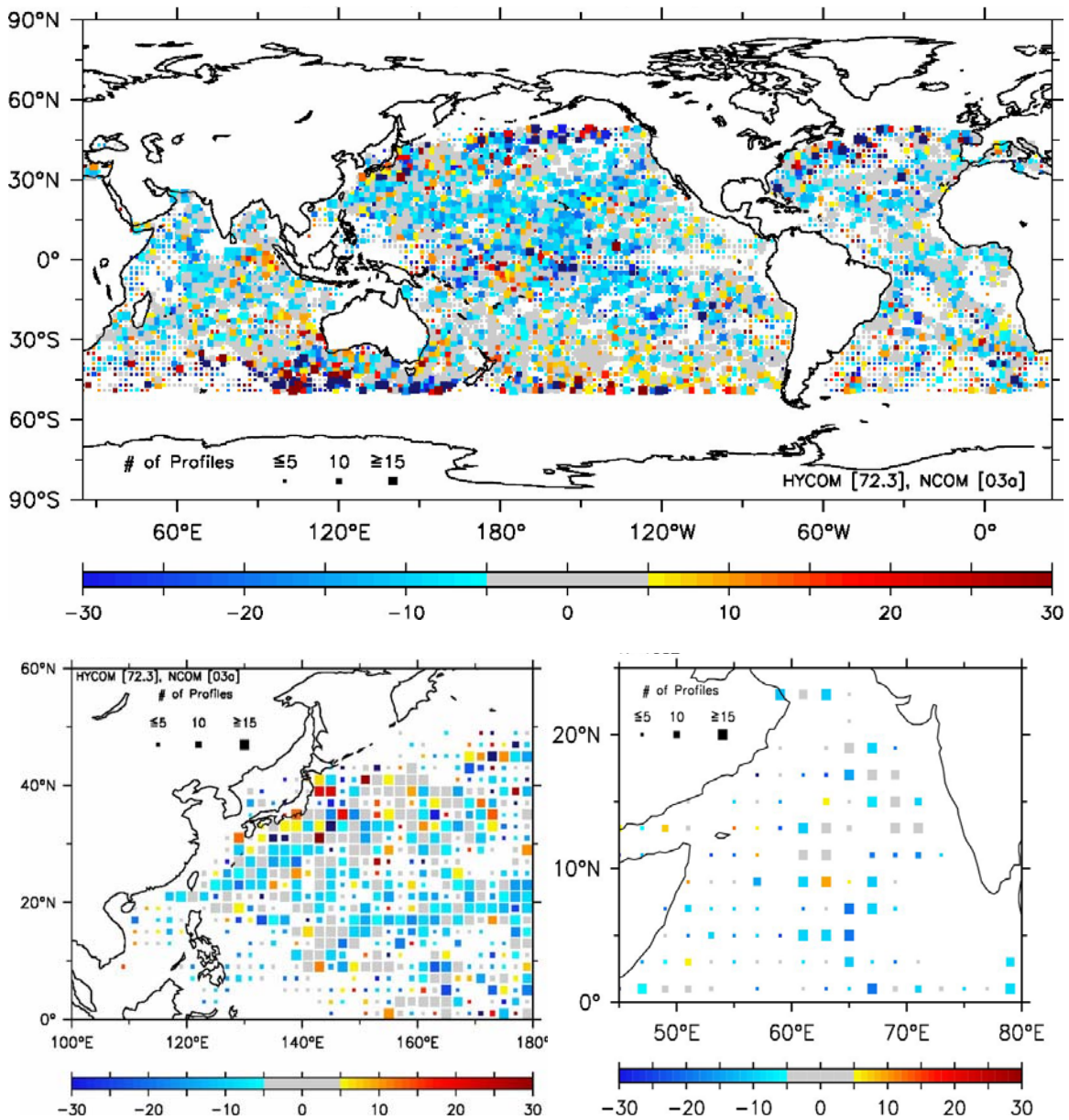


Figure 31: As in Figure 29 except for SLD using (top) 50681 unassimilated profiles, (bottom left) 6792 unassimilated profiles and (bottom right) 1082 unassimilated profiles. The percentage of points where V3.0 has equal or less absolute error than V2.5 is 42% for the whole domain, 34% in the western Pacific and 30% in the Arabian Sea.

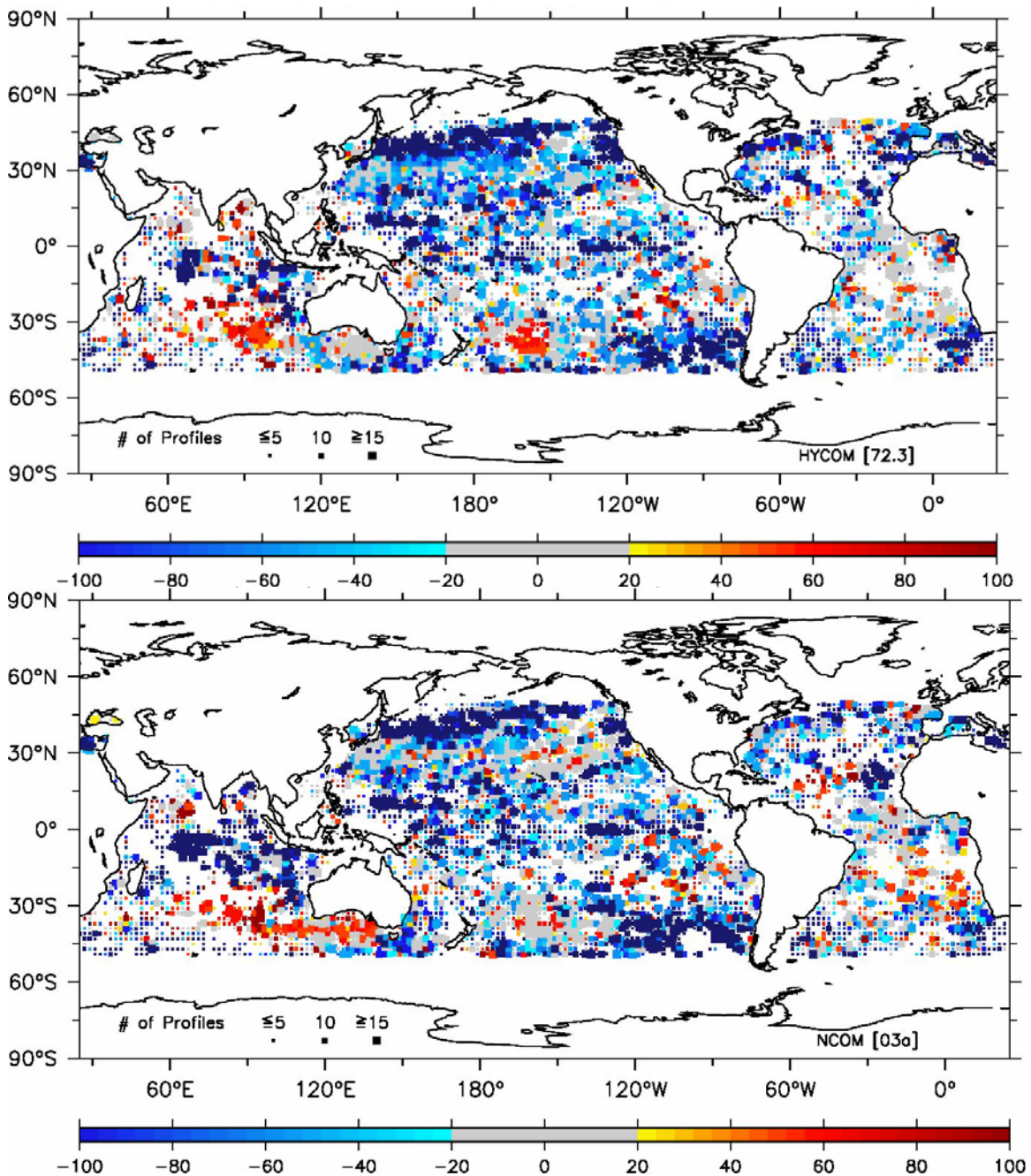


Figure 32: As in Figure 28 except for the DSC and the number of unassimilated profiles is 40187. For V3.0, the basin average MdBE = -49 m, RMSE = 201 m and 23% of the points are within 20 m of the observation. For V2.5, the basin average MdBE = -49 m, RMSE = 187 m and 23% of the points are within 20 m of the observation.

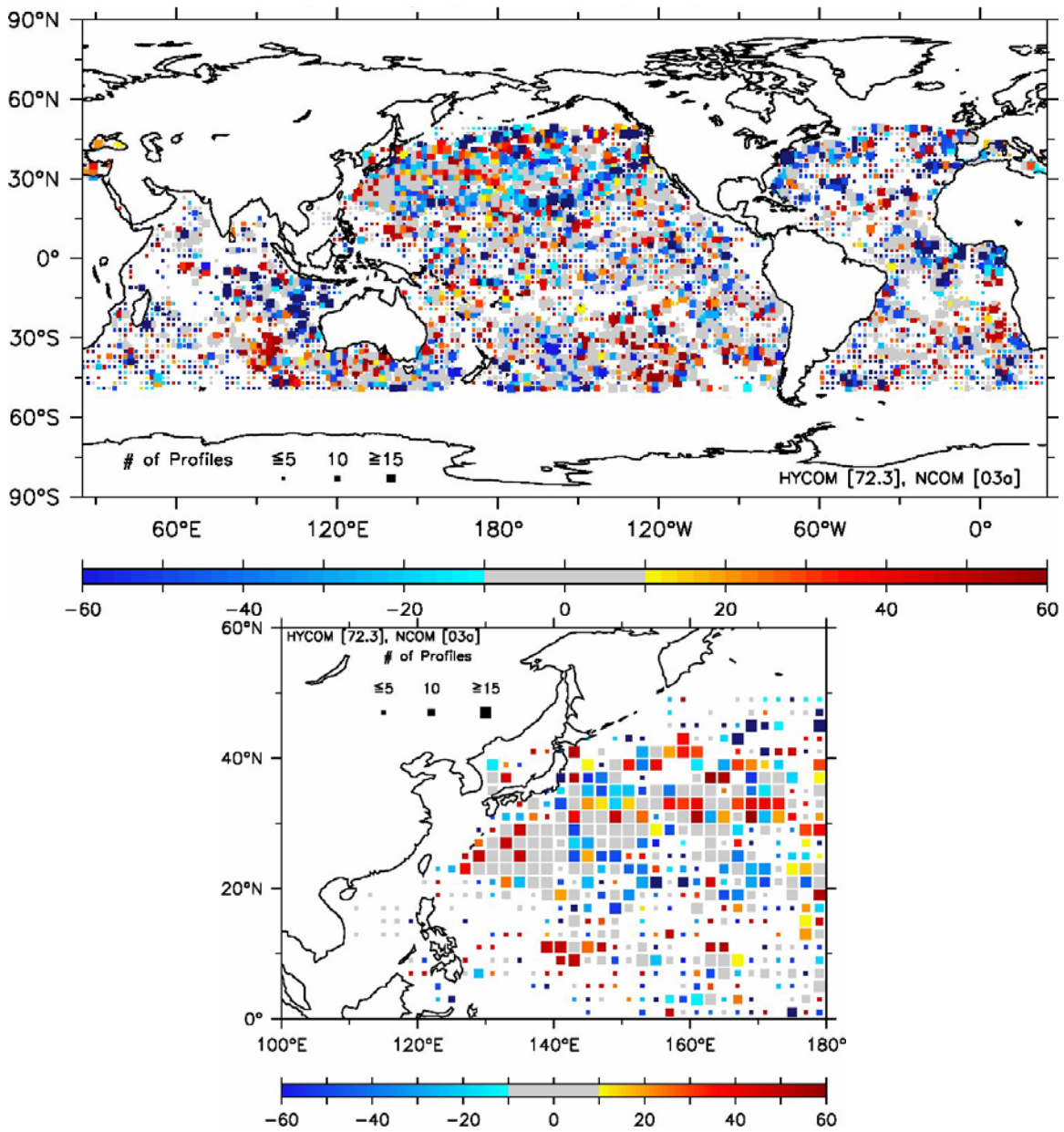
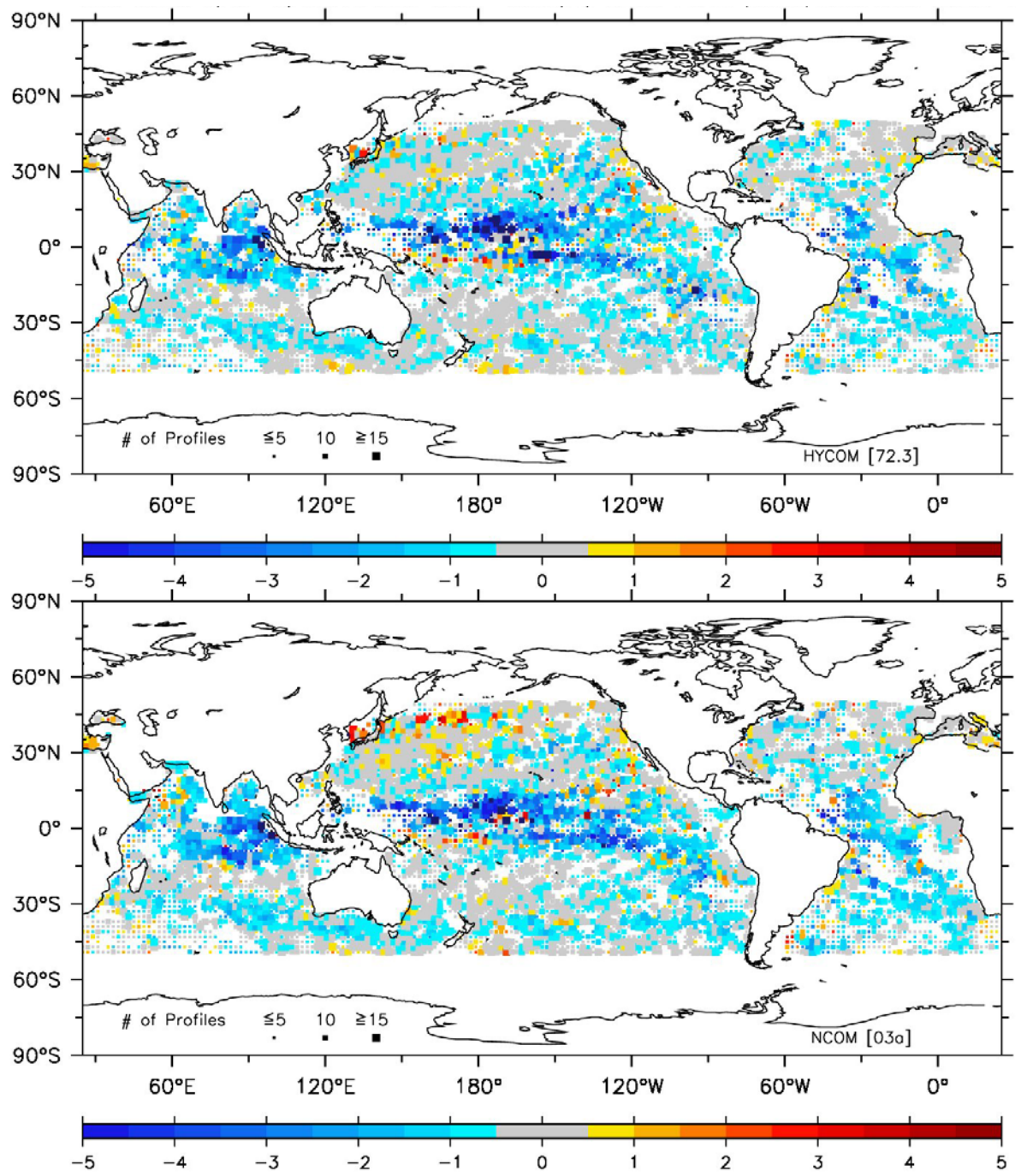


Figure 33: As in Figure 29 except for DSC using (top – whole domain) 40187 unassimilated profiles and (bottom – western Pacific) 6597 unassimilated profiles. The percentage of points where V3.0 has equal or less absolute error than V2.5 is 39% for the whole domain and 40% in the western Pacific.



**Figure 34:** As in Figure 28 except for BLG and the number of unassimilated profiles is 66495. For V3.0, the basin average MdbE = -0.4 m/s/100 ft, RMSE = 2.2 m/s 100 ft and 46% of the points are within 0.5 m/s 100 ft of the observation. For V2.5, the basin average MdbE = -0.4 m/s/100 ft, RMSE = 2.0 m/s 100 ft and 42% of the points are within 0.5 m/s/100 ft of the observation.

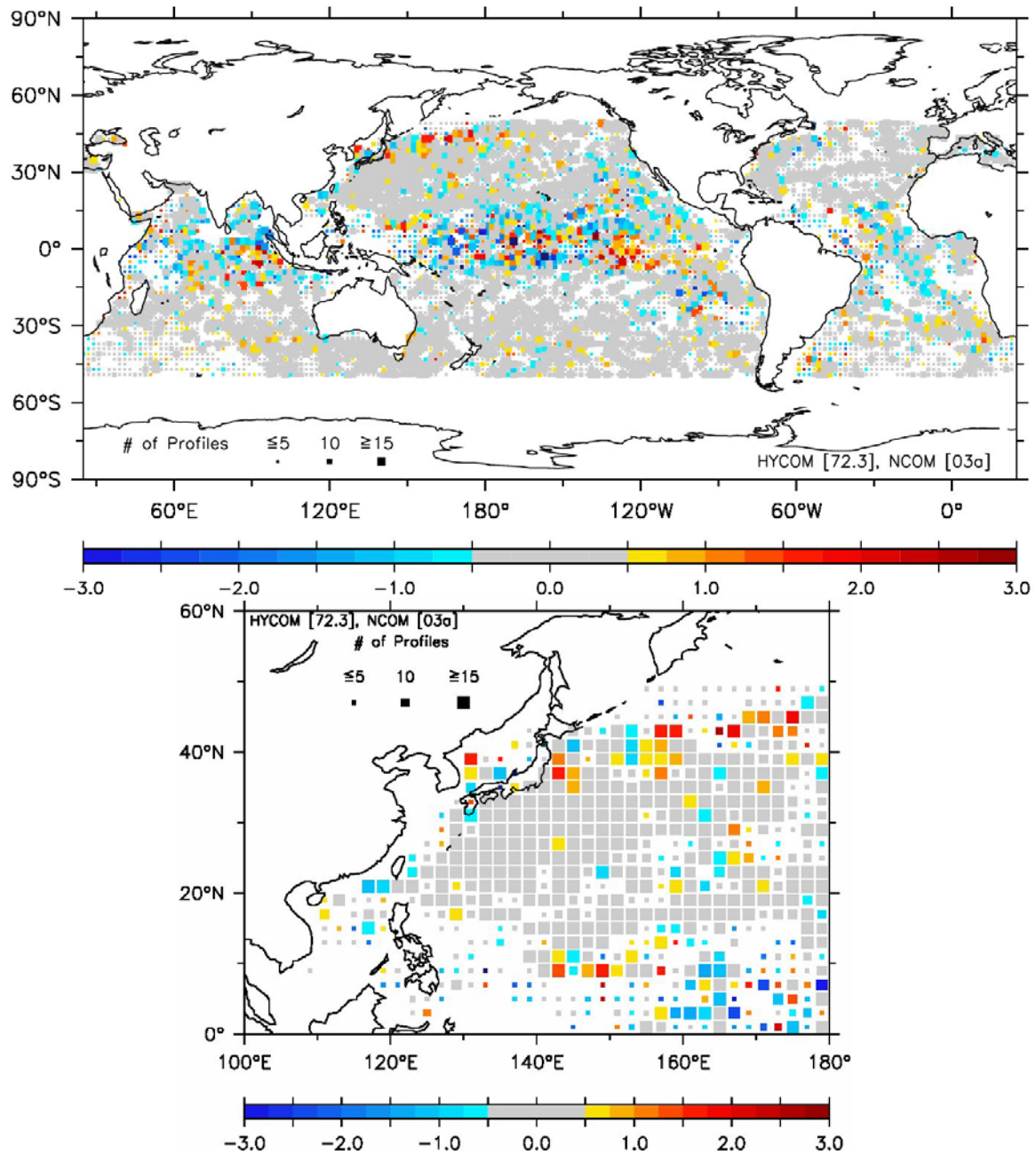


Figure 35: As in Figure 29 except for BLG using (top – whole domain) 66495 unassimilated profiles and (bottom – western Pacific) 9699 unassimilated profiles. The percentage of points where V3.0 has equal or less absolute error than V2.5 is 47% for the whole domain and 46% in the western Pacific.

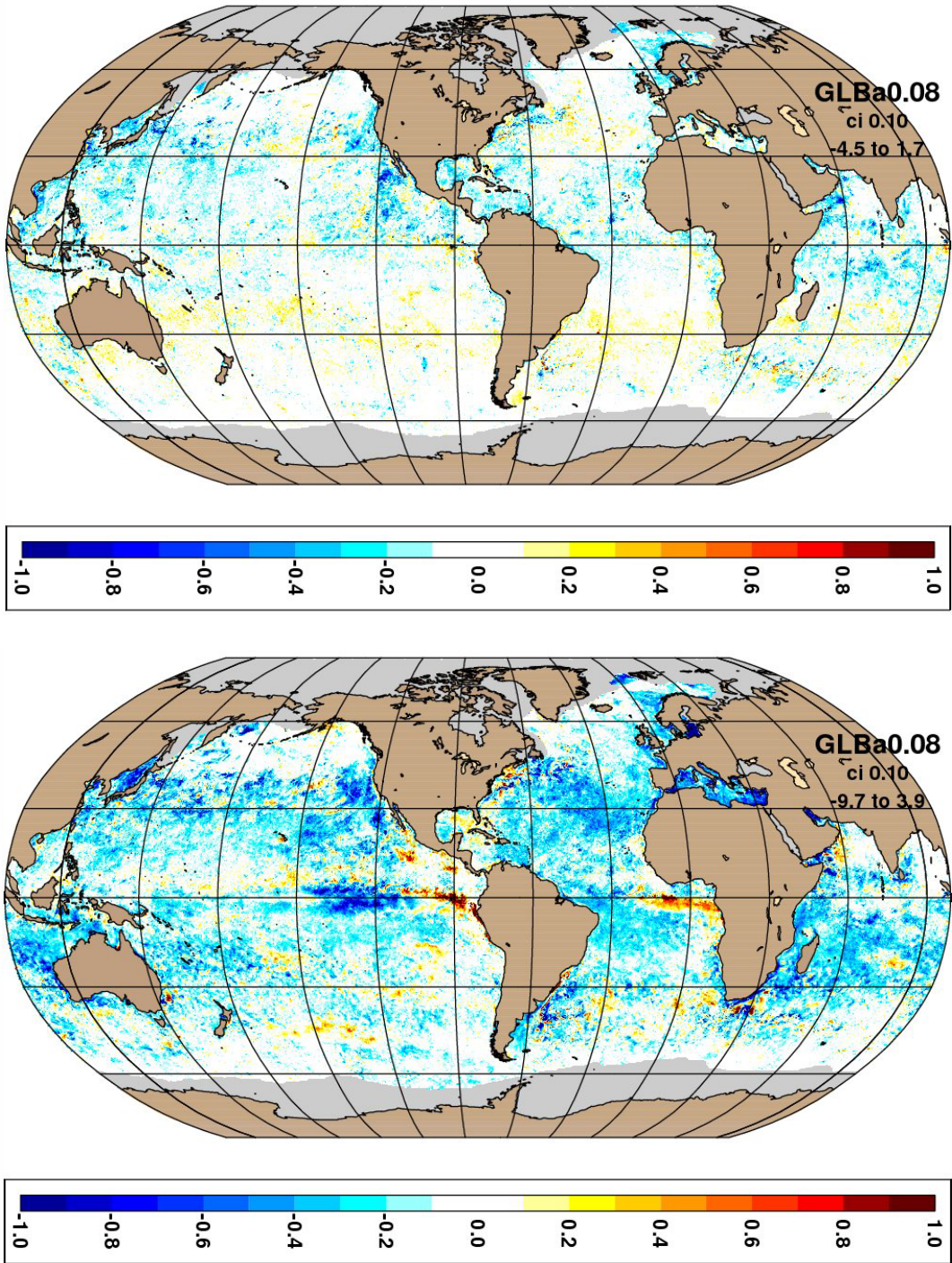


Figure 36: Sea surface temperature (SST) mean error (ME) relative to ~33,000,000 MCSST observations at the analysis time of the hindcasts for V3.0 (top) and V2.5 (bottom). Red (blue) colors indicate simulated SST is warmer (cooler) than observed. Values between  $\pm 0.1^\circ\text{C}$  are white. The gray area near the poles is a 1982-2007 annual mean sea ice coverage mask from the Climate Diagnostics Center optimum interpolation SST analysis.

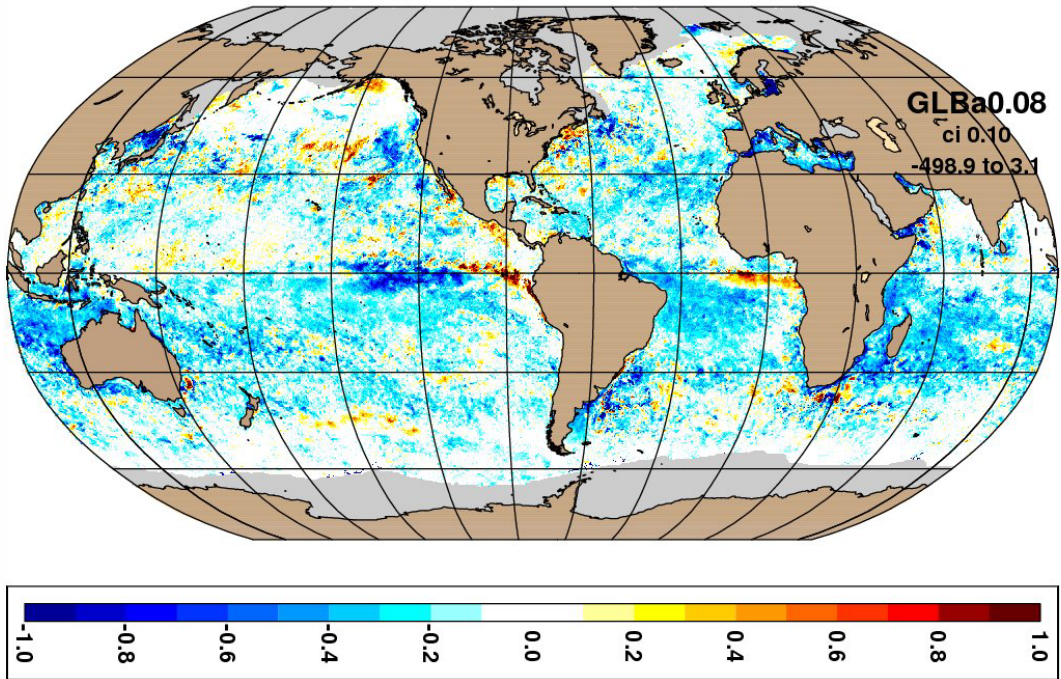
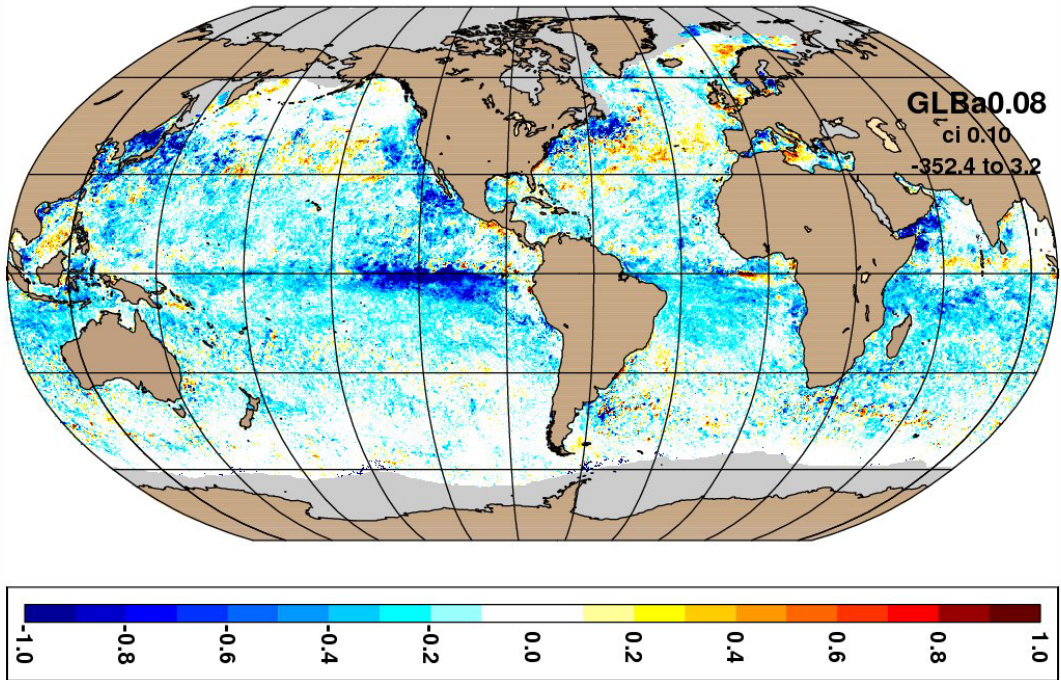
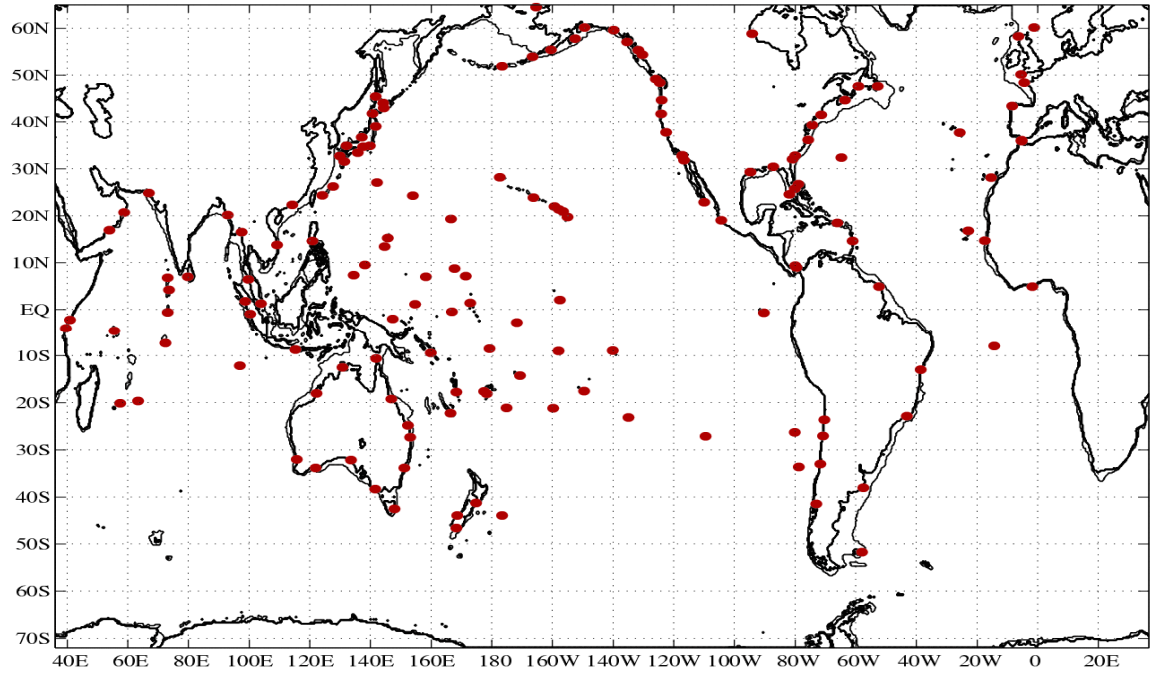
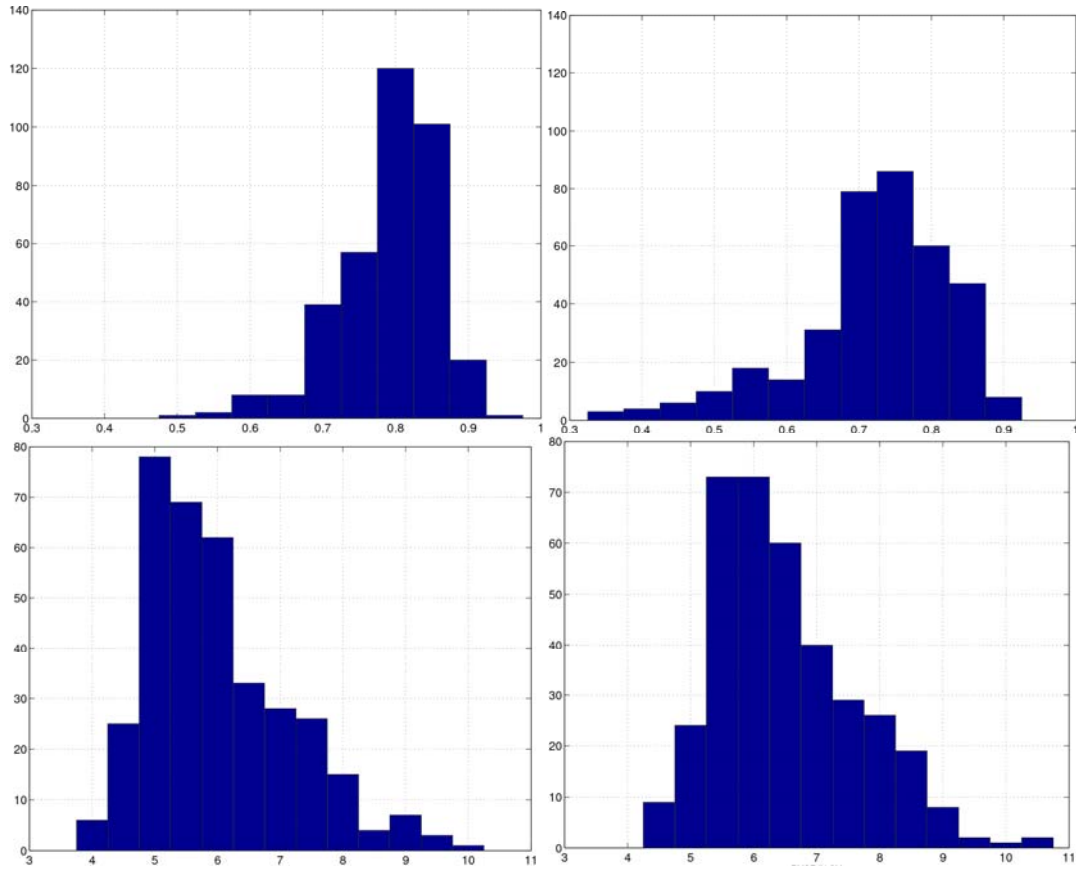


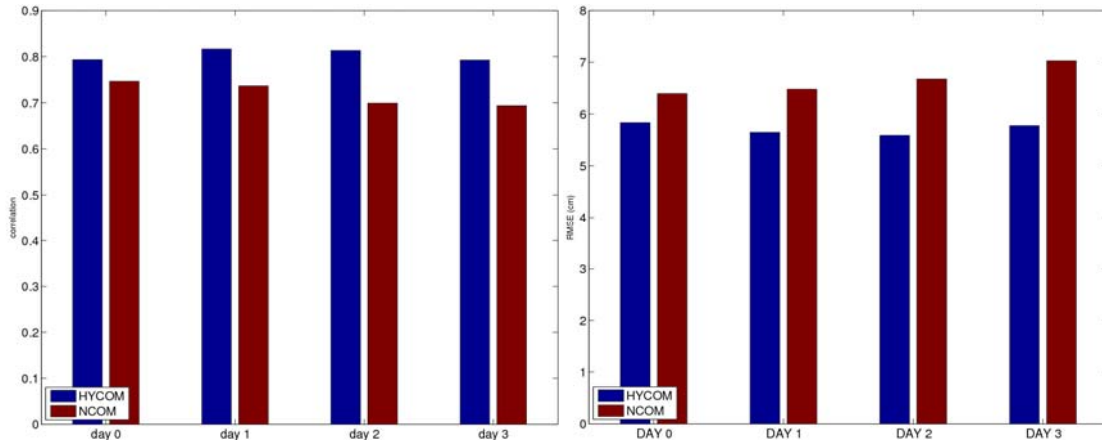
Figure 37: As in Figure 36 except for the 3-day forecast.



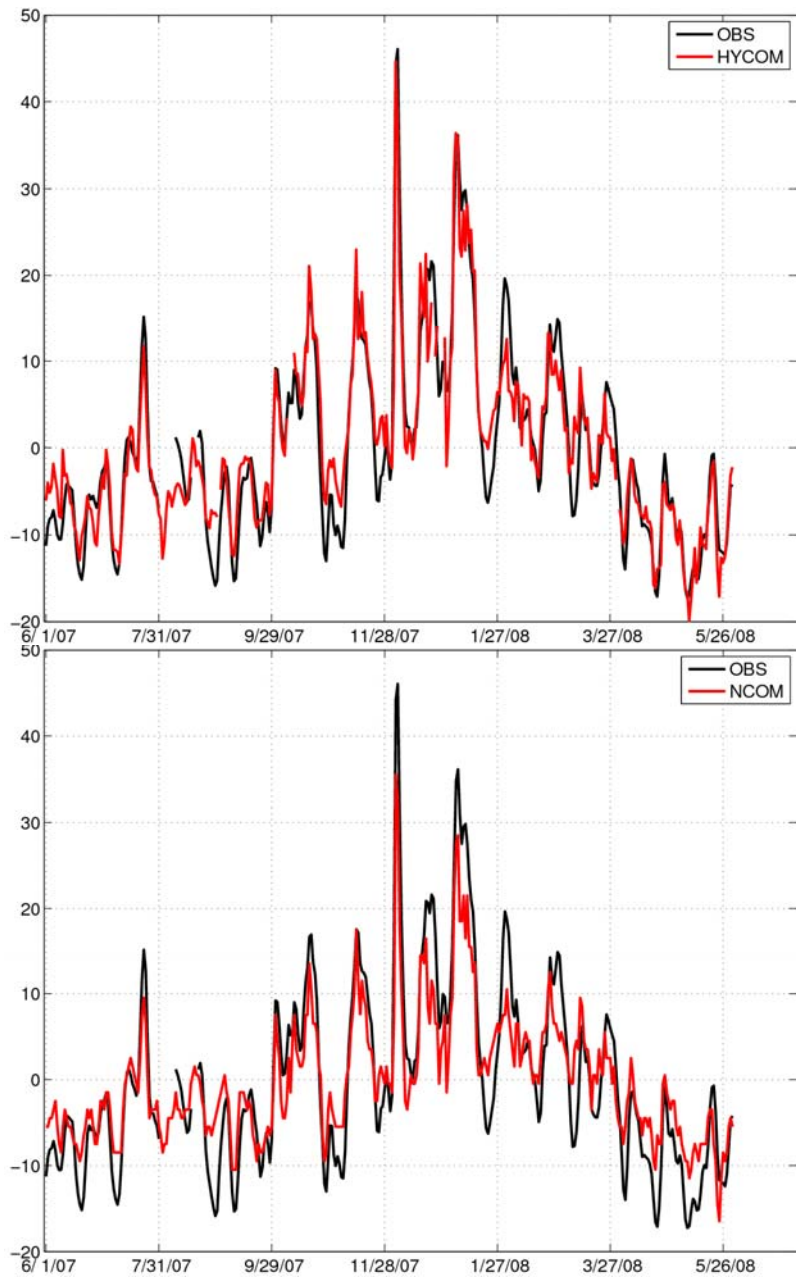
**Figure 38:** Locations of the 147 coastal and island sea level stations used in this analysis. Simulated sea level was sampled at the model gridpoint closest to the observation location.



**Figure 39: Histograms of correlation (top row) and RMSE (bottom row) for simulated vs. observed sea level at the analysis time during the hindcast period 1 June 2007 – 31 May 2008 at the 147 stations shown in Figure 38. The panels in the left column are V3.0 and those in the right column are V2.5. Median correlation is 0.8 (0.73) and median RMSE is 5.8 cm (6.3 cm) for V3.0 (V2.5). The statistics are computed basin-wide at each time point of the hindcast. The y-axis indicates the numbers of days in that bin, .05 for correlation and 0.5 for RMSE, and they sum to 366 days. The percentage of points in V3.0 (V2.5) with correlation higher than or equal to the bar centered on .8 is 68% (31%). The percentage of points in V3.0 (V2.5) with RMSE lower than or equal to the bar centered on 6 cm is 68% (49%).**



**Figure 40: Histograms of correlation (left) and RMSE (right) for simulated vs. observed sea level as a function of forecast length at the 147 stations shown in Figure 38. Each represents a composite correlation or RMSE for all stations. The blue histograms are for V3.0 and the red ones for V2.5. The left-most histograms in each panel are for the analysis time and the right-most histograms are for the 3-day forecast.**



**Figure 41: Observed (black) vs. simulated (red) sea level (cm) at Neah Bay, WA over the period of the hindcast 1 July 2007 – 31 May 2008. The top panel is V3.0 and the bottom panel is V2.5. The observations have been de-tided and atmospheric pressure loading effects removed. Note in early December 2007 the sharp rise of nearly 50 cm in both observed and simulated sea level. The correlation/RMSE is .94/3.7 cm (.93/4.8 cm) in V3.0 (V2.5).**

