



NRL/MR/7531--19-9935

# Earth System Prediction Capability (ESPC) Initial Operational Capability (IOC) Deterministic System

BENJAMIN RUSTON  
CAROLYN REYNOLDS  
TIM WHITCOMB  
MATT JANIGA

*Naval Research Laboratory  
Monterey, CA*

E. JOSEPH METZGER  
JAY SHRIVER

*Naval Research Laboratory  
Stennis Space Center, MS*

MARK COBB

*Fleet Numerical Meteorology and Oceanography Center  
Stennis Space Center, MS*

JOEL FELDMEIER

*Naval Postgraduate School  
Monterey, CA*

December 17, 2019

# 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> 17-12-2019			<b>2. REPORT TYPE</b> NRL Memorandum Report			<b>3. DATES COVERED (From - To)</b>			
<b>4. TITLE AND SUBTITLE</b>  Earth System Prediction Capability (ESPC) Initial Operational Capability (IOC) Deterministic System						<b>5a. CONTRACT NUMBER</b>			
						<b>5b. GRANT NUMBER</b>			
						<b>5c. PROGRAM ELEMENT NUMBER</b> 75-4726-E9-5			
<b>6. AUTHOR(S)</b>  Benjamin Ruston, Joseph Metzger,* Jay Shriver,* Carolyn Reynolds, Tim Whitecomb, Matt Janiga, Mark Cobb,** and Joel Feldmeier***						<b>5d. PROJECT NUMBER</b>			
						<b>5e. TASK NUMBER</b>			
						<b>5f. WORK UNIT NUMBER</b> 4726			
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b>  Naval Research Laboratory 1 University Circle Monterey, CA 93943						<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  NRL/MR/7531--19-9935			
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>  Office of Naval Research One Liberty Center 875 North Randolph Street, Suite 1425 Arlington, VA 22203-1995						<b>10. SPONSOR / MONITOR'S ACRONYM(S)</b>  ONR			
						<b>11. SPONSOR / MONITOR'S REPORT NUMBER(S)</b>			
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  <b>DISTRIBUTION STATEMENT A:</b> Approved for public release; distribution is unlimited.									
<b>13. SUPPLEMENTARY NOTES</b>  *Naval Research Laboratory, 1005 Balch Boulevard, Stennis Space Center, MS 39529; **Naval Postgraduate School, 1 University Circle, Monterey, CA 93943; ***Fleet Numerical Meteorology and Oceanography Center, 1005 Balch Boulevard, Stennis Space Center, MS 39529									
<b>14. ABSTRACT</b>  The Navy Earth System Prediction Capability (ESPC) is a coupled model that includes the NAVGEM atmospheric global forecast model, the HYCOM ocean model, and CICE sea ice model coupled together using the Earth System Modeling Framework. This report details the testing and validation of this system running in deterministic mode, which is a single forecast and data assimilation run beginning in December of 2016 and continuing throughout the calendar year of 2017. This validation and testing report is part of the delivery of the Initial Operational Capability (IOC) to the Navy and can be used to help guide decisions on operational promotion of the Navy ESPC system, which has been designed to provide consistent, high-quality global environmental information for Navy and DoD operations at lead times from weeks to months.									
<b>15. SUBJECT TERMS</b> Navy-ESPC    NAVGEM    HYCOM    CICE Ensemble    Global coupled forecasting    Model validation									
<b>16. SECURITY CLASSIFICATION OF:</b>						<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> Benjamin Ruston	
<b>a. REPORT</b> Unclassified Unlimited	<b>b. ABSTRACT</b> Unclassified Unlimited	<b>c. THIS PAGE</b> Unclassified Unlimited			Unclassified Unlimited			122	<b>19b. TELEPHONE NUMBER (include area code)</b> (831) 656-4020

This page intentionally left blank.

## Executive Summary & Recommendations

The Initial Operational Capability (IOC) delivery of the deterministic portion of the Earth System Prediction Capability (ESPC) has been tested on Department of Defense (DoD) High Performance Computing Modernization Program (HPCMP) architectures and compared to U.S. Navy operational global modeling systems for the atmosphere, ocean, and sea ice. The Navy ESPC at IOC is designed to provide consistent, high-quality global environmental information for Navy and DoD operations at lead times from weeks to months.

The coupled model includes the NAVGEM atmospheric global forecast model, the HYCOM ocean model, and CICE sea ice model coupled together using the Earth System Modeling Framework. Development was carried out using two configurations: a high-resolution deterministic forecast system described in this Validation Test Report (VTR) and a lower-resolution ensemble forecast system described in a separate VTR. While the models are very similar to their uncoupled versions, there are several differences outlined in this report – key among these is the inclusion of an older version of CICE used in Navy ESPC (CICE 4.0) compared to GOFS 3.5 (which uses CICE 5.1.2).

Because of latency and computational cost, the deterministic Navy ESPC system is not a candidate to replace the deterministic uncoupled NAVGEM system. However, Navy ESPC deterministic is a candidate for eventual replacement of the standalone GOFS system once skill parity can be achieved. As outlined in this report, although Navy ESPC performs broadly neutral relative to the uncoupled baselines, there are several key metrics where superior GOFS forecasts demonstrate the need for continued execution of the standalone GOFS system.

Based upon the evaluation of the deterministic system performed by the NRL ESPC team and the validation test panel, we provide the following findings and recommendations:

1. The deterministic Navy ESPC system provides a new deterministic coupled prediction capability with key improvements in certain Navy-relevant parameters (e.g. tropical 10m winds and near surface (upper 30 m) temperature and salinity).
2. Compared to the current operational capability, the performance of the deterministic coupled system is broadly comparable to current operational systems at tactical forecast lead times (0 to 7 days).
3. While comparisons with the uncoupled controls are overall neutral, there are several areas where the current uncoupled systems provide superior guidance (sonic layer depth/below layer gradient for ocean acoustics, and sea ice edge in the polar latitudes).
4. This initial version of Navy ESPC is suitable for transition to operations.
5. It does provide skill over climatology and persistence in the 8-16 day forecast range.

This page intentionally left blank.

## **1 Introduction**

This validation test report (VTR) discusses the Initial Operational Capability (IOC) delivery of the deterministic portion of the Earth System Prediction Capability (ESPC) that will provide global environmental information to meet Navy and DoD operational and planning needs from under the sea to the upper atmosphere. The final version of ESPC is scheduled to be a fully coupled global atmosphere/ocean/sea ice/wave/land/aerosol prediction system. The deterministic IOC delivery is intended to provide daily predictions out to 16 days. A separate VTR will address the performance of the Navy ESPC ensemble system that produces weekly forecasts out to 45 days. The combined deterministic and ensemble IOC capability being delivered in 2019 is a subset of the final system, and targets use at the Navy DoD Supercomputing Resource Center (DSRC).

Navy ESPC was largely developed and run at Navy DSRC using four Cray supercomputers on the unclassified side; however, delivery is targeting the two new HPE SGI 8600 supercomputers with 736 compute nodes and 48 cores/nodes for a total of 35,328 cores per machine. Presently, Commander Naval Meteorology and Oceanography Command (CNMOC) receive 15% of the unclassified cycles (~450 Teraflops (TF), ~5100 cores) at the Navy DSRC for operational systems and ESPC will be required to fit within this allotment.

## **2 Navy Earth System Prediction Capability Forecast Suite**

The IOC components of the Navy ESPC global prediction system consist of: the HYbrid Coordinate Ocean Model (HYCOM), the Community Ice CodE (CICE), and the NAVy Global Environmental Model (NAVGEN). The Navy ESPC IOC delivers the first fully coupled global forecast capability for the Ocean/Sea-Ice/Atmosphere components. The Navy ESPC IOC is accompanied by data assimilation components where the Navy Coupled Ocean Data Assimilation (NCODA) is used for the ocean, sea ice, and wave components and the NRL Atmospheric Variational Data Assimilation System – Accelerated Representer (NAVDAS-AR) is used for the atmosphere.

### **2.1 Global Forecast Systems**

The Navy ESPC IOC couples the ocean, sea-ice, and atmosphere between the HYCOM, CICE, and NAVGEN components.

#### **2.1.1 HYCOM/CICE -- ocean and sea-ice models**

The HYbrid Coordinate Ocean Model (HYCOM) and the Community Ice CodE (CICE) began running in a coupled mode for the Global Ocean Forecasting System (GOFS) v3.1 in 2015. HYCOM is a primitive equation ocean general circulation model capable of nowcasting and forecasting the 3-dimensional temperature, salinity and current structure of the global ocean. Its grid is uniform cylindrical from 78.64°S to 66.0°S, on a Mercator projection from 66.0°S to 47°N and curvi-linear north of this, employing an Arctic dipole patch where the poles are shifted over land to avoid a singularity at the North Pole. A similar configuration is not needed in the Southern Hemisphere since the South Pole is situated over land. It employs potential density referenced to 2000 m and

includes the effects of thermobaricity (Chassignet et al., 2003). Vertical coordinates can be isopycnals (density tracking), often the best coordinate 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) but uses the K-Profile Parameterization. A more complete description of HYCOM physics can be found in Bleck (2002). An application of global HYCOM within the Indonesian Sea can be found in Metzger et al. (2010) and GOFS is described in Metzger et al. (2014). The HYCOM version number delivered for IOC is 2.2.99DHi.

The Los Alamos-developed CICE model (Hunke and Lipscomb, 2008) is coupled within the Navy ESPC system via the Earth System Modeling Framework (ESMF) (Hill et al., 2004). The sea ice and ocean models use the same grid configuration and pass information back and forth every hour. CICE includes sophisticated ice thermodynamics such as multiple ice thickness layers, multiple snow layers and the capability to forecast multi-categories of ice thickness according to World Meteorological Organization definitions. In addition, CICE has several interacting components including a thermodynamic model that computes local growth rates of snow and ice due to snowfall; vertical conductive, radiative and turbulent fluxes; a model of ice dynamics that predicts the velocity field of the ice pack based on a model of the material strength of the ice; a transport model that describes advection of the areal concentration, ice volumes and other state variables; and a ridging parameterization that transfers ice among thickness categories based on energetic balances and rates of strains. The CICE version number delivered for IOC is 4.0.

## **2.1.2 NAVGEM -- atmospheric model**

### **2.1.2.1 NAVGEM forecast model overview**

As of this writing, NAVGEM version 1.4.3 is run operationally at the Fleet Numerical Meteorology and Oceanography Center (FNMOC). The ESPC implementation of the NAVGEM forecast model in the IOC coupled system retains most of the features of the forecast model in NAVGEM 1.4.3 (validated in the NRL/FNMOC NAVGEM v1.4.3 test report - *DoD distribution only*). An important exception is the addition of new model physics, largely developed especially for application in the coupled system. The dynamics framework utilized for the IOC deterministic (T681L60) (18 km equivalent horizontal grid) follows that of NAVGEM 1.4.3 in its use of the same three-time level semi-Lagrangian, semi-implicit dynamical core, perturbation virtual potential temperature for the prognostic temperature variable and use of the reduced Gaussian grid. Differences in the ESPC IOC NAVGEM ensemble configuration will be detailed separately. Of note, however, the ensemble (T359L60) (37 km equivalent horizontal grid) configuration follows NAVGEM 1.2

(T359L50) in retaining the virtual potential temperature as the prognostic temperature variable, and using the full Gaussian grid. The NAVGEM model uses a terrain following hybrid-sigma vertical coordinate and the deterministic and ensemble configurations retain the same 60 vertical levels, with a model top of 0.04 hPa (~ 72km), used in NAVGEM versions 1.3 – 1.4.3. The NAVGEM version number delivered for IOC is NAVGEM v1.4.3; however this version of the model has incorporated latent capabilities to alter the physics which are enabled for Navy ESPC and are detailed in the following section.

### **2.1.2.2 NAVGEM coupled physics**

The implementation of the NAVGEM forecast model in the Navy-ESPC system retains most of the features of the forecast model in NAVGEM 1.4.3, which is the current stand-alone configuration at FNMOC. An important exception being the addition of a new suite of model physics referred to as Coupled Version Physics (CVP) (Table 1). The CVP physics has been tailored for the coupled system with two primary objectives: 1) to improve the representation of the Madden Julian Oscillation (MJO), which is recognized as a key contributor to extended range predictability, and 2) to improve the consistency between surface fluxes computed in NAVGEM and in the ocean model HYCOM. The physics differences are largely associated with atmospheric convection and air-sea fluxes.

Representation of the MJO in the coupled system is enhanced significantly through the implementation of a new treatment of atmospheric convection as part of the CVP. This modified version of the Kain-Fritsch scheme (Kain and Fritsch 1990; Kain and Fritsch 1993) incorporates both turbulence- and dynamically forced modes, and is an extension of the work of Ridout et al. (2005). As in the previous version, the new scheme includes a modified closure formulation for the Kain-Fritsch dynamically forced mode based on an assumed quasi-balance of updraft parcel buoyancy at the cloud base level. This closure formulation requires the scheme to be called at every time step to adjust the cloud base mass flux in a similar manner as in the Emanuel convection scheme (Emanuel 1991; Emanuel and Zivkovic-Rothman 1999). The scheme similarly includes from the 2005 version a modified updraft source level selection procedure, which was adopted from the implementation of the Emanuel scheme developed by Peng et al. (2004), which became the operational NOGAPS version of the scheme. Several other enhancements have been added for the current scheme, including addition of a treatment of convective momentum transport, a modified representation of the rate of updraft-environment mixing adapted from Peng et al. (2004), and a modified cloud top condition that enhances the sensitivity of convection to dry layers and associated thermal inversions. Convective momentum transport is treated using the adaptation of the treatment of Gregory et al. (1997). The turbulence-forced mode of convection is represented using a slightly modified version of the scheme employed for the dynamically forced mode. For this mode the precipitating downdraft code is turned off, and the cloud base mass flux is parameterized based on the mass flux at the lifting condensation level of boundary layer plumes in a slightly modified version of the NAVGEM EDMF scheme. Triggering of the turbulence-forced mode by the plumes is represented using the mixed-layer Richardson number convective trigger formulation described by Ridout and Reynolds (1998).

In regards to surface fluxes in the coupled system, issues with consistency between the component models arise in part due to the hourly exchange of variables (i.e., coupling frequency) currently implemented between NAVGEM and HYCOM and CICE. Given this coupling frequency, to provide for sufficiently responsive updates to the surface fluxes over the ocean in NAVGEM and HYCOM, rather than exchange fluxes between the models, each model computes its own fluxes, with NAVGEM providing 10-m winds and near surface (2-m) temperature and moisture to HYCOM, and HYCOM providing sea surface temperatures (SSTs) and surface currents to NAVGEM. The CVP helps to mitigate resultant differences in surface fluxes felt by the ocean and atmosphere by implementing the HYCOM surface flux scheme in NAVGEM. This scheme is an adaptation Kara et al. (2005) of the COARE 3.0 scheme of Fairall et al. (2003). Notably, the coupled system framework enables for the first time the inclusion of the surface current contribution to the near-surface shear in the NAVGEM surface flux computation. This treatment is consistent not only with the implementation of the COARE 3.0 scheme in HYCOM, but also with the development of the scheme itself, in which surface current corrections were made to the near surface wind speed (Fairall et al. 2003).

As noted, Navy ESPC coupled system modifies the current standalone version of NAVGEM (see summary in Table 1). In addition, within the ESPC IOC there are differences between the NAVGEM version used deterministically, and that used for ensemble forecasts (i.e., as discussed in this VTR). There are also slight physics changes between these resolutions. The dynamics framework utilized for the deterministic (T681L60) (18 km equivalent horizontal grid) configuration follows that of NAVGEM 1.4.3 in its use of the same three-time level semi-Lagrangian, semi-implicit dynamical core, perturbation virtual potential temperature for the prognostic temperature variable and use of the reduced Gaussian grid. The ensemble (T359L60) (37 km equivalent horizontal grid) configuration follows NAVGEM 1.2 (T359L50) in retaining the virtual potential temperature as the prognostic temperature variable, and using the full Gaussian grid. The adiabatic correction option introduced in NAVGEM 1.2.2 is turned off for the ensemble resolution. Configurations for both NAVGEM resolutions used in Navy-ESPC retain the 60-level vertical grid used in NAVGEM versions 1.3 – 1.4.3.

The NAVGEM version used in the Navy-ESPC coupled system benefits from the CVP, which improved the consistency between surface fluxes in the coupled system and improved the representation of the MJO. Prior to this physics change, NAVGEM was not competitive with the best MJO forecast models as shown in the model intercomparison study of Jiang et al. (2015). The CVP led to a very significant improvement in the MJO such that the Navy-ESPC model compares favorably to other state-of-the-art forecast models in representation of the MJO as seen through evaluation of the 17 years of Navy-ESPC forecasts produced for the NOAA SubX project (Janiga et al. 2018).

The CVP suite has a default setting and optional versions that can be specified by setting a new environmental variable "CVP\_TYPE". The current default setting is "CVP\_TYPE=3", which selects the version of the suite used for the current VTR testing. Setting "CVP\_TYPE=2" selects the physics code that is being used in the Navy-ESPC coupled system for the ongoing North American

Multi-Model Ensemble for the Subseasonal eXperiment (SubX) runs, and “CVP\_TYPE=0” selects the NAVGEM 1.4.3 version physics.

**Table 1: Summary of NAVGEM changes between the NAVGEM v1.4 version and the NAVGEM v1.4-CV3 used in the Navy-ESPC system. The main changes between the two atmospheric models include changes in the convection parameterization and boundary layer scheme.**

Parameterization/Scheme	NAVGEM v1.4	NAVGEM v1.4-CV3
Convection Parameterization	SAS (Moorthi et al. 2001)	Modified Kain- Fritsch
Boundary Layer Scheme	Louis et al. (1982)	COARE (Kara et al. 2005)
Prognostic Temperature Variable	Perturbation Virtual Potential Temperature	Virtual Potential Temperature
Grid	Thinned Gaussian	Full Gaussian
Adiabatic Correction	Used	Turned off for ensemble resolution

## 2.2 Data Assimilation Systems

The Navy ESPC system utilizes two data assimilation schemes. The Navy Coupled Ocean Data Assimilation (NCODA) for the ocean, sea-ice, and wave models and the Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System-Accelerated Representer (NAVDAS-AR) for the atmosphere. Currently all components run loosely coupled, though the ocean and sea-ice both use the NCODA system to perform their data assimilation. The increments produced by NCODA for ocean and the NAVDAS-AR atmospheric components are added in unison incrementally to the coupled Navy ESPC system, while the sea ice increments from NCODA are directly inserted.

### 2.2.1 NCODA –ocean/ice assimilation scheme

The Navy Coupled Ocean Data Assimilation (NCODA) is a fully three-dimensional, multivariate (3DVar) data assimilation scheme (Cummings 2005; Cummings and Smedstad 2013) for the following ocean/ice variables: temperature, salinity, geopotential, vector velocity components, and ice concentration; all are analyzed simultaneously. Data are selected for assimilation based on receipt time (i.e., the time the observation is received at the center) instead of the observation time so, any data received since the previous NCODA analysis are used in the next analysis. For each data type, the user defines the maximum age of data to be used in the analysis. All data will not necessarily be from synoptic times, so they can be compared against a time dependent background field using the First Guess at Appropriate Time (FGAT). Hourly forecast fields are used in FGAT for assimilation of Sea Surface Temperatures (SSTs) to maintain the diurnal cycle, whereas daily averaged forecast fields are used in FGAT for profile data, both synthetic and real. NCODA is cycled with HYCOM and CICE to provide updated initial conditions for the next model forecast using an incremental analysis update procedure (Bloom et al. 1996).

In GOFS 3.1, the NCODA ocean analysis increments are inserted into HYCOM over a six hour window, whereas the NCODA ice analysis is directly inserted into CICE. The analysis corrections to the HYCOM and CICE forecasts are based on all observations that have become available since the last analysis, which may include observations made prior to the previous analysis data window. These include surface observations from satellites, including altimeter Sea Surface Height (SSH) anomalies, SST, and sea ice concentration, plus in situ SST observations from ships and buoys as well as temperature and salinity profile data from XBTs, CTDs and Argo floats (Table 3). See Table 13.1 in Cummings and Smedstad (2013) for a complete list of assimilated observations along with typical data counts. All observations are first quality controlled, and this is done via NCODA\_QC (Quality Control), which is operational at FNMOC.

### **2.2.2 NAVDAS-AR -- atmospheric assimilation scheme**

The Naval Research Laboratory (NRL) Atmospheric Variational Data Assimilation System Accelerated Representer (NAVDAS-AR) (Xu et al., 2005; Rosmond and Xu, 2006) has the ability for both weak and strong constraint variational assimilation, and is formulated in the terms of dual variables (“observation space”), the dimension of which is generally much smaller than the corresponding state (“model”) space. This restriction of the observation space is accomplished by discarding the unobservable degrees of freedom in the system (Bennett, 2002).

NAVDAS-AR can process well over 100 million observations in every 6 hour data assimilation window. After quality control and data thinning, a net of approximately 4 million observations are typically assimilated to create the final analysis. Observation types routinely assimilated are listed in Table 3. This list includes: conventional observations (e.g. radiosondes, dropsondes, buoys, etc.); aircraft observations; feature tracked winds; satellite based radiance measurements from sensors such as AMSU-A, MHS, SSMIS, AIRS, ATMS, IASI, CrIS, SAPHIR, GMI, AMSR/2, as well as a suite of geostationary sensors which provide a clear-sky radiance product; bending angles from Global Navigation Satellite System Radio Occultation (GNSS-RO); SSMIS and WindSat wind speeds, scatterometer winds, and synthetic observations of tropical storms. NAVDAS-AR approximates the full non-linear Navy ESPC coupled forecast system with a series of linear coupled Euler-Lagrange equations, which are in essence a simplified version of the stand-alone NAVGEM system and allow for a computationally efficient solve of the assimilation problem. The resolutions of the adjoint and tangent-linear models (and thus the effective analysis resolution) are approximately 100 km (T119). The same vertical model resolution (60 levels, from the surface to 0.04 hPa) is used by the full Navy ESPC system and the adjoint and tangent-linear models (and the resulting analysis). The solution (or best estimate of the atmospheric state) is obtained through a series of TLM and ADJ model integrations. The number of model integrations depends on the number of observations and their fit to the full resolution Navy ESPC coupled model. Typically about 70 iterations are required for the current operational system to meet the pre-specified convergence criterion. This is the most computationally intensive part of the 4D-Var analysis. The computation of the convolution of the background error covariance matrix with the adjoint sensitivity field at the initial time represents a substantial portion of each model integration computational cost.

## 2.3 ESPC coupling and forcing

The Navy ESPC system is coupled via the ESMF. This framework allows components such as the atmospheric, ocean, wave, sea-ice, and land surface models (planned) to dynamically interact at specified coupling timesteps. This allows the different components to adjust for the processes described by the various other components. Further, by keeping these large data in-memory it reduces the file I/O of the system compared to running each component separately. Figure 1 describes the fields passed between the components via the ESMF. Note the stand-alone atmospheric system NAVGEM typically holds the ocean surface temperature and sea ice fixed over the forecast so the coupled system provides a substantial improvement to this forcing particularly for forecasts longer than a few days. No bias corrections are made to any of the fluxes exchanged between model components. This differs from GOMS 3.1 and 3.5 in that bias corrections are made to operational NAVGEM that provides the surface forcing. Specifically, operational NAVGEM wind speed has a scaling factor and offset applied that is based on a regression analysis against past scatterometer data. A scaling factor is also applied to operational NAVGEM radiative fluxes (downward shortwave and longwave) relative to the NASA Clouds and the Earth's Radiant Energy System (CERES) satellite data. Lastly, a heat flux offset is added to HYCOM that is based on 5-day forecast SST error relative to the verifying analysis. These bias corrections vary both spatially and temporally but are one-time corrections and cannot account for interannual variability. They have been used in stand-alone GOMS to reduce surface forcing errors under the assumption that all the error is coming from the atmosphere. However, SST error could also arise from the ocean model due to advection of misplaced currents.

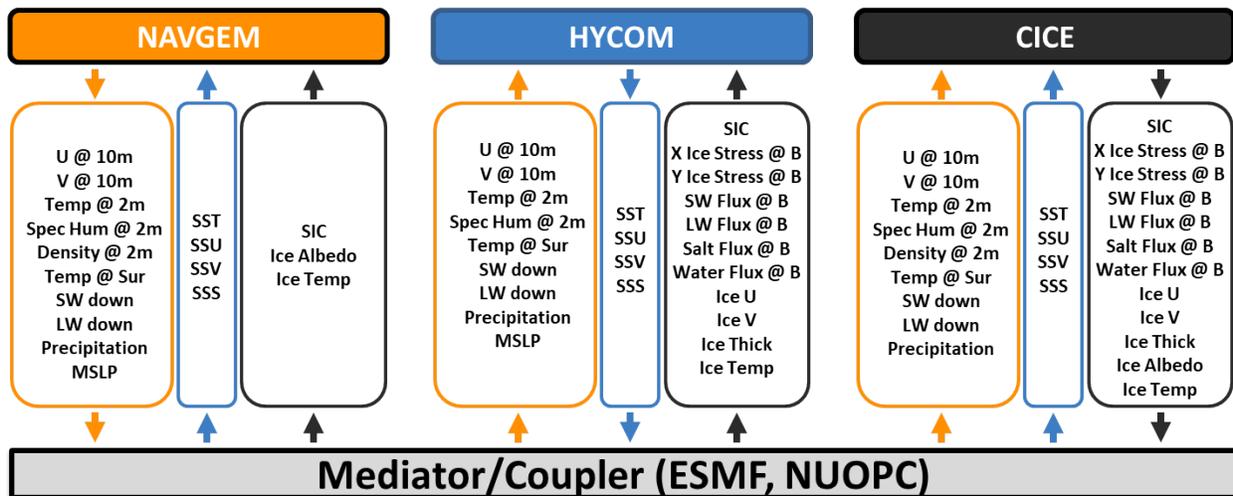


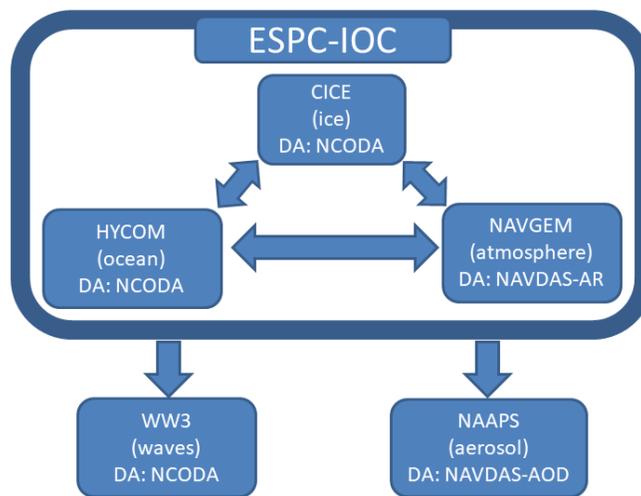
Figure 1. List of fields passed to and from the components (i.e., variables that are coupled). The color represents the component/model the field is from and the arrows represent if the fields are going into or out of the component. Long name for the atmospheric fields include: U @ 10m -> zonal winds at 10 meters, V @ 10m -> meridional winds at 10 meters, Temp @ 2m -> air temperature at 2 meters, Spec Hum @ 2m -> specific humidity at 2 meters, Density @ 2m -> air density at two meters, Temp @ Sur -> temperature of the surface (ground, ice, ocean), SW down -> downward surface shortwave radiative flux, LW down -> downward surface longwave radiative flux, Precipitation -> total precipitation, MSLP -> mean sea level pressure. Long names for the ocean variables include: SST -> sea surface temperature, SSU -> zonal sea surface current, SSV -> meridional sea surface current, SSS -> sea surface salinity. Long names for the sea ice variables include: X Ice Stress @ B -> x direction ice

stress at sea ice basal, Y Ice Stress @ B -> y direction ice stress at sea ice basal, SW Flux @ B -> shortwave radiative flux at sea ice basal, LW Flux @ B -> longwave radiative flux at sea ice basal, Salt Flux @ B -> flux of salt at sea ice basal, Water Flux @ B -> flux of fresh water at sea ice basal, Ice U -> zonal velocity of sea ice at the surface, Ice V -> meridional velocity of sea ice at the surface, Ice Thick -> ice thickness, Ice Albedo -> sea ice albedo, Ice Temp -> temperature of sea ice at surface.

### 3 ESPC IOC Configuration

#### 3.1 Navy ESPC System Component Configuration

A schematic of the Navy ESPC IOC coupled system is shown in Figure 2. Two-way coupling exists between the HYCOM, CICE, and NAVGEM components. A land surface model is embedded inside NAVGEM, but in the future a more advanced land surface model with a specific data assimilation component can be considered to be included in the coupled system. Aerosol loading in the atmosphere is predicted via the Navy Aerosol Analysis and Prediction System (NAAPS), and at this time is run independent of the coupled Navy ESPC. Work had progressed rapidly in including the aerosol capability in-line within NAVGEM which will allow this to be brought into Navy ESPC by FOC. For the IOC, WAVEWATCH III (WW3) is not being delivered as an active component.



**Figure 2. A schematic of the ESPC-IOC coupled system delivered in 2019.**

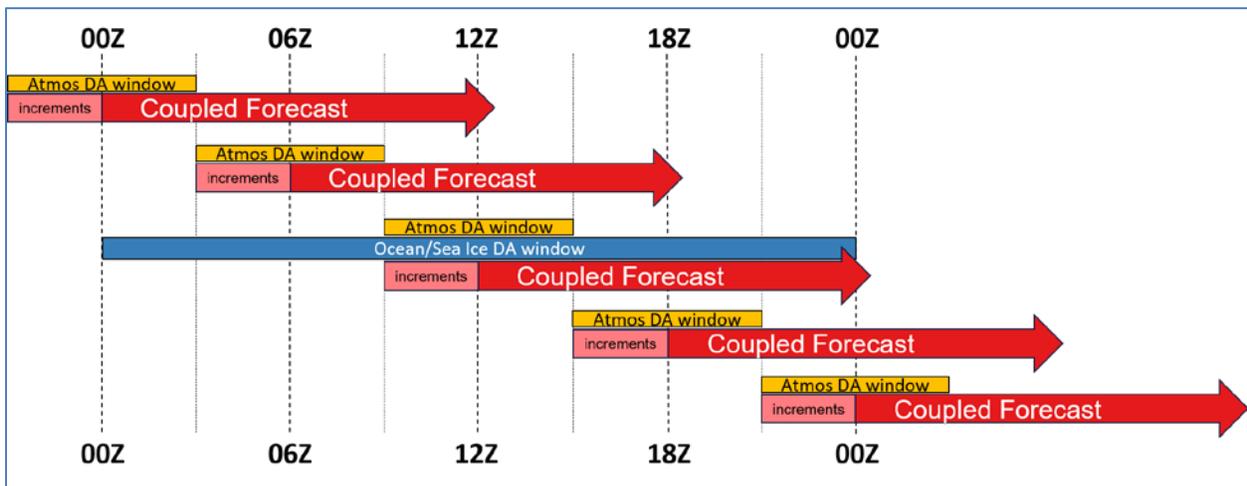
The system components are run at varying horizontal and vertical resolutions. This is described in more detail in Section 2 above, and a summary is presented in Table 2. Here the various Navy ESPC modeling components and their resolutions are shown. The deterministic system is projected to run daily with the full analysis being performed. This will have a subsequent longer forecast submitted off the 12Z cycle point, with this forecast extending to 16 days. Optimally at the other cycle points (00Z, 06Z and 18Z) an atmospheric analysis cycle will be performed. These only require a 9-hour forecast at minimum, while a 30-hour forecast allows both some failover safety and provides for additional diagnostics of system performance. For these cycles, the forecast extending past 9-hours can be removed from the critical pathway of the operational

timeline. For the 12Z cycle point the models must begin at -03 (09Z) hours to incrementally insert the corrections from the assimilation systems, and then a 36 hour forecast is required for the subsequent 12Z assimilation cycle.

Figure 3 shows how the system will cycle when running operationally, and highlights the data latency issues that prohibit the Navy ESPC system being a replacement for the standalone 6-hour-cycling NAVGEM forecasting system. In the current operational system, NAVGEM runs several hours behind the cycle time (i.e. the 00Z forecast would run at ~1 Z for the early data cutoff time, ~3Z for the regular data cutoff time, and ~8Z for the late data cutoff time). As shown in the figure, the 24-hour ocean data assimilation window means that the 12Z forecast for deterministic Navy ESPC would not start running data processing until very late in the data window or even after it's completed (e.g. anywhere from 11 to 12+ hours after the cycle time). While this latency between analysis cycle time (e.g. 12Z) and forecast completion (much later than 00Z) is consistent with current GOFs configurations, it means that the deterministic Navy ESPC forecast would be delayed more than 12 hours at a minimum, and thus is not suitable to replace the current uncoupled NAVGEM system for low-latency forecast requirements (e.g. driving regional models).

**Table 2. Horizontal and vertical resolutions of individual ESPC-IOC components at 2019 delivery.**

Forecast	Time Scale, Frequency	Atmosphere NAVGEM	Ocean HYCOM	Ice CICE	Aerosol NAAPS
ESPC-IOC Deterministic	0-16 days, daily	T681L60 (20km) 60 level	1/25° (4.5km) 41 layers	1/25° (4.5km)	3/16° (21km)



**Figure 3. A schematic of how the Navy ESPC IOC will cycle when running operationally.**

### 3.2 Input Data Streams

Each component system will have its own input and output streams. Both FNMOC-Monterey and -Stennis will be responsible for maintaining this data flow. FNMOC will collect, quality control, and deliver to the operational system atmospheric, oceanic and sea ice observations in the correct format and in a timely manner. Total volume of the data is currently about 100 GB/day. Table 3 lists a summary of the input observations used by NCODA and NAVDAS-AR.

**Table 3. List of observations used in NCODA and NAVDAS-AR. In-Situ and satellite observations used in each system are listed.**

<b>System</b>	<b>Observations</b>	
<b>NCODA</b>	<i>In-situ</i>	<ul style="list-style-type: none"> <li>a. Temperature, salinity, profile data from XBT, CTD, Argo, TAO moorings, gliders, and marine mammals</li> <li>b. Current observations from HF radar, drifting buoy</li> <li>c. Optical data from gliders, AUVs</li> <li>d. Naval Ice Center ice edge</li> </ul>
	Satellite	<ul style="list-style-type: none"> <li>a. Sea Surface Temperature (SST)</li> <li>b. Altimeter Sea Surface Height Anomaly (SSHA)</li> <li>c. Altimeter Significant Wave Height (SWH)</li> <li>d. Microwave and visible sea ice concentration</li> </ul>
<b>NAVDAS-AR</b>	<i>In-situ</i>	<ul style="list-style-type: none"> <li>a. Radiosondes; Pibals; Downsondes; Dropsondes; and Driftsonde</li> <li>b. Land and ship surface observations</li> <li>c. Aircraft observations</li> <li>d. Synthetic observations</li> </ul>
	Satellite	<ul style="list-style-type: none"> <li>a. Surface Winds <ul style="list-style-type: none"> <li>i. Scatterometer, ASCAT and ERS-2</li> <li>ii. SSMI/SSMIS</li> <li>iii. WindSat</li> </ul> </li> <li>b. Feature Tracked Winds <ul style="list-style-type: none"> <li>i. Geostationary (6 satellites)</li> <li>ii. Polar Orbiters (AVHRR and MODIS)</li> <li>iii. Combined polar/geo winds (CIMSS)</li> </ul> </li> <li>c. GNSS-RO Bending Angle <ul style="list-style-type: none"> <li>i. GRACE-A, -B; Terra and TanDEM SAR-X</li> <li>ii. COSMIC FM1-6, COSMIC 2, KOMPSAT-5</li> <li>iii. GRAS MetOp-series</li> </ul> </li> <li>d. IR Sounding Radiances <ul style="list-style-type: none"> <li>i. 3 IASI, AIRS, 2 CrIS</li> <li>ii. Geostationary Clear-Sky Radiances (6 satellites)</li> </ul> </li> <li>e. MW Sounding Radiances <ul style="list-style-type: none"> <li>i. AMSU-A; 3 SSMIS; 4 MHS; 2 ATMS</li> </ul> </li> <li>f. Ozone Retrievals <ul style="list-style-type: none"> <li>i. SBUV/2; GOME-2; OMPS-Nadir Profiler and Total Column</li> </ul> </li> <li>g. Aerosol Optical Depth <ul style="list-style-type: none"> <li>i. 2 MODIS (MOD04)</li> <li>ii. 4 AVHRR (ACSPO)</li> <li>iii. VIIRS</li> </ul> </li> <li>h. Fire Biomass</li> </ul>
<b>NAVDAS-AR (cont)</b>		

		<ul style="list-style-type: none"> <li>i. 2 GOES, Meteosat</li> <li>i. Soil Moisture <ul style="list-style-type: none"> <li>i. AMSR-2, SMOS</li> </ul> </li> <li>j. Ocean Altimeter <ul style="list-style-type: none"> <li>i. Sea Surface Height Anomaly (SSHA)</li> <li>ii. Significant Wave Height (SWH)</li> </ul> </li> </ul>
--	--	---

### 3.3 Navy ESPC-IOC System outputs

In the coupled system, NAVGEM output at T681L60 is approximately 3 GB, HYCOM output is approximately 55 GB and CICE output is 12 GB, per forecast tau. NAVGEM will have output typically written every 3 hours at the earlier times and less frequently (either 6 or 12) hours after 5 days of forecast. Since the ocean contains tidal forcing, 1-hourly output is needed to resolve these frequencies. However, the CONOPS for using global boundary conditions with tides in regional ocean systems has not yet been fully resolved so ocean output will initially be every three hours similar to GOFS 3.1. Sea ice output will be 12-hourly. Thus for the 16-day deterministic run, atmosphere and ocean/ice output on the native grid will total 7788 GB (NAVGEM = 364 GB, HYCOM = 7040 GB, CICE = 384 GB). Along with the other forecasting system mentioned above, the output is also fed to various applications such as the Optimum Path Aircraft Routing System (OPARS), the National Unified Operational Prediction Capability (NUOPC), various tactical decision aid applications such as the Target Acquisition Weapons Systems (TAWs), and many other critical applications. Effective delivery of ESPC IOC model output will need to be ensured by implemented system design.

HYCOM outputs whole domain, 3-dimensional archive files on its hybrid vertical grid. These can be a daily average, or an instantaneous snapshot at a user-defined frequency. Currently, the native grid snapshots are further interpolated in space to uniform latitude (0.02°)/longitude (0.04°) grid from 80.48°S to 90°N and in the vertical to 101 z-levels for the following variables: temperature, salinity, zonal and meridional velocity components. Sea surface height (SSH) is also included in these files. These are output in netCDF format and a single tau is approximately 32 GB. CICE output is further subdivided separately into Arctic and Antarctic domains and these total approximately 2 GB per tau.

To better describe the volumes of data output by the system the history files which describe the system state are presented in Table 4. These are the volumes produced by the system and can be used to generate fields and output products from the model for a particular forecast time.

**Table 4. Output volumes for the individual ESPC-IOC components at 2019 delivery to describe each models state per hour.**

Forecast	Time Scale, Frequency	Atmosphere NAVGEM	Ocean HYCOM	Ice CICE	Aerosol NAAPS
ESPC-IOC Deterministic	per hour	3 GB	32 GB	2 GB	1 GB

### 3.4 Navy ESPC IOC System Computational Timings

The Navy ESPC IOC system has performed robustly on the Navy DSRC systems, with the bulk of testing and the runs for this VTR performed on the Cray XC30 systems Armstrong and Shepard, and the Cray XC40 systems Conrad and Gordon. The forecast model, which runs on approximately 2800 cores, is the most computationally expensive task by far. The next most expensive tasks are the solvers of the data assimilation components NAVDAS-AR for the atmosphere, and NCODA for the sea-ice and ocean. Figure 4 and Figure 5 show various tasks and the core hours used. The atmospheric data assimilation task for NAVDAS-AR is called “4DVARsweep” and the NCODA tasks labeled Analysis1 and Analysis2 represent different basins. For some tasks, especially within the ocean data assimilation, the increased resolution relative to the ensemble configuration adds additional computational requirements. However, the consistent resolution for the atmospheric data assimilation system leads to similar computational requirements for NAVDAS-AR solver independent of the model resolution.

As noted above, the IOC is ultimately targeted at the newer HPE SGI 8600 supercomputers at Navy DSRC. At this time, many task timings between the CrayXC40 and SGI 8600 are comparable, but investigation into more optimization possibilities is in progress. As migration to the SGI system continues, a full assessment of computational characteristics specific to the operational system is ongoing.

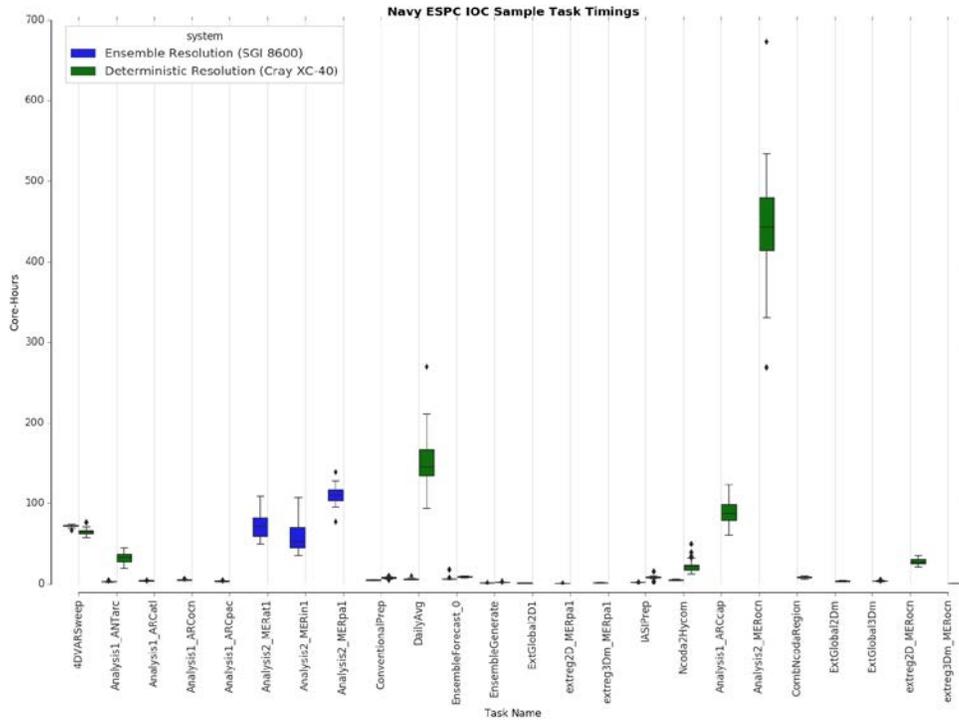


Figure 4. Core hours used for representative non-forecast tasks for the lower ensemble resolution (blue boxes, run on Gaffney, the Navy DSRC SGI 8600) and the higher deterministic resolution (green boxes, run on Gordon, a Navy DSRC Cray XC-40).

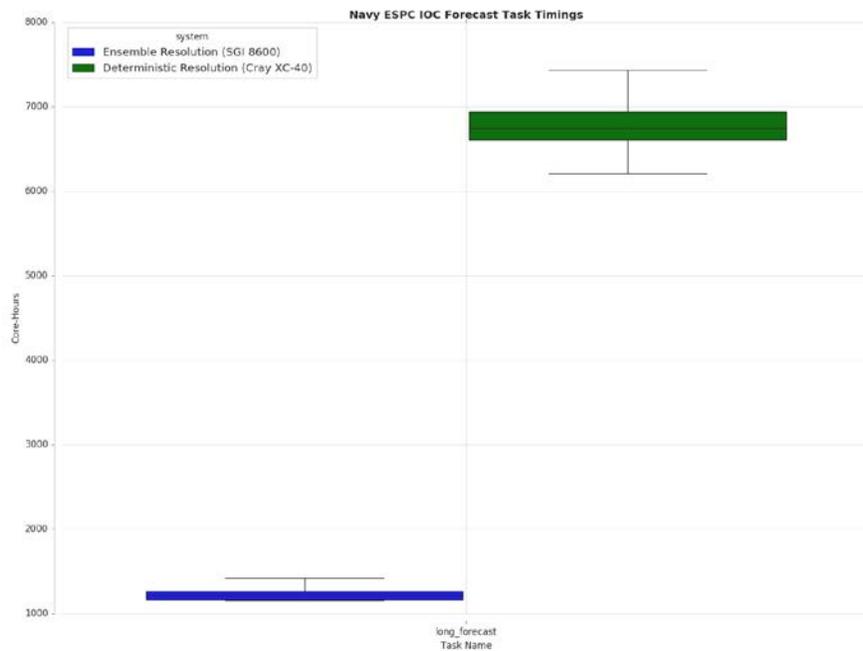


Figure 5. Same as Figure 4, but with the forecast task isolated because of the order-of-magnitude difference between forecast and non-forecast compute usage.

The computational requirements for wallclock time provide a barrier to adoption of Navy ESPC as a replacement for NAVGEM, but would permit replacement of GOFS. Typical operational pace for the global atmospheric model is O(7-8) minutes of wallclock time per forecast day (a number consistent with many other atmospheric modeling centers). The Navy ESPC deterministic system timings are O(nearly 1 hour) per forecast day – significantly longer than NAVGEM, but similar order to the GOFS system. To execute the coupled model in 7-8 minutes/forecast day would require *significantly* more resources in order to reduce the runtime. In addition to the increased resources for the ocean component (the primary driver of computational requirements), that tempo would require many more resources for the atmospheric component (in the current configuration, the atmospheric component uses very few cores relative to the ocean due to the relaxed pace of the integration). Section 3.1 outlines additional challenges for data assimilation-based latency even if the 7-8 minutes/forecast day benchmark could be attained.

## 4 Navy ESPC-IOC System Performance

The Navy ESPC-IOC deterministic system has been exercised over some shorter case studies and over an extended period beginning in December 2016 and is presently running in near real time. These experiments starting in December 2016 were initialized from GOFS 3.5 for the ocean and sea-ice and FNMOC operational NAVGEM v1.4 for the atmosphere. The period used for the bulk of the performance verification is calendar year 2017. The first month, December 2016, is excluded to allow the various system components to spin-up. A series of 7-day forecasts were run on the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, 16<sup>th</sup>, 21<sup>st</sup> and 26<sup>th</sup> of each month to examine skill as a function of forecast length. In addition, 16-day forecasts were performed once per week to enable comparison with the Navy ESPC ensembles and these forecasts are evaluated in section 4.3 for the atmosphere.

Overall, the deterministic Navy ESPC-IOC system has shown robust performance for each component (atmosphere, ocean, sea ice) comparable to performance of the stand-alone operational systems. In particular, the performance at longer forecast lead times, which the system was designed for, is encouraging and the performance is competitive with other forecast centers (ECMWF, NOAA/EMC), placing the Navy predictive capability near the forefront of extended range forecasting.

### 4.1 Ocean Performance

For the ocean, all models are evaluated against unassimilated observations. The baseline system for comparison is the stand-alone Global Ocean Forecast System (GOFS) 3.5, which has the same horizontal and vertical resolution as Navy ESPC IOC Deterministic System. GOFS 3.5 also includes the forcing and physics for astronomical surface tides and baroclinic tides. The current operational ocean and sea ice only system is GOFS 3.1, which has lower horizontal resolution (9 km) and doesn't include the tidal physics. Thus, Navy ESPC ocean performance will be compared to both the operational GOFS 3.1 and the pre-operational GOFS 3.5. The pre-operational GOFS 3.5 is undergoing test and many of the evaluations in this report will be similar in the GOFS 3.5

VTR when it is completed. The GOFs forecasts extend to 7 days using atmospheric forcing from operational NAVGEM forecasts. For this VTR, a shorter Navy ESPC 7-day forecast is launched every 5<sup>th</sup> day in the month, 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, 16<sup>th</sup>, 21<sup>st</sup>, and 26<sup>th</sup>, to correspond with similar 7-day forecasts in GOFs 3.5 and GOFs 3.1. Beyond 7 days, no comparisons exist. Thus, the extended range forecasts, days 7-16, will be compared to the latest ocean climatology, (Generalized Digital Environmental Model, GDEM4, Carnes 2010) ) and persistence of the model analysis. The 16-day forecasts are run once a week on Wednesday to overlap with Navy ESPC ensemble forecasts, but are not compared to the ensembles from an oceanographic perspective in this VTR.

The ocean performance is evaluated relative to standard scorecard metrics in eight geographical regions shown in Figure 6. The metrics are: bias and RMSE of temperature and salinity over depth; bias, mean absolute error and vector correlation of the 15 m velocity; bias and RMSE of mixed layer depth (MLD); and acoustic proxies (sonic layer depth (SLD) and below layer gradient (BLG)), which have been used in previous ocean VTRs. In addition, two new scorecard metrics, bias and RMSE of the depth of the 15°C, 20°C and 26°C isotherms, and skill in frontal placements are evaluated in the global region. Barotropic and baroclinic tides are new features of Navy ESPC and GOFs 3.5. The barotropic tidal elevation amplitudes and phase for the principal components will be evaluated relative to the inverse tidal solution TPX08 (Egbert and Erofeeva, 2002). The first vertical mode stationary baroclinic semi-diurnal tidal elevation will be evaluated relative to satellite altimeter SSH estimates.

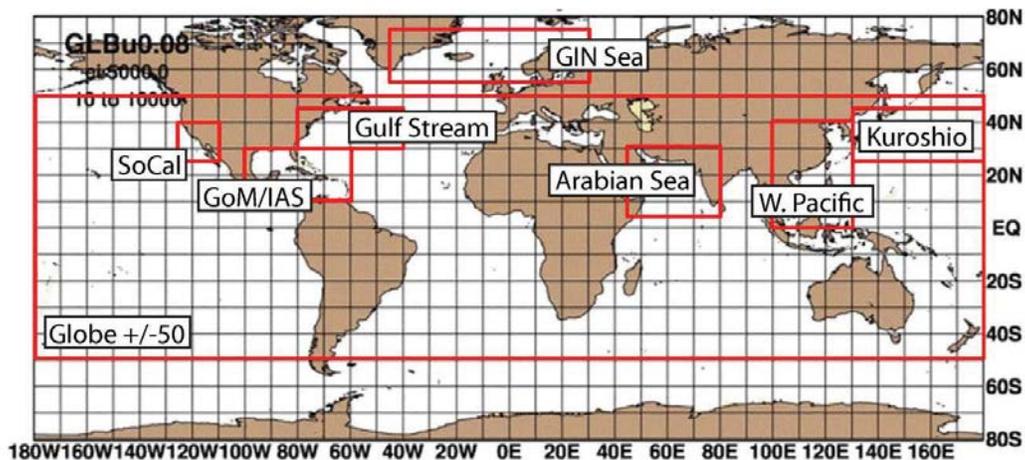


Figure 6. For the standard ocean scorecard metrics, eight regions are used in this report: Globe: 180°W-180°E, 50°S-50°N; West Pacific marginal seas: 100-130°E, 0-40°N; Kuroshio Extension: 130°E-180°, 25-45°N; Arabian Sea/Persian Gulf/Gulf of Aden: 44-80°E, 5-30°N; Gulf Stream: 80-40°W, 30-45°N; Southern California (SoCal): 125-110°W, 25-40°N; Greenland/Iceland/Norwegian (GIN) Seas: 45°W-30°E, 55-75°N; and Gulf of Mexico (GoM)/Intra-Americas Sea (IAS): 100-60°W, 10-30°N.

#### 4.1.1 Temperature and Salinity vs. Depth Error Analysis

An error analysis of temperature and salinity vs depth for the top 500 m of the ocean is performed using unassimilated profiles. For a given observation, the model outputs are sampled at the

nearest model grid point. In the vertical, both the observations and the model outputs are linearly remapped to a common set of depth levels: [0, 2, 4, 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 200, 250, 300, 350, 400 and 500 m]. Model-data differences exceeding three standard deviations (99% confidence level) are excluded from the analysis. Thus, the number of points in the analysis may differ between the models. The error analyses are performed at the “nowcast” time, i.e. tau 000. The NCODA analysis is valid at 12Z on each day with a data window that reaches forward 12 hours. To make sure that only unassimilated profiles are used in the error analysis, profiles from the end of the data window at 00Z to 00Z+1 day are compared to the HYCOM archive file at tau 000. Verification for the forecasts are centered at tau 024, 048, etc. Profiles +/-1.5 hours are compared to the 3-hourly forecast archive files within each 24-hour period.

The mean error (ME) or bias and the root mean square error (RMSE) for the temperature at the “nowcast” time (tau 000) for the eight analysis regions defined in Figure 6 for Navy ESPC, pre-operational GOFS 3.5 and operational GOFS 3.1 are shown in Figure 7. Both systems show a small cold bias at all depths and all regions, except very near the surface in Southern California. Generally, the ME is less than 0.25°C in both systems and smaller than the FNMOC-defined tolerance of 0.5°C for the upper ocean. The largest biases occur in the western boundary current regions of the Gulf Stream and Kuroshio Extension. GOFS 3.5 generally has slightly smaller bias than Navy ESPC and GOFS 3.1, with a few exceptions. The largest difference in bias occurs in the Gulf Stream region, where both GOFS 3.5 and Navy ESPC have smaller biases than GOFS 3.1. For the RMSE, the differences between Navy ESPC and GOFS 3.5 are very small. GOFS 3.5 has slightly smaller RMSE at all depths for every region compared to Navy ESPC. The biggest RMSE difference occurs between 100-200 m (thermocline depth range) in the Gulf Stream region. Both Navy ESPC and GOFS 3.5 have smaller RMSE than the operational lower resolution GOFS 3.1 except in the Western Pacific where GOFS 3.1 has smaller RMSE below 150 m. The largest difference between the operational system, GOFS 3.1, and the pre-operational Navy ESPC and GOFS 3.5 occurs in the western boundary current Gulf Stream and Kuroshio regions over the entire 8-500 m water column.

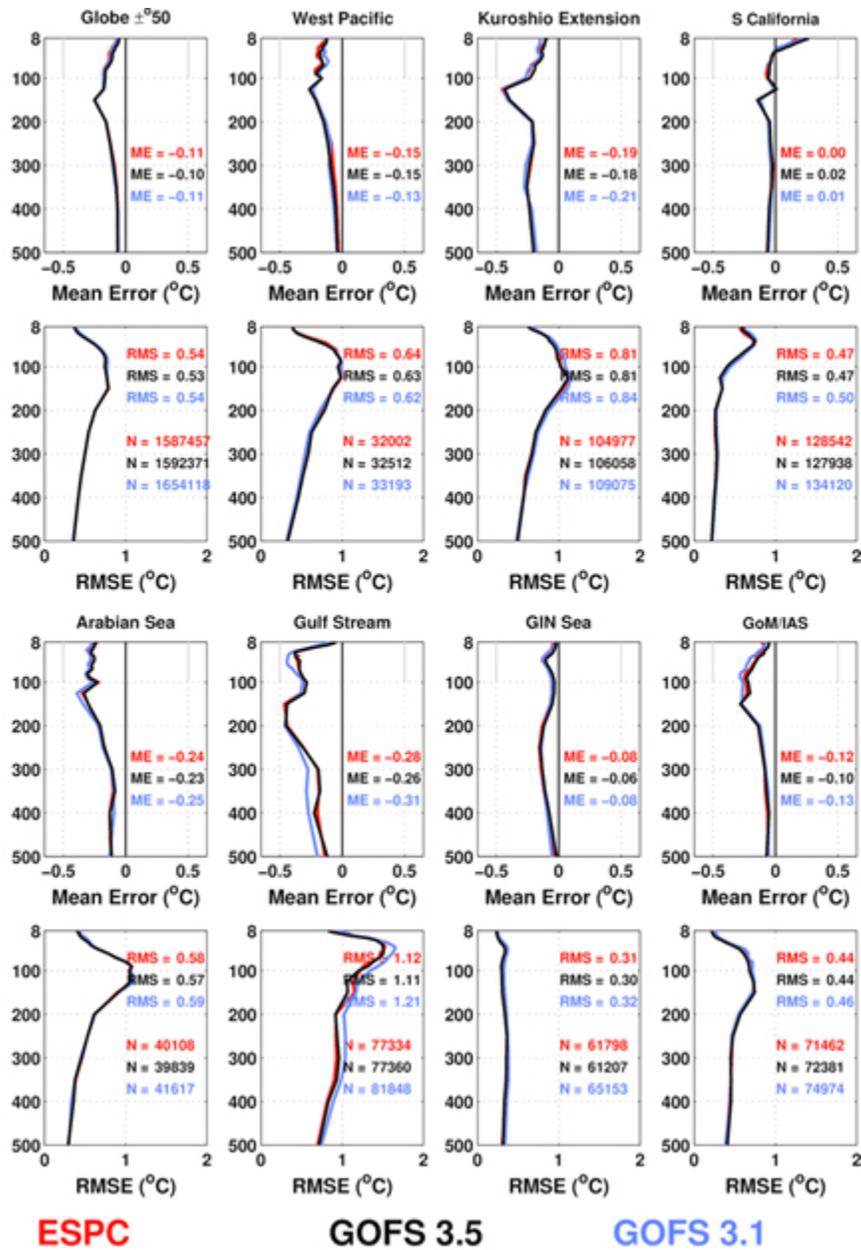


Figure 7. Temperature (°C) versus depth error analysis in the upper 500 m against unassimilated profile observations at the “nowcast” time for the eight regions defined in Figure 6. The first and third rows are the mean error (ME) (model minus observations) and the second and fourth rows are the root mean square error (RMSE). The red curves are the Navy ESPC analyses, the black curves are the GOFS 3.5 pre-operational ocean-only analyses, and the blue curves are the GOFS 3.1 operational ocean-only analyses. The numerical values for ME/RMSE are the averages over 8-500 m. The number of observations is the N=xxxxx values. The number of profiles is approximately N/20. A minimum depth of 8 m is used because that is the typical shallowest depth sample by Argo profiles.

The profiles of upper-ocean mean error (ME) or bias and the root mean square error (RMSE) for temperature at forecast days 2, 4 and 6 from the 7-day forecasts for Navy EPSC (red), pre-operational GOFS 3.5 (black) and operational GOFS 3.1 (blue) for the eight analysis regions are shown in Figure 8, Figure 9, Figure 10 and Figure 11. The trends found at the nowcast time

generally continue throughout the 7-day forecasts. All regions have a cold bias at all depths, except near the surface in Southern California. For the globe, Navy ESPC has the smallest bias near the surface, which decreases as the forecast progresses, while below 200 m GOFS 3.5 has the smallest bias. For most regions, the near-surface bias decreases in Navy ESPC compared to the ocean-only forecasts, which leads to slightly lower 8-500m average biases for Navy ESPC as the forecast proceeds. However, in the western boundary current regions, Gulf Stream and Kuroshio, Navy ESPC tends to have the largest bias between 100 m and 200 m. The biases for all systems are small, except in the western boundary current regions between 100 m and 200 m.

The error growth over the forecast is similar for all three systems. Near the surface, Navy ESPC has smaller RMSE and lower error growth. However, the differences between the systems are small. GOFS 3.5 tends to have smaller RMSE across all regions and all forecast days, except near the surface. The biggest differences between the three systems occur in the western boundary current regions between 50 m and 200 m. However, on a given forecast day, any of the three systems may have the best performance at a particular depth in a particular region.

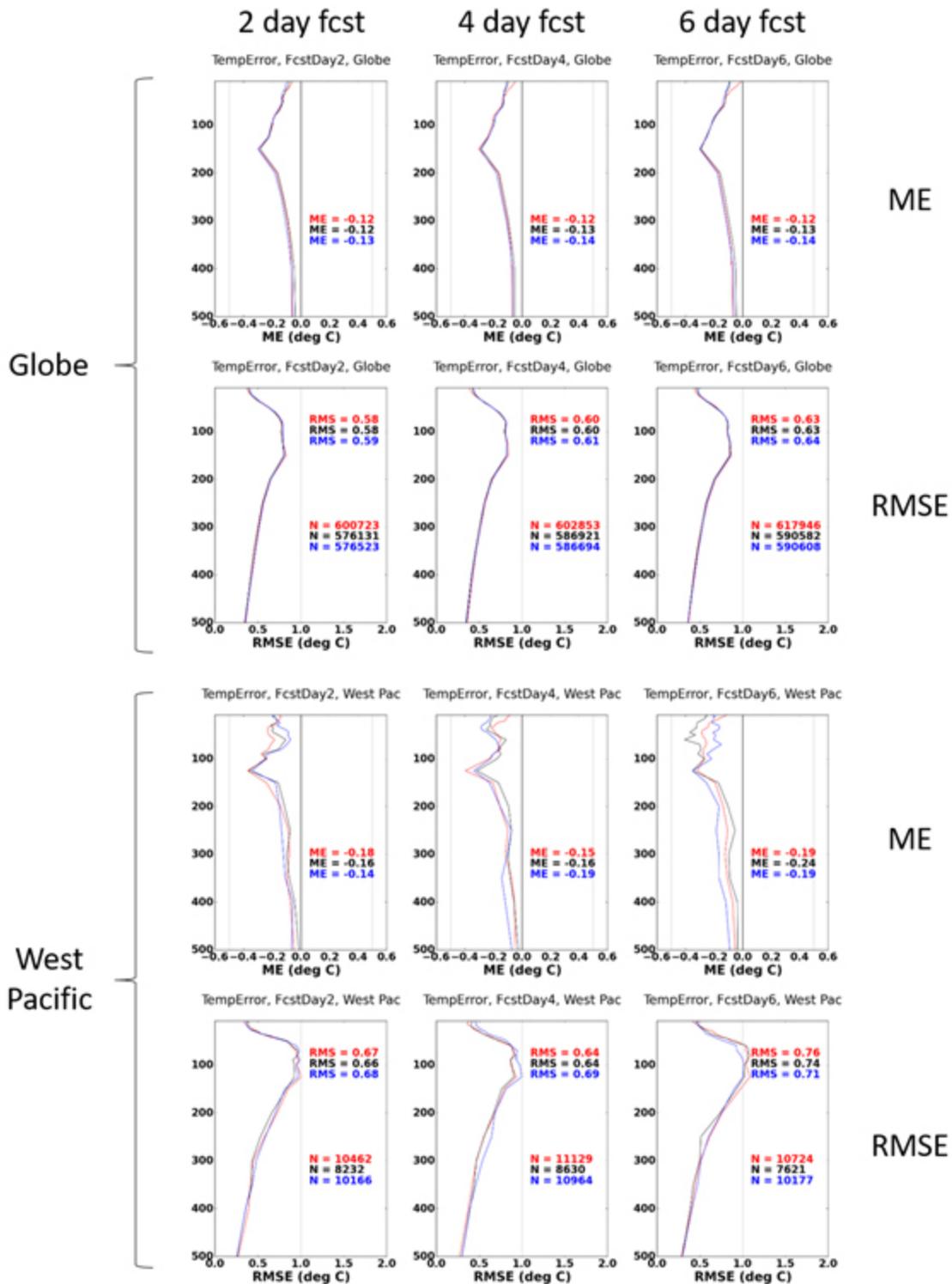


Figure 8: Temperature (°C) versus depth error analysis in the upper 500 m evaluated against unassimilated profile observations at forecast days 2, 4 and 6 for the global and Western Pacific regions. The first and third rows are the mean error (ME) and the second and fourth rows are the root mean square error (RMSE). The first column is forecast day 2, the second column is forecast day 4 and the third column is forecast day 6. The upper six plots are for the Globe, and the lower six plots are for the Western Pacific. The red curves are the Navy ESPC analyses, the

black curves are the GOFS 3.5 pre-operational ocean-only analyses and the blue curves are the operational GOFS 3.1. The numerical values for ME/RMSE are the averages over 8-500 m. The number of observations is the N=xxxxx values. The number of profiles is approximately N/20. A minimum depth of 8 m is used, because that is the typical shallowest depth sample by Argo profiles.

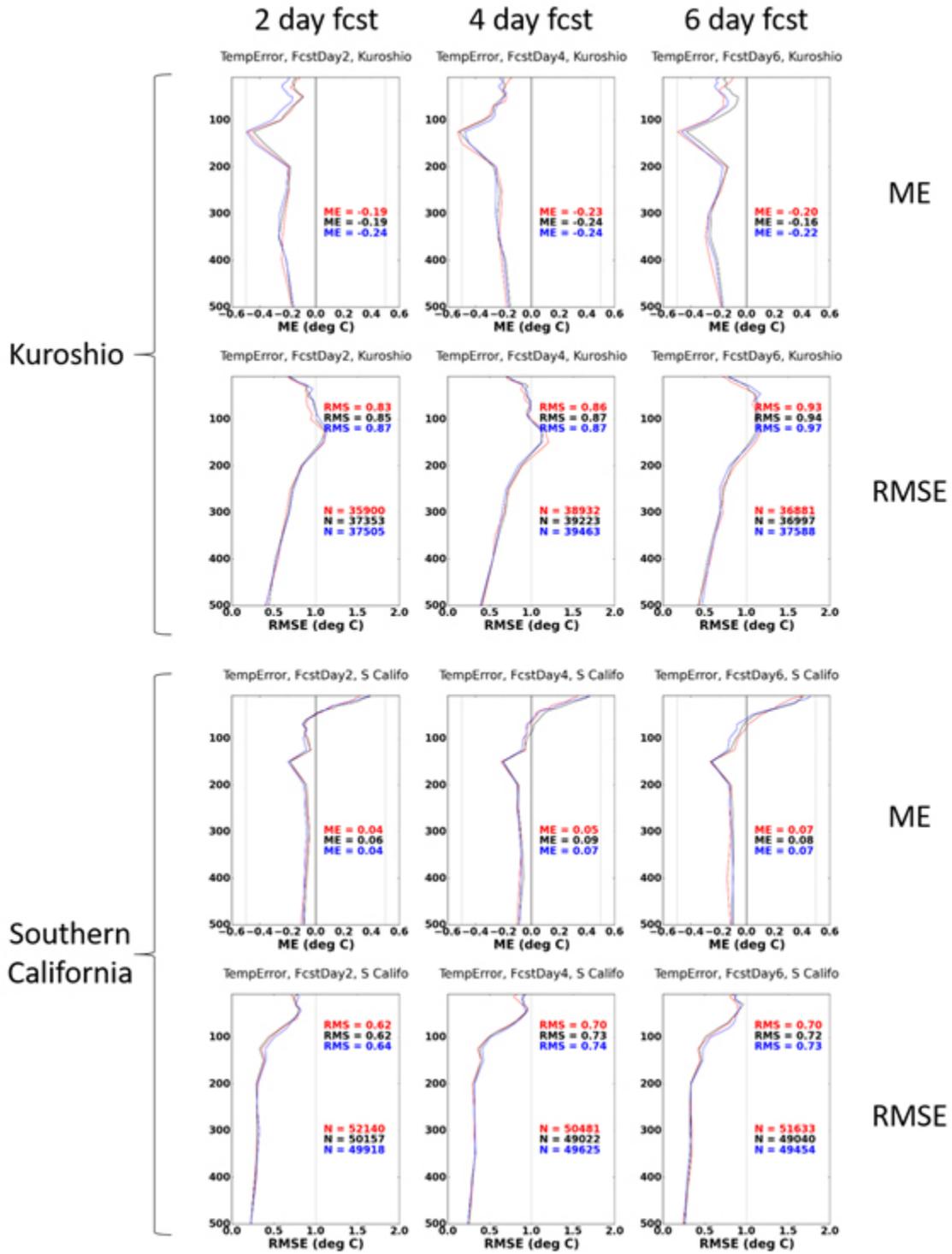


Figure 9: As Figure 8 except for the Kuroshio and Southern California regions, respectively.

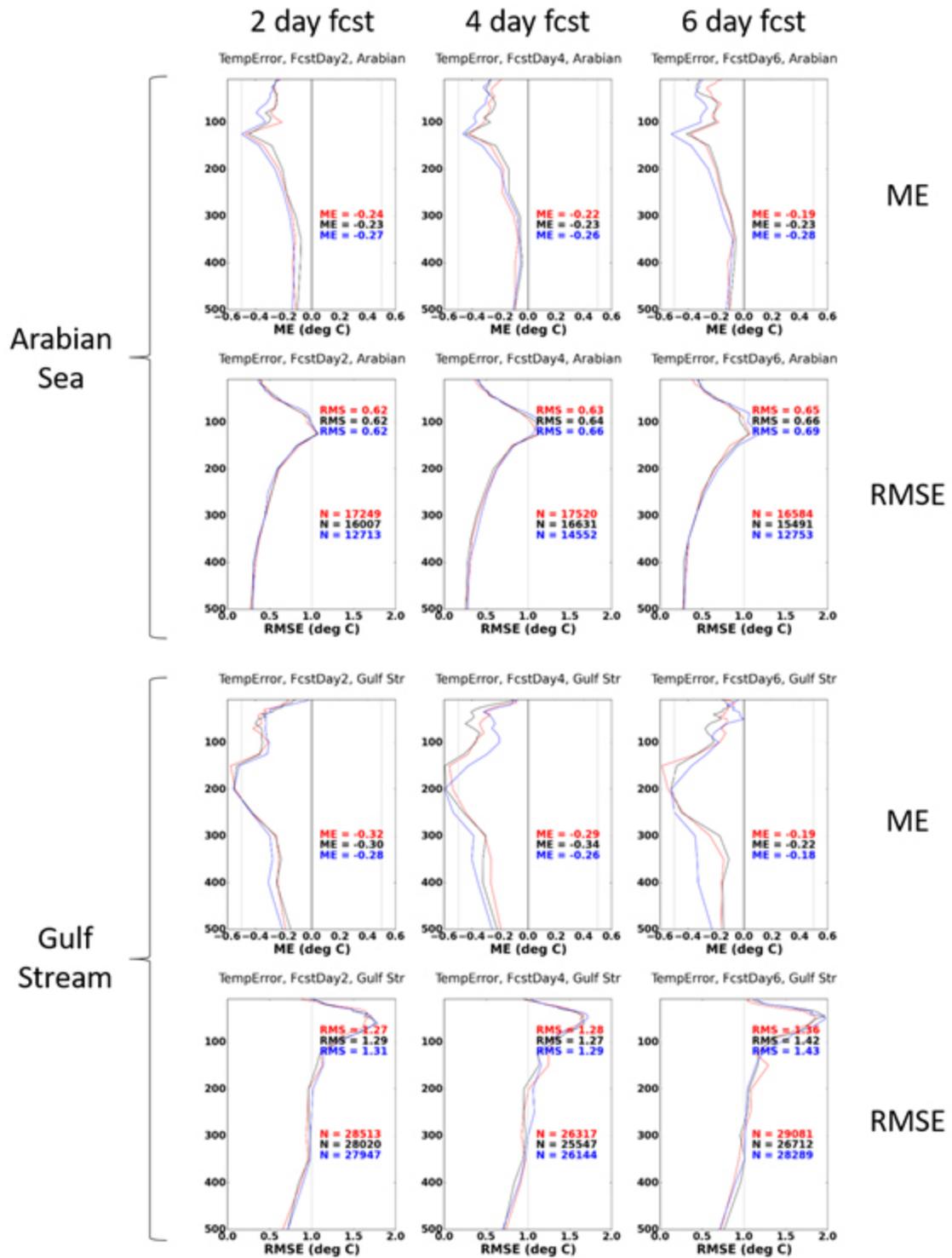


Figure 10: As Figure 8 except for the Arabian Sea and Gulf Stream regions, respectively.

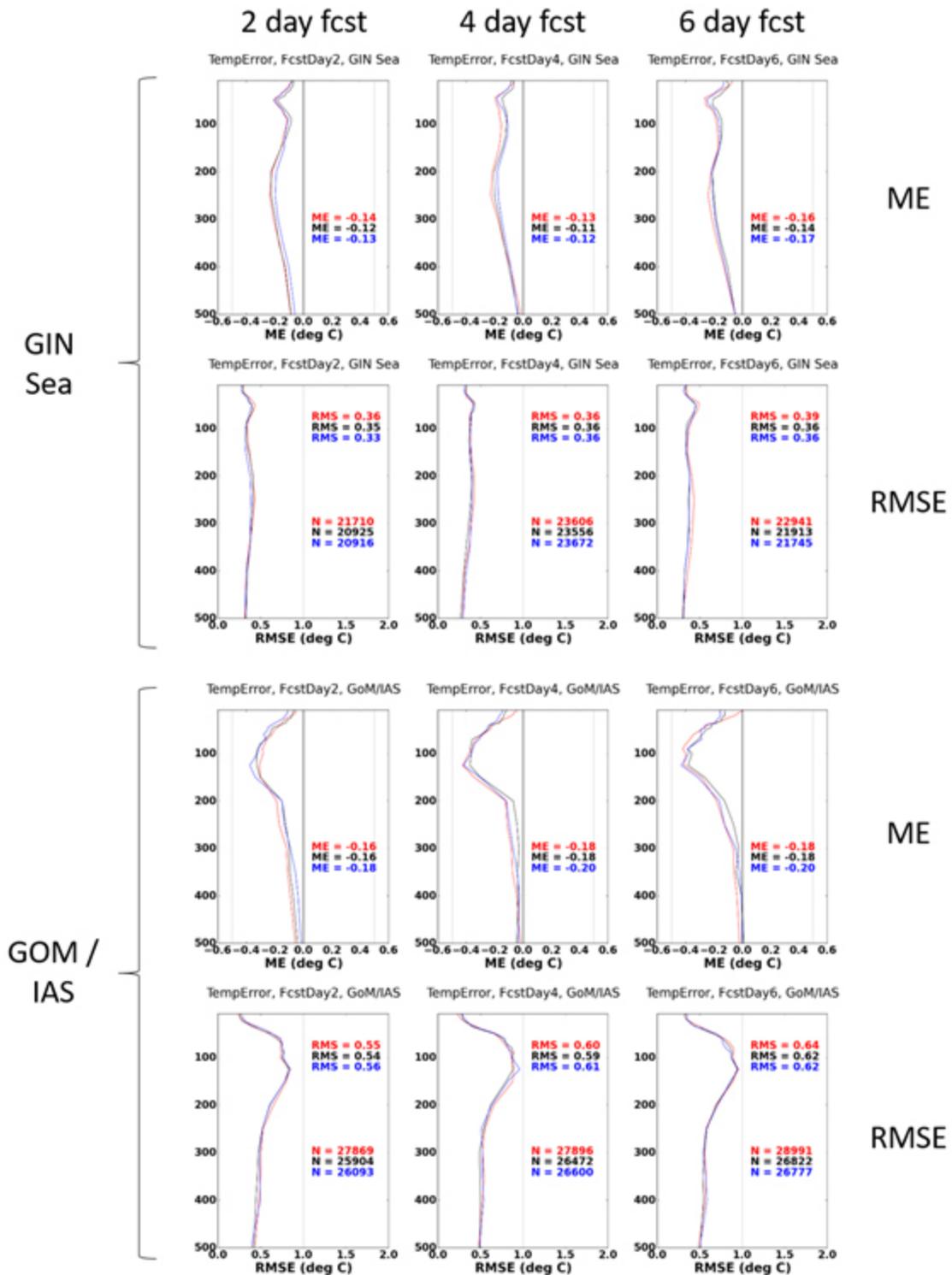


Figure 11: As Figure 8 except for the GIN Sea and GOM/IAS regions, respectively.

The global upper ocean temperature mean error (ME) or bias as a function of forecast length is shown in Figure 12 and Figure 13. For each region, the bias is calculated over 4 depth

ranges, 0-50m, 50-150m, 150-500m and 8-500m. For the globe, initially all systems have a cold bias at all depths. GOFs 3.1 and GOFs 3.5, which have a surface temperature bias correction, have small cooling trend over the upper 50 m of about  $9 \text{ W/m}^2$ . Navy ESPC have a warming trend over the upper 50 m of  $14.4 \text{ W/m}^2$  and has a small warm bias at the end of the 16-day forecast. At deeper depths, all systems have a small cooling trend near the surface of about  $4 \text{ W/m}^2$ . Similar results are found in the other regions with the magnitude of the trends within a factor of 2, except in Southern California, where all systems have a warm bias and warming trends of  $15\text{-}30 \text{ W/m}^2$ , in the Gulf Stream region, where all systems have a small warming trend near the surface of about  $10 \text{ W/m}^2$ , in the GIN Seas, where all systems have a small cooling trend at all depths of about  $10 \text{ W/m}^2$  and in the GOM/IAS, where all systems have a small warming trend below 150 m.

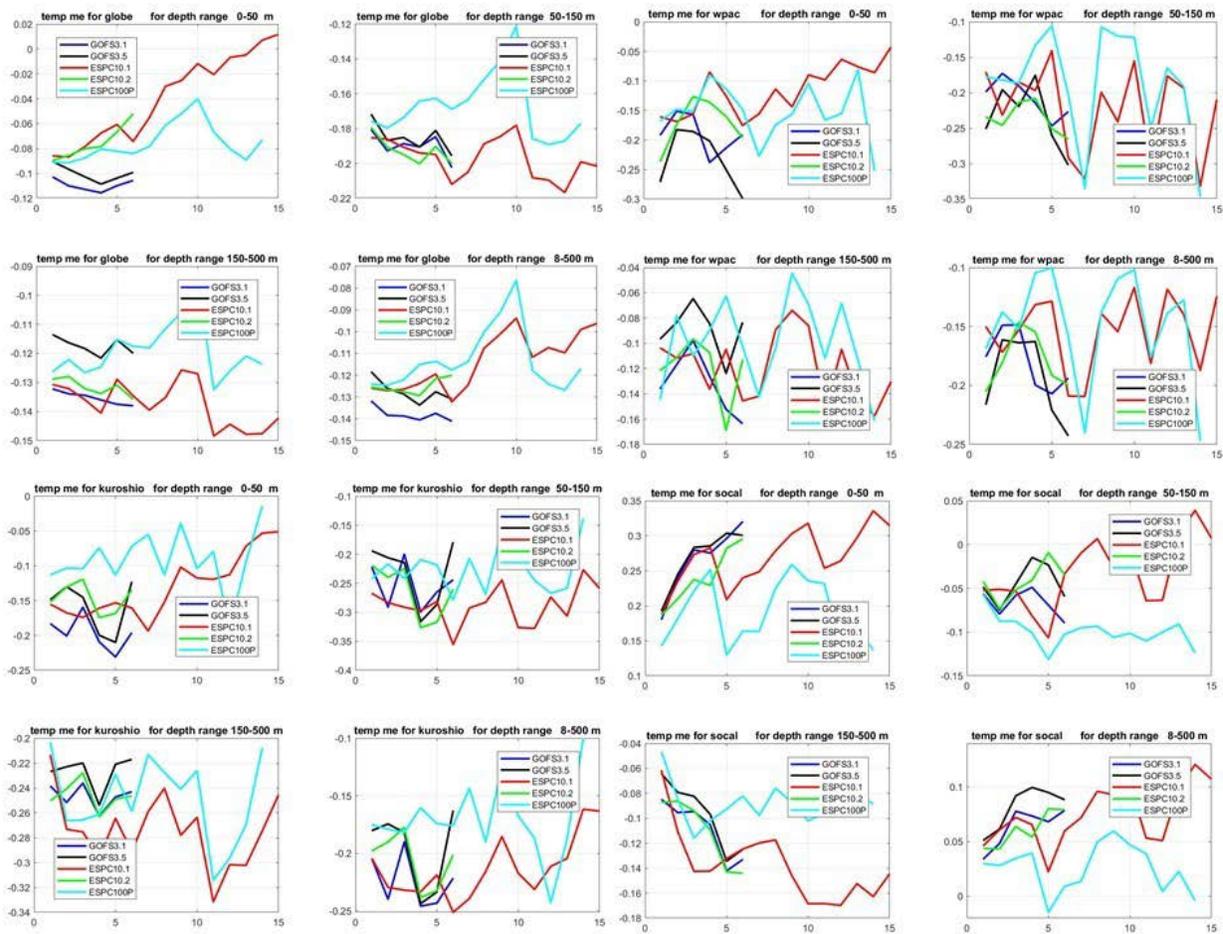


Figure 12: Temperature ( $^{\circ}\text{C}$ ) mean error (ME) or bias analysis in the upper 500 m evaluated against unassimilated profile observations as a function of forecast length for 4 regions over the globe and Pacific Ocean (Globe, Western Pacific, Kuroshio Extension and Southern California). The analyses for a region are grouped over four depth ranges, 0-50m, 50-150m, 150-500m and 8-500m. Thus, the first and third rows are the 0-50m and 50-150m depths and the second and fourth rows are the 150-500m and 8-500m depths. The upper left four plots are for the Globe, the upper right four plots are for the Western Pacific, lower left four plots are for the Kuroshio Extension and the lower right four plots are for Southern California. The red curves are the Navy ESPC 16-day forecast, the green curves are the Navy ESPC 7-day forecasts, the black curves are the GOFs 3.5 pre-operational ocean-only 7-day

forecasts and the blue curves are the operational GOFS 3.1 ocean-only 7-day forecasts. The cyan curves are the persistence of the Navy ESPC nowcast analysis over the 16 days of the forecast.

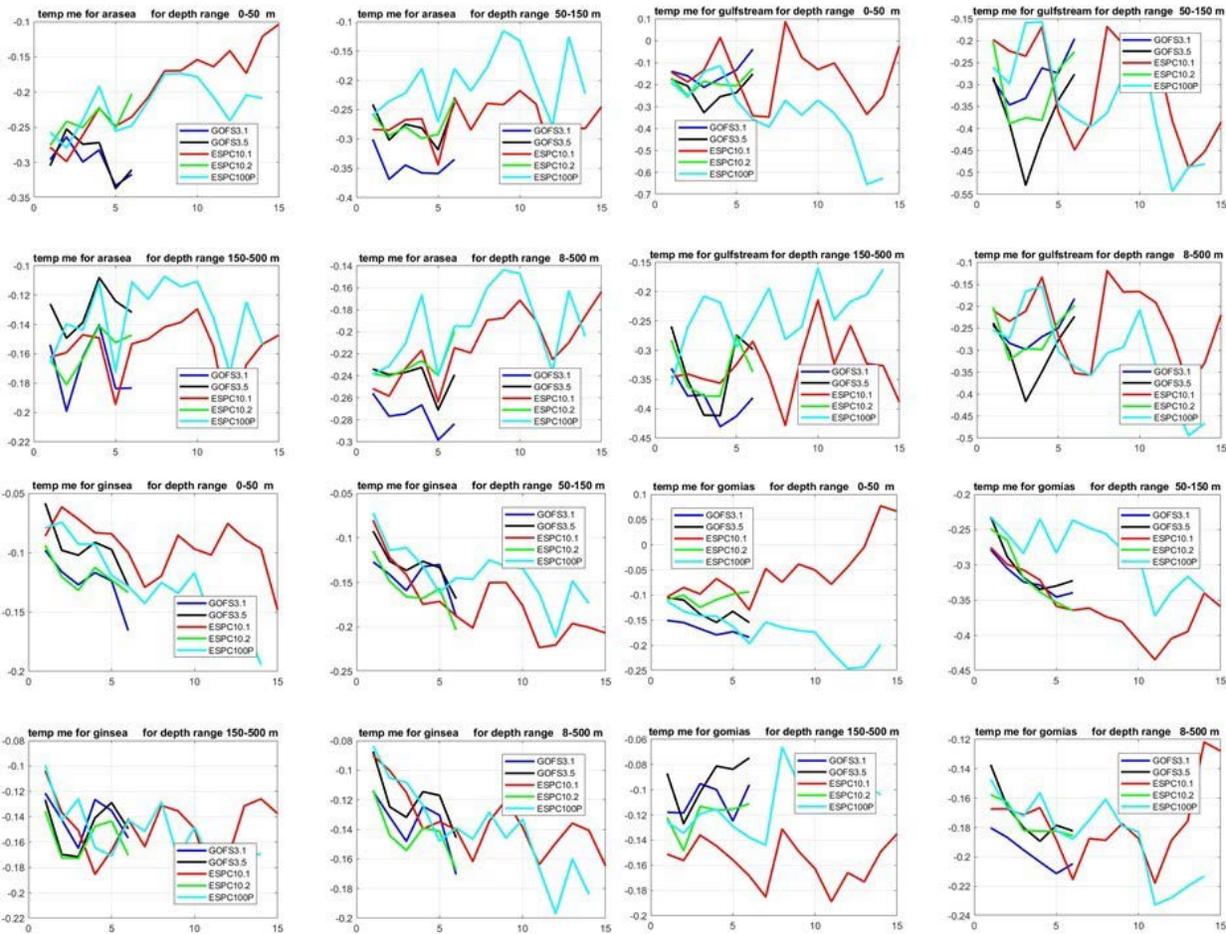
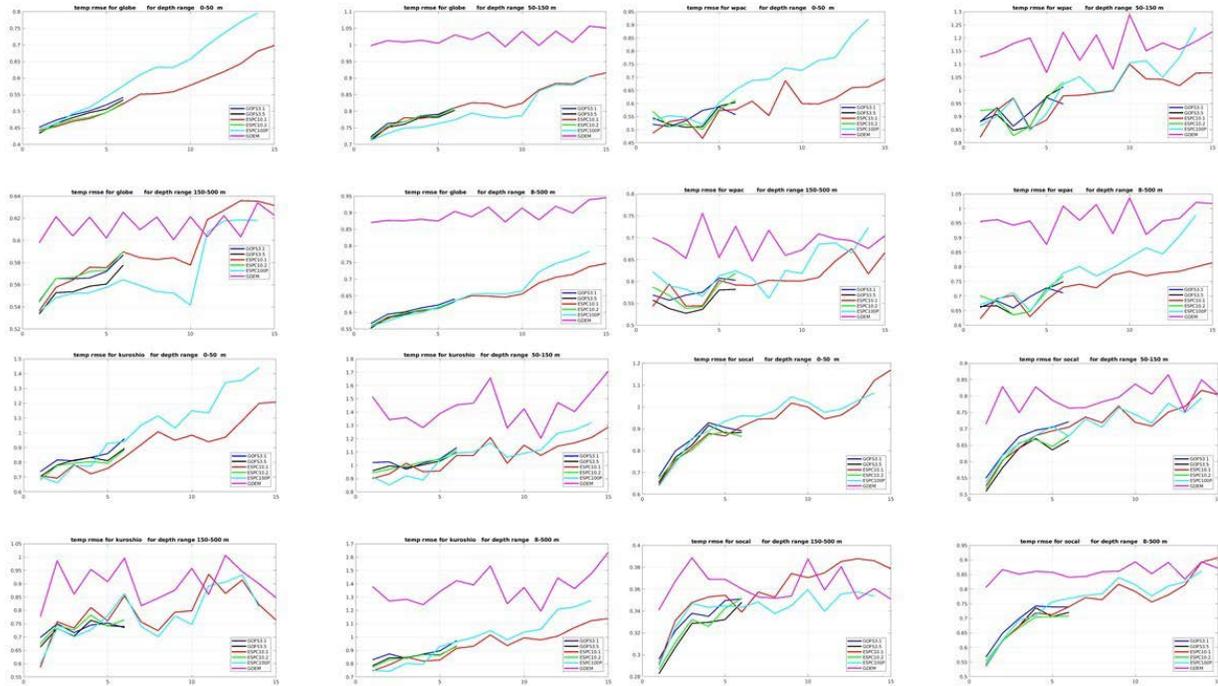


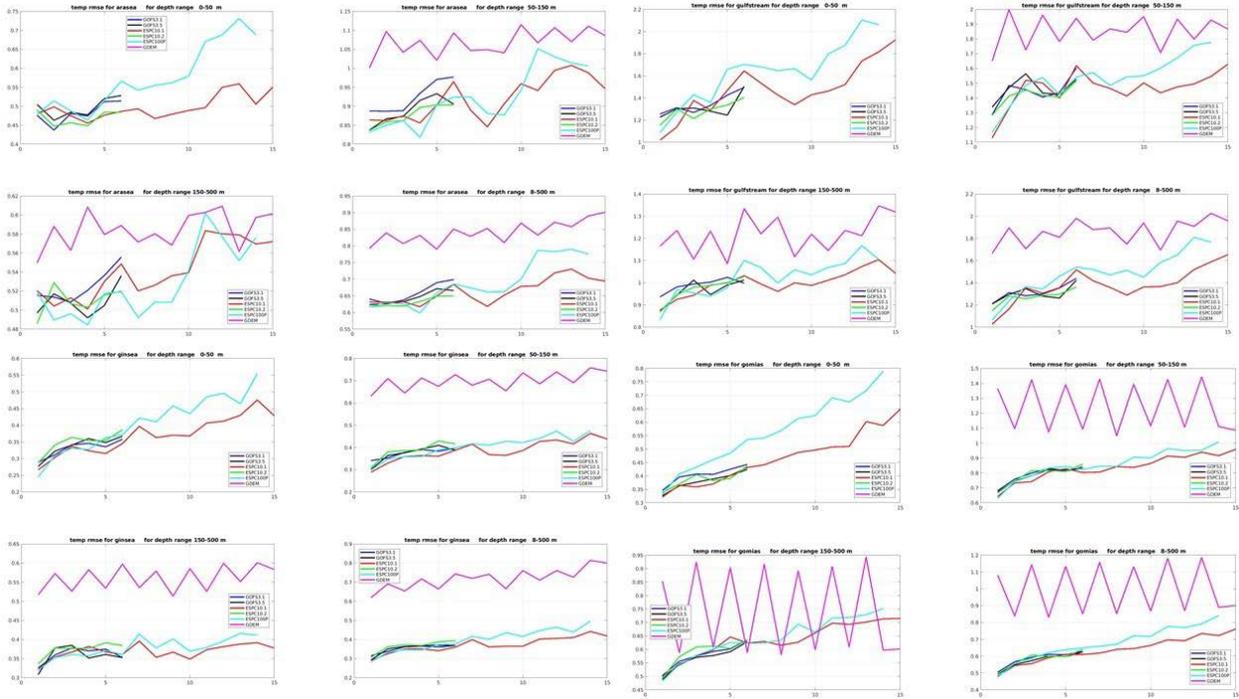
Figure 13: As Figure 12 except the upper left four plots are for the Arabian Sea, the upper right four plots are for the Gulf Stream, lower left four plots are for the GIN Seas and the lower right four plots are for GOM/IAS.

The upper ocean temperature RMSE as a function of forecast length is shown in Figure 14 and Figure 15. For each region, the RMSE is calculated over 4 depth ranges, 0-50m, 50-150m, 150-500m and 8-500m. For the globe at all depths, the trend in RMSE for all models is similar over the forecasts with approximately 16% growth in the error over 6 days and 32% growth in the error for the 16-day extended range forecast. GOFS 3.5 and Navy ESPC have the smallest and nearly identical global RMSE for the 7-day forecasts with pre-operational GOFS 3.1 having the largest, although the difference in RMSE is less than 8%. Navy ESPC performs better than persistence over the entire 16-day forecast. For the globe below 50 m, all models perform better than climatology over the entire forecast length. Below 150 m, GOFS 3.5 has the smallest RMSE. The estimates for the 16-day Navy ESPC, climatology and persistence are calculated from the extended forecast sampling which has fewer profiles and a different spatial sampling pattern than the 7-day forecasts. In general, the ocean temperature profiles are under-observed, which leads to noisy RMSE estimates. The relative performance for the other analysis regions is similar.

The RMSE is largest in the western boundary regions (Gulf Stream and Kuroshio Extension) with error growth of ~50% over the extend range forecast. The error growth is smaller in the Arabian Sea, Western Pacific and Greenland Iceland and Norwegian Seas with ~20% growth over the extended range forecast.



**Figure 14: Temperature (°C) root mean square error (RMSE) analysis in the upper 500 m evaluated against unassimilated profile observations as a function of forecast length for 4 regions over the globe and Pacific Ocean (Globe, Western Pacific, Kuroshio Extension and Southern California). The analyses for a region are grouped over four depth ranges, 0-50m, 50-150m, 150-500m and 8-500m. Thus, the first and third rows are the 0-50m and 50-150m depths and the second and fourth rows are the 150-500m and 8-500m depths. The upper left four plots are for the Globe, the upper right four plots are for the Western Pacific, lower left four plots are for the Kuroshio Extension and the lower right four plots are for Southern California. The red curves are the Navy ESPC 16-day forecast, the green curves are the Navy ESPC 7-day forecasts, the black curves are the GOFS 3.5 pre-operational ocean-only 7-day forecasts and the blue curves are the operational GOFS 3.1 ocean-only 7-day forecasts. The cyan curves are the persistence of the Navy ESPC nowcast analysis over the 16 days of the forecast. The magenta curve is the comparison to climatology (GDEM4).**



**Figure 15:** As Figure 14 except for the 4 regions in the Indian and Atlantic Oceans (Arabian Sea, Gulf Stream, Greenland/Iceland/Norwegian (GIN) Seas and Gulf of Mexico/Intra-Americas Sea(GOM/IAS)). The upper left four plots are for the Arabian Sea, the upper right four plots are for the Gulf Stream, lower left four plots are for the GIN Seas and the lower right four plots are for GOM/IAS.

The upper ocean salinity error versus depth for the eight analysis regions is shown in Figure 16. Globally, all systems have a small fresh bias. The salinity bias is small for all regions with a larger error near the surface and in the western boundary current regions. The Western Pacific has a salty bias in the upper 50 m, while the GIN Seas and Southern California have a salty bias over the upper 500 m.

Navy ESPC and GOM 3.5 have nearly identical RMSE for all regions, except near the surface in the Gulf Stream and GOM/IAS, where Navy ESPC has a slightly smaller RMSE. Both higher resolution systems perform slightly better than the lower resolution, operational GOM 3.1. The largest salinity RMSE occurs in the Gulf Stream region.

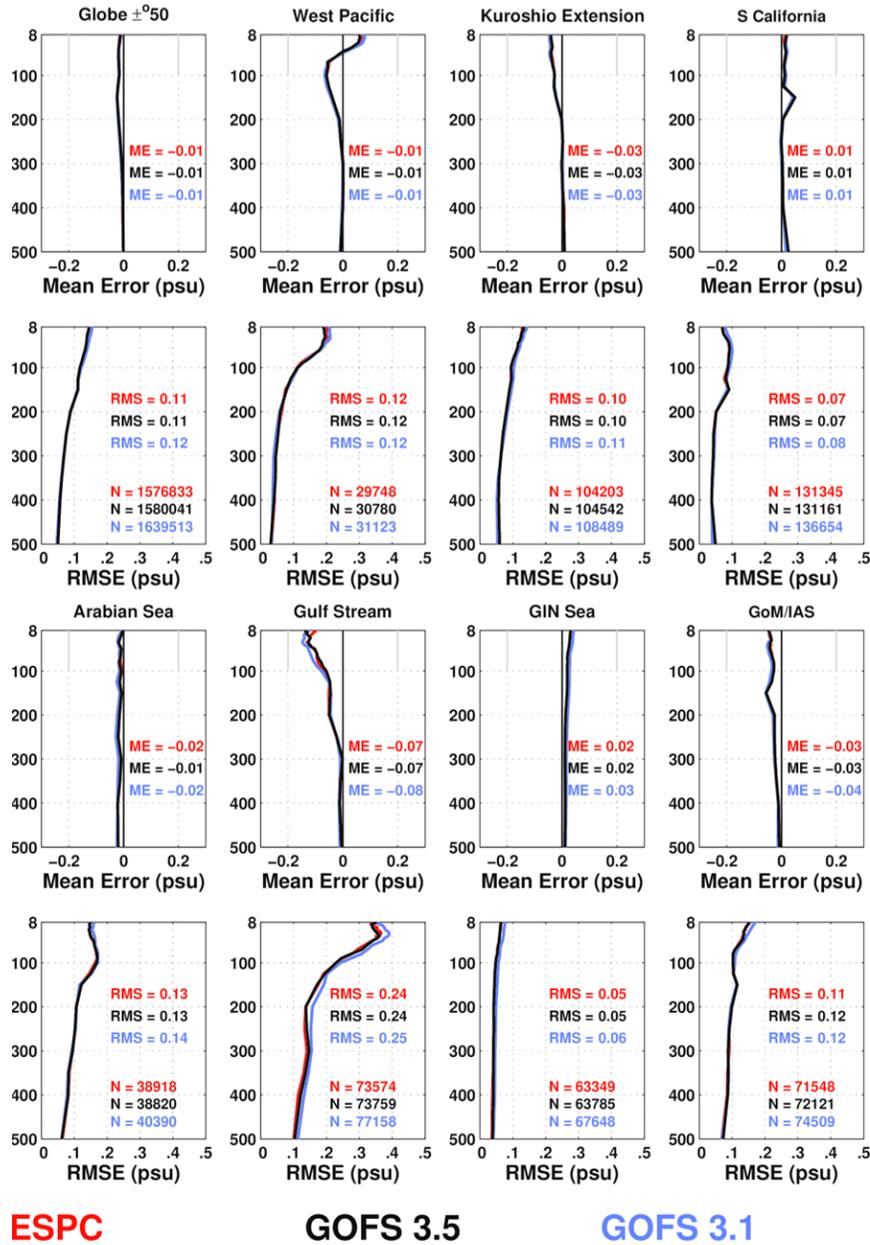


Figure 16. As Figure 7 except for salinity (psu).

The profiles of upper-ocean mean error (ME) or bias and the root mean square error (RMSE) for salinity at forecast days 2, 4 and 6 from the 7-day forecasts for Navy ESPC (red), pre-operational GOFS 3.5 (black) and operational GOFS 3.1 (blue) for the eight analysis regions are shown in Figure 17, Figure 18, Figure 19 and Figure 20. The trends found at the nowcast time generally continue throughout the 7-day forecasts. Globally, all systems have a small fresh bias. The salinity bias is small and similar for all systems for all regions with a larger fresh bias near the surface and in the western boundary current regions. The Western Pacific has a salty bias in the upper 50 m, while the GIN Seas and Southern California have a salty bias over the upper 500 m. The ocean-only forecast have a weak salty tendency of  $\sim 0.002$  psu/day in the upper 50 m, while Navy ESPC

has a little drift or a slight freshening over the forecast. At depth, the trends are small with a tendency towards decreasing fresh bias.

The error growth over the forecast is similar for all three systems. Near the surface, Navy ESPC has smaller RMSE and lower error growth. However, the differences between the systems are small. GOFS 3.5 tends to have smaller RMSE across all regions and all forecast days, except near the surface. The biggest difference between the three systems occurs in the western boundary current region between 50 m and 200 m. However, on a given forecast day, any of the three systems may have the best performance at a particular depth in a particular region.

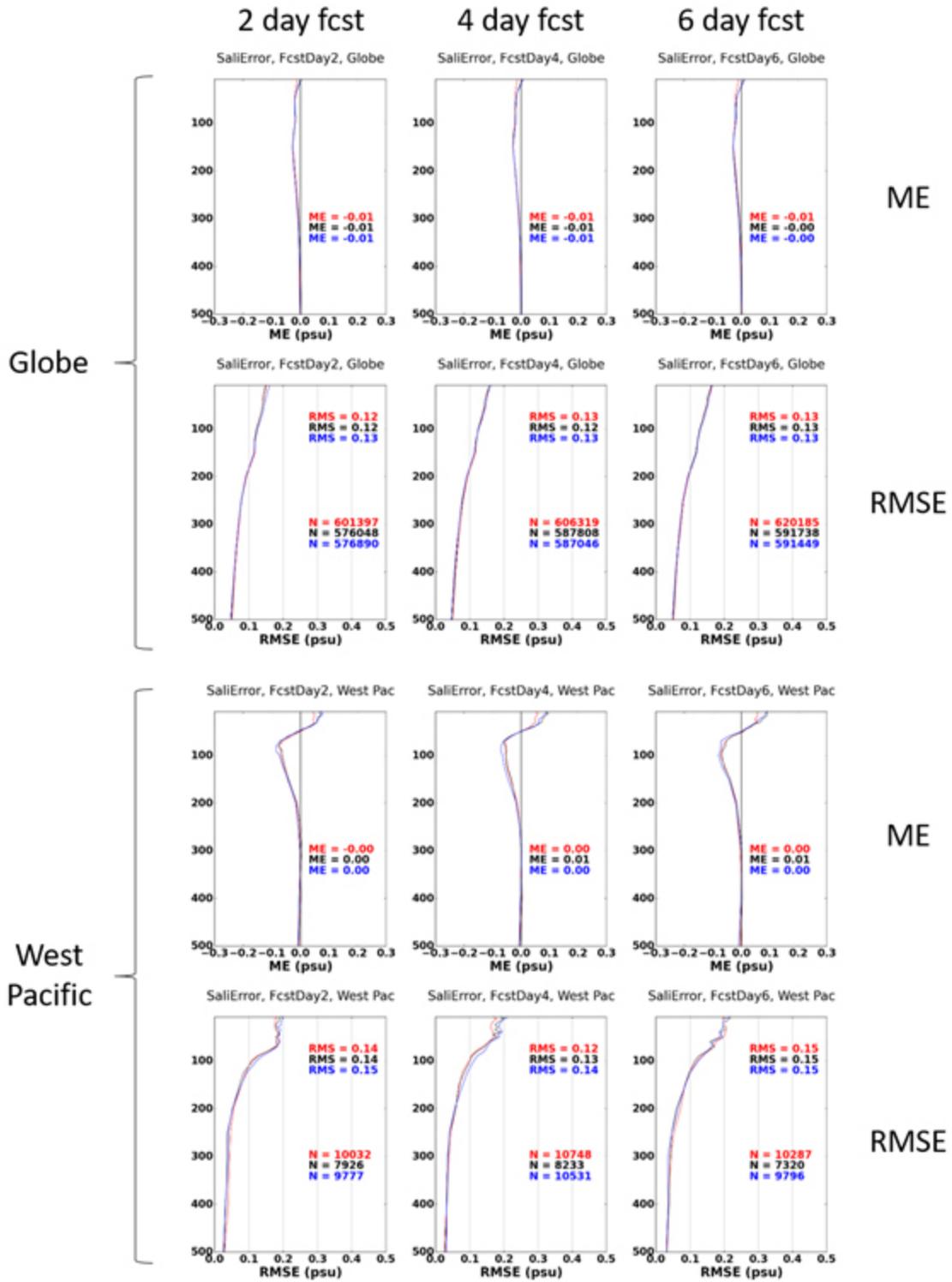


Figure 17: Salinity (psu) versus depth error analysis in the upper 500 m evaluated against unassimilated profile observations at forecast days 2, 4 and 6 for the global and Western Pacific regions. The first and third rows are the mean error (ME) and the second and fourth rows are the root mean square error (RMSE). The first column is forecast day 2, the second column is forecast day 4 and the third column is forecast day 6. The upper six plots are for the Globe, and the lower six plots are for the Western Pacific. The red curves are the Navy ESPC analyses, the

black curves are the GOFS 3.5 pre-operational ocean-only analyses and the blue curves are the operational GOFS 3.1. The numerical values for ME/RMSE are the averages over 8-500 m. The number of observations is the N=xxxxx values. The number of profiles is approximately N/20. A minimum depth of 8 m is used, because that is the typical shallowest depth sample by Argo profiles.

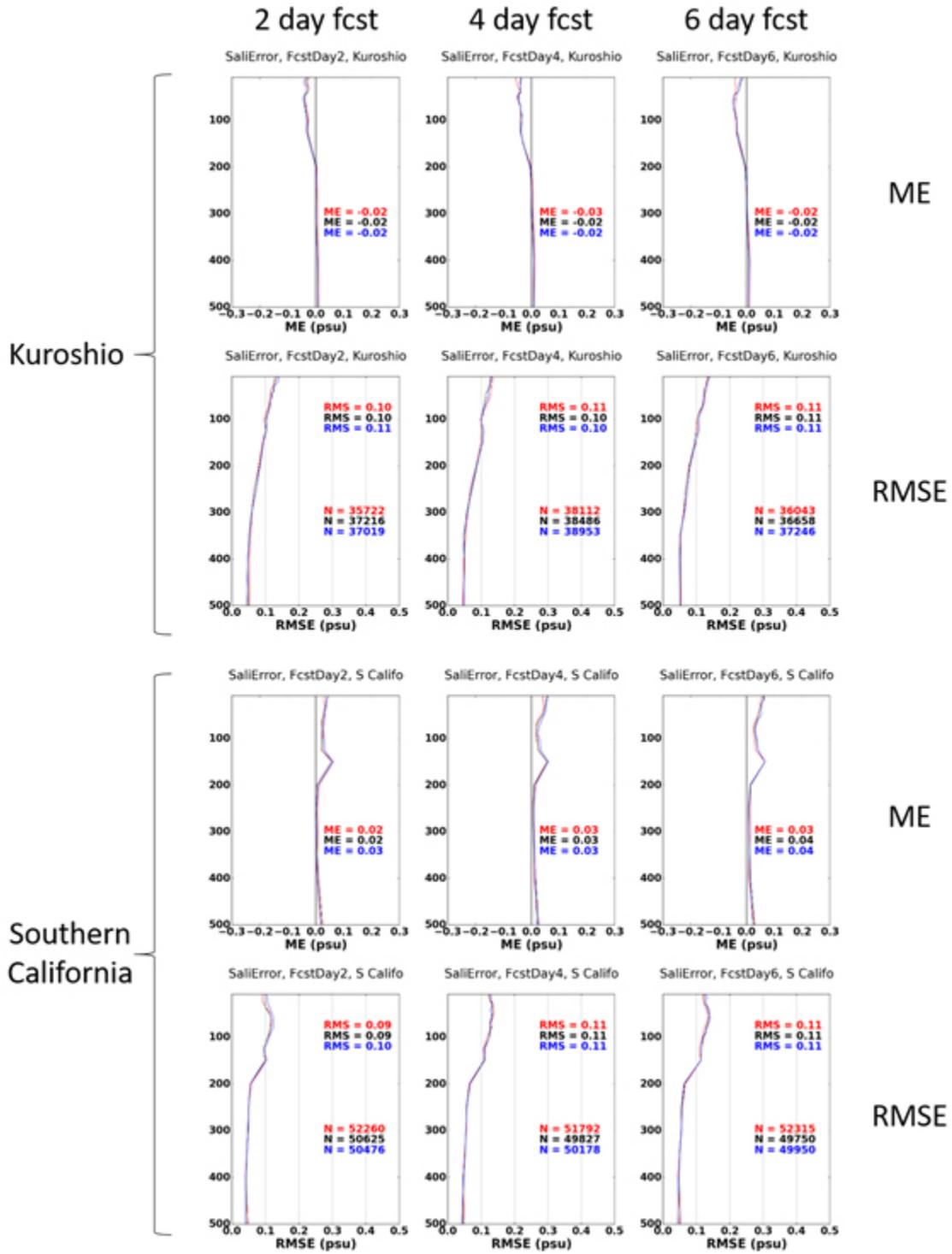


Figure 18: As Figure 17 except for the Kuroshio and Southern California regions, respectively.

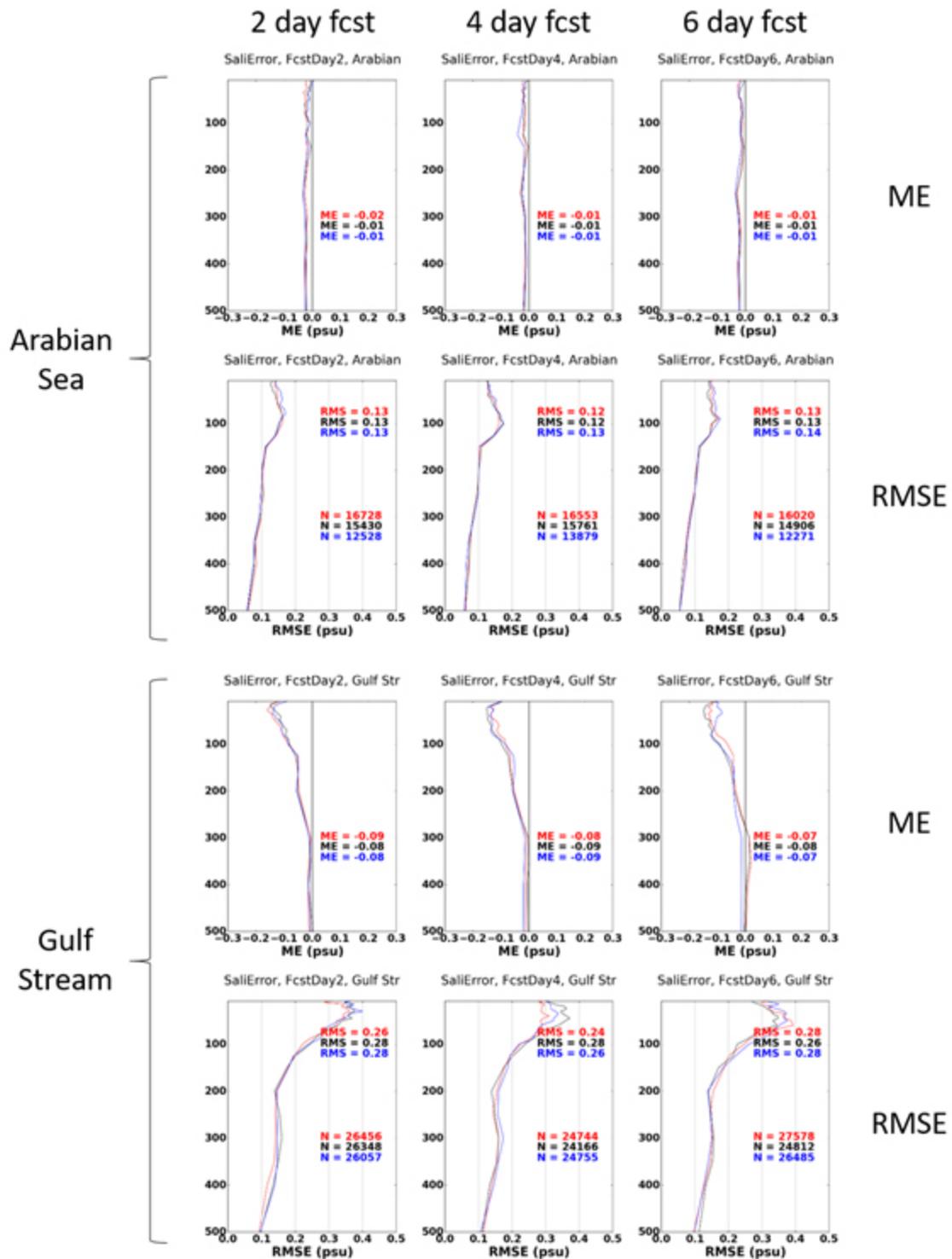


Figure 19: As Figure 17 except for the Arabian Sea and Gulf Stream regions, respectively.

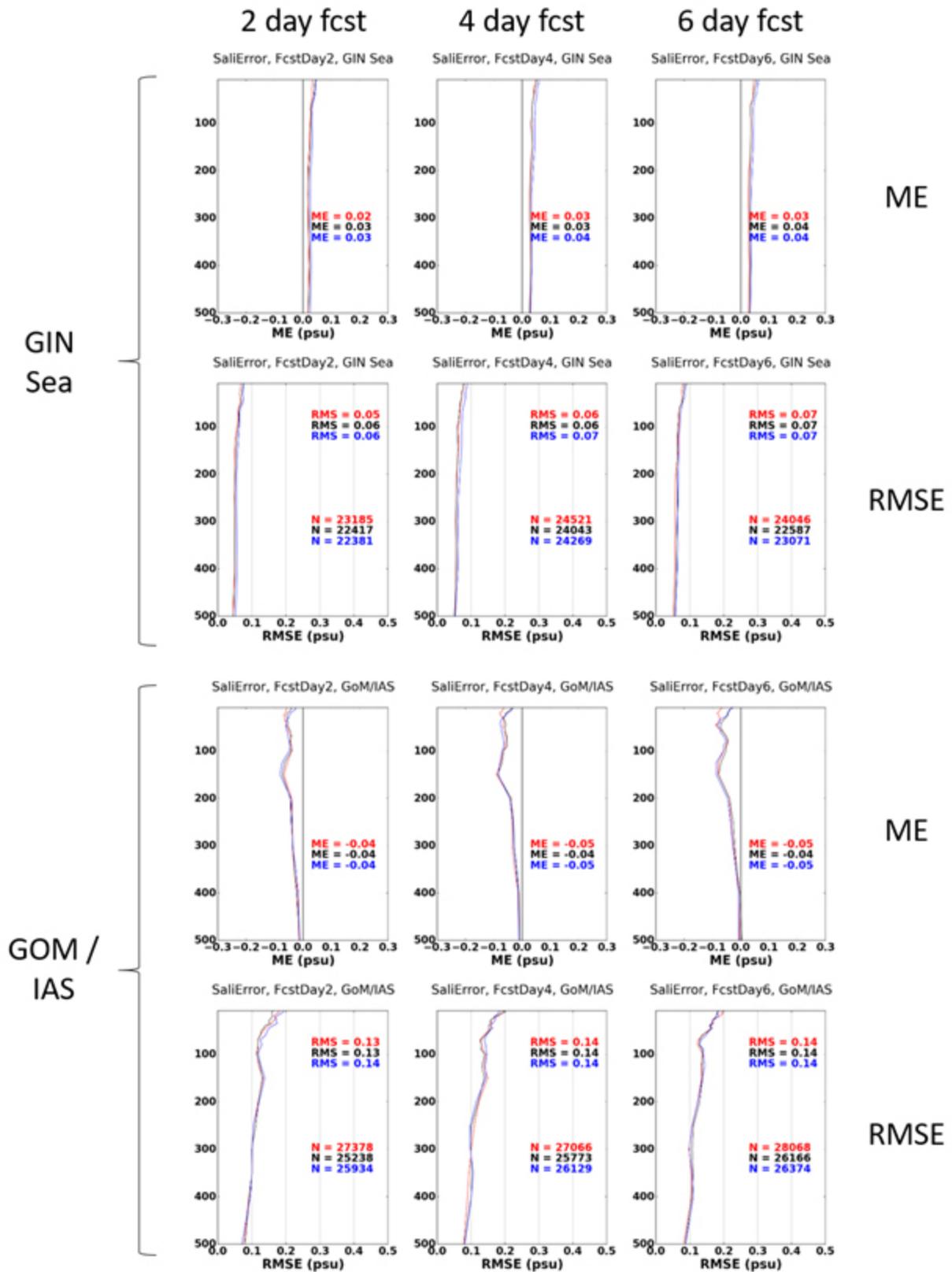


Figure 20: As Figure 17 except for the GIN Sea and GOM/IAS regions, respectively.

The global upper ocean salinity bias as a function of forecast length is shown in Figure 21 and Figure 22. For Navy ESPC, the trends observed in the 7-day forecasts continue in the 16-day forecasts. Near the surface, little change occurs in the bias for the second week of the forecast, while a slight freshening occurs at depth.

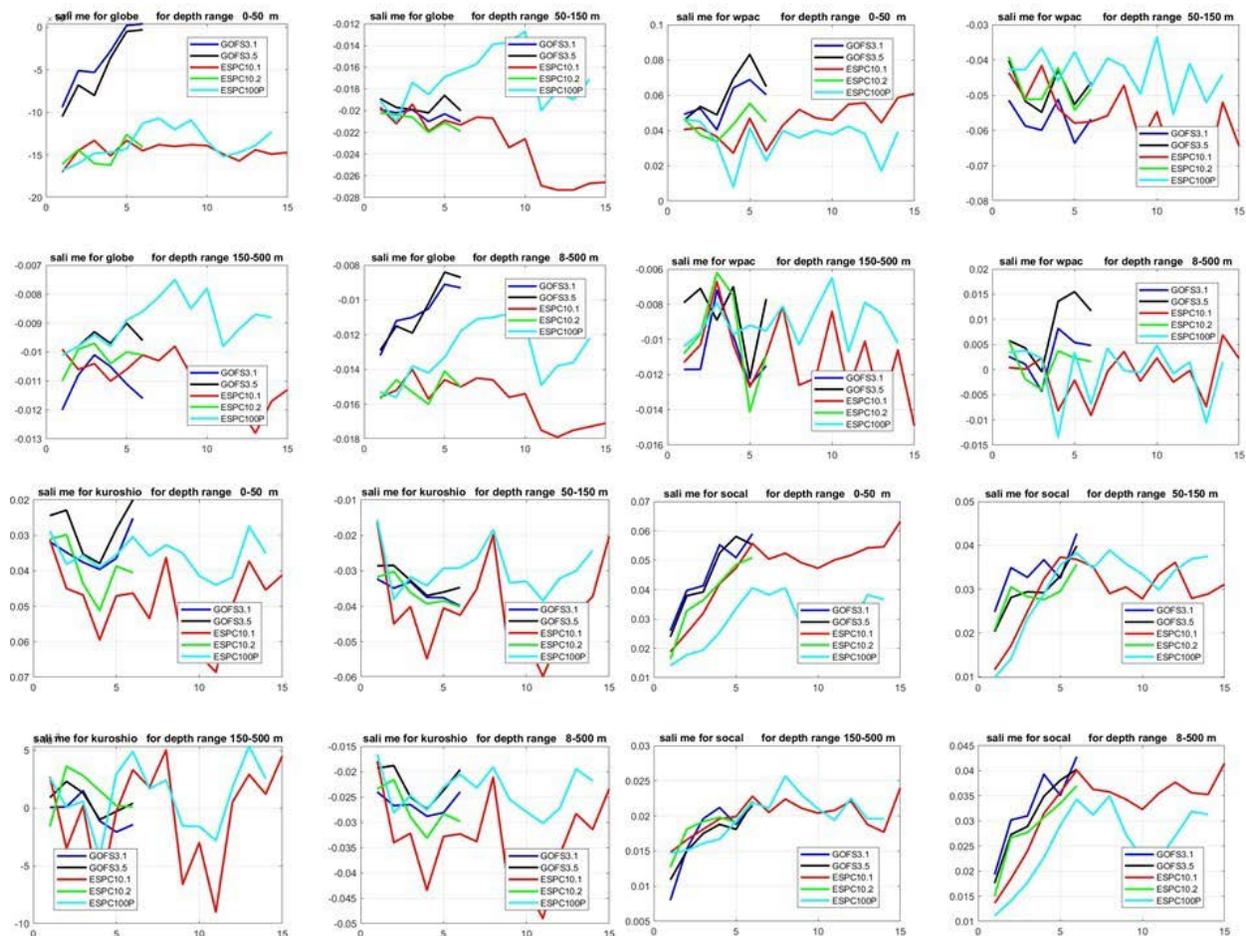


Figure 21: Salinity (psu) mean error (ME) or bias analysis in the upper 500 m evaluated against unassimilated profile observations as a function of forecast length for 4 regions over the globe and Pacific Ocean (Globe, Western Pacific, Kuroshio Extension and Southern California). The analyses for a region are grouped over four depth ranges, 0-50m, 50-150m, 150-500m and 8-500m. Thus, the first and third rows are the 0-50m and 50-150m depths and the second and fourth rows are the 150-500m and 8-500m depths. The upper left four plots are for the Globe, the upper right four plots are for the Western Pacific, lower left four plots are for the Kuroshio Extension and the lower right four plots are for Southern California. The red curves are the Navy ESPC 16-day forecast, the green curves are the Navy ESPC 7-day forecasts, the black curves are the GOF3.5 pre-operational ocean-only 7-day forecasts and the blue curves are the operational GOF3.1 ocean-only 7-day forecasts. The cyan curves are the persistence of the Navy ESPC nowcast analysis over the 16 days of the forecast.

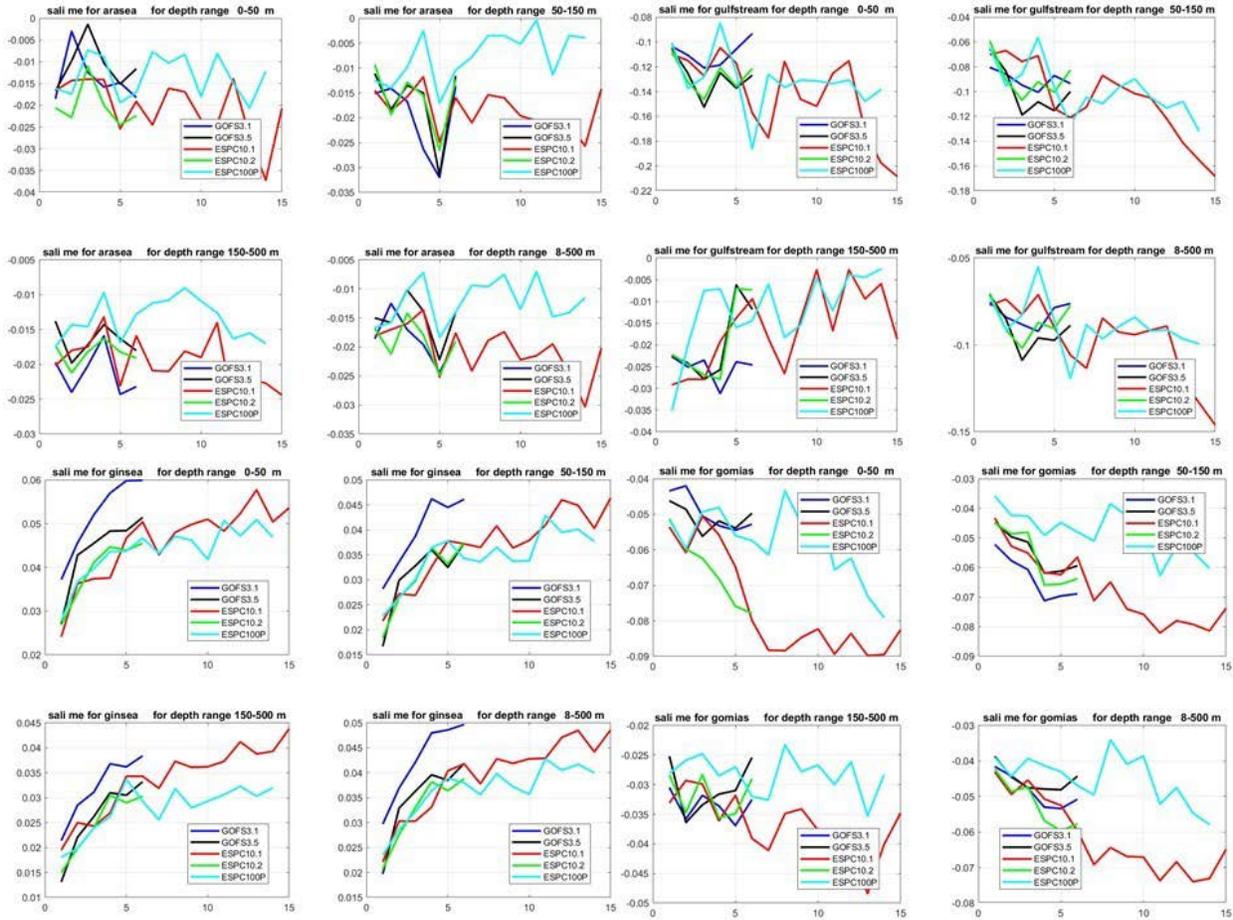
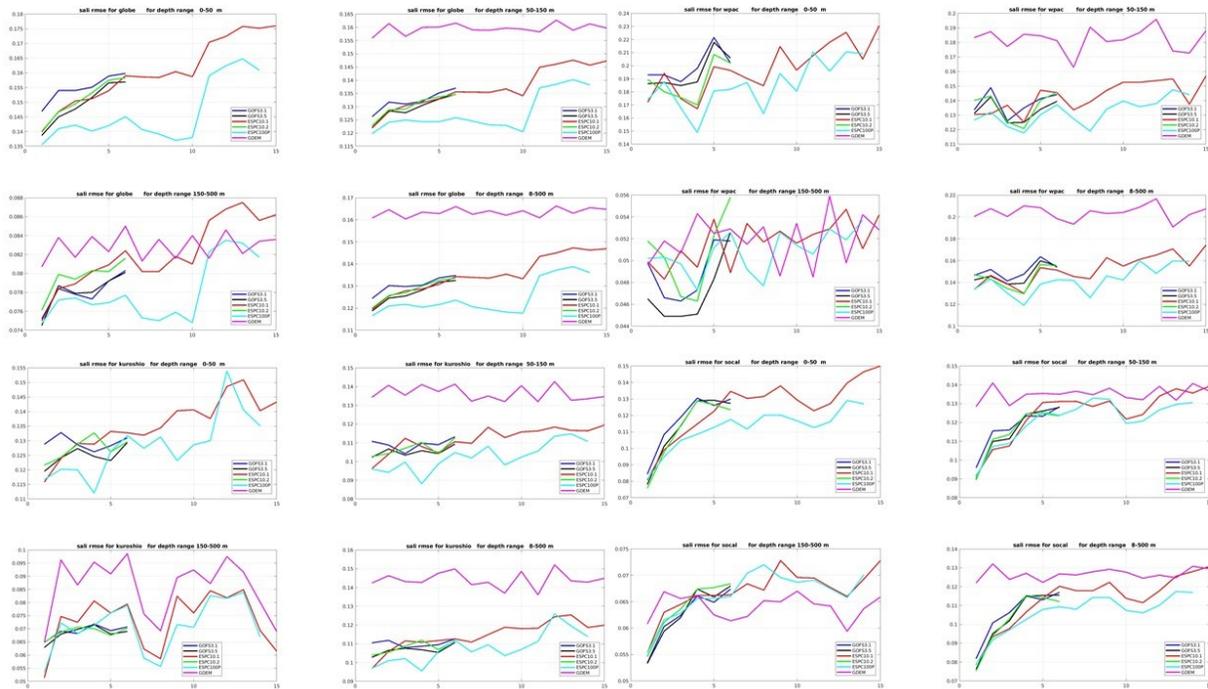


Figure 22: As Figure 21 except the upper left four plots are for the Arabian Sea, the upper right four plots are for the Gulf Stream, lower left four plots are for the GIN Seas and the lower right four plots are for GOM/IAS.

The global upper ocean salinity RMSE as a function of forecast length is shown in Figure 23 and Figure 24. The two higher resolution systems have nearly identical RMSE over the first seven days with similar error growth for Navy ESPC over the second week. The higher resolution systems have slightly small RMSE than the operational system GOF3.1. All three systems perform better than climatology over their entire forecast. The rate of error growth observed by all systems over the 7-day forecasts continues for Navy ESPC during the second week of the extended range forecast in the upper 50 m and the error grows slower in the second week below 50 m.



**Figure 23: Salinity (psu) root mean square error (RMSE) analysis in the upper 500 m evaluated against unassimilated profile observations as a function of forecast length for 4 regions over the globe and Pacific Ocean (Globe, Western Pacific, Kuroshio Extension and Southern California). The analyses for a region are grouped over four depth ranges, 0-50m, 50-150m, 150-500m and 8-500m. Thus, the first and third rows are the 0-50m and 50-150m depths and the second and fourth rows are the 150-500m and 8-500m depths. The upper left four plots are for the Globe, the upper right four plots are for the Western Pacific, lower left four plots are for the Kuroshio Extension and the lower right four plots are for Southern California. The red curves are the Navy ESPC 16-day forecast, the green curves are the Navy ESPC 7-day forecasts, the black curves are the GOFs 3.5 pre-operational ocean-only 7-day forecasts and the blue curves are the operational GOFs 3.1 ocean-only 7-day forecasts. The cyan curves are the persistence of the Navy ESPC nowcast analysis over the 16 days of the forecast. The magenta curve is the comparison to climatology (GDEM4).**

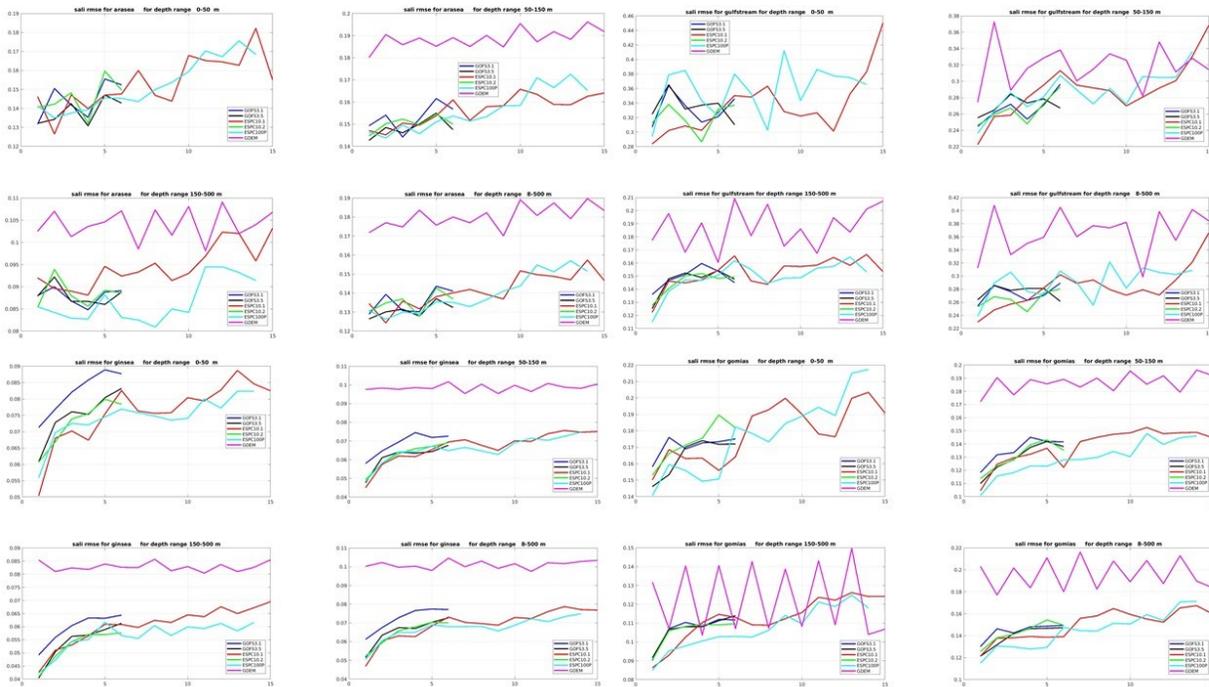


Figure 24: As Figure 23 except for the 4 regions in the Indian and Atlantic Oceans (Arabian Sea, Gulf Stream, Greenland/Iceland/Norwegian (GIN) Seas and Gulf of Mexico/IntraAmericas Sea(GOM/IAS)). The upper left four plots are for the Arabian Sea, the upper right four plots are for the Gulf Stream, lower left four plots are for the GIN Seas and the lower right four plots are for GOM/IAS.

### 4.1.2 Acoustical Proxy Error Analysis

Accurate knowledge of the underwater acoustical environment is important to maintaining tactical advantages during naval operations. The three-dimensional structure of temperature and salinity and the surface mixed layer depth determine the sound speed profile, which characterizes the acoustical ducts in the ocean. Thus, the ocean nowcast/forecast system must be able to accurately predict the mixed layer depth (MLD), sonic layer depth (SLD), below layer sound speed gradient (BLG) and other acoustical proxies. The MLD is defined by a density difference of  $0.15 \text{ kg/m}^3$  between the surface and a given depth if both T and S profiles are available or a difference of  $0.5^\circ\text{C}$  between the surface and a given depth if only T profiles are available. The SLD is the distance between the surface and the depth of the sound speed maximum, which is often, but not always, at the base of the mixed layer. The BLG is the rate of change of the sound speed with depth per 100 feet in the first 300 feet below the SLD or below the surface if the SLD is absent. These quantities are derived from the Naval Oceanographic Office Reference Publication 33 (RP33, 1992).

Histograms of the bias (ME) and RMSE for the MLD at the “nowcast” time separated into shallow mixed layers (MLD < 50 m) and intermediate mixed layers (50 m < MLD < 250 m) are shown in Figure 25. All systems have a shallow bias for both shallow and intermediate MLD. The shallow bias is  $\sim 5 \text{ m}$  (or  $\sim 10\%$ ) for the shallow mixed layers with the largest biases in the Arabian Sea.

The results are mixed across the regions, but Navy ESPC has similar bias as GOFS 3.5. The operational GOFS 3.1 has the smallest bias for most regions. For intermediate mixed layers the shallow bias is larger (~10 m) with the largest bias in the Arabian Sea and Navy ESPC has a slightly smaller bias than GOFS 3.5 across most regions. Operational GOFS 3.1 has the smallest intermediate MLD bias for most regions. For RMSE, the typical errors are ~15 m for shallow mixed layers and ~25 m for intermediate depth mixed layer. The largest RMSE > 20 m occur in the Arabian Sea for shallow mixed layers and RMSE > 30 m for the Arabian and GIN Seas for intermediate depth mixed layers. Overall, Navy ESPC has slightly lower RMSE than GOFS 3.5 across the regions with the exception of the Arabian and GIN Seas where GOFS 3.5 performs better. Again, operational GOFS 3.1 has the smallest RMSE, except in the Gulf of Mexico/IntraAmericas Sea.

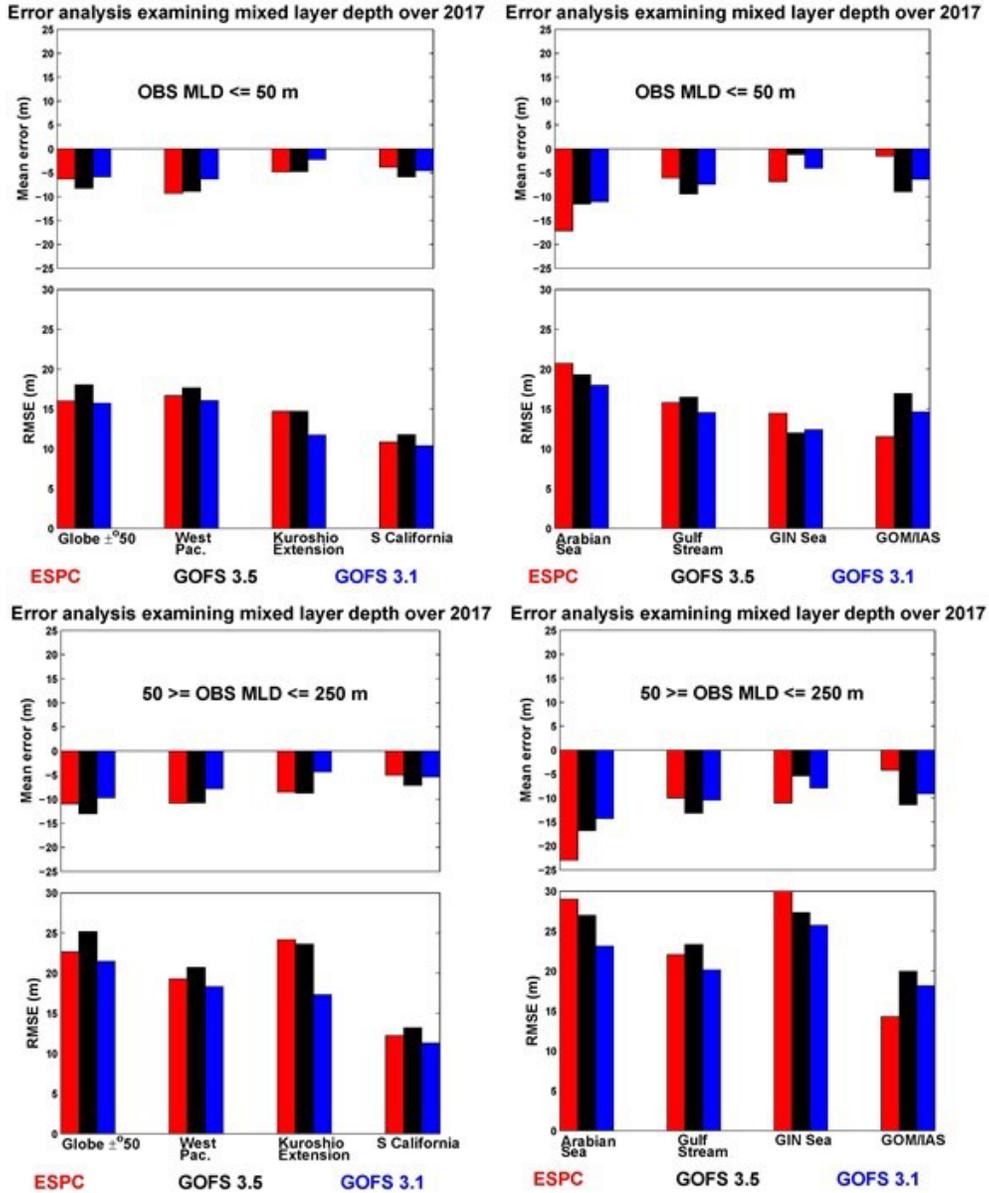


Figure 25. The ME and RMSE of the mixed layer depth (MLD) at “nowcast” time for the eight analysis regions separated into shallow MLD < 50m (top two rows) and intermediate 50 m <MLD <250 m (bottom two rows). The red bars are Navy ESPC, the black bars are the pre-operational GOFS 3.5, and the blue bars are operational GOFS 3.1. All systems show a shallow bias in the MLD for shallow and intermediate depth mixed layers.

The bias of the mixed layer depth as a function of forecast day is shown in Figure 26. The relative performance during the forecast is similar for all systems. If the initial bias is smaller for a system, it tends to remain smaller over the forecast. For the globe the bias for all systems tends to increase during the forecast. However, except for GOM/IAS, little trend in bias is observed for the regions. The global increase in bias appears to be associated with the GOM/IAS regions of the Atlantic Ocean.

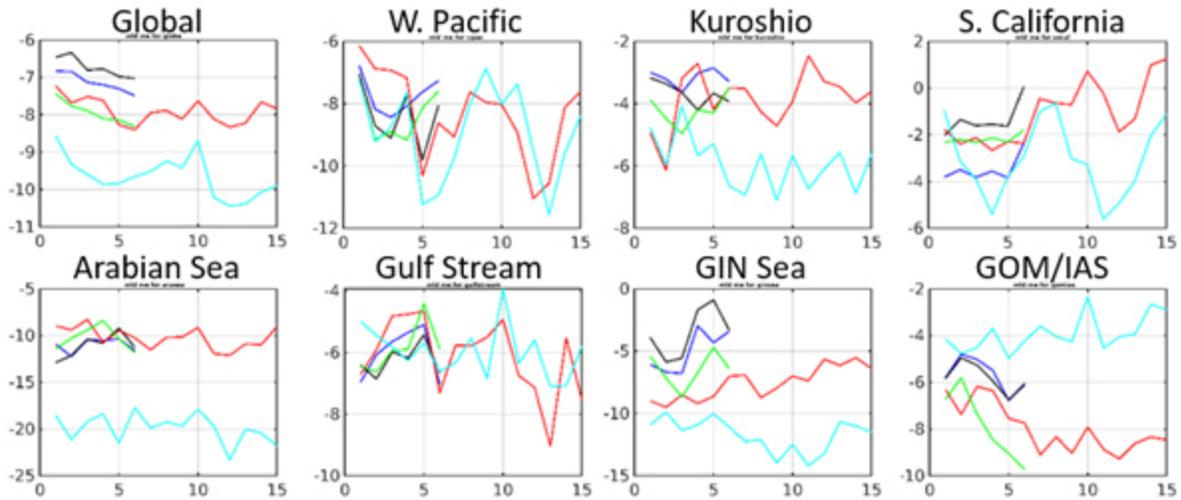


Figure 26: Mixed layer depth (MLD) bias (ME) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFS 3.5, blue lines are operational GOFS 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

The mixed layer depth RMSE as a function of forecast day is shown in Figure 27. All systems have a similar error growth as the forecast progresses. Globally, the RMSE error growth is about 0.5 m/day or about 13%/day. Except for the Arabian Sea, where little error growth is observed, the error growth for each region is about 0.5 m/day representing 13-20%/day. The largest MLD RMSE is observed in the Arabian and GIN Seas.

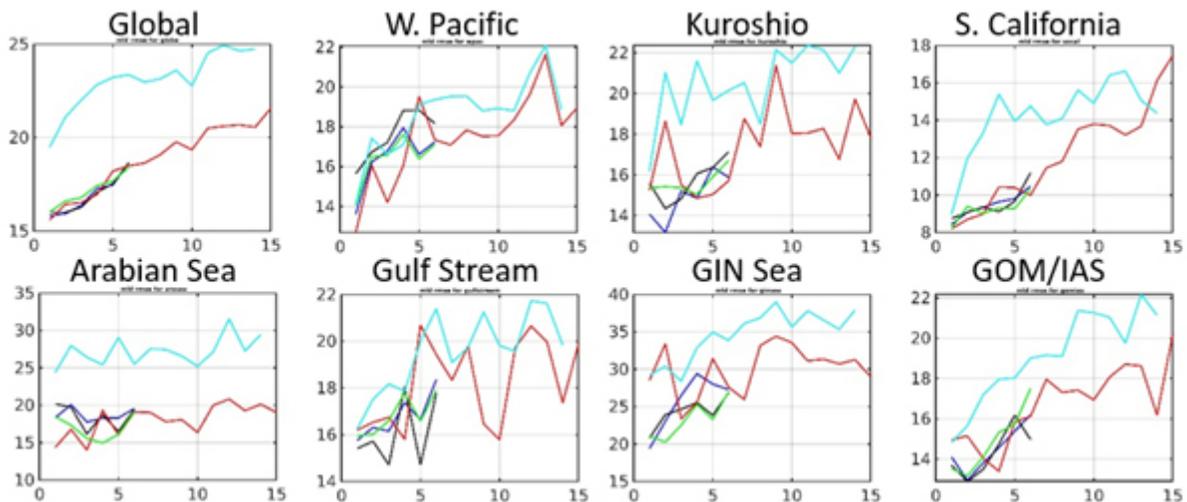


Figure 27: Mixed layer depth (MLD) root mean square error (RMSE) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFS 3.5, blue lines are operational GOFS 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

Histograms of the bias (ME) and RMSE for the SLD at the “nowcast” time separated into shallow sonic layers (SLD < 50 m) and intermediate depth sonic layers (50 m < SLD < 250 m) are shown in Figure 28. All systems have a shallow bias for both shallow and intermediate SLD. The shallow bias is ~ 5 m (or ~10%) for the shallow sonic layers with the largest biases in the Arabian Sea. GOFS 3.5 has slightly smaller bias than Navy ESPC and GOFS 3.1 across most regions. For intermediate sonic layers the shallow bias is larger (~10 m) with the largest bias in the Arabian Sea and GOFS 3.5 has a slightly smaller bias than Navy ESPC and GOFS 3.1 across the regions, although the differences generally are small. For RMSE, the typical errors are ~15 m for shallow sonic layers and ~20 m for intermediate depth sonic layer. For shallow sonic layers the RMSE is similar across the regions with NAVY ESPC having the poorest performance in the Arabian Sea. The largest RMSE > 30 m for intermediate depth sonic layers occurs in the western boundary current regions of the Gulf Stream and Kuroshio Extension. Overall, GOFS 3.5 has slightly lower RMSE than Navy ESPC and GOFS 3.1 across the regions with the poorest performance by Navy ESPC in the Arabian Sea.

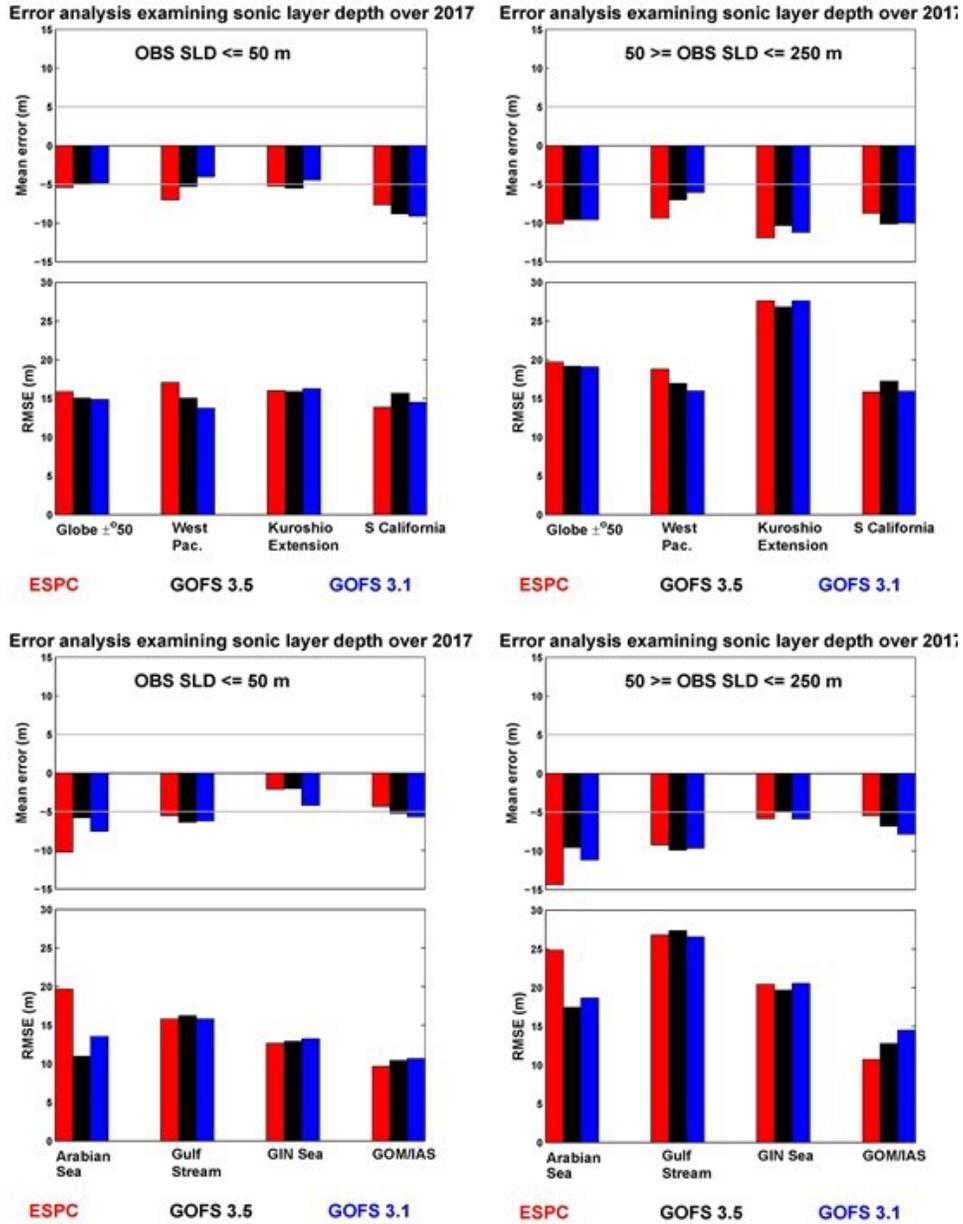


Figure 28. The ME and RMSE of the sonic layer depth (SLD) at “nowcast” time for the eight analysis regions separated into shallow SLD < 50m (left column) and intermediate depth 50 m <SLD <250 m (right column). The red bars are Navy ESPC, the black bars are the pre-operational GOFS 3.5, and the blue bars are operational GOFS 3.1. Both models show a shallow bias in the SLD for shallow and intermediate sonic layers.

The bias of the sonic layer depth as a function of forecast day is shown in Figure 29. The relative performance during the forecast is similar for all systems. If the initial bias is smaller for a system, it tends to remain smaller over the forecast. For the globe the bias for the ocean only systems tends to increase during the forecast, but changes very little for Navy ESPC. For most region, little trend in bias is small, approximately 0.2-0.5 m/day. Navy ESPC has a large increasing trend in SLD bias in GOM/IAS.

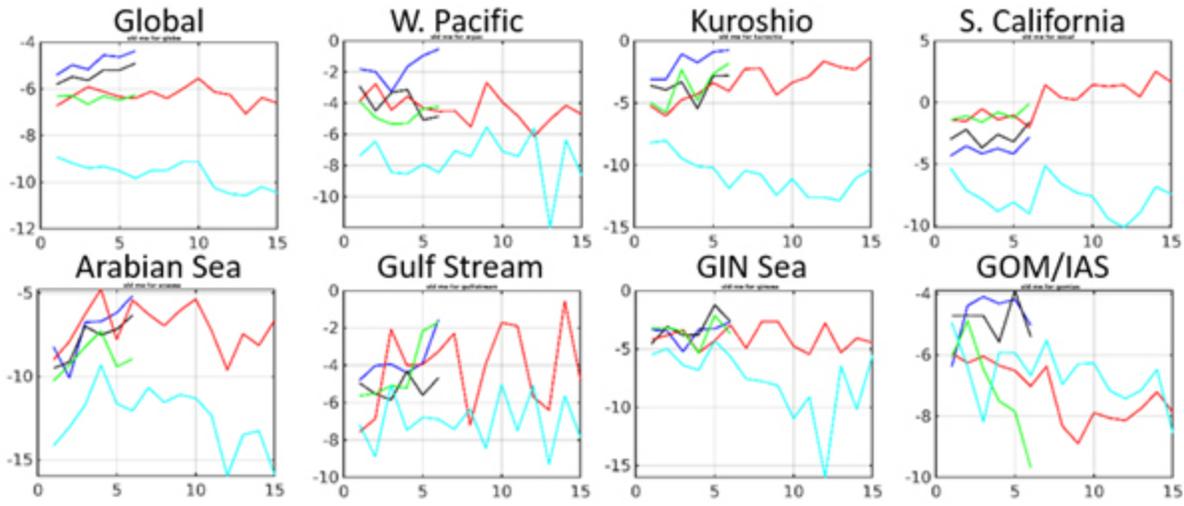


Figure 29: Sonic layer depth (SLD) bias (ME) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFS 3.5, blue lines are operational GOFS 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

The sonic layer depth RMSE as a function of forecast day is shown in Figure 30. All systems have a similar small error growth as the forecast progresses. Globally, the RMSE error growth is about 0.2 m/day or about 7%/day. Except for the Arabian Sea, where little error growth is observed, the error growth for each region is about 0.5 m/day representing 13-20%/day. The largest SLD RMSE is observed in the Gulf Stream.

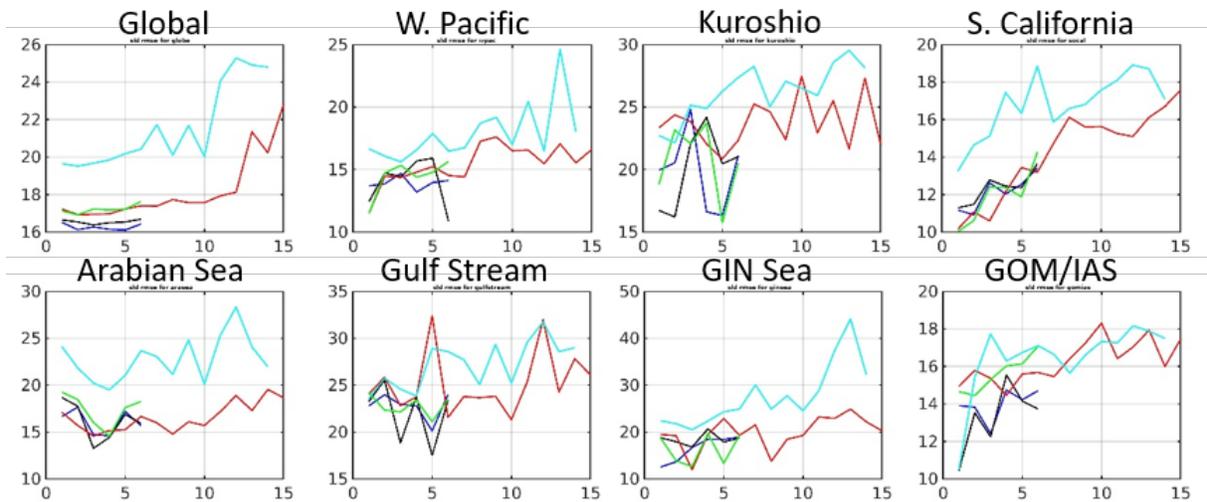


Figure 30: Sonic layer depth (SLD) root mean square error (RMSE) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFS 3.5, blue lines are operational GOFS 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

Histograms of the bias (ME) and RMSE for the below layer sound speed gradient (BLG) at the “nowcast” time are shown in Figure 31. In general, no consistent pattern emerges for the bias in BLG with both positive and negative ME across the regions. GOFS 3.5 has slightly lower bias than Navy ESPC and GOFS 3.1. The largest bias in the BLG is observed in the Arabian Sea. For RMSE, GOFS 3.5 has slightly better performance than Navy ESPC and GOFS 3.1. The largest RMSE occur in the western boundary current regions of the Gulf Stream and Kuroshio Extension, the Western Pacific and Arabian Sea. The largest RMSE occurs for Navy ESPC in the Arabian Sea.

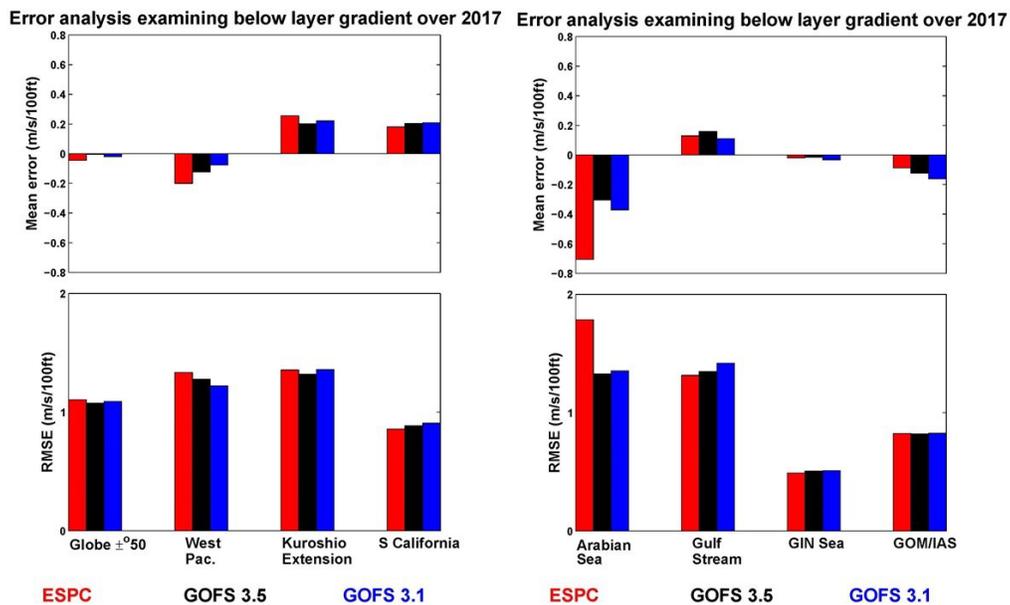


Figure 31. The ME and RMSE of the below layer gradient (BLG) at “nowcast” time for the eight analysis regions. The red bars are Navy ESPC, the black bars are the pre-operational GOFS 3.5, and the blue bars are operational GOFS 3.1. GOFS 3.5 has slightly lower ME and RMSE than Navy ESPC across the regions.

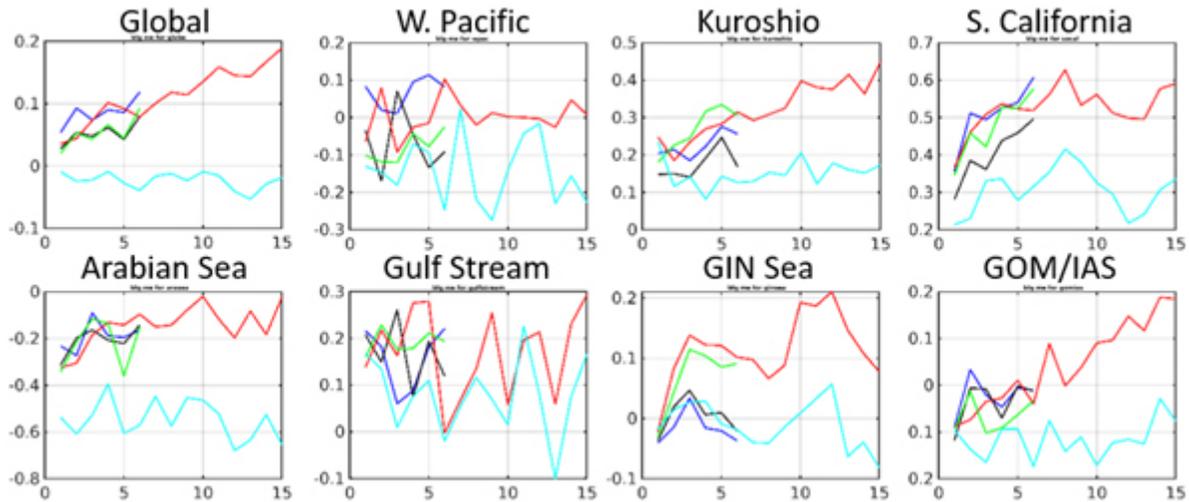


Figure 32: Below layer sound speed gradient (BLG) bias (ME) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFs 3.5, blue lines are operational GOFs 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

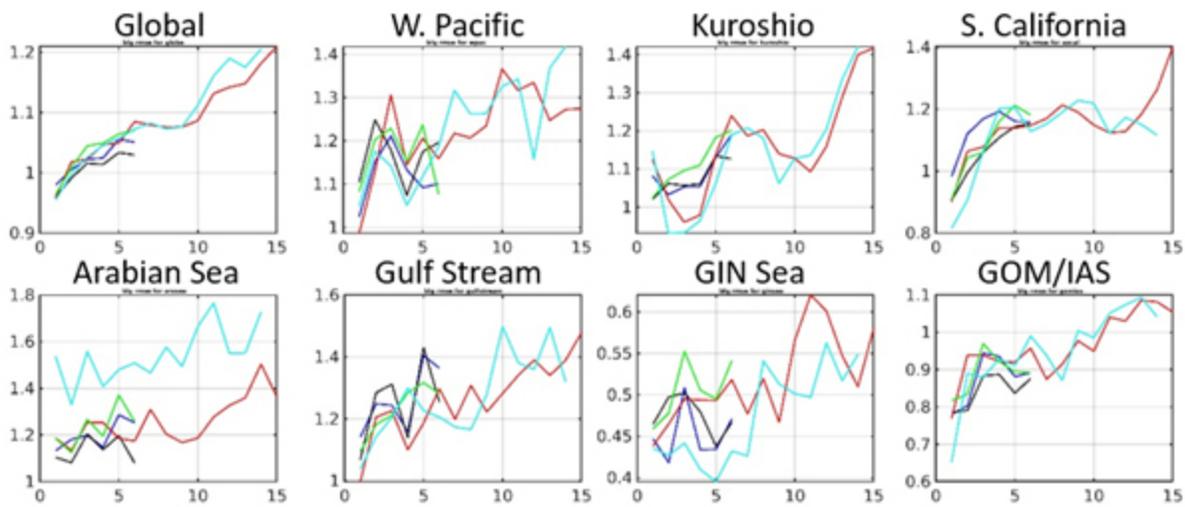


Figure 33: Below layer sound speed gradient (BLG) root mean square error (RMSE) against unassimilated observations as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecasts, black lines are pre-operational GOFs 3.5, blue lines are operational GOFs 3.1 and cyan lines are persistence of “nowcast” Navy ESPC analysis.

#### 4.1.3 Upper Ocean Velocity Validation

Upper ocean velocity is compared against drifting buoys obtained from the National Oceanic and Atmospheric Administration Global Drifter Program. Observed SST from these drifters may be assimilated into both systems (if it is reported in a timely fashion and appears on Global Telecommunications System), but the velocities are not and thus they are an independent validation data set. The drifters have a drogoue at 15 m and instantaneous velocity output from

all systems is extracted at this depth. The error analyses used are mean error (ME) and RMSE of 15 m speed and vector correlation (Crosby et al., 1993) to provide directional error.

Figure 34 shows monthly averaged values of these three metrics for GOFS 3.5 (black), Navy ESPC (red) and GOFS 3.1 (blue) over the entire year. All three systems have a positive current speed bias which is consistent with the 10 m winds in both Navy ESPC and operational NAVGEM being stronger than observed (Figure 70). For most months, GOFS 3.1 has lower ME and RMSE than Navy ESPC and GOFS 3.1. Navy ESPC tends to have the largest ME and RMSE over the year, which may reflect the use of wind bias correction in the GOFS systems which is not used in Navy ESPC. For vector correlation, the results are mixed, although Navy ESPC tends to have the lowest correlation indicating that surface wind direction may not be as accurate as operational NAVGEM.

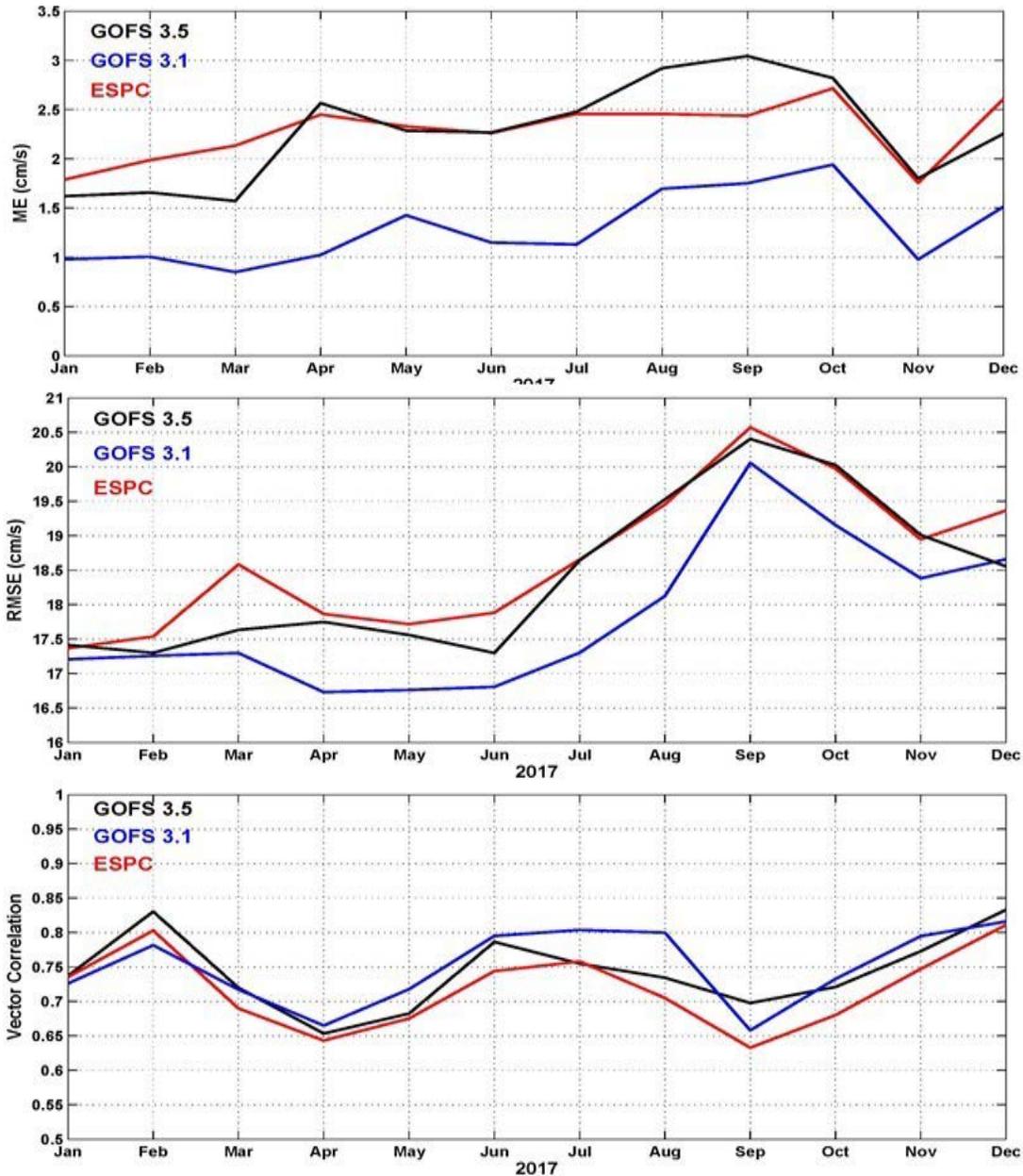


Figure 34. Monthly speed mean error (ME, cm/s, top) and root mean square error (RMSE, cm/s, middle) and direction vector correlation (bottom) of GOFS 3.1 (blue), GOFS 3.5 (black) and Navy ESPC (red) at the nowcast time versus independent drifting buoys at 15 m depth from the NOAA Global Drifter Program. The vector correlation is based on Crosby et al. (1993) with zero indicating no correlation and two indicating perfect correlation.

The upper ocean current speed mean error (ME) relative to drifting buoy observations as a function of forecast day is shown in Figure 35. All systems tend to have a small positive bias in the current speed except in GIN Seas and GOM/IAS. The trends are also small indicating relatively consistent surface wind forcing over the entire length of the forecast. GOFS 3.1 has a negative speed bias in the Western Pacific and GIN Seas. Operational GOFS 3.1 often has the smallest bias

and Navy ESPC the largest bias. Overall, the largest speed biases are found in the western boundary current regions, where the observed mean speeds are the largest.

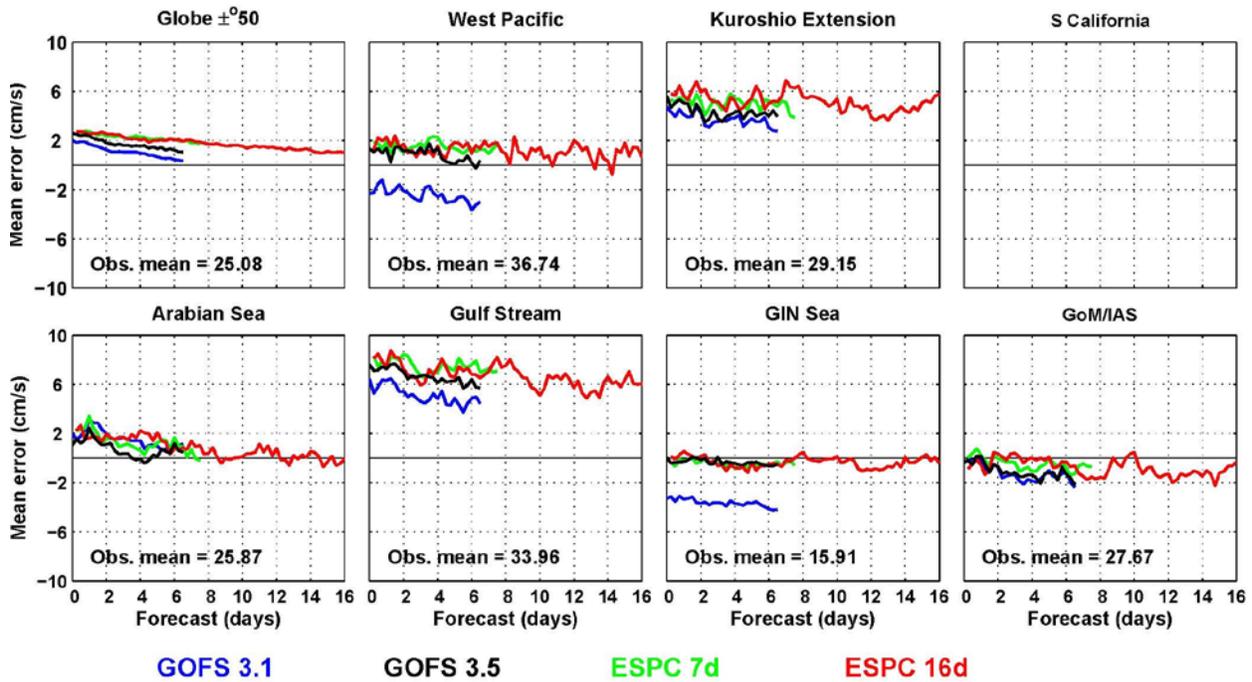


Figure 35: The 15m current speed mean error (ME, cm/s) versus independent drifting buoys as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecast, black lines are pre-operational GOFS 3.5 and blue lines are operational GOFS 3.1. The Southern California region is blank because the observations were too limited for robust statistics. The mean observed speed is noted for each region.

The upper ocean current speed RMSE compared to drifting buoys is shown in Figure 36. In general, all systems have similar performance. The RMSE for most of the forecast and most of the regions is less than the mean speed (noted in the panels of Figure 35). The largest RMSE is observed in the Gulf Stream region, where the RMSE is close to the mean speed. The error growth is small, around 0.5 (m/s)/day or about 2%/day. The error growth tendencies for the 7-day forecasts are continued into the second week of the Navy ESPC 16-day forecasts. GOFS 3.1 has smaller RMSE in several regions and Navy ESPC tends to have slightly larger RMSE.

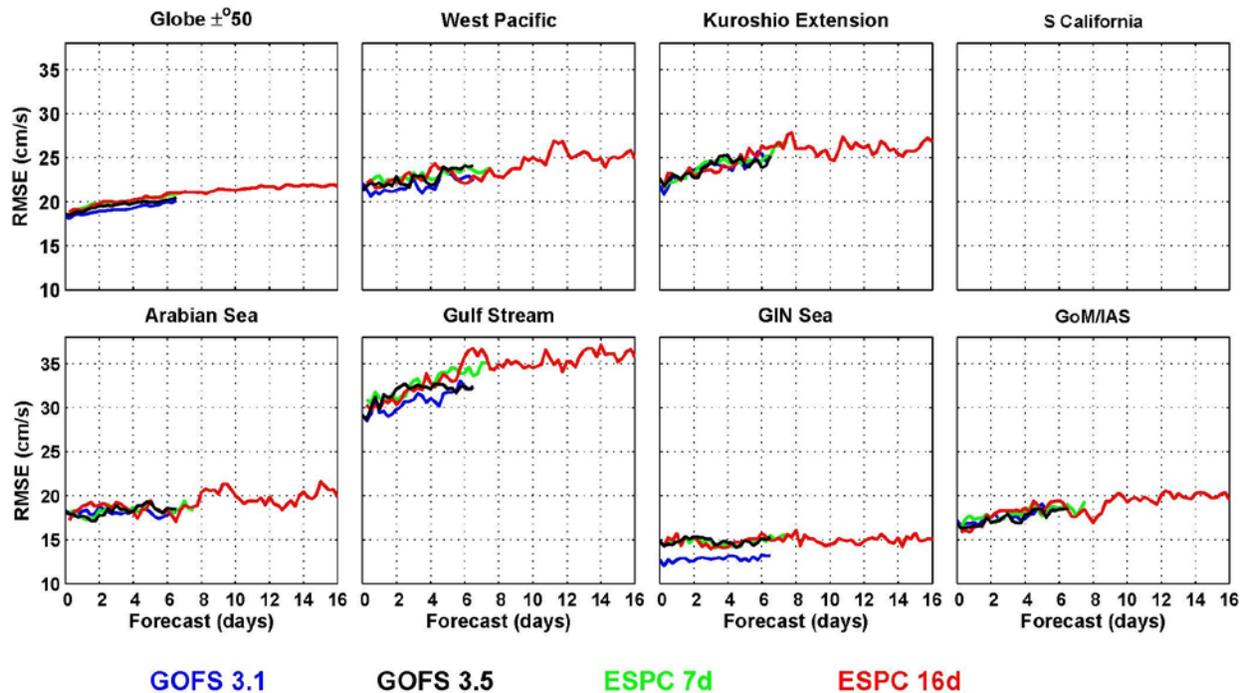


Figure 36: The 15m current speed root mean square error (RMSE, cm/s) versus independent drifting buoys as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecast, black lines are pre-operational GOFS 3.5 and blue lines are operational GOFS 3.1. The Southern California region is blank because the observations were too limited for robust statistics.

The 15m directional current vector correlation relative to drifting buoys as a function of forecast day is shown in Figure 37. A perfect correlation has a value of 2 and no correlation has a value of 0. The vector correlations generally are low, less than 1.0 except for early in the forecasts in the Western Pacific and GOM/IAS, where the vector correlation is above 1.0 for 2-3 days. All systems have similar performance with Navy ESPC slightly higher in some regions. The tendency of the vector correlation decrease for Navy ESPC is similar between weeks 1 and 2, although no significant correlation is observed in the second week.

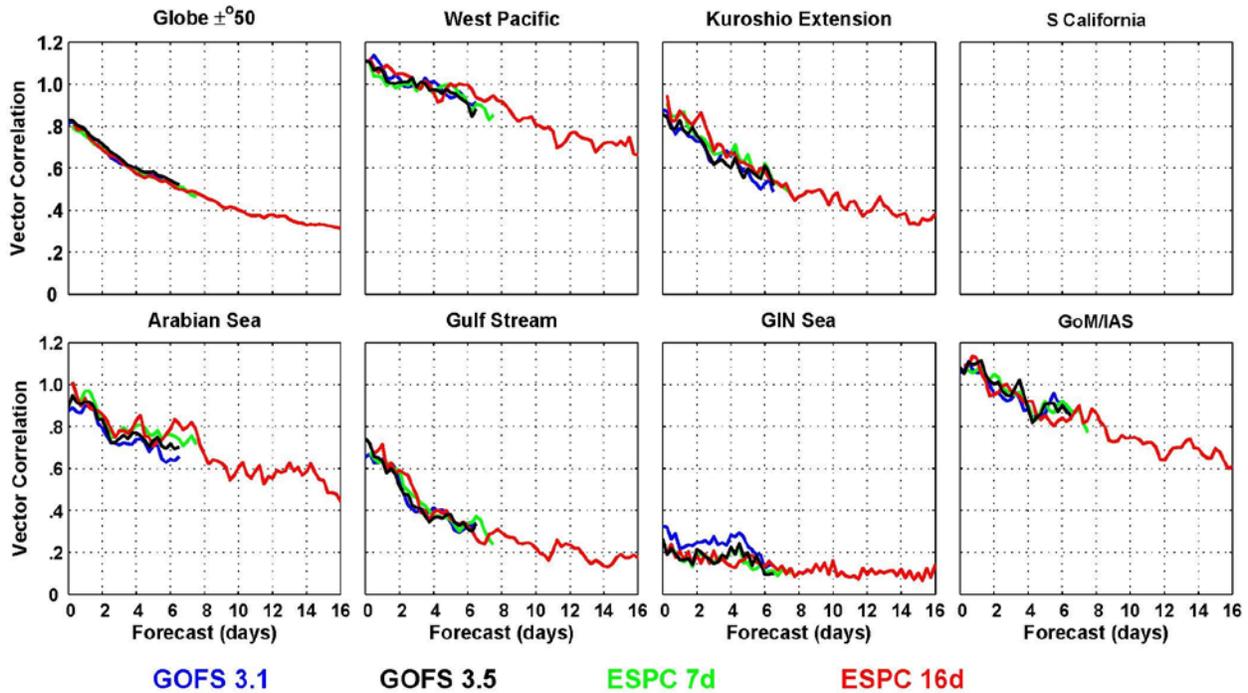
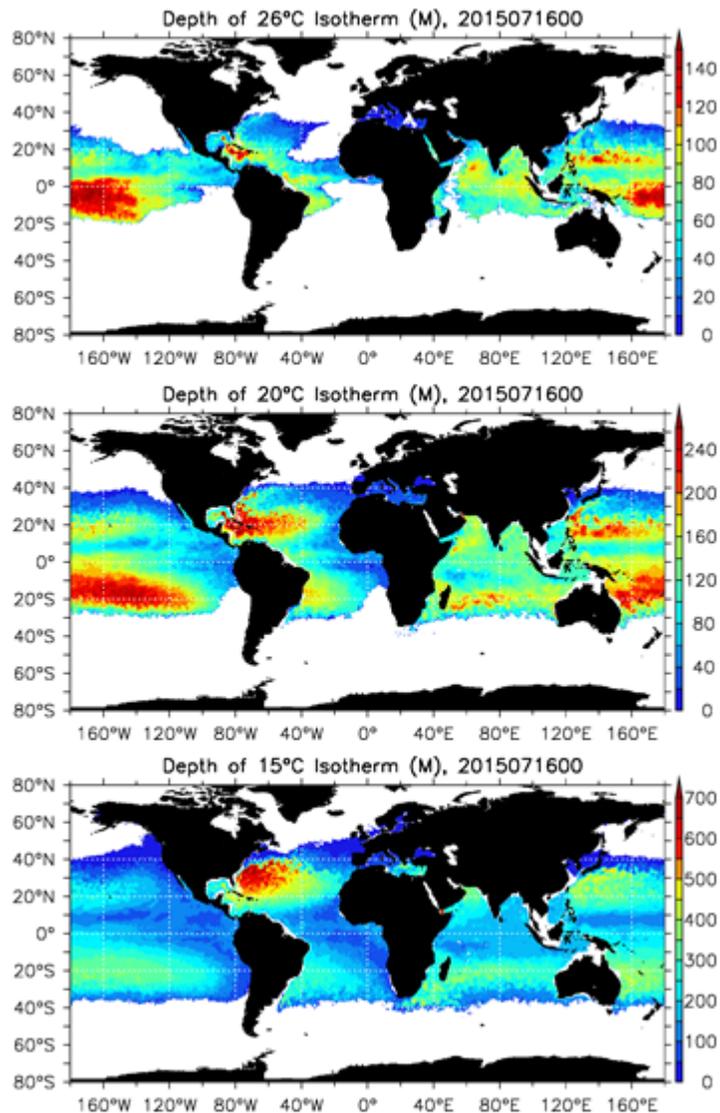


Figure 37: The 15m current directional vector correlation versus independent drifting buoys as a function of forecast length. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecast, black lines are pre-operational GOFS 3.5 and blue lines are operational GOFS 3.1. The Southern California region is blank because the observations were too limited for robust statistics.

#### 4.1.4 Isotherm Depth Error Analysis

A new metric for this report is the errors in the depths of selected isotherms. For this report three isotherms have been selected, 26°C, 20°C and 15°C. Typical depths for these isotherms are shown in Figure 38. The 26°C isotherm is shallow with depths less than 150 m and surfacing poleward of 30°. The 20°C isotherm is typically in the upper thermocline with depths less than 250 m and surfacing poleward of 40°. The 15°C isotherm is typically in the lower thermocline present everywhere in the tropics and subtropics with depths less than 500 m except in the western North Atlantic where the thermocline is much deeper and this isotherm is found around 700 m.



**Figure 38. Maps of the typical depths of the upper ocean shallow 26°C isotherm (upper), upper thermocline 20°C isotherm (middle) and lower thermocline 15°C isotherm (bottom). This example is derived from GOFS 3.1.**

The bias (ME), RMSE and mean absolute error (MAE) relative to observations for the isotherm depths at the “nowcast” time for the ocean stand-alone GOFS 3.1 (blue) and GOFS 3.5 (black), Navy ESPC (red) and GDEM climatology (gray dashed) are shown in Figure 39. All models have a small shallow bias (<2%) for the isotherm depths. The RMSE increases in magnitude for the deeper isotherms. GOFS 3.5 has slightly better performance than the other models, but the differences are small.

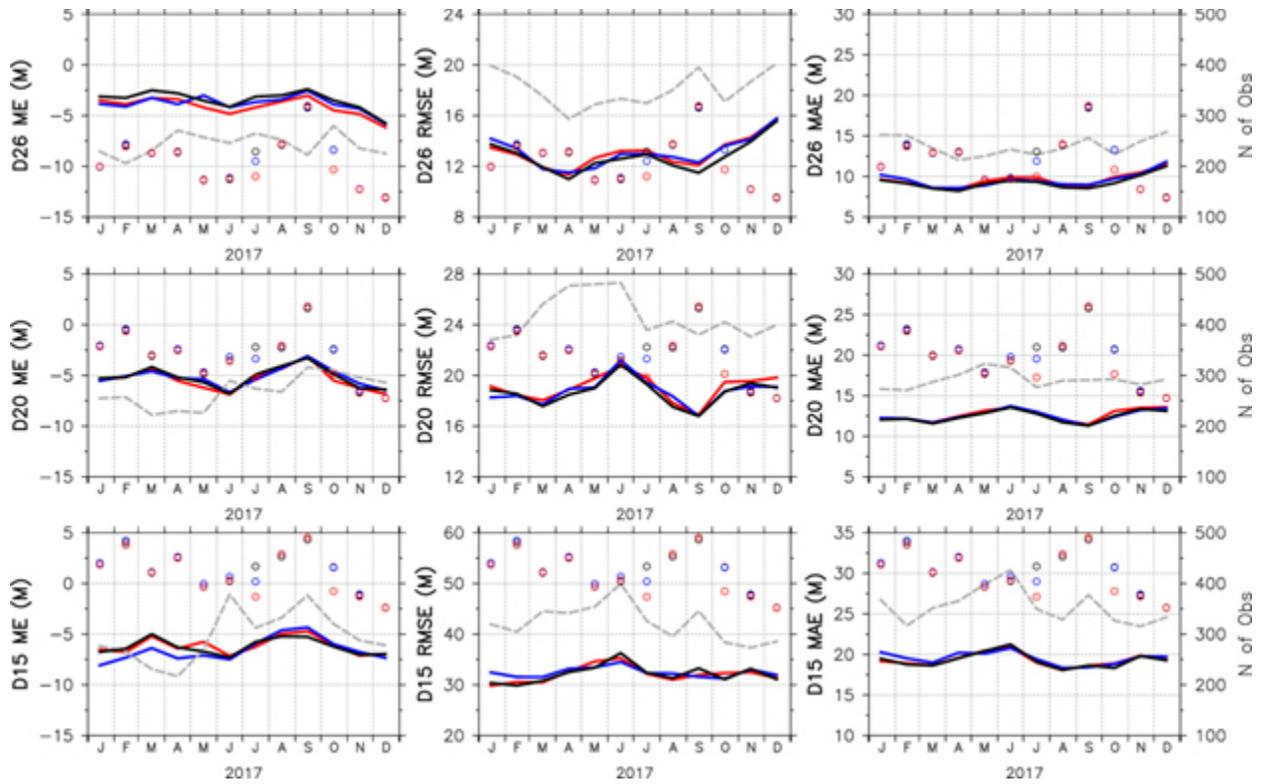


Figure 39. Monthly bias (ME, left), RMSE (middle) and mean absolute error (MAE, right) of isotherm depths relative to observations at the nowcast time. Isotherm depth of 26°C (upper), 20°C (middle) and 15°C (bottom) are shown for GOFS 3.1 (blue), GOFS 3.5 (black), ESPC (red). The gray dashed line represents GDEM climatology. The average number of observations/day are shown on the right axis and represented by the open circles in the respective colors.

The bias (ME), RMSE and mean absolute error (MAE) relative to observations for the isotherm depths as a function of forecast day for the ocean standalone GOFS 3.1 (blue) and GOFS 3.5 (black), Navy ESPC (7-day, green and 16-day, red) and GDEM climatology (gray dashed) are shown in Figure 40. The higher resolution models GOFS 3.5 and Navy ESPC perform slightly better than the lower resolution operational model GOFS 3.1, but no model consistently performs better than the others for all error metrics and all depths. The rate of growth of the errors in the 16-day forecast is similar to the shorter 7-day forecasts. The 16-day forecasts for the shallower 26°C and 20°C isotherm depths are better than climatology for the entire forecast, while the lower thermocline 15°C isotherm depth performance is nearly the same as climatology at the end of the forecast.

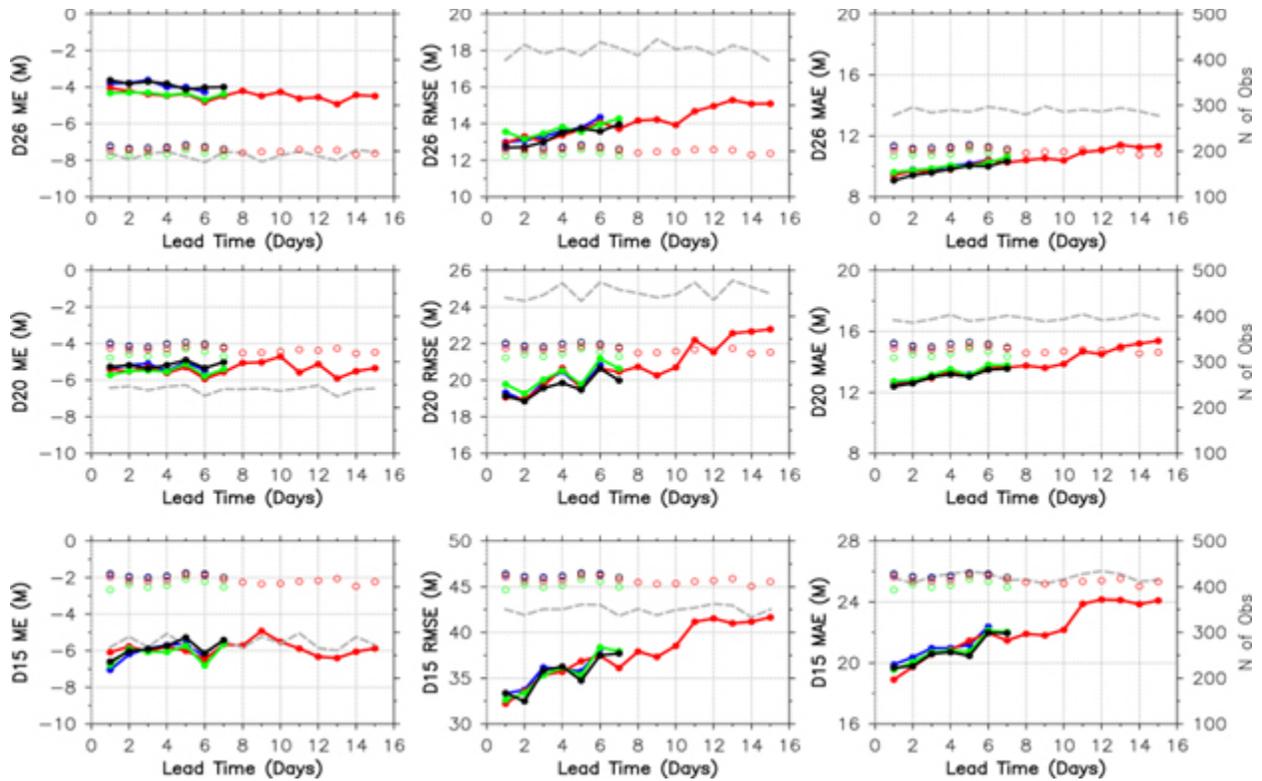
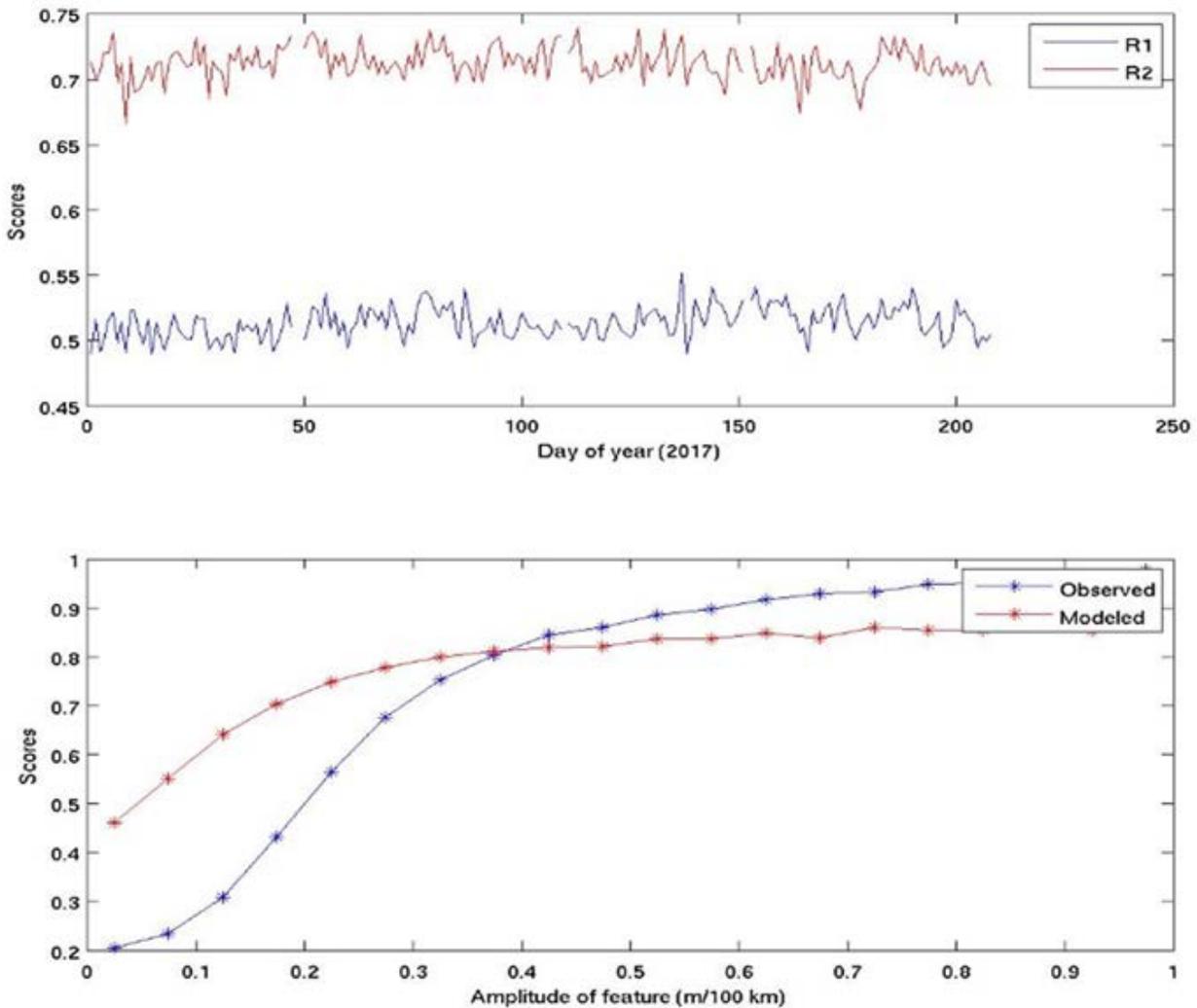


Figure 40. Bias (ME, left), RMSE (middle) and mean absolute error (MAE, right) of isotherm depths relative to observations as a function of forecast time. Isotherm depth of 26°C (upper), 20°C (middle) and 15°C (bottom) are shown for GOFS 3.1 (blue), GOFS 3.5 (black), ESPC 7-day forecast (green) and ESPC 16-day forecast (red). The gray dashed line represents GDEM climatology. The average number of observations/day are shown on the right axis and represented by the open circles in the respective colors. All models beat climatology, although the 15°C isotherm depth reaches climatology at the end of the forecast.

#### 4.1.5 Frontal Placement Validation

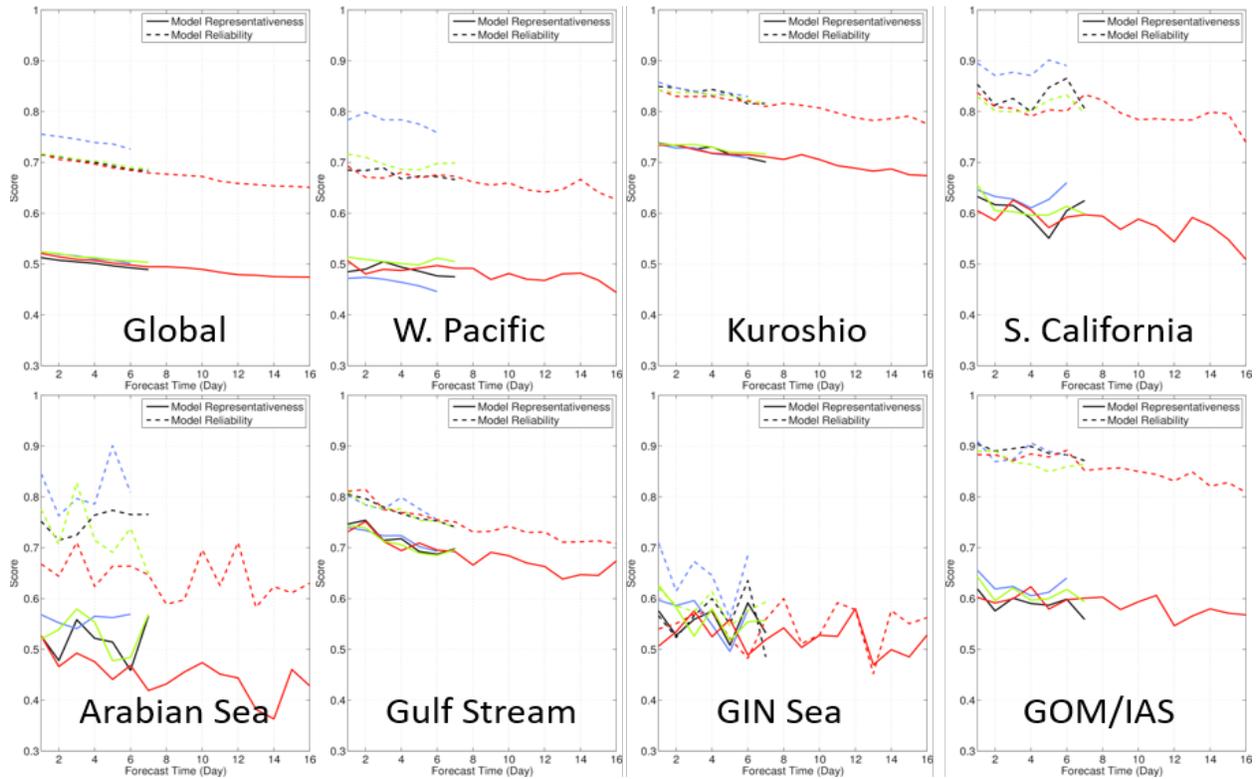
A new metric has been developed to identify fronts in SSH from altimetry and models (Douglass and Mask, 2019). This technique has two metrics: Representativeness, which is how many fronts found in the altimeter are found in the model and Reliability, which is how many fronts found in the model are found in the altimeter. In the upper panel of Figure 41, the two statistics are shown for the ESPC analyses from Jan-July, 2017. On any given day around the globe, approximately 72% of the modeled fronts are found in the altimeter data, while approximately 51% of the altimetric fronts are found in the model. In the lower panel, the same metrics are sorted by the strength of the fronts. For large frontal amplitudes, greater than 0.4m/100 km, the two metrics are nearly identical. The differences between the different approaches occur for small amplitude fronts with more small amplitude fronts in the altimetric data than in the model.



**Figure 41. Top panel: Frontal Placement Representativeness (R1, blue), the number of fronts found in the altimeter which also are found in the model and Frontal placement Reliability (R2, red), the number of fronts found in the model which also are found in the altimeter. The altimeter has more fronts than the model produces. Lower panel: The reliability (red) and representativeness (blue) scores sorted by the strength of the fronts in m/100km. For fronts stronger than 0.4m/100km, the two scores are similar, while the altimeter has more weaker-fronts leading to a lower representativeness score.**

In Figure 42 the performance of the frontal placement over the forecast time is shown for the eight validation regions and the full suite of forecasts from GOFS 3.1, GOFS 3.5 and Navy ESPC. For Navy ESPC, two sets of forecasts were performed, a 7-day forecast launched every 5 days to correspond with the GOFS 3.1 and 3.5 forecasts and a 16-day forecast launched every week on Wednesday to correspond with the ESPC ensembles. In most regions, the current operational ocean and sea ice forecasting system, GOFS 3.1, has a higher reliability score which is consistent with the lower resolution model producing fewer weak fronts. Navy ESPC forecasts have slightly higher representativeness scores than GOFS, but the difference is not statistically significant.

Over the extended range forecasts, the reliability and representativeness scores have a modest degradation in the frontal location skill of 2-4%.



**Figure 42. Frontal placement skill scores for the eight analysis regions. Green lines are Navy ESPC 7-day forecast, red lines are Navy ESPC 16-day forecast, black lines are pre-operational GOFS 3.5 and blue lines are operational GOFS 3.1. Solid lines indicate model representativeness (R1) while dashed lines indicate model reliability (R2). GOFS 3.1 has a higher reliability score consistent with the lower resolution model producing fewer weak fronts. Navy ESPC has a slightly higher representativeness score, but the difference between models is not statistically significant. Over the extended range forecasts, the reliability and representativeness scores have a modest degradation in the frontal location skill of 2-4%.**

#### 4.1.6 Tidal Elevation Error Analysis

The ocean models in Navy ESPC Deterministic and GOFS 3.5 have a major change in their physics compared to previous ocean operational models (GOFS 3.1 and earlier) with the addition of the astronomical gravitational forcing of the sun and moon, self-attraction and loading of the ocean tidal displacement (SAL), topographic wave drag for the barotropic tidal currents and a statistical error correction for model bathymetry, elevation and drag deficiencies based upon an augmented state ensemble Kalman filter (Ngodock et al. 2016). These models now predict the tidal currents and displacements concurrently with the ocean circulation. The amplitudes and phases of the semi-diurnal lunar ( $M_2$ ) barotropic tide are shown in Figure 43. Visually the maps are very similar with the biggest differences in the structure of the North Atlantic and Western Pacific amphidromes. In Figure 44, the errors between the models and TPXO8 inverse tidal solutions are shown. The total RMSE is 2.57 cm for Navy ESPC and 2.59 cm for GOFS 3.5 as shown

in the upper row with ESPC on the left and GOFS on the right. The RMS amplitude of the deep water  $M_2$  tidal elevation is 33 cm and the best inverse tidal solutions have RMSE of 0.5 cm relative to 151 deep water pressure gauges. Following Shriver et al (2012), the RMSE can be broken into errors associated with the amplitude of the tide (middle row) and errors associated with the amplitude-weighted phase (bottom row). The RMSE is approximately split evenly between amplitude only and amplitude-weighted phase errors with the amplitude only errors of 1.73 cm and 1.74 cm for ESPC and GOFS 3.5, respectively, and amplitude-weighted phase errors of 1.90 cm and 1.93 cm for ESPC and GOFS 3.5, respectively. The differences in the two model tidal solutions are not significant.

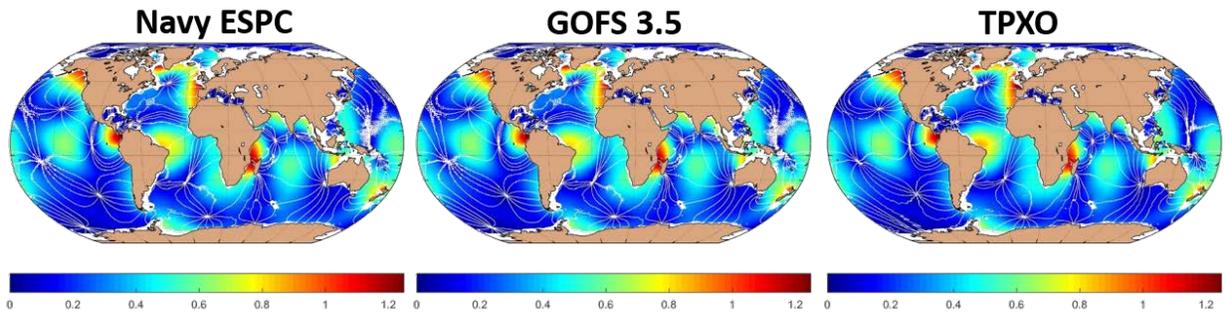


Figure 43. The amplitude (m) and phase of the  $M_2$  barotropic tidal elevation for Navy ESPC (left), GOFS 3.5 (middle) and an inverse tidal solution TPXO8 (right).

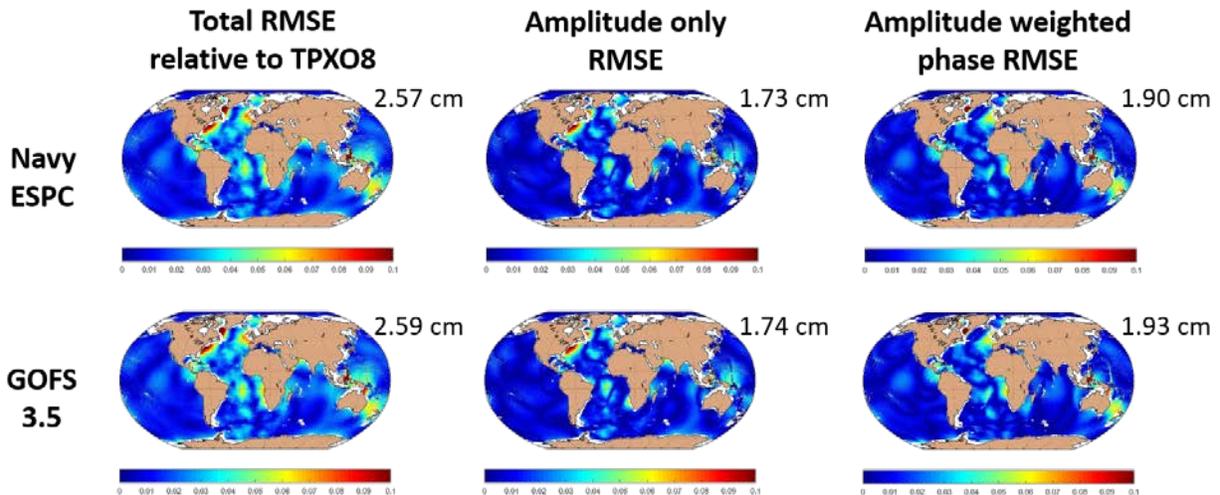
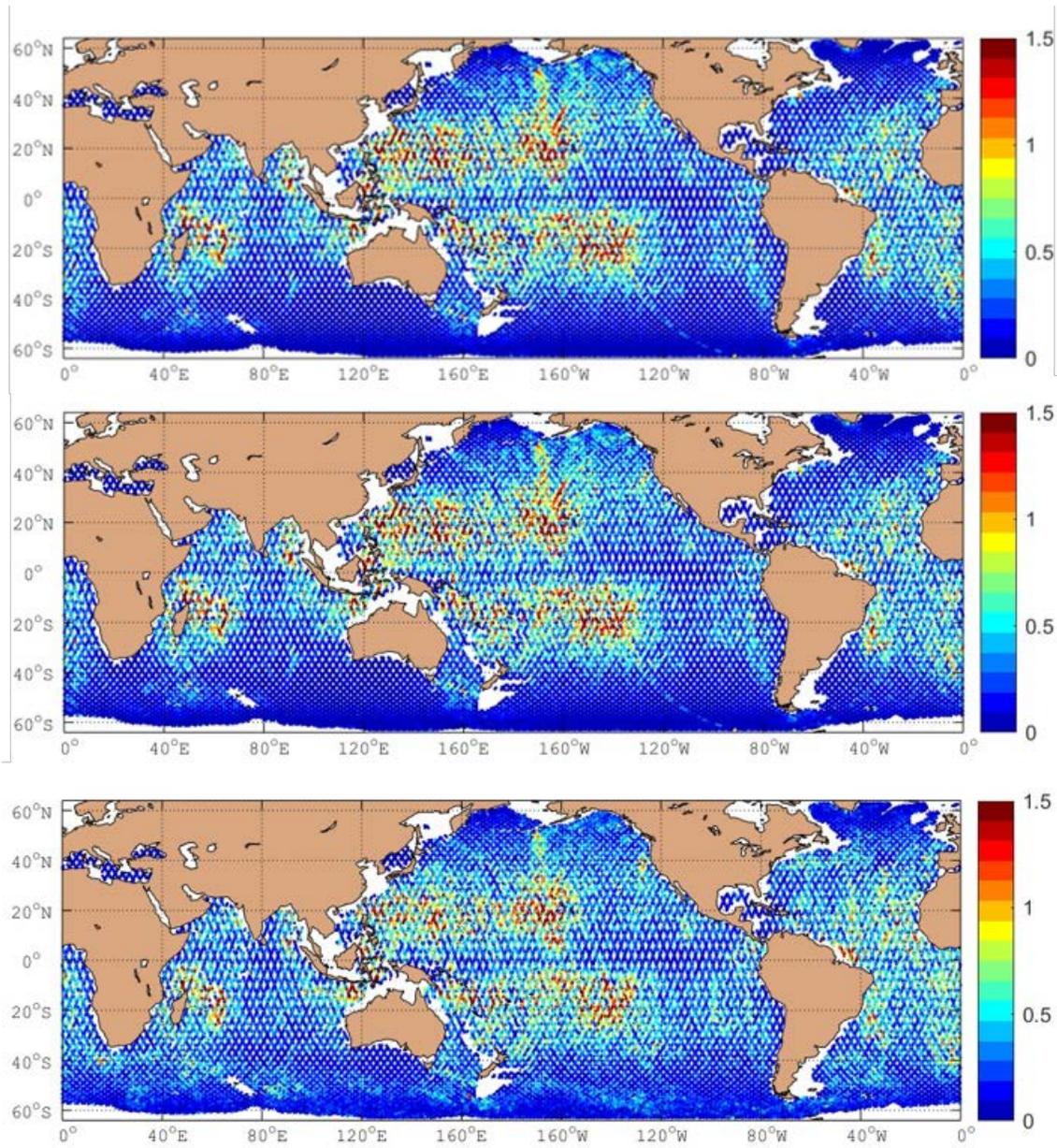


Figure 44. The differences between the Navy ESPC  $M_2$  tidal elevation (top row) and the GOFS 3.5  $M_2$  tidal elevation (bottom row) and the inverse tidal solution TPXO8. The left column is the total RMSE with ESPC having an error of 2.57 cm and GOFS 3.5 having an error of 2.59 cm. The middle column is the RMSE associated with the amplitude only of the  $M_2$  tidal coefficients and the right column is the amplitude weighted phase error (see Shriver et al., 2012 for the definitions).

The barotropic tides generate internal tides through an interaction with bathymetry. The SSH displacements associated with the internal tides generally are small around 1 cm and can be

estimated using satellite altimetry (Shriver et al 2014). The internal tides are strongest near a limited number of generation regions such as Hawaii, French Polynesia and Madagascar as can be seen in Figure 45 for the models and altimetry. In Table 5, the areal averaged amplitudes for regions around the generation hotspots are given. Generally, the model amplitudes are approximately 10 % greater than the satellite altimetric estimates. Based on the results of Ansong et al. (2015) this is an expected bias. They found that areal averaged amplitudes decrease as a function of record length, asymptoting near a record length of about 4 years.



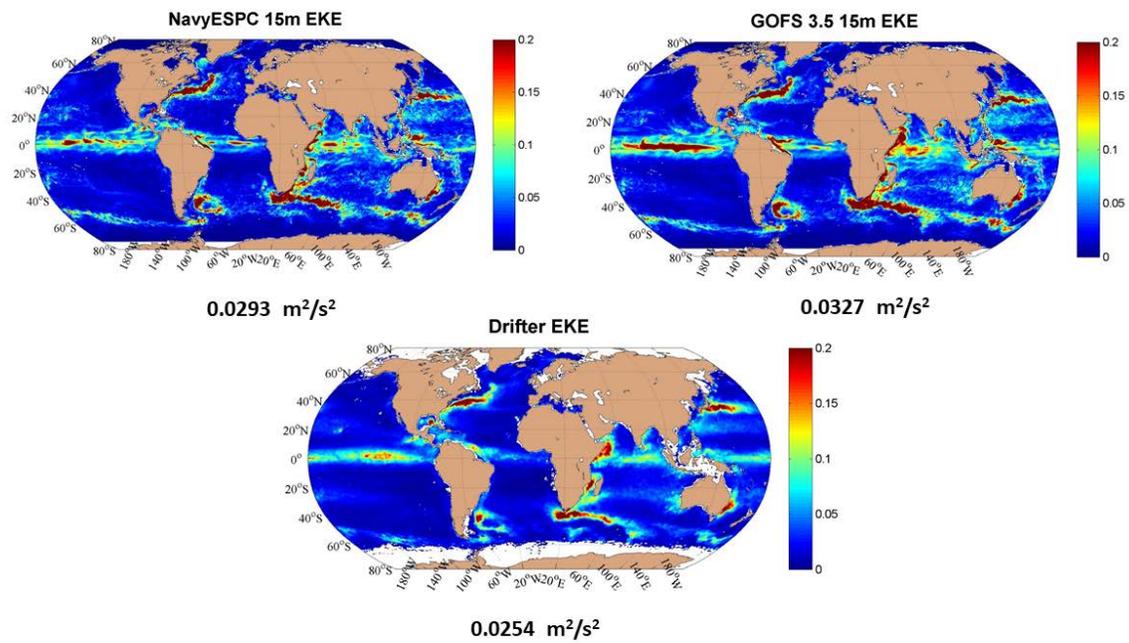
**Figure 45.** Amplitude of the stationary  $M_2$  baroclinic tidal elevation (cm) for Navy ESPC (top), GOFS 3.5 (middle) and satellite altimetry SSH estimate (bottom, Shriver et al., 2014). The two model amplitudes are nearly identical and approximately 10% greater than the satellite estimate (see Table 5).

**Table 5. Amplitudes (cm) of the M2 internal tidal elevation from satellite altimetry SSH, Navy ESPC and GOFS 3.5. The model amplitudes are about 10% greater than the satellite estimates.**

Box locations	Satellite based	ESPC	GOFS3.5
NE Pacific	0.79	0.86	0.86
NW Pacific	0.79	0.99	0.99
SE Pacific	0.84	0.88	0.88
SW Pacific	0.73	0.75	0.74
Madagascar	0.71	0.81	0.81

#### **4.1.7 Eddy Kinetic Energy Climatology**

The upper ocean (15 m) eddy kinetic energy (EKE) for the two high resolution systems, Navy ESPC and GOFS 3.5, is compared to an independently estimated climatology of EKE from nearly 20,000 drifting buoys deployed over 40 years (Figure 46). The model EKE estimates are obtained from a single year, 2017. The spatial patterns of the EKE are similar between the models and the drifting buoy climatology, although the models have stronger western boundary currents and tropical currents. Overall the models have a higher globally averaged EKE than the drifting buoys. Navy ESPC is 17 % higher than the drifting buoys, GOFS 3.5 is 29% higher and GOFS 3.1 (not shown) is 9% higher. Some of the differences may be due to sampling. The drifting buoy climatology is based upon 40 years of observations with variable spacing, while the model estimates are averages for a single year. Also, drifting buoys tend to be dispersed in strong current regions, leading to an underestimate of the EKE in strong currents such as the western boundary currents and tropics. Scatterplots (Figure 47) of observed vs model EKE indicate that GOFS 3.5 is more highly correlated with the observations than Navy ESPC.



**Figure 46: Maps of the mean Eddy Kinetic Energy (EKE,  $\text{m}^2/\text{s}^2$ ) at 15 m from Navy ESPC (top left), GOFS 3.5 (top right) and drifting buoy archive (bottom). The basin-wide average is listed below each panel. Both models have more EKE than the drifting buoys with Navy ESPC 17% greater and GOFS 3.5 29% greater. Some of the differences may be due to sampling. The model estimates cover one year, 2017, while the observations include nearly 20,000 buoys over more than 40 years.**

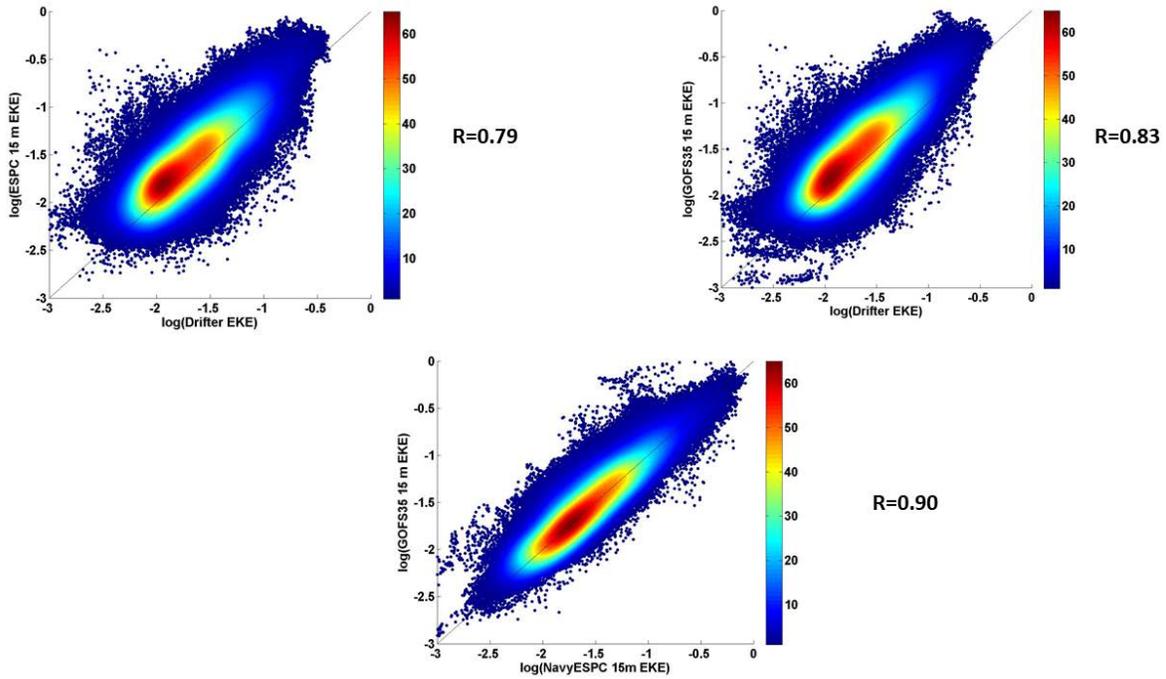


Figure 47: Scatterplots of 15 m EKE for Navy ESPC and Drifter (top left), GOFS 3.5 and Drifter (top right) and Navy ESPC and GOFS 3.5 (bottom). GOFS 3.5 has a higher spatial correlation with the drifters than Navy ESPC, while the two models are highly correlated.

#### 4.1.8 Ocean metric scorecard

To quantify the performance of GOFS 3.1, GOFS 3.5 and Navy ESPC, an ocean score card developed by Zamudio et al. (2015) is employed. To track the system’s performance of most ocean metrics, errors are translated into scores via the following. Let  $Error\_model(i)$  be the model error for the individual ocean metrics ( $i$ ), and model is one of the three systems. And, let

$$\begin{aligned}
 Maximum\_error(i) &= \text{largest of } Error\_model(i), \\
 Score\_model(i) &= 1 - (Error\_model(i) / Maximum\_error(i)) \\
 Total\_score\_model &= \text{mean } (Score\_model(i))
 \end{aligned}$$

be the maximum error over 2017 for each individual region, the scores of the models per each individual metric, and the total score per each model, respectively. A model score of 1 indicates perfect skill relative to the observations, whereas a model score of 0 indicates no skill. The difference of the scores shows the relative performance to each other and allows a determination as to which system is more skillful. RMSE is the metric that contributes to the score for T vs. depth, S vs. depth, MLD, SLD, BLG, and depths of the 26°C, 20°C and 15°C isotherms. The isotherm error is only used for the global domain, not the regional domains. The drifting buoy score uses RMSE of speed and additionally the inverse of vector correlation.

Figure 48 shows the monthly scores for the global analysis region at the “nowcast” time. The upper left panel shows *Total\_score\_model* while the other panels show *Score\_model(i)*. Note that *Score\_model(i)* is not an absolute value (i.e. skill of 0.1 does not imply it is only accurate 10% of the time) but a relative measure comparing the three systems. Because they are normalized by *Maximum\_error(i)*, one curve always has a value of zero at one point along the x-axis. The systems perform similarly at the nowcast time with a slight edge to GOFS 3.5.

Figure 49 shows *Total\_score\_model* for all eight analysis regions at the nowcast time. The higher horizontal resolution Navy ESPC and GOFS 3.5 perform better in the western boundary current regions (Gulf Stream and Kuroshio) than lower resolution GOFS 3.1. GOFS 3.5 performs better than Navy ESPC for the globe and some Navy relevant regions (Western Pacific, Arabian Sea, and GIN Sea). Navy ESPC outperforms the other systems in the Gulf Stream, Southern California and Gulf of Mexico regions. Note that because each region is normalized by the maximum error of that specific region, it is not possible to compare the relative skill of one analysis region with another. That can be done by normalizing by the maximum error over all regions (not shown), but that can give the false impression of high skill if one region happens to be particularly unskillful.

Similar to Figure 48, Figure 50 shows *Total\_score\_model* and *Score\_model(i)* over the global analysis region for each of the metrics, but here as a function of forecast length. For the global domain, GOFS 3.5 typically outperforms Navy ESPC at most forecast days. This is especially true for sonic layer depth.

Figure 51 shows *Total\_score\_model* for the eight analysis regions as a function of forecast length. Here *Score\_model(i)* is normalized by *Maximum\_error(i)* across all forecast days. *Maximum\_error(i)* is typically largest for the day six forecast. The results are somewhat mixed with Navy ESPC performing broadly similar to GOFS 3.5.

For the interested reader, Appendix B shows *Score\_model(i)* as a function of forecast length for various metrics and all analysis regions.

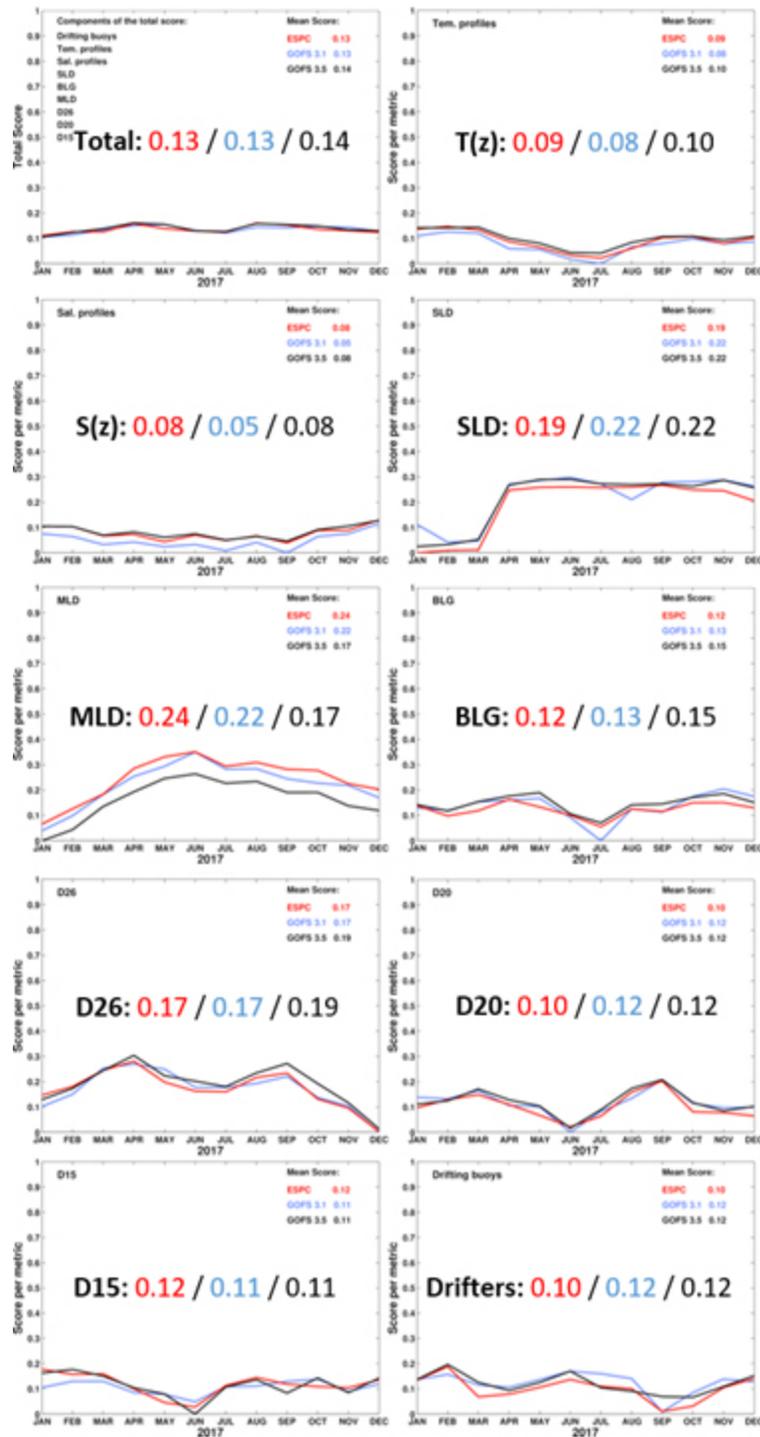


Figure 48: Monthly scores of the global (50°S–50°N) analysis region at the “nowcast” time for *Total\_score\_model* (the composite of all nine metrics) in the upper left panel, and *Score\_model(i)* in the other panels for each metric: T(z) = temperature versus depth, S(z) = salinity versus depth, SLD = sonic layer depth, MLD = mixed layer depth, BLG = below layer gradient, D26 = depth of 26°C isotherm, D20 = depth of 20°C isotherm, D15 = depth of 15°C isotherm, and Drifters = combined speed and vector correlation. Higher numbers are better. The yearly average is noted in each panel for GOFS 3.1 (blue), GOFS 3.5 (black) and Navy ESPC (red). This analysis is based on all days of 2017.

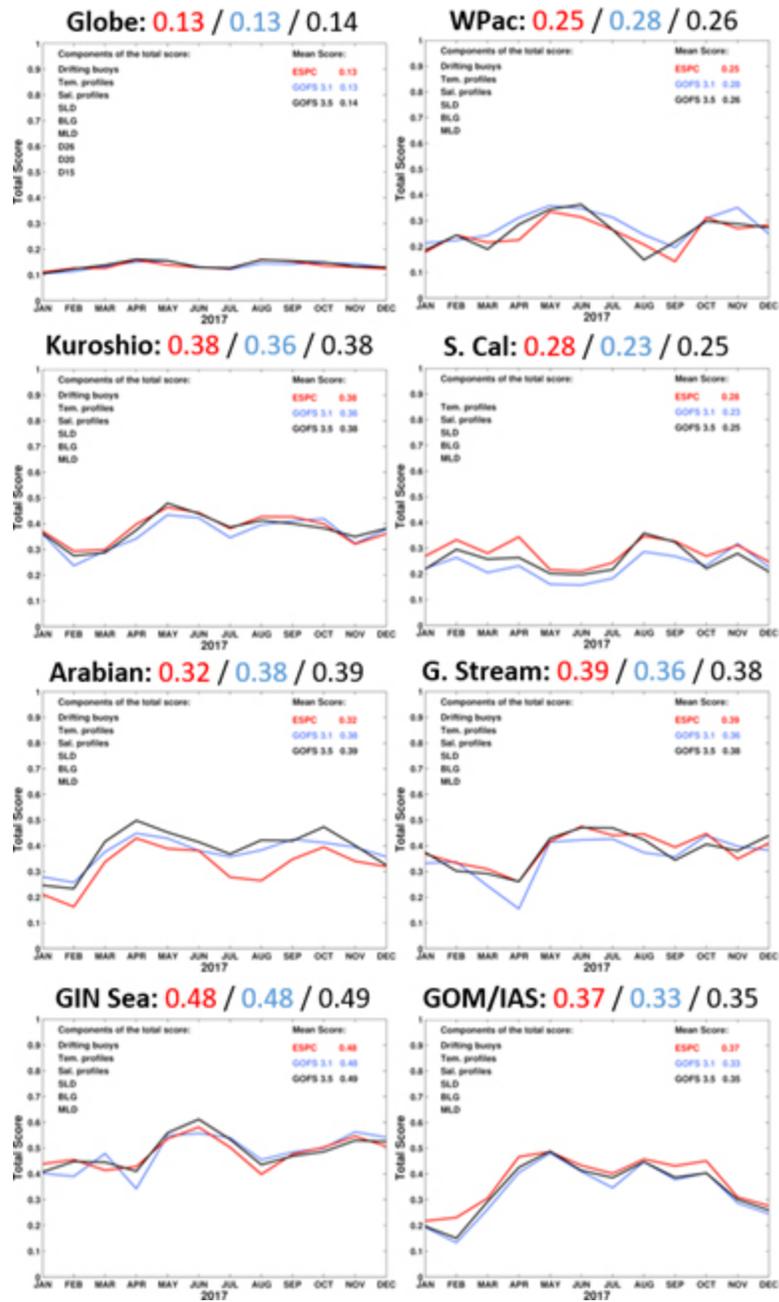


Figure 49: Monthly *Total\_score\_model* for the various analysis regions at the “nowcast” time. The yearly average is noted above each panel for GOF3.1 (blue), GOF3.5 (black) and Navy ESPC (red). Because each region is normalized by the maximum error of that specific region, these do not compare the relative skill of one analysis region with another. This analysis is based on all days of 2017.

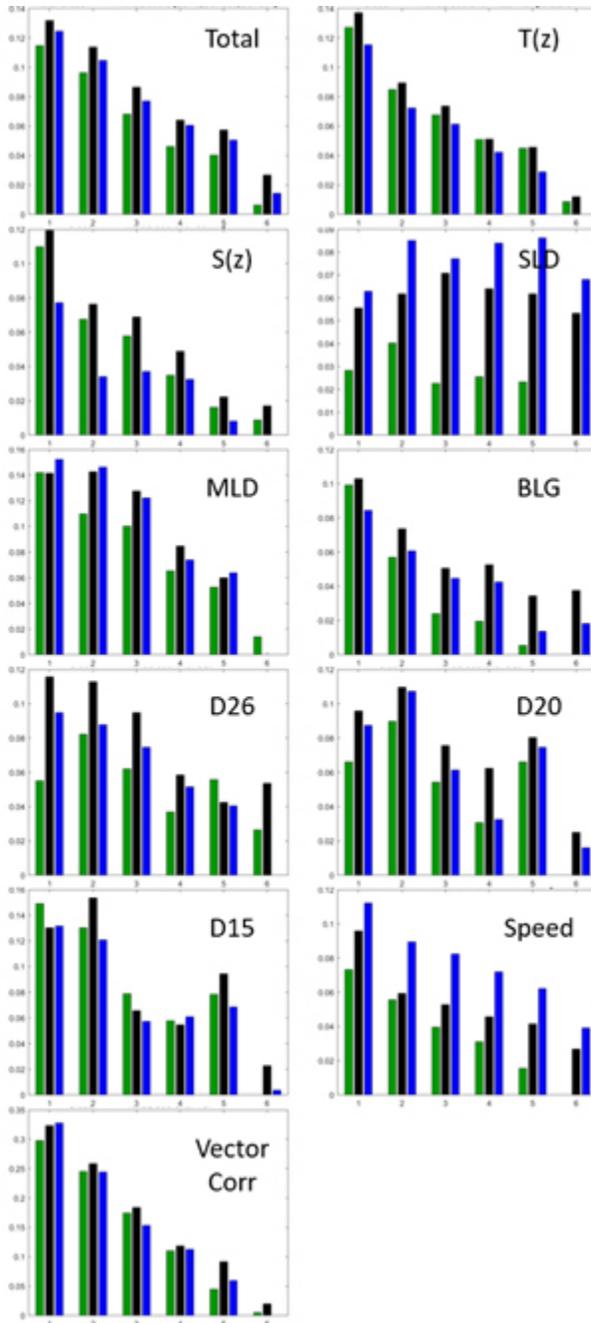


Figure 50: Scores of the global (50°S-50°N) analysis region as a function of forecast length (days) for *Total\_score\_model* (the composite of all nine metrics) in the upper left panel, and *Score\_model(i)* in the other panels for each metric: T(z) = temperature versus depth, S(z) = salinity versus depth, SLD = sonic layer depth, MLD = mixed layer depth, BLG = below layer gradient, D26 = depth of 26°C isotherm, D20 = depth of 20°C isotherm, D15 = depth of 15°C isotherm, Speed = ocean drifter speed and Vector Corr = inverse of ocean vector correlation. Higher numbers are better. The colors are GOFS 3.1 (blue), GOFS 3.5 (black) and Navy ESPC (green). This analysis is based on 72 forecasts starting on the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, 16<sup>th</sup>, 21<sup>st</sup> and 26<sup>th</sup> of each month.

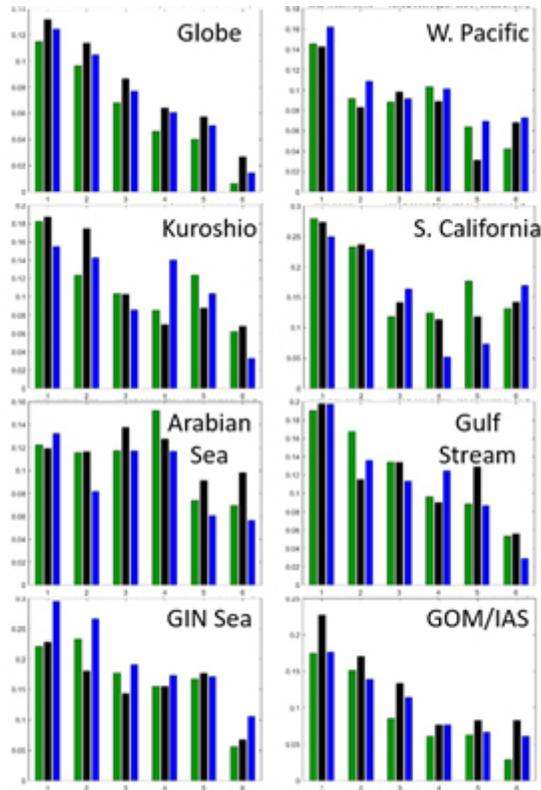
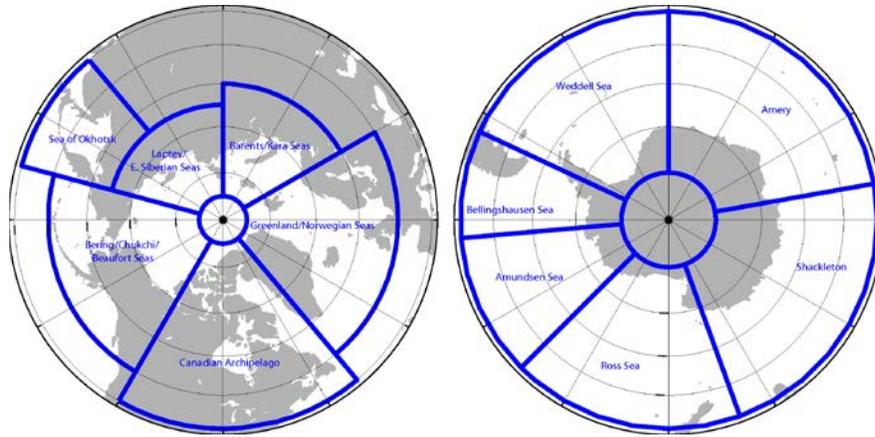


Figure 51: *Total\_score\_model* for the various analysis regions as a function of forecast length (days) for GOFS 3.1 (blue), GOFS 3.5 (black), and Navy ESPC 7-day forecasts (green). Because each region is normalized by the maximum error of that specific region, these do not compare the relative skill of one analysis region with another. This analysis is based on 72 forecasts starting on the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, 16<sup>th</sup>, 21<sup>st</sup> and 26<sup>th</sup> of each month.

## 4.2 Sea Ice Performance

Sea ice distribution and properties are modeled using the Community Ice Code (CICE) in GOFS 3.1, GOFS 3.5 and Navy ESPC. However, the versions of the code are different between the two systems. The Navy ESPC reanalyses were started for the VTR before issues with the latest version of CICE, version 5.1.2, had been resolved. Thus, Navy ESPC uses an older version of CICE, version 4.0. GOFS 3.1 also uses CICE v4.0. As will be seen below, the differences between the versions of CICE is significant with GOFS 3.5 and CICE v5.1.2 performing better almost everywhere than Navy ESPC and GOFS 3.1 with CICE v4.0 for the sea ice edge metric.

For the sea ice performance evaluation, the Arctic and Antarctic will be divided into multiple regions, shown in Figure 52, chosen to correspond to the regional seas as well as the entire Pan-Arctic and Pan-Antarctic domains.



**Figure 52. Arctic (left) and Antarctic (right) ice validation analysis regions chosen to correspond to the regional sea in each hemisphere. The Central Arctic region (not labeled) is the small circle from 85°N to the North Pole. It is not used in the sea ice edge analysis but is used in the sea ice drift analysis**

Both GOFs and ESPC assimilate the same sea ice concentration data, thus at the “nowcast” time the differences between the two systems are small. The analysis corrections are directly inserted into the model fields at 09Z and forecasts run forward from there. By 00Z, just outside the data assimilation window, differences begin to appear and increase with forecast length. In Figure 53, Figure 54, Figure 55 and Figure 56 the Arctic sea ice concentration (top row) and thickness (bottom row) are shown for four snapshots at 12Z for late winter March 15, late spring June 15, end of melt season September 15, and early winter December 15 for GOFs 3.5 (left column) and ESPC (right column). Small differences in the sea ice concentration can be seen in the Beaufort and Barents Seas and the thick ice along the Canadian Archipelago, but overall the sea ice fields are very similar at the analysis time. Because ESPC was initialized from GOFs 3.5 in mid-December 2016, the sea ice thickness is very similar and remains consistent through the year.

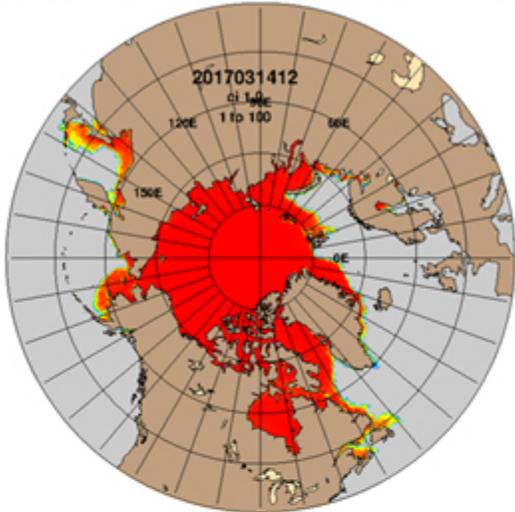
#### 4.2.1 Sea ice edge error

Maritime operations in the polar regions rely upon an accurate prediction of the sea ice edge. The differences in ice concentration in the marginal ice zones affect the location of the ice edge. The National Ice Center (NIC) produces a daily ice edge analysis for the Arctic and Antarctic using a variety of satellite sources (visible images, infrared, scatterometer, passive microwave, and synthetic aperture radar (SAR) data) to define the ice edge as areas of <10% sea ice concentration. Model sea ice edge locations are defined as those grid points that exceed a certain threshold value (5%) for ice concentration and have a neighboring point with ice concentration below the threshold. The distances between each NIC observed point and the nearest model-derived ice edge location are calculated. In Table 6, the differences in the ice edge error are listed for GOFs 3.5 and ESPC for 15 hours after the analysis insertion (00Z, tau 12) and 39 hours after analysis insertion (00Z +1 day, tau 36). This analysis is performed for every day of 2017. GOFs with CICE v5.1.2 performs better than ESPC with CICE v4.0 over the Pan-Arctic and regional seas with an average reduction in the sea ice edge error of ~7% at tau 12 and ~10% at tau 36. The biggest differences are along the Siberian coast and Laptev Sea and the Canadian Archipelago.

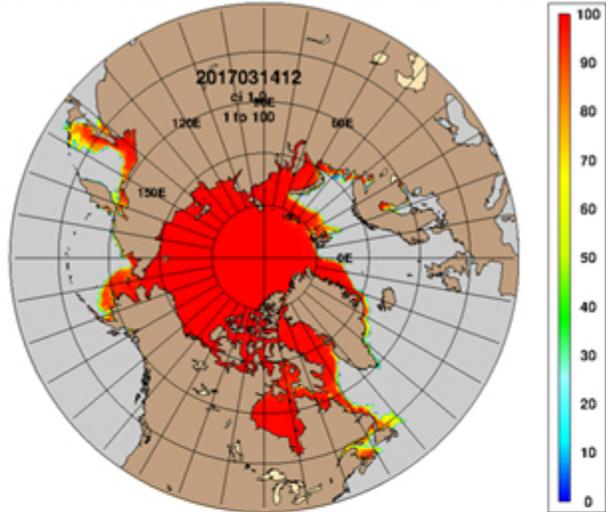
As seen in Figure 57, the differences in the 24-hour forecast ice edge errors have a seasonal variation. The largest Pan-Arctic differences occur during the fall freeze-up and winter with much smaller differences in the spring and summer. Similar results are found for the regional seas. The Sea of Okhotsk is ice-free during the late summer/early fall months, and ice edge error is exceptionally high with the first new ice of the season. The Laptev Sea is completely ice covered throughout most of the year except for the late summer/fall months. Thus sea ice edge error can only be computed in these months.

In Figure 58, the sea ice edge error as a function of forecast day is shown. This analysis uses the 72 7-day forecasts for GOFS 3.1/3.5 and ESPC, and the 52 16-day forecasts for ESPC. GOFS 3.5 using CICE v5.1.2 generally performs better than any of the GOFS 3.1 or ESPC CICE v4.0 forecasts. However, at taus 024 and 048 NAVY ESPC has similar skill to GOFS 3.5 for a few regions (GIN Sea, Barents/Kara Seas, and Beaufort Sea). Note for the Pan-Arctic and some regions the error is higher at the nowcast time and decreases as forecast length increases (out to about tau 048) before eventually increasing again. We hypothesize this is due to the order in which the Interactive Multisensor Snow and Ice Mapping System (IMS) sea ice mask is applied, i.e. after the NCODA sea ice analysis completes. IMS typically doesn't extend as far equatorward as the NIC sea ice edge and so the nowcast time may have higher error. Work is ongoing to bring IMS inside NCODA and only apply it to passive microwave sensors that are known to suffer from non-differentiation of sea ice and summertime melt ponds. The trend of ice edge error for the 16-day forecasts is similar to the 7-day forecasts.

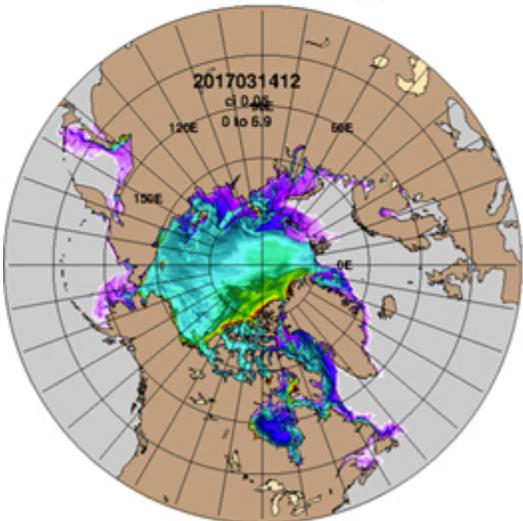
GLBc0.04-21.6 Ice Concentration (%): 20170315



GLBc0.04-10.0 Ice Concentration (%): 20170315



GLBc0.04-21.6 Ice Thickness (m): 20170315



GLBc0.04-10.0 Ice Thickness (m): 20170315

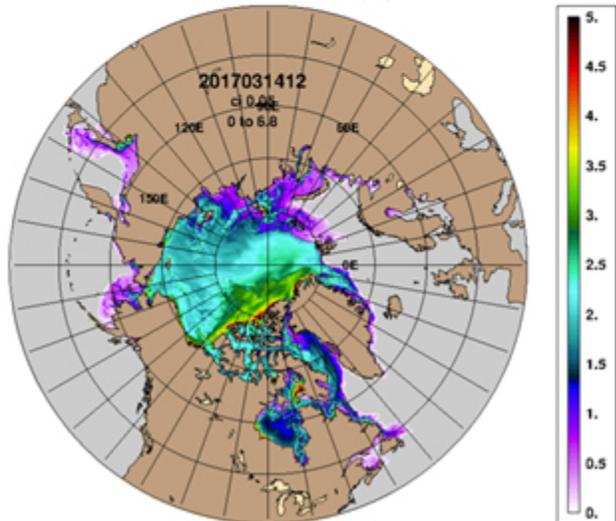
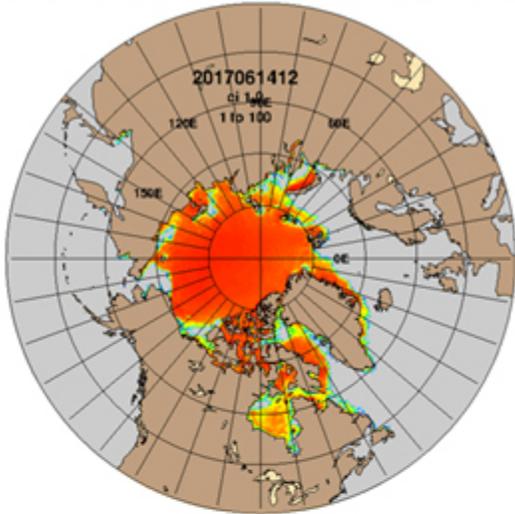
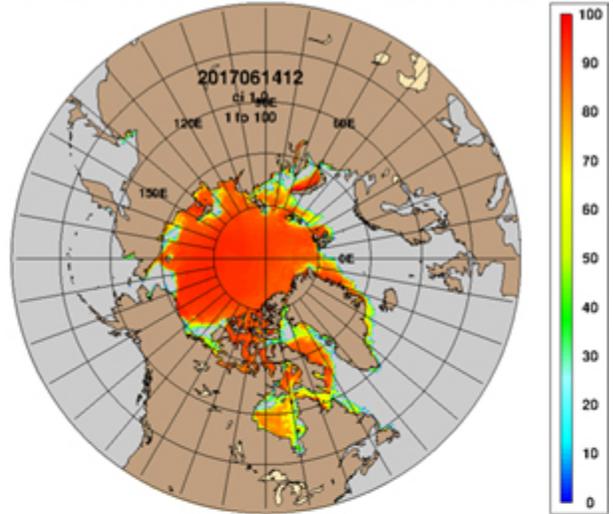


Figure 53. Pan-Arctic sea ice concentration (% top) and thickness (m, bottom) on 15 March 2017 for GFS 3.5 (left) and Navy ESPC (right).

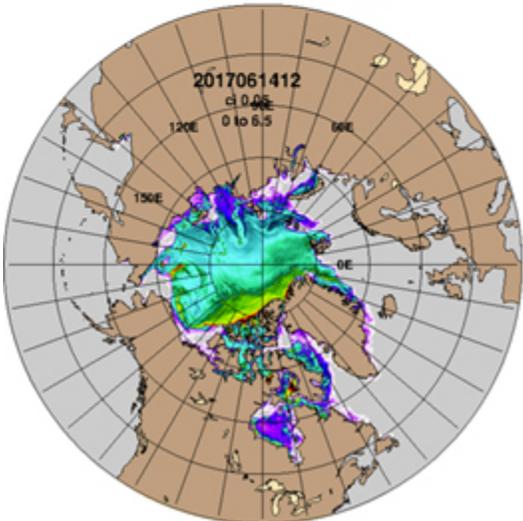
GLBc0.04-21.6 Ice Concentration (%): 20170615



GLBc0.04-10.0 Ice Concentration (%): 20170615



GLBc0.04-21.6 Ice Thickness (m): 20170615



GLBc0.04-10.0 Ice Thickness (m): 20170615

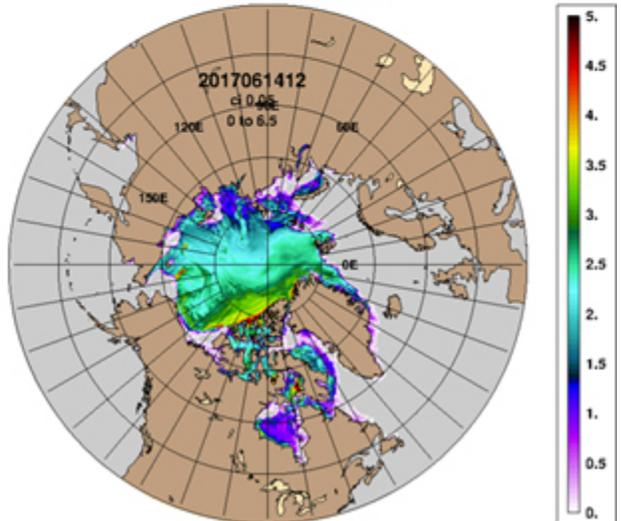
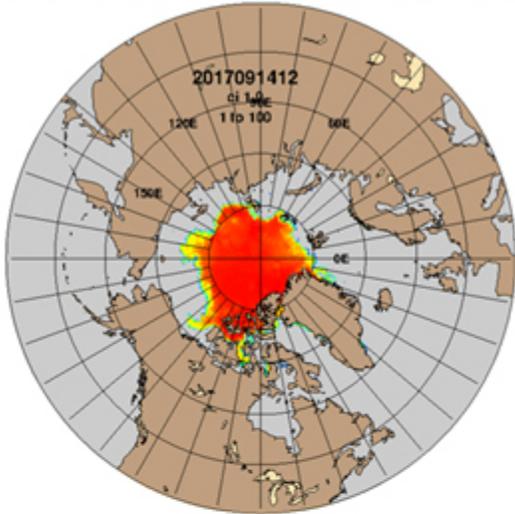
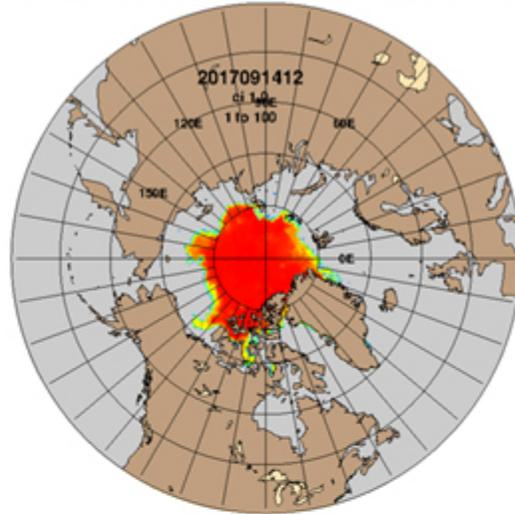


Figure 54: Same as Figure 53 except for 15 June 2017.

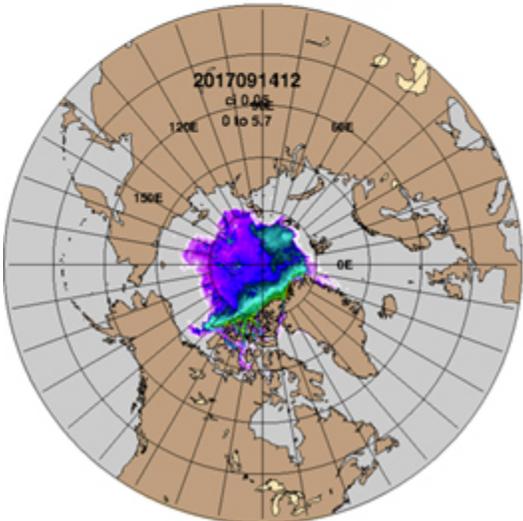
GLBc0.04-21.6 Ice Concentration (%): 20170915



GLBc0.04-10.0 Ice Concentration (%): 20170915



GLBc0.04-21.6 Ice Thickness (m): 20170915



GLBc0.04-10.0 Ice Thickness (m): 20170915

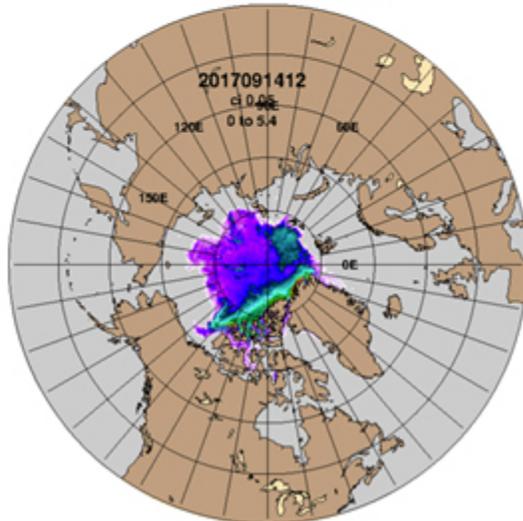
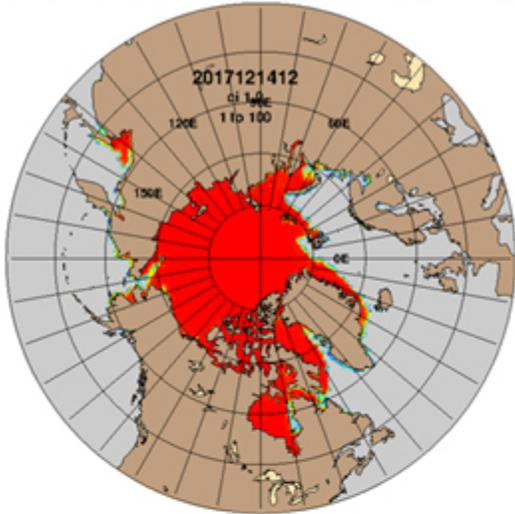
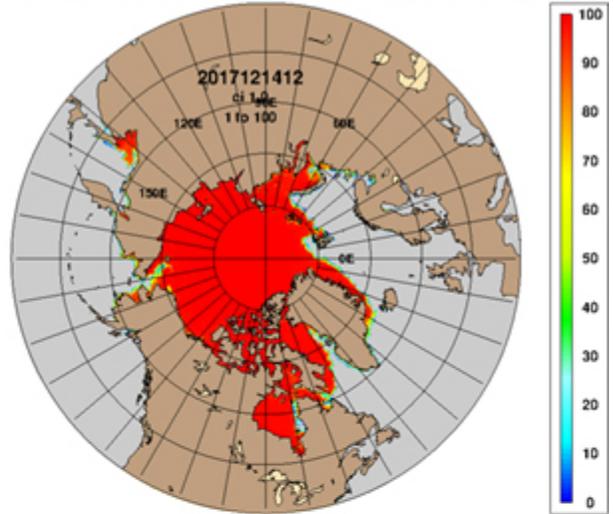


Figure 55: As Figure 53 except for 15 September 2017.

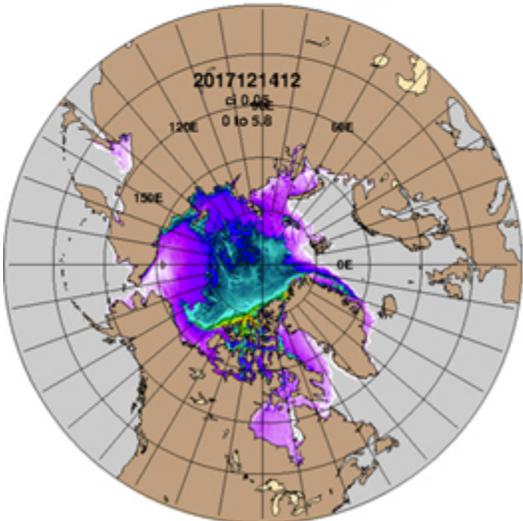
GLBc0.04-21.6 Ice Concentration (%): 20171215



GLBc0.04-10.0 Ice Concentration (%): 20171215



GLBc0.04-21.6 Ice Thickness (m): 20171215



GLBc0.04-10.0 Ice Thickness (m): 20171215

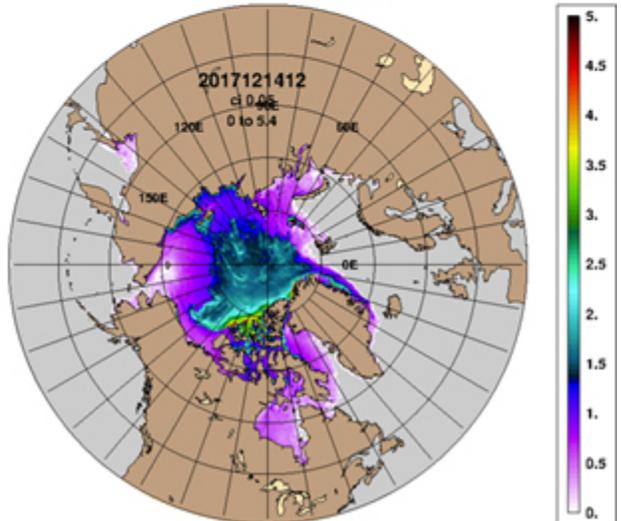


Figure 56: As Figure 53 except for 15 December 2017.

**Table 6. Sea ice edge error (km) relative to the independent NIC ice edge for the Pan-Arctic and the regional seas show in Figure 52. Lower error is highlighted in green. These statistics are based on every day of the 2017 reanalysis for each system.**

<b>Region</b>	<b>Tau</b>	<b>GIFS 3.5 Error (km)</b>	<b>Navy ESPC Error (km)</b>	<b>Percent Difference</b>
Pan-Arctic	12	23.5	25.2	-7.3
	36	21.4	23.6	-10.3
Greenland/Norwegian	12	24.6	25.9	-5.5
	36	20.8	22.1	-6.6
Barents/Kara	12	21.8	22.3	-2.3
	36	21.5	22.7	-5.5
Laptev/E. Siberian	12	25.0	29.6	-18.5
	36	23.4	30.6	-31.6
Sea of Okhotsk	12	18.6	19.8	-6.4
	36	17.1	19.2	-12.5
Bering/Chukchi /Beaufort	12	20.6	21.1	-2.4
	36	19.9	20.6	-3.5
Canadian Archipelago	12	26.5	31.0	-17.0
	36	22.6	27.3	-21.2
“Wins”		14	0	

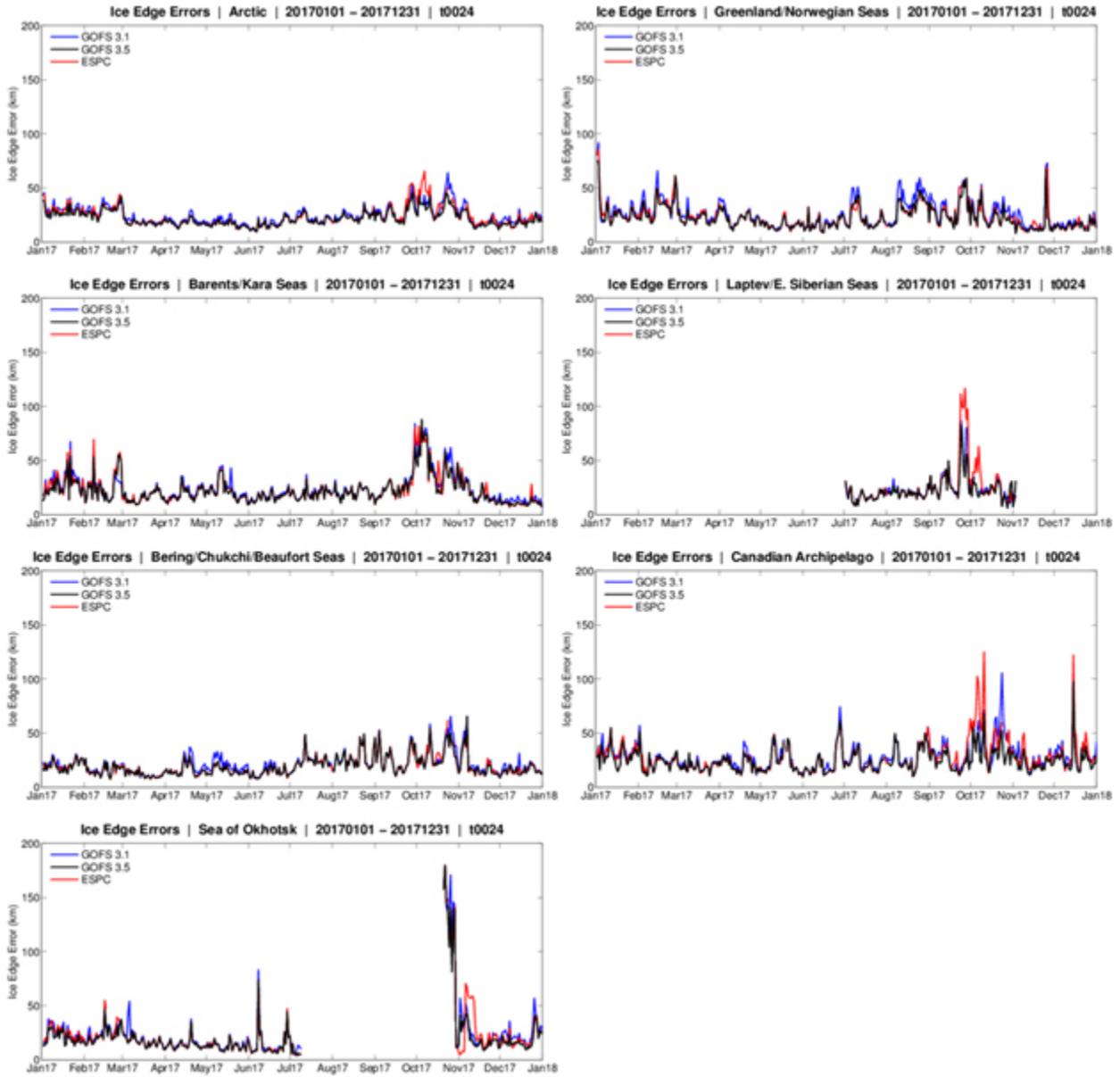
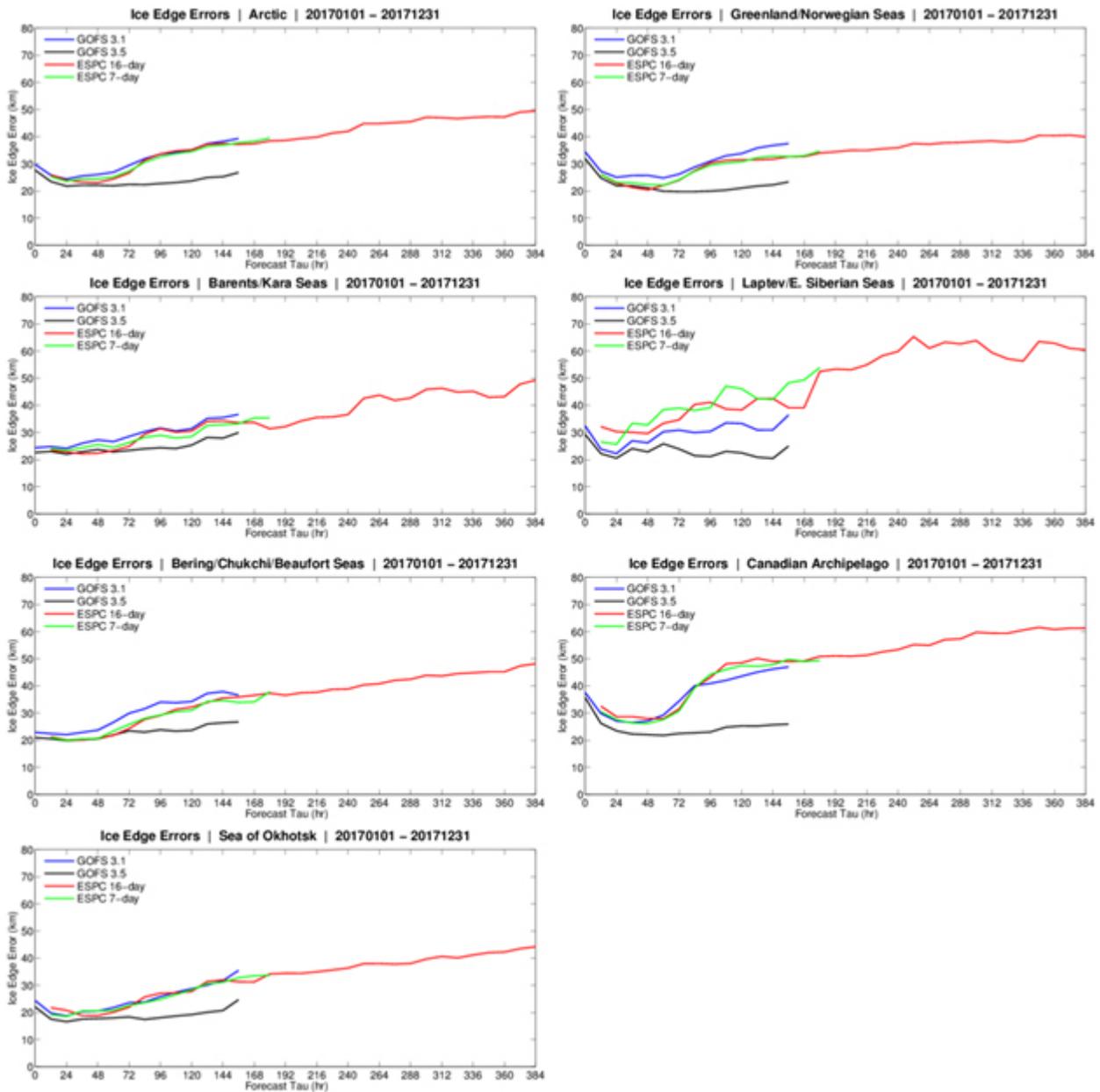


Figure 57. Daily ice edge errors (km) for the Pan-Arctic and regional domains shown in Figure 52 for GOFs 3.1 (blue), GOFs 3.5 (black) and Navy ESPC (red) at the 24-hour forecast time.



**Figure 58: Northern Hemisphere sea ice edge error (km) vs. the independent NIC ice edge as a function of forecast length for calendar year 2017 for GOF3.1 (blue), GOF3.5 (black), deterministic ESPC 7-day forecasts (green) and 16-day forecasts (red) for the Pan-Arctic and regional domains shown in Figure 52. ESPC and GOF3.1 use CICE v4.0 whereas GOF3.5 uses CICE v5.1.2.**

Similar analyses can be performed for the Antarctic. In Figure 59, Figure 60, Figure 61, and Figure 62 the sea ice concentration and thickness for late summer (March 15), fall (June 15), winter (September 15) and spring (December 15) are shown for GOF3.5 and Navy ESPC, respectively. Thickness differences occur in the thin first year ice around the Shackleton and Amery regions, but they are small. As shown in Figure 63, there is a seasonal variation in the differences between the models. The Pan-Antarctic sea ice edge error differences are typically largest in austral summer and fall. Unlike the Arctic, GOF3.5 and CICE v5.1.2 don't always perform better in all

regions. GOFS 3.5 generally outperforms ESPC at short forecasts (Table 7), but GOFS 3.1 often beats GOFS 3.5 for unknown reasons (Figure 63) in this hemisphere.

As the forecast length increases, the differences between GOFS 3.5 with CICE v5.1.2 and ESPC with CICE v4.0 increase (Figure 64). But for several regions, the error of operational GOFS 3.1 is on par with GOFS 3.5. This suggests that in the Southern Ocean, the deficiencies of the untuned higher resolution atmosphere in the coupled system are having a negative impact on sea ice edge errors.

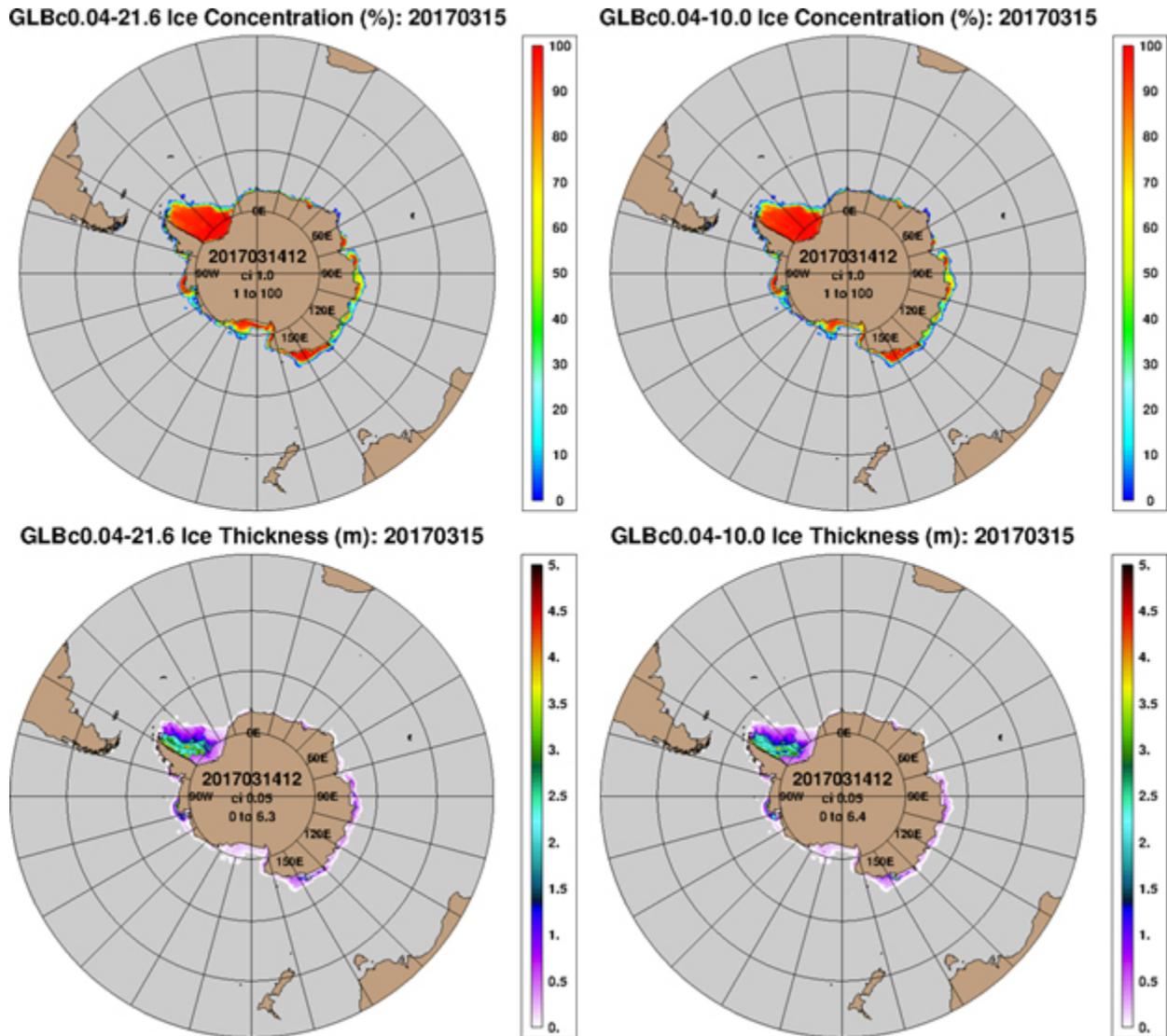
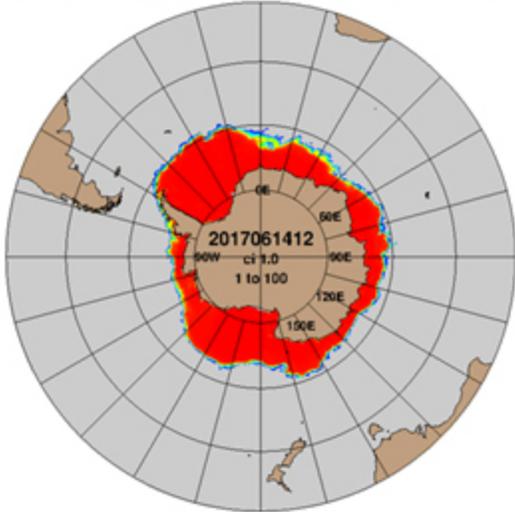
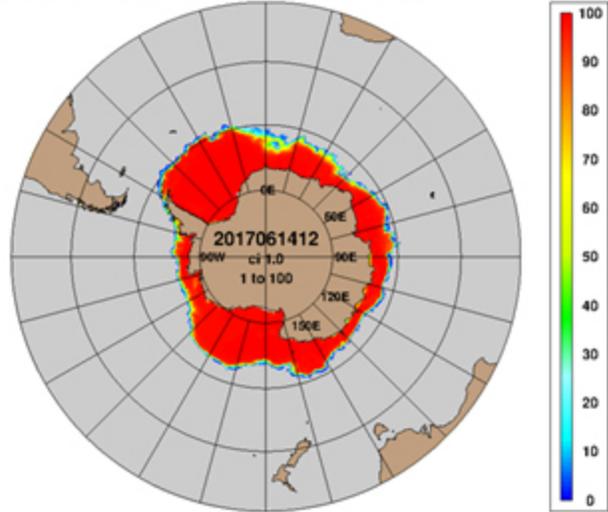


Figure 59. Antarctic sea ice concentration (% , top) and thickness (m, bottom) on 15 March 2017. Shown are GOFS 3.5 (left column) and Navy ESPC (right column).

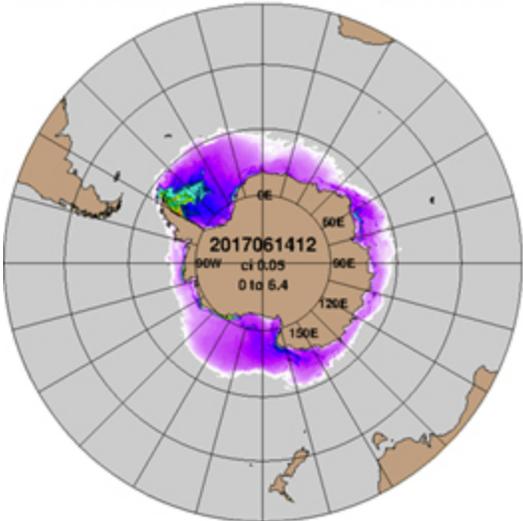
GLBc0.04-21.6 Ice Concentration (%): 20170615



GLBc0.04-10.0 Ice Concentration (%): 20170615



GLBc0.04-21.6 Ice Thickness (m): 20170615



GLBc0.04-10.0 Ice Thickness (m): 20170615

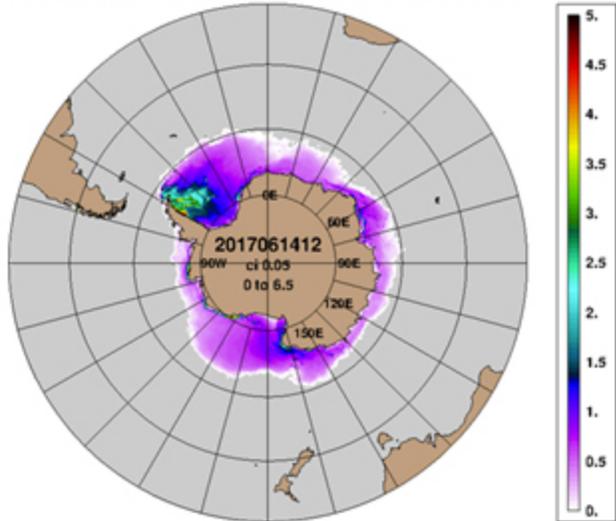
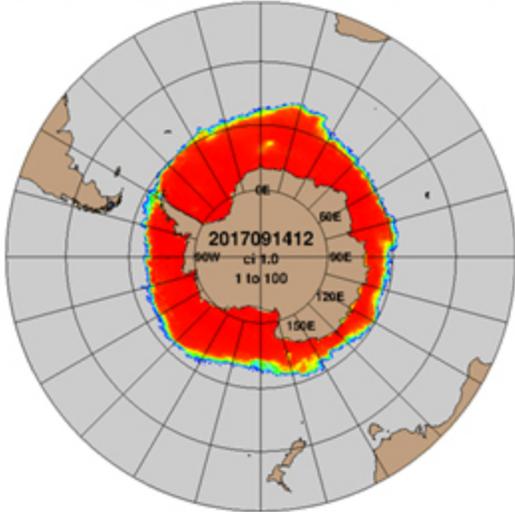
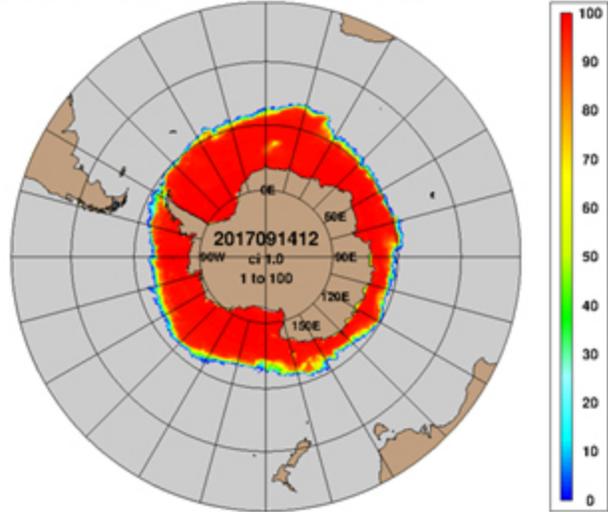


Figure 60: As Figure 59 but for 15 June 2017.

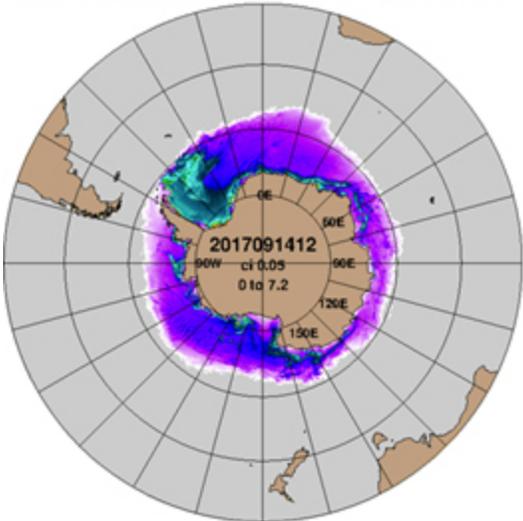
GLBc0.04-21.6 Ice Concentration (%): 20170915



GLBc0.04-10.0 Ice Concentration (%): 20170915



GLBc0.04-21.6 Ice Thickness (m): 20170915



GLBc0.04-10.0 Ice Thickness (m): 20170915

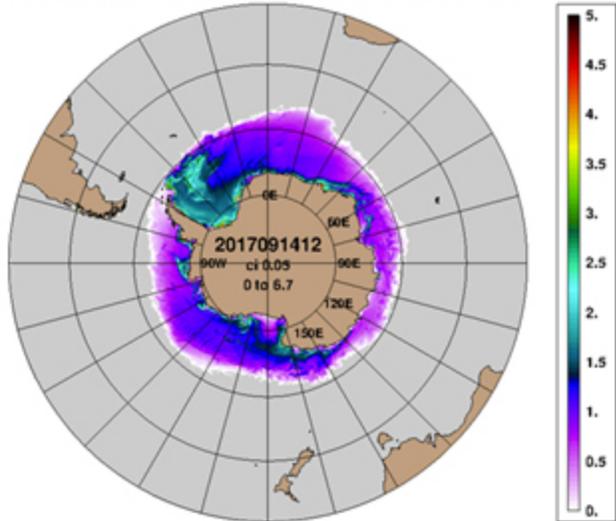
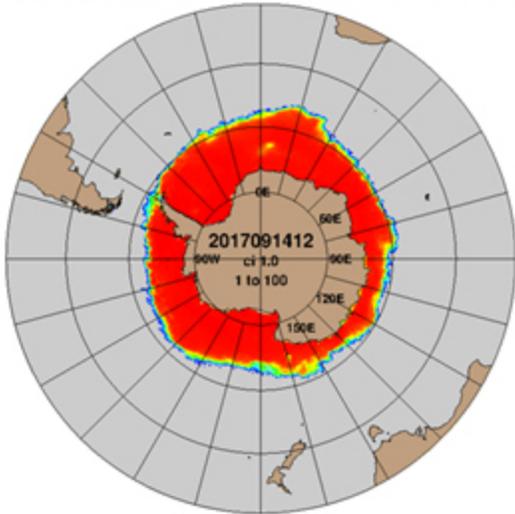
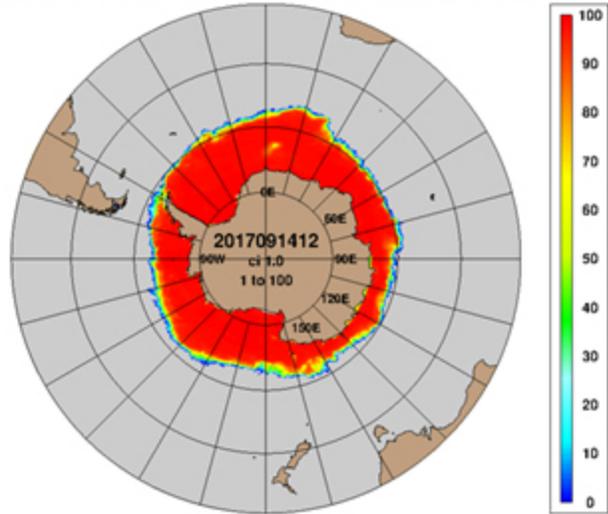


Figure 61: As Figure 59 but for 15 September 2017.

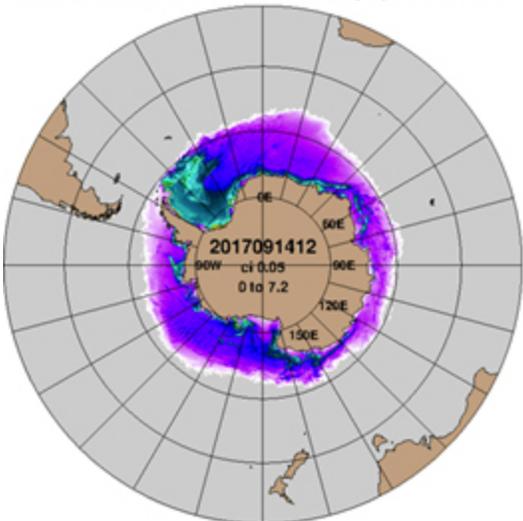
GLBc0.04-21.6 Ice Concentration (%): 20170915



GLBc0.04-10.0 Ice Concentration (%): 20170915



GLBc0.04-21.6 Ice Thickness (m): 20170915



GLBc0.04-10.0 Ice Thickness (m): 20170915

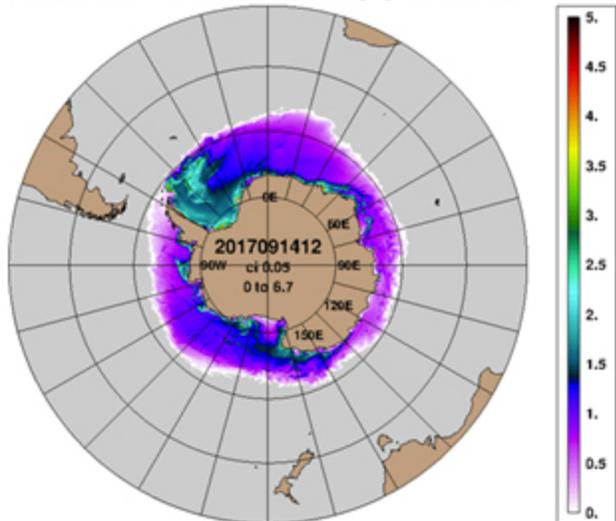


Figure 62: As Figure 59 but for 15 December 2017.

**Table 7. Sea ice edge error (km) relative to the independent NIC ice edge for the Pan-Antarctic and the regional seas shown in Figure 52. Lower error is highlighted in green. These statistics are based on every day of the 2017 reanalysis for each system.**

<b>Region</b>	<b>Tau</b>	<b>GOFs 3.5 Error (km)</b>	<b>Navy ESPC Error (km)</b>	<b>Percent Difference</b>
Antarctic	12	36.8	39.4	-7.0
	36	35.9	40.4	-12.5
Amery	12	39.5	46.2	-16.9
	36	38.1	46.8	-22.8
Shackleton	12	27.2	30.1	-10.7
	36	26.1	31.7	-21.3
Ross	12	40.0	39.3	1.7
	36	39.2	39.5	-0.8
Amundsen	12	41.4	41.8	-0.8
	36	40.9	43.7	-6.8
Bellingshausen	12	35.7	37.8	-5.9
	36	35.7	39.6	-10.8
Weddell	12	39.7	42.0	-5.8
	36	38.7	42.8	-10.7
"Wins"		13	1	

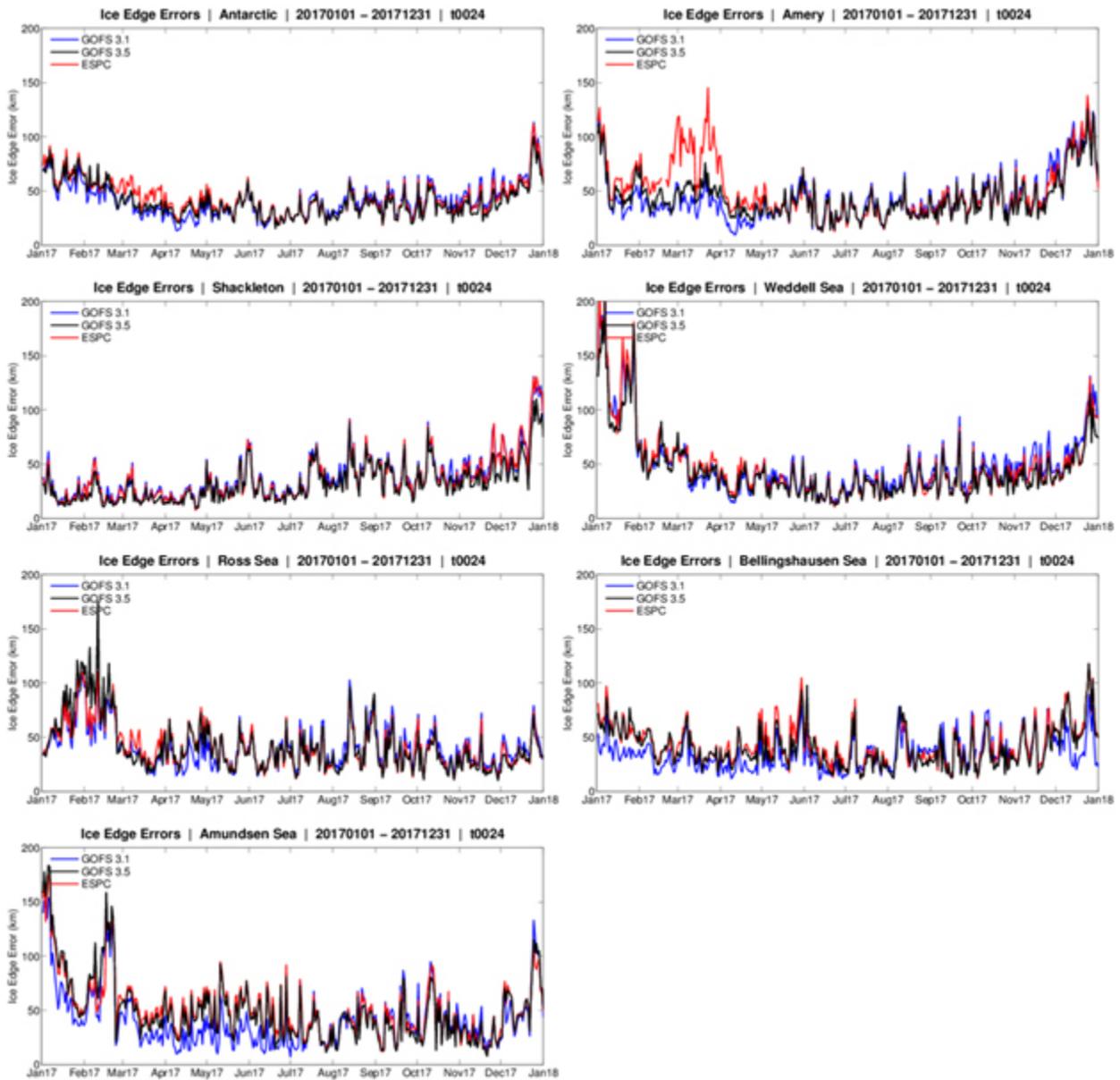


Figure 63: Daily ice edge errors (km) for the Pan-Antarctic and regional domains shown in Figure 52 for GOF3.1 (blue), GOF3.5 (black) and Navy ESPC (red) at the 24-hour forecast time.

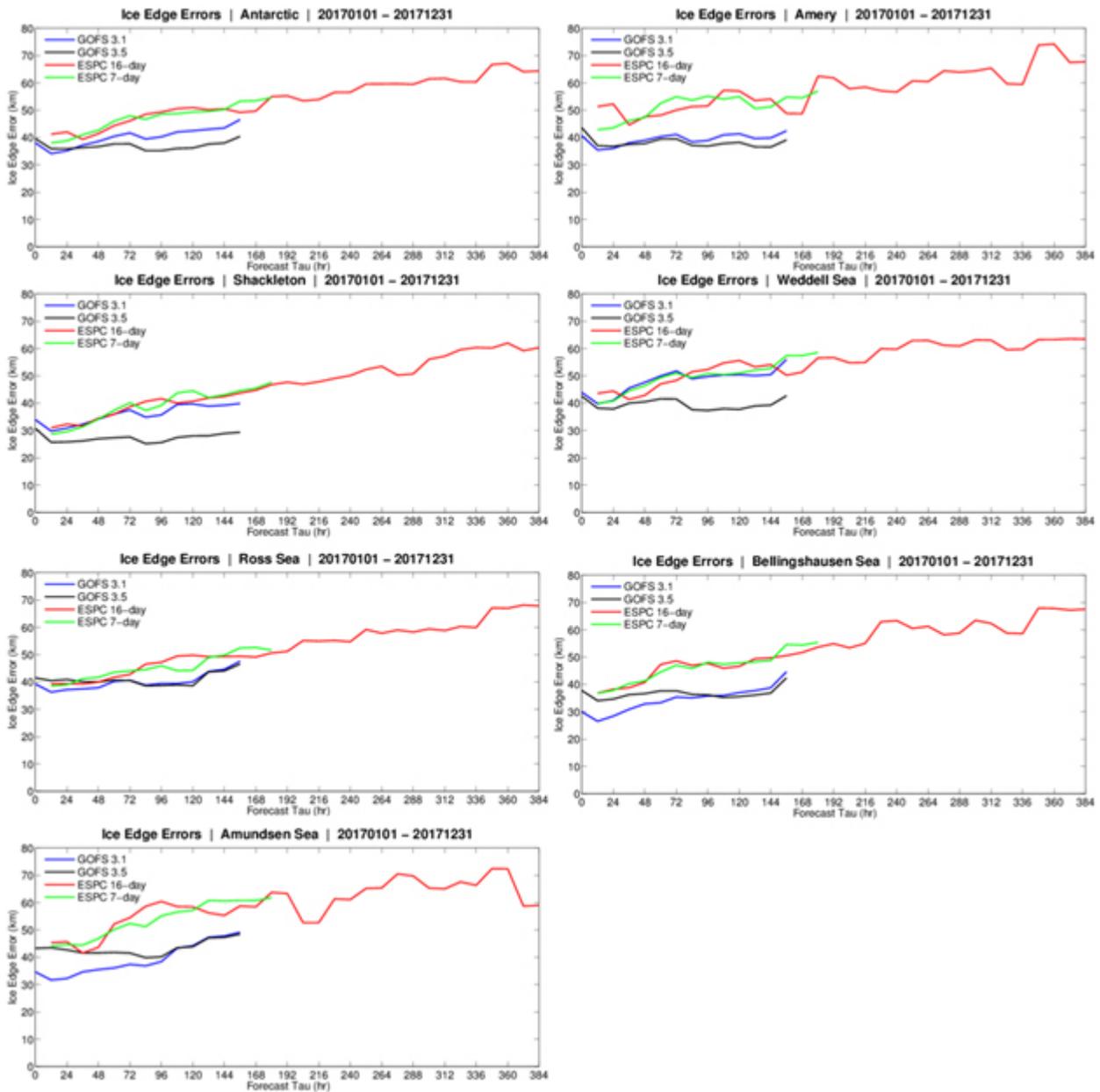
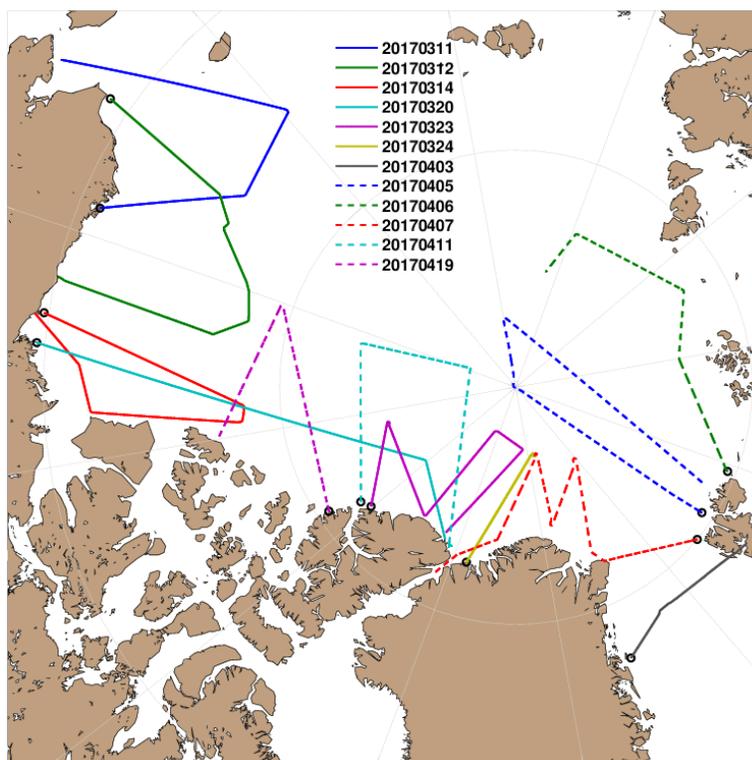


Figure 64: Southern Hemisphere sea ice edge error (km) vs. the independent NIC ice edge as a function of forecast length for calendar year 2017 for GOFs 3.1 (blue), GOFs 3.5 (black), deterministic ESPC 7-day forecasts (green) and 16-day forecasts (red) for the Pan-Antarctic and regional domains shown in Figure 52. ESPC and GOFs 3.1 use CICE v4.0 whereas GOFs 3.5 uses CICE v5.1.2.

#### 4.2.2 Sea ice thickness

The NASA Operation IceBridge (OIB) mission (Kurtz et al., 2013) collects airborne remote sensing measurements to bridge the gap between NASA’s Ice, Cloud and land Elevation Satellite (ICESat) and ICESat-2 missions. Although limited in spatial extent (north of Alaska and the Canadian Archipelago), short in length (typically one day) and only containing a small number (12) of flights,

IceBridge data is used in validating sea ice thickness. Figure 65 shows the IceBridge flight paths for 2017.



**Figure 65: Operation IceBridge flight paths for 2017. The black circle indicates the starting point of each flight.**

The IceBridge flight data have very high temporal resolution. The data were filtered via a running mean that smoothed the time series sufficiently. Several time filters were examined and a six-minute filter produced accurate results. This equates to approximately 45 km of flight path. GOFS 3.1, GOFS 3.5 and ESPC output at the nowcast time were interpolated to these flight paths for the following error analysis.

Figure 66 shows sea ice thickness for eight OIB flights along with GOFS 3.1, GOFS 3.5 and ESPC. The IceBridge data exhibit significantly more spatial variability than any of the forecast systems. Table 8 summarizes the error statistics for these flights. GOFS 3.1 has an overall thin bias whereas GOFS 3.5 and ESPC are on average slightly too thick. Because ESPC was initialized with GOFS 3.5 sea ice on 15 December 2016, there are not large thickness differences between these two systems even though they use a different version of CICE. Sea ice thickness evolves slowly and this comparison is still relatively near the initialization time.

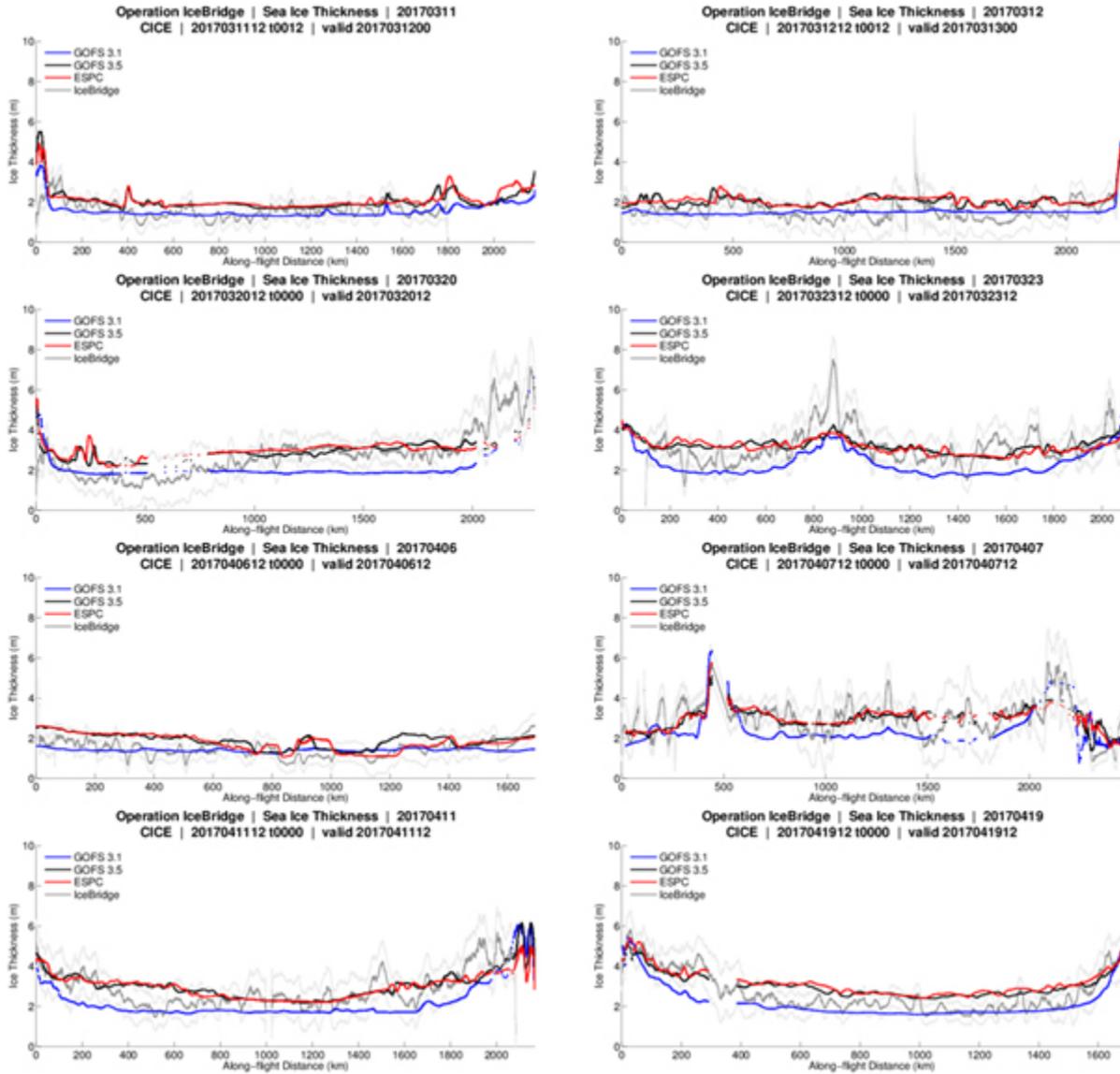


Figure 66: Ice thickness (m) versus time for GOFS 3.1 (blue), GOFS 3.5 (black) and ESPC (red) relative to Operation IceBridge flight data (gray with uncertainty estimate as lighter gray). Each panel represents a different flight as shown in Figure 65.

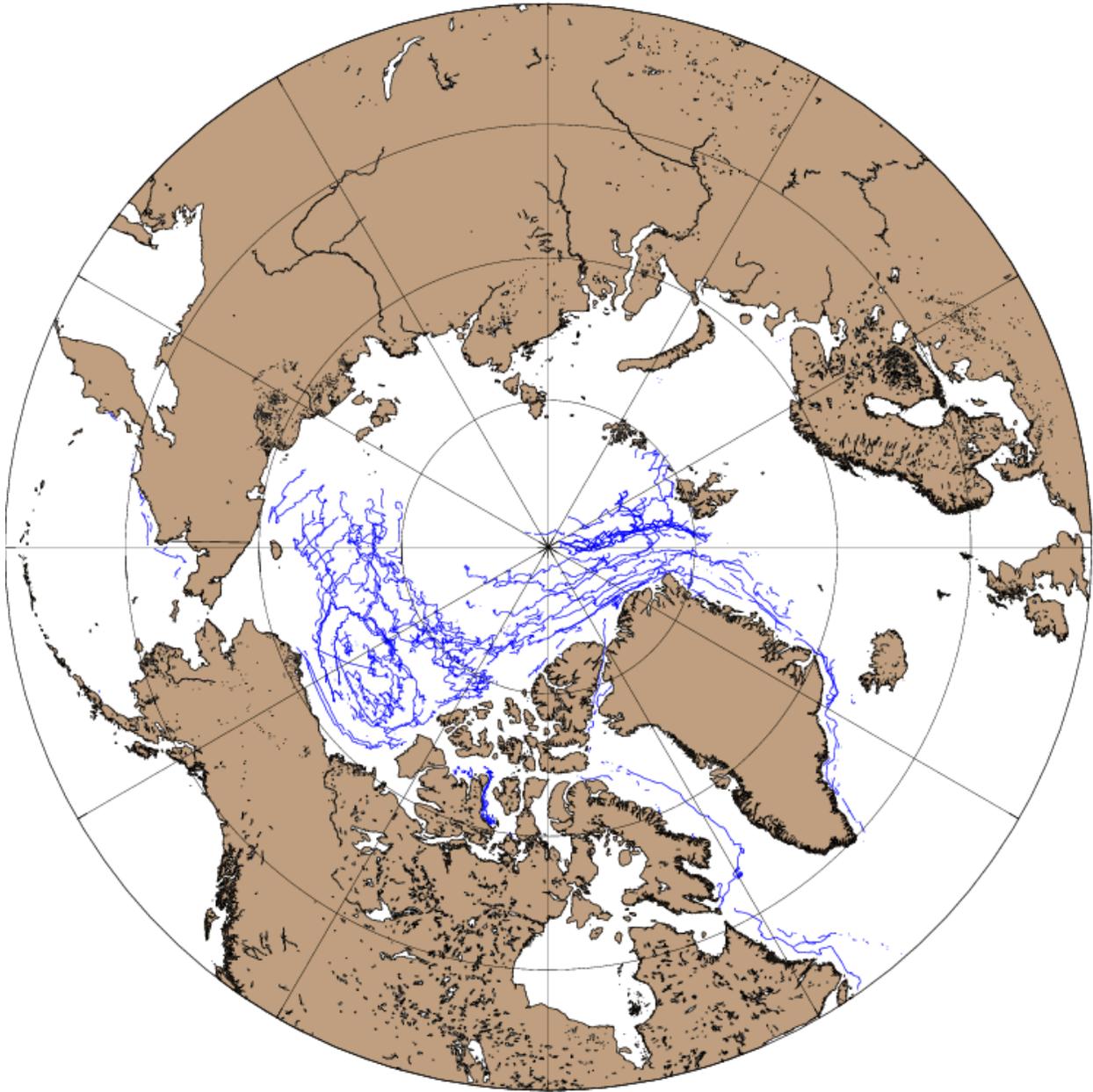
**Table 8: Operation IceBridge sea ice thickness error against GOFs 3.1, GOFs 3.5 and ESPC sampled along the same flight paths and dates for 2017. Units are meters. Lower errors/total “wins” are highlighted in green.**

Flight	Mean Error			Mean Absolute Error			RMSE		
	GOFs 3.1	GOFs 3.5	ESPC	GOFs 3.1	GOFs 3.5	ESPC	GOFs 3.1	GOFs 3.5	ESPC
20170311	-0.10	0.42	0.44	0.32	0.48	0.51	0.46	0.69	0.66
20170312	-0.07	0.44	0.48	0.36	0.52	0.54	0.45	0.61	0.63
20170314	-0.72	-0.24	-0.10	0.92	0.56	0.49	1.08	0.82	0.73
20170320	-0.60	0.33	0.45	0.76	0.46	0.50	0.88	0.61	0.69
20170323	-0.75	0.20	0.17	0.79	0.54	0.55	0.94	0.68	0.73
20170324	-1.12	-0.50	-0.48	1.12	0.50	0.48	1.19	0.58	0.55
20170403	-0.50	-0.14	-0.02	1.06	1.03	1.01	1.42	1.30	1.36
20170405	-0.82	0.05	0.02	0.83	0.40	0.41	0.98	0.54	0.54
20170406	-0.05	0.42	0.34	0.28	0.48	0.43	0.36	0.56	0.54
20170407	-0.59	0.03	0.07	0.86	0.50	0.50	0.99	0.65	0.65
20170411	-0.54	0.35	0.35	0.60	0.50	0.50	0.74	0.65	0.65
20170419	-0.38	0.51	0.62	0.41	0.61	0.68	0.54	0.67	0.75
“Wins”	4	3	6	4	5	5	4	6	5

### 4.2.3 Sea ice drift

Sea ice drift from the three systems is compared against drifting buoys from the International Arctic Buoy Program (IABP; <http://iabp.apl.washington.edu/index.html>). A total of 207 drifting buoys (Figure 67) exist over 2017. Note that the majority of the drifters are north of Alaska and the Canadian Archipelago and so the following regions are used in the analysis: Pan-Arctic, Greenland/Iceland/Norwegian Seas, Canadian Archipelago, Bering/Chukchi/Beaufort Seas and the Central Arctic. From the daily latitude/longitude pairs of the ice-bound drifters, observed ice drift components are derived using the Haversian formula to determine the x- and y-direction distances travelled each day. Results are then converted from km/day to m/s. Model sea ice velocity components are interpolated via cubic splines to the observed positions and model drift is derived from the daily mean ice velocity valid at tau 012.

Similar to the ocean drift error analysis (4.1.3), sea ice speed absolute mean error and RMSE, and sea ice direction vector correlation is computed. Figure 68 shows the monthly variation of these quantities over the Pan Arctic region for the 12-hour forecasts at all ice concentrations. ESPC tends to have the lowest speed bias and often the lowest RMSE, but the vector correlation of direction is always lower than either GOFs.



**Figure 67: The International Arctic Buoy Program ice-bound drifting buoy tracks over calendar year 2017 used in the sea ice drift error analysis. A total of 207 buoys are shown.**

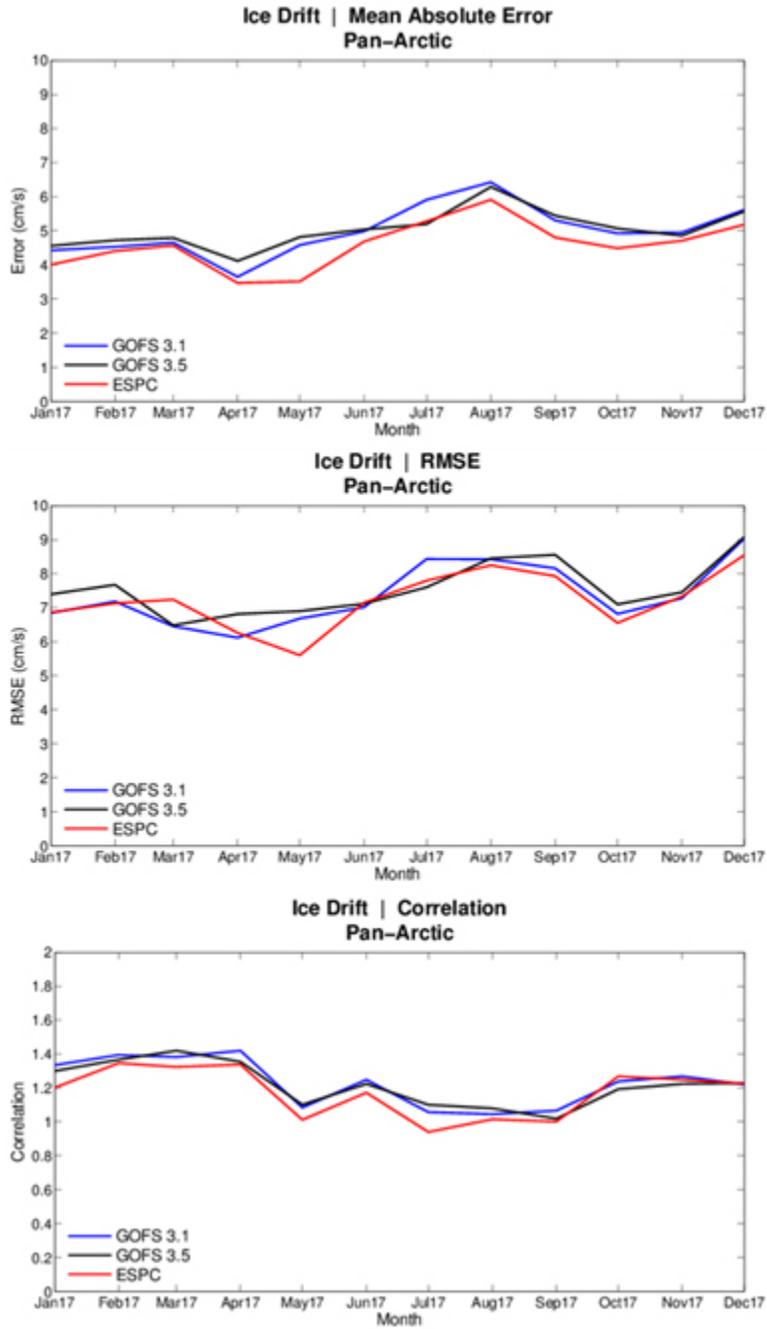


Figure 68: Monthly (top) speed absolute mean error (cm/s), (middle) RMSE (cm/s), and (bottom) direction vector correlation of unassimilated IABP ice-bound drifting buoys for the Pan Arctic region compared against GOFS 3.1 (blue), GOFS 3.5 (black) and ESPC (red) for tau 012 at all ice concentrations. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. This analysis spans all days of 2017.

Table 9 breaks this analysis down by region and again shows that sea ice speed is lowest for ESPC, but the direction correlation is also the worst. Northern Hemisphere sea ice drift is also categorized by ice concentration to evaluate system performance within both the pack ice and the marginal ice zone (Table 10 and Table 11). However, the analysis is only performed for the Pan-Arctic domain since the number of ice bound drifters gets very small or non-existent in the sub-regions. Table 10 shows observed ice drift speed along with the model speed and the bias. Note that in the 0.1-0.3 category there are only six drifters so the statistics may not be robust. Observed sea ice speed increases with decreasing ice concentration and generally the model ice drift is too fast. ESPC has the lowest bias at higher concentrations whereas GOFS 3.1 performs best at the lowest two concentration categories. While Table 9 showed that ESPC had the lowest speed bias, the same is not true when broken down into smaller sea ice concentration categories (Table 11). Here GOFS 3.1 and GOFS 3.5 generally outperform ESPC for all ice drift metrics.

**Table 9: Sea ice drift statistics of speed absolute mean error and RMSE (both in cm/s), and direction vector correlation for IABP ice-bound drifting buoys versus GOFS 3.1 GOFS 3.5 and ESPC for the 12-hour forecast over 2017 at all ice concentrations. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. Lower errors/total “wins” are highlighted with green.**

Region	Absolute Mean Error			RMSE			Vector Correlation		
	GOFS 3.1	GOFS 3.5	ESPC	GOFS 3.1	GOFS 3.5	ESPC	GOFS 3.1	GOFS 3.5	ESPC
Pan-Arctic	4.88	4.94	4.47	7.27	7.44	7.14	1.27	1.26	1.21
Central Arctic	4.45	4.41	3.75	5.47	5.38	4.78	1.65	1.65	1.60
GIN Seas	5.71	5.85	5.79	9.06	9.53	10.27	1.01	0.97	0.90
Canadian Arch.	5.09	5.14	4.45	8.24	8.32	7.40	0.92	0.94	0.92
Bering Beaufort	4.57	4.61	4.16	6.31	6.43	5.93	1.43	1.43	1.39
“Wins”	1	0	4	1	0	4	4	3	0

**Table 10: Sea ice drift speed mean error (cm/s) categorized by ice concentration for IABP ice-bound drifting buoys over the Pan-Arctic region versus GOFs 3.1, GOFs 3.5 and ESPC for the 12-hour forecast over 2017. The sum of the buoys is greater than 207 buoys because the same buoy might appear in more than one ice category. “Wins” are highlighted in green.**

Concentration (# of buoys)	Buoy speed	GOFs 3.1		GOFs 3.5		ESPC	
		Speed	Bias	Speed	Bias	Speed	Bias
0.8 < X ≤ 1.0 (107)	8.62	11.16	2.54	11.29	2.68	10.09	1.47
0.6 < X ≤ 0.8 (44)	13.79	14.58	0.79	15.00	1.22	13.03	-0.76
0.3 < X ≤ 0.6 (28)	13.04	17.46	4.46	20.77	7.74	17.31	4.27
0.1 < X ≤ 0.3 (6)	25.29	33.35	7.76	36.37	10.78	42.66	17.07
0.0 < X ≤ 0.1 (25)	17.81	18.16	0.35	14.66	-3.15	14.86	-2.95
“Wins”			2		0		3

**Table 11: Sea ice drift statistics of speed absolute mean error and RMSE (both in cm/s), and direction vector correlation categorized by ice concentration for IABP ice-bound drifting buoys over the Pan-Arctic domain versus GOFs 3.1, GOFs 3.5 and ESPC for the 12-hour forecast over 2017. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. “Wins” are highlighted with green.**

Concentration (# of buoys)	Absolute ME			RMSE			Vector Correlation		
	G3.1	G3.5	ESPC	G3.1	G3.5	ESPC	G3.1	G3.5	ESPC
0.8 < X ≤ 1.0 (107)	4.43	4.44	3.96	6.37	6.29	5.86	1.41	1.43	1.39
0.6 < X ≤ 0.8 (44)	6.91	6.98	6.98	9.48	9.54	9.52	0.95	0.87	0.81
0.3 < X ≤ 0.6 (28)	9.25	11.65	10.32	11.88	14.65	13.45	0.98	0.78	0.77
0.1 < X ≤ 0.3 (6)	15.01	10.78	17.32	19.18	16.46	24.96	1.05	1.50	1.33
0.0 < X ≤ 0.1 (25)	11.81	15.34	14.42	17.71	19.53	19.75	0.80	0.53	0.61
“Wins”	3	1	1	3	1	1	3	2	0

### 4.3 Atmospheric Performance

Here we compare the Navy ESPC deterministic forecast to the operational NAVGEM forecasts. Error statistics corresponding to the FNMOC operational score card (described in Table 12) were calculated for the 2017 test period. There are several reasons why the performance of these two systems will differ, aside from the fact that one is coupled to ocean and ice models. As noted in section 2.1.2, the stand-alone NAVGEM physics suite was deficient in representing the Madden Julian Oscillation (MJO) a key driver of extended-range prediction skill. Much effort was spent in developing a coupled physics suite that would produce skillful MJO forecasts, and the success in producing a forecast model with high MJO skill is detailed in the Navy ESPC Ensemble VTR. Also, the two systems were run at different resolutions. The operational forecasts run at FNMOC in 2017 used NAVGEM v1.4.1 at T425, while the Navy ESPC deterministic forecasts were run at T681. One might assume that higher resolution would result in increased skill; however, the physics suite was tuned for the ensemble resolution, not the higher resolution of the deterministic system, and there was limited experience running the NAVGEM v1.4 code at T681. For the next ESPC upgrade, we will incorporate NAVGEM 2.0 which includes a version of the coupled physics that has been optimized for the T681 resolution and we expect a significant performance upgrade with the inclusion of NAVGEM 2.0 into the ESPC system.

Table 12 shows the results from the FNMOC score card. For the standard scorecard considering forecasts lead times of 96 and 120 h, the overall score is -5. However, these degradations are generally small and become even smaller at longer leads. In addition, for a score card considering the same variables but for lead times of 144 h the overall score is +1. The Navy ESPC significantly outperforms the NAVGEM operational forecasts for 200-hPa and 850-hPa winds in the tropics and has lower biases for 2-m temperatures compared to fixed buoys. Navy ESPC has a slight positive but stable bias in tropical 10-m wind speeds, as opposed to the significant bias trend in NAVGEM, which starts off with a substantial high wind speed bias at analysis time, but has a significant low wind speed bias by 144 h. The improvements in tropical winds are consistent with the significantly improved MJO performance. Taking a broader view of the performance of the two systems, the skill between the two is comparable, with each system having distinct strengths and weaknesses compared to the other.

**Table 12. Operational NAVGEM scorecard comparing Navy ESPC deterministic forecasts to NAVGEM operational forecasts for the 2017 VTR test period, weightings have been assigned and approved by the Administrative Modeling Oversight Panel (AMOP).**

Reference	Level	Region	Lead time	Variable	Level type	Metric	Weight	Score	144-h Score
Fixed Buoy	None	Northern Hemisphere	96	Wind Speed	surface	Mean Error	2	0	0
Fixed Buoy	None	Tropics	96	Wind Speed	surface	Mean Error	2	0	0
Manual Sfc Land	None	Northern Hemisphere	96	Air Temperature	surface	Mean Error	1	0	0
Radiosondes	100	Global	96	Geopotential Height	pressure	RMS Error	1	-1	0
Radiosondes	250	Global	96	Air Temperature	pressure	RMS Error	1	0	0
Radiosondes	250	Global	96	Wind	pressure	Vector RMS Error	1	-1	0
Radiosondes	500	Global	96	Geopotential Height	pressure	RMS Error	1	-1	0
Radiosondes	700	Global	96	Relative Humidity	pressure	Mean Error	1	0	0
Radiosondes	850	Global	96	Air Temperature	pressure	Mean Error	1	-1	0
Radiosondes	850	Global	96	Air Temperature	pressure	RMS Error	1	-1	0
Radiosondes	850	Global	96	Wind	pressure	Vector RMS Error	1	0	0
Self Analysis	200	Northern Hemisphere	96	Wind	pressure	Vector RMS Error	1	0	0
Self Analysis	200	Southern Hemisphere	96	Wind	pressure	Vector RMS Error	1	0	0
Self Analysis	200	Tropics	96	Wind	pressure	Vector RMS Error	1	1	1
Self Analysis	500	Northern Hemisphere	120	Geopotential Height	pressure	Anomaly Correlation	2	0	0
Self Analysis	500	Southern Hemisphere	120	Geopotential Height	pressure	Anomaly Correlation	1	-1	-1
Self Analysis	500	Northern Hemisphere	120	Geopotential Height	pressure	RMS Error	2	0	0
Self Analysis	500	Southern Hemisphere	120	Geopotential Height	pressure	RMS Error	1	-1	0
Self Analysis	850	Northern Hemisphere	96	Wind	pressure	Vector RMS Error	1	0	0
Self Analysis	850	Southern Hemisphere	96	Wind	pressure	Vector RMS Error	1	0	-1
Self Analysis	850	Tropics	96	Wind	pressure	Vector RMS Error	2	2	2
Self Analysis	1000	Northern Hemisphere	120	Geopotential Height	pressure	Anomaly Correlation	1	0	0
Self Analysis	1000	Southern Hemisphere	120	Geopotential Height	pressure	Anomaly Correlation	1	-1	0
total								-5	1

Figure 69 shows the geopotential height AC (Anomaly Correlation) coefficient, the spatial correlation between the forecast anomaly and the anomaly in the verifying “self” analysis (using the analysis of the forecast system for verification). These anomalies are computed with respect to a model climate. Also shown in the lower two panels of Figure 69 are the RMSE of forecasts of geopotential height with respect to radiosondes for both Navy ESPC deterministic and NAVGEM operational. Shading on these plots indicates a 95% confidence interval. The differences between the two runs are small with slightly better AC and RMSE in NAVGEM operational. Note that while the performance in terms of RMSE at 500-hPa and 100-hPa is slightly worse for Navy ESPC at 24-120 h (and suffers a -1 on the scorecard for each of these), performance between the two systems is very similar at 144 h, reflecting the fact that the Navy ESPC physics suite has been designed to perform well at longer lead times.

As Navy ESPC is a coupled system, we hope to see improved performance near the ocean surface. Evaluation of 10-m wind speed biases (Figure 70, left panels) with respect to fixed buoys shows comparable error between the two systems in the NH, larger biases in the SH (though not statistically significant at 144 h), and improved performance for the Navy ESPC system at most forecast lead times in the tropics. Operational NAVGEM exhibits a strong weakening trend in wind speed with forecast lead time, while the Navy ESPC has small positive biases that are fairly stable with lead time. This is not reflected in the scorecard category since Navy ESPC has larger biases at 96 h. However, Navy ESPC has smaller biases at 144-h; which would likely be maintained at longer lead times. For 2-m temperature compared to fixed buoys, Navy ESPC outperforms NAVGEM operational forecasts in both the NH and the tropics at most forecast lead times. For the NH land surface comparison, NAVGEM outperforms Navy ESPC out to 96 h. However, at 144-

h, the system biases are equivalent, and the Navy ESPC biases at longer lead times are quite small. We note that when using ECMWF analyses for verification, the Navy ESPC deterministic forecast wind speed bias is substantially smaller than the wind speed bias of the NAVGEM ensemble forecasts in the Northern Extratropics, Tropics, and Southern Extratropics (see the Navy ESPC Ensemble VTR for details). The 2-m temperature biases for Navy ESPC are comparable to the NAVGEM ensemble for the tropics, but larger in the extratropics.

For 200-hPa and 850-hPa vector wind RMSE vs. self-analysis (Figure 71), the Navy ESPC and NAVGEM systems have very similar performance in the NH and SH extratropics. However, in the tropics, the Navy ESPC performance is clearly superior to NAVGEM. The improved winds in both the lower and upper troposphere are likely a result of tuning the physics suite to better forecast the MJO and other tropical phenomena as shown below.

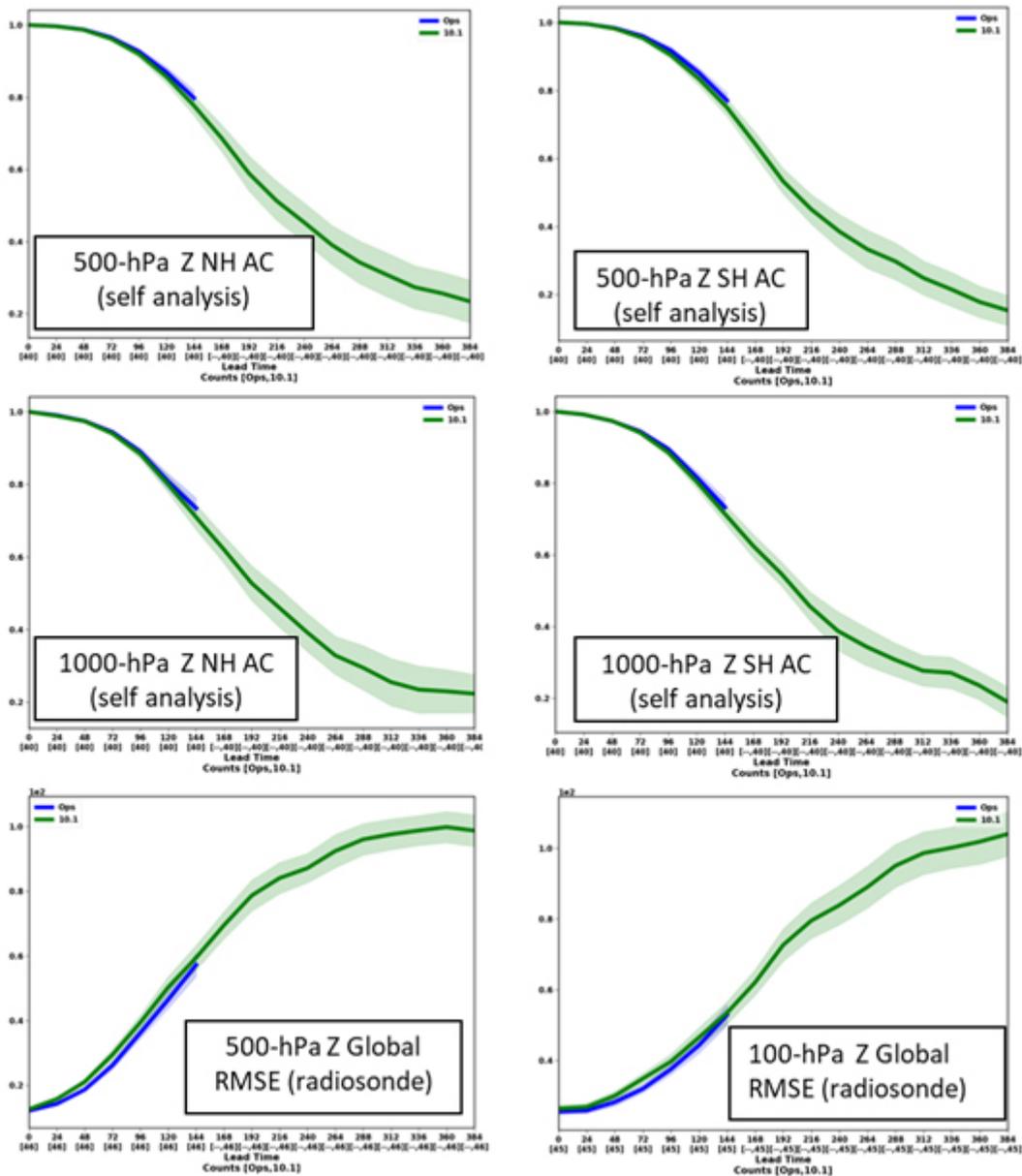


Figure 69. Comparison of the NAVGEM operational forecasts (blue) to the Navy ESPC deterministic forecasts (green) for 500-hPa geopotential height AC vs. self-analysis (top panels, NH left, SH right), 1000-hPa geopotential height AC vs. self-analysis (middle panels, NH left, SH right), 500-hPa global geopotential height RMSE vs. radiosondes (lower left), and 100-hPa global geopotential height RMSE vs. radiosondes (lower right). A 95% confidence interval is shown by the shaded regions.

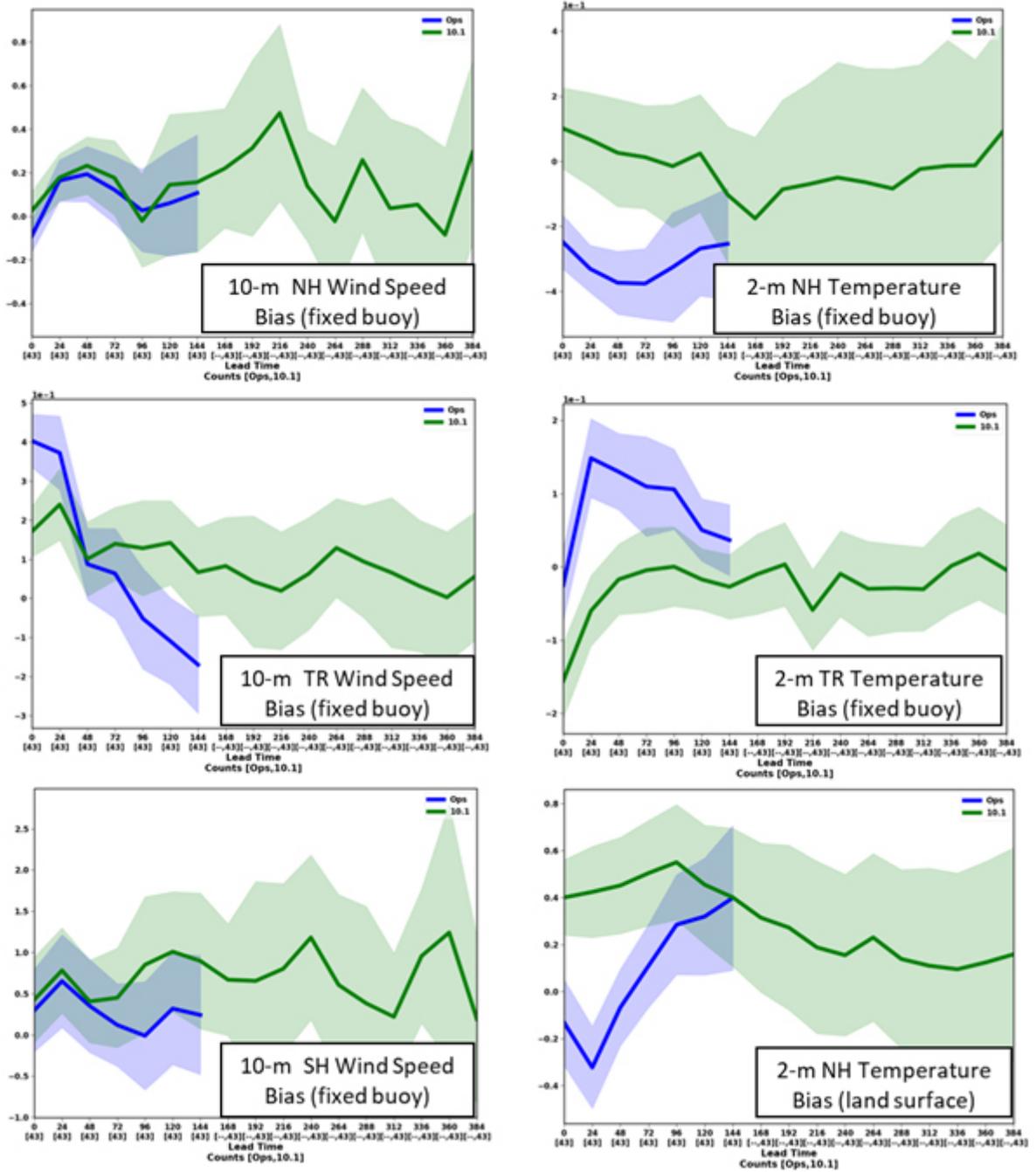


Figure 70. Comparison of the NAVGEM operational forecasts (blue) to the Navy ESPC deterministic forecasts (green) biases as a function of forecast lead time verified against fixed buoys and land surface station observations for 10-m NH ocean wind speed (upper left), 2-m NH ocean temperature (upper right), 10-m tropical ocean wind speed (middle left), 2-m tropical ocean temperature (middle right), 10-m SH ocean wind speed (lower left) and 2-m NH land temperature (lower right). A 95% confidence interval is shown by the shaded regions.

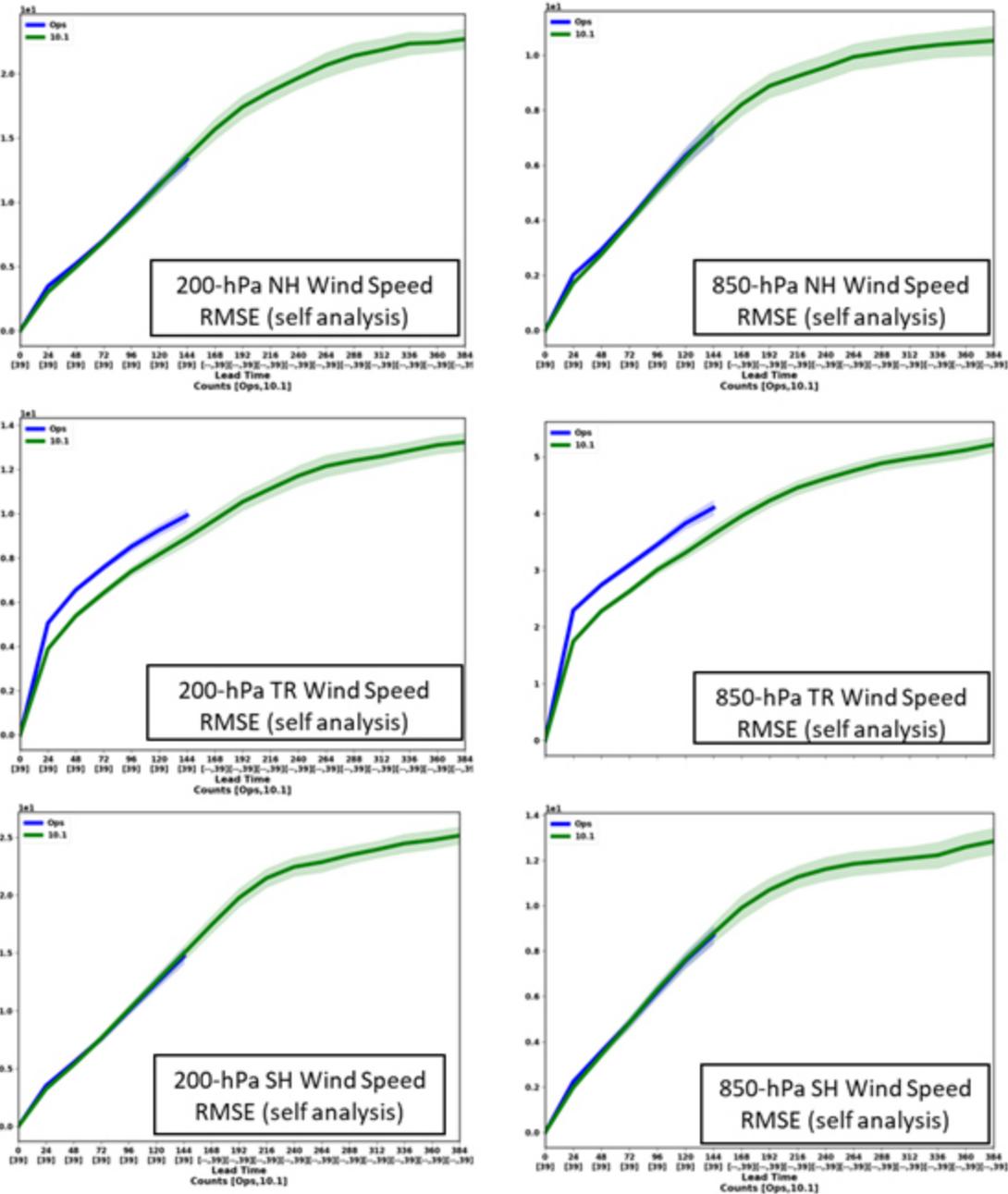


Figure 71. Comparison of the NAVGEM operational forecasts (blue) to the Navy ESPC deterministic forecasts (green) vector wind speed RMSE as a function of forecast lead time verified against self-analysis at 200-hPa (left panels) and 850-hPa (right panels) for NH (top), Tropics (middle) and SH (bottom). A 95% confidence interval is shown by the shaded regions.

Radiosonde verification for other variables is shown in Figure 72. The two systems perform comparably for 250-hPa temperature bias and RMSE while, NAVGEM performs slightly better for 250-hPa global wind RMSE. Navy ESPC has a stronger dry bias at 700-hPa and stronger warm bias at 850-hPa than NAVGEM, which contributes to slightly higher 850-hPa temperature RMSE in

Navy ESPC. These lower-tropospheric biases will be a focus area for improvements with the next implementation.

While the sample size was not large enough to be significant, we did calculate tropical cyclone track errors. Figure 73 shows the tropical cyclone track errors for NAVGEM and Navy ESPC. In this small sample size, the track errors are larger for Navy ESPC than for NAVGEM; however, the differences are not statistically significant. The tropical storm sampling which can be obtained from weekly long forecasts is not adequate to robustly define the error space. To augment the tropical storms evaluation, Figure 74 compares the TC track errors for the 48-h forecast which had been run for the continuous data assimilation cycling. This comparison yields a larger sample size and we find the TC track errors are very similar between the two systems; Navy ESPC is slightly worse at early forecast lead times and slightly better at 48 h.

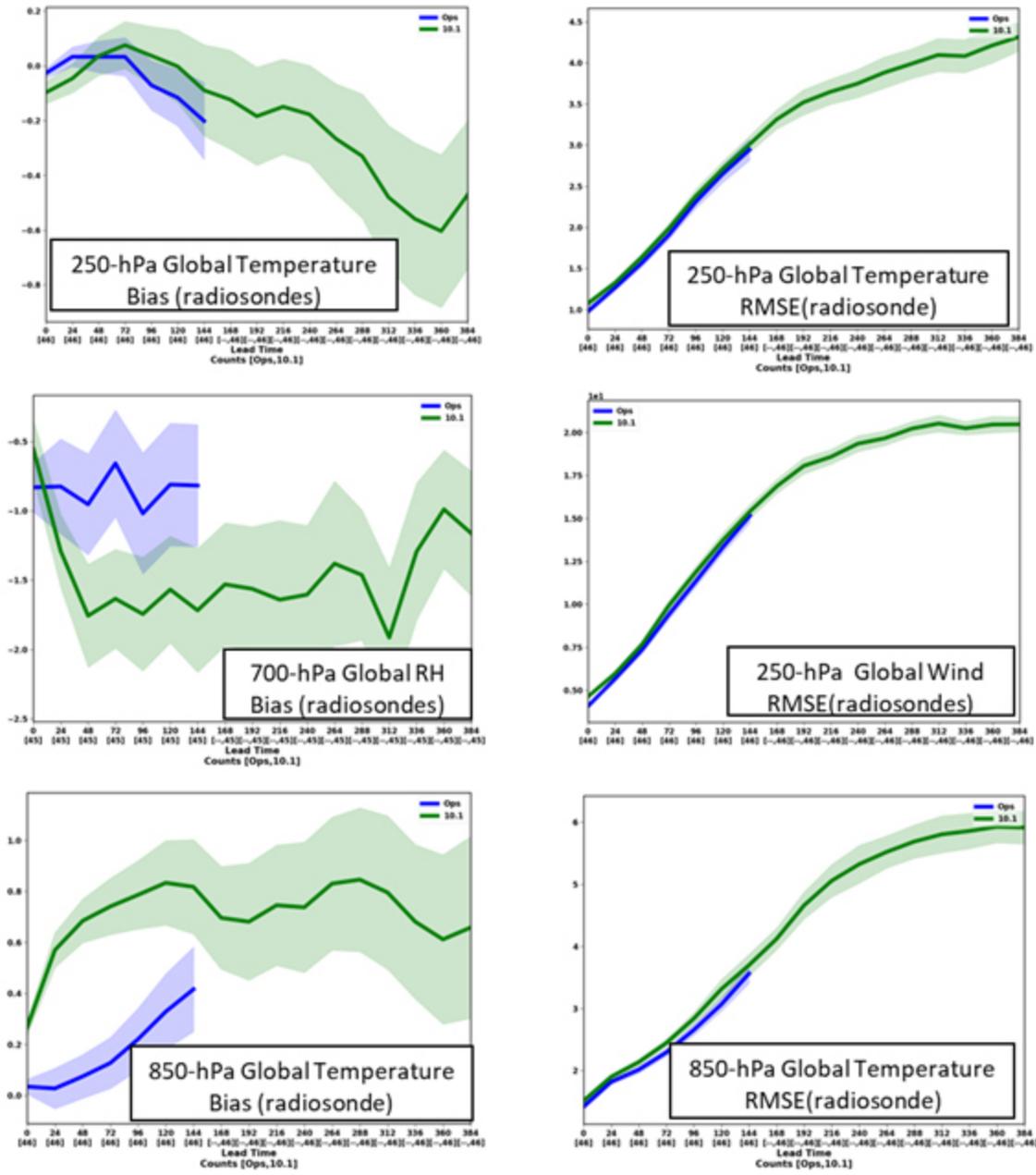


Figure 72. Comparison of the NAVGEM operational forecasts (blue) to the Navy ESPC deterministic forecasts (green) verified globally against radiosondes for 250-hPa temperature bias (upper left), 250-hPa temperature RMSE (upper right), 700-hPa RH bias (middle left), 250-hPa wind RSME (middle right), 850-hPa temperature bias (lower left), and 850-hPa temperature RMSE (lower right).

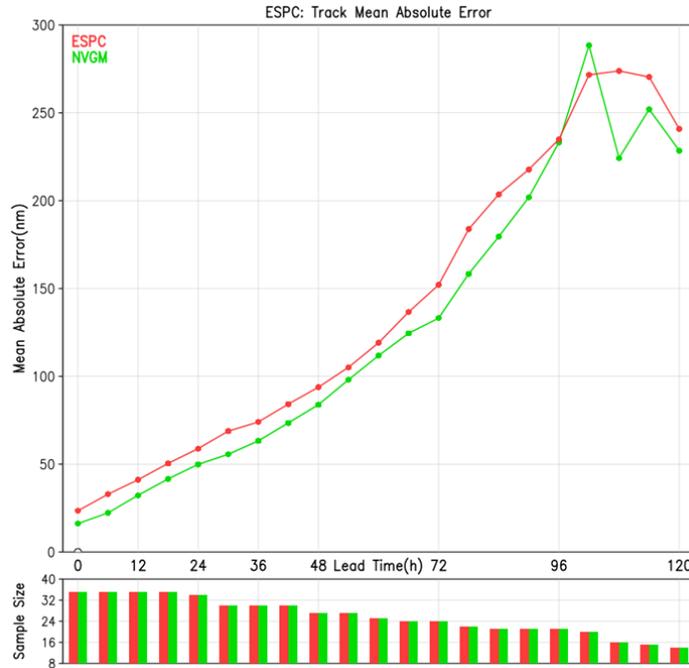


Figure 73. Tropical cyclone track errors for a homogenous comparison of Navy ESPC (red) and NAVGEM (green) for the long forecast. The sample size is provided in the lower portion of the figure. The differences are not statistically significant.

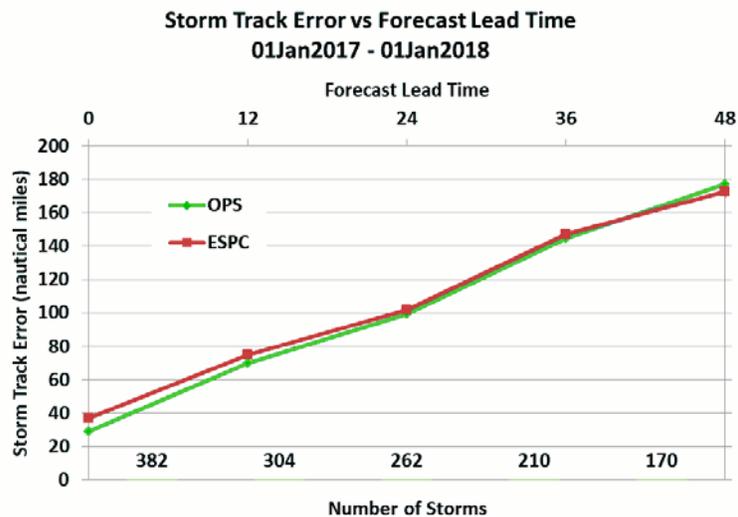


Figure 74. Tropical cyclone track errors for a homogenous comparison of Navy ESPC (red) and operational NAVGEM (green) for the 48-h forecasts associated with the DA update cycle. The number of storms sampled is listed on the lower x-axis while the upper x-axis displays the forecast lead time. The differences are not statistically significant.

Changes to the physical parameterizations in Navy ESPC were designed to improve the performance of the model specifically for the Madden-Julian Oscillation (MJO), one of the primary modes of variability in the tropics, and one of the few global phenomena with potential predictability on extended-range timescales. The phase of the MJO is known to have substantial impact on many other phenomena that impact sensible weather important to both DoD and

civilian interests, such as tropical cyclone genesis and frequency, atmospheric rivers, and even the Arctic and Antarctic atmospheric circulation. As such, it is important for extended-range forecasting systems to be capable of realistic simulations and accurate predictions of the MJO. Changes to the physical parameterizations led to a very significant improvement in the MJO such that the Navy ESPC model compares favorably to other state-of-the-art forecast models in representation of the MJO as seen through evaluations of the 17 years of Navy ESPC forecasts produced for the NOAA Subseasonal Experiment (SubX) (Janiga et al. 2018; Kim et al. 2019).

Here we show results comparing the prediction of the MJO by the Navy systems to other state-of-the-art systems, focusing on the first 16 forecast days (Figure 75). A more extensive comparison valid out to longer lead times and including forecasts from additional centers is provided in the ensemble VTR. In order to evaluate the skill of the MJO forecasts, we use the Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004). This commonly used index has two components, which we evaluate separately. RMM1 is positive when the MJO is active over the Maritime Continent and negative when the MJO is active over South America to Africa. RMM2 is positive when the MJO is active over the Pacific Ocean and negative when the MJO is active over the Indian Ocean.

This evaluation of MJO performance compares the skill of Navy ESPC deterministic forecasts and operational deterministic NAVGEM forecasts. We also examine the skill of individual members of the Navy ESPC ensemble, operational NAVGEM ensemble (NAVGEM ET), NOAA Climate Forecast System version 2 (CFSv2) ensembles, and a 16 member subset of the European Center for Medium-range Forecasting (ECMWF) ensemble. The forecasts from the other centers were downloaded from the Subseasonal to Seasonal (S2S) database (Vitart et al, 2017).

The best performer, on an individual member basis, is Navy ESPC run at the ensemble resolution. This makes sense as the NAVGEM coupled physics was tuned to perform well at the Ensemble resolution. The Navy ESPC deterministic forecasts do not perform quite as well for this index as the Navy ESPC ensemble forecasts. This reflects the fact that the Navy ESPC coupled physics was not tuned for higher resolution. However, Navy ESPC at both deterministic and ensemble resolutions perform quite well. Navy ESPC ensemble actually outperforms ECMWF. However, this is not true when considering the ensemble mean, as detailed in the ensemble VTR. Navy ESPC deterministic outperforms the NOAA CFSv2 system for all forecast lead times for RMM1 and for lead times past 5 days for RMM2. The AC at initial time is smaller for Navy ESPC deterministic than for NAVGEM deterministic, but the skill drops off faster for NAVGEM, such that Navy ESPC deterministic outperforms NAVGEM deterministic at lead times longer than 4 days for RMM1, and is comparable to the NAVGEM deterministic at lead times of 6 days for RMM2. Both Navy ESPC deterministic and Navy ESPC ensemble outperform the individual NAVGEM ensemble members substantially at all lead times.

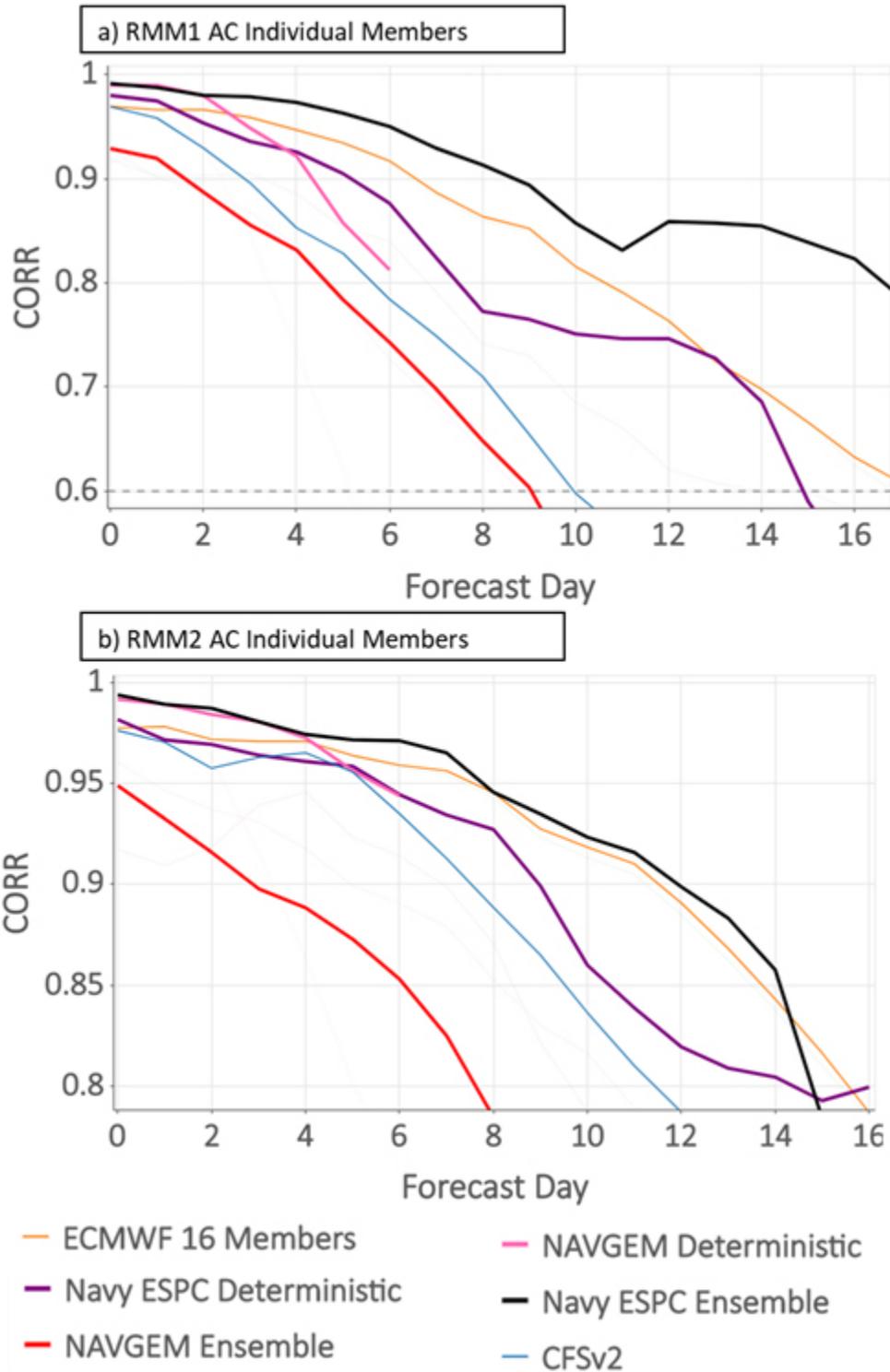


Figure 75. AC for RMM1 (a) and RMM2 (b). Scores are calculated for individual forecast members and averaged for each center. Model colors denoted in key are the ECMWF 16-member ensemble, NOAA (CFSv2), Navy ESPC high-resolution deterministic, Navy ESPC ensemble, NAVGEM deterministic, and NAVGEM ensemble.

In order to highlight the excellent performance of the Navy ESPC system at long lead times, we include here a figure that is also included in the ensemble VTR (Figure 76). It is anticipated that large-scale teleconnection patterns will be predictable on longer time scales than fields at individual points due to the relatively large scale and slower temporal variability of these patterns. We evaluate the skill of the Navy ESPC ensemble and deterministic forecasts and benchmark their performance relative to NAVGEM and systems from other centers, including the Australian Bureau of Meteorology (BOM), Meteo-France, ECMWF, and NOAA. The teleconnection patterns considered include the Arctic Oscillation (AO) and Antarctic Oscillation (AAO), which provide information on the zonal symmetry of the northern and southern jet streams, and the Pacific North American Oscillation (PNA) and North Atlantic Oscillation (NAO), which provide information on dominant modes of variability over these respective regions. Evaluating the skill of the forecasts for these teleconnection patterns is a useful way to gauge the ability of the model to capture these relatively slowly varying atmospheric modes of variability, which are tied to anomalous sensible weather conditions. In addition, the Naval Information Warfighting Development Center has said that Navy forecasters are interested in forecasts of the AO to provide information as to the magnitude of anomalies in weather patterns that might be expected in the Northern Hemisphere middle and high latitudes.

Figure 76 shows “Violin” plots illustrating the distributions of the forecast day when the AC drops below 0.6 for the individual members of the ensemble systems. There is only one point for the Navy ESPC deterministic model since it is not an ensemble system. The Navy ESPC deterministic point should be compared with the centers of the rectangles inside the violin plots. We do not include the NAVGEM deterministic system here since the forecasts generally are not long enough to reach an AC of 0.6. The day that the ensemble mean forecast skill falls below 0.6 is also included for these systems, and is indicated by the “X”. The performance of the Navy ESPC ensemble system (ESPC-E) is the focus of the discussion of this figure in the ensemble VTR. Here we focus on the performance of the Navy ESPC deterministic system (ESPC-D). We see that the Navy ESPC deterministic forecast is comparable to the average score for the Navy ESPC ensemble forecasts for the AO, AAO, and NAO, and not quite as good for PNA. Comparing Navy ESPC deterministic forecasts (ESPC-D) with the skill of the individual NAVGEM ensemble forecast members (NAVGEM) indicates clearly superior performance of the Navy ESPC for all four teleconnection patterns. In fact, the Navy ESPC deterministic forecast provides on average an extra two days of forecast skill above the NAVGEM ensemble members. Comparison to the other centers also indicates that Navy ESPC performs comparably to ECMWF, generally considered the best system in the world, and is comparable to the NOAA CFSv2 for the AO, PNA, and NAO and superior to CFSv2 for the AAO.

The examination of Navy ESPC in terms of the MJO and teleconnection patterns reveals that the changes to the physical parameterizations accomplish what they were designed to do. That is, while NAVGEM outperforms the Navy ESPC for the 4-day and 5-day NAVGEM score-card metrics, the Navy ESPC forecasts substantially outperform the NAVGEM deterministic and ensemble forecasts in terms of metrics more appropriate for longer time ranges. As the Navy ESPC forecasts will not replace the NAVGEM forecasts in the near future (due to latency issues), we believe this

superior performance in forecast skill at longer forecast lead times supports transition of the Navy ESPC to operations.

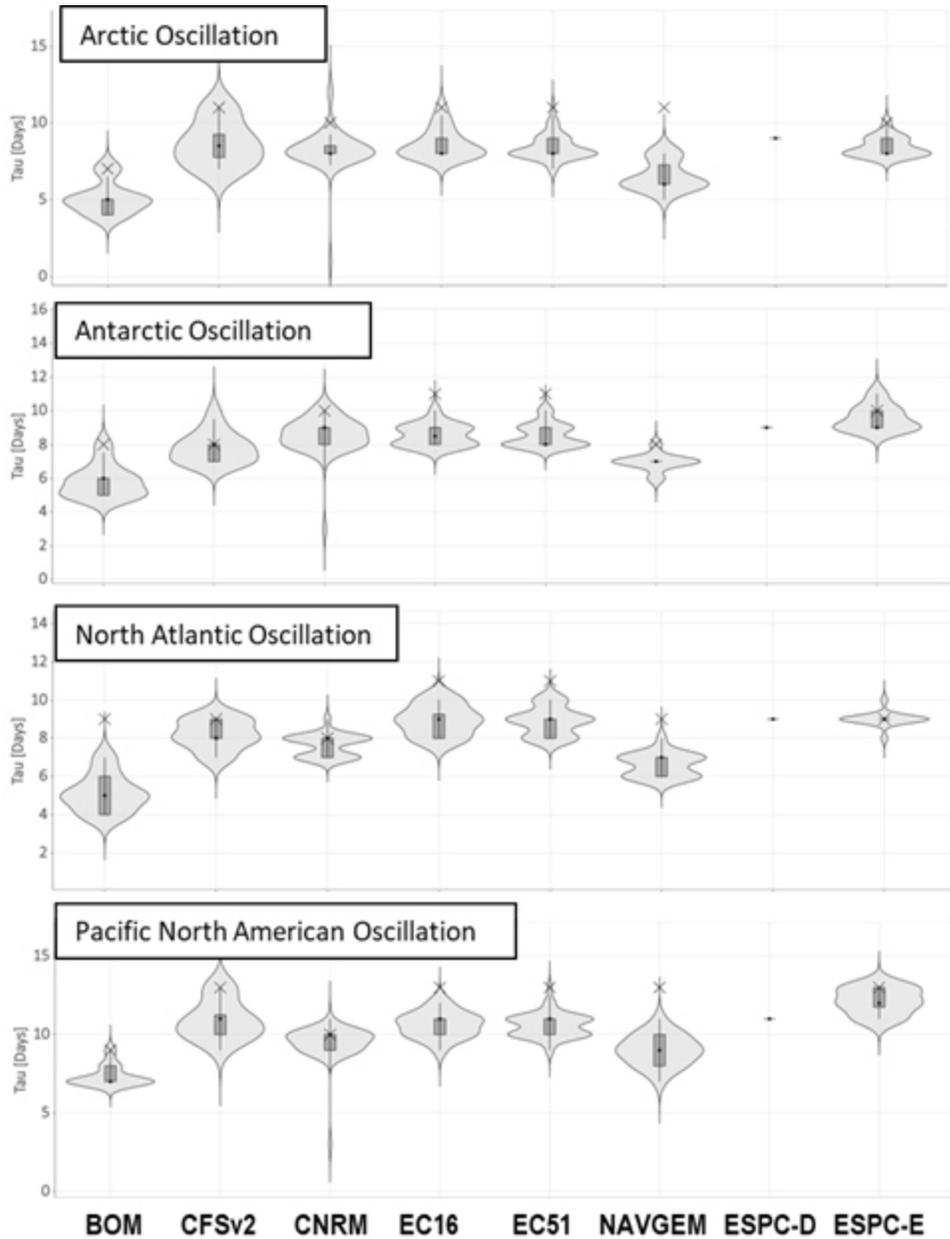


Figure 76. “Violin” plots showing the distribution of the day that the forecast of the large-scale oscillation (as denoted in panels) AC falls below 0.6. Each ensemble member is treated as an individual (deterministic) forecast, and the width of the violin is proportional to the number of ensemble members that fall below 0.6 on that day. The forecasts considered come from the Australian Bureau of Meteorology (BOM), NOAA (CFSv2), Meteo-France (CNRM), ECMWF 16-member (EC16) and 51-member (EC51) ensembles, NAVGEM ET (NAVGEM), Navy ESPC high-resolution deterministic forecasts (ESPC-D), and the Navy ESPC ensemble (ESPC-E). The day on which the ensemble mean forecast AC falls below 0.6 is denoted by the “x”.

In summary, the performance of the two systems is comparable and reflects the different priorities in performance. NAVGEM is slightly better than Navy ESPC on the standard forecast

score card which has been optimized for; however, this advantage disappears at 144 h. Overall, Navy ESPC performs comparably or better for most near surface metrics and tropospheric tropical winds. In addition, for many metrics the improvement over NAVGEM increases with forecast lead time. Verification of long-lead forecasts of teleconnections and the MJO show that the Navy ESPC deterministic and Navy ESPC ensemble forecasts perform very well compared to state-of-the-art systems from other centers such as ECMWF and NOAA (see the Navy ESPC Ensemble VTR for more details).

#### **4.3.1 Atmospheric Coupling diagnostics**

In this section we evaluate the ability of Navy ESPC to reproduce atmosphere-ocean coupling on intraseasonal time scales. The diagnostics used are based on lag correlation between indices capturing the large scale intraseasonal variability of precipitation associated with the MJO and local SST and air-sea fluxes for various regions. The lag correlations for this index are calculated for both observations from a satellite based precipitation reanalysis and the model forecasts separately.

These indices measure the intraseasonal variability of precipitation in the Indo-Pacific warm pool during boreal summer and are based on the first 2 EOFs of eastward propagating Tropical Rainfall Measurement Mission (TRMM) precipitation (Figure 77 a, b) and describe the monsoon active/break cycle. Similarly to other intraseasonal variability measures such as Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004) or Boreal Summer Intra-Seasonal Oscillation (BSISO) index (Lee et al 2013), the precipitation indices are created by projecting observed and forecasted anomalies on the first two EOFs of the variable used in creating the index (such as wind, OLR or precipitation). The monsoon active/break transition is usually associated with strong upper ocean anomalies and the coupling diagnostic are designed to measure the atmosphere/ocean feedbacks, especially for the areas where the large intraseasonal variability is observed. For this evaluation we use precipitation index because of its direct relationship to convection.

The analysis shown in Figure 77 uses precipitation, SST and surface fluxes from 45 days Navy ESPC hindcasts for June-September 2012-2016 period and corresponding observed surface fluxes and SSTs from the TropFlux project (Kumar et al, 2012). TropFlux analysis provides consistent air-sea heat and momentum flux data for the entire 30°N-30°S region.

The results in Figure 78 compare the correlations between first and second PCs with local variables. The left panels (Figure 78a, c, and e) against SST and the right panels (Figure 78b, d and f) against net downward shortwave radiation (or just radiation). We also show the correlation between the precipitation indices and SST for CFSv2 forecasts (left, panel, Figure 78g, h and i). The observation (solid) lines and the model (dashed) show that the correlations Navy ESPC system for the entire 2012-2016 period are similar at all lags and compare favorably with similar correlations for CFSv2. Based on this we can conclude that Navy ESPC represents well the interaction between the large scale intraseasonal precipitation variability and local ocean variability, in spite of somewhat weaker correlations between intraseasonal modes and net

downward shortwave surface radiation. However when examining the results for single seasons 2012 (Figure 78c and d) and 2015 (Figure 78e and f) this similarity may not always be as robust. For example the skill for 2015 is significantly lower than the skill for 2012, and similarly for the entire 2012-2016 period. The lower skill of the lagged correlation of the SST with the precipitation index are consistent with lower skill of lagged correlation of radiation with the precipitation index. Therefore, the errors in coupling may be caused by the surface radiation biases. In the model the same technique can be used for different indices of intraseasonal variability and different components of the surface flux.

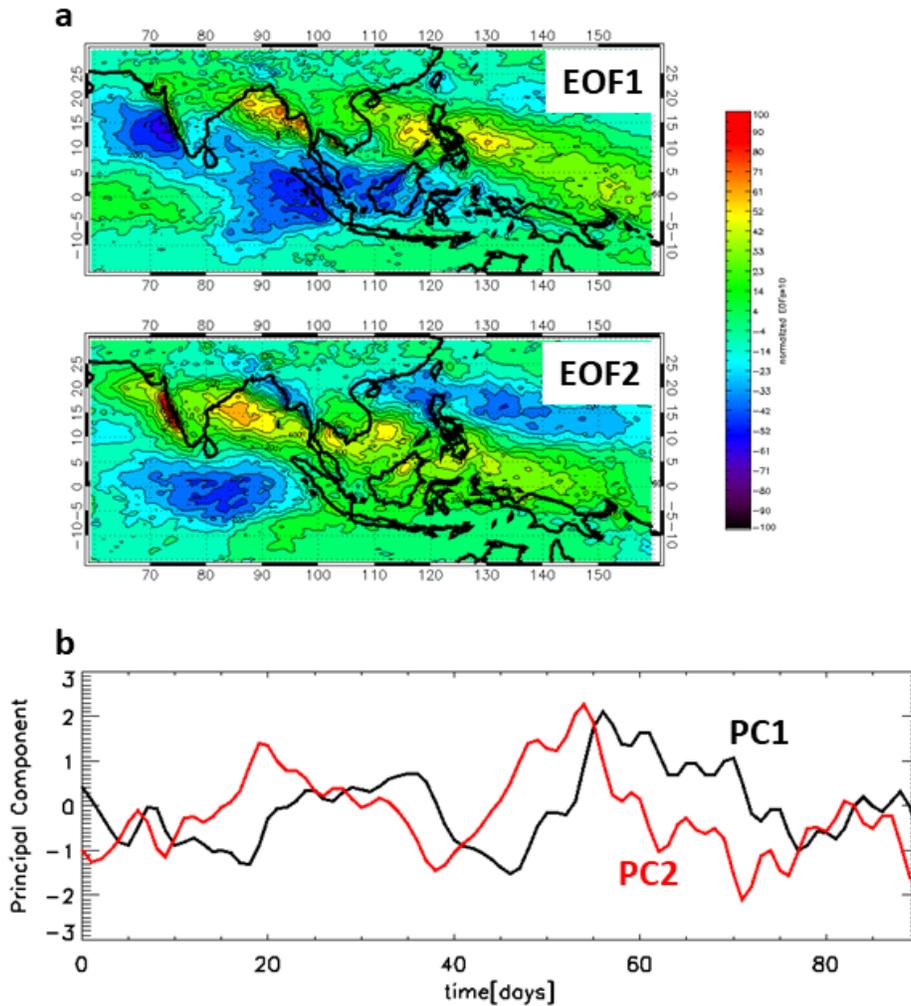
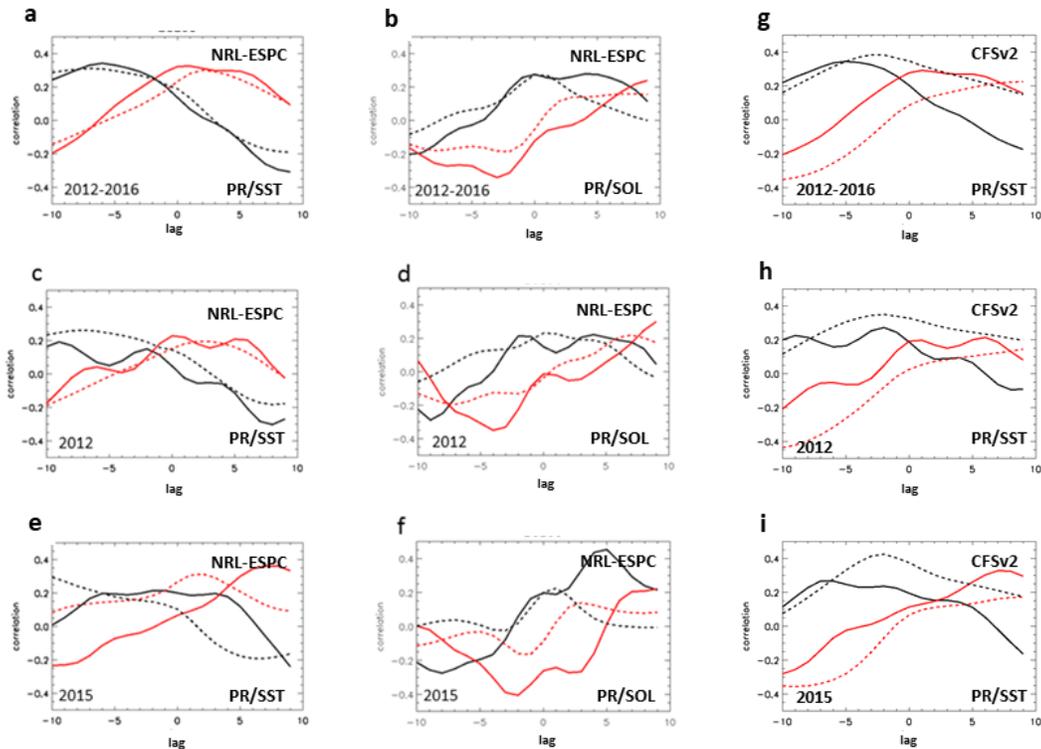


Figure 77. (a) Normalized TRMM precipitation EOFs used for evaluation of intraseasonal variability during boreal summer monsoon (June-September). The indices (normalized principal components) are created by projecting the model or observed precipitation anomalies these 2 EOFs and are shown in (b).



**Figure 78.** Lag correlation between precipitation modes defined by EOFs shown in Figure 77 and local SST (a, c, and e) and surface net solar radiation (b, d, and f) in the Bay of Bengal for June-September. Panels (g) (h) and (i) show the same as (a) (c) and (d) but for the CFSv2. The black lines indicate the correlations for the principal component (PC) for the first EOF, red for the second PC. Solid lines show the observed correlations, the dashed lines indicate the correlations for ESPC system forecasts.

## 5 Summary and Future Work

The Navy ESPC deterministic and ensemble systems represent powerful new capabilities for the Navy. While the Navy ESPC atmospheric deterministic forecasts were out-performed by the NAVGEM operational system for the standard NAVGEM scorecard, we note that the new coupled physics developed to improve NAVGEM performance for the MJO, for near-surface fields, and for extended ranges has done exactly that. In fact, applying the standard score card metrics at a 144 h lead time instead of 96 or 120 h lead times results in near-neutral scores. In addition, Navy ESPC outperforms the operational NAVGEM forecast model by a considerable margin for tropical lower and upper-tropospheric winds. Comparisons of large-scale teleconnection pattern and MJO forecasts from Navy ESPC with forecasts from systems from other centers indicates deterministic forecast skill comparable to that of the top centers in the world, such as ECMWF (see Navy ESPC ensemble VTR for more details).

For the ocean, the coupled Navy ESPC deterministic performance is generally similar to the similar resolution pre-operational uncoupled GOFs 3.5. Both high-resolution systems perform slightly better than the lower resolution operational GOFs 3.1, except for some of the acoustic metrics. In the top 30 m of the upper ocean, the benefits of the coupled system can be observed.

The coupled system with no flux correction has smaller temperature biases than the flux-corrected ocean-only forecasts and those biases often are reduced over the forecast. Similarly, the trends in bias and root mean square error growth found in the 7-day forecasts are continued into the second week of the 16-day forecasts.

For sea ice, the coupled Navy ESPC system presently uses an older version of the sea ice model, CICE v 4.0, compared to the pre-operational GOFs 3.5, which uses CICE v5.1.2. This has a large impact on accurately forecasting the sea ice edge with GOFs 3.5 clearly outperforming both GOFs 3.1 and Navy ESPC. The current operational GOFs 3.1 uses the older version of the sea ice model. Sea ice thickness performance is similar for all systems, but it must be noted that the observational data used in this analysis is limited to a small number of flights over a short time frame. With regard to the sea ice drift metric, Navy ESPC often shows the lowest speed bias and RMSE, however the direction vector correlation values were typically the lowest. The performance of Navy ESPC is slightly worse in the Antarctic and the degradation may be due to performance issues of the uncalibrated higher resolution atmosphere in Navy ESPC. It is encouraging that the error growth trends in the first week of the forecasts are continued into the second week in the 16-day forecasts.

Because of latency and computational cost, the deterministic Navy ESPC system is not a candidate to replace the deterministic uncoupled NAVGEM system. However, Navy ESPC deterministic is a candidate for eventual replacement of the standalone GOFs system once skill parity can be achieved. As outlined in this report, although Navy ESPC performs broadly neutral relative to the uncoupled baselines, there are several key metrics where superior GOFs forecasts demonstrate the need for continued execution of the standalone GOFs system.

Based upon the evaluation of the deterministic system performed by the NRL ESPC team and the validation test panel, we provide the following findings and recommendations:

1. The deterministic Navy ESPC system provides a new deterministic coupled prediction capability with key improvements in certain Navy-relevant parameters (e.g. tropical 10m winds and near surface (upper 30 m) temperature and salinity).
2. Compared to the current operational capability, the performance of the deterministic coupled system is broadly comparable to current operational systems at tactical forecast lead times (0 to 7 days).
3. While comparisons with the uncoupled controls are overall neutral, there are several areas where the current uncoupled systems provide superior guidance (sonic layer depth/below layer gradient for ocean acoustics, and sea ice edge in the polar latitudes).
4. This initial version of Navy ESPC is suitable for transition to operations.
5. It does provide skill over climatology and persistence in the 8-16 day forecast range.

We look forward to increased performance with the next Navy ESPC upgrade which will include performance upgrades in all of the component systems as well as the addition of WW3 as detailed below.

**Waves:** The initial implementation of the Navy ESPC does not include ocean surface waves. Wave Watch 3 (WW3) has recently been incorporated into the Navy ESPC at the ensemble resolution ( $1/4^\circ$ ), and is showing comparable performance to the control WW3 forecasts, but this was not incorporated in time for the VTR runs. For FOC, we plan to incorporate WW3 into the Navy ESPC system at  $1/4^\circ$  for the ensemble component and  $1/8^\circ$  for the deterministic component. This will be the first Navy global wave model with resolution higher than  $1/4^\circ$ , surface currents in forcing, and ensembles that take into account uncertainty associated with errors in ice edge position. The inclusion of WW3 into Navy ESPC will allow, for the first time, prediction of anomalous ocean surface wave conditions out to 45 days. Feedback from WW3 to the other components of the system will enhance extended-range forecast skill in all Navy ESPC component models. Accounting for ice-edge uncertainty will improve prediction of environmental conditions in the marginal ice zone.

**Sea Ice:** The current Navy ESPC system includes CICE v4. Comparison with GOFs 3.5, which includes CICE v5.1.2, indicates that the more advanced version of CICE provides more accurate sea ice prediction, especially during winter freeze up. This is due to the inclusion of important processes for sea ice thermodynamics, such as melt ponds, snow cover, and landfast sea ice, which are currently missing in CICEv4. We plan to incorporate CICE v6 into the Navy ESPC system for FOC, as this will include the improvements of CICE v5.1.2 as well as other advantages, including the fact that the model will not have to be recompiled to be run at different resolutions.

**Ocean:** The current version of Navy ESPC has 41 vertical hybrid layers in HYCOM and does not account for tides in the ensemble configuration. For FOC, we anticipate increased fidelity and performance through the addition of tides to ensemble configuration, increased vertical levels (50-60). Higher order (more accurate) advection, proper treatment of rivers and evaporation and precipitation, more accurate vertical mixing and flux calculations will also be incorporated. We expect these improvements to result in increased fidelity in ocean structure, particularly in regions of strong vertical gradients, which will lead to improved sound speed profiles and representation of the acoustical environment. The addition of tidal forcing provides a good representation of internal waves that is important to the submarine community.

**Atmosphere:** The current version of NAVGEM in ESPC is based on NAVGEM 1.4.3. Upgrading to the refactored version of NAVGEM 2.0 along with attendant physics upgrades at FOC should result in significant atmospheric forecast improvements in the Navy ESPC. The current NAVGEM version in Navy ESPC IOC has a top at 0.01hPa (approximately 72 km), with rudimentary middle-atmospheric physics. For FOC, we plan to incorporate NAVGEM-HA (High-Altitude), which has a model top at 100 km and sophisticated middle-atmosphere physics. This capability will allow for operational forecasts (for the first time, for any lead time) of the middle atmosphere, including the mesosphere and lower thermosphere. This will provide accurate lower boundary conditions for the ionospheric forecasts, protect assets traveling in and through the middle atmosphere, and improve S2S tropospheric weather prediction as the stratosphere has been shown as a source of long-lead tropospheric predictability. A new in-line aerosol capability will also be included in Navy ESPC for FOC.

## 6 Appendix A

The NAVGEM component of the Navy ESPC is compared to the operational NAVGEM runs for the VTR test period for the calendar year of 2017. During this time period, 16-day Navy ESPC deterministic forecasts are run once per week such that they may be compared to the Navy ESPC ensemble forecast (see the ESPC Ensemble VTR) for teleconnection pattern and MJO forecast performance. The scorecard items are listed in Table 12. The columns are the source used for evaluation (field means the analysis field), the level, the area, the field, the type of error, the forecast hour used for the evaluation, and the weight or score. A positive score indicates that the test forecast system is statistically better; a negative score indicates that the control is statistically better, and a zero score means that neither are statistically better. In addition to the standard scorecard which considers performance at the 96-h and 120-h forecast times, we also include scores for the 144-h forecast time.

The evaluation criteria for the various items are:

1. Tropical Cyclone Track: An improvement in the 96-hour TC track at a 95% statistical confidence level using a Student T test.
2. Anomaly Correlation: An improvement in the 120-hour height anomaly correlation at a 95% statistical confidence level.
3. Root Mean Square Height Error: A 5% or greater improvement in the 120-hour root mean square error at a 95% statistical confidence level.
4. Root Mean Square Temperature Error: A 5% or greater improvement in the 96-hour root mean square error at a 95% statistical confidence level
5. Vector Root Mean Square Error: A 5% or greater improvement in the 96-hour wind-vector root mean square error at a 95% statistical confidence level.
6. Mean Speed Error: A 0.25 m/s or greater improvement in the 96-hour mean error at a 95% statistical confidence level.
7. Radiosonde Mean Bias Error for Temperature: A 0.5 degree C or greater improvement in the 96-hour mean temperature error at a 95% statistical confidence level.
8. Radiosonde Mean Bias Error for Relative Humidity: A 5% or greater improvement in the 96-hour mean relative humidity error at a 95% statistical confidence level.

## 7 Appendix B

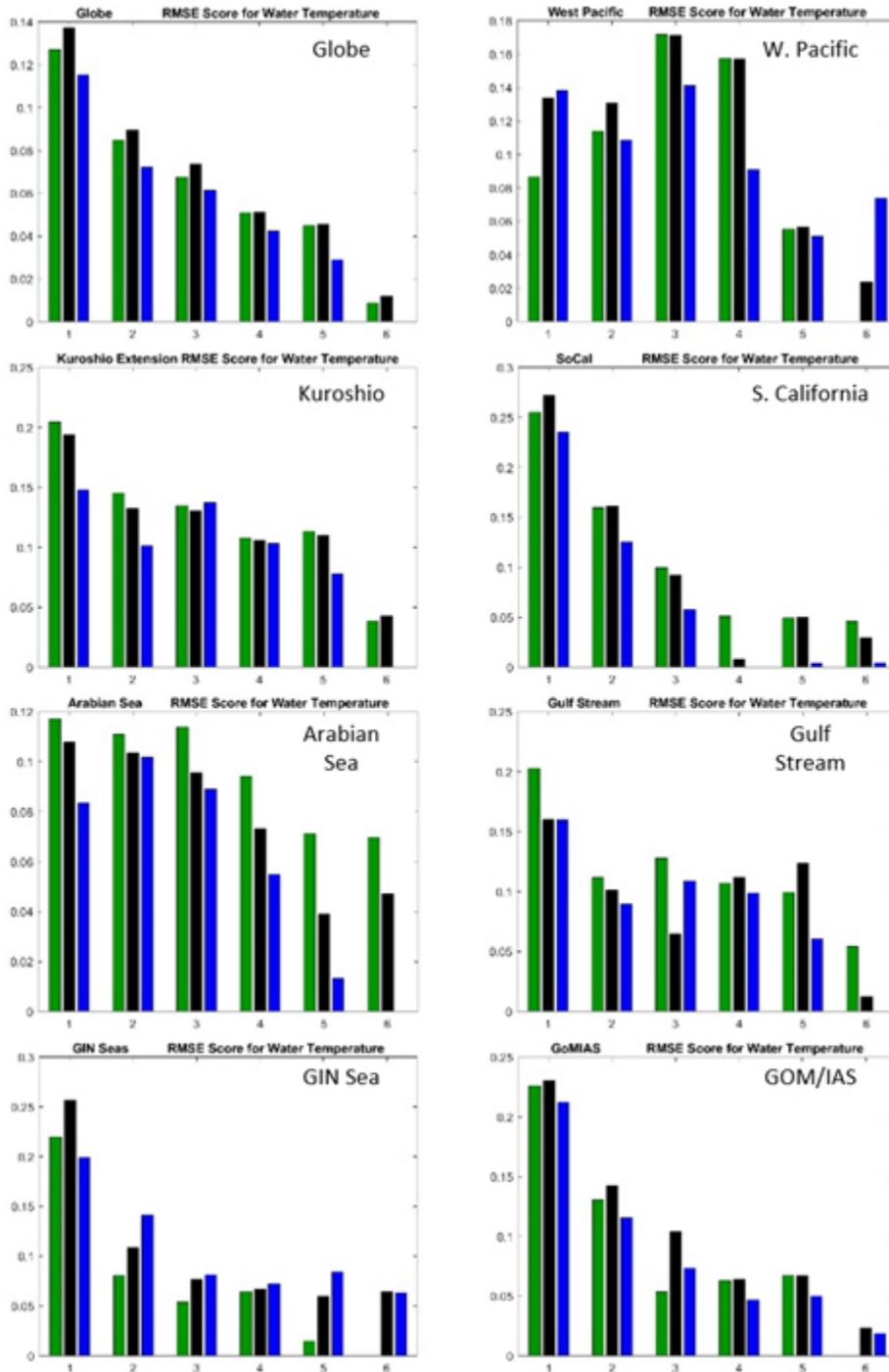


Figure 79:  $Score\_model(i)$  for temperature vs. depth as a function of forecast length. Higher numbers are better. The colors are GOFs 3.1 (blue), GOFs 3.5 (black) and Navy ESPC (green). This analysis is based on 72 forecasts starting on the 1<sup>st</sup>, 6<sup>th</sup>, 11<sup>th</sup>, 16<sup>th</sup>, 21<sup>st</sup> and 26<sup>th</sup> of each month.  $Score\_model(i)$  is normalized by  $Maximum\_error(i)$  across all forecast days, hence one histogram bar will always be zero, typically at day six.

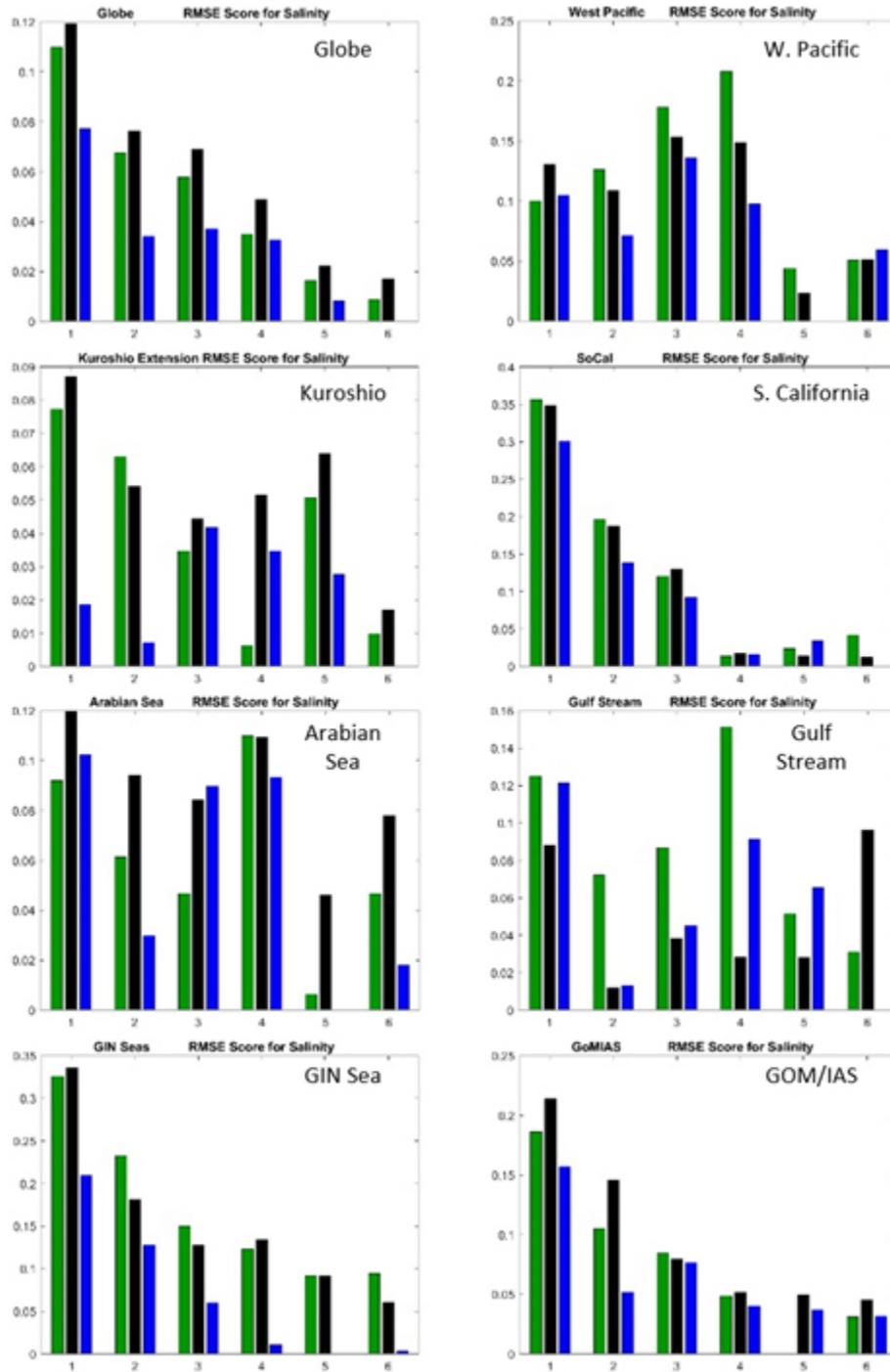


Figure 80: As Figure 79 except for salinity vs. depth as a function of forecast length.

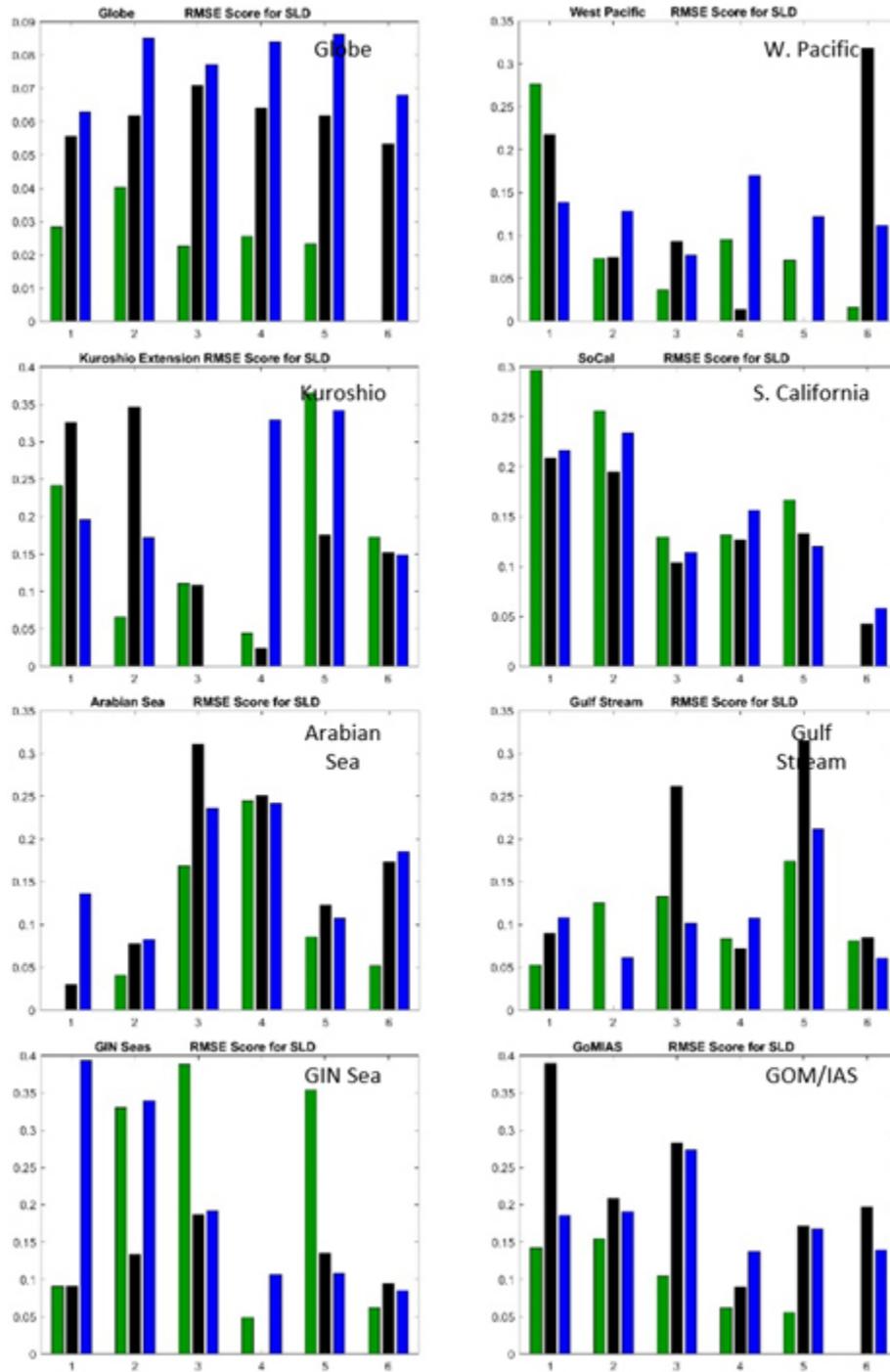


Figure 81: As Figure 79 except for sonic layer depth as a function of forecast length.

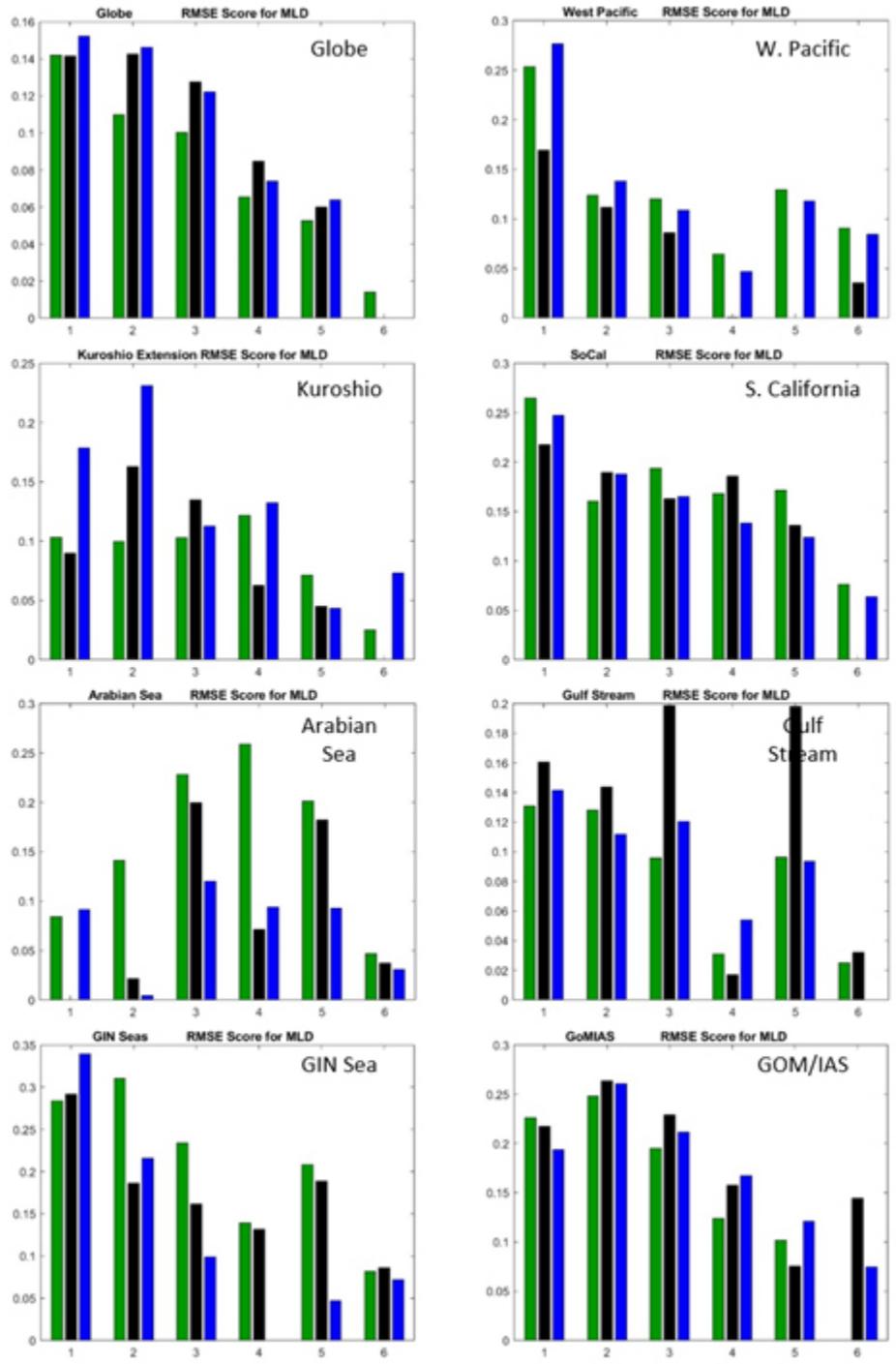


Figure 82: As Figure 79 except for mixed layer depth as a function of forecast length.

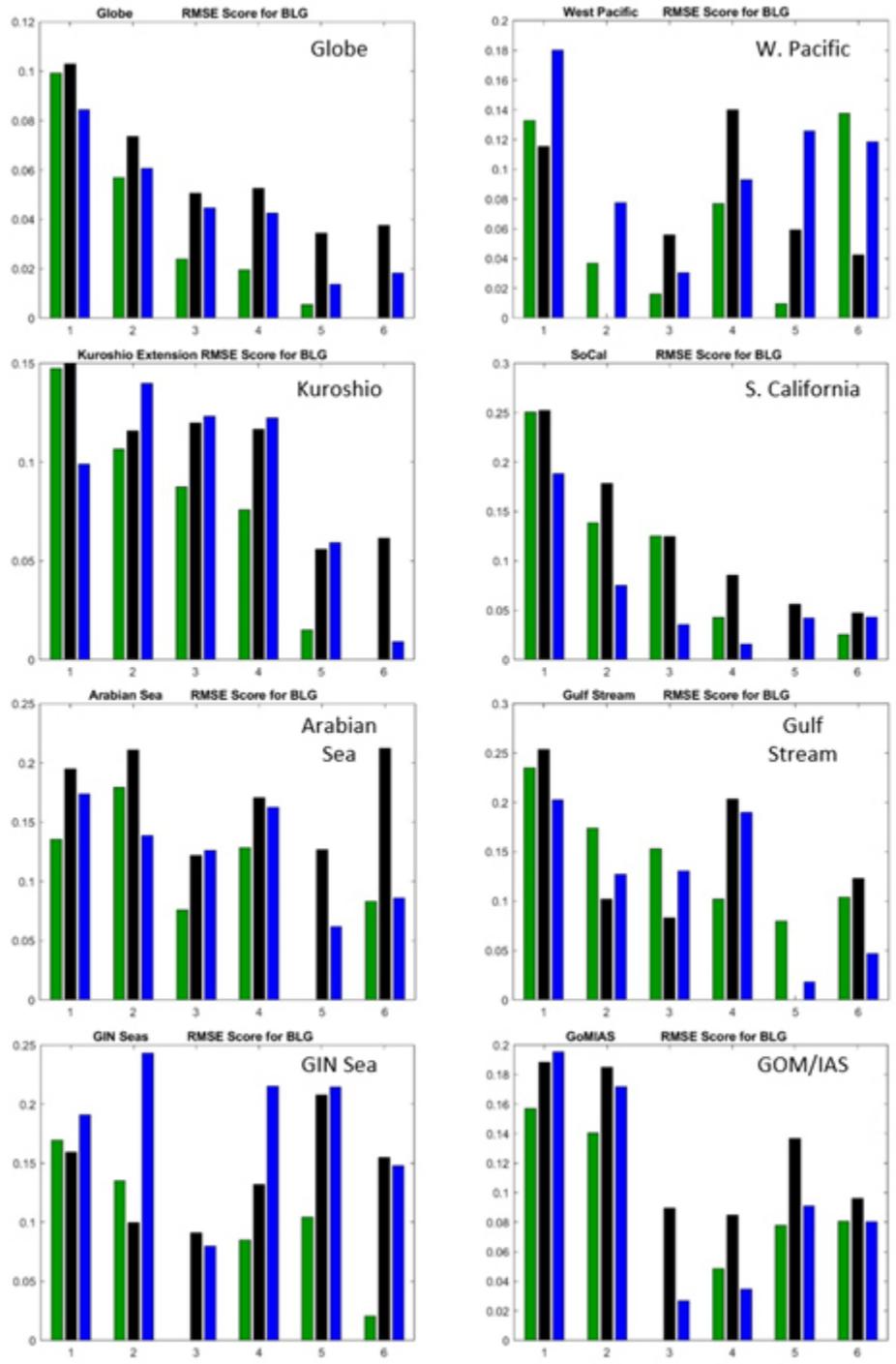


Figure 83: As Figure 79 except for below layer gradient as a function of forecast length.

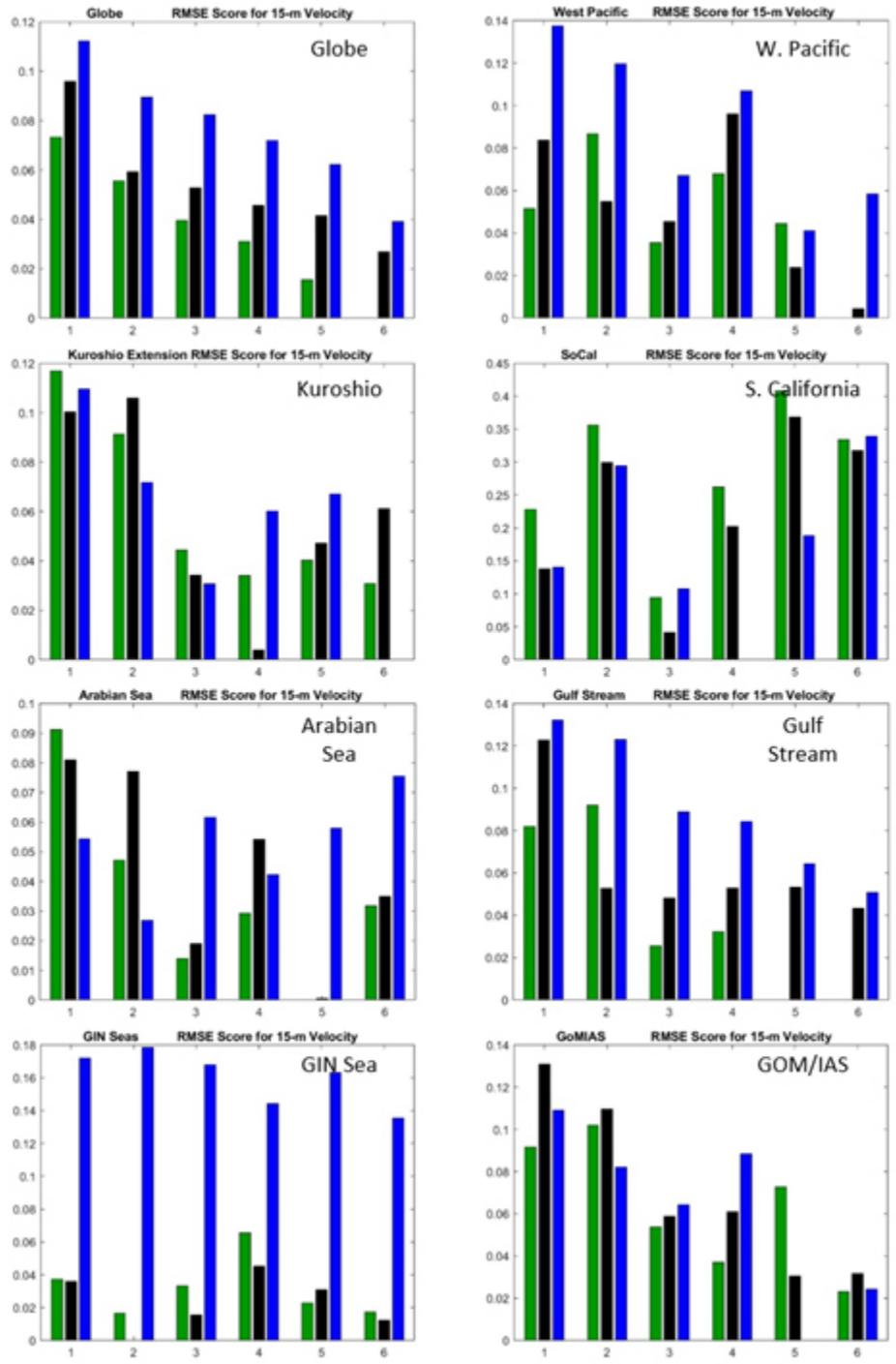


Figure 84: As Figure 79 except for 15 m ocean speed as a function of forecast length.

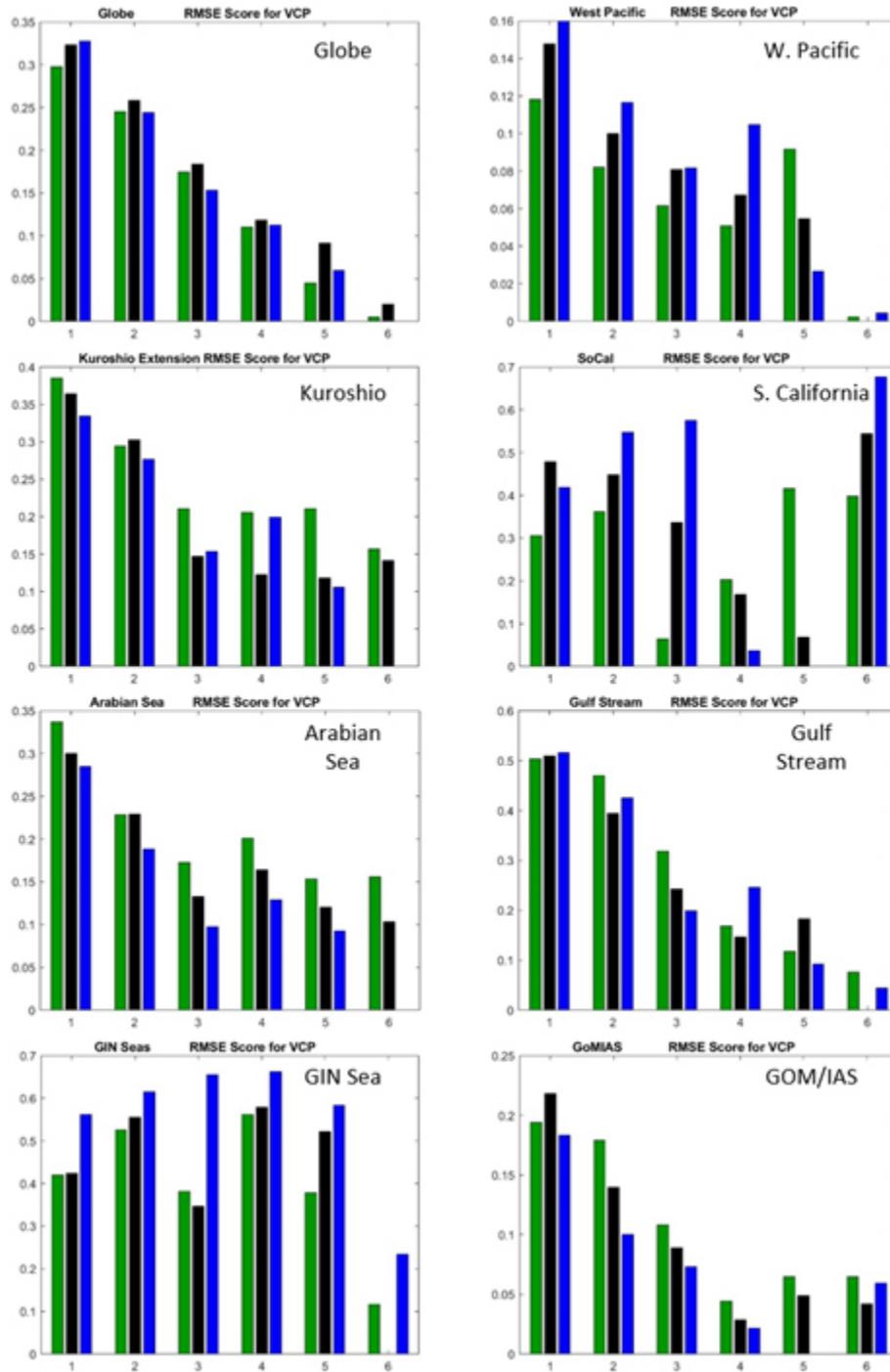


Figure 85: As Figure 79 except for 15 m ocean drift direction as a function of forecast length.

## 8 Appendix C (Acronyms)

AMSU - Advanced Microwave Sounding Unit A

MHS - Microwave Humidity Sounder

SSMIS - Special Sensor Microwave Imager Sounder

AIRS - Atmospheric InfraRed Sounder

ATMS - Advanced Technology Microwave Sounder

IASI - Infrared Atmospheric Sounder Interferometer

CrIS - Cross-track Infrared Sounder

SAPHIR - Sondeur Atmospherique du Profil d'Humidite Intertropicale par Radiometrie

GMI - Global Precipitation Mission (GPM) Microwave Imager

AMSR/2 - Advanced Microwave Scanning Radiometer - 2

GNSS-RO - Global Navigation Satellite System Radio Occultation

ECMWF - European Centre for Medium-Range Weather Forecasts

NOAA - National Oceanic and Atmospheric Administration

EMC - Environmental Modeling Center

## 9 References

- Ansong, J. K., B. K. Arbic, M. C. Buijsman, J. G. Richman, J. F. Shriver, and A. J. Wallcraft, 2015: Indirect evidence for substantial damping of low-mode internal tides in the open ocean. *J. Geophys. Res. Oceans*, **120**, 6057–6071, doi:10.1002/2015JC010998.
- Bleck, R., 2002: An oceanic general circulation model framed in hybrid isopycnic-Cartesian coordinates. *Ocean Model.*, **4**, 55-88, doi:10.1016/S1463-5003(01)00012-9.
- Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 1256-1271, doi:10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.
- Booij, N., R. C. Ris, and L. H. Holthuijsen, 1999: A third-generation wave model for coastal regions: 1. Model description and validation, *J. Geophys. Res.*, **104**, 7649-7666, doi:10.1029/98JC02622.

- Carnes, M. R., R. W. Helber, C. N. Barron and J. M. Dastugue, 2010: Validation Test Report for GDEM4. NRL Memo. Report, NRL/MR/7330--10-9271, 66 pp, doi:10.21236/ADA530343.
- 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**, 2504-2526, doi:10.1175/1520-0485(2003)033<2504:NASWTH>2.0.CO;2.
- Chawla, A., H. L. Tolman, J. L. Lanson, E. M. Devaliere, and V. M. Gerald, 2009: Validation of a multi-grid WAVEWATCH IIITM modeling system. NOAA/MMAB Contribution 281, 15 pp, [http://polar.ncep.noaa.gov/mxab/papers/tn281/multi\\_hindanalysis.pdf](http://polar.ncep.noaa.gov/mxab/papers/tn281/multi_hindanalysis.pdf).
- Chua, B., L. Xu, T. Rosmond, and E. Zaron, 2009: Preconditioning representer-based variational data assimilation systems: Application to NAVDAS-AR. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, S. K. Park and L. Xu, Eds., Springer-Verlag, 307–319.
- Clough, S. A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown. 2005: Atmospheric radiative transfer modeling: A summary of the AER codes. *J. Quant. Spectrosc. Ra.*, **91**, 233–244, doi:10.1016/j.jqsrt.2004.05.058.
- Crosby, D. S., L. C. Breaker and W. H. Gemmill, 1993: A proposed definition for vector correlation in geophysics: Theory and application. *J. Atmos. Oceanic. Technol.*, **10**, 355-367, doi:10.1175/1520-0426(1993)010<0355:APDFVC>2.0.CO;2.
- Cummings, J. A., 2005: Operational multivariate ocean data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3583-3604, doi:10.1256/qj.05.105.
- \_\_\_\_\_, and O. M. Smedstad, 2013: Variational data assimilation for the global ocean. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, S. K. Park and L. Xu, Eds., Vol. II, Springer, 303–343.
- Dee, D., 2004: Variational bias correction of radiance data in the ECMWF system. In Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June-1 July 2004, Reading, UK, pp. 97-112.
- Douglass, E. M. and A. C. Mask, 2019: Detection of fronts as a metric for numerical model accuracy. *J. Atmos. Oceanic. Technol.*, doi:10.1175/JTECH-D-18-0106.1.
- Dykes, J. D. and W. E. Rogers, 2013: Implementation of the multiple grid system of WAVEWATCH III at NAVOCEANO. NRL Memorandum Report: NRL/MR/7320-12-9494, 19 pp.
- Egbert, G. D. and S. Y. Erofeeva, 2002: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic. Technol.*, **19**, 183–204, doi:10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.
- Emanuel, K. A., 1991: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.*, **48**, 2313-2335 doi:10.1175/1520-0469(1991)048<2313:ASFRCC>2.0.CO;2.
- \_\_\_\_\_, and M. Zivkovic-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.*, **56**, 1766-1782, doi: 10.1175/1520-0469(1999)056<1766:DAEOAC>2.0.CO;2.
- Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm. *J. Climate*, **16**, 571-591, doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.
- Gregory, D., R. Kershaw and P. M. Inness, 1997: Parameterization of momentum transport by convection. II: Tests in single-column and general circulation models. *Quart. J. Roy. Meteor. Soc.*, **123**, 1153-1183, doi:10.1002/qj.49712354103.

- Halliwel, G. R., 2004: Evaluation of vertical coordinate and vertical mixing algorithms in the HYbrid Coordinate Ocean Model (HYCOM), *Ocean Model.*, **7**, 285–322, doi:10.1016/j.ocemod.2003.10.002.
- Han, J. and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520-533, doi:10.1175/WAF-D-10-05038.1.
- Heymsfield, A. J., and L. J. Donner, 1990: A scheme for parameterizing ice-cloud water content in general circulation models. *J. Atmos. Sci.*, **47**, 1865-1877, doi: 10.1175/1520-0469(1990)047<1865:ASFPIC>2.0.CO;2.
- Hill, C., C. DeLuca, V. Balaji, M. Suarez, A. da Silva, 2004: The architecture of the Earth System Modeling Framework. *Comput. Sci. Eng.*, **6**, 18-28, doi:10.1109/MCISE.2004.1255817.
- Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmospheric Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.*, **125**, 1414-1430, doi:10.1175/1520-0493(1997)125<1414:TNRLSC>2.0.CO;2.
- Hogan, T. F., 2007: Land surface modeling in the Navy Operational Global Atmospheric Prediction System; extended abstract. *AMS 22nd Conference on Weather Analysis and Forecasting / 189th Conference on Numerical Weather Prediction*; Park City, Utah; <https://ams.confex.com/ams/pdfpapers/123403.pdf>.
- Hunke, E. C., and W. H. Lipscomb, 2008: CICE: The Los Alamos sea ice model documentation and software user's manual, Version 4.0, Tech. Rep. LA-CC-06–012, Los Alamos Natl. Lab., Los Alamos, N. M., [http://www.cesm.ucar.edu/models/ccsm4.0/cice/ice\\_usrdoc.pdf](http://www.cesm.ucar.edu/models/ccsm4.0/cice/ice_usrdoc.pdf).
- Hyer, E. J., J. S. Reid, and J. Zhang, 2011: An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals. *Atmos. Meas. Tech.*, **4**, 379-408, doi:10.5194/amt-4-379-2011.
- Janiga, M. A., C. Schreck, J. A. Ridout, M. Flatau, N. Barton, E. J. Metzger, and C. Reynolds, 2018: Subseasonal forecasts of convectively coupled equatorial waves and the MJO: Activity and predictive skill. *Mon. Wea. Rev.*, **146**, 2338-2360, doi:10.1175/MWR-D-17-0261.1.
- Jensen, R. E., P. A. Wittmann, and J. D. Dykes, 2002: Global and regional wave modeling activities. *Oceanography*, **15**, 57-66, doi:10.5670/oceanog.2002.36.
- Kain, J. S., and M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802, doi:10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.
- \_\_\_\_\_, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 24, Amer. Meteor. Soc., 165–170.
- \_\_\_\_\_, 2004: The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor. Climatol.*, **43**, 170-181, doi:10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.
- Kara, A. B., H. E. Hurlburt, and A. J. Wallcraft, 2005: Stability-dependent exchange coefficients for air-sea fluxes. *J. Atmos. Oceanic Technol.*, **22**, 1080-1094, doi:10.1175/JTECH1747.1.
- Kim, H., M. A. Janiga., K. Pegion, 2019: MJO propagation processes and mean biases in the SubX and S2S reforecasts. *Submitted to J. Geophys. Res.-Atmos.*
- Kumar, B. P., J. Vialard, M. Lengaigne, V. S. N. Murty, and M. J. McPhaden, 2012. TropFlux: Air-sea fluxes for the global tropical oceans—description and evaluation. *Clim. Dynam.*, **38**, 1521-1543, doi:10.1007/s00382-011-1115-0.

- Kurtz, N. T., S. L. Farrell, M. Studinger, N. Galin, J. P. Harbeck, R. Lindsay, V. D. Onana, B. Panzer and J. G. Sonntag, 2013: Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data. *Cryosphere*, **7**, 1035-1056, doi:10.5194/tc-7-1035-2013.
- Lee J. Y., B. Wang, M. C. Wheeler, X. Fu, D. E. Waliser, I. S. Kang, 2013: Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Clim. Dynam.*, **40**, 493-509, doi:10.1007/s00382-012-1544-4.
- Louis, J., M. Tiedtke, and J. Geleyn, 1982: A short history of the PBL parameterization at ECMWF. *Proc. ECMWF Workshop on Planetary Boundary Layer Parameterization*, Reading, United Kingdom, ECMWF, 59–80.
- McCormack, J. P., S. D. Eckermann, D. E. Siskind, and T. E. McGee, 2006: CHEM2D-OPP: A new linearized gas-phase ozone photochemistry parameterization for high-altitude NWP and climate models. *Atmos. Chem. Phys.*, **6**, 4943-4972, doi:10.5194/acp-6-4943-2006.
- Moorthi, S., H.-L. Pan, P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Technical Procedures Bulletin, 484, pp 14, Available at <http://www.nws.noaa.gov/om/tpb/484.htm>.
- Metzger, E. J., H. E. Hurlburt, X. Xu, J. F. Shriver, A. L. Gordon, J. Sprintall, R. D. Susanto and H. M. van Aken, 2010: Simulated and observed circulation in the Indonesian Seas: 1/12° global HYCOM and the INSTANT observations. *Dynam. Atmos. Oceans*, **50**, 275-300, doi:10.1016/j.dynatmoce.2010.04.002.
- \_\_\_\_\_, O. M. Smedstad, P. G. Thoppil, J. A. Cummings, A. J. Wallcraft, L. Zamudio, D. S. Franklin, P. G. Posey, M. W. Phelps, P. J. Hogan, F. L. Bub and C. J. DeHann, 2014: US Navy operational global ocean and Arctic ice predictions systems. *Oceanography*, **27**, 32-43, doi:10.5670/oceanog.2014.66.
- Mitchell, K., 2005: The community Noah land-surface model. User's Guide Public Release Version 2.7.1, Environmental Modeling Center, NOAA/NCEP, 26 pp.
- Ngodock, H. E., I. Souopgui, A. J. Wallcraft, J. G. Richman, J. F. Shriver and B. K. Arbic, 2016: On improving the accuracy of the M2 barotropic tides embedded in a high-resolution global ocean circulation model. *Ocean Model.*, **97**, 16-26, doi:10.1016/j.ocemod.2015.10.011.
- Peng, M. S., J. A. Ridout, and T. F. Hogan, 2004: Recent modifications of the Emanuel convective scheme in the Navy Operational Global Atmospheric Prediction System. *Mon. Wea. Rev.*, **132**, 1254–1268, doi:10.1175/1520-0493(2004)132<1254:RMOTEC>2.0.CO;2.
- Pincus, R., H. W. Barker, and J.-J. Morcrette, 2003: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous clouds. *J. Geophys. Res.*, **108**, 4376, doi:10.1029/2002JD003322
- Reid, J. S., and Coauthors, 2009: Global monitoring and forecasting of biomass-burning smoke: Description of and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) program, *IEEE J.-S.T.A.R.S.*, **2**, 144-162, doi:10.1109/JSTARS.2009.2027443.
- Ridout, J. A., and C. A. Reynolds, 1998: Western Pacific warm pool region sensitivity to convective triggering by boundary layer thermals in the NOGAPS atmospheric GCM. *J. Climate*, **11**, 1553–1573, doi:10.1175/1520-0442(1998)011<1553:WPWPRS>2.0.CO;2.
- \_\_\_\_\_, Y. Jin, and C.-S. Liou, 2005: A cloud-base quasi-balance constraint for parameterized convection: Application to the Kain–Fritsch cumulus scheme. *Mon. Wea. Rev.*, **133**, 3315–3334, doi:10.1175/MWR3034.1.

- Ritchie, H., 1987: Semi-Lagrangian advection on a Gaussian grid. *Mon. Wea. Rev.*, **115**, 608-619, doi:10.1175/1520-0493(1987)115<0608:SLAOAG>2.0.CO;2.
- \_\_\_\_\_, 1988: Application of the semi-Lagrangian method to a spectral model of the shallow water equations. *Mon. Wea. Rev.*, **116**, 1587-1598, doi:10.1175/1520-0493(1988)116<1587:AOTSLM>2.0.CO;2.
- \_\_\_\_\_, 1991: Application of the semi-Lagrangian method to a multilevel spectral primitive equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 91-106.
- \_\_\_\_\_, C. Temperaton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud. 1995: Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, 489-514, doi:10.1175/1520-0493(1995)123<0489:IOTSLM>2.0.CO;2.
- Rogers, W. E., and T. J. Campbell, 2009: Implementation of curvilinear coordinate system in the WAVEWATCH-III Model. NRL Memorandum Report: NRL/MR/7320-09-9193, 42 pp, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a507120.pdf>.
- \_\_\_\_\_, J. D. Dykes, D. Wang, S. N. Carroll, and K. Watson, 2012: Validation Test Report for WAVEWATCH III. NRL Memorandum Report: NRL/MR/7320-12-9425, 73 pp, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a576706.pdf>.
- RP33: Fleet Oceanographic and Acoustic Reference Manual, 1992: Naval Oceanographic Office Reference Publication 33, Naval Oceanographic Office, Stennis Space Center, MS.
- Shriver, J. F., B. K. Arbic, J. G. Richman, R. D. Ray, E. J. Metzger, A. J. Wallcraft and P. G. Timko, 2012: An evaluation of the barotropic and internal tides in a high resolution global ocean circulation model. *J. Geophys. Res.*, **117**, C10024. doi:10.1029/2012JC008170.
- \_\_\_\_\_, J. G. Richman and B. K. Arbic, 2014: How stationary are the internal tides in a high-resolution global ocean circulation model? *J. Geophys. Res.-Oceans*, **119**, 2769–2787. doi:10.1002/2013JC009423.
- Slingo, J. 1987: The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **113**, 899-927, doi:10.1002/qj.49711347710.
- Sušelj, K., J. Teixeira and D. Chung, 2013: A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *J. Atmos. Sci.*, **70**, 1929-1953, doi:10.1175/JAS-D-12-0106.1.
- Teixeira, J. and T. F. Hogan, 2002: Boundary layer clouds in a global atmospheric model: Simple cloud cover parameterizations. *J. Climate*, **15**, 1261-1276, doi:10.1175/1520-0442(2002)015<1261:BLCIAG>2.0.CO;2.
- Tolman, T. L., B. Balasubramanian, L. D. Burroughs, D. V. Chalikov, Y. Y. Chao, H. S. Chen, and V. M. Gerald, 2002: Development and implementation of wind generated ocean surface wave models at NCEP. *Wea. Forecasting*, **17**, 311-333, doi:10.1175/1520-0434(2002)017<0311:DAIOWG>2.0.CO;2.
- \_\_\_\_\_, 2007: Toward a third release of WAVEWATCH III; a multi-grid model version, Tech. Note 251, NOAA/NWS/NCEP/MMAB, 12 pp.
- Vitart, F. and Coauthors, 2017: The Sub-seasonal to Seasonal Prediction (S2S) project database. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-16-0017.1.
- Webster, S., A. R. Brown, D.R. Cameron, and C.P. Jones, 2003: Improvements to the representation of orography in the Met Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **133**, 1989-2010, doi:10.1256/qj.02.133.

- Wheeler, M. C. and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, doi:10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.
- Wittmann, P. A., 2002: Implementation of WAVEWATCH III at Fleet Numerical Meteorology and Oceanography Center. Conf. Proceedings: *MTS/IEEE: Conference and Exposition*. Nov 5-8, 2001 Honolulu, HI, 1474-1479.
- Zamudio, L., P. Spence and G. A. Jacobs, 2015: Verification and validation report for Navy operational regional forecast models. Part II: Evaluation. NRL internal report, 30 April 2015.
- Zhang, J. L., and J. S. Reid, 2006: MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals. *J. Geophys. Res.-Atmos.*, **111**, D22207, doi:10.1029/2005JD006898.
- \_\_\_\_\_, J. S. Reid, and B. N. Holben, 2005: An analysis of potential cloud artifacts in MODIS over ocean aerosol optical thickness products. *Geophys. Res. Lett.*, **32**, L15803, doi: 10.1029/2005GL023254.
- \_\_\_\_\_, J. S. Reid, D. L. Westphal, N. L. Baker, and E. J. Hyer, 2008: A system for operational aerosol optical depth data assimilation over global oceans. *J. Geophys. Res.-Atmos.*, **113**, D10208, doi:10.1029/2007JD009065.
- Zhao, Q. Y., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931-1953, doi:10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2.