Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/7323--19-9975

Validation Test Report for the Global Ocean Forecast System 3.5 – 1/25º HYCOM/CICE with Tides

E.J. Metzger
P.J. Hogan
J.F. Shriver
P.G. Thoppil
E. Douglass
Z. Yu
R.A. Allard
C.D. Rowley

Ocean Dynamics & Prediction Branch
Ocean Sciences Division

O.M. Smedstad
D.S. Franklin
M.W. Phelps

Perspecta, Inc.
Chantilly, VA

L. Zamudio
A.J. Wallcraft
J.R. Richman

Florida State University
Tallahassee, FL

October 27, 2020

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
27-10-2020	NRL Memorandum Report	

4. TITLE AND SUBTITLE

Validation Test Report for the Global Ocean Forecast System 3.5 – 1/25º HYCOM/CICE with Tides

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

E.J. Metzger, P.J. Hogan, J.F. Shriver, P.G. Thoppil, E. Douglass, Z. Yu, R.A. Allard, C.D. Rowley, O.M. Smedstad*, D.S. Franklin*, M.W. Phelps*, L. Zamudio**, A.J. Wallcraft**, and J.R. Richman**

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
1L30

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/7323--19-9975

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)

ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

*Perspecta Inc., 14295 Park Meadow Drive, Chantilly, VA 20151
**Florida State University, 600 W College Ave, Tallahassee, FL 32306

14. ABSTRACT

The Global Ocean Forecast System (GOFS) 3.5 is comprised of the 1/25° HYbrid Coordinate Ocean Model that is two -way coupled to the Community Ice CodE in a daily update cycle assimilating real-time ocean/sea ice observations using the Navy Coupled Ocean Data Assimilation. GOFS predicts the ocean's "weather", which includes the three-dimensional ocean temperature, salinity and current structure, the surface mixed layer, the location of mesoscale features, and sea ice concentration, thickness and drift in both hemispheres. GOFS 3.5 is scheduled to replace GOFS 3.1 which is the existing operational system at Fleet Numerical Meteorology and Oceanography Center (FNMOC). This report describes the system comparisons against unassimilated observations at both the nowcast time and as a function of forecast length. An ocean scorecard provides the relative performance of GOFS 3.5 vs. GOFS 3.1. Overall, GOFS 3.5 is performing equal to or better than GOFS 3.1, and is recommended to move forward for operational testing by FNMOC.

15. SUBJECT TERMS

HUCOM	NCODA	Global ocean nowcast/forecast system
CICE	Model validation and verification	Internal tides

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Joe Metzger
Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	Unclassified Unlimited	70	**19b. TELEPHONE NUMBER** *(include area code)* (228) 688-4762

This page intentionally left blank.

Contents

1. INTRODUCTION 4

2. GOFS 3.1 / GOFS 3.5 CONFIGURATION DIFFERENCES 4

2.1 HYCOM 4

2.2 CICE 6

2.3 NCODA 7

2.4 GOFS runstream configuration 7

2.5 NAVGEM forcing 8

2.6 GOFS 3.5 timings 9

3. OCEAN VALIDATION TESTING RESULTS 9

3.1 Ocean validation 9

3.2 Temperature/salinity vs. depth error analysis 10

3.3 Isotherm depth error analyses 19

3.4 Acoustical proxy error analyses 22

3.5 Surface layer trapping of acoustic frequencies 28

3.6 Upper ocean velocity validation 30

3.7 Frontal placement validation 32

3.8 Tidal elevation error analysis 33

3.9 Eddy kinetic energy climatology 36

3.10 Ocean score card 38

4. SEA ICE VALIDATION TESTING RESULTS 49

4.1 Sea ice validation 49

4.2 Sea ice edge location error analyses 51

4.3 Sea ice thickness error analyses 57

4.4 Sea ice drift error analyses 62

5. SUMMARY AND RECOMMENDATIONS 65

6. ACKNOWLEDGMENTS 66

7. REFERENCES 67

8. TABLE OF ACRONYMS 71

1. INTRODUCTION

This Validation Test Report (VTR) documents the continuing development, validation, verification and transition of the Global Ocean Forecast System (GOFS) for operational use at the Fleet Numerical Meteorology and Oceanography Center (FNMOC) and run on the Navy DoD Supercomputing Resource Center (DSRC) supercomputers. GOFS 3.5 is based on the 1/25° global HYbrid Coordinate Ocean Model (HYCOM) that is two-way coupled to the Community Ice CodE (CICE) and uses the Navy Coupled Ocean Data Assimilation (NCODA). Pending successful validation by NRL and OPerational TESTing (OPTEST) by FNMOC and the National Ice Center (NIC), it is scheduled to replace 1/12° GOFS 3.1 detailed in Metzger et al. (2014, 2017).

This VTR will focus on ocean and sea ice error analyses for GOFS 3.1 and 3.5 against unassimilated observations. The metrics used are similar to previous VTRs and for the ocean include: temperature (T) and salinity (S) vs. depth, acoustical proxies [sonic layer depth (SLD), mixed layer depth (MLD), below layer gradient (BLG), cut-off frequency (COF)], isotherm depths, upper ocean velocity and frontal location. Sea ice edge, thickness and drift are the cryospheric metrics.

2. GOFS 3.1 / GOFS 3.5 CONFIGURATION DIFFERENCES

2.1 HYCOM

The ocean model software in GOFS 3.1 is HYCOM version 2.2.99Gi vs. HYCOM version 2.2.99DHi for GOFS 3.5. The latter is consistent with the code base used in the Earth System Prediction Capability (ESPC) project. While the software in GOFS 3.1 is more up-to-date than in GOFS 3.5, all upgrades are associated with the research version of HYCOM and are not implemented in the operational system. Thus HYCOM uses the same underlying physics package here in both GOFS 3.1 and 3.5.

The first configuration difference is associated with the horizontal resolution. Ocean models require finer spatial resolution and more computer time than atmospheric models (for example) in part because the space scales for variability due to flow instabilities (oceanic mesoscale vs. atmospheric highs and lows) are about 20-30 times smaller than found in the atmosphere. We need to resolve the oceanic eddy space-scale very well because (1) it is relevant

Manuscript approved October 27, 2020.

for most Navy applications, (2) GOFS needs to provide high resolution boundary conditions to even higher resolution coastal models, (3) upper ocean - topographic coupling via flow instabilities has a major impact on the pathways of many upper ocean currents (including mean pathways) and very fine resolution of the flow instabilities is required to get sufficient coupling, (4) very fine resolution is required to obtain (a) inertial jets and sharp oceanic fronts which span major ocean basins as observed and (b) the associated nonlinear recirculation gyres which affect the shape of large-scale ocean gyres, (5) it is necessary to resolve small islands and narrow passages in many regions, (6) in data assimilative mode we do not want the ocean model to "fight" the data because the natural behavior of the ocean model is inconsistent with the observations, and (7) a very high resolution model is needed to help get an accurate mean sea surface height field to add to the deviations from satellite altimetry (observations alone do not provide sufficient resolution to determine this). NRL research suggests that about 3.5 km mid-latitude resolution is needed globally. Research is on-going to study how increased vertical resolution impacts the water column, especially with regard to sound speed.

GOFS 3.1 is on a 4500 x 3298 tripole global grid (0.08° or 8.9 km equatorial resolution, 6.8 km mid-latitude resolution and ~3.5 km resolution at the North Pole), whereas GOFS 3.5 is on a 9000 x 7055 grid (0.04° or 4.4 km equatorial resolution, 3.4 km mid-latitude resolution and 1.8 km resolution at the North Pole). The model coastline in both is the 0.1m isobath but with depths shallower than 5 m set to 5 m and depths greater than 6500 m scaled by 0.2 (maximum resulting depth is 7200 m). The southernmost latitude in GOFS 3.5 extends to 86.0°S (vs. 78.64°S in GOFS 3.1), which includes the ocean under the floating Antarctic ice shelves and is needed for improved tidal simulation. Both systems contain 41 vertical hybrid coordinate layers.

The second configuration difference is the implementation of astronomical tidal forcing in GOFS 3.5 that generates internal gravity waves at tidal frequencies. Also known as internal tides, they are generated by large-scale barotropic tidal flow over topographic features that generate vertical motion. In the stratified fluid in GOFS 3.5, this means the interfaces between layers will oscillate vertically. Internal tides have smaller sea surface height signatures and smaller horizontal scales than barotropic tides, large undulations at depth and vertically sheared velocity fields. See Arbic et al. (2018) for a detailed discussion of the development of tidal forcing within HYCOM.

GOFS 3.5 includes the geopotential tidal forcing for the three largest semidiurnal constituents (M_2, S_2, and N_2) and two largest diurnal constituents (K_1 and O_1). The tidal forcing is augmented with a self-attraction and loading (SAL) term accounting for the load deformations of the solid earth and the self-gravitation of the tidally deformed ocean and solid earth (Hendershott, 1972; Ray, 1998). An Augmented State Ensemble Kalman Filter technique is applied to reduce errors in the model's barotropic tidal sea surface elevations (Ngodock et al., 2016). The wave drag parameterization by Jayne and St. Laurent (2001) is used here. Following Buijsman et al. (2015), this wave drag is tuned to minimize tidal sea surface elevation errors with respect to the TPXO8-atlas (Egbert et al., 1994).

2.2 CICE

The Los Alamos-developed CICE model (Hunke and Lipscomb, 2008) is two-way coupled to HYCOM 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.

A major difference with regard to cryospheric prediction is that GOFS 3.1 uses CICE v4.0 whereas GOFS 3.5 employs CICE v5.1.2 (but still uses CICE v4.0 parameters). This newer version includes the correction of known software bugs, two new melt pond parameterizations, an improved parameterization for form drag, use of sea ice velocity in the coupling updates for high frequency coupling, and Elastic-Anisotrophic-Plastic rheology to name a few. As will be noted below, this version difference has an especially big impact on sea ice edge location error.

2.3 NCODA

NCODA is designed to represent the 3D ocean mesoscale by creating synthetic profile data that projects surface information downward into the water column for assimilation into the forecast model. This is done using the Improved Synthetic Ocean Profile (ISOP) system (Helber et al., 2013) and is used by both GOFS 3.1 and 3.5. The ISOP system employs a versatile 1D variational approach to utilize historical vertical covariances that not only preserves the relationship between the surface and subsurface ocean properties but also the vertical gradients of temperature and salinity. Assimilation of synthetic profiles with more accurate vertical structure results in more accurate prediction of the ocean vertical structure which in turn results in more accurate acoustic propagation/transmission loss predictions.

GOFS 3.1 uses NCODA 3.99 whereas GOFS 3.5 uses NCODA 4.1. The latter includes support for more recent satellite sources. Another significant change is related to profile thinning, which removes near duplicate profiles based on the correlation length scale, rather than the grid spacing. The effect is to remove more profiles, making the analysis less likely to have convergence problems in the solver and thus more robust.

2.4 GOFS runstream configuration

The GOFS 3.1 and 3.5 runstream in an operational environment (or when run as a reanalysis) is depicted in Figure 1. The NCODA analysis is performed at 12Z and uses the First Guess at Appropriate Time (FGAT) capability. Comparing ocean observations to model background fields valid at different times is a source of error for the NCODA analysis. This error is reduced by comparing observations against time-dependent background fields using the FGAT technique (Cummings and Smedstad, 2014). Thus, late arriving observations can be assimilated without the need to do a complete NCODA analysis N days back from the nowcast time (as was done in GOFS 3.0). HYCOM/CICE is run forward starting at 09Z with the incremental analysis update (Bloom et al., 1996) window for the ocean between 09Z → 12Z, i.e. starting before the analysis time. Thus at 12Z HYCOM has fully assimilated all the observational data and the nowcast time (tau 000) and NCODA analysis time are synchronized.

NCODA uses direct insertion for the sea ice analysis into CICE at 09Z. While this creates a three hour offset in the nowcast time between the ocean and sea ice, the sea ice nowcast time is

arbitrarily shifted to 12Z for consistency with the ocean because the ice fields do not undergo large changes over this three hour window.

In an operational environment, the models run forward for 7.5 days (tau 180) each day. For this VTR, we did not produce 7.5 day forecasts every single day due to computational resource constraints. Instead the system cycled each day and ran forward long enough to create the first guess field (i.e. the 24 hour average between tau 012 and tau 036) needed for the next day's assimilation cycle. Then on the 1st, 6th, 11th, 16th, 21st and 26th of each month, multi-day forecasts were created for a total of 72 total forecasts. The GOFS 3.5 forecasts extended 7.5 days but GOFS 3.1 forecasts stopped at 6.5 days due to an oversight in the scripting.

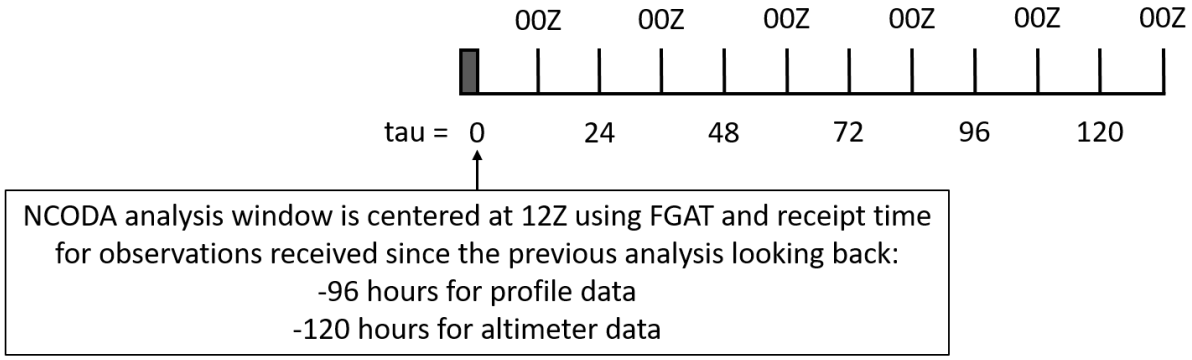

Figure 1: The GOFS 3.1/3.5 runstream configuration in an operational environment. The forecast length is only shown to 5.5 days (tau 132) but actually extends to 7.5 days (tau 180). The gray box indicates the ocean incremental insertion window between 09Z – 12Z.

2.5 NAVGEM forcing

For the reanalysis year (2017) and 72 multi-day forecasts, both GOFS 3.1 and 3.5 use atmospheric forcing from NAVGEM 1.4 that has a spectral resolution of T425 (~31 km grid spacing). The GOFS forecasts used forecast quality NAVGEM forcing. The 3-hourly atmospheric output are interpolated to the HYCOM/CICE grids and the following fields are used: 2 m air temperature, 2 m specific humidity, downward surface shortwave and longwave radiation, mean sea level pressure, total (convective and large-scale) precipitation, surface temperature and 10 m wind velocity. Bulk formulae are used to compute the net surface heat fluxes using the scheme of

Kara et al. (2005) that is an adaptation of the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean Atmosphere Response Experiment (COARE) 3.0 scheme of Fairall et al. (2003). Because of the tidal forcing in GOFS 3.5, it additionally uses dynamic mean sea level pressure forcing while this is not used in GOFS 3.1.

2.6 GOFS 3.5 timings

As GOFS 3.5 cycles forward in time each day, a series of scripts are executed to perform individual tasks such as gathering the ocean/sea ice observations, interpolating the atmospheric forcing to the model grids, performing the NCODA analyses for various subdomains, running the HYCOM/CICE forecast, and post-processing model output. Some of these tasks can run in parallel, while others must successfully complete before subsequent tasks can be executed. The majority of these tasks only require a single computational core (processer), but the NCODA tasks may need O(500) cores and the forecast model O(3000) cores. HYCOM and CICE run simultaneously on separate cores with HYCOM using more cores for load balance, and both have good scalability so more cores can be used to reduce overall walltime if necessary. On the Navy DSRC Hewlett Packard Enterprise SGI 8600 supercomputer, HYCOM/CICE is presently configured to run using 2646/600=3246 cores and takes ~9.2 walltime hours to complete the forecast through tau 180.

3. OCEAN VALIDATION TESTING RESULTS

A year-long reanalysis was integrated for both GOFS 3.1 and GOFS 3.5 to emulate a real-time operational system, assimilating all available observational data. (In NRL nomenclature these are GLBb0.08-57.9 and GLBc0.04-21.6, respectively.) The system cycled forward one day at a time over calendar year 2017 and multi-day forecasts were computed as described in Section 2.4. No ocean/sea ice data were assimilated into the forecasts, thus all observational data used for validation are independent.

3.1 Ocean validation

The ocean error analyses used for validation here are similar to those used in previous GOFS VTRs, namely: 1) temperature/salinity vs. depth, 2) depths of specific isotherms, 3) acoustical proxies, 4) predictions of surface layer trapping of acoustical frequencies, 5) validation of upper ocean currents against drifting buoys, 6) frontal placement, 7) tidal elevation, and 8) eddy

kinetic energy. The analysis regions (Figure 2) have been chosen to be both Navy relevant and dynamically diverse.

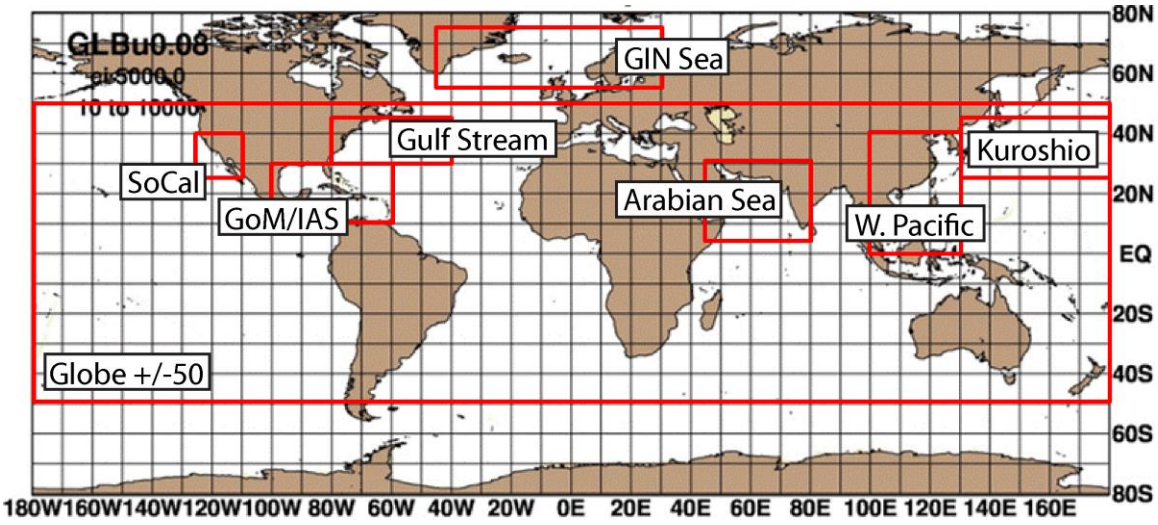

Figure 2: The ocean analysis regions used in this report: Globe: 180°W-180°E, 50°S-50°N; West Pacific marginal seas: 100-130°E, 0-40°N; Kuroshio Extension: 130°E-180°, 25-45°N; Arabian Sea/Persian Gulf/Gulf of Aden: 44-80°E, 5-30°N; Gulf Stream: 80-40°W, 30-45°N; Southern California (SoCal): 125-110°W, 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.

3.2 Temperature/salinity vs. depth error analysis

A temperature/salinity vs. depth error analysis within the top 500 m of the water column is performed using unassimilated profiles. For a given observation, both GOFS 3.5 and GOFS 3.1 are sampled at the nearest model grid point. In the vertical, both the observations and model output are 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 that exceed three standard deviations are excluded, i.e. the 99% confidence interval is used, and this is the reason why the observation counts differ between the two systems. Figure 3 illustrates how profile data are sampled. In order to make sure only unassimilated profiles are used in the error analysis, profiles over a 12-hour window from 00Z (tau 012) to 12Z (tau 024) are compared back to the nowcast time (tau 000). When sampling the forecasts, profiles ±1.5 hours on either side of the 3-hourly HYCOM archive files centered ±12 hours over 12Z are used.

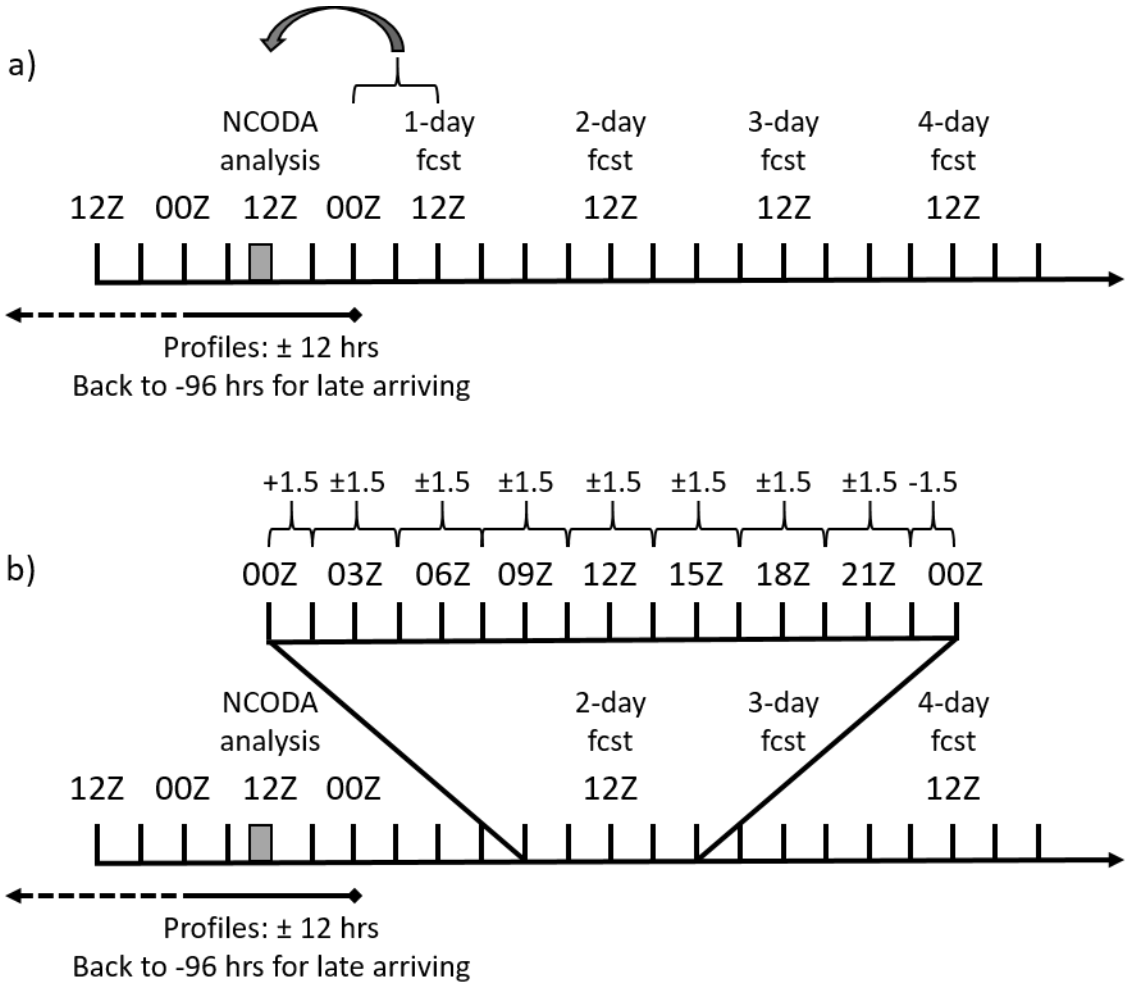

Figure 3: Methodology for sampling profile data from HYCOM archive files a) at the nowcast time and b) at a forecast time. Here the 2-day forecast is shown. The gray box shows the time window for incremental insertion.

Figure 4 and Figure 5 show mean error (ME) (bias) and root mean square error (RMSE) for temperature in the eight analysis regions at the nowcast time. In general, the ME is small (<0.25°C) in both systems and always well within the FNMOC-defined tolerances for the upper ocean (0.5°C). Over most regions and depths there is a small overall cool bias in both systems. With regard to the individual analysis regions, the ME and RMSE results are mixed with one system performing slightly better than the other depending upon the region that is selected. The largest differences are in the western boundary current regions (Kuroshio Extension and Gulf Stream) with GOFS 3.5 outperforming GOFS 3.1. These resulted will be quantified below when

the scores are computed. As expected, the highest RMSE is seen at the depth level of the thermocline.

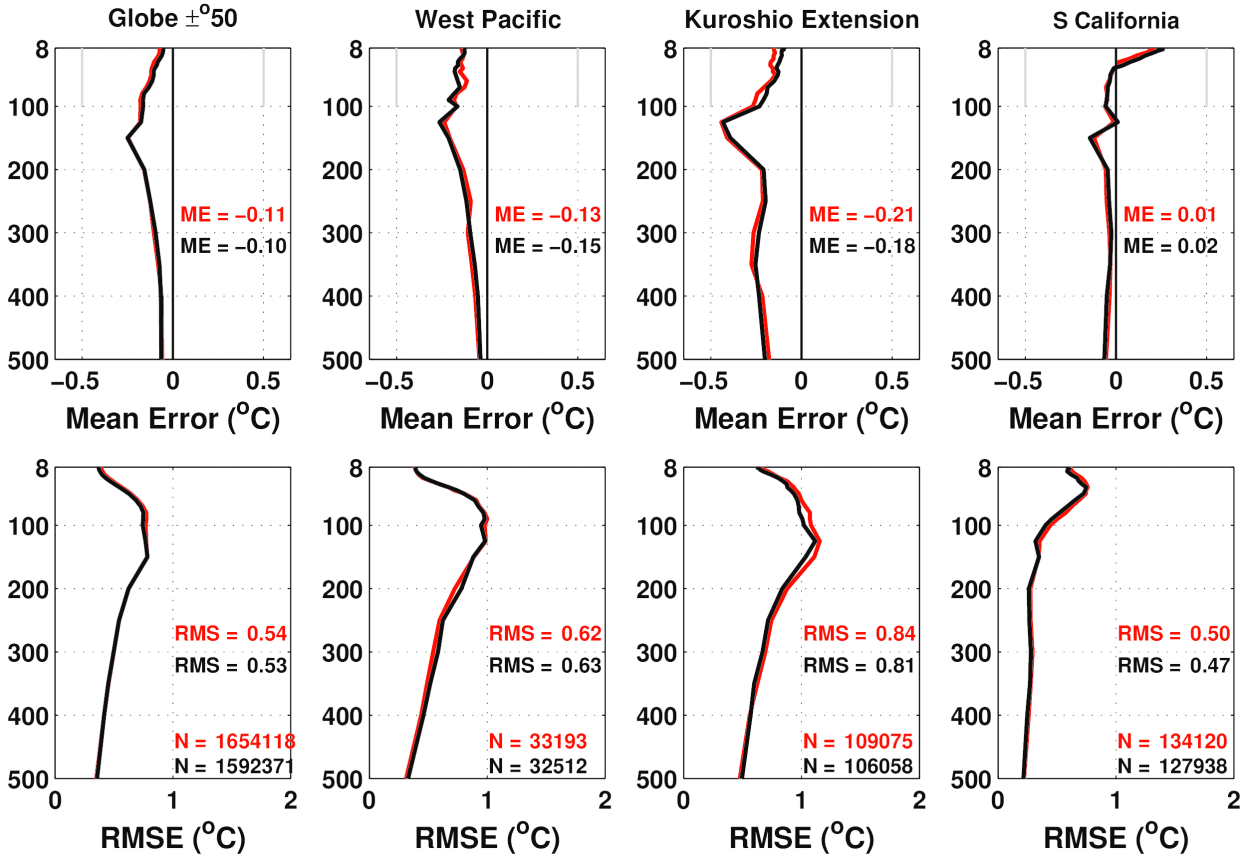

Figure 4: Temperature (°C) vs. depth error analysis in the upper 500 m against unassimilated profile observations at the nowcast time for four regions (Global, West Pacific, Kuroshio Extension, and Southern California in the columns from left to right, respectively) defined in Figure 2 spanning the reanalysis period January – December 2017. Top row is mean error (ME) and bottom row is root mean square error (RMSE). The red curves are from GOFS 3.1 and the black curves are GOFS 3.5. The numeric values for ME/RMSE are averages from 8-500 m. N represents the number of observations used in each region, e.g. if a single profile samples 18 depths, that number is added to the others to define the total. N/24 gives the approximate number of profiles in each region. The minimum depth is 8 m because that is often the last depth sampled by Argo profiles. The gray lines in the ME plots are the tolerances set by FNMOC for the temperature bias.

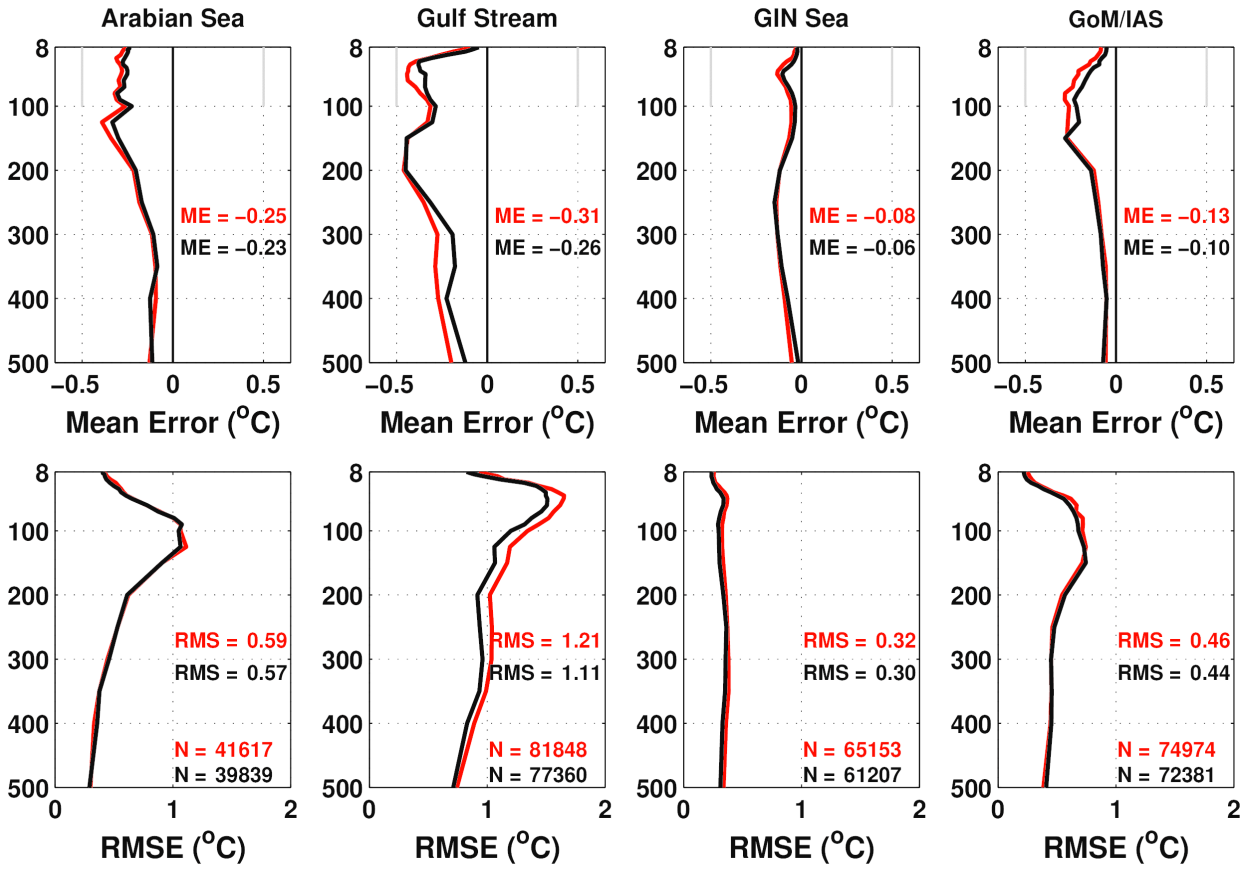

Figure 5: As Figure 4 except for the Arabian Sea, Gulf Stream, GIN Sea and Gulf of Mexico/Intra-Americas Sea regions from left to right, respectively.

A salinity vs. depth error analysis is shown in Figure 6 and Figure 7. Similar to temperature, the two systems have comparable error at most depths with the exception of the very near surface (approximately the top 40 m) where GOFS 3.5 has lower error than GOFS 3.1.

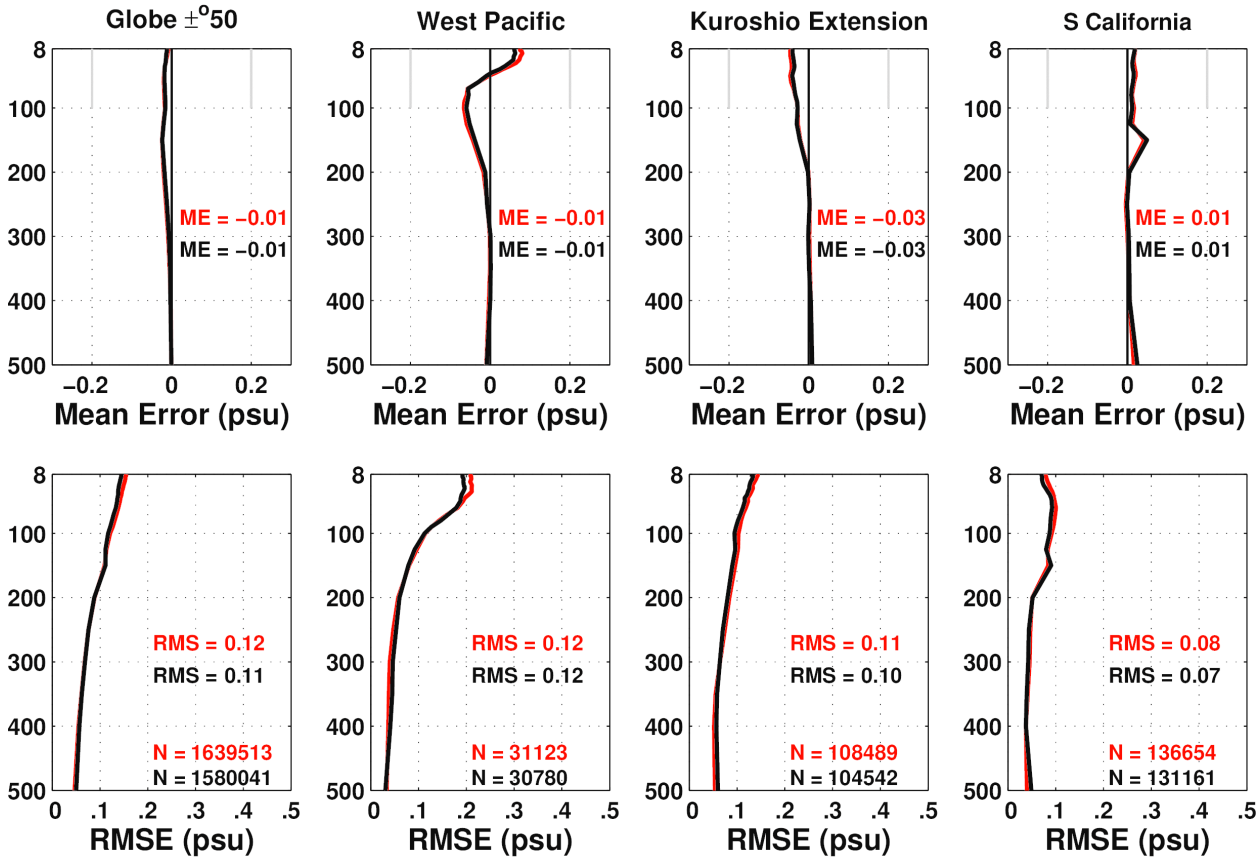

Figure 6: As in Figure 4 except for salinity (psu).

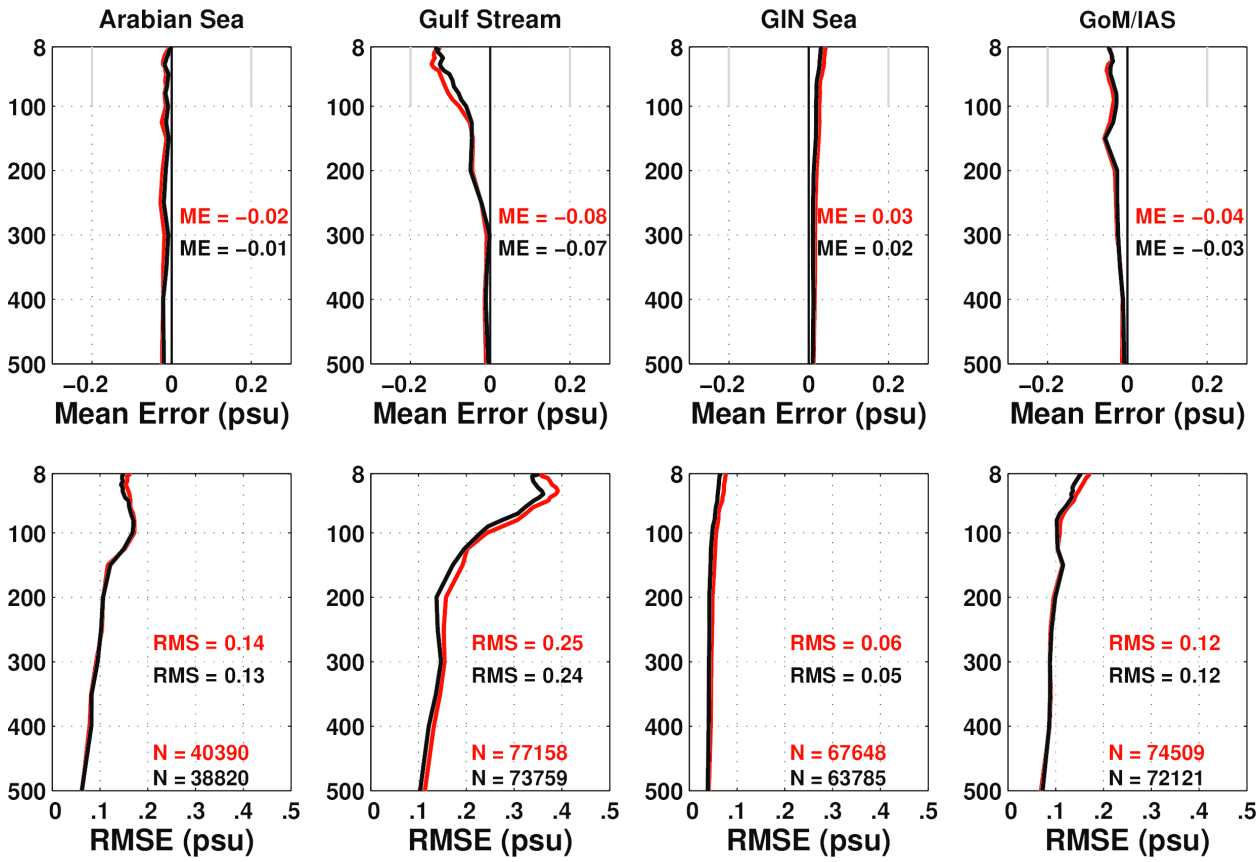

Figure 7: As in Figure 5 except for salinity (psu).

Upper ocean temperature and salinity error analyses as a function of tactical forecast length (7-days) are shown for GOFS 3.5 in Figure 8 - Figure 11. The black/red/cyan/green curves show the 1-day/3-day/5-day/7-day forecasts, respectively. The change in ME vs. forecast length is typically largest near the surface, but the degradation in skill as the forecast length increases is still small nonetheless. The error growth with forecast length is more obvious for RMSE, and averaged over the top 500 m, the global domain error between the nowcast time and the 7-day forecast grows by 10% (0.58 to 0.64 °C) and 8% (0.12 to 0.13 psu) for temperature and salinity, respectively. The error growth is less consistent for the regional domains and there are times when the longer forecast error is smaller than the 1-day forecast error, although, the difference is generally small in such cases. The largest spread for both T and S is in the Gulf Stream domain. Forecast error growth for GOFS 3.1 is not shown here, but will be quantified against GOFS 3.5 in the ocean score section.

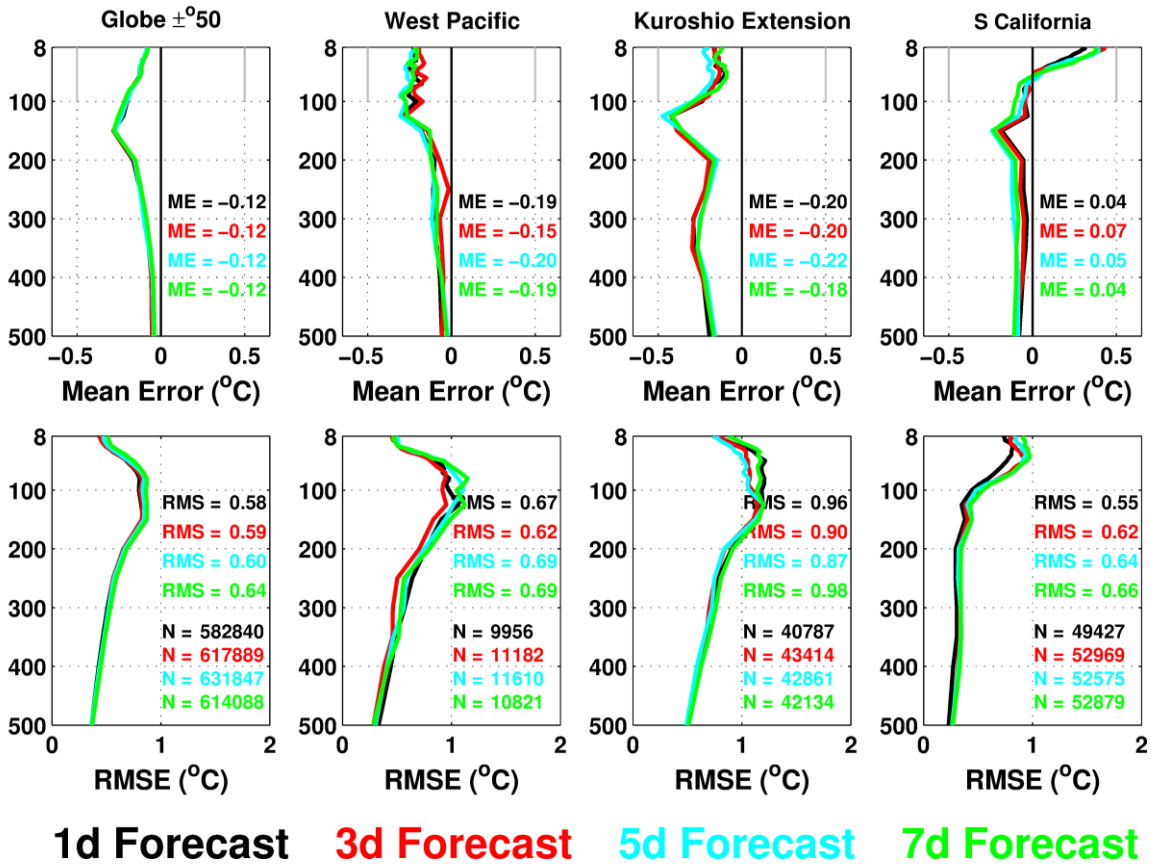

Figure 8: Temperature (°C) vs. depth error analysis of GOFS 3.5 in the upper 500 m against unassimilated profile observations at four forecast lengths (1-day: black, 3-day: red, 5-day: cyan and 7-day: green) for four regions (Global, West Pacific, Kuroshio Extension, and Southern California in the columns from left to right, respectively) spanning the reanalysis period January – December 2017. Top row is mean error (ME) and bottom row is root mean square error (RMSE).

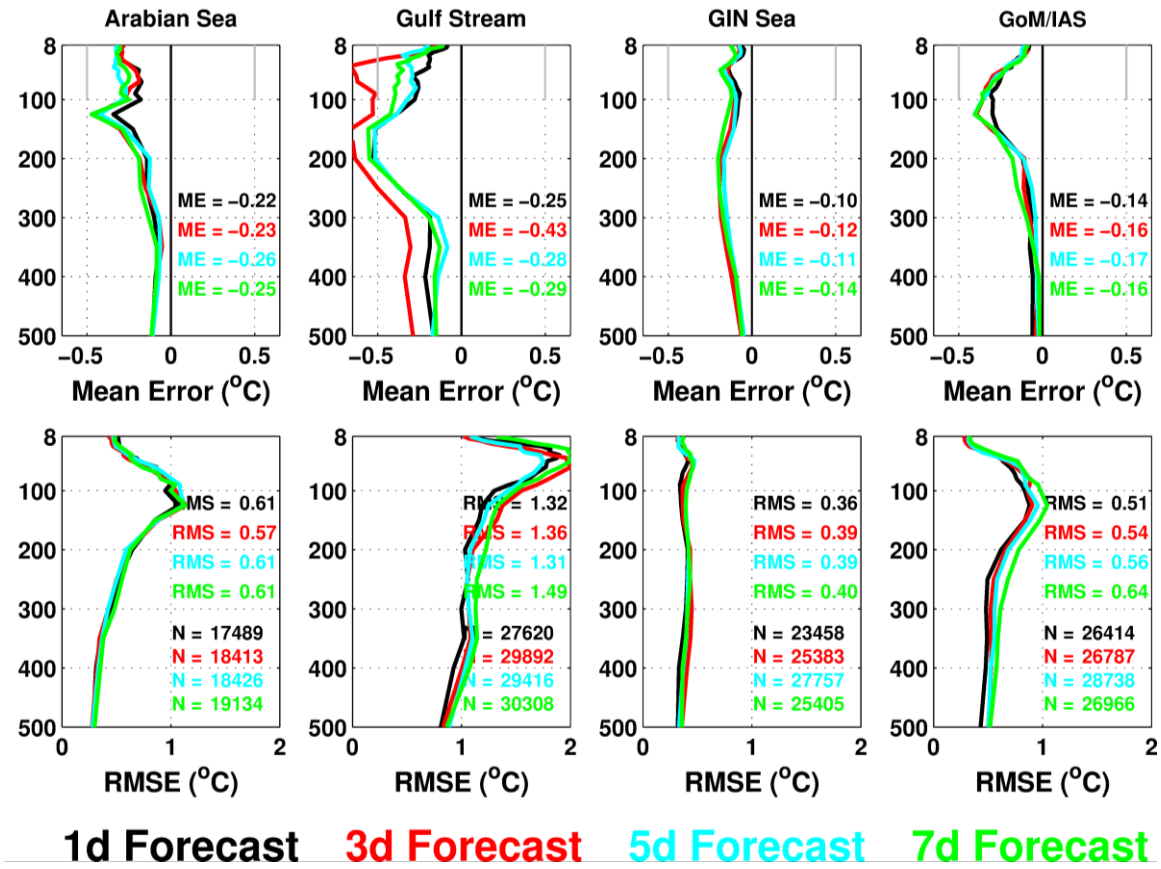

Figure 9: As Figure 8 except for the Arabian Sea, Gulf Stream, GIN Sea and Gulf of Mexico/Intra-Americas Sea regions from left to right, respectively.

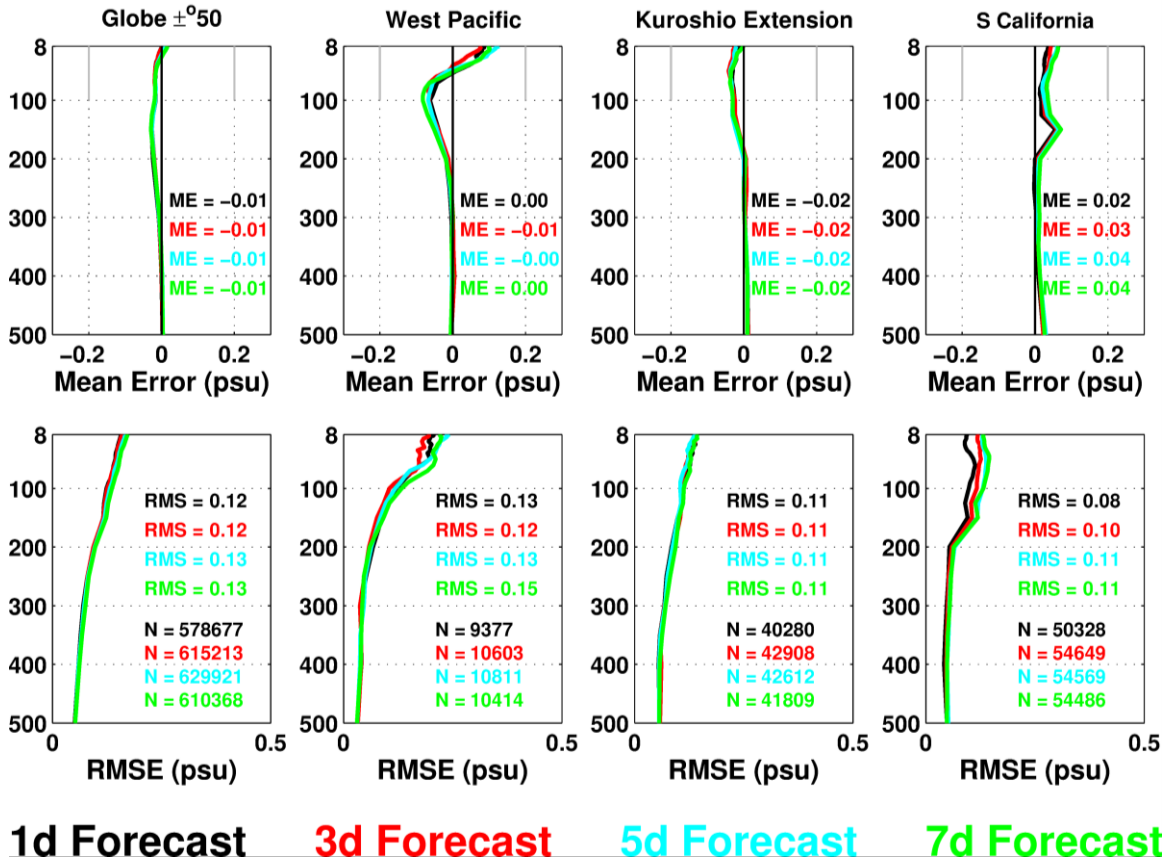

Figure 10: As Figure 8 except for salinity (psu).

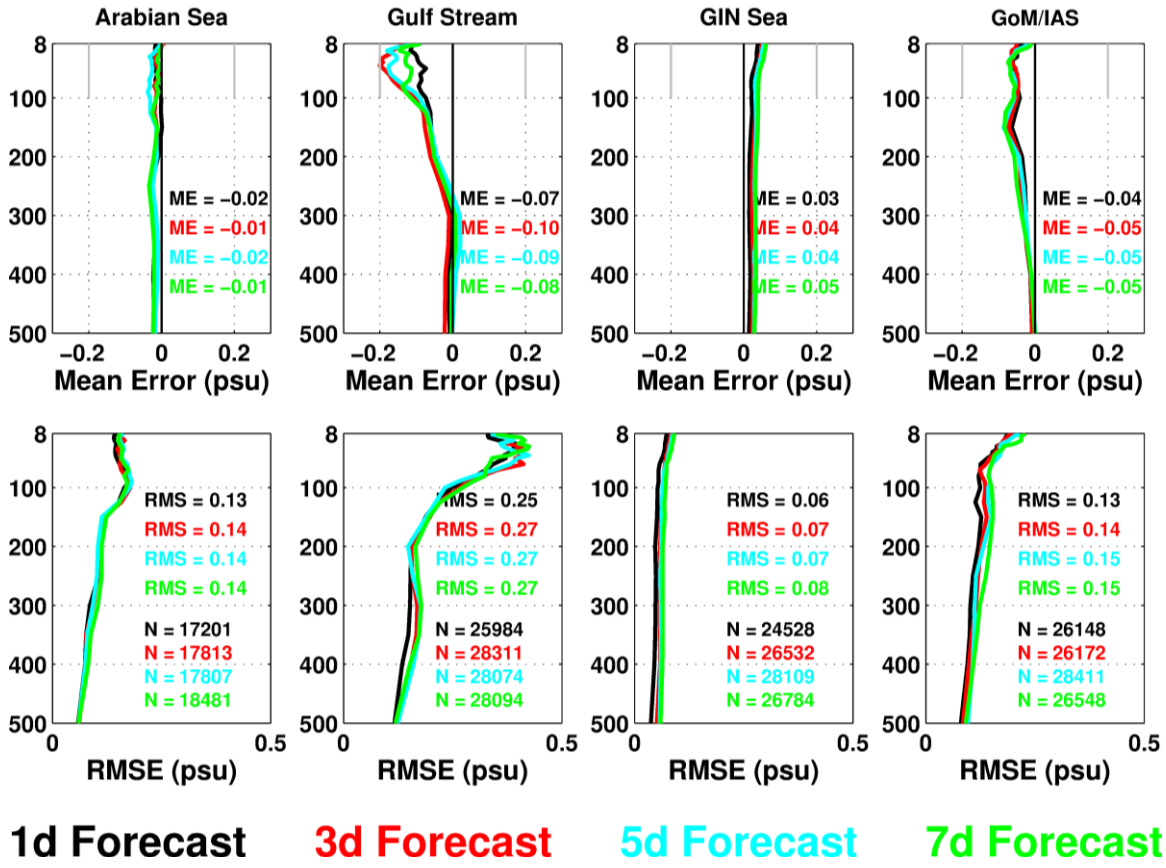

Figure 11: As Figure 9 except for salinity (psu).

3.3 Isotherm depth error analyses

A new metric compared to previous GOFS VTRs is the error in the depth of selected isotherms (26°C, 20°C and 15°C). Typical depths for these isotherms are shown in Figure 12. 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°. Lastly, the 15°C isotherm is generally in the lower thermocline and 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 around 700 m.

The bias, RMSE and mean absolute error (MAE) for GOFS 3.1, GOFS 3.5 and the GDEM4 climatology (Carnes et al., 2010) relative to the observations for the isotherm depths at the nowcast time are shown in Figure 13 and as a function of forecast length in Figure 14. GOFS 3.1 and GOFS 3.5 generally show similar performance and outperform climatology except for bias of the 15°C

isotherm. When examining the error vs. forecast length, GOFS 3.1 typically has higher RMSE and MAE for most days of the forecast. This metric will be quantified in the ocean score card section.

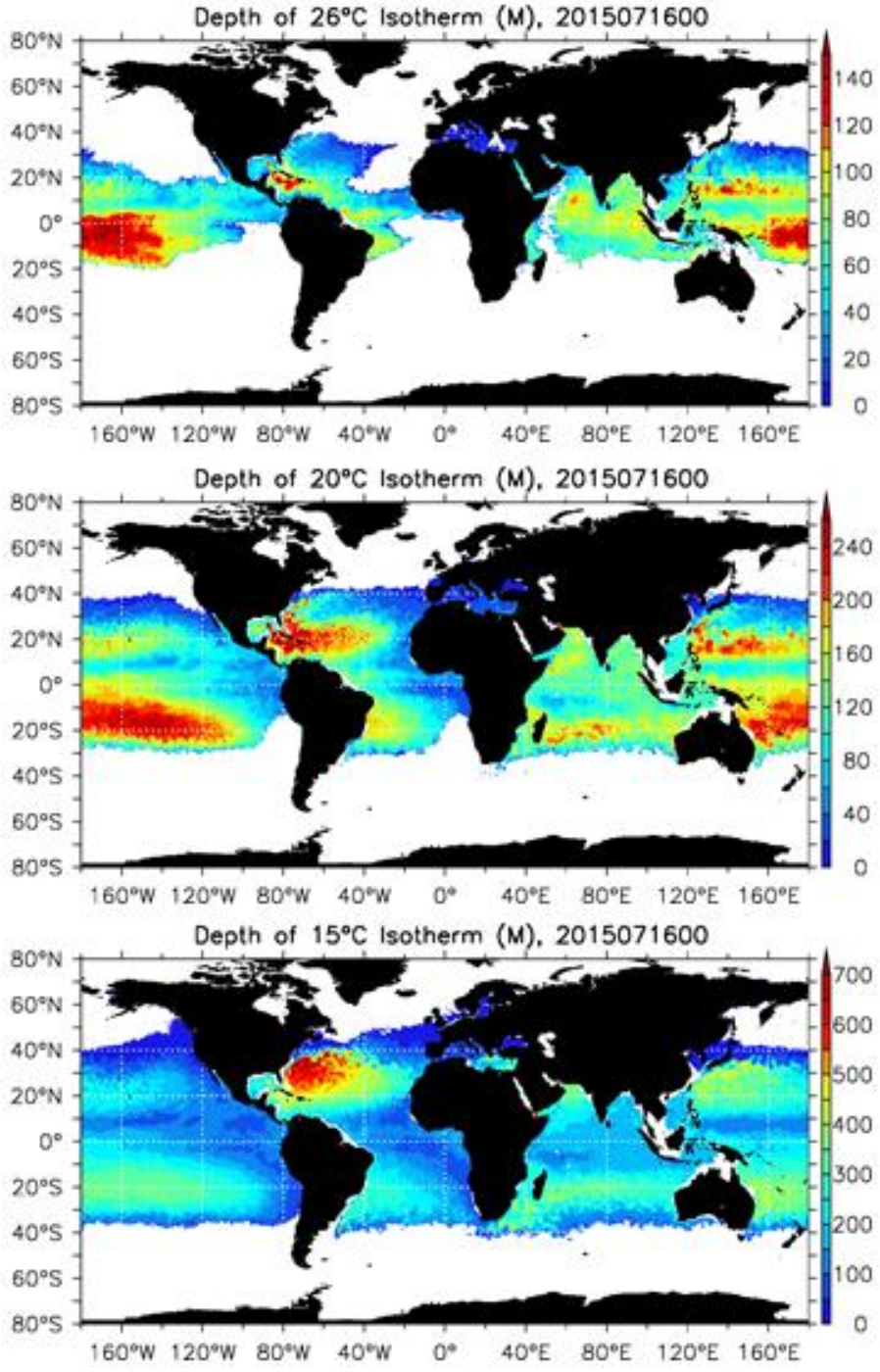

Figure 12: Maps of typical depths (m) of the 26°C (top), 20°C (middle) and 15°C (bottom) isotherms. This example is derived from GOFS 3.1 on 16 July 2015 00Z.

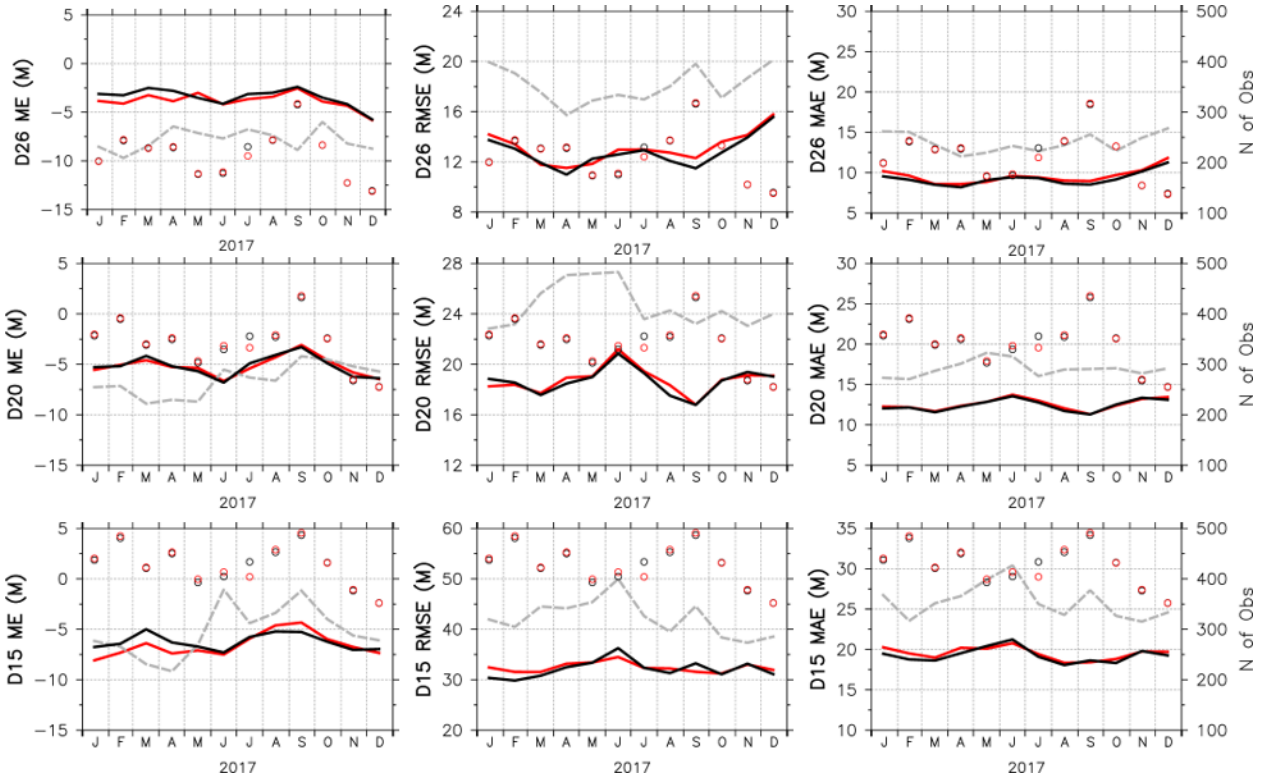

Figure 13: Monthly mean error (left column), RMSE (middle column) and mean absolute error (right column) of isotherm depth (m) error relative to the observations at the nowcast time for GOFS 3.5 (black), GOFS 3.1 (red) and the GDEM4 climatology (gray dashed). The top/middle/bottom rows are for the 26°C/20°C/15°C isotherms, respectively. The average number of observations/day are shown on the right axis and represented by open circles in the respective colors.

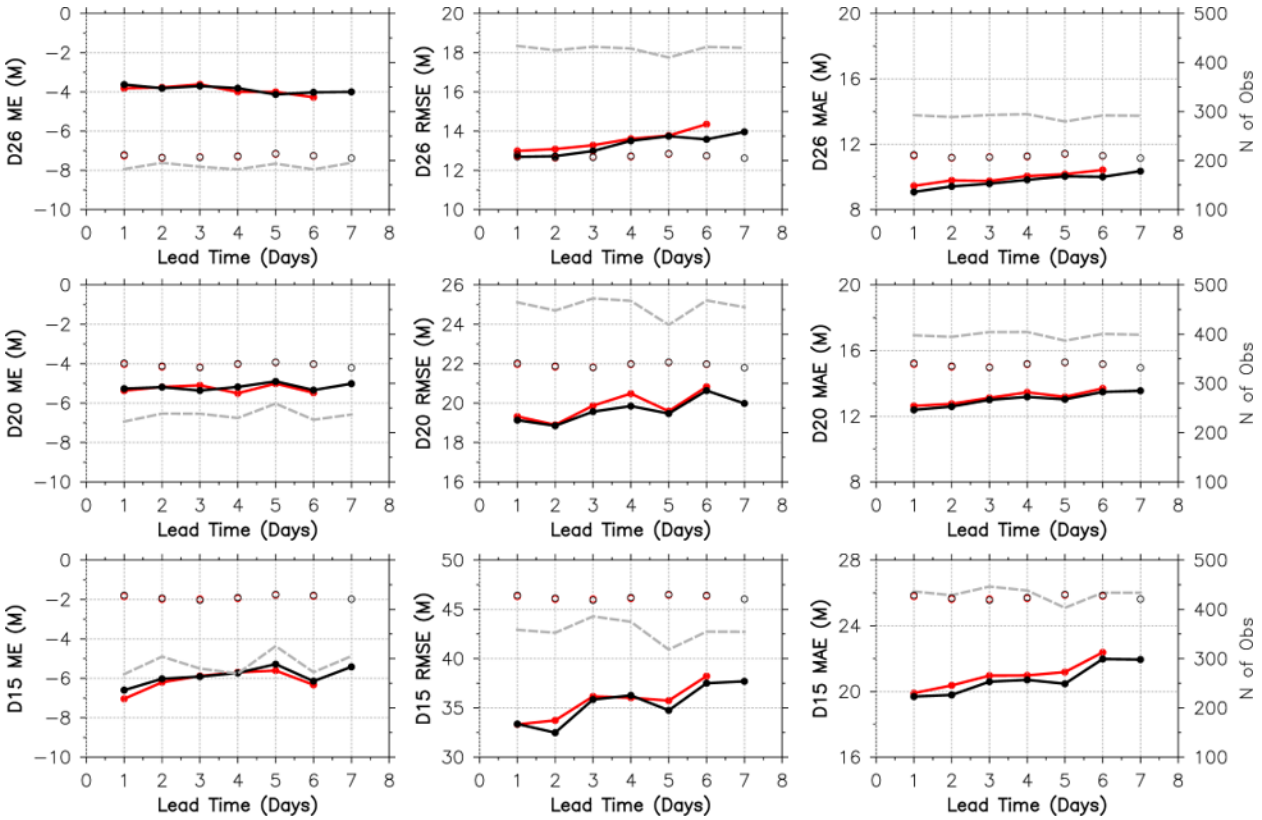

Figure 14: As in Figure 13 but as a function of forecast length.

3.4 Acoustical proxy error analyses

Accurate knowledge of the underwater acoustical environment can lead to significant tactical advantages during naval operations. The 3D T and S structure and the upper ocean mixed layer largely determine the sound speed profile that characterizes the acoustical ducts within the water column. Thus, an ocean nowcast/forecast system must be able to accurately simulate mixed layer depth (MLD), sonic layer depth (SLD), and below layer gradient (BLG). Here MLD is defined by a positive density difference of 0.15 kg/m^3 between the surface and a given depth if both T and S profiles are available, or a negative temperature difference of 0.5°C between the surface and a given depth if only T profiles are available. SLD is the vertical distance from the surface to the depth of the sound speed maximum, often but not always at the base of the mixed layer. Lastly, BLG is defined as the sound speed rate of change with depth per 100 feet in the first 300 feet below the SLD or below the surface if the SLD is absent. These quantities are derived from Naval Oceanographic Office Reference Publication 33 (RP33, 1992), with the exception that

the sound speed equation by Chen and Millero (1977) and later correction by Millero and Li (1994) is used rather than that by Wilson (1960).

Histograms of ME and RMSE for MLD, SLD and BLG are shown in Figure 15 - Figure 17, respectively for the nowcast time. The first two metrics examine all occurrences but are further broken down into groups when the observed values were shallow (≤ 50 m), mid-range (50 m ≤ X ≤ 250 m) or deep (≥ 250 m). For MLD and SLD, bias (RMSE) is typically 10-20% (20-40%) for the shallower two depth ranges, but significantly larger for the deepest range. GOFS 3.1 generally has lower error than GOFS 3.5 for MLD, but GOFS 3.5 outperforms GOFS 3.1 for SLD. BLG performance at the nowcast time is similar for both systems.

Figure 18 shows acoustical proxy error as a function of forecast length for GOFS 3.5. As expected, error generally grows slowly with increasing forecast length. However, for reasons not fully understood, the error for the 1-day forecasts in a few regions is higher than that for the longer forecasts. Similar to the other metrics, these will be quantified in the ocean score card section.

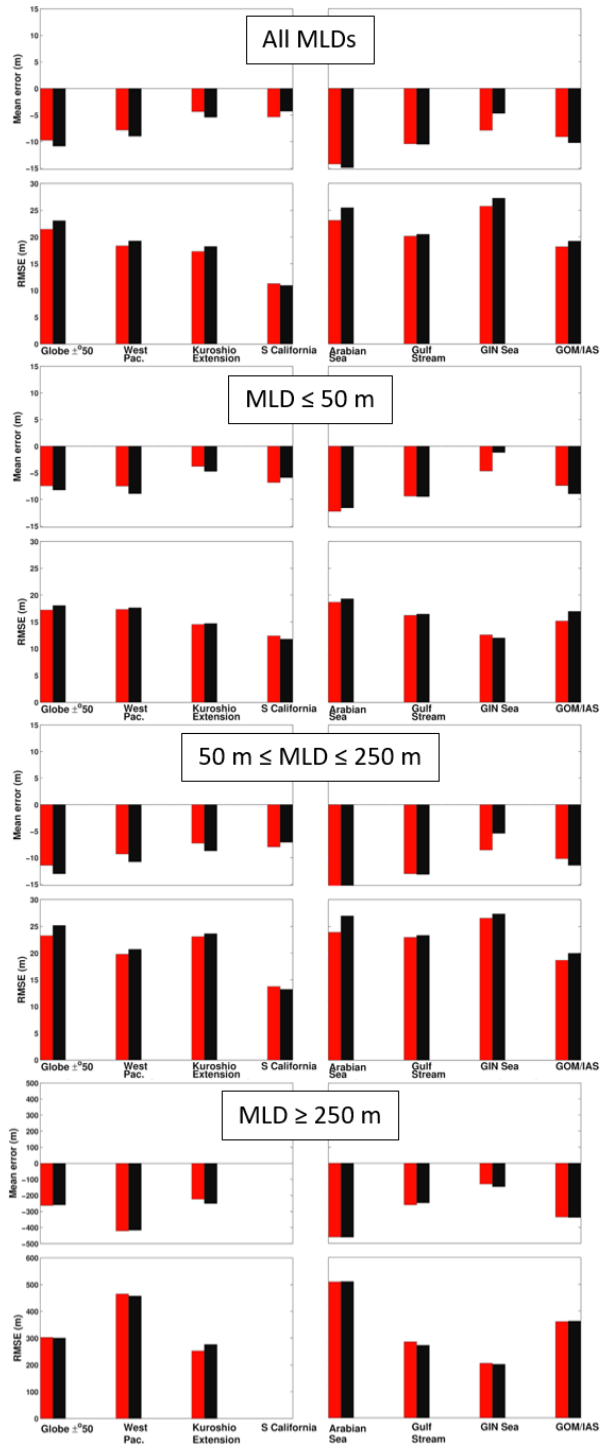

Figure 15: Histograms of mixed layer depth mean error (1st, 3rd, 5th and 7th rows) and root mean square error (2ns, 4th, 6th and 8th rows) against unassimilated observations for GOFS 3.1 (red) and GOFS 3.5 (black) at the nowcast time for the eight analysis regions (columns) spanning calendar year 2017. Units are meters. The top two rows are statistics for all MLD ranges, rows three and four are for MLD ≤ 50 m, rows five and six are for 50 m ≤ MLD ≤ 250 m, and rows seven and eight are for MLD ≥ 250 m.

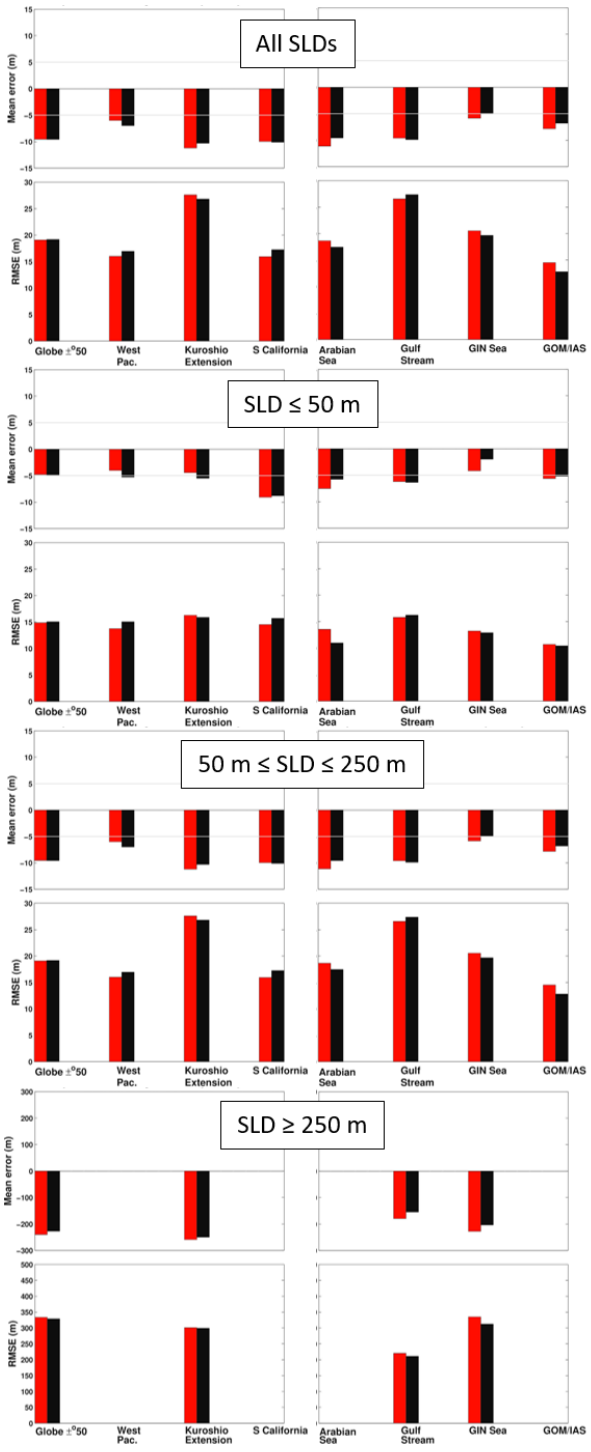

Figure 16: As Figure 15 except for sonic layer depth.

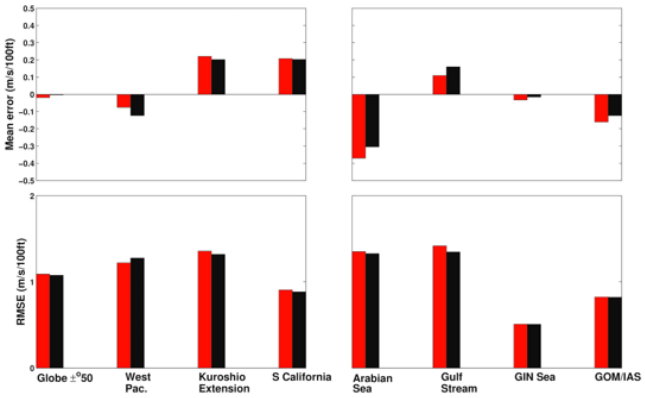

Figure 17: Histograms of below layer gradient mean error (top row) and root mean square error (bottom row) against unassimilated observations for GOFS 3.1 (red) and GOFS 3.5 (black) at the nowcast time for the eight analysis regions (columns) spanning calendar year 2017. Units are m/s/100 ft.

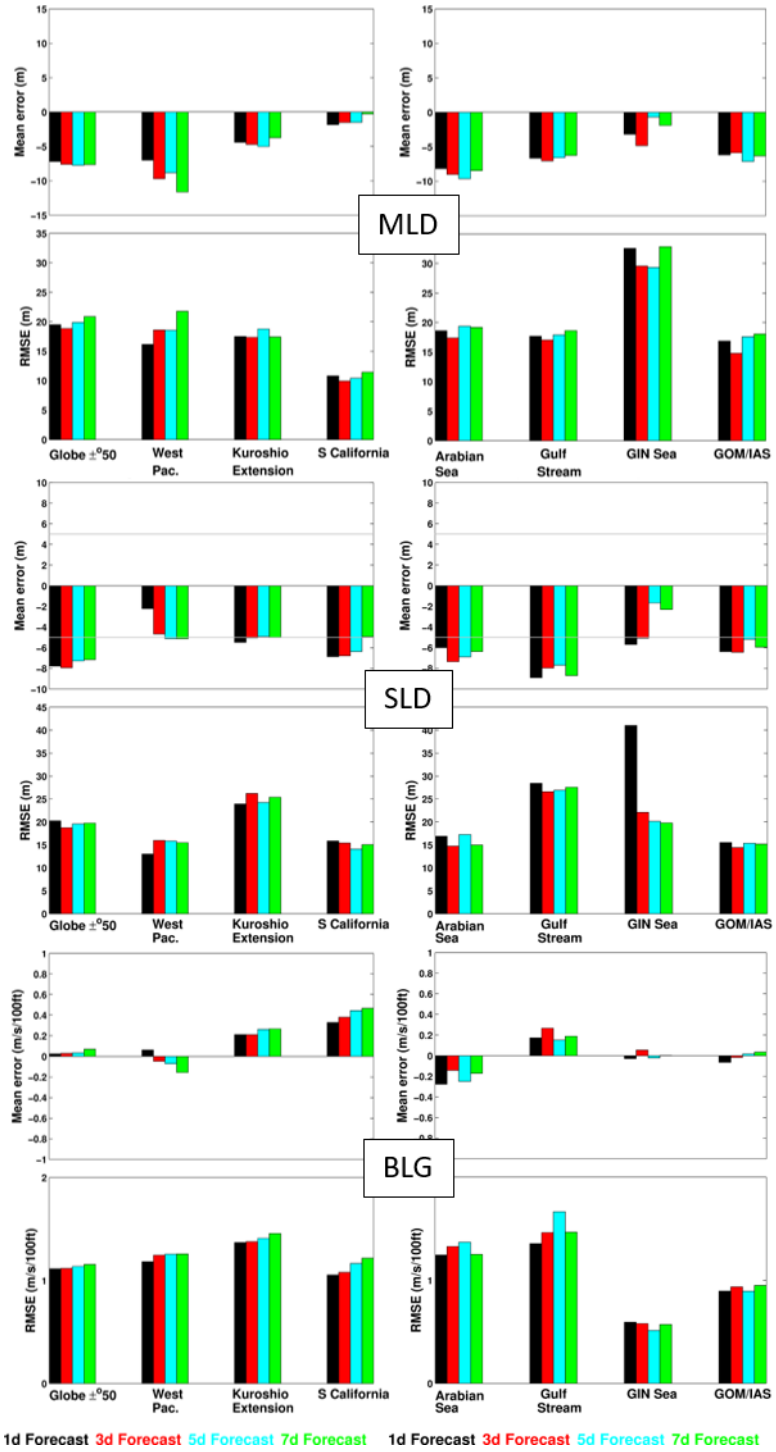

Figure 18: Mixed layer depth (m), sonic layer depth (m) and below layer gradient (m/s/100 ft) mean error (1st, 3rd and 5th rows) and RMSE (2nd, 4th, 6th rows) against unassimilated observations for GOFS 3.5 for the eight analysis regions (columns) spanning calendar year 2017. Histograms are colored black for the 1-day forecasts, red for the 3-day forecasts, cyan for the 5-day forecasts and green for the 7-day forecasts. For MLD and SLD, errors are computed across all depth ranges.

3.5 Surface layer trapping of acoustic frequencies

Under certain conditions, the surface layer may trap acoustic energy in a surface duct in which sound rays of sufficiently high frequencies refract back toward and bounce off the surface repeatedly. The result is that sound beams spread cylindrically in the horizontal direction with intensity decreasing as the inverse of the range (Helber et al., 2010). The cutoff frequency (COF) is the critical frequency above which sound is trapped in a surface duct or below which sound is not trapped. SLD determines the frequency of sound that can be trapped, with shallower SLD leading to higher COF. SLD and COF are computed from unassimilated T and S profiles and compared against both forecast systems.

Consistent with the NCODA-3DVAR VTR (Smith et al., 2011), "stoplight" curves for acoustical trapping are produced in Figure 19 (for the same profile set used in the T/S vs. depth error analyses). Green (black) signifies true-positive (true-negative) predictions of acoustic trapping, i.e. both the data and GOFS indicate (do not indicate) trapping. Red (yellow) is used for false-positives (false-negatives) where the observations (models) show no acoustical trapping but the model (observations) shows trapping. The y-axis is the percent of occurrence for each of the four curves and thus adds to 100% for a given frequency. Good predictions are when the green/black curves occur more frequently than the yellow/red curves. The better performing system will have higher true positives/negatives and lower false positives/negatives and a smaller ratio of false to true predictions.

In the GIN Sea region, the black and green curves cross each other at low frequencies (deep sonic layers) as both the observations and systems indicate surface trapping across most of the frequency range. The opposite is true for the SoCal subdomain where shallow sonic layers dictate a slow rise in the true positive curves. For most of the analysis regions, the true positives/negatives occur more often than the false positives/negatives across most of the frequency range. A comparison of the two systems is summarized in Table 1 that lists the average occurrence of acoustical trapping across the entire frequency range, i.e. the average of the individual curves in Figure 19. Performance is similar in most regions, but GOFS 3.1 has slightly smaller errors than GOFS 3.5.

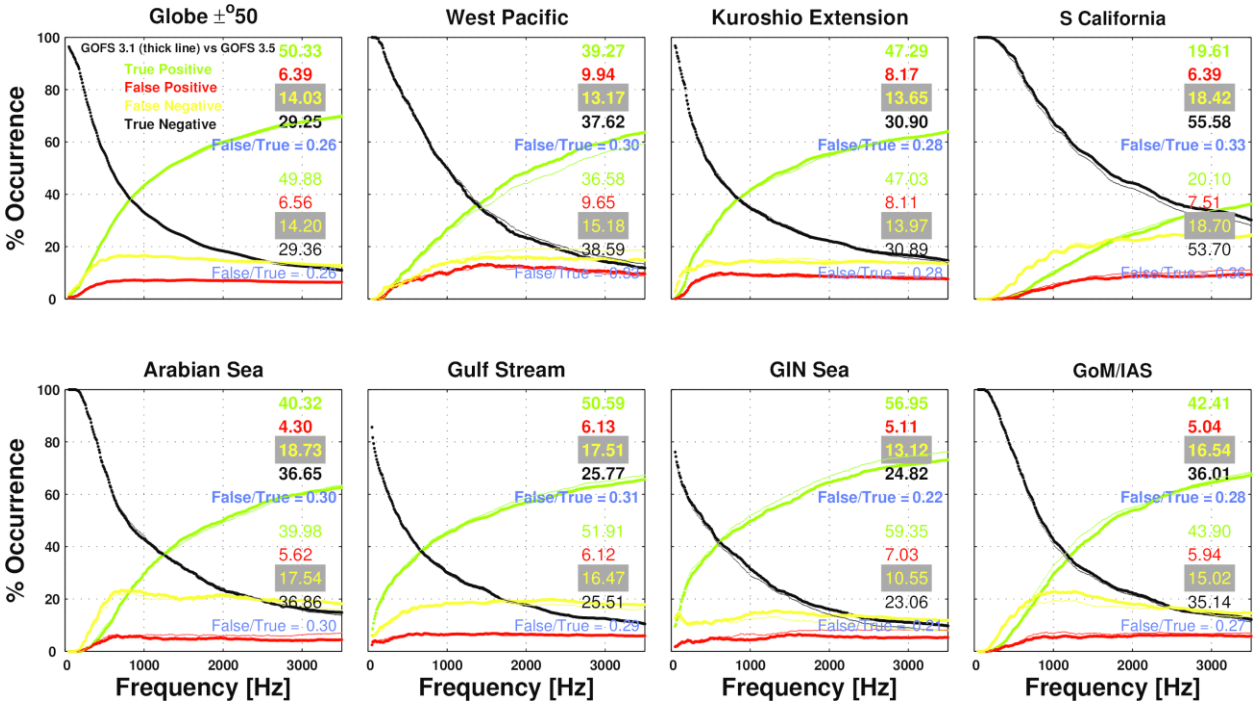

Figure 19: Percent occurrence of acoustical trapping as a function of frequency (Hz) using unassimilated profile observations compared against GOFS 3.5 (thin lines) and GOFS 3.1 (thick lines) for the eight analysis regions. Green (black) signifies true-positive (true-negative) predictions of acoustic trapping, i.e. both the data and GOFS indicate (do not indicate) trapping. Red (yellow) is used for false-positives (false negatives) where the observations (model) show acoustical trapping but the model (observations) does not show trapping. This comparison is made at the nowcast time and spans calendar year 2017.

Table 1: Percent occurrence of acoustical trapping averaged across all frequencies between 0-3500 Hz for the eight analysis regions for GOFS 3.1 and 3.5 broken down by true positives (TP), true negatives (TN), false positives (FP), false negatives (FN) and the ratio of false/true (F/T). The better performing system will have higher true positives/negatives and lower false positives/negatives. The TP, TN, FP, FN boxes are shaded green if the percentage difference exceeds 1%. The smaller of the F/T ratio boxes are shaded green.

	Globe		Western Pacific		Kuroshio		So. Cal.		Arabian Sea		Gulf Stream		GIN Sea		GoM/ IAS	
	3.1	3.5	3.1	3.5	3.1	3.5	3.1	3.5	3.1	3.5	3.1	3.5	3.1	3.5	3.1	3.5
TP	50	50	39	37	47	47	20	20	40	40	51	52	57	59	42	44
TN	29	29	38	39	31	31	56	54	37	37	26	26	25	23	36	35
FP	6	7	10	10	8	8	6	8	4	6	6	6	5	7	5	6
FN	14	14	13	15	14	14	19	19	19	18	18	16	13	11	17	15
F/T	.26	.26	.30	.33	.28	.28	.33	.36	.30	.30	.31	.29	.22	.21	.28	.27

3.6 Upper ocean velocity validation

Upper ocean velocity from GOFS 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 for this velocity error analysis. The drifters have a drogue at 15 m and instantaneous GOFS velocity output are extracted at this depth. The error analyses are ME and RMSE of 15 m speed, and vector correlation (Crosby et al., 1993) to provide directional error.

Figure 20 shows drifter speed bias and RMSE as well as directional vector correlation for both systems. Speed biases are generally positive but small and less than 10% of the observed mean speed with the exception of the western boundary current regions (Kuroshio Extension and Gulf Stream). Speed RMSE is similar for both systems with relatively slow error growth (11% for the global domain) from the nowcast time through the 7-day forecast. The growth of directional error over the forecasts is larger with the vector correlation decreasing by 32% for the global domain, but performance is similar between both systems.

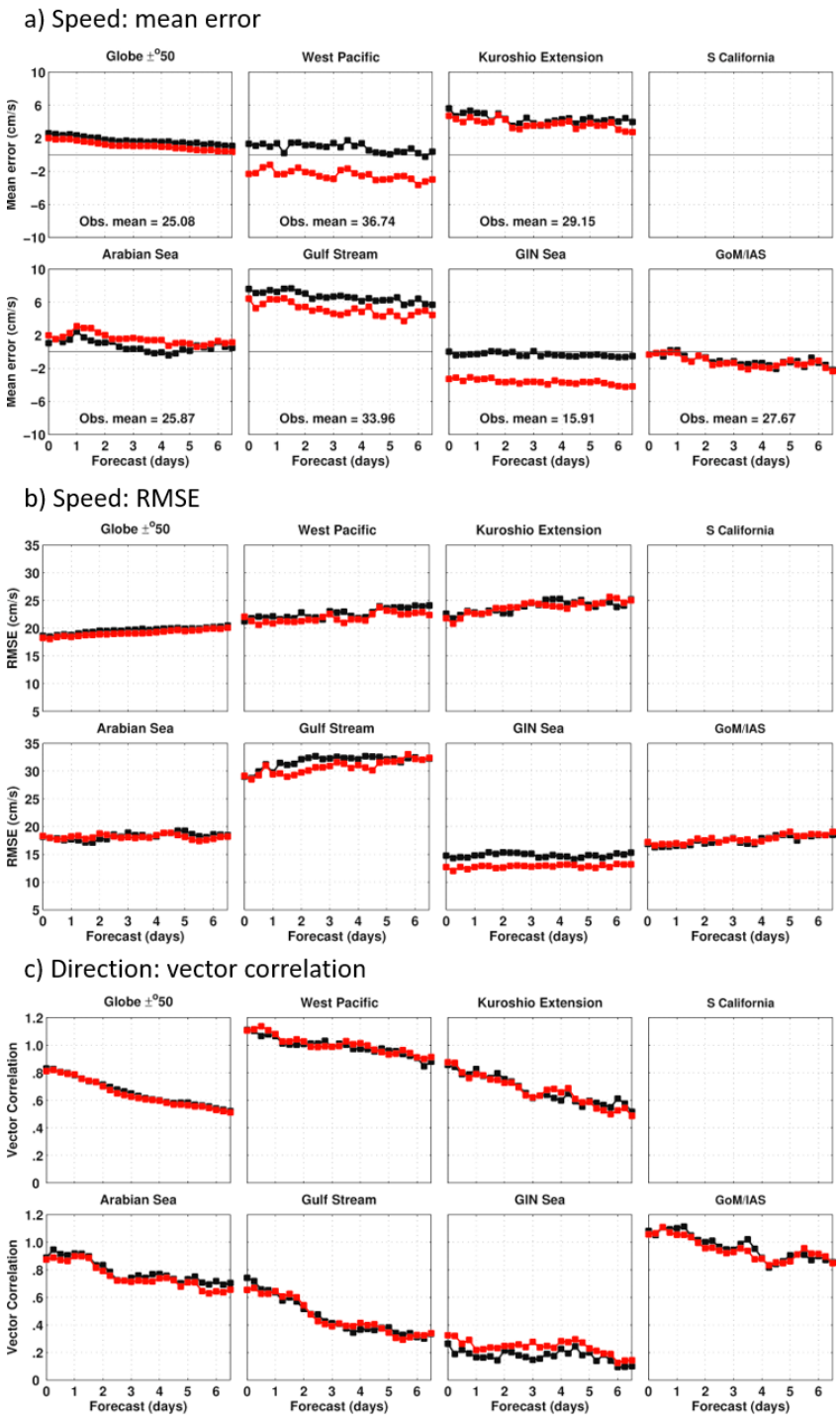

Figure 20: Speed (a) mean error (cm/s), (b) RMSE (cm/s), and (c) direction vector correlation of unassimilated drifting buoys at 15 m depth compared against GOFS 3.5 (black) and GOFS 3.1 (red) as a function of forecast length for the eight analysis regions. The mean observed drifting buoy speed for each region is noted in the top eight subpanels. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. This analysis spans calendar year 2017. The Southern California region is blank because the sample size was too small for meaningful statistics.

3.7 Frontal placement validation

A new metric has been developed to identify fronts in sea surface height (SSH) from altimetry and models (Douglass and Mask, 2019). Fronts are identified as gradients in alongtrack altimetric height more than one standard deviation greater than the one-year average of gradients in that location. Modeled SSH is interpolated to the locations of altimetric measurements, and fronts are identified in the model using the same technique as in observations. This metric is calculated daily, using all altimetry available that day and comparing with a daily SSH snapshot. Two metrics are calculated: 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. On any given day around the globe, approximately 70-75% of the modeled fronts are found in the altimeter data, while approximately 50% of the altimetric fronts are found in the model.

In Figure 21 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 and GOFS 3.5. In most regions, GOFS 3.1, has significantly higher reliability scores which is consistent with the lower resolution system producing fewer weak fronts. This is consistent with Sandery and Sakov (2017) that showed a high-resolution submesoscale permitting system does not always match the skill of a lower resolution system in forecasting the mesoscale circulation. This is due to the inverse cascade of the kinetic energy spectrum that lowers predictability of the mesoscale. Representativeness scores are similar in the western boundary current regions, but GOFS 3.1 outperforms GOFS 3.5 in the other regions.

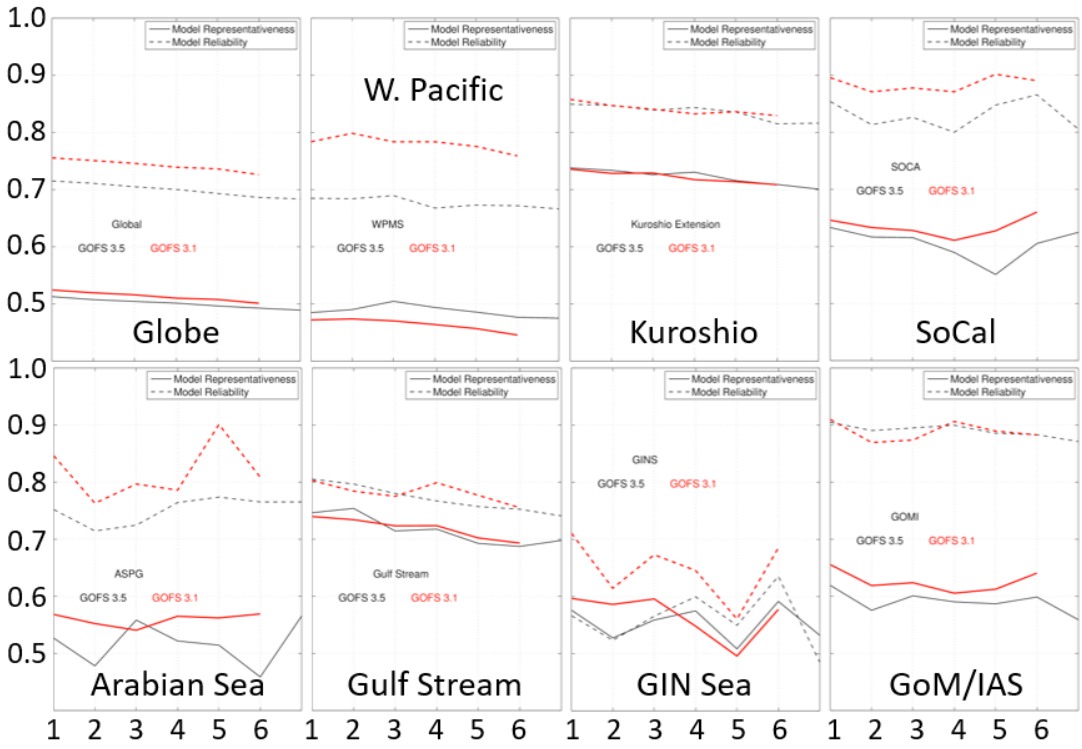

Figure 21: Frontal placement skill scores vs. forecast length (days) for the eight analysis regions. Red lines are GOFS 3.1 and black lines are GOFS 3.5. Solid lines indicate model representativeness while dashed lines indicate model reliability. GOFS 3.1 has a higher reliability score consistent with the lower resolution model producing fewer weak fronts.

3.8 Tidal elevation error analysis

GOFS 3.5 has a major change in its 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, 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). GOFS 3.5 now predicts 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 22. Visually the maps are very similar with the biggest differences in the structure of the North Atlantic and Western Pacific amphidromes.

In Figure 23, the errors between GOFS 3.5 and TPXO8 inverse tidal solutions are shown with the domain averaged values to the upper 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 panel) and errors associated with the amplitude-weighted phase (right panel). The RMSE is approximately split evenly with the amplitude only errors of 1.74 cm, and amplitude-weighted phase errors of 1.93 cm.

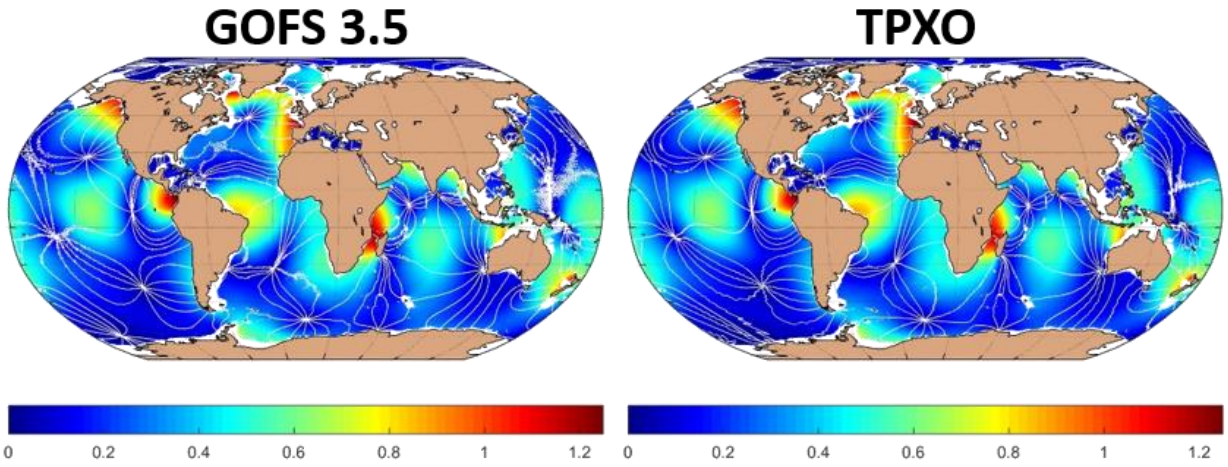

Figure 22: The amplitude (m) and phase of the M_2 barotropic tidal elevation for GOFS 3.5 (left) and an inverse tidal solution TPXO8 (right).

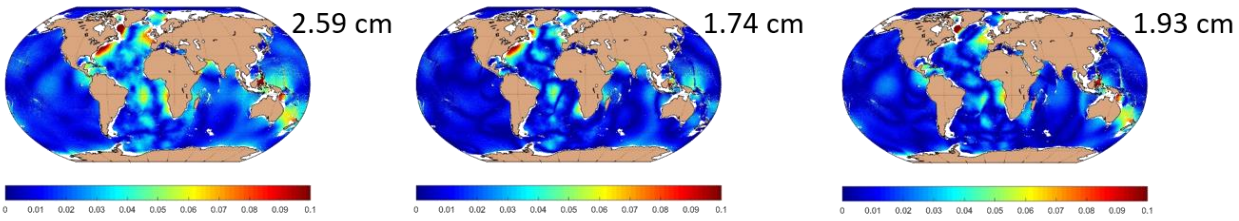

Figure 23: The difference between the GOFS 3.5 M_2 tidal elevation and the inverse tidal solution TPXO8. The left panel is the total RMSE, the middle panel is the RMSE associated with the amplitude only of the M_2 tidal coefficients and the right panel is the amplitude weighted phase error (see Shriver et al., 2012 for the definitions). The domain averaged value is note to the upper right of each panel.

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., 2012). 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 24 for GOFS 3.5 and altimeter data. In Table 2, 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. This is an expected bias based

on the results of Ansong et al. (2015) that found the areal averaged amplitudes decrease as a function of record length, asymptoting near a record length of about 4 years.

Table 2: Amplitudes (cm) of the M_2 internal tidal elevation from satellite altimetry SSH and GOFS 3.5. The model amplitudes are about 10% greater than the satellite estimates.

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

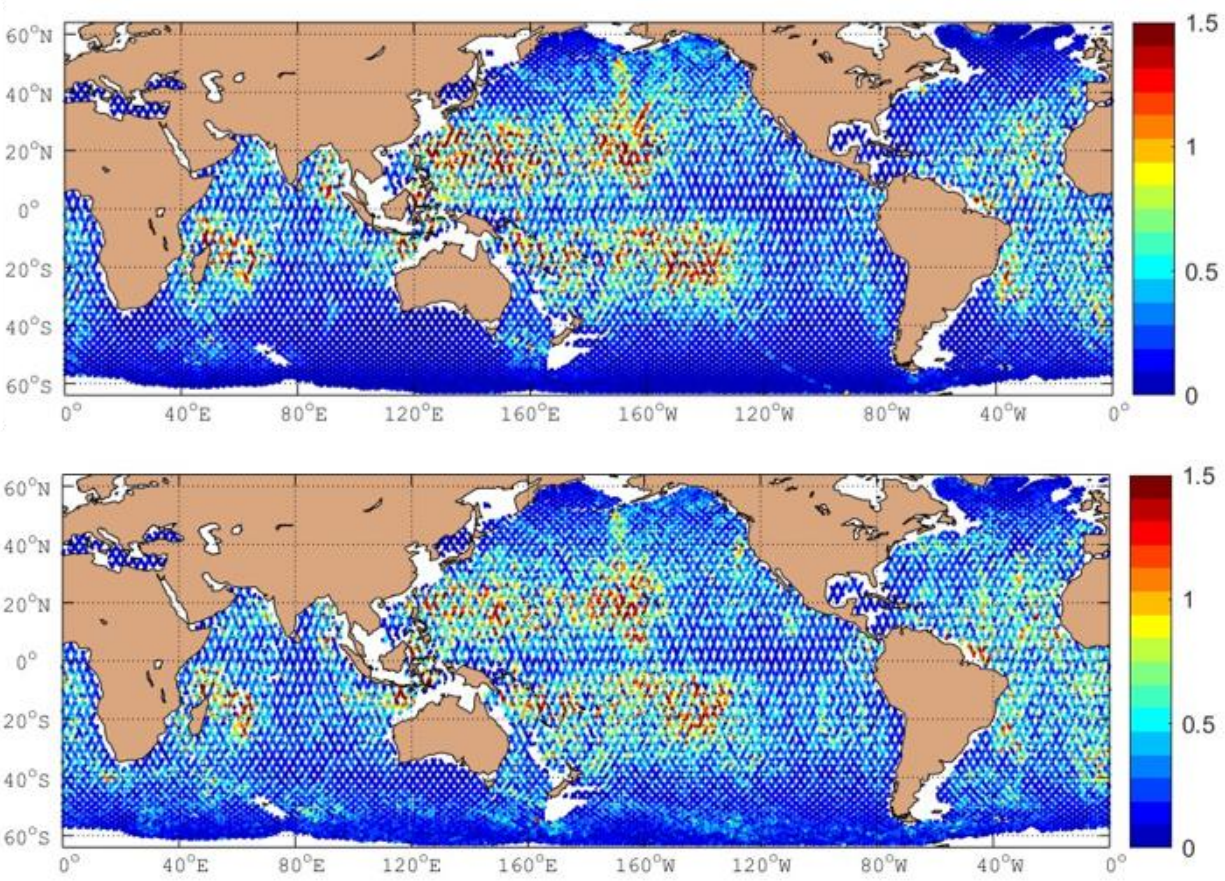

Figure 24: Amplitude of the stationary M_2 baroclinic tidal elevation (cm) for GOFS 3.5 (top) and satellite altimetry SSH estimate (bottom, Shriver et al., 2012). The GOFS 3.5 amplitudes are approximately 10% greater than the satellite estimate.

Similar to Figure 6 in Shriver et al. (2012), an example of the stationary part of the internal tide amplitude and phase from GOFS 3.5 is compared to a long reanalysis of altimeter SSH data (Figure 25). Even with the difference in the record length (2017 for GOFS 3.5 vs. 17 years for the altimeter data), the amplitudes of the observed altimetric and model internal tides agree well, 0.49 cm vs. 0.56 cm, respectively.

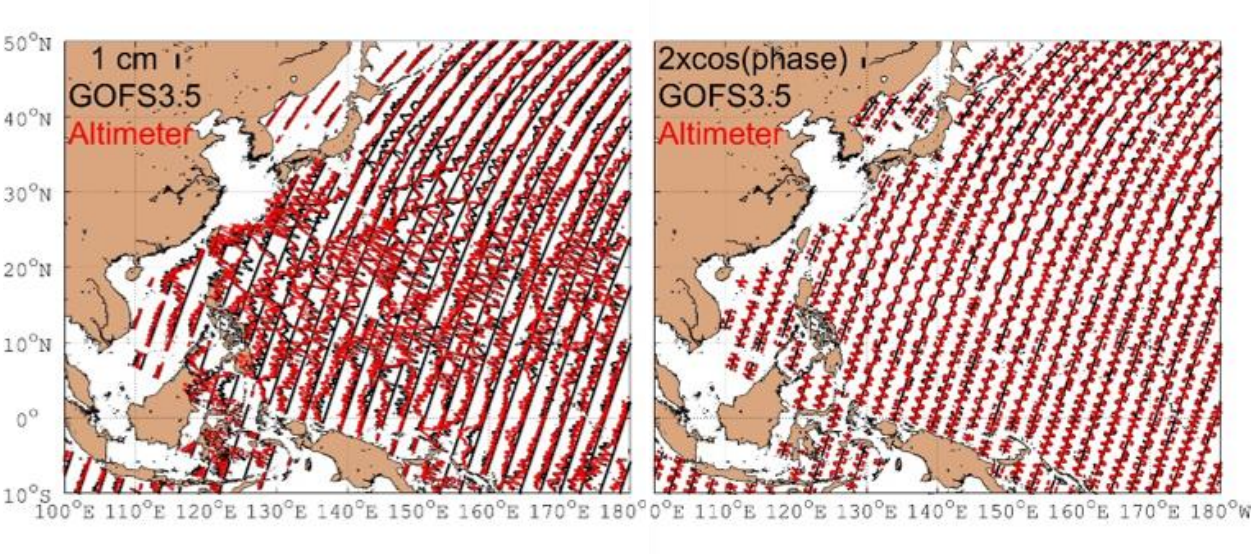

Figure 25: M_2 internal tide amplitude (cm, left) and phase (2x cos, right) along ascending tracks from GOFS 3.5 (black) and altimeter data (red) in the western Pacific Ocean. For amplitude, the line showing the coordinates of the track represents zero amplitude for the tides on that track.

3.9 Eddy kinetic energy climatology

The upper ocean (15 m) eddy kinetic energy (EKE) for GOFS 3.5 is compared to an independently estimated climatology of EKE from nearly 20,000 drifting buoys deployed over 40 years (Lumpkin and Pazos, 2007) (Figure 26). 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 with GOFS 3.5 29% higher and GOFS 3.1 30% higher. Some of the differences may be due to sampling. The drifting buoy climatology is based upon 40 years of observations with variable spacing, whereas the model estimates are averages for a single year (2017). 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 27) of observed vs. model EKE indicate the models have similar correlation. GOFS 3.1 and 3.5 are highly correlated with each other (0.96) (not shown).

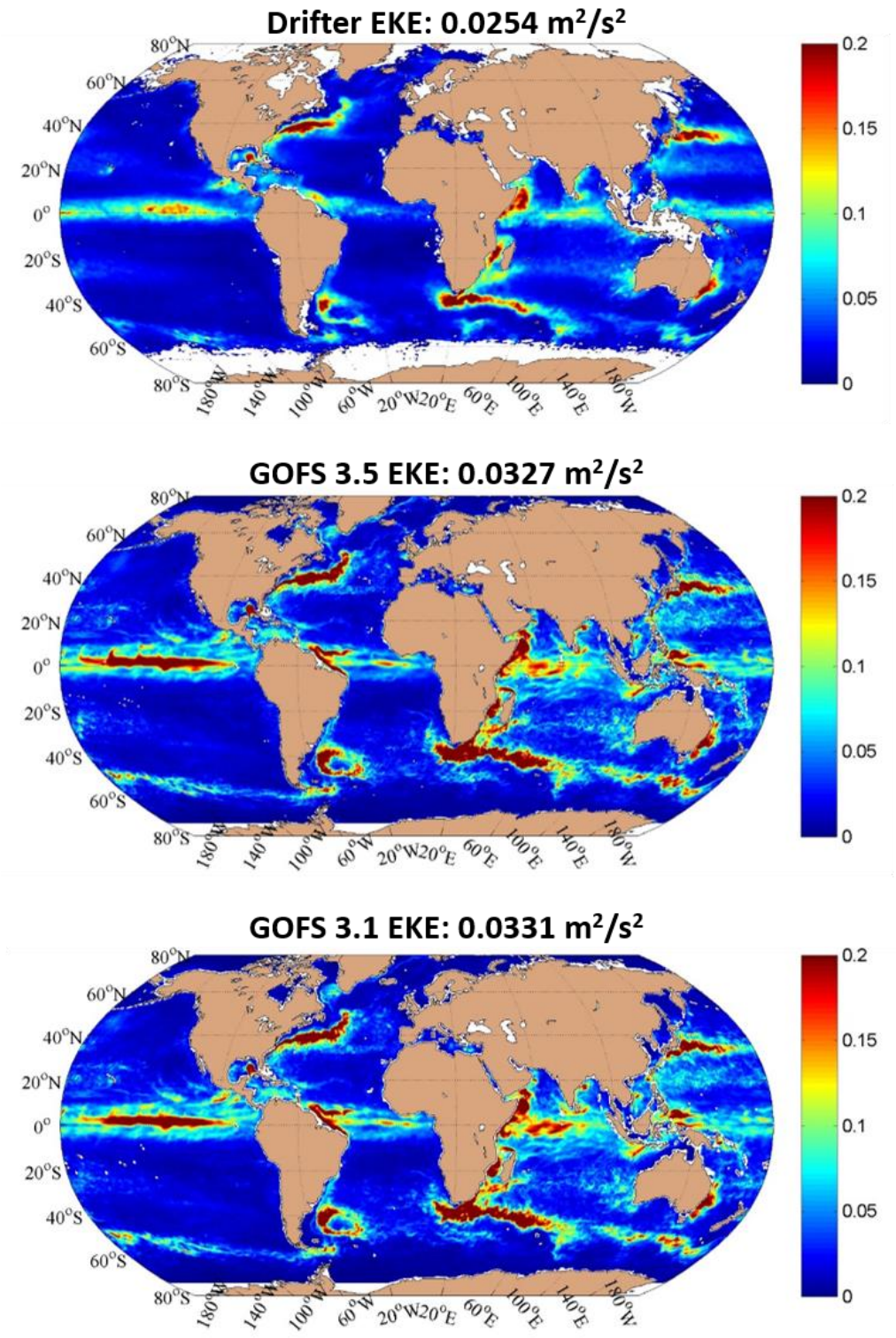

Figure 26: Mean Eddy Kinetic Energy (EKE, m^2/s^2) at 15 m derived from the drifting buoy archive (top), GOFS 3.5 (middle) and GOFS 3.1 (bottom). The basin-wide average EKE is listed above each panel.

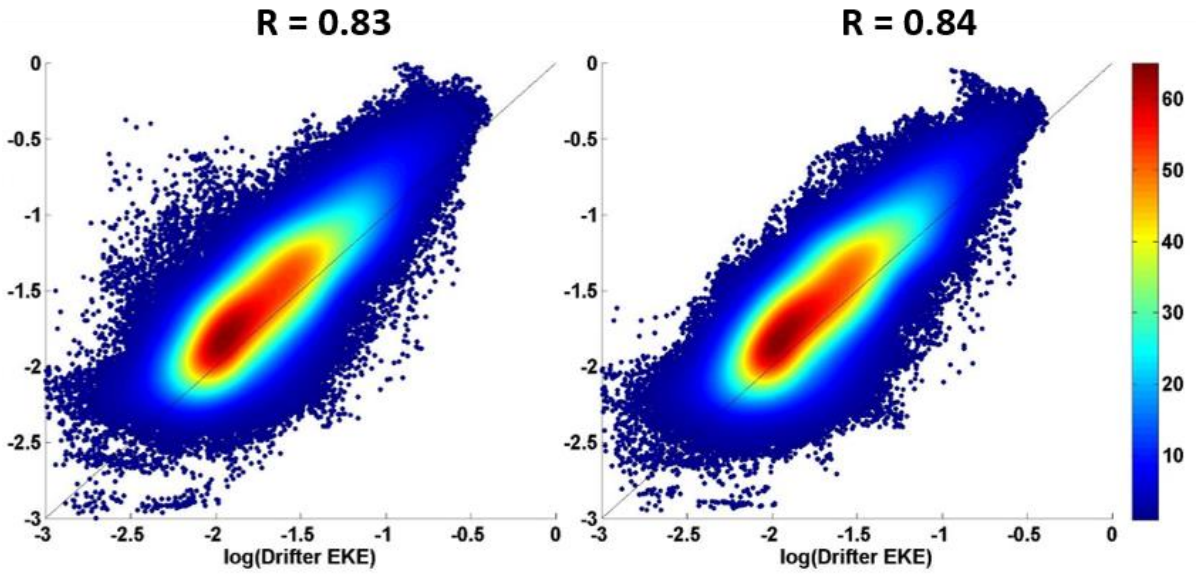

Figure 27: Scatterplots of 15 m EKE for GOFS 3.5 vs. the drifter climatology (left), and GOFS 3.1 vs. the drifter climatology (right). The correlation coefficient is listed above each plot. Both GOFS 3.1 and 3.5 have a positive bias relative to the drifter climatology.

3.10 Ocean score card

To quantify the performance of GOFS 3.1 and GOFS 3.5, 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 two prediction systems, and let

Maximum_error(i) = largest of *Error_model(i,t),*

Score_model(i,t) = 1 – (Error_model(i,t) / Maximum_error(i))

Total_score_model = mean *(Score_model(i,t))*

be the maximum error over time (*t*) in 2017 for each individual region, the scores of the models per each individual metric, and the total score per each model, respectively. 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, depths of the 26°C, 20°C and 15°C isotherms, and MLD, SLD, BLG. The isotherm error is

only used for the global domain, not the regional domains. For the cutoff frequency metric, the false/true ratio is used as *Error_model(i)*. The drifting buoy score uses RMSE of speed and additionally the inverse of vector correlation.

Figure 28 shows *Total_score_model* for the eight analysis regions as a function of forecast length. Here *Score_model(i,t)* is normalized by *Maximum_error(i)* across all forecast days which is typically largest for the day 6 forecast. GOFS 3.5 has higher overall score in four regions (Arabian Sea, Gulf Stream, Southern California and Gulf of Mexico/Intra-Americas Sea), GOFS 3.1 has higher score in three regions (Globe, Western Pacific and GIN Sea), and the scores are equal in the Kuroshio Extension region.

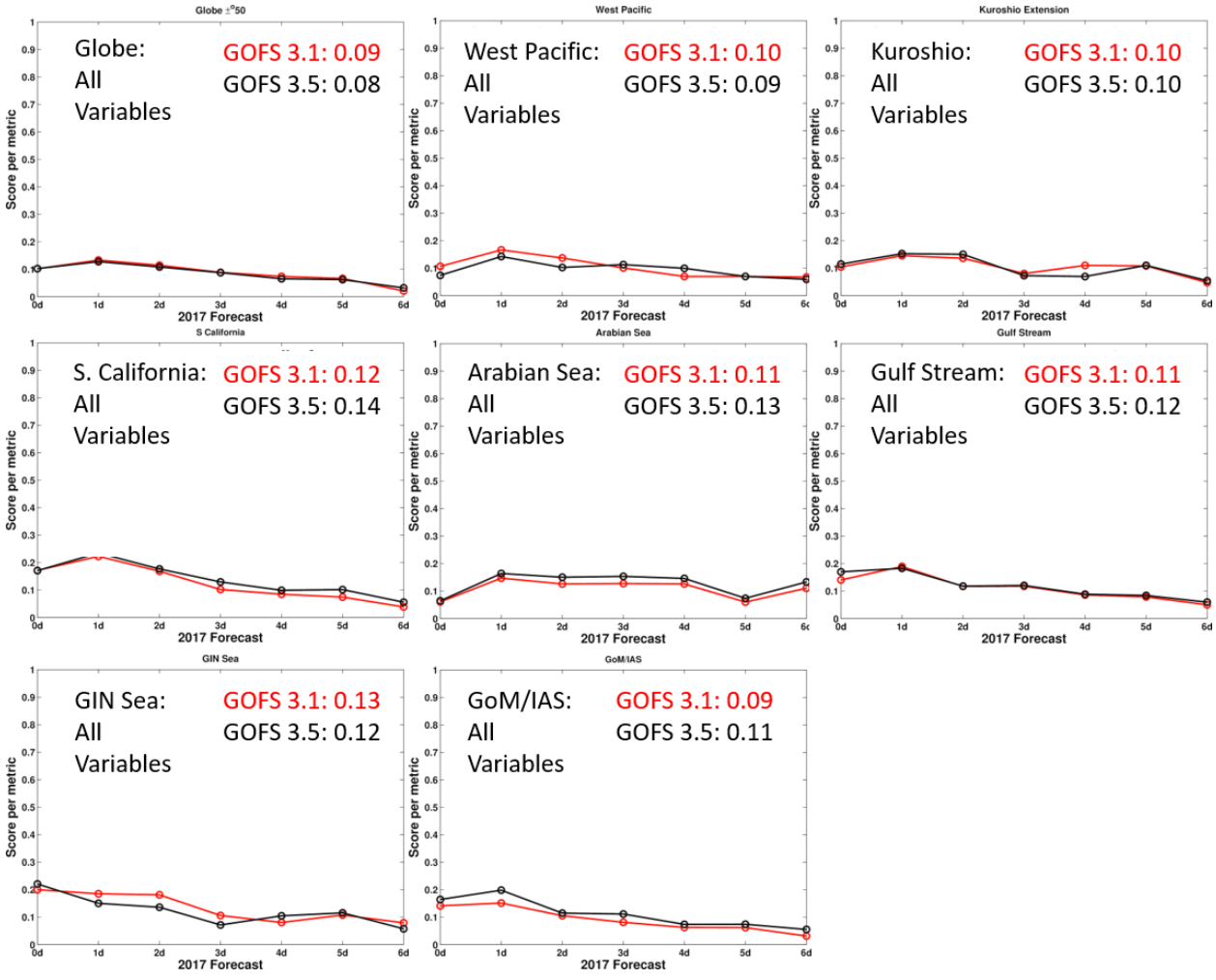

Figure 28: *Total_score_model* for the eight analysis regions as a function of forecast length (days) for GOFS 3.1 (red) and GOFS 3.5 (black). Because each region is normalized by the maximum error of that specific region these do not compare the relative skill of one analysis region with another, but within each region they do illustrate the relative performance of GOFS 3.1vs. GOFS 3.5.

Figure 29 through Figure 36 show *Total_score_model* and *Score_model(i,t)* of each metric for the eight analysis regions as a function of forecast length. Note again that *Score_model(i,t)* 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 two systems. Because each metric is normalized by *Maximum_error(i),* one curve always has a value of zero at one point along the x-axis. Because skill decreases slowly with increasing forecast length, the curves tend to be not that far above zero.

Note that for many regions the MLD/SLD/BLG/COF scores at the nowcast time are lower than the score for (say) the 1-day forecast (see Figure 29 as an example). This is due to the manner in which profiles were sampled. Recall in Figure 3a that in order to use unassimilated observations at the nowcast time, only profiles from tau 012 to 024 can be used. Hence the profiles are not centered over the nowcast time (as they are for the forecasts) and the sampling is skewed toward the future leading to lower acoustical parameter scores.

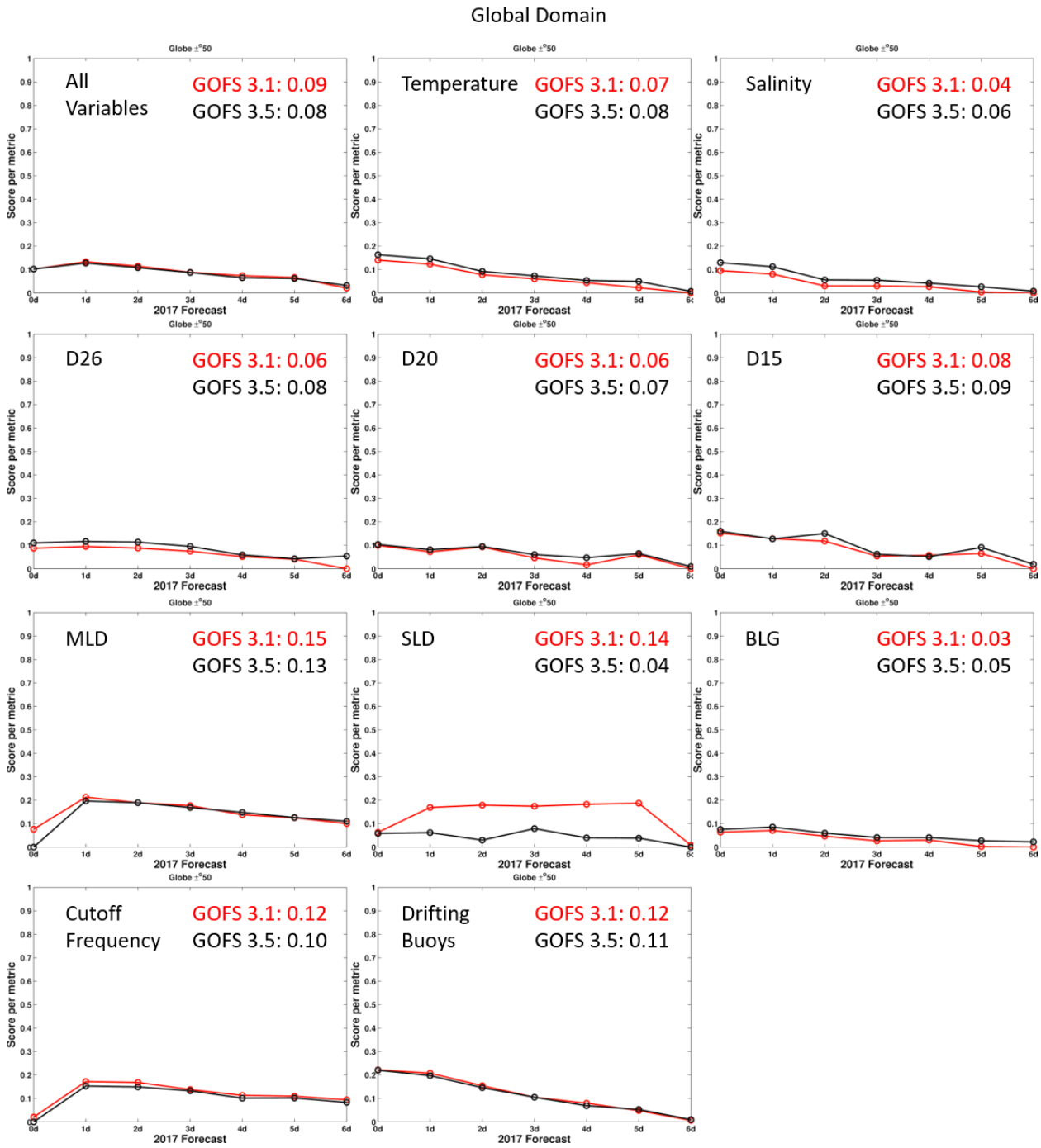

Figure 29: Scores of the global analysis region as a function of forecast length (days) for *Total_score_model* (labeled All Variables), and *Score_model(i,t)* in the other panels for each metric: Temperature = temperature vs. depth, Salinity = salinity vs. depth, D26 = depth of the 26°C isotherm, D20 = depth of the 20°C isotherm, D15 = depth of the 15°C isotherm, MLD = mixed layer depth, SLD = sonic layer depth, BLG = below layer gradient, Cutoff Frequency = cutoff frequency and Drifting buoys = ocean drifter speed and direction. GOFS 3.1 are red curves while GOFS 3.5 are black curves. Higher numbers are better.

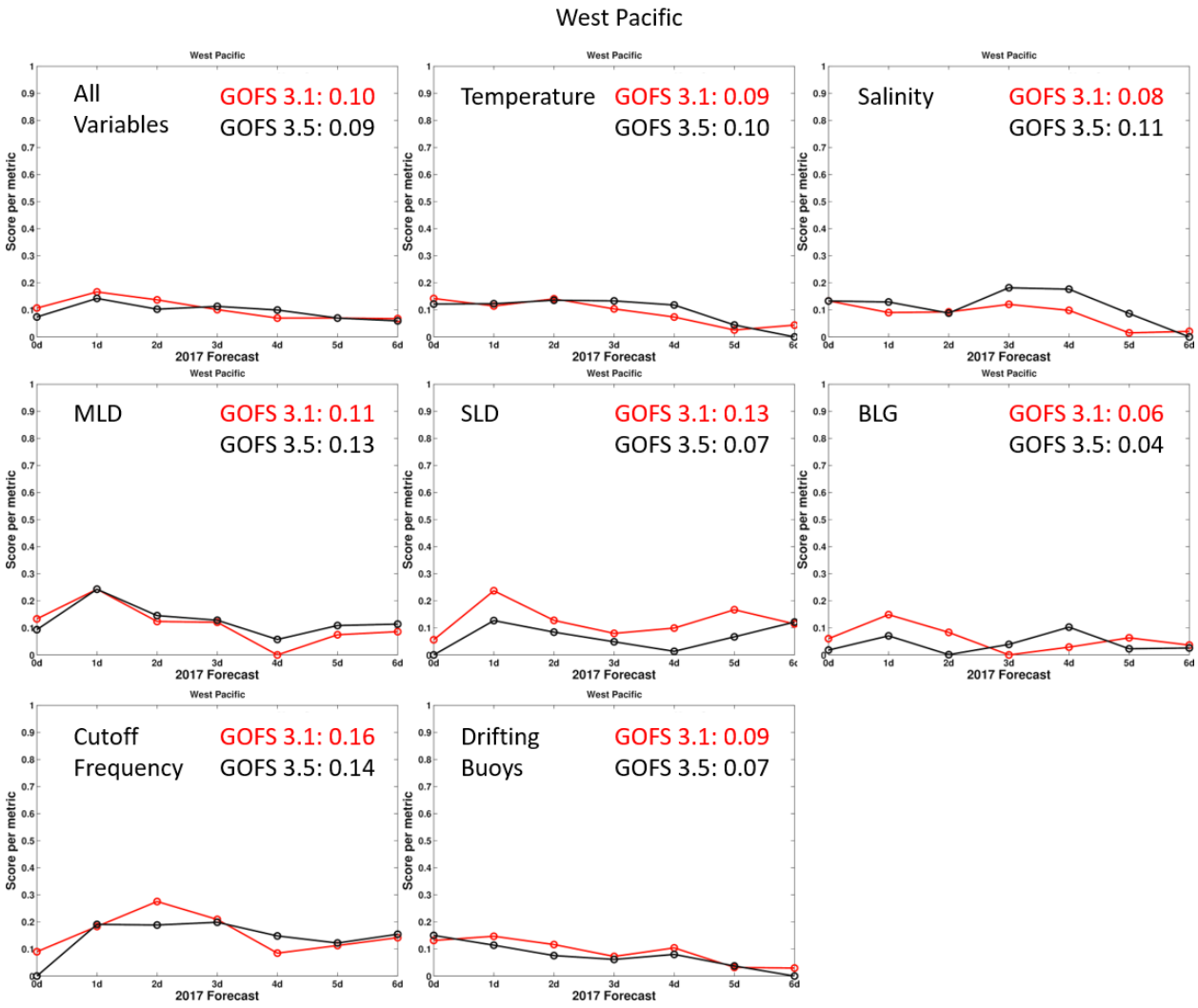

Figure 30: As Figure 29 except for the Western Pacific region, and without the D26, D20, D15 metrics that only apply to the global domain.

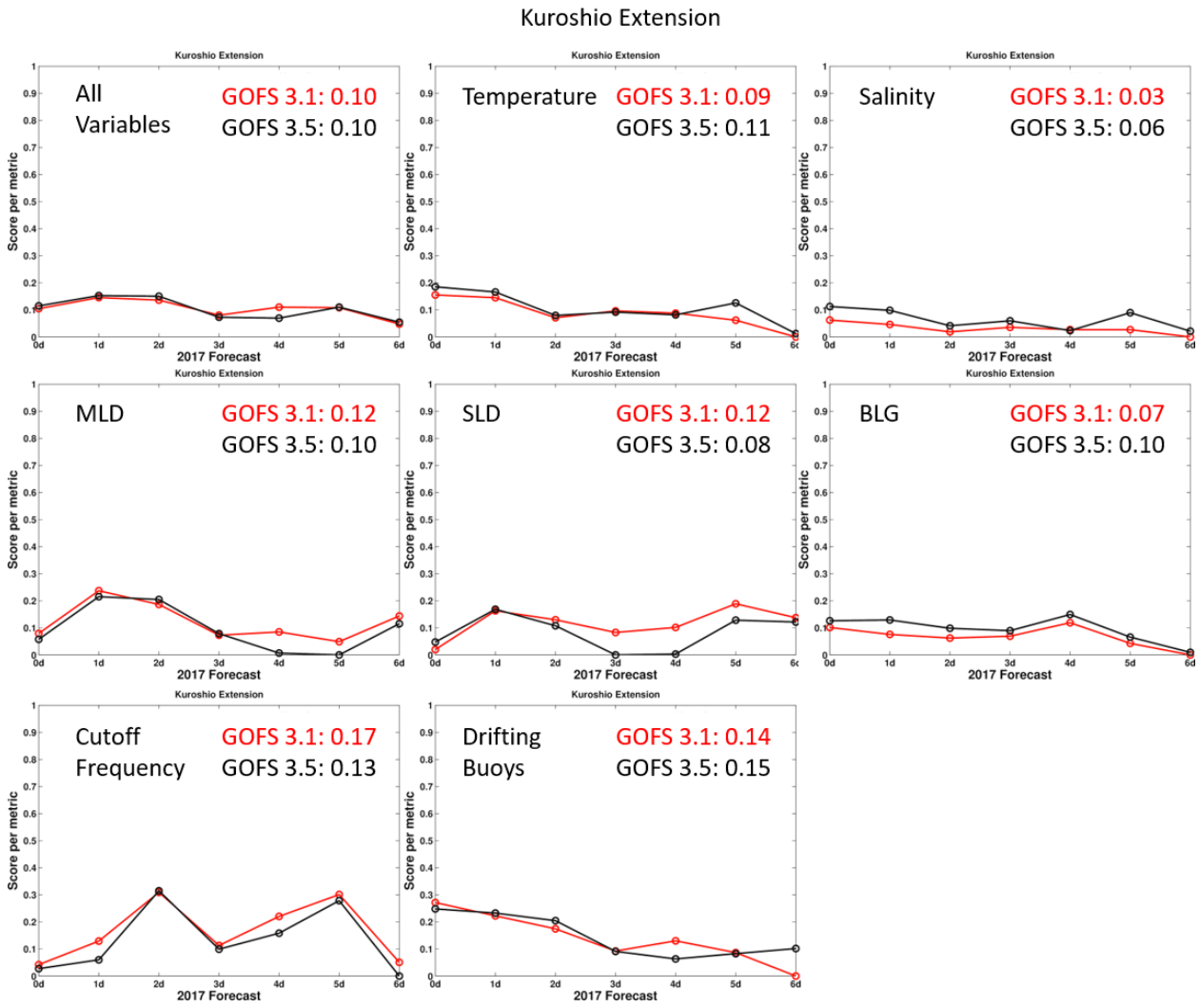

Figure 31: As Figure 29 except for the Kuroshio Extension region, and without the D26, D20, D15 metrics that only apply to the global domain.

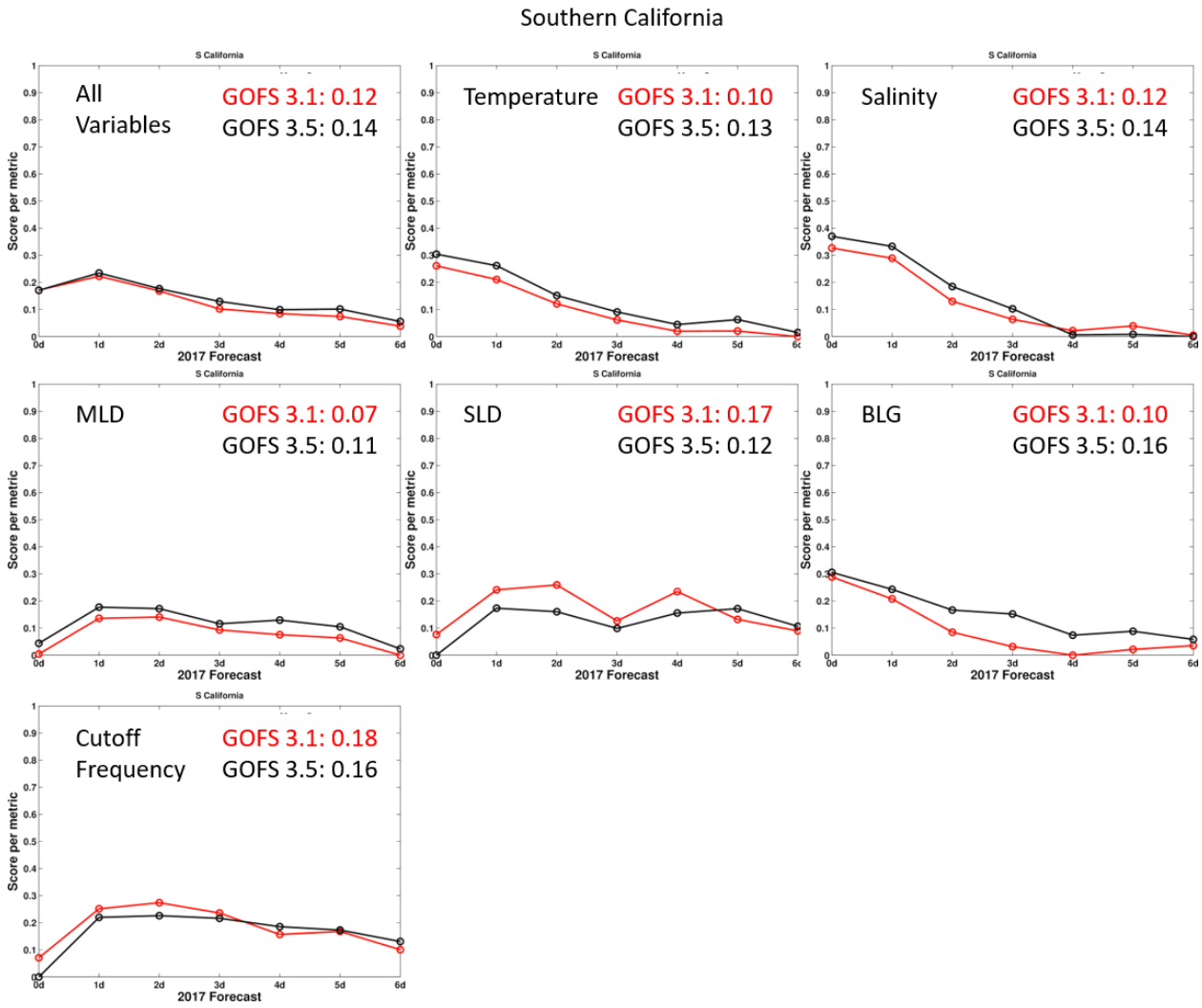

Figure 32: As Figure 29 except for the Southern California region, and without the D26, D20, D15 metrics that only apply to the global domain, and the drifting buoy metrics.

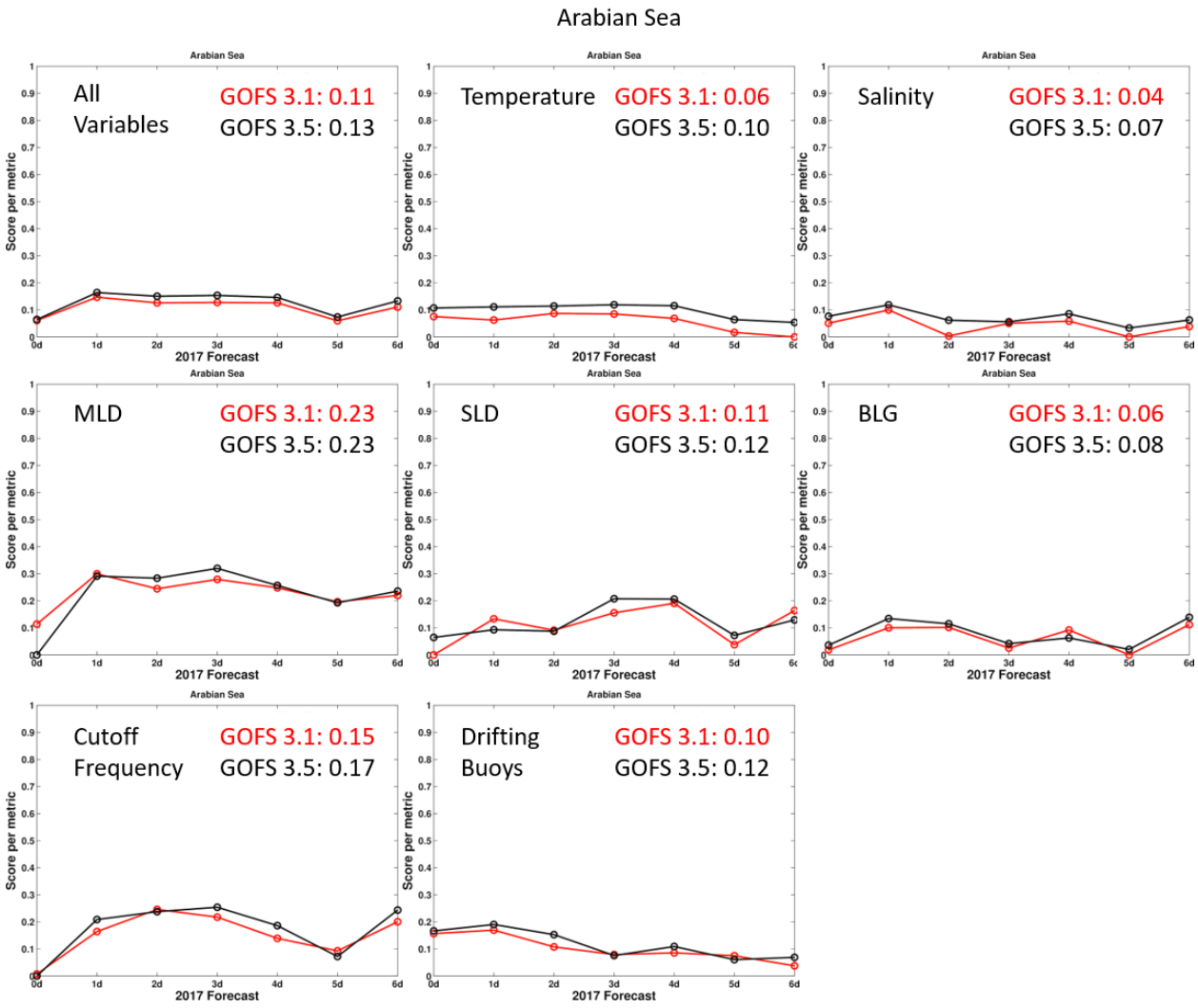

Figure 33: As Figure 29 except for the Arabian Sea region, and without the D26, D20, D15 metrics that only apply to the global domain.

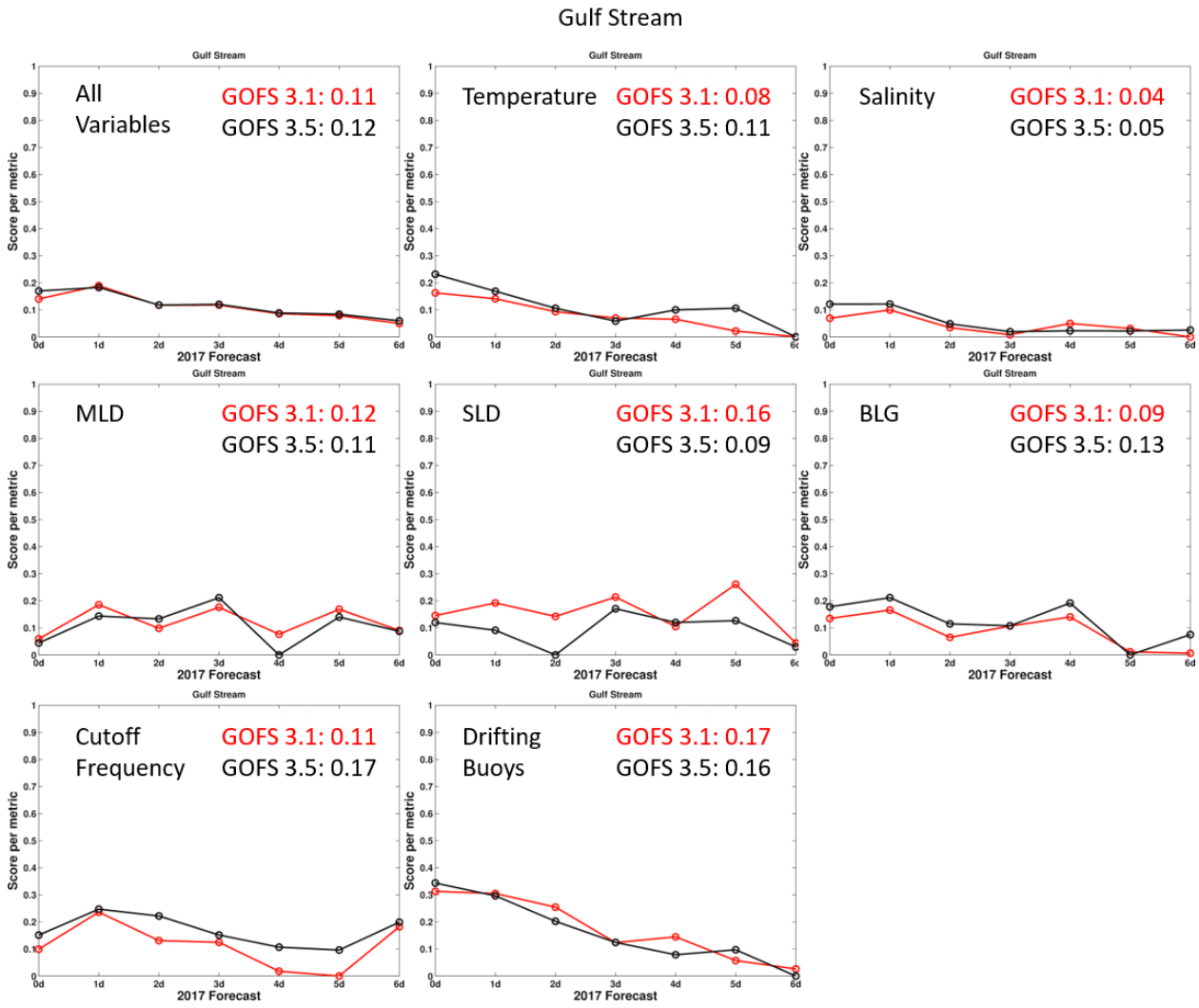

Figure 34: As Figure 29 except for the Gulf Stream region, and without the D26, D20, D15 metrics that only apply to the global domain.

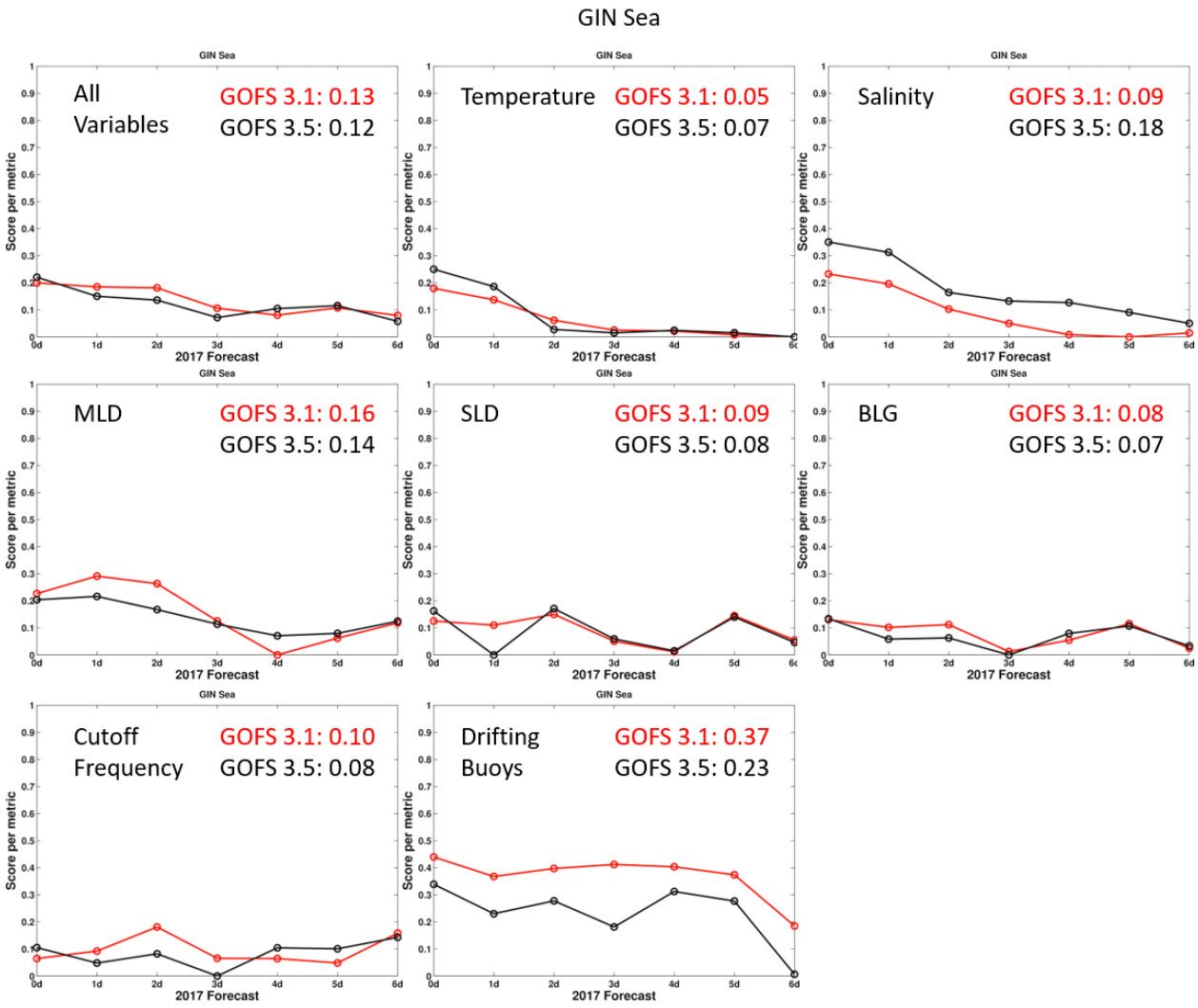

Figure 35: As Figure 29 except for the GIN Sea region, and without the D26, D20, D15 metrics that only apply to the global domain.

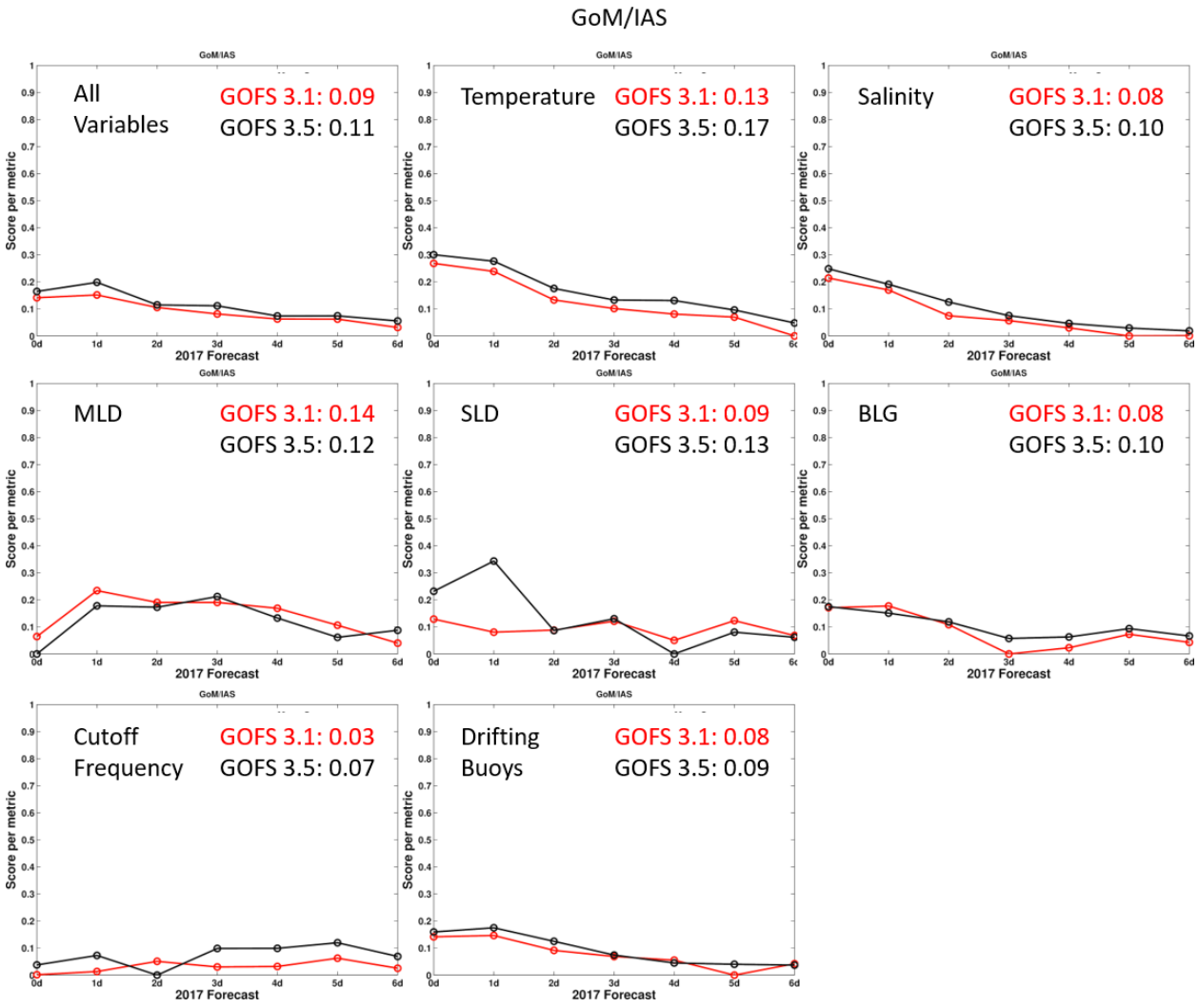

Figure 36: As Figure 29 except for the Gulf of Mexico/Intra-Americas Sea region, and without the D26, D20, D15 metrics that only apply to the global domain.

Table 3 consolidates *Score_model(i,t)* of the individual metrics averaged across all forecasts. GOFS 3.5 has higher scores for T/S vs. depth (and hence the isotherm depth metric) than GOFS 3.1 for all eight analysis regions. However, GOFS 3.1 generally has higher scores with respect to the acoustical metrics. GOFS 3.1 also performs slightly better with regard to upper ocean velocity. This is likely due to the fact that higher horizontal resolution GOFS 3.5 has more unconstrained small scale features than GOFS 3.1.

Table 3: *Score_model(i,t)* for GOFS 3.1 and GOFS 3.5 averaged over all forecasts. The higher score is shaded green.

	Globe		Western Pacific		Kuroshio		So. Cal.		Arabian Sea		Gulf Stream		GIN Sea		GoM/ IAS	
Metric	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**	**3.1**	**3.5**
T(z)	.07	.08	.09	.10	.09	.11	.10	.13	.06	.10	.08	.11	.05	.07	.13	.17
S(z)	.04	.06	.08	.11	.03	.06	.12	.14	.04	.07	.04	.05	.09	.18	.08	.10
D26	.06	.08	—	—	—	—	—	—	—	—	—	—	—	—	—	—
D20	.06	.07	—	—	—	—	—	—	—	—	—	—	—	—	—	—
D15	.08	.09	—	—	—	—	—	—	—	—	—	—	—	—	—	—
MLD	.15	.13	.11	.13	.12	.10	.07	.11	.23	.23	.12	.11	.16	.14	.14	.12
SLD	.14	.04	.13	.07	.12	.08	.17	.12	.11	.12	.16	.09	.09	.08	.09	.13
BLG	.03	.05	.06	.04	.07	.10	.10	.16	.06	.08	.09	.13	.08	.07	.08	.10
COF	.12	.10	.16	.14	.17	.13	.18	.16	.15	.17	.11	.17	.10	.08	.03	.07
Buoy	.12	.11	.09	.07	.14	.15	—	—	.10	.12	.17	.16	.37	.23	.08	.09

4. SEA ICE VALIDATION TESTING RESULTS

4.1 Sea ice validation

As noted earlier, sea ice distribution and properties are modeled using CICE in GOFS 3.1 and GOFS 3.5. However, the versions of the code are different between the two systems with GOFS 3.1 using CICE v4.0 while GOFS 3.5 employs CICE v5.1.2. As will be shown below, the differences between the versions of CICE are significant with GOFS 3.5 outperforming GOFS 3.1 everywhere in the Northern Hemisphere with regard to the sea ice edge metric.

For the sea ice performance evaluation, the Arctic and Antarctic will be divided into multiple regions (Figure 37), chosen to correspond to the regional seas as well as the Pan-Arctic and Pan-Antarctic domains.

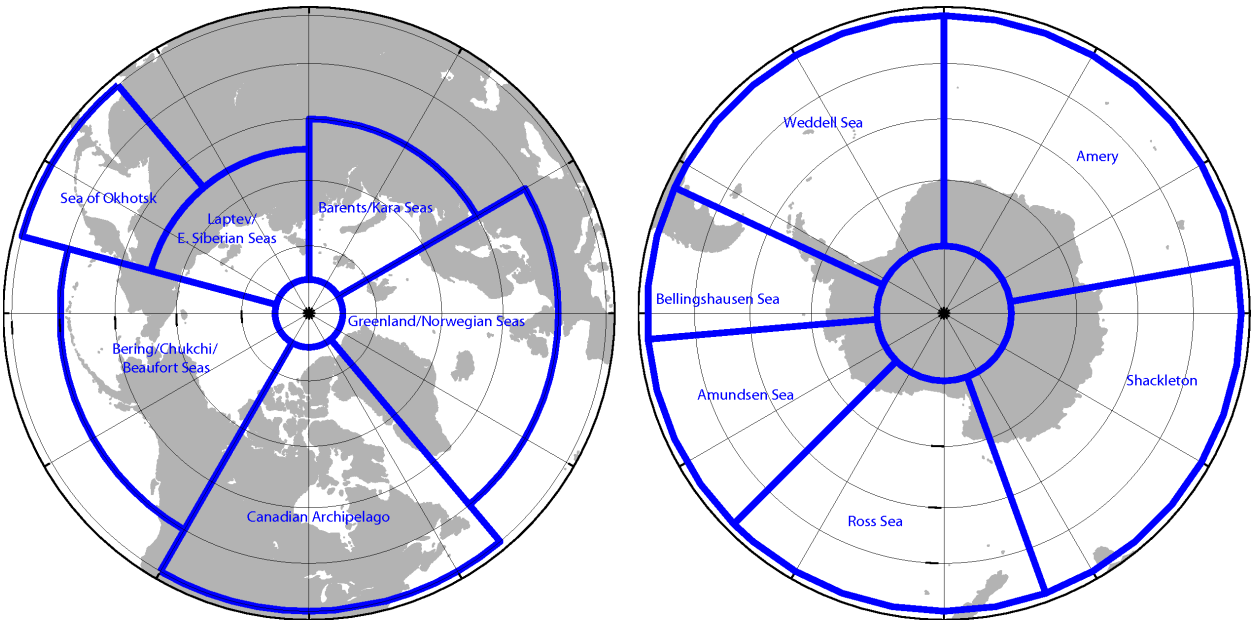

Figure 37: Arctic (left) and Antarctic (right) ice validation analysis regions defined by the regional seas in each hemisphere.

Sea ice extent and concentration are largely controlled by assimilation of satellite sea ice concentration observations. Because both systems assimilate the same data streams there is little difference in the concentration fields (not shown). The systems have a longer "memory" with regard to sea ice thickness and differences are attributed to the initial conditions that started each reanalysis. Satellite measurements (e.g. CryoSat-2) of Pan-Arctic sea ice thickness are available during certain months of the year, but neither the GOFS 3.1 nor 3.5 reanalyses were initialized this way. Figure 38 shows the sea ice thickness for mid-March and mid-September 2017 for both systems. GOFS 3.5 has thicker sea ice across most of the Arctic, but given the limited thickness observations it is not known which is more realistic. In the Antarctic, the two systems have similar sea ice thickness because of the seasonal nature of the first year ice.

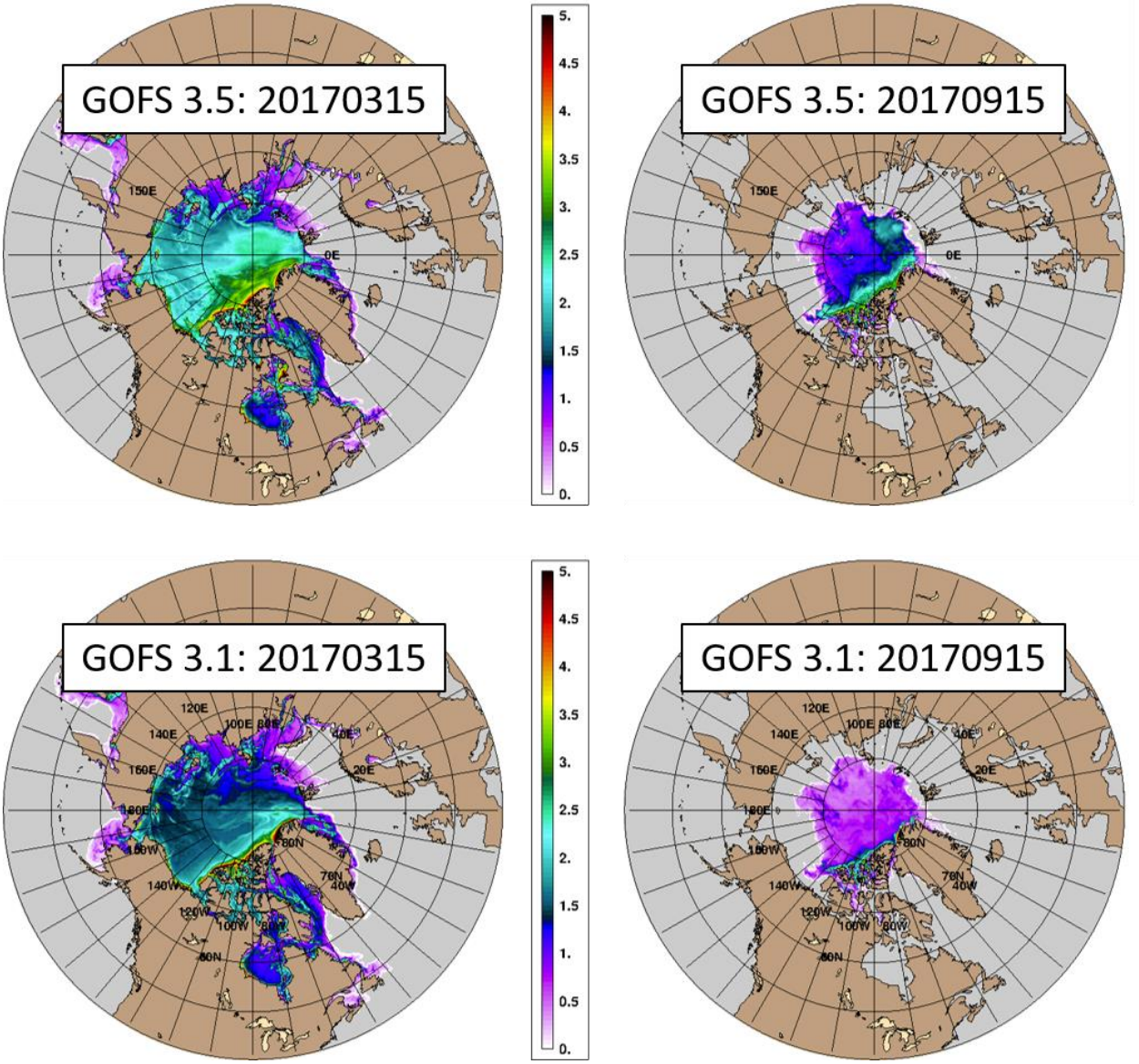

Figure 38: Sea ice thickness (m) for the Pan-Arctic on 15 March 2017 (left column) and 15 September 2017 (right column) for GOFS 3.5 (top row) and GOFS 3.1 (bottom row).

4.2 Sea ice edge location error analyses

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 NIC produces a daily sea 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 4, the differences in the sea ice edge error are listed for GOFS 3.1 and 3.5 at tau 000 and tau 024, i.e. three and 27 hours after direct insertion of the NCODA sea ice analysis at 09Z, respectively. This analysis is performed for every day of 2017. GOFS 3.5 with CICE v5.1.2 performs better than GOFS 3.1 with CICE v4.0 over the Pan-Arctic and regional seas with an average reduction in the sea ice edge error of ~6% at tau 000 and ~11% at tau 024.

Table 4: Sea ice edge error (km) relative to the independent NIC ice edge for the Pan-Arctic and the regional seas show in Figure 37. Lower error is highlighted in green. These statistics are based on every day of the 2017 reanalysis.

Region	Tau	GOFS 3.1	GOFS 3.5	% Difference
Pan-Arctic	000	29.7	27.9	5.9
	024	24.6	23.5	11.5
GIN Sea	000	33.7	31.7	6.0
	024	25.3	22.0	13.0
Barents/Kara	000	24.5	23.1	5.6
	024	23.5	21.4	9.3
Laptev/E. Siberian	000	29.1	27.6	5.2
	024	25.0	23.1	7.6
Sea of Okhotsk	000	25.0	22.9	8.5
	024	19.8	17.0	13.9
Bering/Chukchi /Beaufort	000	23.2	21.3	7.9
	024	21.8	19.7	9.7
Canadian Archipelago	000	37.6	35.9	4.5
	024	27.6	23.5	15.0
“Wins”		0	14	

As seen in Figure 39, 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 sea 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.

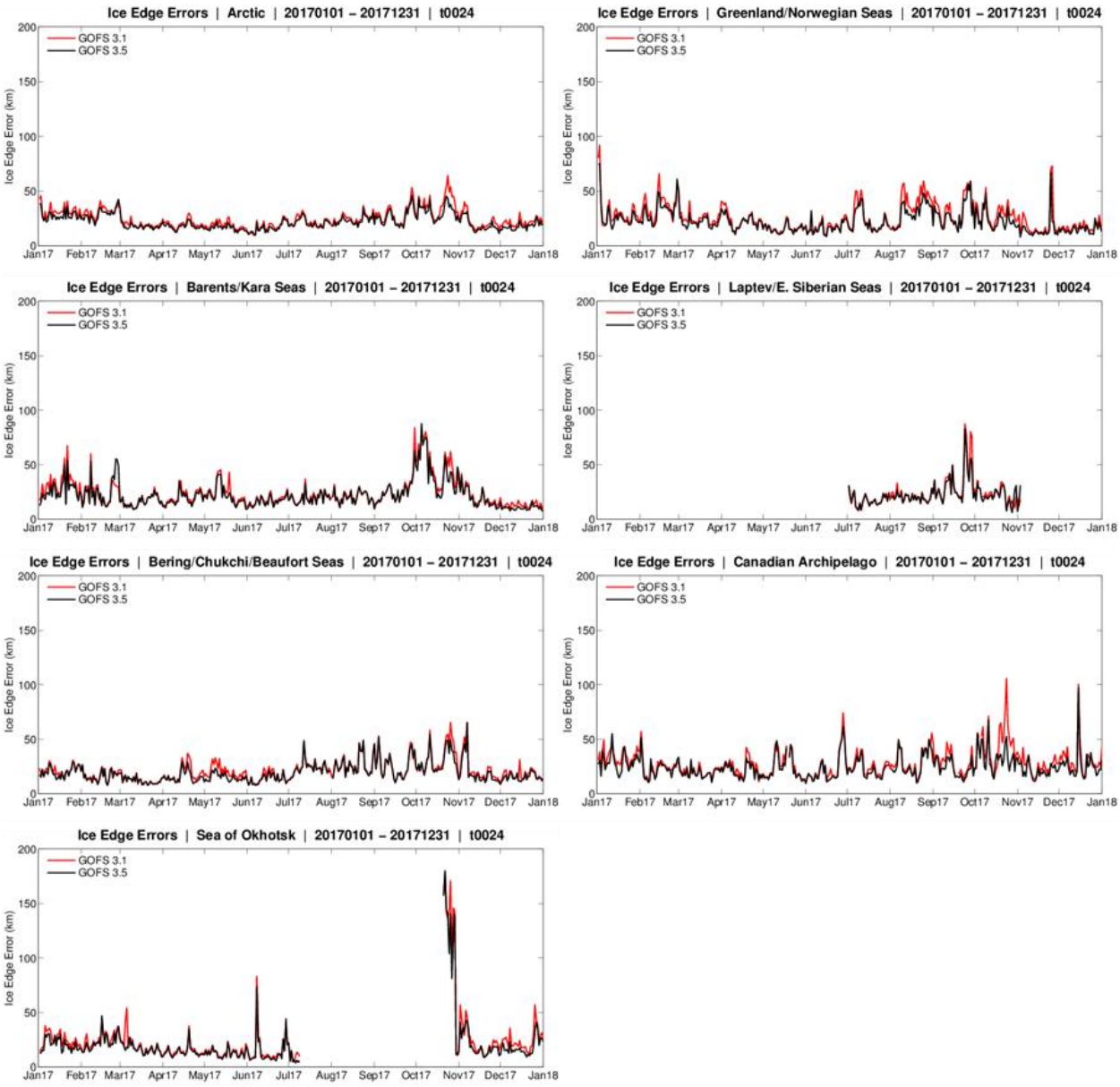

Figure 39: Daily sea ice edge errors (km) for the Pan-Arctic and regional domains shown in Figure 37 for GOFS 3.1 (red) and GOFS 3.5 (black) at tau 024.

In Figure 40, the sea ice edge error as a function of forecast day is shown. This analysis uses the 72 7-day forecasts. GOFS 3.5 using CICE v5.1.2 always performs better than the GOFS 3.1 using CICE v4.0 forecasts. 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.

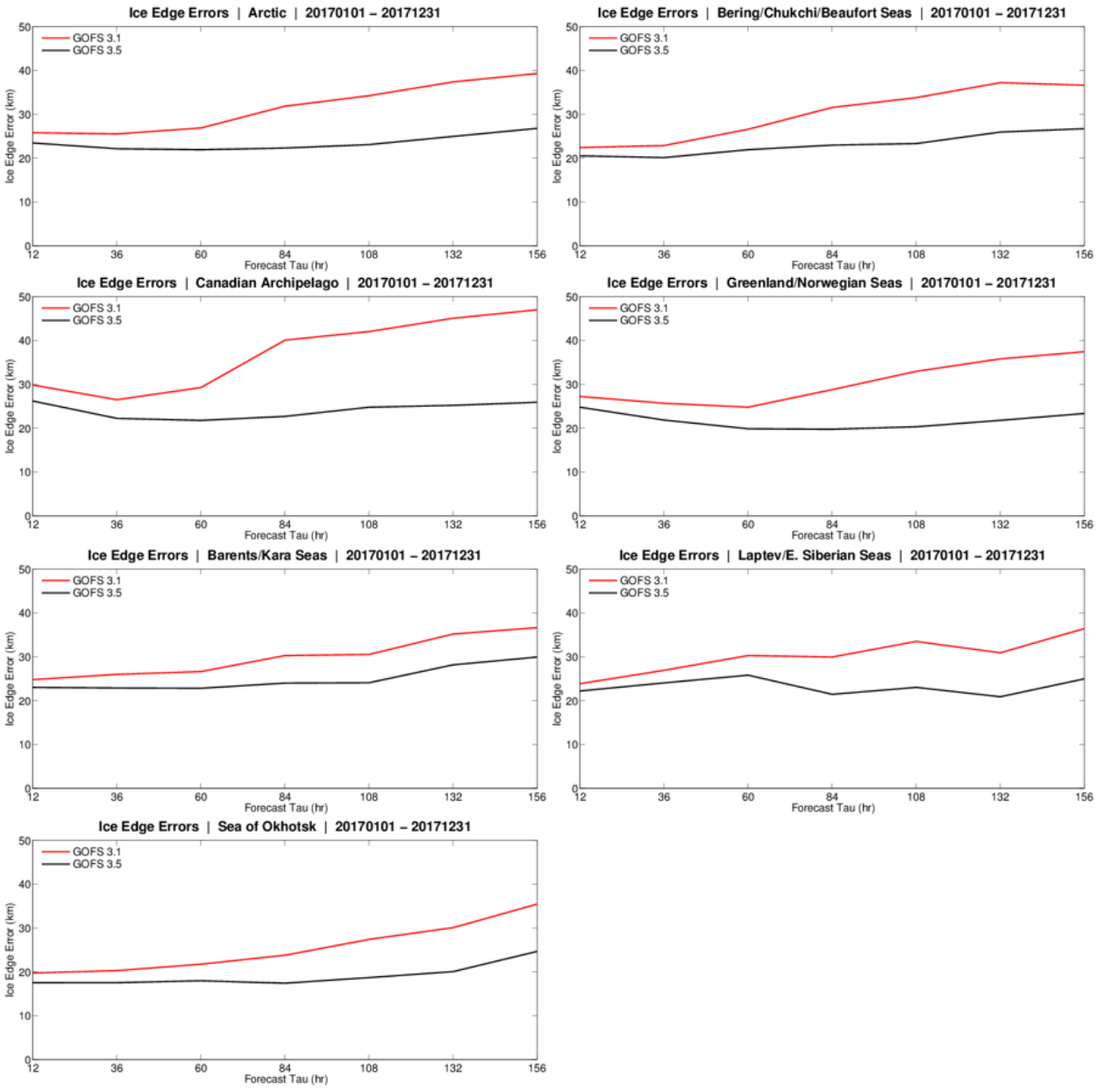

Figure 40: Northern Hemisphere sea ice edge error (km) vs. the independent NIC ice edge as a function of forecast length for the 72 forecasts made over calendar year 2017 for GOFS 3.1 (red) and GOFS 3.5 (black) for the Pan-Arctic and regional domains shown in Figure 37.

Similar analyses can be performed for the Antarctic. In Table 5, the differences in the sea ice edge error are listed and Figure 41 shows daily sea ice edge error at tau 024. This analysis is performed for every day of 2017. Unlike the Arctic, GOFS 3.5 underperforms GOFS 3.1 in many regions at these short forecast lengths. GOFS 3.5 generally has lower error during the melt season (September through December) but higher error during the freeze season (January through May).

Table 5: Sea ice edge error (km) relative to the independent NIC ice edge for the Pan-Antarctic and the regional seas show in Figure 37. Lower error is highlighted in green. These statistics are based on every day of the 2017 reanalysis for each system.

Region	Tau	GOFS 3.1	GOFS 3.5	% Difference
Pan-Antarctic	000	37.7	38.9	-3.1
	024	36.9	36.9	0.0
Amery	000	40.0	42.2	-5.6
	024	38.6	39.7	-2.8
Shackleton	000	33.3	29.6	11.1
	024	32.7	27.4	16.1
Ross	000	39.7	41.8	-5.2
	024	37.7	39.9	-5.7
Amundsen	000	33.7	42.4	-26.0
	024	32.9	40.9	-24.3
Bellinghausen	000	29.1	36.8	-26.8
	024	29.2	36.1	-23.8
Weddell	000	44.0	41.8	5.1
	024	44.1	39.5	10.4
“Wins”		9	4	

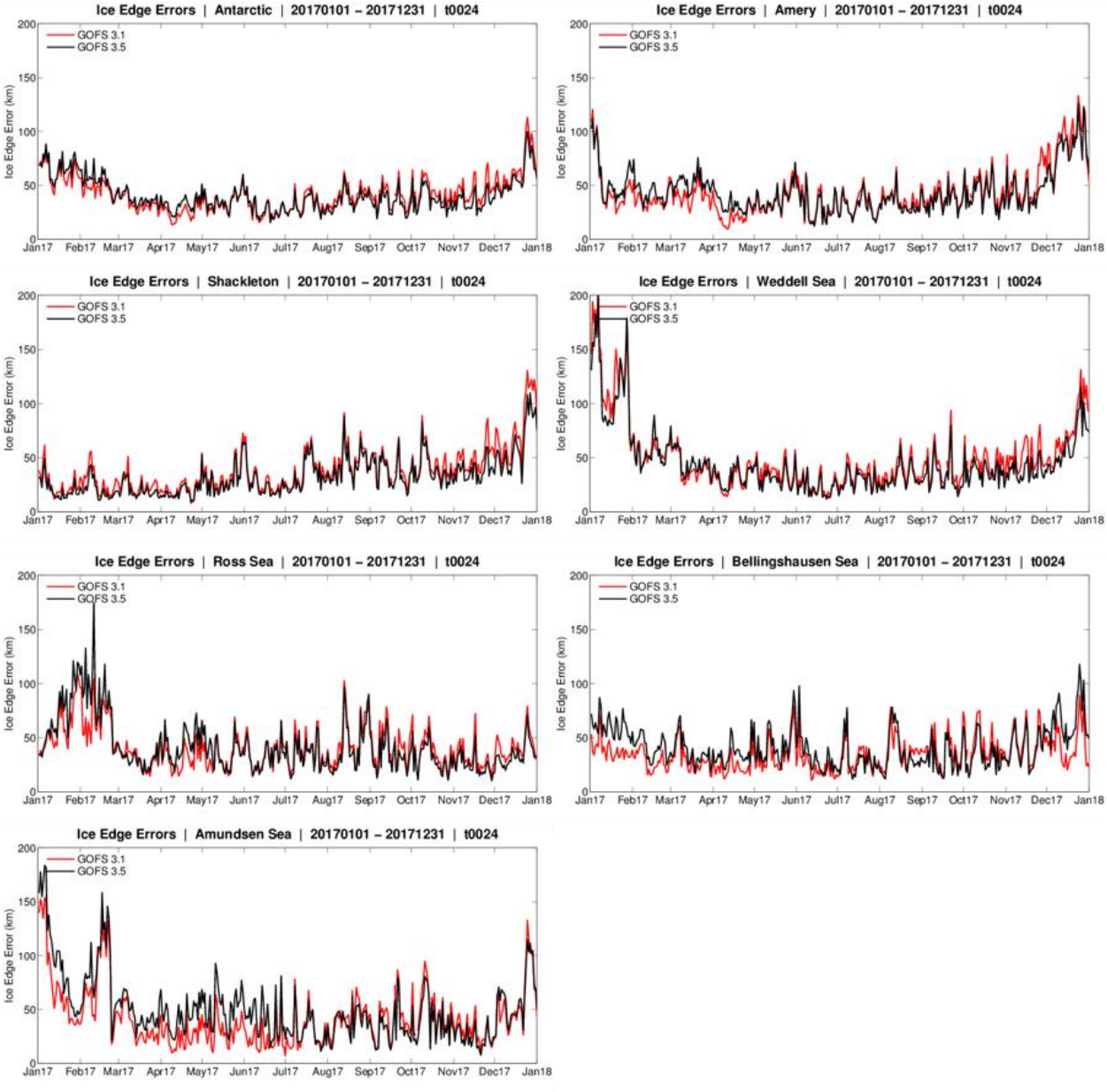

Figure 41: Daily ice edge errors (km) for the Pan-Antarctic and regional domains shown in Figure 37 for GOFS 3.1 (red) and GOFS 3.5 (black) at the 24-hour forecast time.

Figure 42 shows sea ice edge error as a function of forecast length for the Southern Hemisphere. For the Pan-Antarctic, GOFS 3.1 generally has lower error at the short forecast lengths but then underperforms GOFS 3.5 beyond tau 024. Other strong regional differences are shown.

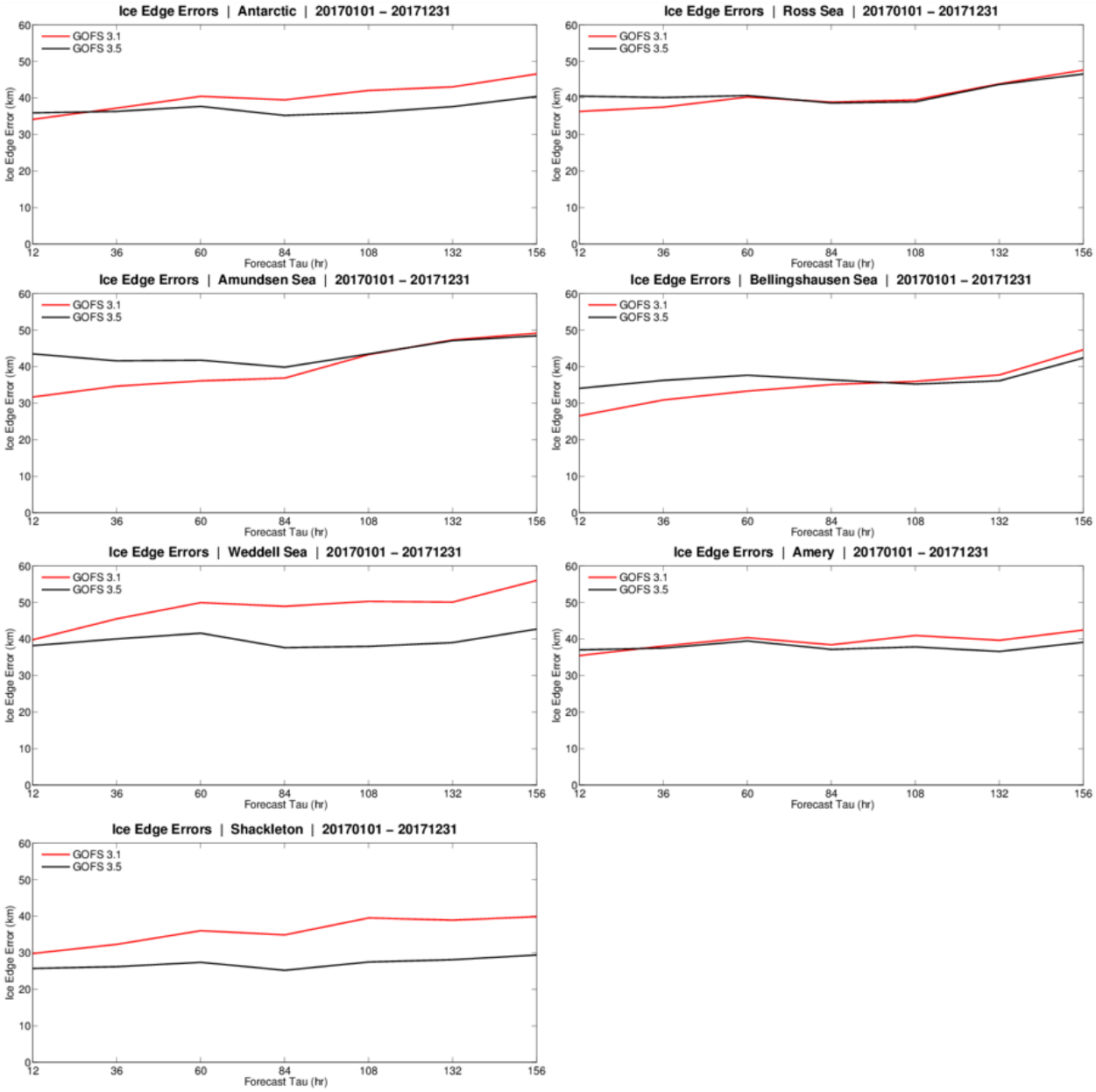

Figure 42: Southern Hemisphere sea ice edge error (km) vs. the independent NIC ice edge as a function of forecast length for the 72 forecasts made over calendar year 2017 for GOFS 3.1 (red) and GOFS 3.5 (black) for the Pan-Antarctic and regional domains shown in Figure 37. GOFS 3.1 use CICE v4.0 whereas GOFS 3.5 uses CICE v5.1.2.

4.3 Sea ice thickness error analyses

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 each) and only containing a small

number (14) of flights, IceBridge data are used in validating sea ice thickness. Figure 43 shows the OIB flight paths for 2017. The flight data have very high temporal resolution but were filtered via a running mean that smoothed the time series sufficiently. Several time filters were examined and a six-minute filter produced reasonable results. This equates to approximately 45 km of flight path. GOFS 3.1 and GOFS 3.5 output at the nowcast time were interpolated to these flight paths for the following error analysis.

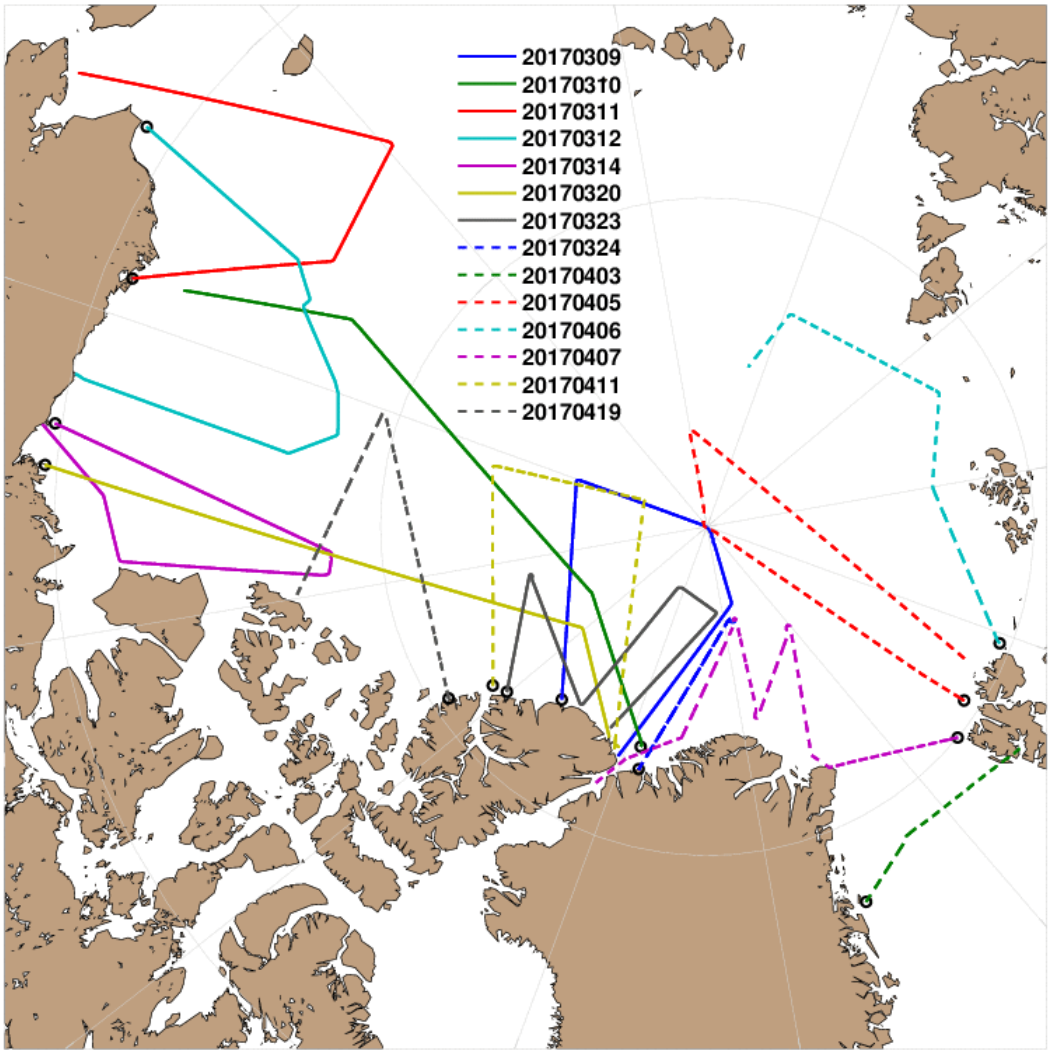

Figure 43: Operation IceBridge flight paths for 2017. The black circle indicates the starting point of each flight.

Figure 44 shows sea ice thickness for 12 OIB flights along with GOFS 3.1 and GOFS 3.5. The IceBridge data exhibit significantly more spatial variability than any of the forecast systems,

even after the spatial smoothing is applied. Table 6 summarizes the error statistics for these flights and GOFS 3.5 outperforms GOFS 3.1 for this limited set of observations. GOFS 3.1 has an overall thin bias whereas GOFS 3.5 and ESPC are on average slightly too thick.

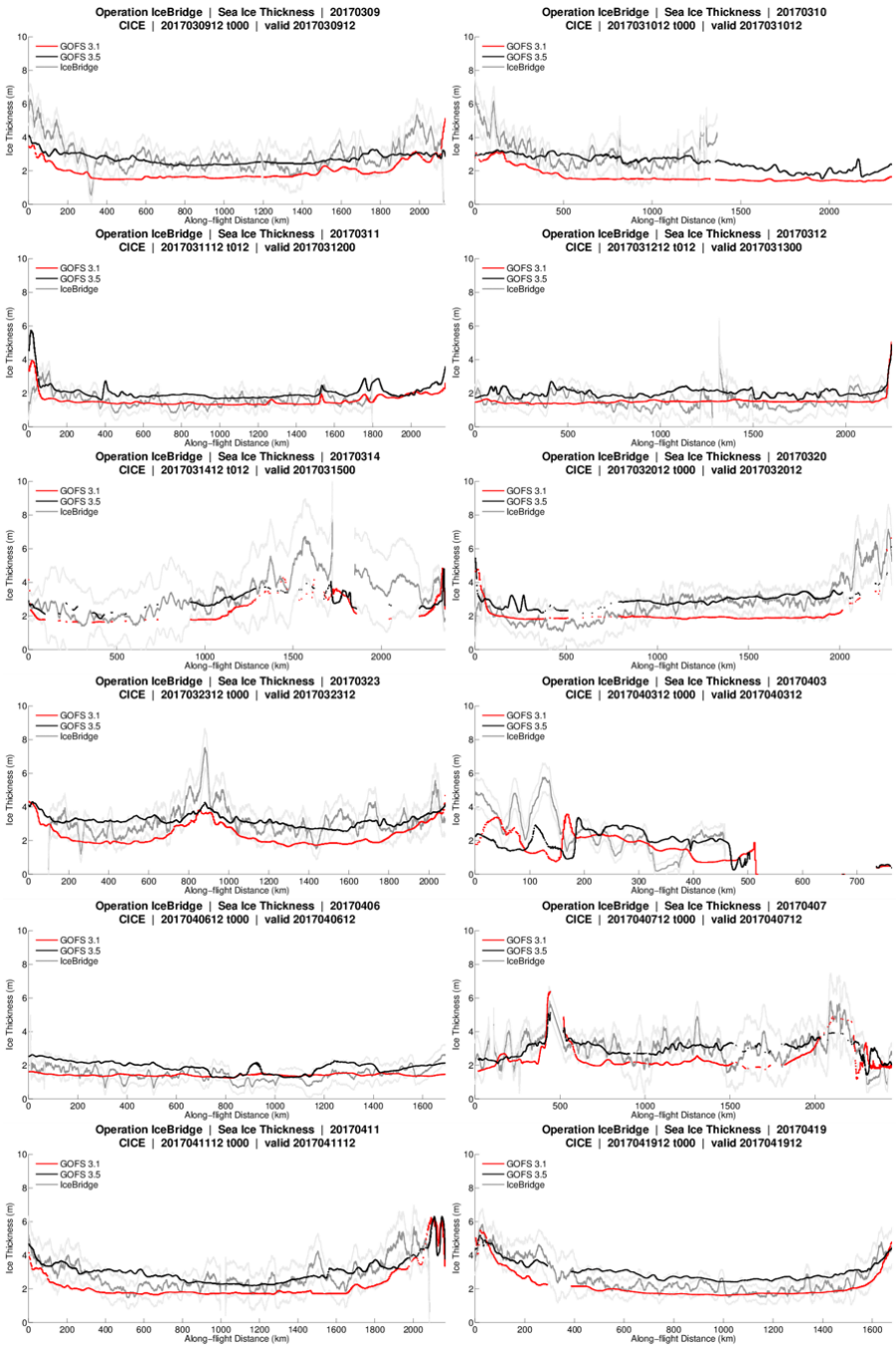

Figure 44: Sea ice thickness (m) versus time for GOFS 3.1 (red) and GOFS 3.5 (black) relative to Operation IceBridge flight data (gray with uncertainty estimate as lighter gray). Each panel represents a different flight as shown in Figure 43.

Table 6: Operation IceBridge sea ice thickness error against GOFS 3.1 and GOFS 3.5 sampled along the same flight paths and dates for 2017. Units are meters. Lower errors/total "wins" are highlighted in green.

	Mean Error		Mean Absolute Error		RMSE	
Flight	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5
20170309	-0.91	-0.10	0.97	0.49	1.11	0.67
20170310	-1.07	-0.04	1.07	0.58	1.19	0.72
20170311	-0.10	0.42	0.32	0.48	0.46	0.69
20170312	-0.07	0.44	0.36	0.52	0.45	0.61
20170314	-0.72	-0.24	0.92	0.56	1.08	0.82
20170320	-0.60	0.33	0.76	0.46	0.88	0.61
20170323	-0.75	0.20	0.79	0.54	0.94	0.68
20170324	-1.12	-0.50	1.12	0.50	1.19	0.58
20170403	-0.50	-0.14	1.06	1.03	1.42	1.30
20170405	-0.82	0.05	0.83	0.40	0.98	0.54
20170406	-0.05	0.42	0.28	0.48	0.36	0.56
20170407	-0.59	0.03	0.86	0.50	0.99	0.65
20170411	-0.54	0.35	0.60	0.50	0.74	0.65
20170419	-0.38	0.51	0.41	0.61	0.54	0.67
"Wins"	4	10	4	10	4	10

A second source of sea ice thickness observations are from the Upward Looking Sonar (ULS) ice draft data from the Woods Hole Oceanographic Institution Beaufort Gyre Exploration Project (Krishfield and Proshutinsky, 2006). Figure 45 shows the location of the three moorings north of Alaska in the Beaufort Sea. Similar to the OIB data, the spatial extent is limited but these observations spanned all of 2017 and thus provide better temporal resolution, albeit at only three distinct locations. Ice draft is converted to sea ice thickness by multiplying the draft by 0.89 (Rothrock et al., 2003). In this region of the Arctic, GOFS 3.1 sea ice thickness is closer to the observational data than GOFS 3.5.

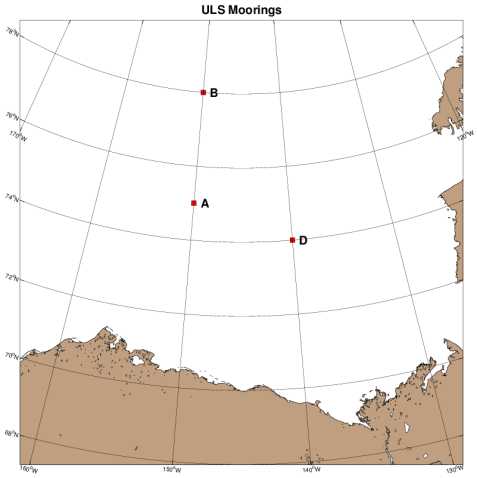

Figure 45: Location of the three Upward Looking Sonar moorings maintained the Woods Hole Oceanographic Institution in the Beaufort Sea north of Alaska.

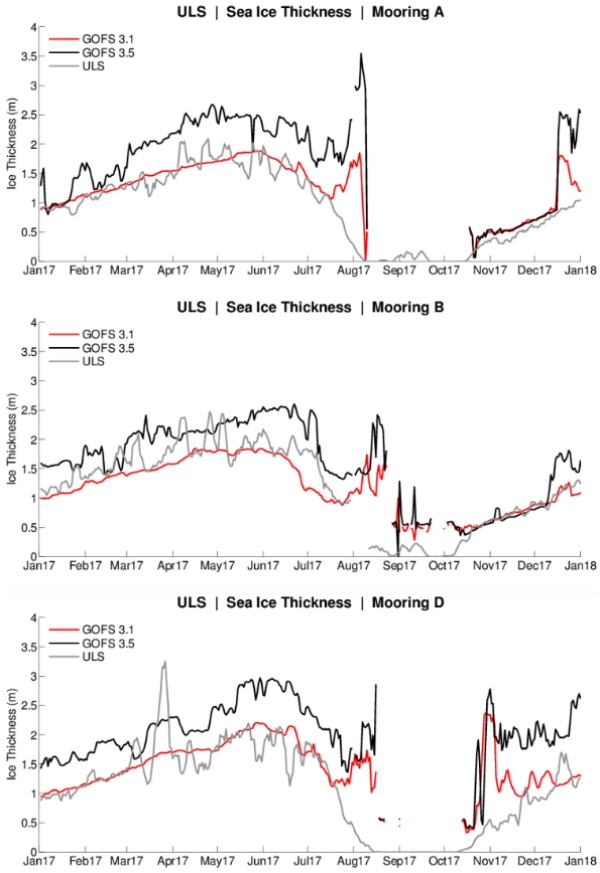

Figure 46: Sea ice thickness (m) versus time for GOFS 3.1 (red) and GOFS 3.5 (black) relative to ULS data (gray) for the moorings shown in Figure 45.

Table 7: ULS sea ice thickness error against GOFS 3.1 and GOFS 3.5 at the three moorings in the Beaufort Sea for 2017. Units are meters. Lower errors/total "wins" are highlighted in green.

	Mean Error		Mean Absolute Error		RMSE	
Mooring	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5
A	0.16	0.62	0.22	0.62	0.36	0.84
B	-0.04	0.36	0.24	0.40	0.36	0.44
D	0.23	0.78	0.36	0.82	0.56	0.94
"Wins"	3	0	3	0	3	0

No sea ice thickness observations are known to exist for 2017 for the Antarctica, thus this thickness analysis is exclusive to the Northern Hemisphere.

4.4 Sea ice drift error analyses

Sea ice drift from the two 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 47) 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, sea ice speed absolute mean error and RMSE, and sea ice direction vector correlation are computed. Figure 48 shows the monthly variation of these quantities over the Pan Arctic region for the 12-hour forecasts at all ice concentrations levels and Table 8 the yearly averaged statistics. GOFS 3.1 has slightly lower MAE and RMSE, and slightly higher vector correlation than GOFS 3.5, although not significantly so.

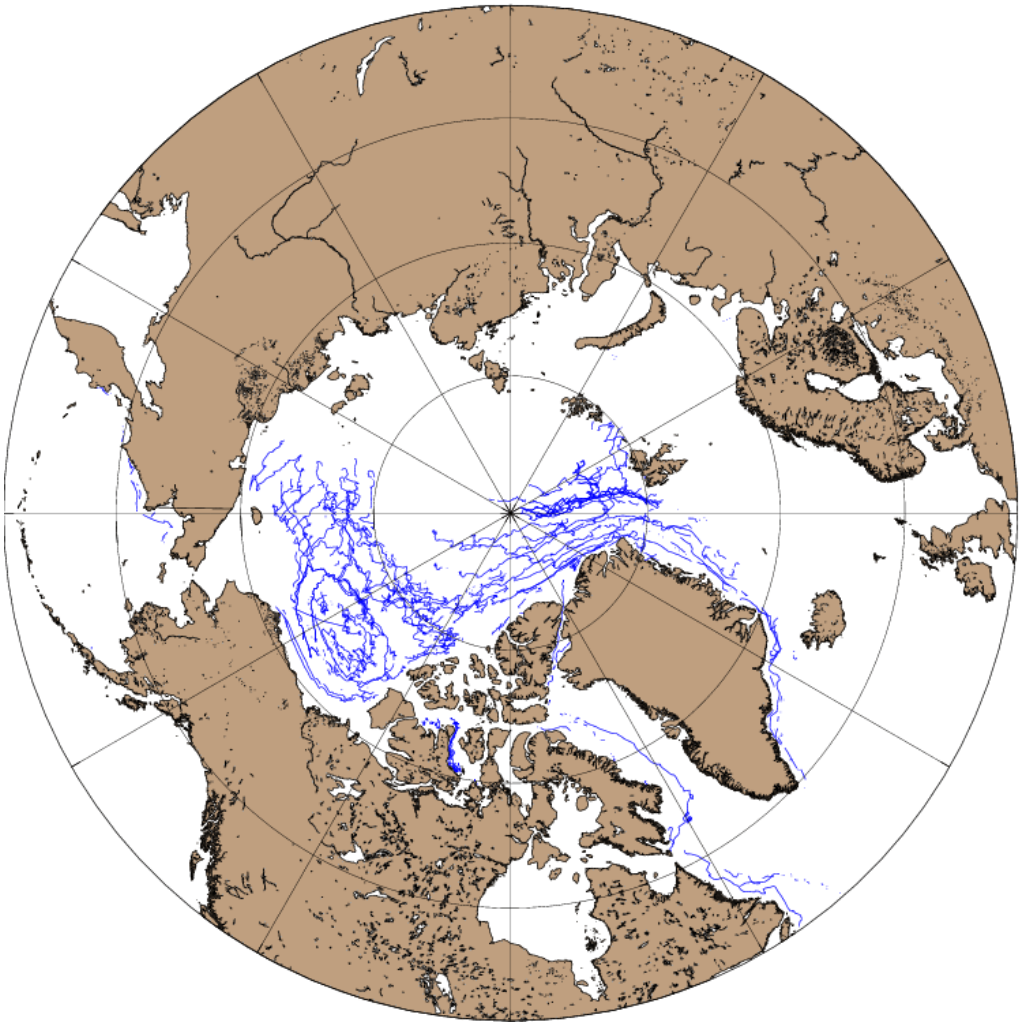

Figure 47: The International Arctic Buoy Program ice-bound drifting buoy tracks over calendar year 2017 used in the sea ice drift error analysis. A total of 207 buoys are shown.

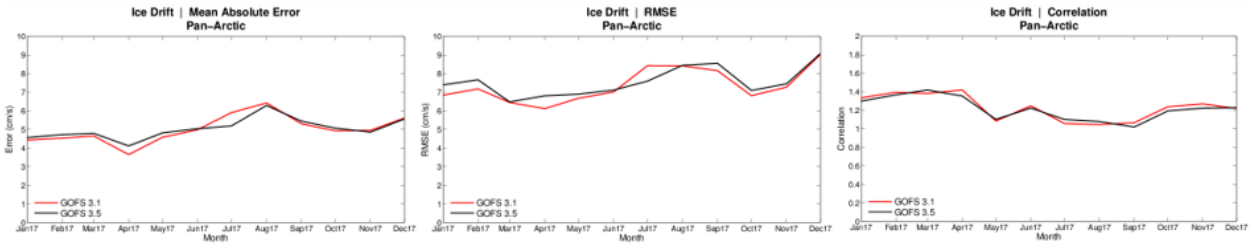

Figure 48: Monthly speed (left) mean absolute error (cm/s), (middle) RMSE (cm/s), and (right) direction vector correlation of unassimilated IABP ice-bound drifting buoys for the Pan Arctic region compared against GOFS 3.1 (red) and GOFS 3.5 (black) for tau 012 at all ice concentrations. The vector correlation 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 8: Error statistics for GOFS 3.1 and GOFS 3.5 compared against IABP drifting buoys spanning all of 2017. Units of MAE and RMSE are cm/s. The vector correlation is based on Crosby et al. (1993) with 0 indicating no correlation and 2 indicating perfect correlation. Lower errors/higher vector correlation/higher total "wins" are highlighted in green.

	Mean Absolute Error		RMSE		Vector Correlation	
Region	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5	GOFS 3.1	GOFS 3.5
Pan-Arctic	4.9	4.9	7.3	7.4	1.27	1.26
Central Arctic	4.4	4.4	5.5	5.4	1.65	1.65
GIN Sea	5.7	5.9	9.1	9.5	1.01	0.97
Canadian Archipelago	5.1	5.1	8.2	8.3	0.92	0.94
Bering/Chukchi	4.6	4.6	6.3	6.4	1.43	1.43
"Wins"	1	0	5	0	2	1

Northern Hemisphere sea ice drift is additionally categorized by sea ice concentration to evaluate system performance within both the pack ice and the marginal ice zone (Table 9 and Table 10). 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 subregions. Table 9 shows observed sea 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. Generally the model sea ice drift is too fast compared to the buoys and GOFS 3.1 has lower bias than GOFS 3.5.

Table 9: 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 and GOFS 3.5 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	
		Speed	Bias	Speed	Bias
$0.8 < X \leq 1.0$ (107)	8.6	11.2	2.6	11.3	2.7
$0.6 < X \leq 0.8$ (44)	13.8	14.6	0.8	15.0	1.2
$0.3 < X \leq 0.6$ (28)	13.0	17.5	4.5	20.8	7.8
$0.1 < X \leq 0.3$ (6)	25.3	33.4	8.1	36.4	11.1
$0.0 < X \leq 0.1$ (25)	17.8	18.2	0.4	14.7	-3.1
"Wins"			5		0

Table 10: Sea ice drift statistics of speed MAE 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 and GOFS 3.5 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)	MAE		RMSE		Vector Correlation	
	G3.1	G3.5	G3.1	G3.5	G3.1	G3.5
$0.8 < X \leq 1.0$ (107)	4.4	4.4	6.4	6.3	1.41	1.43
$0.6 < X \leq 0.8$ (44)	6.9	7.0	9.5	9.5	0.95	0.87
$0.3 < X \leq 0.6$ (28)	9.3	11.7	11.9	14.7	0.98	0.78
$0.1 < X \leq 0.3$ (6)	15.0	10.8	19.2	16.5	1.05	1.50
$0.0 < X \leq 0.1$ (25)	11.8	15.3	17.7	19.5	0.80	0.53
"Wins"	3	1	3	2	3	2

5. SUMMARY AND RECOMMENDATIONS

GOFS 3.5 represents the existing state-of-the-art Navy global prediction system, with three main differences compared to GOFS 3.1. First, the horizontal resolution doubles from 0.08° to 0.04° degrees at the equator. Fine resolution is needed to obtain inertial jets and sharp ocean fronts and the associated nonlinear recirculation gyres that affect the shape of large-scale ocean gyres. Additionally, it allows for more accurate upper ocean – topographic coupling via flow instabilities that can have major impacts on current pathways. Secondly, GOFS 3.5 includes astronomical tidal forcing that generates internal gravity waves at tidal frequencies, i.e. internal tides. Lastly, a more up-to-date sea ice model is implemented in GOFS 3.5.

Numerous ocean metrics are presented that include T and S vs. depth error, isotherm depth error, acoustical proxy (MLD, SLD, BLG, and COF) error and upper ocean velocity error. These metrics are quantified in section 3.10 and show the relative performance of GOFS 3.1 vs. GOFS 3.5 via an ocean scorecard (Figure 28 through Figure 36 and Table 3). With regard to T/S vs. depth error and isotherm depth error, GOFS 3.5 clearly outperforms GOFS 3.1 across all eight analysis regions. The acoustical proxy error results are mixed with GOFS 3.1 generally performing better

for SLD, but GOFS 3.5 generally performing better for BLG. The upper ocean velocity errors are also mixed.

Tides in GOFS 3.5 are also analyzed and are shown to be in good agreement with data assimilative inverse tidal solutions. The generation locations of the M_2 internal tides also agrees well with satellite altimeter data. The amplitude of the internal tides in GOFS 3.5 is also consistent with estimates derived from altimeter data.

Sea ice edge error shows a clear win for GOFS 3.5 using CICE v5.1.2 vs. GOFS 3.1 using CICE v4.0 in the Arctic. Ice edge error as a function of forecast length also grows more slowly in GOFS 3.5 providing more confidence at longer forecast lengths. Sea ice thickness error analyses in the Northern Hemisphere are included for completeness, but are very limited in spatial and temporal extent and thus not applicable to the entire Arctic domain. Sea ice drift error analyses are limited to the Western Hemisphere and show mixed results. GOFS 3.1 typically has lower error, but is not significantly different from GOFS 3.5.

In the aggregate, GOFS 3.5 performs equal to or better than GOFS 3.1 and it is recommended that GOFS 3.5 replace GOFS 3.1 as the U.S. Navy's stand-alone ocean/sea ice forecast system.

6. ACKNOWLEDGMENTS

This work was funded as part of the 6.4 Large Scale Ocean Prediction and 6.4 Ocean Data Assimilation projects, managed by the Office of Naval Research under program element 0603207N. The numerical reanalyses and forecasts were performed on the Navy DSRC Cray XC40 computers (Conrad and Gordon) at Stennis Space Center, Mississippi using grants of computer time from the DoD High Performance Computing Modernization Program. The Validation Test Panel consisted of: Mark Cobb (FNMOC), Sofia Montalvo (National Ice Center), James Potemra (International Pacific Research Center), Pat Hogan, and Joe Metzger (NRL). This is NRL contribution NRL/MR/7320--19-9975. It is approved for public release, distribution is unlimited.

7. 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.

Arbic, B.K., et al., 2018: A primer on global internal tide and internal gravity wave continuum modeling in HYCOM and MITgcm. In "*New Frontiers in Operational Oceanography*", E. Chassignet, A. Pascual, J. Tintoré, and J. Verron, Eds.,

Bloom, S.C., L.L. Takacs, A.M. da Silva, D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.* 124:1256–1271.

Buijsman, M. C., B. K. Arbic, J. A. M. Green, R. W. Helber, J. G. Richman, J. F. Shriver, P. G. Timko, and A. J. Wallcraft, 2015: Optimizing internal wave drag in a forward barotropic model with semidiurnal tides, *Ocean Modell.*, 85, 42–55.

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.

Chen, C.T. and F.J. Millero, 1977: Speed of sound in seawater at high pressures. *J. Acoust. Soc. Am*., 62, 1129-1135.

Crosby, D.S., L.C. Breaker and W.H. Gemmill, 1993: A proposed definition for vector correlation in geophysics: Theory and application. *J. Atm. Ocean. Tech*., 10, 355-367.

Cummings, J.A. and O.M. Smedstad, 2014. Ocean data impact in global HYCOM. *J. Atmos. Ocean. Technol.*, 31, 1771-1791, doi: 10.1175/jtech-d-14-00011.1.

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.

Egbert, G. D., A. Bennett, and M. Foreman, 1994: Topex/poseidon tides estimated using a global inverse model, *J. Geophys. Res*., 99, 821–852.

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.

GODAE OceanView, 307-392, doi:10.17125/gov2018.ch13.Buijsman, M. C., B. K. Arbic, J. A. M. Green, R. W. Helber, J. G. Richman, J. F. Shriver, P. G. Timko, and A. J. Wallcraft, 2015: Optimizing internal wave drag in a forward barotropic model with semidiurnal tides, *Ocean Modell.,* 85, 42–55.

Helber, R.W., C.N. Barron, M. Gunduz, P.L. Spence and R.A. Zingarelli, 2010: Acoustic metric assessment for UUV observation system simulation experiments. *U.S. Navy Journal of Underwater Acoustics,* 60(1), 101-123.

Helber, R.W., T.L. Townsend, C.N. Barron, J.M. Dastugue and M.R. Carnes, 2013: Validation Test Report for the Improved Synthetic Ocean Profile (ISOP) System, Part I: Synthetic Profile Methods and Algorithm. *NRL Memo. Report*, NRL/MR/7320—13-9364.

Hendershott, M. C., 1972: The effects of solid earth deformation on global ocean tides, *Geophys. J. R. Astron. Soc.,* 29, 389–402.

Hill, C., C. DeLuca, V. Balaji, M. Suarez, and A. da Silva, 2004: The Architecture of the Earth System Modeling Framework. *Computing in Science and Engineering*, 6:18-28.

Hunke, E.C. and W. 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 National Laboratory, Los Alamos, NM. (http://climate.lanl.gov/models/cice/index.htm).

Jayne, S. R., and L. C. St. Laurent, 2001: Parameterizing tidal dissipation over rough topography, *Geophys. Res. Lett*., 28, 811–814.

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.

Lumpkin, R., and M. Pazos, 2007: Measuring surface currents with Surface Velocity Program drifters: The instrument, its data and some recent results. In *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, edited by A. Griffia et al., pp. 39-67, Cambridge Univ. Press, Cambridge, U.K.

Krishfield, R., and A. Proshutinsky, 2006: BGOS ULS Data Processing Procedure. Woods Hole Oceanographic Institution, March 2006, 14 pp.

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.

Metzger, E.J., R.W. Helber, P.J. Hogan, P.G. Posey, P.G. Thoppil, T.L. Townsend, A.J. Wallcraft, O.M. Smedstad, D.S. Franklin, L. Zamudio and M.W. Phelps, 2017: Validation Test Report II for the Global Ocean Forecast System 3.1 - 1/12º HYCOM/NCODA/CICE/ISOP. *NRL Memo. Report,* NRL/MR/7320--17-9718.

Metzger, E.J., O.M. Smedstad, P.G. Thoppil, H.E. Hurlburt, 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. DeHaan, 2014: US Navy operational global ocean and Arctic ice prediction systems. *Oceanography*, Vol. 27. No. 3, 32-43, http://dx.doi.org/10.5670/oceanog.2014.66.

Millero, F.J. and X. Li, 1994: Comments on "On equations for the speed of sound in seawater". *J. Acoust. Soc. Am*., 95(5), 2757-2759.

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 barotropic tides embedded in a high-resolution global ocean circulation model, *Ocean Modell*., 97, 16–26.

Ray, R. D., 1998: Ocean self-attraction and loading in numerical tidal models, *Mar. Geod*., 21, 181–192.

Rothrock, D.A., J. Zhang, and Y. Yu, 2003: The Arctic ice thickness anomaly of the 1990s: a consistent view from observations and models. *J. Geophys. Res*. 108, 3083. https://doi.org/10.1029/2001JC001208.

RP33: Fleet Oceanographic and Acoustic Reference Manual, 1992: *Naval Oceanographic Office Reference Publication 33*, Naval Oceanographic Office, Stennis Space Center, MS.

Sandery, P.A., and P. Sakov, 2017: Ocean forecasting of mesoscale features can deteriorate by increasing model resolution towards the submesoscale. *Nature Communications*, 8:1566, doi:10.1038/s41467-017-01595-0.

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.

Smith, S., J.A. Cummings, C. Rowley, P. Chu, J. Shriver, R. Helber, P. Spence, S. Carroll, O.M. Smedstad and B. Lunde, 2011: Validation Test Report for the Navy Coupled Ocean Data Assimilation 3D Variational Analysis (NCODA-VAR) System, Version 3.43. *NRL Memo. Report*, NRL/MR/7320--11-9363.

Wilson, W.D., 1960: Equation for the speed of sound in sea water. *J. Acoust. Soc. Amer*., 32**,** 1357.

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.

8. TABLE OF ACRONYMS

3DVAR	3-Dimensional VARiational analysis
BLG	Below Layer Gradient
CICE	Community Ice CodE
COF	CutOff frequency
DoD	Department of Defense
DSRC	DoD Supercomputing Resource Center
EKE	Eddy Kinetic Energy
ESMF	Earth System Modeling Framework
ESPC	Earth System Prediction Capability
FGAT	First Guess at Appropriate Time
FNMOC	Fleet Numerical Meteorology and Oceanography Center
GDEM	Generalized Digital Environment Model
GIN	Greenland/Iceland/Norwegian
GOFS	Global Ocean Forecast System
GoM	Gulf of Mexico
HYCOM	HYbrid Coordinate Ocean Model
IABP	International Arctic Buoy Program
IAS	Intra-Americas Sea
IMS	Interactive Multisensor Snow and Ice Mapping System
ISOP	Improved Synthetic Ocean Profiles
MAE	Mean Absolute Error
ME	Mean Error
MLD	Mixed Layer Depth
NAVGEM	NAVy Global Environmental Model
NAVOCEANO	Naval Oceanographic Office
NIC	National Ice Center
NCODA	Navy Coupled Ocean Data Assimilation
NRL	Naval Research Laboratory
OIB	Operation IceBridge
OPTEST	Operational Test
RMSE	Root Mean Square Error
S	Salinity
SAL	Self Attraction and Loading
SAR	Synthetic Aperture Radar
SLD	Sonic Layer Depth
SoCal	Southern California
SSH	Sea Surface Height
SSMIS	Special Sensor Microwave Imager/Sounder
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
T	Temperature
ULS	Upward Looking Sonar
VTP	Validation Test Panel
VTR	Validation Test Report