

## Skill tests of three-dimensional tidal currents in a global ocean model: A look at the North Atlantic

Patrick G. Timko,<sup>1,2,3</sup> Brian K. Arbic,<sup>2</sup> James G. Richman,<sup>4</sup> Robert B. Scott,<sup>5,6</sup>  
E. Joseph Metzger,<sup>4</sup> and Alan J. Wallcraft<sup>4</sup>

Received 22 September 2011; revised 22 June 2012; accepted 26 June 2012; published 9 August 2012.

[1] We present and apply several skill tests to assess tidal currents generated within a global ocean model compared to an archive of current meter records spanning 40 years. Within the North Atlantic we have identified over 1800 velocity records available for comparison to model output. The skill tests identify those regions where tidal ellipse parameters (semi-major/minor axis, inclination and Greenwich phase) fall within 95% confidence intervals derived from observations. Additional skill tests examine the vertical and horizontal structure of the tidal currents using the correlation and root mean square error between the observations and model. The skill tests account for grid misrepresentation within the model by comparing model values in a local 9-point neighborhood. Both the three-dimensional structure of model tidal currents and barotropic model tidal currents are compared to the altimetry based barotropic model TPXO 7.2. Our results indicate that the non-data assimilative ocean model performance is comparable to TPXO 7.2. The semi-major axes of the semi-diurnal constituents lie within the 95% confidence intervals between 40% and 60% of the time in some regions of the North Atlantic. The performance degrades when all tidal ellipse parameters are considered. Root mean square errors between the model and observations distributed through the water column are generally less than  $1 \text{ cm s}^{-1}$  representing an error of less than  $\sim 10\%$  for constituent  $M_2$ . The model tides also exhibit high correlation with observations in the horizontal planes. The skill tests could be adapted to other sets of observations and other models.

**Citation:** Timko, P. G., B. K. Arbic, J. G. Richman, R. B. Scott, E. J. Metzger, and A. J. Wallcraft (2012), Skill tests of three-dimensional tidal currents in a global ocean model: A look at the North Atlantic, *J. Geophys. Res.*, *117*, C08014, doi:10.1029/2011JC007617.

### 1. Introduction

[2] In this paper we present and apply skill tests to assess the ability of a non-data assimilative global ocean general circulation model to simulate three-dimensional tidal

currents within the North Atlantic. Model skill is assessed against an archive of current meters spanning  $\sim 40$  years. The various skill tests include evaluation of model performance using 95% confidence intervals of the individual observations as well as tests based upon the root mean square error and correlation between the model and observations in both vertical and horizontal directions. The tests are applied at the model value nearest to the observations and also within a 9-point neighborhood surrounding the observations in order to test for grid misrepresentation. The model used in this study is the global  $1/12.5^\circ$  Hybrid Coordinate Ocean Model (HYCOM) which has recently implemented a forward tide algorithm (based upon astronomical arguments) [Arbic *et al.*, 2010]. The inclusion of tides in the ocean general circulation model is still under development and hence the results presented regarding the skill of the model represent the current state of the tidal algorithm at time of writing.

[3] The version of global HYCOM [Chassignet *et al.*, 2007] used in this paper has a horizontal resolution of  $1/12.5^\circ$  and 32 hybrid layers in the vertical. During model execution, the model spins up for a period of 10 years

<sup>1</sup>Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, USA.

<sup>2</sup>Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA.

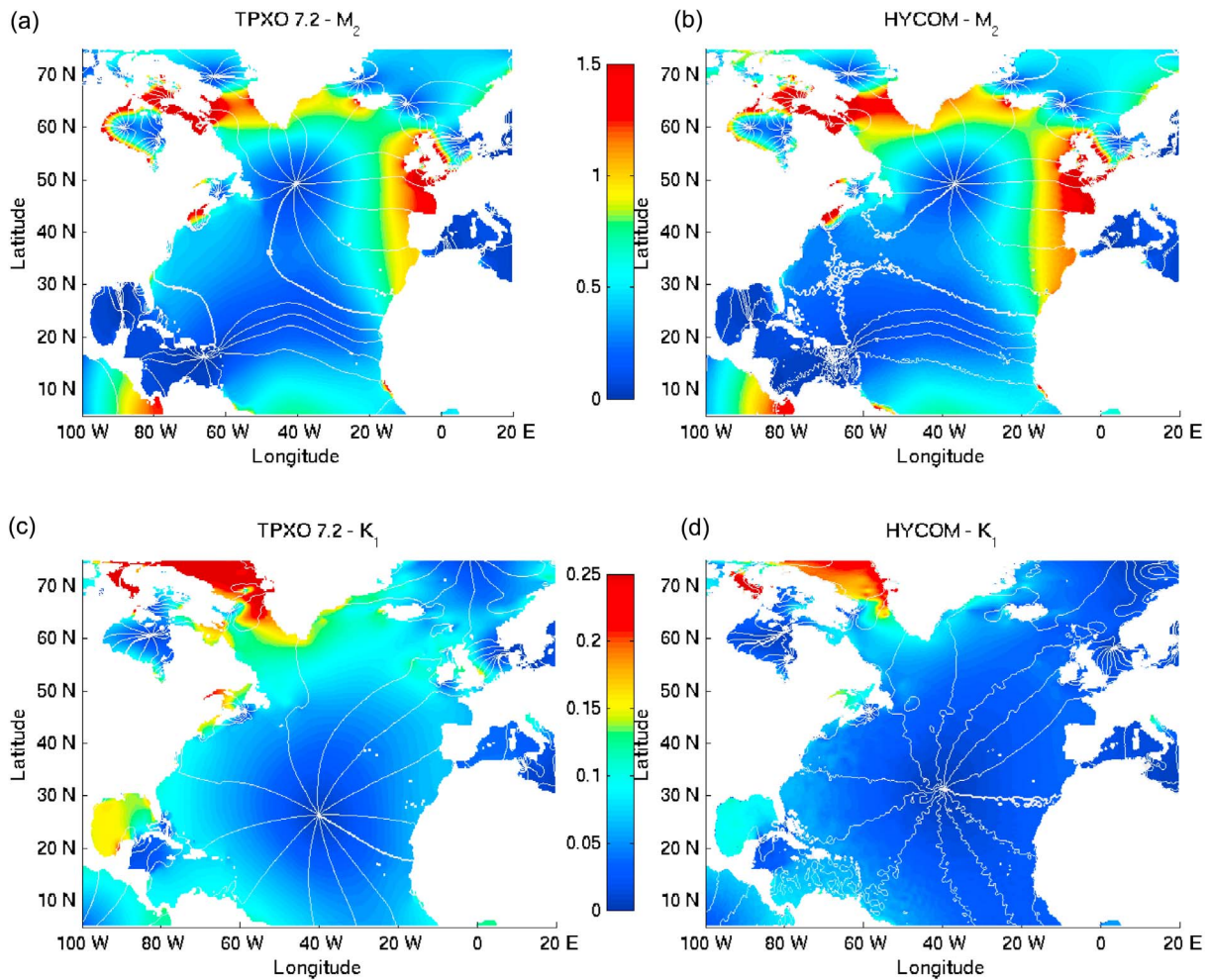
<sup>3</sup>Now at Centre for Applied Marine Sciences, Marine Science Laboratories, Menai Bridge, UK.

<sup>4</sup>Oceanography Division, Naval Research Laboratory, Stennis Space Center, Mississippi, USA.

<sup>5</sup>Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.

<sup>6</sup>Département de Physique and LPO, Université de Bretagne Occidentale, CNRS, Brest, France.

Corresponding author: P. G. Timko, Centre for Applied Marine Sciences, Marine Science Laboratories, Menai Bridge, Anglesey LL59 5AB, UK. (p.timko@bangor.ac.uk)



**Figure 1.** Comparison of sea surface amplitude (m) and phase for constituents (a and b)  $M_2$  and (c and d)  $K_1$  in HYCOM (Figures 1b and 1d) to TPXO 7.2 (Figures 1a and 1c). Contours for the Greenwich phase are drawn at intervals of  $30^\circ$ . HYCOM values used in the figure are interpolated from the model to the TPXO grid. HYCOM values are estimated from September 2004 model output.

forced by climatology and then proceeds for 6 years with inter-annual forcing after the spin up period. Meteorological forcing was Fleet Numerical Meteorological and Oceanography Center (FNMOC) 3-hourly  $0.5^\circ$  Navy Operational Global Atmospheric Prediction System (NOGAPS) [Rosmond *et al.*, 2002]; wind speeds were scaled to be consistent with QuickSCAT observations. Additional details of global HYCOM are described by Metzger *et al.* [2010].

[4] In addition to meteorological forcing at the surface, the model is forced by the semi-diurnal tidal constituents  $M_2$ ,  $S_2$ ,  $N_2$ ,  $K_2$  and diurnal constituents  $K_1$ ,  $O_1$ ,  $Q_1$  and  $P_1$  using the tidal potential. The astronomical tidal potential is corrected for the effects of solid earth body tides [Hendershott, 1972]. As in Arbic *et al.* [2010], the effects of self-attraction and loading (SAL) [Hendershott, 1972; Ray, 1998] are with the “scalar approximation” [Ray, 1998], in which the SAL is taken to be a constant times the sea surface height. The topographic wave drag scheme used within the model is tuned to minimize  $M_2$  surface elevation errors and captures 93% of the surface elevation variance at a standard set of 102 pelagic tide gauges [Shum *et al.*, 1997]

distributed throughout the global ocean [Arbic *et al.*, 2010, and references therein]. Baroclinic tidal elevations were validated in a comparison of high-passed HYCOM  $M_2$  elevations in a  $50^\circ \times 35^\circ$  box to those in altimetry data [Ray and Mitchum, 1996, 1997]. The HYCOM amplitude of baroclinic tides were comparable to (actually slightly larger than) those in the altimeter, confirming that the model is able to resolve baroclinic tides to some level of accuracy (see Figures 7–9 and related text in Arbic *et al.* [2010]).

[5] Figure 1 shows a comparison of amplitude and phase in the North Atlantic for constituents  $M_2$  and  $K_1$  in HYCOM to those of the altimetry based model TPXO 7.2 [Egbert *et al.*, 1994; Egbert and Erofeeva, 2002]. Elevation errors of data assimilative models such as TPXO when compared to the set of 102 pelagic tide gauges are of order 1–1.5 cm for  $M_2$  [Shum *et al.*, 1997] and tend to be much smaller than the errors associated with non-data assimilative tide models such as HYCOM. With the HYCOM wave drag scheme tuned to the  $M_2$  constituent, the agreement between HYCOM and TPXO 7.2 is very good. To measure the accuracy of the sea surface heights we interpolate the HYCOM output to the

**Table 1.** The Sea-Surface Height Discrepancy, Signal Strength, Percent Sea-Surface Height Variance Captured, Integrated Barotropic Kinetic Energy for HYCOM and TPXO, and the Barotropic Kinetic Energy Ratio<sup>a</sup>

Constituent	D (cm)	S (cm)	VC (%)	KE <sub>HYCOM</sub> (10 <sup>15</sup> Joules)	KE <sub>TPXO</sub> (10 <sup>15</sup> Joules)	KE Ratio (%)
M <sub>2</sub>	18.3	44.3	82.8	46.975	44.713	105.1
S <sub>2</sub>	2.6	7.7	88.6	5.077	4.830	105.1
N <sub>2</sub>	1.4	3.4	82.7	1.927	2.127	90.6
K <sub>1</sub>	2.8	5.8	76.9	0.704	1.483	47.5
O <sub>1</sub>	2.7	4.0	54.9	0.625	0.774	80.7
Q <sub>1</sub>	0.17	0.27	58.2	0.070	0.090	77.3

<sup>a</sup>D, sea-surface height discrepancy; S, signal strength; VC, sea-surface height variance captured; KE<sub>HYCOM</sub>, integrated barotropic kinetic energy for HYCOM; KE<sub>TPXO</sub>, integrated barotropic kinetic energy for TPXO; 100\*KE<sub>HYCOM</sub>/KE<sub>TPXO</sub>, barotropic kinetic energy ratio. Values in table are for the North Atlantic and HYCOM values are based upon the September 2004 model output.

TPXO grid and use the same diagnostic as *Arbic et al.* [2004] for the sea-surface height discrepancy, D,

$$D = \sqrt{\frac{\langle \iint (\eta_{HYCOM} - \eta_{TPXO})^2 dA \rangle}{\iint dA}}, \quad (1)$$

where  $\eta_{HYCOM}$  and  $\eta_{TPXO}$  are the sea-surface elevations of HYCOM and TPXO 7.2, respectively and for each constituent, the angle brackets represent the average over one tidal period for that constituent. The percent sea surface height variance captured is given by  $100*[1 - (D/S)^2]$  where  $S$  represents the TPXO signal strength

$$S = \sqrt{\frac{\langle \iint (\eta_{TPXO})^2 dA \rangle}{\iint dA}}. \quad (2)$$

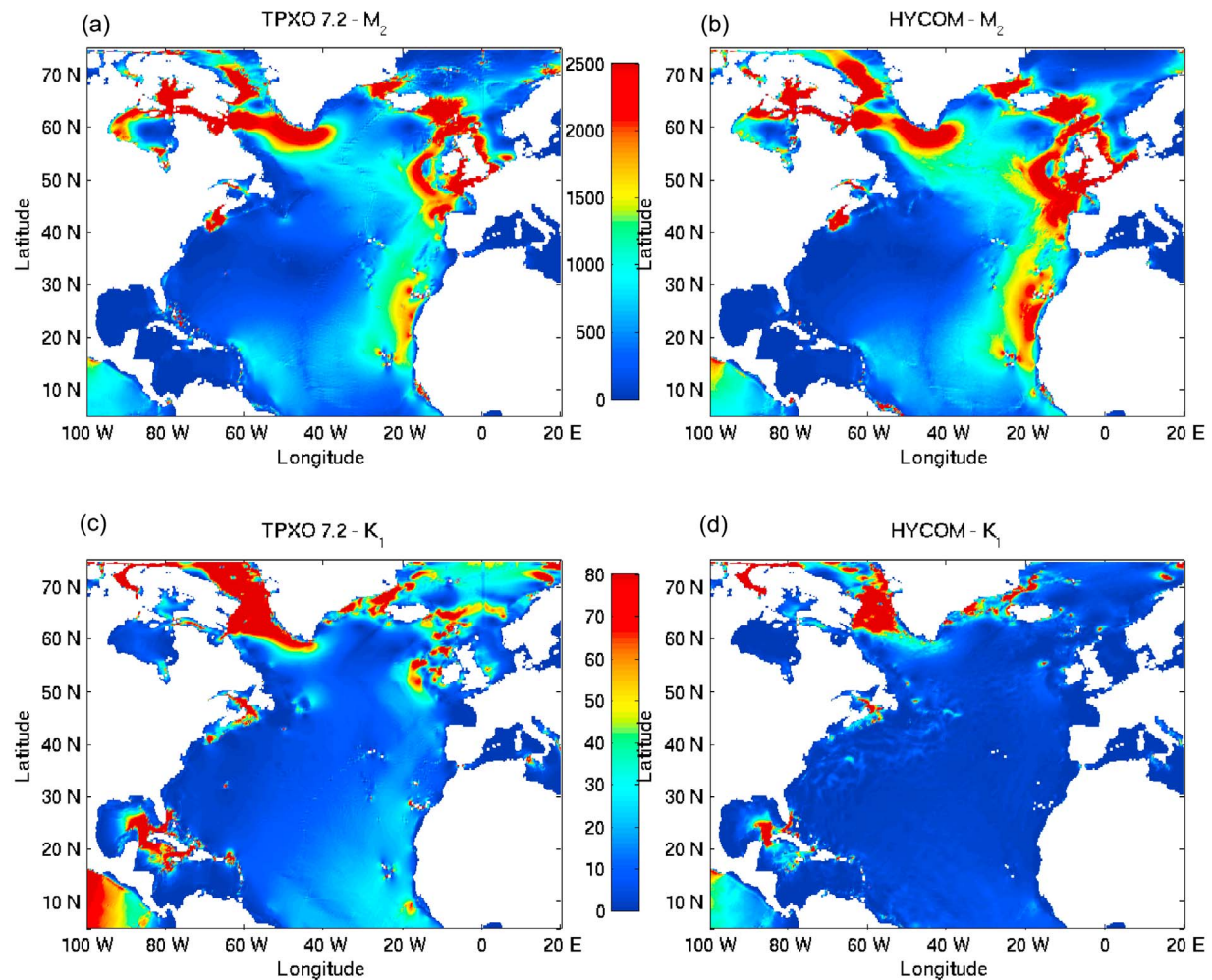
Compared to the 102 pelagic tide gauges used for the model run of *Arbic et al.* [2010] the current model run captures 94.5% of the elevation variance for M<sub>2</sub> and 95.2% of the elevation variance for K<sub>1</sub>. Comparison of the global model to only 102 data points, however, does not fully capture the variance of the tidal signal that may be found in a basin such as the North Atlantic. For the region bounded by 80°W, 20°E, 5°N, and 75°N as shown in Figure 1 the percent sea surface height variance captured by HYCOM when compared to TPXO 7.2 is 83% for constituent M<sub>2</sub> and 77% for constituent K<sub>1</sub>. The sea surface height discrepancy, signal strength and percent variance captured for the six tidal constituents M<sub>2</sub>, S<sub>2</sub>, N<sub>2</sub>, K<sub>1</sub>, O<sub>1</sub>, and Q<sub>1</sub> for the region depicted in Figure 1 are provided in Table 1.

[6] While surface elevations are well represented in the model it is not immediately obvious that tidal currents are likewise well represented. Figure 2 shows a comparison of barotropic tidal kinetic energy per unit area,  $E = \rho_0 H(a^2 + b^2)/4$ , (where  $\rho_0 = 1035 \text{ kg m}^{-3}$ ,  $H$  is the depth of the water column, and  $a$  and  $b$  are respectively the semi-major and semi-minor axis of the tidal ellipse) in HYCOM compared to TPXO for constituents M<sub>2</sub> and K<sub>1</sub>. For M<sub>2</sub> the barotropic kinetic energy in HYCOM appears slightly stronger than in TPXO along the coastlines of Europe and Africa and also over the mid-Atlantic Ridge at midlatitudes. Overall, however, HYCOM provides a good representation of the barotropic kinetic energy field when compared to TPXO. HYCOM clearly replicates the higher barotropic tidal kinetic energy in the eastern North Atlantic as well as the higher

energy in the higher latitudes. For K<sub>1</sub> the barotropic tidal kinetic energy in HYCOM is significantly weaker than TPXO. The weak barotropic K<sub>1</sub> kinetic energy is most noticeable in the Labrador Sea, the Icelandic basin and the Gulf of Mexico. Integrating over the area shown in Figure 2 we find that total M<sub>2</sub> barotropic kinetic energy is  $4.698 \times 10^{16}$  Joules in HYCOM compared to  $4.471 \times 10^{16}$  Joules in TPXO. The energy ratio  $100*KE_{HYCOM}/KE_{TPXO}$  for M<sub>2</sub> is 105.1% compared to an energy ratio of 47.5% for K<sub>1</sub>. The barotropic kinetic energy ratio for the other constituents ranges between 77% and 105%, with the details included in Table 1. The cause of the weakness in K<sub>1</sub> is currently a subject of investigation and will not be discussed in the work presented here. The HYCOM values used to estimate the tidal kinetic energy are based upon model output from September. Some of the weakness in K<sub>1</sub> may be attributed to the destructive interference from constituent P<sub>1</sub> which exists in the model forcing but is not resolved during the tidal analysis of only 30 days of model output and produces a minimum value for K<sub>1</sub> during September.

[7] In addition to the astronomical forcing, differences in stratification, meteorological effects and bathymetry may affect both the magnitude and direction of the barotropic tidal currents producing baroclinic tides. The vertical profile of the tidal currents between the sea surface and ocean floor cannot be measured by satellites and we must rely on moored or moving instruments that profile the water column if we wish to gain knowledge of the vertical structure. Typically such measurements are limited in both their spatial and temporal scales, being confined by the interest of those undertaking the task of observation.

[8] The ocean current observations used to assess model performance are a subset of the Current Meter Archive (CMA) previously used by *Scott et al.* [2010] to compare low frequency ocean currents in global ocean models to observations. The CMA has been expanded since the description provided in the citation. In addition to data sourced from the OSU Buoy Group (<http://kepler.oce.orst.edu>), *Wunsch* [1997], and the Woods Hole Oceanographic Institution Upper Ocean Processes Group (<http://uop.whoi.edu/index.html>), it also now contains data sourced from the British Oceanographic Data Centre (<http://www.bodc.ac.uk>). Within this archive exist approximately 9000 unique time series covering a period of approximately 40 years. The individual records vary in length from 10 days to years and observation intervals range from 5 min to several hours. We have chosen to be inclusive in our selection of current meter records by imposing minimal constraints on the data we use for comparison. The observations used in our analysis are



**Figure 2.** Comparison of barotropic tidal kinetic energy ( $\text{J m}^{-2}$ ) for constituents (a and b)  $M_2$  and (c and d)  $K_1$  in HYCOM (Figures 2b and 2d) to TPXO 7.2 (Figures 2a and 2c). HYCOM values used in the figure are interpolated from the model to the TPXO grid. Model values are based upon September 2004 model output.

taken from past experiments with little attention paid to the objectives of the original observer. More often than not, ocean observations are conducted to study the circulation along a particular coastline or near one or more significant bathymetric features such as seamounts, canyons and ridges. The observations, in some ways, represent a biased view of the ocean.

[9] It is possible to estimate the tides by harmonic analysis. For a scalar quantity the analysis results in an amplitude and phase that describe the strength and the timing of the tidal response at any point to the various tidal forcing frequencies. For a vector quantity such as tidal currents the results may be described by an ellipse. The current ellipse is defined by 5 parameters: (1) magnitude of the semi-major axis; (2) magnitude of the semi-minor axis or eccentricity; (3) inclination or the angle between the semi-major axis and a fixed reference axis (typically the east-west axis); (4) Greenwich Phase or timing of the flow along the semi-major axis (referenced to Greenwich); and (5) direction of rotation (whether the vector that describes the current rotates in a clockwise or counter-clockwise direction). A more detailed description of harmonic tidal analysis and

the description of tidal currents as ellipses may be found in *Godin [1972]*, *Foreman [2004]*, or one of several other references on the subject of tidal analysis.

[10] Changes in local stratification produce a natural variability in observed tidal currents. Most of this variability is found in surface regions and in shallow water where changes in solar radiation, fresh-water inflow and meteorological forces may significantly alter local conditions. The process of harmonic analysis produces only one estimate for each constituent of a given time series. For a long time series, the estimate provided represents the average value of that constituent over the length of the signal and does not reveal the natural variability that may or may not exist at that location. The variability may be estimated and expressed in terms of confidence intervals as part of the harmonic tidal analysis procedure.

[11] We compare model output to observations that are not coincident in time under assumptions that extreme changes in stratification, extreme meteorological events and catastrophic changes in bathymetry are rare and represent outliers. Our primary purpose in this paper is to present methods for assessing model skill in reproducing the observed three-

dimensional tidal structure within the ocean. Tidal algorithms in global HYCOM are still under development and for this reason we limit our interpretation of tidal currents within the model and focus our discussion on model-observation comparisons which can be used to highlight regions where model skill is relatively high and also where model skill is relatively low. The results presented here provide a quantitative measurement of model skill in terms of the likelihood that HYCOM is able to predict tidal currents. This information may be used to assess how well HYCOM is presently able to predict tidal currents and also may be used to guide future investigations into the cause of model-observation mismatches. While observations exist for the global ocean we have chosen to also limit our discussion to the North Atlantic, which has the greatest density of observations. A subsequent paper currently in preparation will present model-observation comparisons on a global/ocean basin scale.

[12] In section 2 we describe the current meter archive, the basis for selecting which records to use for comparison to the model, and the processing of the observations and model output for development of the statistical framework used in the skill tests. Section 3 describes skill tests used to assess model performance. In Section 4, the skill tests are applied to HYCOM within the North Atlantic basin. Section 5 summarizes the results of the skill tests and discusses how the skill tests presented in this paper should be interpreted.

## 2. Pre-Processing of Model, Observations, and Tidal Analysis

[13] Global HYCOM output at hourly intervals was limited to 30 consecutive days (September 2004) due to data storage constraints. The three-dimensional global model output contains both scalar and vector quantities requiring approximately 8 terabytes of disk space. From 30 days of model output it is possible to resolve tidal constituents  $M_2$ ,  $S_2$ ,  $N_2$ ,  $K_1$ ,  $O_1$  and  $Q_1$  using harmonic tidal analysis. Tidal forcing frequencies  $K_2$  and  $P_1$  require a data length of 183 days to resolve due to the Rayleigh Criterion that the differences between resolved frequencies must be greater than the inverse of the data length. That is to say a signal of length  $N$  hours is required to separate signals of frequency  $\omega_1$  and  $\omega_2$  which must satisfy  $|\omega_1 - \omega_2| < R/N$ . The Rayleigh Criterion constant,  $R$ , is typically set equal to one but may be set to other values.

[14] Velocity data from the 32 model layers that represent the vertical coordinate is interpolated to a 44 z-level coordinate centered on the pressure points of the C-grid. The z levels are not uniformly distributed. Resolution in z-space varies from 5 m in the top 50 m to 500 m below 2000 m depth. Since HYCOM uses an isopycnal coordinate system in deep water there is little, if any, benefit to resolving deep layers at higher vertical resolution than the number of vertical layers used within the model itself.

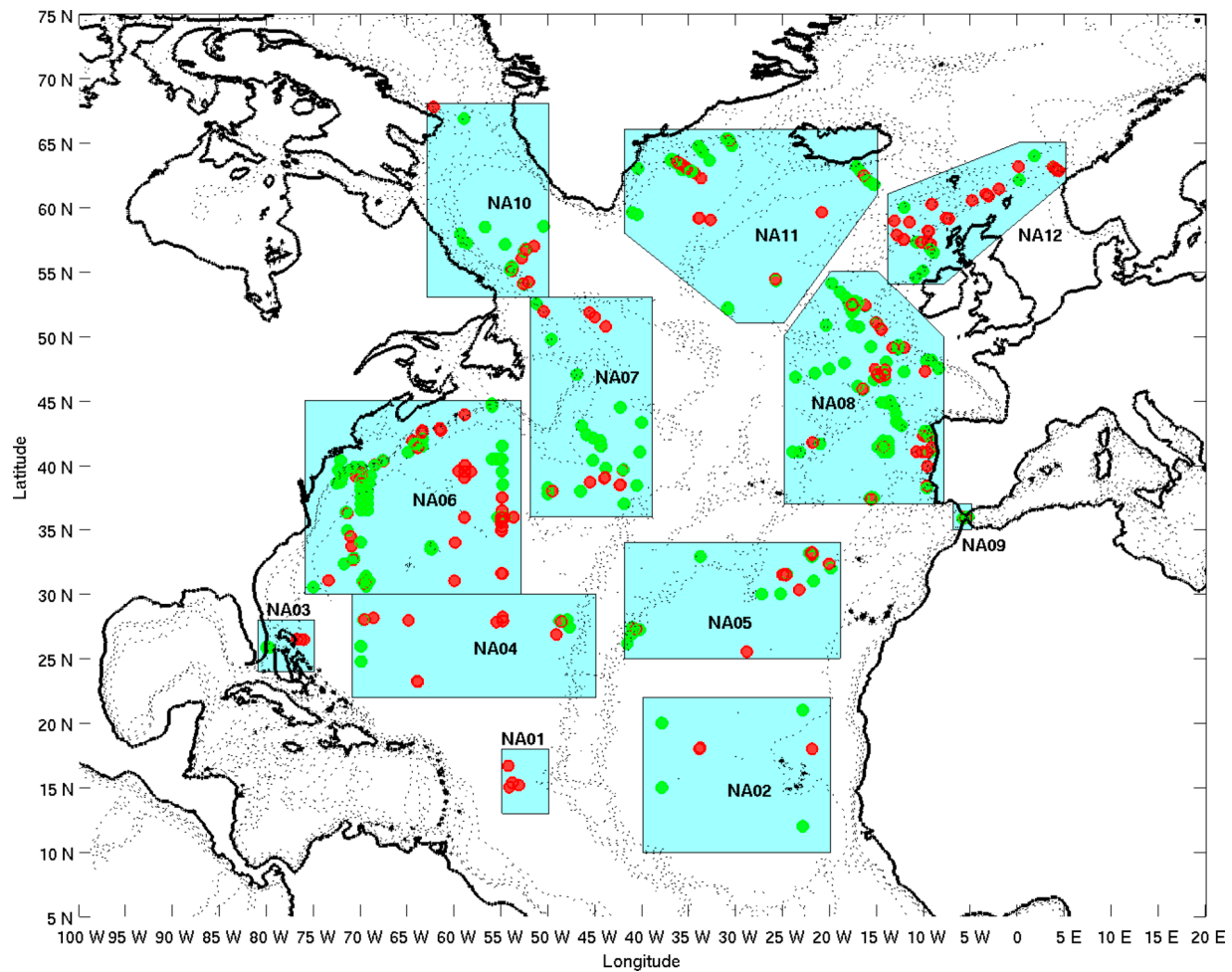
[15] Harmonic tidal analysis [Foreman, 2004] is applied to the interpolated velocity records to produce an estimate of the tidal current ellipses over the entire model domain. For any mooring at a given latitude and longitude the model grid may be searched to locate the nearest model grid pressure point (nearest neighbor). In addition to the global tidal analysis, for each observation (at a given depth) the velocity field of the model nearest neighbor on the z-level coordinate system is

linearly interpolated to observation depth. The velocity field for the 8 model grid pressure points surrounding the nearest neighbor is also linearly interpolated to observation depth. This interpolation is performed prior to harmonic analysis of the model velocity records used in this study.

[16] Within the CMA we identified 1871 velocity records from the North Atlantic (NA) that satisfied the following criteria: (1) Time intervals between recorded values must not exceed one hour. (2) Records must contain a minimum of 105 days of observations. (3) Discontinuous records must provide a minimum of 105 days of data. Eleven records that failed to meet the record length criteria were edited by linearly interpolating a few isolated missing values to meet one of the two record length criteria. In total 1882 unique time series were identified. The velocity records (also referred to as instruments) are distributed across 653 moorings. Only 276 of the moorings have velocity records at 3 or more distinct depths. The number of records with a minimum of 4383 h of observations is 1370. For these 1370 records it is possible to obtain estimates of all eight tidal constituents used to force HYCOM. Of the 1882 individual velocity records identified within the NA there are 999 records distributed across 399 moorings (152 moorings have 3 or more velocity records) that have continuous 30 day records for the month of September. The 1882 velocity records were grouped into 12 geographic regions. The regions (Figure 3 and Table 2) represent: tropical NA (Regions 1 and 2); the Florida Current system (Region 3); subtropical NA (Regions 4 and 5); Gulf Stream and North Atlantic Current system (Regions 6 and 7); eastern midlatitude NA (Region 8); Strait of Gibraltar (Region 9); Labrador Sea (Region 10); central and high-latitude eastern NA (Regions 11 and 12). Note that in some cases, the individual regions contain only a limited number of September observations (Table 3).

[17] The Matlab routine `t_tide` [Pawlowicz *et al.*, 2002] was used for analyzing the observations and their corresponding model counterparts. This routine is based upon the Foreman [2004] FORTRAN routine used for the analysis of the global model data on the 44 z-levels. As with the Foreman FORTRAN code, the direction of rotation is indicated by the sign of the semi-minor axis. A positive value for the semi-minor axis indicates counter-clockwise rotation; a negative sign indicates clockwise rotation. In addition to providing estimates of the ellipse parameters, error estimates of the ellipse parameters using 95% bootstrapped confidence intervals based upon an uncorrelated colored-noise model [Pawlowicz *et al.*, 2002] are also output as part of the analysis.

[18] The signal-to-noise ratio of a tidal constituent estimated from the semi-major axis is  $SNR = A_{maj}^2/E_{maj}^2$ ; where  $A_{maj}$  and  $E_{maj}$  are the estimated semi-major axis and error of the semi-major axis, respectively. When  $SNR < 1$  the signal is too weak to provide a meaningful estimate of that constituent and confidence intervals for such constituents tend to be overly large. In order to minimize the number of false positive outcomes any constituent for which  $SNR < 1$  is not used to determine if the model values lie within the 95% confidence for that constituent. The alternative would be to attempt to remove a low-passed signal from the current meter records to remove secular trends and low frequency signals in order to increase the value of SNR. However, it is not obvious that such a process would produce useful



**Figure 3.** Location of the 12 regions within the North Atlantic along with locations of the observations. Red circles identify mooring locations with 3 or more instruments (velocity records) in the vertical. Green circles identify mooring locations with only 1 or 2 velocity records.

information regarding the weak constituents (R. Pawlowicz, personal communication, 2012). For a current meter record in which some constituents satisfy  $SNR < 1$  while others satisfy  $SNR > 1$  we only discard the constituents for which  $SNR < 1$  and retain the other constituents for comparison to model values. Table 4 provides an accounting of the number of current meter records used for comparison to the model

values for each constituent within each ocean region depicted in Figure 3 and referenced by name in Table 2. The total number of observations used for comparison in basin wide comparisons is the sum of the number of observations in each region.

[19] Another issue that needs to be considered is that in addition to the six tidal constituents  $M_2, S_2, N_2, K_1, O_1,$  and

**Table 2.** A List of the Defining Latitude and Longitude Boxes for the 12 Regions Defined in the North Atlantic<sup>a</sup>

Region	Latitude		Longitude		Region
	Min	Max	Min	Max	
NA01	13	18	305	310	western tropical North Atlantic
NA02	10	22	320	340	eastern tropical North Atlantic
NA03	24	28	279	285	Florida Current system
NA04	22	30	289	315	western subtropical North Atlantic
NA05	25	34	318	341	eastern subtropical North Atlantic
NA06	30	45	284	307	Gulf Stream system
NA07	36	53	308	321	North Atlantic Current
NA08	37	55	335	352	eastern midlatitude North Atlantic
NA09	35	37	353	355	Strait of Gibraltar
NA10	53	68	297	310	Labrador Sea
NA11	51	66	318	345	central high-latitude North Atlantic (Icelandic and Irminger Basins)
NA12	54	65	346	365	eastern high-latitude North Atlantic (Rockall Plateau to Norwegian Sea)

<sup>a</sup>See Figure 3. Also included for reference is a list of seas, large bathymetric features, and/or geographic region.

**Table 3.** Number of Instruments, Moorings, and Moorings With 3 or More Velocity Records (Instruments) for Each of the 12 Ocean Regions Shown in Figure 3 and Described in Table 2<sup>a</sup>

Region	Instruments		Moorings		Moorings With 3+ Instruments	
	Sep	Total	Sep	Total	Sep	Total
NA01	17	18	7	7	4	4
NA02	13	36	7	13	1	5
NA03	1	19	1	7	0	3
NA04	48	65	20	26	9	12
NA05	90	167	28	50	15	27
NA06	300	623	107	188	46	84
NA07	18	91	13	40	0	13
NA08	184	271	80	114	22	36
NA09	27	114	4	11	1	6
NA10	94	118	41	54	14	16
NA11	139	199	65	84	25	31
NA12	68	161	26	59	15	39

<sup>a</sup>Values show records containing September observations (Sep) and the total number of observations for the region that are available.

$Q_1$  that may be estimated from 30 days of model output, the model is also forced by tidal constituents  $K_2$  and  $P_1$ . Analysis of velocity records shorter than 4383 h ( $\sim 183$  days) may not be able to distinguish between the constituent pairs ( $S_2, K_2$ ) and ( $K_1, P_1$ ) because the pairs are too close together in frequency (Rayleigh Criterion). If one year of data is separated into 30-day windows which are individually subjected to harmonic analysis we find that tidal constituents  $S_2$  and  $K_1$  are modulated by constituents  $K_2$  and  $P_1$ , respectively, and this modulation appears as a six-month beating of the estimates of  $S_2$  and  $K_1$  with  $S_2$  ( $K_1$ ) having maximal (minimal) values for the months of March and September. Foreman and Henry [1989] suggest that  $K_2$  and  $P_1$  may be inferred from  $S_2$  and  $K_1$  when models are forced by tides only. The model run used in this paper is not forced by tides only. Using one year of model output for the surface velocity field we attempted to infer  $K_2$  and  $P_1$  on the model surface velocity field using inference parameters based upon TPXO 7.2, 1 year of the model barotropic velocity field, and also using the tidal potential. Our attempts to use inference produced inconsistent results when analyzing 30-day windows of the surface velocity field compared to the yearlong tidal analysis of the model surface velocity field. At some locations our attempts to apply inference improved the 30-day estimate, in other locations the estimates of the constituents  $K_1$  and  $S_2$  had larger errors (compared to the estimates from the yearlong signal) or remained approximately the same when inference was applied. Our examination of the observations available for comparison to the model indicate that the change in the magnitude of the semi-major axis values for  $S_2$  and  $K_1$  due to modulation by  $K_2$  and  $P_1$  may be as large as or larger than the magnitude of the uncertainty associated with the estimates of constituents  $S_2$  and  $K_1$ . For this reason we have chosen to use only the September observations when testing if the model values lie within the confidence intervals of the observations.

### 3. Model Skill Tests

[20] For all of our skill tests presented below we test the model *nearest neighbor* (model grid point closest to the recorded mooring location) and the 8 grid points

surrounding the nearest neighbor on the model grid. The group of 9 grid points associated with an instrument (mooring) location will be referred to as the *instrument (mooring) neighborhood*. In coastal regions an instrument or mooring neighborhood may contain land points reducing the number of model grid cells for comparison to the observations. A grid point in the instrument/mooring neighborhood that is not the nearest neighbor will be referred to as a *neighbor*. Our reference to an *instrument* refers to an individual velocity record at a single depth. In some cases the term instrument refers to an individual bin from an ADCP record. By comparing to the nearest neighbor and surrounding cells we are able to allow for uncertainties that may exist in the grid representation within the model and also for small uncertainties that may even exist in recorded mooring locations. The difference in distance between the mooring location and the nearest neighbor on the model grid and the next closest model neighbor is typically less than 1–2 km. When testing within an instrument neighborhood we consider the test successful if at least one neighbor in the neighborhood is able to satisfy the 95% confidence intervals. If multiple neighbors satisfy the confidence intervals the neighborhood test is successful but still only counts as one successful outcome.

[21] Our first skill test investigates the number of model values satisfying 95% bootstrapped confidence intervals for each of the ellipse parameters. Confidence intervals used for the semi-major and semi-minor axes, ellipse inclination and Greenwich Phase are those generated by *t\_tide* [Pawlowicz *et al.*, 2002]. This skill test is based upon the observations from September only.

[22] Additional skill tests based upon the root mean square error and correlation between model and observations are also presented. These tests are based upon the set of all 1882 observations available for comparison and allow us to assess model performance through the water column at a given mooring location and along horizontal planes in the different ocean regions irrespective of the strength of the signal or its SNR.

[23] For our first test we identify the model values of the individual ellipse parameters that lie within their respective confidence intervals in order to form subsets of the

**Table 4.** The Number of Instruments (September Observations) Within Each Region for Which the Error Associated With the Semi-major Axis,  $E_{maj}$ , is Less Than the Magnitude of the Semi-major Axis,  $A_{maj}$ <sup>a</sup>

Region	Number of Instruments With $E_{maj}/A_{maj} < 1$						Total
	M2	S2	N2	K1	O1	Q1	
NA01	17	17	17	3	2	1	17
NA02	13	13	13	6	7	8	13
NA03	1	0	0	1	1	1	1
NA04	48	48	48	48	46	47	48
NA05	90	89	86	88	87	85	90
NA06	296	258	270	239	237	209	300
NA07	18	18	18	17	18	12	18
NA08	184	182	182	143	133	114	184
NA09	27	27	27	27	27	27	27
NA10	94	91	86	66	61	40	94
NA11	117	84	63	50	56	54	139
NA12	68	64	62	55	57	47	68

<sup>a</sup>Also included is the total number of instruments within each region.

September observations. We intersect these subsets to determine those locations in the model where two or more model ellipse parameters are within their 95% bootstrapped confidence intervals. This is not to say that the model values satisfy the joint probability distribution of the multiple ellipse parameters but simply identifies those locations where the model values of two or more ellipse parameters simultaneously lie within their individual 95% confidence intervals. The joint 95% confidence region for a complex signal,  $Ae^{-i\theta}$ , is determined by an F-distribution [Foreman and Henry, 1989]. By intersecting the results of the individual ellipse parameters the overall level of significance may be lower than 95%. We do not attempt to address the issue of the family wise error rate that may exist as a result of this procedure. The first intersection we consider identifies those locations where the model values are able to satisfy the confidence intervals of both the semi-major and semi-minor axes. We continue by intersecting the subset resulting from this first intersection with those locations where the model ellipse inclination and then the Greenwich phase also fall within their corresponding 95% confidence intervals. Tidal kinetic energy is proportional to the sum of the squares of the semi-major and semi-minor axis. This hierarchy first determines those locations in the model where the tidal kinetic energy is expected to match the observations. Tidal kinetic energy is significant as it drives mixing within the ocean and if energy levels are not correct the influence of the tides on local stratification cannot be expected to be correct. Second it determines those locations where the direction and timing of maximum tidal flow is within the expected range of values based upon available observations.

[24] The assessment of model skill using the above method highlights those locations where model skill is able to satisfy 95% confidence intervals at a given location. It does not provide a means to assess model skill where the model does not satisfy the specified criteria. Another approach in assessing model skill is to examine the root mean square error (RMSE) between the model and observations. Such an approach maximizes the amount of information that can be obtained based upon the available observations and model output. When a minimum of 3 observations are grouped together we may also calculate the correlation between the model and observations. Our methodology in assessing model skill using the RMSE and correlation are outlined below.

[25] At each mooring we calculate the mooring root mean square error

$$RMS_{moor} = \sqrt{\frac{1}{K} \sum_{k=1}^K [A_{obs}(k) - A_{mod}(k)]^2}, \quad (3)$$

where  $\{A_{obs}(k)\}_{k=1}^{k=K}$  are the  $K$  semi-major axis values estimated from the  $K$  instruments on a given mooring and  $\{A_{mod}(k)\}_{k=1}^{k=K}$  represents the corresponding model semi-major axis value for a given model neighbor.  $RMS_{moor}$  is representative of the difference between model and observations for all instruments located on a single entire mooring. This measure provides an estimate of how well the model fits the observations through the water column at a mooring location. Within each mooring neighborhood we

define the *mooring best fit location* as that neighbor with the lowest  $RMS_{moor}$  score.

[26] A high value of  $RMS_{moor}$  may occur if there is a mismatch in model bathymetry and true bathymetry near such features as seamounts, channels or coastal features. A high value of  $RMS_{moor}$  may also occur if the model stratification does not match the stratification at the time the observations were taken. This may occur in cases of mooring blow-over in the presence of strong tidal currents or near the surface where the mixed layer in the model may lie above or below the mixed layer at the time of observation.

[27] Restricting our attention to moorings with at least 3 instruments at distinct depths we may also calculate the correlation skill score [von Storch and Zwiers, 2003] between the model and the observations at each mooring. We define the mooring correlation skill score

$$CSS_{moor} = \frac{\sum_{k=1}^{k=K} (x_k^{mod} - \bar{x}^{mod})(y_k^{obs} - \bar{y}^{obs})}{\sqrt{\sum_{k=1}^{k=K} (x_k^{mod} - \bar{x}^{mod})^2 \sum_{k=1}^{k=K} (y_k^{obs} - \bar{y}^{obs})^2}}, \quad (4)$$

where  $x_k^{mod}$ ,  $y_k^{obs}$  represent the ranks of the paired  $K$  model and observed semi-major axis values  $\{A_{mod}(k), A_{obs}(k)\}_{k=1}^K$  for a given mooring, respectively, and  $\bar{x}^{mod}$ ,  $\bar{y}^{obs}$  are the means of the ranks. Our definition of the correlation skill score,  $CSS_{moor}$ , uses the nonparametric Spearman rank correlation coefficient as opposed to the Pearson correlation coefficient defined by von Storch and Zwiers [2003]. The range of possible scores is between  $-1$  and  $1$ . A positive score indicates that the model is related to the observations by a monotonic increasing function. A negative score indicates that the model is related to the observations by a monotonic decreasing function.

[28] When the skill scores  $RMS_{moor}$  and  $CSS_{moor}$  are combined at a single mooring we are able to determine those locations where the model values closely match the observations and also determine how well the velocity shear in the model matches the shear in the observations. If  $RMS_{moor}$  is small and  $CSS_{moor}$  is high then we can be assured of a good match between the model and observations up to the level of significance for  $CSS_{moor}$ . If  $CSS_{moor}$  does not meet a desired level of significance then model tidal currents may not reflect the correct amount of velocity shear in the water column.

[29] To test model skill in horizontal planes defined by vertical depth ranges we can also form subsets with the set of all instruments by grouping instruments within geographic region and specified depth range. We define a regional root mean square error

$$RMS_{reg} = \sqrt{\frac{1}{I_{reg}(D)} \sum_{i=1}^{i=I_{reg}(D)} [A_{obs}(i) - A_{mod}(i)]^2}, \quad (5)$$

where the set  $\{A_{obs}(i)\}_{i=1}^{i=I_{reg}(D)}$  represents  $I_{reg}(D)$  instruments found within a specified depth bin  $D = \{z_0 \leq z \leq z_1\}$  for a predetermined ocean region and  $\{A_{mod}(i)\}_{i=1}^{i=I_{reg}(D)}$  are the corresponding model values. When a depth range contains at



least 3 instruments we also define the regional correlation skill score

$$CSS_{reg} = \frac{\sum_{k=1}^{k=K} (x_k^{mod} - \bar{x}^{mod})(y_k^{obs} - \bar{y}^{obs})}{\sqrt{\sum_{k=1}^{k=K} (x_k^{mod} - \bar{x}^{mod})^2 \sum_{k=1}^{k=K} (y_k^{obs} - \bar{y}^{obs})^2}}, \quad (6)$$

where  $x_k^{mod}$ ,  $y_k^{obs}$  are the ranks of the sets  $\{A_{mod}(i)\}_{i=1}^{i=I_{reg}(D)}$  and  $\{A_{obs}(i)\}_{i=1}^{i=I_{reg}(D)}$  (the sets defined over the same depth bins used to define  $RMS_{reg}$ ), respectively and  $\bar{x}^{mod}$ ,  $\bar{y}^{obs}$  are the means of the ranks.

[30] The regional root mean square allows us to assess model performance in the horizontal plane. It also allows for comparison of data spanning several years to be compared to a single set of model output. The choice of depth bins allows us to assess the overall regional representation of the model in the mixed layer, mid-layers, and abyssal layers.

[31] When evaluating model performance over the depth bins we use both a nearest neighbor and also an instrument best fit approach. The instrument best fit is the neighbor with the smallest absolute difference between the model value and the observation value of the semi-major axis, i.e.,  $inst_{best-fit} = \min_i \{|A_{obs} - A_{mod}(i)|\}_{i=1}^{i=9}$ . The use of an instrument best fit to determine regional scores in a depth bin is justified as the instruments used in the evaluation of  $RMS_{reg}$  are, typically, grouped together from different locations and/or times and there is no reason to assume a relationship between the individual instruments and their location in the instrument neighborhood. Typically when we apply our test to examine model skill over a large geographic region we are mixing observations from several different ocean observing experiments that often span several different years.

## 4. Application of the Skill Tests to 1/12.5° Global HYCOM

### 4.1. Assessing Model Skill Using 95% Confidence Intervals

[32] Bootstrapped confidence intervals are estimated for the semi-major and semi-minor axis, inclination and phase of the tidal ellipses. Figure 4 shows the percent of all September observations in the NA for which the model ellipse parameters, estimated from the three-dimensional velocity field at the model nearest neighbor and within a 9-point instrument neighborhood, lie within the 95% confidence intervals of the observations. Also shown are the percent of all September observations for which the tidal ellipse parameters from TPXO 7.2 at the mooring locations and the HYCOM barotropic mode at the nearest neighbor lie within the 95% confidence intervals. The TPXO 7.2 is data assimilative and therefore is expected to have an advantage over the barotropic HYCOM mode. TPXO 7.2 is gridded at a resolution of 1/4° whereas the horizontal resolution of HYCOM is 1/12.5°. A 9-point model neighborhood centered on the nearest neighbor represents approximately the same area as represented by a single TPXO 7.2 grid cell. Three-dimensional HYCOM fields are able to simulate the baroclinic structure of the ocean and therefore, ideally,

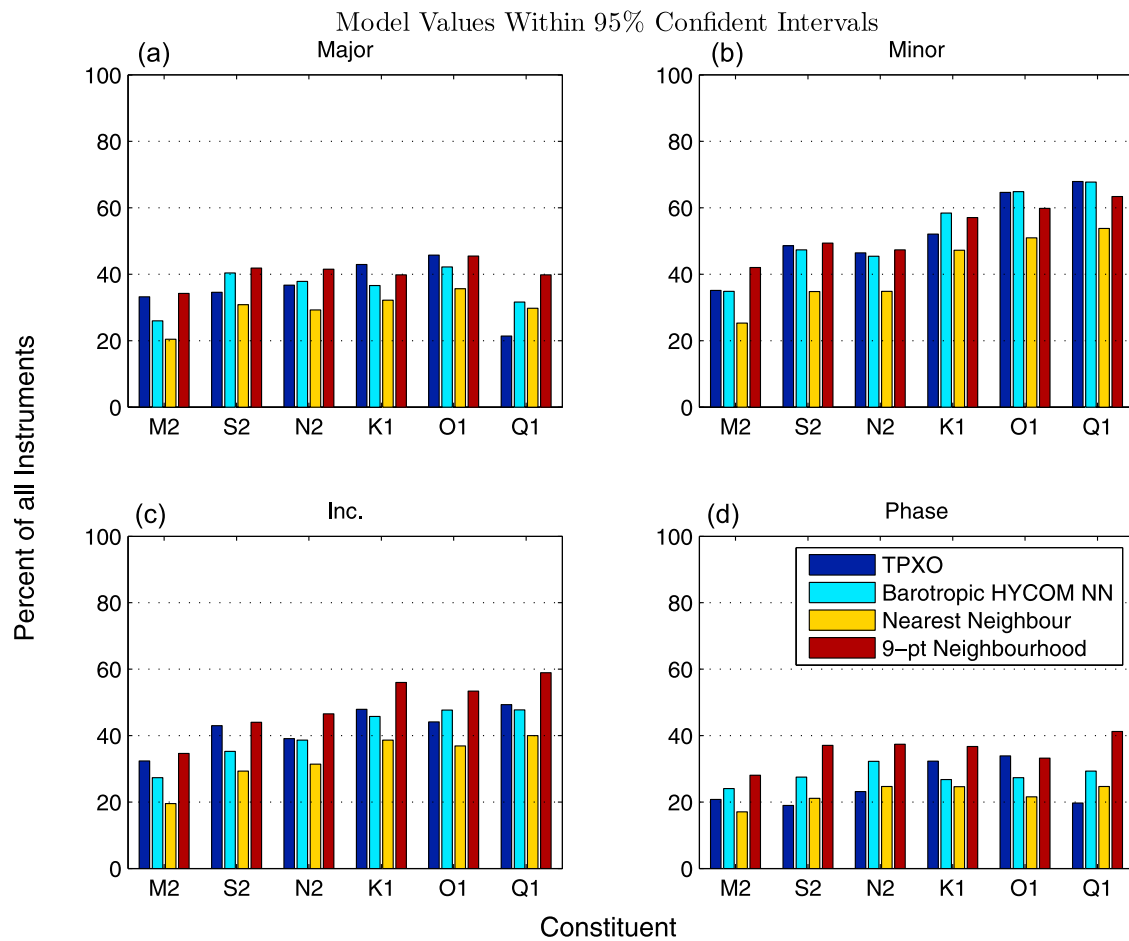
should be able to provide a better estimate of the tidal currents as they vary with depth within a given neighborhood.

[33] The mean and median values (from the observations in the NA) of the error to signal ratio ( $ESR = E_{maj}/A_{maj}$ ) for the semi-major and semi-minor axis along with the mean and median errors of the inclination and Greenwich phase are provided in Table 5. Also included in Table 5 are the mean and median values from the set of all observations in the NA. In Figure 4, the success rates for the semi-minor axis appear higher than those of the semi-major axis and may contain a significant number of false positive outcomes due to the larger mean and median values of the ESR (Table 5). This same type of overestimation of model success may also occur for weaker tidal constituents. When applying harmonic tidal analysis to the full set of observations, the full time series were used regardless of their differing lengths. The large discrepancy between the mean and median ESR values for the semi-minor axis results from approximately 10%–20% of the signals with  $ESR > 5$ . The mean of the absolute value of the semi-minor axis for those instruments with  $ESR > 5$  is less than 0.1 cm s<sup>-1</sup>. If we eliminate those semi-minor axis values for which  $ESR > 5$  we find that the mean and median ESR values lie much closer together.

#### 4.1.1. Basin Scale Analysis

[34] Over the entire NA the percent of model semi-major axis values able to satisfy the 95% confidence intervals (success rate) for the barotropic TPXO 7.2 is (for each constituent): M<sub>2</sub> (33%), S<sub>2</sub> (35%), N<sub>2</sub> (37%), K<sub>1</sub> (43%), O<sub>1</sub> (46%), and Q<sub>1</sub> (21%). For the HYCOM barotropic mode we have success rates of: M<sub>2</sub> (26%), S<sub>2</sub> (40%), N<sub>2</sub> (38%), K<sub>1</sub> (37%), O<sub>1</sub> (42%), and Q<sub>1</sub> (32%). At the model nearest neighbor the success rate for three-dimensional HYCOM is: M<sub>2</sub> (20%), S<sub>2</sub> (31%), N<sub>2</sub> (29%), K<sub>1</sub> (32%), O<sub>1</sub> (36%) and Q<sub>1</sub> (30%). If we extend our comparison to the instrument neighborhoods the success rates increase to: M<sub>2</sub> (34%), S<sub>2</sub> (42%), N<sub>2</sub> (42%), K<sub>1</sub> (40%), O<sub>1</sub> (45%) and Q<sub>1</sub> (40%). For all constituents except Q<sub>1</sub> both the TPXO 7.2 and the barotropic HYCOM mode have higher success rates than those of the three-dimensional HYCOM at the nearest neighbor. The increase of success rates when the model is evaluated within the instrument neighborhoods as opposed to at the nearest neighbor alone suggests that some grid misrepresentation does occur in the model since the baroclinic vertical modes are dependent upon the depth of the water column and the density field. However, a closer inspection of the individual signals from the model would be necessary to determine the actual reason why a model value from a neighbor matches the observed value more closely than the model value from nearest neighbor. It is most likely that the reasons for the mismatch will vary from instrument to instrument.

[35] For the semi-minor axis, the success rate for the semi-diurnal constituents for TPXO 7.2 is: M<sub>2</sub> (35%), S<sub>2</sub> (49%), and N<sub>2</sub> (46%) which is comparable to the success rate for the barotropic HYCOM mode: M<sub>2</sub> (35%), S<sub>2</sub> (47%), and N<sub>2</sub> (45%). For the three-dimensional HYCOM we find success rates of: M<sub>2</sub> (25%), S<sub>2</sub> (35%), and N<sub>2</sub> (35%) at the nearest neighbor and 42%, 49%, and 47%, respectively within the instrument neighborhoods. For the diurnal constituents both the mean and median ESR is greater than one and success rates of 47%–68% and higher for these constituents are expected to be the result of false positive outcomes,



**Figure 4.** The values indicated in the figure represent the percent of model values that lie within the 95% confidence intervals of the tidal ellipse parameters: (a) semi-major axis, (b) semi-minor axis, (c) inclination (ellipse orientation), and (d) Greenwich phase. Dark blue bars represent the results for TPXO 7.2 and light blue bars represent the barotropic HYCOM mode at the nearest neighbor (NN). Yellow bars represent three-dimensional D HYCOM at the nearest-neighbor and red bars represent the three-dimensional HYCOM model values where at least one-neighbor lies within the 95% bootstrapped confidence intervals. The results are shown for the six tidal constituents available from 30-days of model output.

especially when  $ESR > 5$  which happens for 10%–20% of the signals.

[36] For the ellipse inclination success rates for TPXO 7.2 and the barotropic HYCOM mode are 32%–43% and 27%–39%, respectively. The three-dimensional HYCOM has success rates of 20%–31% (34%–46%) at the nearest neighbor (within an instrument neighborhood) for the semi-diurnal constituents. The success rates at the nearest neighbor are lower than the success rates of both the TPXO 7.2 and the barotropic HYCOM mode while the success rates within the 9-point instrument neighborhoods are greater than or equal to TPXO 7.2. The diurnal constituents tend to have higher success rates than the semi-diurnal constituents: TPXO 7.2 (44%–49%), barotropic HYCOM mode (46%–48%), three-dimensional HYCOM nearest neighbor (36%–40%), and instrument neighborhoods (53%–58%). An examination of the inclination error (Table 5) for the diurnal constituents indicates that the error associated with these constituents is relatively large compared to the range of possible values for inclination ( $0^\circ$ – $179^\circ$ ). As with the semi-

**Table 5.** Mean and Median Values of Error to Signal Ratios for the Semi-major/minor Axis and Mean and Median Values for Errors in Tidal Ellipse Inclination and Greenwich Phase for Each Tidal Constituent, Computed From the Complete Set of Observations and the September Observations

	Major		Minor		Inclination		Phase	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
M <sub>2</sub> total	0.14	0.08	1.33	0.36	12	6	13	6
M <sub>2</sub> September	0.18	0.11	5.31	0.49	15	8	15	8
S <sub>2</sub> total	0.30	0.24	3.57	0.83	29	17	29	18
S <sub>2</sub> September	0.29	0.21	3.13	0.78	26	17	26	18
N <sub>2</sub> total	0.31	0.25	5.75	0.90	30	20	30	21
N <sub>2</sub> September	0.32	0.26	3.55	0.80	30	21	31	23
K <sub>1</sub> total	0.45	0.43	6.85	1.61	37	32	38	34
K <sub>1</sub> September	0.42	0.38	9.35	1.22	38	28	38	31
O <sub>1</sub> total	0.53	0.55	13.57	1.69	48	45	49	49
O <sub>1</sub> September	0.48	0.48	8.61	1.62	43	38	43	40
Q <sub>1</sub> total	0.51	0.51	10.36	1.50	51	50	51	53
Q <sub>1</sub> September	0.54	0.55	5.90	1.62	49	43	48	46

minor axis we anticipate that the successful outcomes for the ellipse inclination of the diurnal constituents likely contains a large number of false positive outcomes due to the relatively large errors associated with these constituents. When inclination values lie close to the east-west axis, there is also the possibility of false negatives occurring. This results from the ambiguity in the specification of ellipse inclination. By convention the ellipse inclination is reported as the northward directed semi-major axis. If the observed ellipse inclination is close to  $180^\circ$ , model values for the ellipse inclination may be reported as being close to  $0^\circ$  (when they may be close to  $180^\circ$ ) and therefore fail to lie within the confidence intervals for the observed value. We have corrected for this by adding  $180^\circ$  to the model values when the model ellipse inclination was less than  $10^\circ$  and the observed value was greater than  $170^\circ$ . The same situation may occur if observation and model values are reversed in the above explanation in which case we have corrected by subtracting  $180^\circ$  from the model values. This effect may be significant in places such as the Strait of Gibraltar where the semi-major axis of the tidal ellipse is expected to lie along the east-west direction.

[37] Success rates for the Greenwich phase are: 20%–34% for TPXO 7.2; 24%–32% for the barotropic HYCOM mode; 17%–25% at the nearest neighbor, and 28%–41% in the instrument neighborhoods for three-dimensional HYCOM. Within the instrument neighborhoods  $Q_1$  is the weakest of all constituents being assessed in this test and typically has the largest amount of uncertainty associated with it. The mean and median errors for  $Q_1$  are  $48^\circ$  and  $46^\circ$ , respectively, indicating that for at least 1/2 of the observations the error associated with this constituent covers 1/4 of all possible values ( $0^\circ$ – $359^\circ$ ). The success rates for  $Q_1$  may include several false positives due to the large errors associated with the Greenwich phase. False negatives are also possible for the Greenwich phase which is measured from the angle of inclination of the tidal ellipse. When ellipse inclination values lie close to the east-west axis, the ambiguity in the determining ellipse inclination may result in  $180^\circ$  discontinuities appearing in the reported Greenwich phase values.

[38] From a basin scale analysis such as the one described above there seems little benefit to assessing model performance using any ellipse parameters other than the semi-major axis. Our test results indicate that due to the large amount of uncertainty that may exist for the semi-minor axis, inclination and Greenwich phase of individual constituents, the number of successful outcomes may contain a large number of false positives for these ellipse parameters. It is possible to reduce the number of false positive outcomes by increasing the required SNR. Another option would be to place constraints for each of the ellipse parameters. However, signals with higher SNR also tend to have less uncertainty and hence smaller error for the other ellipse parameters as well. Increasing SNR is the most practical option when trying to reduce the number of false positives, however if the SNR is chosen to be too high, very few signals would qualify for comparison to the model. We have not yet investigated the effect of increasing the SNR used to determine which observations to use for comparison to the model.

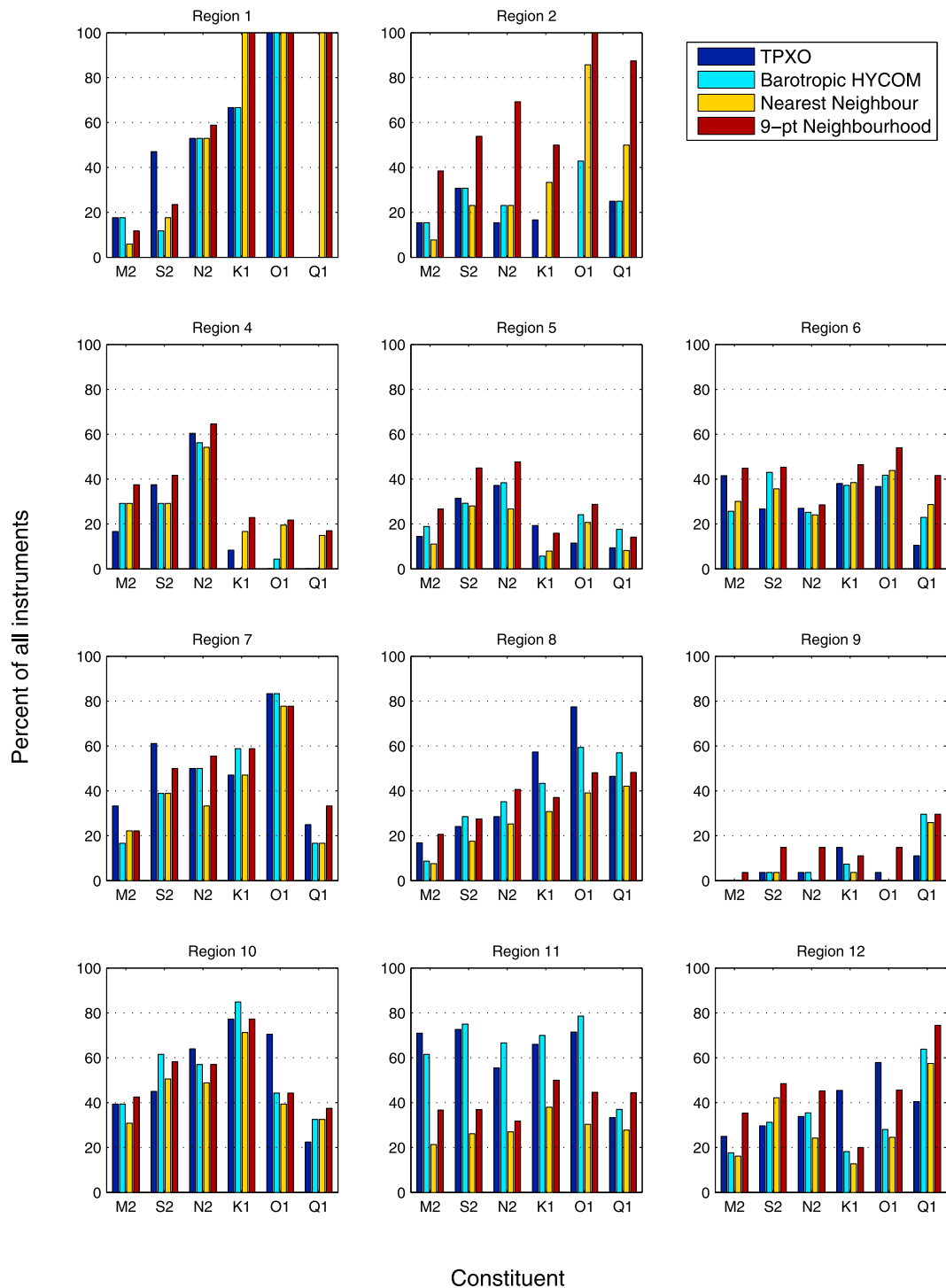
#### 4.1.2. Regional Scale Analysis

[39] We can also apply a similar analysis to each of the twelve ocean regions we have identified. Figure 5 shows the results of such an analysis for the semi-major axis. From Figure 5 it is clear that significant variability in model performance exists not only between the individual regions but also with regards to which constituents perform best within a specific region. We have not included the Florida Current System (region 3) in this analysis as it contains only a single current meter record for the month of September. Noise levels ( $ESR > 1$ ) in the western tropical NA (region 1) result in only 1–3 observations being available for comparison of the diurnal constituents. There are 17 observations available for comparison of the semi-diurnal constituents in this region. In the eastern tropical NA (region 2) only 6–8 observations are available for comparison of the diurnal constituents; 13 observations are available for comparison of the semi-diurnal constituents. Due to the limited number of observations available for comparison we do not consider the 100% success rates for the diurnal constituents in the western tropical NA to be significant. The small number of observations available for comparison in the eastern tropical NA also casts doubts on the significance of the results in that region.

[40] From a regional perspective all six tidal constituents perform above their basin success rates for TPXO 7.2 in the North Atlantic Current, Labrador Sea and central high-latitude North Atlantic (regions 7, 10, and 11). All six constituents of the HYCOM barotropic mode are also higher than their basin success rates in the Labrador Sea and central high-latitude NA. The only region in which all six constituents perform above their basin success rates at the nearest neighbor is the Labrador Sea. There are no regions in which all six constituents are higher than their basin success rates within their instrument neighborhoods. All constituents have low success rates in the Strait of Gibraltar (region 9) for all model values. Although the version of global HYCOM used for comparison has a horizontal resolution of  $1/12.5^\circ$  this is not a high enough resolution to resolve the Strait of Gibraltar properly and most observations from this region are associated with a single model point as their nearest neighbor. It is clear from the results of this study that the model point used for comparison is not representative of the tidal flow in this very active ocean region.

[41] All of the semi-diurnal constituents have success rates at or above their NA basin success rates in the North Atlantic Current system and Labrador Sea (regions 7 and 10) when tested at the nearest neighbor of the three-dimensional model. When we examine the semi-diurnal constituents in their instrument neighborhoods all three have success rates at or above their NA basin success rates in the eastern tropical NA, Labrador Sea and eastern high-latitude NA (regions 2, 10, and 12). Excluding the Strait of Gibraltar, constituent  $M_2$  has TPXO 7.2 success rates ranging between 14% and 71% with a maximal value in the central high-latitude NA and minimal values of 14%–18% in the tropical, subtropical and eastern midlatitude NA (region 1, 2, 4, 5, and 8). HYCOM barotropic mode success rates range from 9% to 62% with maximum in the central high-latitude NA and success rates of less than 20% in the tropical, eastern subtropical NA and the North Atlantic Current system. The HYCOM barotropic mode has higher success rates than

Model Semi-Major Axis Within 95% Confidence Intervals by Region



**Figure 5.** As in Figure 4 but for each of the 12 ocean regions depicted in Figure 3 and for the semi-major axis only.

TPXO 7.2 in the subtropical NA (regions 4 and 5) and lower success rates than TPXO 7.2 in the Gulf Stream and North Atlantic current systems, eastern midlatitude NA, and central and eastern high-latitude NA. Success rates for the three-dimensional HYCOM range between 6% and 31% at the nearest neighbor and are lower than TPXO 7.2 in all regions

except the western subtropical NA (region 4). Within the instrument neighborhoods success rates range from 12% to 45% and are more than 10% higher than TPXO 7.2 in the eastern tropical, subtropical NA, and eastern high-latitude NA (regions 2, 4, 5, and 12). HYCOM instrument neighborhoods are 3%–4% higher than TPXO 7.2 in the Gulf

Stream system, eastern midlatitude NA and Labrador Sea (regions 6, 8 and 10). Three-dimensional HYCOM success rates are 11% and 34% lower than TPXO 7.2 in the North Atlantic Current system and central high-latitude NA (regions 7 and 11), respectively. HYCOM success rates are also 6% lower than TPXO 7.2 in the western tropical NA.

[42] Constituent  $S_2$  has success rates of: 24%–73%, 12%–75%, 18%–51%, and 24%–58% for TPXO 7.2, HYCOM barotropic mode, at the nearest neighbor, and within an instrument neighborhood, respectively. The HYCOM barotropic mode success rates are 16% higher than TPXO 7.2 in the Gulf Stream system and Labrador Sea (regions 6 and 10) and 35%, 8% and 22% lower than TPXO 7.2 in the western tropical, western subtropical NA and North Atlantic Current system (regions 1, 4, and 7), respectively. Differences of  $\pm 5\%$  or less between TPXO 7.2 and the HYCOM barotropic mode exist in other regions. At the nearest neighbors the three-dimensional HYCOM model success rates are 3%–46% lower than TPXO 7.2 in all regions except the Gulf Stream system, Labrador Sea and eastern high-latitude NA (regions 6, 10, and 12). Within the instrument neighborhoods HYCOM success rate are 3%–23% higher than TPXO 7.2 in the eastern tropical, subtropical, Gulf Stream system, eastern midlatitude NA, Labrador Sea, and eastern high-latitude NA (regions 2, 4, 5, 6, 8, 10, and 12).

[43]  $N_2$  success rates, within the individual regions, range between 15%–64% for TPXO 7.2 and 23%–67% for the HYCOM barotropic mode. Success rates for TPXO 7.2 and the HYCOM barotropic mode differ by at most 11% in any region. For the three-dimensional velocity field at the nearest neighbor, HYCOM success rates are 3%–28% lower than TPXO 7.2 in all regions except the tropical NA (regions 1 and 2). Within the instrument neighborhoods, however, the only regions in which HYCOM success rates are lower than TPXO 7.2 are the Labrador Sea and central high-latitude NA (regions 10 and 11).

[44] A similar detailed analysis may be conducted for the diurnal constituents. In general, if we ignore the results from the tropical NA (regions 1 and 2) where we have a limited number of observations, the success rates for the diurnal constituent  $K_1$  are: 8%–77%, 0%–85%, 8%–71%, and 16%–77% for TPXO 7.2, HYCOM barotropic mode, at the nearest neighbor, and within the instrument neighborhoods, respectively. For  $O_1$  we have success rates of: 0%–83%, 4%–83%, 20%–78%, and 22%–78% and for  $Q_1$  the success rates are: 0%–46%, 0%–64%, 8%–57%, and 14%–74%.

[45] The poorest HYCOM model performance appears in the central high-latitude NA (region 11) where the success rate for all constituents, except  $Q_1$ , at both the nearest neighbor and within the instrument neighborhoods are lower than the success rates for TPXO 7.2. It is curious that the HYCOM barotropic mode in this region compares reasonably well to TPXO 7.2. The cause of the much lower performance of the baroclinic tides in this region needs to be investigated but such an investigation goes beyond the scope of this paper. We speculate that stratification may not be properly represented in this region. If model bathymetry was inaccurate we would not expect the barotropic mode to be comparable to TPXO 7.2. Diurnal constituents also perform poorly in the subtropical NA (regions 4 and 5) but HYCOM tends to have higher success rates than TPXO 7.2 in these regions.

[46] It is also possible to conduct a regional analysis for the other ellipse parameters. Our results (not shown) indicate that the variability seen in the results for the semi-major axis also exists for the other ellipse parameters. However, although similarities between the success rates of the other ellipse parameters and the semi-major axis do exist, the variability of the other ellipse parameters may differ from that found for the semi-major axis from region to region and within a single region. Some, but not all, of these differences may be attributed to large errors for the various ellipse parameters within a given region and the results of such a regional analysis should be compared to the mean and median values of the errors to determine the likelihood of false positive outcomes. The differences that do exist between the outcomes for the semi-major axis and the other ellipse parameters suggest that assessment of the other ellipse parameters is likely to be a useful method when applied on a regional scale. The differences that do exist may be exploited to determine those locations where the model is able to match not only the semi-major axis but also the semi-minor axis, direction of the principal axis of the tidal ellipse (ellipse inclination) and the time of maximal tidal flow (Greenwich phase) as we demonstrate next.

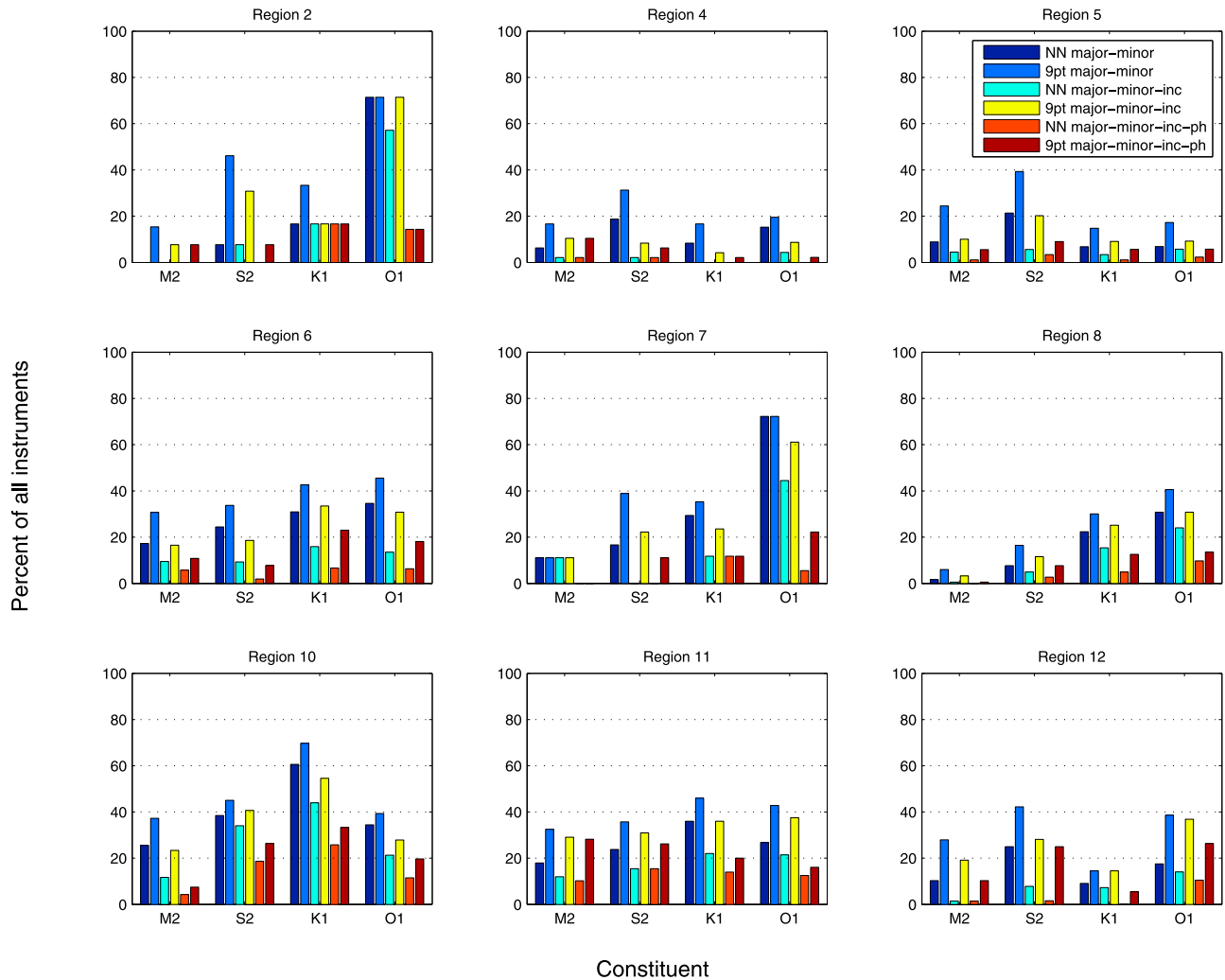
[47] We extend our regional analysis by intersecting the subsets containing positive outcomes for the individual ellipse parameters in order to determine those locations where the model values are able to satisfy more than one ellipse parameter at a time. Figure 6 shows the results of intersecting the subsets of positive outcomes for the nine ocean regions that contain the largest geographic areas. We have only included the results for constituents  $M_2$ ,  $S_2$ ,  $K_1$  and  $O_1$  for reasons of clarity.

[48] For our purposes we will consider the number of positive outcomes of the intersected subsets of ellipse parameters within a given region to be significant if they equal or exceed the basin wide success rates for the semi-major axis of a constituent. This test allows us to identify those regions in which model skill is above the average model skill for the NA. For constituent  $M_2$  this means that the percent of all instruments that coincidentally satisfy two or more ellipse parameters must be equal to or greater than 20% (34%) in order to be considered significant at the nearest neighbor (within the instrument neighborhoods). The corresponding values for  $S_2$ ,  $K_1$ , and  $O_1$  are 31%, 32% and 36% (42%, 40%, and 45%), respectively.

[49] When we require that the model semi-major and semi-minor axis both satisfy their 95% confidence intervals, constituents  $M_2$  and  $S_2$  have significant success rates of 26% and 38% in the Labrador Sea (region 10) at the nearest neighbor.  $K_1$  has nearest neighbor success rates of 61% and 36% in the Labrador Sea and central high-latitude NA (region 11), respectively. Constituent  $O_1$  has success rates of 71% and 72% in the eastern tropical NA and North Atlantic Current system (regions 2 and 7) when tested at the nearest neighbor.

[50] Within the instrument neighborhoods the success rate of 37% for  $M_2$  in the Labrador Sea (region 10) is significant. For  $S_2$  significant success rates of 46%, 45%, and 42% occur in the eastern tropical, Labrador Sea and eastern high-latitude NA (regions 2, 5 and 12). Success rates that are significant within the instrument neighborhoods for the diurnal constituent  $K_1$  are 43%, 70%, and 46% in the Gulf Stream

Model Values Satisfying Two or More Ellipse Parameter 95% Confidence Intervals by Region



**Figure 6.** Percent of all September observations in the nine largest regions (by geographic area) for which the model values lie within the intersections of the subsets of those model values satisfying their respective 95% confidence intervals of the September observations. Dark blue (blue) represent intersections of the semi-major and semi-minor axis subsets when model values are tested against the nearest neighbor, NN, (within the 9-point instrument neighborhoods, 9pt). Light blue (yellow) represent the intersections of semi-major, semi-minor axis and ellipse inclination at the nearest neighbor (within 9-point neighborhoods). Red (dark red) represent the intersections of semi-major, semi-minor axis, inclination and Greenwich phase.

system, Labrador Sea and central high-latitude NA (regions 6, 10, and 11). For  $O_1$ , within the instrument neighborhoods the success rates were 71%, 46%, and 72% in the eastern tropical, Gulf Stream and North Atlantic Current systems (regions 2, 6, and 7).

[51] The only regions in which the success rates were considered significant when the subsets of semi-major axis, semi-minor axis and ellipse inclination are intersected at the nearest neighbor occur for constituent  $S_2$  and  $K_1$  in the Labrador Sea (34% and 44%, respectively) and constituent  $O_1$  in the eastern tropical NA and North Atlantic Current system (57% and 44%, respectively). Within the instrument neighborhoods the  $S_2$  success rate for the Labrador Sea was not considered significant and  $K_1$  had a success rate 55%.

The neighborhood success rates for  $O_1$  were 71% and 61%, respectively.

[52] For intersections of the semi-major, semi-minor axis, inclination and Greenwich phase there are no significant success rates within any region for any constituent. The most consistent model performance appears in the high-latitude NA (regions 10, 11, and 12) where the success rates within the instrument neighborhoods for constituent  $M_2$  were: 7%, 28%, and 10% for the Labrador Sea, central high-latitude, and eastern high-latitude NA, respectively. For  $S_2$  the respective success rates are: 26%, 26%, and 25%; for  $K_1$ : 33%, 20%, and 5%; and for  $O_1$ : 20%, 16%, and 23%.

[53] So far we have restricted our attention to identifying those locations where model values are within their respective 95% confidence intervals. To require that the model

satisfy such a high level of confidence is very restrictive in the sense that those model values that satisfy such a high level of confidence typically have a median ESR less than 0.30 for the semi-diurnal constituents. In open water, away from continental margins and also at depths below the mixed layer this usually translates into an error of less than  $1 \text{ cm s}^{-1}$  for the semi-major axis.

#### 4.2. Assessing Model Skill Using Vertical Profiles of Velocity

[54] When assessing model skill at mooring locations we use all available observations within the NA even if those observations did not occur during the month of September.  $RMS_{moor}$  and  $CSS_{moor}$  are calculated at the 276 moorings with 3 or more instruments. The observed semi-major axis values used here result from harmonic tidal analysis of the complete time series regardless of its length. The combined scores of  $RMS_{moor}$  and  $CSS_{moor}$  (equations (3) and (4)) for constituent  $M_2$  evaluated at the nearest neighbor and also at the mooring best fit location are shown in Figure 7. Those moorings for which  $CSS_{moor}$  satisfies the 95% significance level (i.e.,  $p < 0.05$ , one-tailed) are indicated by solid dots. Open circles are used to show the  $RMS_{moor}$  values when  $CSS_{moor}$  is not significant. The significance level of the correlation is obtained by counting permutations for sample sizes less than 10 [Gibbons, 1985] and estimated using an Edgeworth expansion for larger samples [Best and Roberts, 1975]. We have not included the Strait of Gibraltar in this figure because high values of  $RMS_{moor}$  ( $\sim 30 \text{ cm s}^{-1}$ ) do not scale well. The Strait of Gibraltar contains only 2% of all moorings with 3 or more instruments.

[55] Our tests of the vertical profile reveal that 53% (63%) of all moorings with 3 or more instruments have  $RMS_{moor} \leq 1 \text{ cm s}^{-1}$  for constituent  $M_2$  at the nearest neighbor (best fit location). Another 21% (20%) of all moorings satisfy  $1 \text{ cm s}^{-1} < RMS_{moor} \leq 2 \text{ cm s}^{-1}$  at the nearest neighbor (best fit location). Of those moorings at the nearest neighbor (best fit location) with  $RMS_{moor} > 2 \text{ cm s}^{-1}$  23% (38%) are located in the Gulf Stream system, 18% (16%) are located in the eastern midlatitude NA, and another 35% (25%) are located in the eastern high-latitude NA. For the other tidal constituents 85%–89% satisfy  $RMS_{moor} < 1 \text{ cm s}^{-1}$  at the both the nearest neighbors and best fit locations. Overall, we find that for those regions containing at least 10 moorings the condition that  $RMS_{moor} < 1 \text{ cm s}^{-1}$  was satisfied between 67% and 100% of the time in the subtropical NA, the Gulf Stream/North Atlantic current systems and the Labrador Sea (regions 4, 5, 6, 7, and 10). The percentage of instruments with  $RMS_{moor} < 1 \text{ cm s}^{-1}$  in the eastern midlatitude NA, central and eastern high-latitude NA was only 13%–33%. All other regions had fewer than 6 moorings available for comparison.

[56] Only 7% (11%) of all moorings at the nearest neighbor (best fit location) have correlations significant at the 95% level for constituent  $M_2$ . The other constituents have significant correlations for 4%–6% (5%–9%) of all moorings with three or more instruments. The low percentage of significant correlations, not much higher than may be expected by pure chance, can be partially attributed to the 48% (28%) of moorings with only 3(4) instruments. The percent of moorings with significant correlations does not increase significantly until the level of significance is decreased to

80% ( $p < 0.20$ ). At a significance level of 80% the number of significant correlations for  $M_2$  increases to 24% (32%) at the nearest neighbor (best fit location). For the other constituents the percent of significant correlations at the 80% significance level is 17%–26% (24%–32%) at the nearest neighbor (best fit location). Of the 59 moorings with 5 or more instruments 20% have significant nearest neighbor  $CSS_{moor}$  scores at both the 95% and 90% levels; 27%, and 34% are significant at the 85% and 80% levels, respectively. If we evaluate  $CSS_{moor}$  at the best fit locations, the percentage of significant correlations increase to 31%, 34%, 37%, and 46% at the 95%, 90%, 85%, and 80% levels.

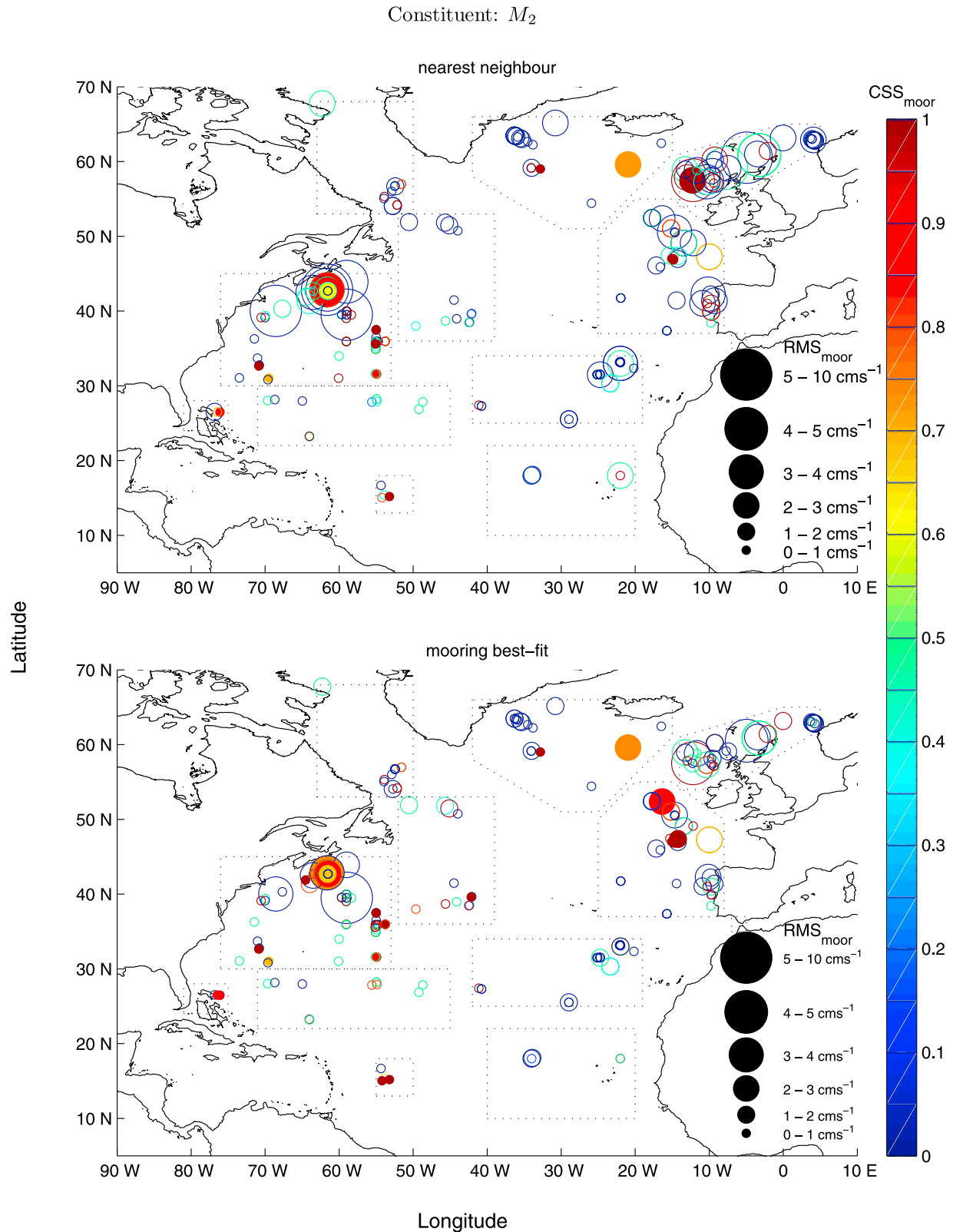
[57] Within the 12 regions we have identified, the percentage of moorings that have significant correlations ( $p < 0.05$ , one-tailed) is less than 20% in most regions (not shown). In the tropical North Atlantic (regions 1 and 2) and the Florida Current system (region 3) there are 5 or fewer moorings with 3 or more instruments in each region (Table 4). With such a small sample it is difficult to conclude that the percent of moorings with significant correlations ( $p < 0.05$ , one-tailed) is itself significant.

[58] The usefulness of  $CSS_{moor}$  in this study is limited due to the small number of instruments on the individual moorings. However, we believe that  $CSS_{moor}$  may still provide useful information particularly when comparing model data to velocity observations with a higher vertical resolution such as those provided by an Acoustic Doppler Current Profiler (ADCP). If observations are available with a high vertical resolution it is more likely that the baroclinic structure of the water column will be revealed and  $CSS_{moor}$  should demonstrate if the model baroclinic structure matches the observed baroclinic structure. Significant correlations exist in the eastern subtropical, midlatitude and high-latitude NA for the semi-diurnal constituents. Diurnal constituents are found to be significantly correlated for the North Atlantic Current system and Labrador Sea.

#### 4.3. Assessing Model Skill Using Horizontal Profiles of Velocity

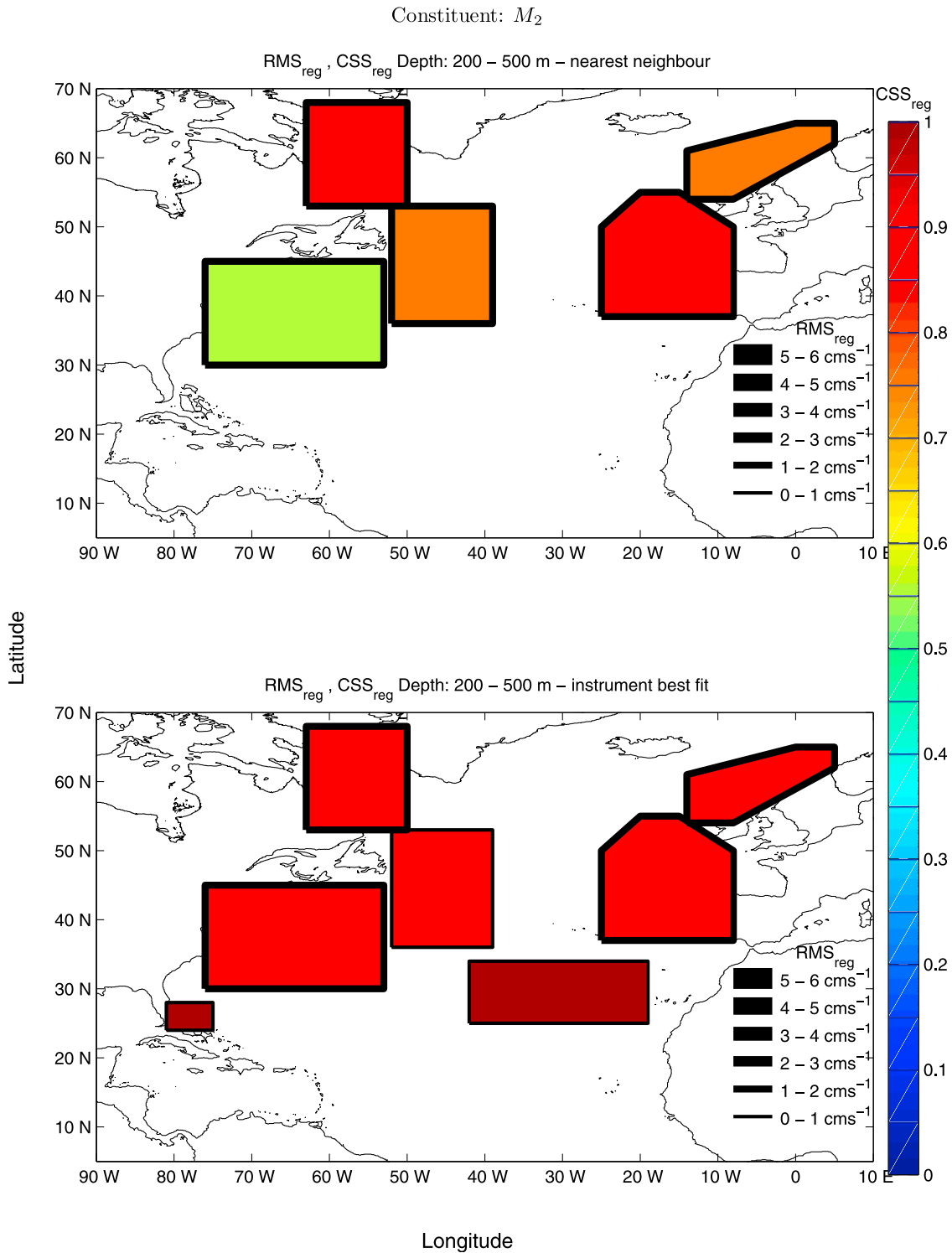
[59] Our final approach in assessing model skill examines each individual region using depth bins. For this analysis we select all instruments within a pre-determined depth bin for a given region and then calculate  $RMS_{reg}$  and  $CSS_{reg}$  (equations (5) and (6)) provided there are at least 3 or more instruments within the depth bin. While it is possible to adjust the depth bins for each individual region we have chosen to illustrate our method by using a single depth bin from 200 to 500 m for all regions within the NA. Figure 8 shows the combined score of  $RMS_{reg}$  and  $CSS_{reg}$  for those regions that have significant correlations at the 95% level ( $p < 0.05$ , one-tailed).  $RMS_{reg}$  and  $CSS_{reg}$  are evaluated at both the nearest neighbor (top panel) and using the instrument best fit location (bottom panel). The Strait of Gibraltar is not included in this figure as high  $RMS_{reg}$  ( $\sim 30 \text{ cm s}^{-1}$ ) scores in this region do not scale well.

[60] For constituent  $M_2$  significant correlations are found in the Gulf Stream system (region 6,  $RMS_{reg} = 1.71$ ,  $CSS_{reg} = 0.556$ ,  $p = 9.734 \times 10^{-10}$ ,  $df = 98$ ), North Atlantic Current system (region 7,  $RMS_{reg} = 1.36$ ,  $CSS_{reg} = 0.755$ ,  $p = 3.298 \times 10^{-3}$ ,  $df = 10$ ), eastern midlatitude NA (regions 8,  $RMS_{reg} = 1.91$ ,  $CSS_{reg} = 0.900$ ,  $p = 1.014 \times 10^{-3}$ ,  $df = 7$ ), Labrador Sea (region 10,  $RMS_{reg} = 1.98$ ,



**Figure 7.** Location of moorings with 3 or more instruments are represented with  $RMS_{moor}$  as the scaled radius of the circle and the color representing,  $CSS_{moor}$ , the correlation between model and observations at that mooring, for tidal constituent  $M_2$  evaluated (top) at the nearest neighbor and (bottom) at the mooring best fit. Those moorings having correlations significant at the 95% level are filled with solid color. The Strait of Gibraltar is not included.  $RMS_{moor}$  and  $CSS_{moor}$  are calculated for all observations in the North Atlantic.

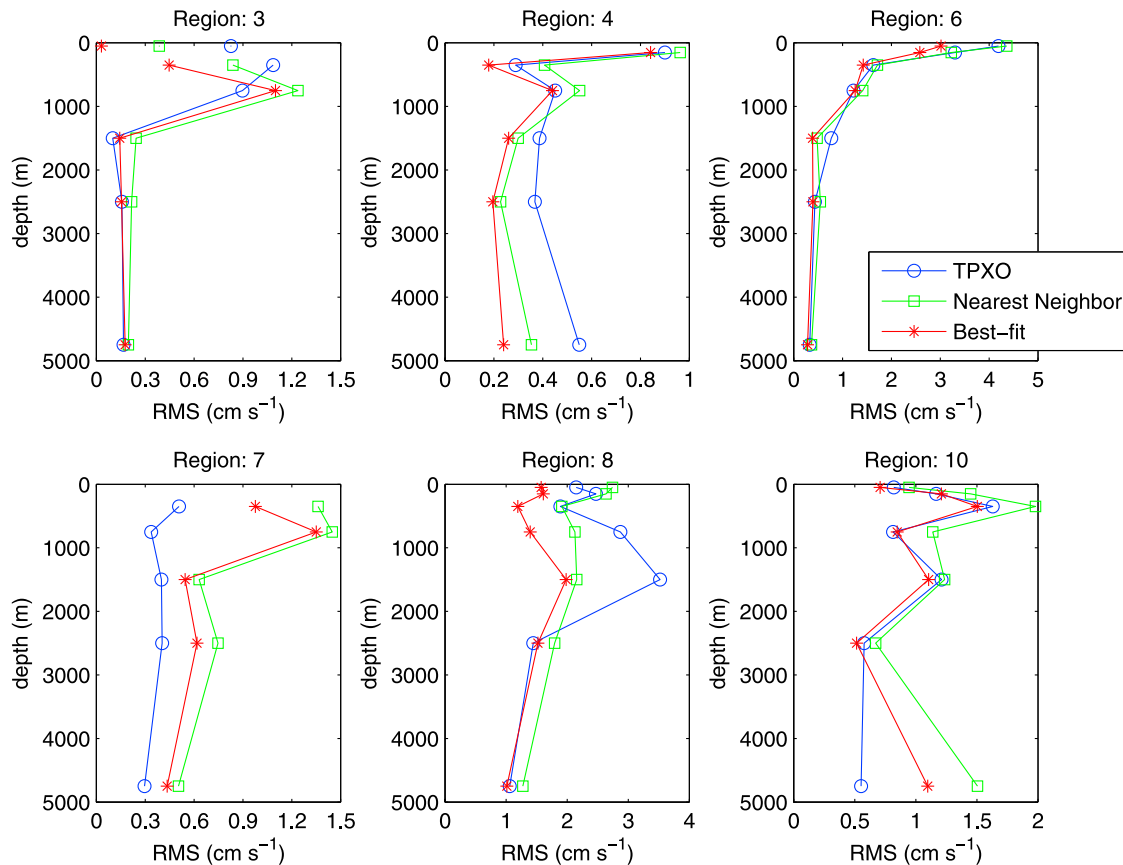




**Figure 8.** Regional depth bins containing at least 3 instruments between 200 and 500 m. The thickness of the line outlining the box represents a scaled  $RMS_{reg}$  with the color representing  $CSS_{reg}$ , the correlation between all instruments within the depth bin and the corresponding model values, for tidal constituent  $M_2$  evaluated (top) at the nearest neighbor and (bottom) at the instrument best fit.

$CSS_{reg} = 0.871$ ,  $p = 1.187 \times 10^{-5}$ ,  $df = 13$ ), and eastern high-latitude NA (region 12,  $RMS_{reg} = 1.59$ ,  $CSS_{reg} = 0.76$ ,  $p = 9.686 \times 10^{-11}$ ,  $df = 47$ ) for the depths 200–500 m when evaluated at the nearest neighbor. When evaluated at the instrument best fit  $RMS_{reg}$  ( $CSS_{reg}$ ) tends to decrease

(increase). When using the instrument best fit locations the Florida Current system (region 3,  $RMS_{reg}(best\ fit) = 0.45$ ,  $CSS_{reg}(best\ fit) = 1.000$ ,  $p = 4.167 \times 10^{-2}$ ,  $df = 2$ ) and eastern subtropical NA (region 5,  $RMS_{reg}(best\ fit) = 0.15$ ,



**Figure 9.** Vertical profiles of  $RMS_{reg}$  ( $\text{cm s}^{-1}$ ) for those regions where observations exist over the entire water column.  $RMS_{reg}$  values for TPXO 7.2 (blue with circles), three-dimensional HYCOM at the nearest neighbor (green with squares), and three-dimensional HYCOM using the instrument best fit from the instrument neighborhoods (red with asterisks) are plotted at the mean depth for each depth bin represented by circles, squares and asterisks. Depth bins used are: 0–100 m, 100–200 m, 200–500 m, 500–1000 m, 1000–2000 m, 2000–3000 m, and 3000–6500 m.

$CSS_{reg}(best\ fit) = 1.000$ ,  $p = 1.984 \times 10^{-4}$ ,  $df = 5$ ) also have significant correlations.

[61] The other constituents have similar behavior in the 200–500 m layer although they generally have fewer regions with significant correlations (not shown).  $S_2$  only has significant correlations in the Gulf Stream and North Atlantic Current systems, Labrador Sea and eastern high-latitude NA (regions 6, 7, 10, and 12);  $N_2$  has significant correlations in the mid and high latitudes (regions 7, 8, 10 and 12) when evaluated at the nearest neighbor. When evaluated at the best fit locations significant correlations also exist for  $S_2$  in the Florida Current system, eastern subtropical and eastern midlatitude NA (regions 3, 5, and 8) and for  $N_2$  in the subtropical NA, Gulf Stream systems (regions 4, 5, and 6).

[62] All three diurnal constituents have significant correlations between 200 and 500 m depth in the Labrador Sea when evaluated at the nearest neighbor. In addition to this region, when  $CSS_{reg}$  is evaluated at the nearest neighbor  $K_1$  also has a significant correlation in the western tropical NA;  $O_1$  has significant correlations in the Gulf Stream and North Atlantic Current systems, eastern subtropical, midlatitude, and high-latitude NA; and  $Q_1$  has significant correlation in the eastern high-latitude NA. As with the semi-diurnal constituents, when evaluated at the best fit locations the number

of regions with significant correlations for the diurnal constituents tends to increase along with increased values for  $CSS_{reg}$  in each region.

[63] Our regional analysis using the values of  $RMS_{reg}$  and  $CSS_{reg}$  indicates that tidal performance is significantly correlated across large geographic regions. Correlations between the model and observations tend to be higher in the mid and high-latitude ocean regions. Regional correlation scores in active ocean regions such as the Gulf Stream system and North Atlantic Current system also appear to be lower than in the eastern half of the NA when the model is evaluated at the nearest neighbor.

[64] In the absence of high vertical resolution observations it is possible to use  $RMS_{reg}$  to construct a regional view of the ability of the model to represent the vertical structure of the tidal currents. Figure 9 shows the vertical profile of  $RMS_{reg}$  of constituent  $M_2$  for those regions in which we are able to profile the entire water column. As can be seen in Figure 9 differences exist when  $RMS_{reg}$  scores are calculated from model values at the nearest neighbor and also at the best fit locations. These differences tend to be small and for the most part  $RMS_{reg}(nearest-neighbor)$  tends to track  $RMS_{reg}(best\ fit)$  through the water column. In most of the following discussion  $RMS_{reg}(best\ fit)$  could be replaced by

$RMS_{reg}(\text{nearest-neighbor})$  with only slight changes to the score values.

[65] In all the regions depicted, except the North Atlantic Current system (region 7),  $RMS_{reg}(TPXO)$  is shown to be larger or comparable to  $RMS_{reg}(\text{best fit})$  in the top 500 m of the water column. This suggests that three-dimensional HYCOM is, for the most part, able to provide a better representation of the upper level tides than the barotropic TPXO 7.2 model. In the Florida Current and Gulf Stream systems (regions 3 and 6) HYCOM and TPXO 7.2 both perform with approximately the same amount of skill below 500 m depth. In the western subtropical NA (region 4) HYCOM has  $RMS_{reg}(\text{best fit}) \sim 0.2 \text{ cm s}^{-1}$  compared to TPXO 7.2 with  $RMS_{reg}(TPXO) \sim 0.4\text{--}0.5 \text{ cm s}^{-1}$  below 1000 m depth. In the eastern midlatitude NA (region 8) between 500 and 2000 m depth,  $RMS_{reg}(\text{best fit}) \sim 1.4\text{--}2.0 \text{ cm s}^{-1}$  compared to  $RMS_{reg}(TPXO) \sim 2.9\text{--}3.5 \text{ cm s}^{-1}$ . In the Labrador Sea (region 10)  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  are comparable between 0 and 3000 m but below 3000 m  $RMS_{reg}(TPXO) = 0.6 \text{ cm s}^{-1}$  is lower than  $RMS_{reg}(\text{best fit}) = 1.1 \text{ cm s}^{-1}$ . TPXO 7.2 also has lower  $RMS_{reg}$  scores than HYCOM in the North Atlantic Current system (region 7). The cause of the poor performance of HYCOM in this region when evaluated by depth bins is not known at present but should be investigated. However, such an investigation is beyond the scope of the work presented here.

[66] A similar analysis can also be done for the other constituents. For constituent  $S_2$ , between 200 and 1000 m depth in the Florida Current system (region 3) we find  $RMS_{reg}(TPXO) \sim 0.15\text{--}0.36 \text{ cm s}^{-1}$  is much lower than  $RMS_{reg}(\text{best fit}) \sim 0.81\text{--}0.85 \text{ cm s}^{-1}$ . In the Gulf Stream system (region 6) differences between  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  are of order  $0.1 \text{ cm s}^{-1}$  through most of the water column except in the top 100 m where  $RMS_{reg}(TPXO) \sim 1.2 \text{ cm s}^{-1}$  and  $RMS_{reg}(\text{best fit}) \sim 0.8 \text{ cm s}^{-1}$ . In the North Atlantic Current system (region 7), between 200 and 1000 m depth  $RMS_{reg}(TPXO) \sim 0.17 \text{ cm s}^{-1}$  compared to  $RMS_{reg}(\text{best fit}) \sim 0.40\text{--}0.63 \text{ cm s}^{-1}$ . In the eastern midlatitude NA (region 8)  $RMS_{reg}(\text{best fit}) = 0.29\text{--}0.48 \text{ cm s}^{-1}$  between 100 and 1000 m depth is lower than  $RMS_{reg}(TPXO) = 0.63\text{--}0.87 \text{ cm s}^{-1}$ . In the western subtropical NA (region 4) differences between  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  are of order  $0.1 \text{ cm s}^{-1}$  through the entire water column with  $RMS_{reg}(\text{best fit})$  tending to be smaller than  $RMS_{reg}(TPXO)$  except between 2000 and 3000 m depth. In Labrador Sea (region 10)  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  differences are of order  $0.2 \text{ cm s}^{-1}$  with the smaller value alternating between TPXO 7.2 and HYCOM.

[67] For constituent  $N_2$  differences between  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  are typically less than  $0.1 \text{ cm s}^{-1}$  below 500 m depth for all regions considered except the eastern midlatitude NA (region 8). In the top 500 m of the Florida Current system  $RMS_{reg}(\text{best fit})$  is smaller than  $RMS_{reg}(TPXO)$  by  $\sim 0.2 \text{ cm s}^{-1}$ .  $RMS_{reg}(\text{best fit})$  is also  $0.2\text{--}0.4 \text{ cm s}^{-1}$  smaller than  $RMS_{reg}(TPXO)$  in the top 200 m of the Gulf Stream system.  $RMS_{reg}(TPXO)$  is  $\sim 0.2 \text{ cm s}^{-1}$  smaller than  $RMS_{reg}(\text{best fit})$  between 200 and 500 m in the Labrador Sea. In the eastern midlatitude NA  $RMS_{reg}(\text{best fit}) \sim 0.04\text{--}0.38 \text{ cm s}^{-1}$  in the top 2000 m with the smallest value in the top 100 m and increasing with depth. This is in contrast to  $RMS_{reg}(TPXO) \sim 0.44\text{--}0.66$  with no obvious trend related to increasing depth.

[68] Diurnal constituent  $K_1$  has differences between  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  typically ranging between  $0.05$  and  $0.20 \text{ cm s}^{-1}$  through the entire water column with TPXO and HYCOM having alternating smaller values except for the western subtropical NA and eastern midlatitude NA (regions 4 and 8). In the western subtropical NA between 500 and 2000 m depth  $RMS_{reg}(\text{best fit}) \sim 0.2\text{--}0.3 \text{ cm s}^{-1}$  is much lower than  $RMS_{reg}(TPXO) \sim 0.5\text{--}0.7 \text{ cm s}^{-1}$ . In the eastern midlatitude NA  $RMS_{reg}(TPXO) \sim 0.1 \text{ cm s}^{-1}$  is lower than  $RMS_{reg}(\text{best fit}) \sim 0.2\text{--}0.4 \text{ cm s}^{-1}$  below 200 m depth.

[69] For constituent  $O_1$   $RMS_{reg}(TPXO) \sim 0.6\text{--}2.9 \text{ cm s}^{-1}$  is lower than  $RMS_{reg}(\text{best fit}) \sim 0.4\text{--}3.3 \text{ cm s}^{-1}$  through the entire water column of the Florida Current system. Differences in  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  of greater than  $0.3 \text{ cm s}^{-1}$  also exist between: 200 and 1000 m depth in the western subtropical NA; 500 and 2000 m depth in the Gulf Stream system; and 200 and 500 m depth of the Labrador Sea. In the Gulf Stream system  $RMS_{reg}(\text{best fit}) \sim 0.08\text{--}0.61 \text{ cm s}^{-1}$  is consistently smaller than  $RMS_{reg}(TPXO) \sim 0.08\text{--}1.90 \text{ cm s}^{-1}$  through all depths. In the Labrador Sea  $RMS_{reg}(TPXO)$  is smaller than  $RMS_{reg}(\text{best fit})$  with a maximum difference of  $0.3 \text{ cm s}^{-1}$  occurring between 200 and 500 m depth.

[70] For constituent  $Q_1$   $RMS_{reg}(\text{best fit})$  is consistently lower than  $RMS_{reg}(TPXO)$  for all depth bins of the Gulf Stream, North Atlantic Current system, eastern midlatitude NA and the Labrador Sea. In the North Atlantic Current system and eastern midlatitude NA the differences are less than  $0.1 \text{ cm s}^{-1}$  for all depths. In the Gulf Stream system differences between  $RMS_{reg}(TPXO)$  and  $RMS_{reg}(\text{best fit})$  are of order  $0.1\text{--}0.3 \text{ cm s}^{-1}$  in the top 2000 m and less than  $0.1 \text{ cm s}^{-1}$  below that depth. In the Labrador Sea differences in the top 500 m are of order  $0.1\text{--}0.2 \text{ cm s}^{-1}$ . In the Florida Current system  $RMS_{reg}(TPXO)$  is less than  $RMS_{reg}(\text{best fit})$  by  $0.4\text{--}1.7 \text{ cm s}^{-1}$  between 200 and 3000 m depth but is larger than  $RMS_{reg}(\text{best fit})$  by  $1.9 \text{ cm s}^{-1}$  in the top 100 m of the water column. Below 200 m depth in the western subtropical NA differences between the  $RMS_{reg}$  scores of the two models is of order  $0.1\text{--}0.2 \text{ cm s}^{-1}$  with no clear advantage shown by either model. Between 100 and 200 m depth, however,  $RMS_{reg}(TPXO) = 0.35 \text{ cm s}^{-1}$  compared to  $RMS_{reg}(\text{best fit}) = 1.63 \text{ cm s}^{-1}$ .

## 5. Summary

[71] This paper evaluates the performance of tidal currents in non-data assimilative HYCOM measured against observations of the current velocity at depth and the data assimilative barotropic TPXO 7.2. For the six tidal constituents analyzed here, HYCOM captures 60%–81% the variance in the sea-surface elevations in the NA exhibited by TPXO 7.2. While sea surface height is well represented, estimates of barotropic tidal kinetic energy indicate that constituent  $K_1$  is weak compared to tidal kinetic energy estimated from TPXO. The barotropic tidal kinetic energy was calculated from only 30 days of model output for September 2004. Some of the discrepancy between the barotropic tidal kinetic energy of TPXO 7.2 and HYCOM for  $K_1$  may be attributed to the destructive interference of constituent  $P_1$  which results in minimum values for  $K_1$  in September. When constituent  $K_1$  is compared to current meter observations it appears to

perform as well as constituent  $M_2$  in most of the skill tests presented above. This apparent contradiction may be due to a bias in the observations which tend to be concentrated in less energetic regions of the ocean.

[72] The use of 95% bootstrapped confidence intervals to estimate model performance on a basin scale for ellipse parameters other than the semi-major axis is likely to produce numerous false positive outcomes due to the relatively large errors that may be associated with the semi-minor axis, ellipse inclination and Greenwich phase for the weaker tidal constituents. Our results indicate that, within the North Atlantic Ocean, the barotropic HYCOM mode when evaluated at the nearest neighbor is comparable to data assimilative TPXO 7.2 in most regions of the ocean. Three-dimensional HYCOM is able to satisfy 95% confidence intervals for the semi-major axis between 20% and 35% of the time when tested at the model nearest neighbor and between 35% and 45% of the time when tested within a 9-point instrument neighborhood. Model performance is also found to differ significantly between regions within the NA with some regions having success rates of 40%–60% for some constituents. Testing of other ellipse parameters may also produce useful results on a regional perspective provided that the errors associated with the observations in those regions are relatively small. When evaluated at the nearest neighbor three-dimensional HYCOM appears weaker than both TPXO 7.2 and the barotropic HYCOM mode. However, three-dimensional non-data assimilative HYCOM appears to perform as well as and slightly better than data assimilative TPXO 7.2 for most ocean regions when we examine the model values within their 9-point instrument neighborhoods. The search of model values within their 9-point neighborhoods allows us to account for uncertainties in model versus true bathymetry and/or differences in model stratification versus true stratification that may exist between model output and the time of observation.

[73] The intersection of the subsets of model values able to satisfy the confidence intervals for the individual ellipse parameters within a region may also provide additional information identifying locations where model skill is above the basin average and able to satisfy two or more tidal ellipse parameters. When subsets of model values able to satisfy individual ellipse parameters were intersected we found no regions where there were a significant number of positive outcomes where model values were able to satisfy all four ellipse parameters simultaneously and only a few instances where HYCOM was able to satisfy both the semi-major axis and semi-minor axis simultaneously.

[74] It should be noted that demanding that model values satisfy the 95% confidence intervals is a very high standard to achieve. We chose to use 95% confidence intervals as it is a common reference used in many fields and we feel that such a high standard should be achievable. The median value of ESR for the semi-major axis of  $M_2$  is  $\sim 0.1$ . In order to satisfy the 95% confidence intervals for such signals, model values typically must be with 10%–20% of the observed value. For  $M_2$ , which has a basin average current of  $3\text{--}4\text{ cm s}^{-1}$ , this requires a difference of  $\sim 0.5\text{ cm s}^{-1}$  between the model and observations. The other tidal constituents typically have much weaker currents associated with them but the ESR tends to be larger. However, even when  $\text{ESR} \sim 1.0$ , for these constituents, differences between

model values are still expected to be no greater than  $0.5\text{--}1.5\text{ cm s}^{-1}$  in order to satisfy the 95% confidence intervals. We have chosen to select records for which  $\text{SNR} > 1$  for all constituents. In order to control the number of false positive outcomes the required SNR for the individual constituents may be adjusted. We have not yet investigated the effect of increasing SNR on the various constituents.

[75] The baroclinic and barotropic tides are not separable for point measurements such as current meter records. One challenge in attempting to assess model performance for baroclinic tides on a global scale is the lack of high vertical resolution in historical current meter records. In the absence of sufficient vertical resolution it is not always possible to separate the barotropic and baroclinic components of the tidal signals. Additional research is currently being done to investigate and exploit the observations available in the CMA to determine how much information is available and also how it may be used to produce additional methods for assessing model performance.

[76] The results of the analysis using root mean square error and correlation values enables us to assess model performance through the water column and across geographic regions; this is not possible when testing if model values satisfy the 95% confidence intervals for the individual observations. In the vertical direction our tests of the vertical profile reveal that 53% (63%) of all moorings with 3 or more instruments have  $\text{RMS}_{\text{moor}} \leq 1\text{ cm s}^{-1}$  for constituent  $M_2$  at the nearest neighbor (best fit location). Those moorings with  $\text{RMS}_{\text{moor}} > 2\text{ cm s}^{-1}$  for  $M_2$  are typically located in the Gulf Stream system, eastern mid- and high-latitude NA. The influence of the Gulf Stream may explain the large relative errors within the Gulf Stream system. The set of observations in the midlatitude NA includes many coastal observations while the eastern high-latitude NA contains many prominent seamounts. Therefore it seems likely that coastal effects and bathymetric representation may play an important role in these regions. For the other constituents up to 90% of the instruments have  $\text{RMS}_{\text{moor}} < 1\text{ cm s}^{-1}$ , however, these constituents are weaker than  $M_2$  and may have larger relative errors.

[77] Less than 11% of all moorings have correlations between the model and observations at the 95% significance level. The number of instruments distributed through the water column at an individual mooring limits the ability of our correlation skill test to achieve a 95% significance level. Of the 276 mooring locations with 3 or more instruments 76% have only 3 or 4 instruments. 24%–32% of all moorings have significant correlations at the 80% level which may be a more appropriate test statistic given the large number of moorings with only 3 or 4 instruments. Those moorings with at least 5 instruments were found to be significant at the 95% level 20% of the time.

[78] By examining the match between the model and observations along a mooring line with 3 or more instruments we are able to not only able to assess how well velocity in the model matches the observations but also if the vertical shear within the model matches the vertical shear at the time of observation. Our results indicate that even in the abyssal plains, where the tide is expected to be strongly barotropic, HYCOM does not always reflect the observed vertical shear. One possible explanation for this may be the isopycnal coordinate system which provides a single velocity value to

represent very thick layers over the abyssal plains. The model may not be able to properly represent differences in velocity if the observations all lie within the same isopycnal layer. A few instruments were also found to have a recorded depth located below model depth. In the results presented here all such locations were counted as not satisfying the test being applied.

[79] In the horizontal direction our test groups together current meter records into depth bins. The observations that belong to an individual depth bin can span several different time periods during which stratification can vary. The results of our skill test between 200 and 500 m shows that the model tidal currents are highly correlated, in the horizontal directions, to observations below the typical mixed layer depth of 200 m.

[80] By grouping the instruments into depth bins we are able to assess model performance over different depth ranges. We have used pre-determined layers to represent the mixed layer, mid ocean levels and abyssal depths. While the mooring locations represent the vertical structure present during a single period of observation, the depth bins often combine instruments from different periods of time and therefore reflect model performance compared to ocean conditions spanning several years. Low scores of  $RMS_{reg}$  and high scores of  $CSS_{reg}$  when observations span different time periods indicate good performance compared to a variety of possible ocean conditions.

[81] In five of the six regions for which we are able to profile  $RMS_{reg}$  over the entire water column three-dimensional HYCOM was found to have lower  $RMS_{reg}$  scores than TPXO 7.2 in the upper 500 m for constituent M<sub>2</sub>. HYCOM also had lower scores than TPXO 7.2 in the western subtropical eastern midlatitude NA below 500 m depth. TPXO 7.2 had higher scores than three-dimensional HYCOM in the North Atlantic Current system and also in the Labrador Sea below 3000 m depth. HYCOM is able to reflect some of the baroclinic structure of the tidal currents when averaged over larger geographic regions even though HYCOM's ability to match the exact baroclinic structures present in the ocean appears to be limited as indicated by the poor  $CSS_{moor}$  scores. The ability to model the baroclinic structure is not possible with TPXO 7.2 since it is a barotropic model.

[82] Misrepresentation of local bathymetry and seasonal differences in mixed layer depth between the model and observations can both influence local tidal currents. Such differences may explain some of the model/observation mismatches revealed in this study. Since three-dimensional HYCOM is capable of modeling the vertical structure of the ocean currents the baroclinic modes that may be represented in the model are dependent on the number of the vertical layers used by the model (and their thicknesses) and also the horizontal grid resolution. If one attempts to calculate the vertical modes at a particular location by solving the appropriate Sturm-Liouville problem the vertical modes are dependent on water column depth and stratification. By searching the model results within a 9-point instrument neighborhood we allow for small changes in water column depth and stratification that may influence the vertical modes produced within the model. Other possible misrepresentations exist such as the exact location of the Gulf Stream and other prominent currents as well as the representation of

meteorological forces that may influence upwelling and downwelling. A closer inspection of individual regions is required to determine the cause of differences between the model and observations.

[83] Our skill tests are intended to serve as a general guide to overall model performance. A more detailed assessment of each individual region is possible but such a task goes beyond the scope of the present work. With the results presented here it is possible to determine those locations where model skill is high and also those regions where model skill is low. This assessment can be done at the instrument level, mooring level and also on a regional and basin scale.

[84] Some of the biggest challenges in assessing model skill for global ocean models are the computational expense and data sizes involved. This is especially true when these considerations make it impossible to execute the model or manage the data to cover the full range of dates from which the observations are taken. For this study we rely on statistical inference in order to derive as much information as possible from the available model output and observations to assess the model performance. Although direct comparison of the model to instrument values provides success rates of only 20%–40% for the North Atlantic basin we find some regions in the ocean where the success rates are 40%–60%. Overall model performance based upon root mean square errors and correlations indicate that HYCOM is able to replicate the tidal currents for many regions of the North Atlantic.

[85] **Acknowledgments.** We would like to thank those individuals involved in the review of this manuscript. They provided many helpful and useful comments leading to significant improvements in both the presentation and analysis of this paper. PGT and BKA acknowledge support from a University of Texas Jackson School of Geosciences Development grant, Naval Research Laboratory (NRL) contract N000173-06-2-C003, Office of Naval Research grants N00014-09-1-1003 and N00014-11-1-0487, and National Science Foundation grant OCE-0924481. RBS acknowledges support from NSF grants OCE-0960834 and OCE-0851457, a contract with the National Oceanography Centre, Southampton, and a NASA subcontract to Boston University. This work was supported in part by a grant of computer time from the DOD High Performance Computing Modernization Program at the Navy DSRC. JGR, EJM, and AJW acknowledge support from the Office of Naval Research (ONR) through the following project: ONR 6.2 Eddy Resolving Global Ocean Prediction Including Tides. This is contribution NRL/JA/7320-11-0889 and has been approved for public release. This is UTIG contribution 2453.

## References

- Arbic, B. K., S. T. Garner, R. W. Hallberg, and H. L. Simmons (2004), The accuracy of surface elevations in forward global barotropic and baroclinic tide models, *Deep Sea Res., Part II*, 51, 3069–3101, doi:10.1016/j.dsr2.2004.09.014.
- Arbic, B. K., A. J. Wallcraft, and E. J. Metzger (2010), Concurrent simulation of the eddy general circulation and tides in a global ocean model, *Ocean Modell.*, 32, 175–187, doi:10.1016/j.ocemod.2010.01.007.
- Best, D. J., and D. E. Roberts (1975), Algorithm AS 89: The upper tail probabilities of Spearman's rho, *Appl. Stat.*, 24, 377–379, doi:10.2307/2347111.
- Chassignet, E. P., H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck (2007), The HYCOM (Hybrid Coordinate Ocean Model) data assimilative system, *J. Mar. Syst.*, 65, 60–83, doi:10.1016/j.jmarsys.2005.09.016.
- 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:EIMOB>2.0.CO;2.
- Egbert, G. D., A. F. Bennett, and M. G. Foreman (1994), TOPEX/POSEIDON tides estimated using a global inverse model, *J. Geophys. Res.*, 99, 24,821–24,852, doi:10.1029/94JC01894.

- Foreman, M. G. G. (2004), Manual for tidal currents analysis and prediction, *Pac. Mar. Sci. Rep.* 78-6, 65 pp., Inst. of Ocean Sci., Sidney, B. C., Canada.
- Foreman, M. G. G., and R. F. Henry (1989), The harmonic analysis of tidal model time series, *Adv. Water Resour.*, 12, 109–120, doi:10.1016/0309-1708(89)90017-1.
- Gibbons, J. D. (1985), *Nonparametric Statistical Inference*, 2nd ed., 408 pp., M. Dekker, New York.
- Godin, G. (1972), *The Analysis of Tides*, Univ. of Toronto Press, Toronto, Ont., Canada.
- Hendershott, M. C. (1972), The effects of solid earth deformation on global ocean tides, *Geophys. J. R. Astron. Soc.*, 29, 389–402, doi:10.1111/j.1365-246X.1972.tb06167.x.
- 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, *Dyn. Atmos. Oceans*, 50, 275–300, doi:10.1016/j.dynatmoce.2010.04.002.
- Pawlowicz, R., B. Beardsley, and S. Lentz (2002), Classical tidal harmonic analysis including error estimates in MATLAB using T-TIDE, *Comput. Geosci.*, 28, 929–937, doi:10.1016/S0098-3004(02)00013-4.
- Ray, R. D. (1998), Ocean self-attraction and loading in numerical tidal models, *Mar. Geod.*, 21, 181–192, doi:10.1080/01490419809388134.
- Ray, R. D., and G. T. Mitchum (1996), Surface manifestation of internal tides generated near Hawaii, *Geophys. Res. Lett.*, 23, 2101–2104, doi:10.1029/96GL02050.
- Ray, R. D., and G. T. Mitchum (1997), Surface manifestation of internal tides in the deep ocean: Observations from altimetry and tide gauges, *Prog. Oceanogr.*, 40, 135–162, doi:10.1016/S0079-6611(97)00025-6.
- Rosmond, T. E., J. Teixeira, M. Peng, T. F. Hogan, and R. Pauley (2002), Navy Operational Global Atmospheric Prediction System (NOGAPS): Forcing for ocean models, *Oceanography*, 15, 99–108, doi:10.5670/oceanog.2002.40.
- Scott, R. B., B. K. Arbic, E. P. Chassignet, A. C. Coward, M. Maltrud, W. J. Merryfield, A. Srinivasan, and A. Varghese (2010), Total kinetic energy in four global eddying ocean circulation models and over 5000 current meter records, *Ocean Modell.*, 32, 157–169, doi:10.1016/j.ocemod.2010.01.005.
- Shum, C. K., et al. (1997), Accuracy assessment of recent ocean tide models, *J. Geophys. Res.*, 102, 25,173–25,194, doi:10.1029/97JC00445.
- von Storch, H., and F. W. Zwiers (2003), *Statistical Analysis in Climate Research*, 484 pp., Cambridge Univ. Press, Cambridge, U. K.
- Wunsch, C. (1997), The vertical partition of oceanic horizontal kinetic energy and the spectrum of global variability, *J. Phys. Oceanogr.*, 27, 1770–1794, doi:10.1175/1520-0485(1997)027<1770:TVPOOH>2.0.CO;2.