Detection of Fronts as a Metric for Numerical Model Accuracy

ELIZABETH M. DOUGLASS

Naval Research Laboratory, Stennis Space Center, Mississippi

ANDREA C. MASK

Naval Oceanographic Office, Stennis Space Center, Mississippi

(Manuscript received 27 June 2018, in final form 22 January 2019)

ABSTRACT

As numerical modeling advances, quantitative metrics are necessary to determine whether the model output accurately represents the observed ocean. Here, a metric is developed based on whether a model places oceanic fronts in the proper location. Fronts are observed and assessed directly from along-track satellite altimetry. Numerical model output is then interpolated to the locations of the along-track data, and fronts are detected in the model output. Scores are determined from the percentage of observed fronts correctly simulated in the model and from the percentage of modeled fronts confirmed by observations. These scores depend on certain parameters such as the minimum size of a front, which will be shown to be geographically dependent. An analysis of two models, the Hybrid Coordinate Ocean Model (HYCOM) and the Navy Coastal Ocean Model (NCOM), is presented as an example of how this metric might be applied and interpreted. In this example, scores are found to be relatively stable in time, but strongly dependent on the mesoscale variability in the region of interest. In all cases, the metric indicates that there are more observed fronts not found in the models than there are modeled fronts missing from observations. In addition to the score itself, the analysis demonstrates that modeled fronts have smaller amplitude and are less steep than observed fronts.

1. Introduction and background

With the advancement of ocean modeling and the proliferation of high-resolution models comes a need for standardized methods of evaluating model output and determining its accuracy. Metrics such as mean error, root-mean-square error, normalized bias, or normalized standard deviation present an objective test of whether a model is accurately representing observations (e.g., Kara and Hurlburt 2006; Stopa and Cheung 2014). These metrics measure the variability averaged over time and space. For example, standard deviation of the sea surface height (SSH) or temperature over a period of time lets the user know whether the model shows roughly the proper magnitude of variability, while a time series of root-mean-square difference between model temperature and observations can demonstrate the presence of drift. While these metrics give a valuable generalized view into biases and variability of model

output, for operational purposes, they may be insufficient. In an operational situation, the user may need accurate knowledge of the precise location of mesoscale features, such as the edge of the Gulf Stream, or the location of a warm-core eddy. These features, and the currents and subsurface structures associated with them, are important aspects of the environment in which the user is operating. Since many ocean models now have the resolution necessary to resolve eddies and other mesoscale features, it is important to understand whether these features are present in the proper locations. Having an eddy field that is statistically comparable to measurements over a period of time is a different challenge than placing a specific mesoscale feature in the right location on the right day, as well as subsequently maintaining the feature on days when it may not be directly observed, such that it is still present when the next observation is available. Operational forecasts assimilate all available data in order to meet the need for accurate, high-resolution regional forecasts with mesoscale features such as eddies and fronts located where they should be; metrics to evaluate these forecasts with

Corresponding author: Elizabeth M. Douglass, elizabeth.douglass@nrlssc.navy.mil

regard to specific feature placement are a useful addition to the assessments already in place.

Many algorithms have been developed to locate ocean features. For example, finding and tracking mesoscale eddies is a well-researched subject (Chelton et al. 2007, 2011; Chaigneau et al. 2008; Beron-Vera et al. 2013). Determining the specific location of western boundary currents such as the Gulf Stream or the Kuroshio is another topic of interest (Qiu 1994; Kelly 1991; Dong et al. 2006; Sokolov and Rintoul 2009). Finding fronts in mapped satellite data such as sea surface temperature or ocean color has been explored as well (Belkin 2009; Cayula and Cornillon 1992, 1995; Kazmin and Rienecker 1996). These algorithms are useful identifying features, describing their characteristics, in some cases tracking their formation, propagation and dissipation. Using these algorithms, the mesoscale variability in a region can be characterized. Algorithms that focus on specific features are less common. For example, Ziegeler et al. (2012) uses a method that determines the displacement of features relative to a mapped product. Zhu et al. (2016) looks at distribution of SST fronts but does not define a quantitative metric for evaluation of their accuracy. However, these methods and metrics are not appropriate for application to altimetric measurements. Currently, satellite altimetry is a nadir measurement, and thus the observations provide a line of data rather than a two-dimensional map or image. While these methods could be applied if the altimetry were mapped to a two-dimensional grid, the results will rely to a certain extent on the assumptions inherent in the mapping process. To simplify the process and to avoid the necessity for such assumptions, the metric developed here will be limited to along-track altimetric measurements rather than a gridded product created from those measurements.

Along-track altimetric measurements (rather than gridded or interpolated products) have high resolution and can distinguish abrupt changes in SSH that are generally indicative of distinct changes in subsurface structure. These fronts are accompanied by a density gradient and the associated geostrophic flow. In some cases, such as the location of the Gulf Stream or the Kuroshio, the presence of a large front and associated current is known, but the precise location can be pinpointed using altimetry. Other fronts may arise from the passage of mesoscale eddies or other transient features. The repeating orbits of the altimetric satellites allow features to be measured more than once while they persist, and the movement of coherent features like eddies can be tracked in time. Using these data to verify and validate model output in a quantifiable manner is the goal of the present manuscript.

2. Description of data

As noted, the data used here are SSH anomalies (the difference between a satellite observation and the long-term mean sea surface height) as measured by satellite altimeters. These data are taken from the Ocean QC dataset (Cummings 2011), which is composed of all data used in the Navy Coupled Ocean Data Assimilation (NCODA) data assimilating system (Cummings and Smedstad 2013). Specifically, we use SSH anomaly as measured by the *Jason-2* altimeter and the *Satellite with Argos Data Collection System (Argos)* and *Ka-Band Altimeter (AltiKa; SARAL)*. These data are assimilated operationally by the Naval Oceanographic Office (NAVOCEANO), in both models described below. Resolution is relatively high, with along-track spacing of roughly 5–6 km. *Jason-2* has a repeat cycle of roughly 10 days with 127 revolutions per cycle. *AltiKa* has a repeat cycle of 35 days with 501 revolutions per cycle. The ground tracks are interleaved for best coverage.

To ensure that observed sea surface height anomaly (SSHA) is comparable to modeled SSH, the mean SSH from a 20-yr Hybrid Coordinate Ocean Model (HYCOM) reanalysis is added to the altimetry. The details of the reanalysis are summarized in Yu et al. (2015). The choice was made to add a mean to the data rather than subtracting a mean from the model due to the operational nature of the results being tested. The anomaly + mean product most closely represents the environment in which users are operating.

3. Description of models

The method described in this manuscript will be used to compare the performance of two models used by the NAVOCEANO to provide operational forecasts. The first is NCOM, the Navy Coastal Ocean Model (Rowley and Mask 2014; Barron et al. 2006; Martin 2000). NCOM is a primitive equation, free surface model. The horizontal resolution, in the cases used here, is roughly 3 km. Its vertical structure has sigma-levels near the surface and z levels at depth. It uses three-dimensional variational assimilation (3DVAR) to assimilate NCODA data mentioned above, which includes but is not limited to the altimetric height measurements. NCOM is used by NAVOCEANO for regional forecasts. As such, boundary conditions are necessary, and these are provided by NAVOCEANO's global operational model, HYCOM (Metzger et al. 2014). HYCOM is the other model used in this analysis. It is also a primitive-equation, free-surface model. It has horizontal resolution of $1/12.5^\circ$ (nominally 9 km at the equator) and 41

FIG. 1. Snapshot of model SSH for the western Pacific and the Mediterranean Sea on 30 Apr 2017.

vertical levels. The hybrid vertical coordinate, which gives the model its name, can be either isopycnal levels that follow density, depth levels, or sigma levels that follow terrain, and can adjust as the model runs. Like NCOM, 3DVAR is used to assimilate the NCODA data.

Since NCOM is a regional model, analysis will be done in two of the regions for which data are operationally produced. Figure 1 shows snapshots of SSH in the western Pacific and Mediterranean regions for both HYCOM and NCOM. The day shown is 30 April 2017. The mean SSH (which will be added to the data to make them comparable) is also shown. Black boxes show the regions where the metric was calculated. Gray regions including the Black Sea and part of the Atlantic Ocean are excluded from the NCOM Mediterranean Sea region and are masked for this calculation. Snapshots

show that even though NCOM and HYCOM assimilate the same data using the same assimilation method, and have generally similar model descriptions, the results have distinct differences.

4. Detection and matching of fronts

The first step in frontal detection is to determine the area and time period of interest. In the present analysis, one 24-h day was examined at a time. Due to the orbits of satellites, the precise number and location of data points vary from day to day. Once the data for a given day are extracted and the model mean is added as described above, the data are smoothed using a running mean over 15 along-track points, or approximately 85 km (see section 7 for more discussion of smoothing scale).

Differentiation amplifies noise, so smoothing must be applied to get meaningful results. Using these smoothed data, the gradient is calculated at each point by dividing the change in SSH by the along-track distance between locations. The gradients are then smoothed, again using a 15-point running mean. Finally, each location where the slope indicates a change of more than $\Delta\eta$ per 100 km is designated as a front. Parameter $\Delta\eta$ is the minimum amplitude change necessary for a feature to be designated as a front.

The metric is sensitive to the choice of threshold, and thus it should be made in a methodical and quantitative way. Thus, the region where the metric is applied is divided into boxes of 1° longitude $\times$ 1° latitude. For 1 year of data, the slope is recorded each time an altimetric track passes through each box. Thus, we obtain the distribution of altimetric slopes over a 1-yr time period at each location. The mean and standard deviation of the slope in each 1° box are determined. At each location, the distribution is approximately Gaussian (not shown). Given this distribution, determining whether the slope on a given day is more than one standard deviation away from the mean is a meaningful measure of whether a gradient is statistically large. The maps of standard deviation of the gradient σ_{gradient} , for the western Pacific and Mediterranean regions, are shown in Fig. 2. There is a large amount of spatial variation in this measure, with high values close to Japan where the Kuroshio passes through, and lower values farther from the western boundary current region. Note the different color scales in Figs. 2a and 2b; the western Pacific color scale saturates at a standard deviation of 0.35 cm km^{-1} while the maximum in the Mediterranean Sea is only 0.1 cm km^{-1} .

Developing a map of σ_{gradient} as described above is an a priori requirement of the metric calculation. At each time step, when the model output is interpolated to the location of the altimetric track, the 1° maps of σ_{gradient} and the mean of the gradient are also interpolated to the exact track location, to find both the mean gradient of sea surface height in that location, and the standard deviation. Then, at each point, a gradient is found to be a “front” if it is more than one standard deviation away from the mean. In other words, a gradient is identified as a front if it is larger than 68% of gradients that occur in a year at that location.

Once a front is identified, several characteristics are determined. The two most important are the dimensions by which the front is defined, the horizontal extent, and the vertical magnitude. In determining horizontal extent, first, the number of consecutive points where slope exceeds the threshold is determined. To take the smoothing into account, half of the smoothing distance

FIG. 2. Standard deviation of SSH gradients (cm km^{-1}) in (a) the western Pacific and (b) the Mediterranean Sea.

is added to each side. Thus, if the slope exceeds the limiting value at only 1 point, then the front is defined as spanning 15 points, 7 points prior to the location of the limiting value, the location of the limiting value itself, and 7 points after it. Once the horizontal extent has been determined, frontal magnitude is defined as the difference between maximum and minimum SSH within the along-track limits of the front.

The process described above is the same for model output as it is for data. Model output is interpolated to the locations of the observations, and then the same method of smoothing, differentiating, smoothing, and finding locations where the smoothed gradient exceeds the frontal threshold is followed again. The size and magnitude of the fronts are determined. At this point, two sets of fronts have been defined, one for the data and one for the model.

Figure 3 shows an example of all the steps described above. Figure 3a shows HYCOM output, interpolated to the locations of the altimetric data, on 26 April 2017. The data shown are in the western Pacific Ocean. One segment of the data is highlighted in Fig. 3a; that is, the segment for which results are shown in Figs. 3b–d. Figure 3b shows the altimetry for this track segment, in its original form (green) and after smoothing (red). Model output for the same segment is shown in blue; the unsmoothed model output is indistinguishable from the smoothed result. In Fig. 3c, the gradient of both the altimetry and model are shown, before and after

FIG. 3. Example of identifying and measuring a front. (a) Along-track HYCOM, with a box around an example segment. (b) HYCOM SSH (blue), altimetric SSH (green), and smoothed altimetric SSH (red). (c) Gradients of unsmoothed HYCOM SSH (black), smoothed HYCOM SSH (blue), unsmoothed altimetric SSH (green), and smoothed altimetric SSH (red). Gray dashed lines show the feature threshold of σ_{gradient} . (d) HYCOM SSH with the five fronts identified by a gradient larger than σ_{gradient} labeled. Vertical and horizontal lines show observed vertical extent (magnitude) and horizontal extent (size) of each front (labeled as $\Delta\eta$ and Δx).

smoothing. The gray dashed line indicates the threshold of σ_{gradient} used to determine the presence of a front. At the beginning of the segment, the threshold is high; near the location of the Kuroshio, as seen in Fig. 2a, gradients are higher. In this case, the gradient of the model SSH is just higher than the threshold to be identified as a front, and the gradient of altimetry is below the threshold. Farther along the segment, the threshold decreases. About 400 km from the start of the segment, the model gradient of SSH again exceeded the threshold for a second time, and a front is identified. At 500-km along-track distance, both data and model exceed the threshold, clearly identifying a

front in both model and observations. At around 700 and 900 km, two more fronts are identified in both modeled and observed SSH. Figure 3d, shows the same blue line as in Fig. 3b, with the five modeled fronts and their dimensions marked. As shown, the magnitude $\Delta\eta$ is the vertical rise of the front while the size Δx is the along-track width.

Once all fronts in both the model and the data have been located, the algorithm determines whether a given observed front is “matched” by a front in the model output. Figure 4 shows an example of frontal matching from 26 April 2017. For two fronts to be considered “matched,” the primary consideration is proximity. If

FIG. 4. Example of matching fronts on 26 Apr 2017. (a) All fronts identified in altimetry; matched fronts are circled. (b) All fronts identified in HYCOM, with matched fronts circled.

the center of a HYCOM front is within the horizontal extent of an altimetric front, and the direction of the fronts is consistent, the two are found to be matched. Note that in this example, there are more unmatched altimetric fronts than HYCOM fronts.

5. Description of metric

Two scores will be used to evaluate the model representation of fronts. The first metric, R1, is a measure of how well the ocean model represents those fronts observed using along-track satellite altimetry. As noted, if the frontal detection algorithm finds a front in the same location and with the same orientation in both the model output and the along-track altimetry, this is a “matched” front. R1 is defined as the number of matched fronts divided by the total number of altimetric fronts. A perfect model would correctly show every observed front, for a score of 1; a model product representing only half of the observed fronts would have a score of 0.5. The second metric, R2, describes the reliability of the model. That is, how confident can the user be that a front predicted by the model will be observed by the altimetry? R2 is defined as the number of matched fronts divided by the number of modeled fronts. A perfect model would find an observed match for every front, for a score of 1. A model that predicted twice as many fronts as are observed would have an R2 score of 0.5.

To get a sense of this metric, note that for the example shown in Figs. 3a and 4, R1 is equal to 0.63, and R2 is equal to 0.70. Of the 11 observed altimetric fronts, 7 appear in the model output, and 7 of the 10 fronts identified in the model are also observed. As noted previously, there are more unmatched altimetric fronts than HYCOM fronts, and thus R1 is smaller than R2. These metrics provide the user with a sense of how likely the model is to replicate an observed front, and the likelihood that a modeled front is also found in observations.

6. Examples

To understand these metrics, they will be applied to HYCOM and NCOM in the western Pacific Ocean and the Mediterranean Sea over a 1-yr period. The scores are calculated every day from 12 September 2016 to 11 September 2017. This will demonstrate the variability of scores in time, as well as how scores vary according to the details of the region in which they are calculated. Additionally, scores are shown for $\Delta\eta$ of σ_{gradient} , $1.5\sigma_{\text{gradient}}$, and $2\sigma_{\text{gradient}}$. The example in Fig. 3 used σ_{gradient} , indicating that a gradient is larger than 68% of gradients at that location. The higher thresholds of $1.5\sigma_{\text{gradient}}$ and $2\sigma_{\text{gradient}}$ indicate that a gradient is higher than 86% and 95% of local gradients, respectively. This will demonstrate the effect of more stringent thresholds on the metric scores.

a. Western Pacific

In the western Pacific, fronts are likely to be strong and well defined. First and foremost, the Kuroshio is the western boundary current of the North Pacific Ocean. It is a strong current with a large signal in SSH. The Kuroshio Extension region, to the east of Japan where the Kuroshio extends into the open ocean, is a region of high eddy activity; eddies will produce large frontal signals (Qiu 2002). This is clear in Figs. 1b and 1c, with the very clear front associated with the Kuroshio and other features such as cold-core eddies evident as well. Thus in this region, one would expect to observe large fronts, but would need a model with skill to properly predict the location.

Both metrics for both models are shown in Fig. 5. Average values are listed in Table 1. Scores are shown for $\Delta\eta$ of σ_{gradient} , $1.5\sigma_{\text{gradient}}$, and $2\sigma_{\text{gradient}}$. Figure 5 shows that scores are highly variable. In the most restrictive case, with fronts limited to $2\sigma_{\text{gradient}}$, both R1 and R2 are likely to decrease to 0 or increase to 1, indicating days in which few to no outlier fronts are located in the model or observations. A single outlier front found in both model and output would give $R1 = 1$ and

FIG. 5. R1 (black) and R2 (red) scores in the western Pacific for HYCOM (*) and NCOM (○). Time series from September 2016 to September 2017. Scores are shown for thresholds of (a) σ_{gradient} , (b) $1.5\sigma_{\text{gradient}}$, and (c) $2\sigma_{\text{gradient}}$.

R2 = 1; if that front were slightly below the threshold in the model, for example, the result would be R1 = 0 and R2 = undefined. On the other hand, with the smaller frontal threshold of σ_{gradient} , scores are consistently above 0.5. R2 scores are higher than R1 scores in all cases. Finally, the results for HYCOM R2 have a slight but consistent advantage over NCOM.

Figure 6 shows the representativeness and reliability of each model as a function of front magnitude. For a given magnitude of front, this shows the likelihood that an observed front is modeled (R1), and the likelihood that a modeled front is observed (R2). Uncertainty is highest for the smallest fronts, with R1 and R2 both

relatively low, and decreases with increasing front size. In general, as shown in Table 1, R2 scores are higher than R1. However, this shows that these differences are most evident for smaller fronts, while when fronts are

TABLE 1. Western Pacific R1 and R2.

Model	$\Delta\eta$	R1	R2
HYCOM	$1.0\sigma_{\text{gradient}}$	0.71	0.78
	$1.5\sigma_{\text{gradient}}$	0.63	0.71
	$2.0\sigma_{\text{gradient}}$	0.50	0.60
NCOM	$1.0\sigma_{\text{gradient}}$	0.68	0.74
	$1.5\sigma_{\text{gradient}}$	0.59	0.66
	$2.0\sigma_{\text{gradient}}$	0.47	0.55

FIG. 6. The representativity (R1) and reliability (R2) of the HYCOM and NCOM models as a function of the magnitude of the fronts for the western Pacific.

larger, R1 and R2 scores are more similar. HYCOM R2 is higher than all other scores for almost every front size. NCOM R2 is higher than R1 for smaller fronts, but for fronts larger than about 0.4 m, NCOM R2 scores are very consistent with both R1 scores. All scores are above 0.7 for fronts larger than 0.4 m. HYCOM R1 is noticeably better than NCOM R1 for large fronts of 50–70 cm, but both are still relatively representative. HYCOM scores slightly exceed NCOM scores in both R1 and R2 in all sizes larger than 0.15 m (sizes at which neither model is particularly skilled).

b. Mediterranean Sea

The Mediterranean Sea presents a very different environment. Fronts are smaller and more variable, as is clear in Figs. 1e and 1f. Gradients are smaller and features are less well defined than in the western Pacific region. Both metrics for both models are shown in Fig. 7. As for the western Pacific, a threshold of $2\sigma_{\text{gradient}}$ is too high, and leads to a high number of days with values of either 0 or 1 in R1 and R2 (Fig. 7c). In most cases, these indicate the presence of only one or two fronts, in altimetry or model output. Table 2 shows the average skill for each of the three thresholds tested.

The metrics indicate that both NCOM and HYCOM are less reliable in the Mediterranean Sea than in the

western Pacific. Even in the least restrictive case, R1 does not exceed 0.55 and R2 does not exceed 0.65. When the frontal threshold is greater than σ_{gradient} , R1 and R2 are both below 0.5 except in the case of HYCOM at a threshold of $1.5\sigma_{\text{gradient}}$, indicating that more often than not, fronts are not paired for either model or observations. It is also interesting to note that this is not solely a function of the number of fronts found by the metric. In the western Pacific, using a threshold of σ_{gradient} , an average of 18 HYCOM fronts and 20 altimetric fronts are found per day. In the Mediterranean Sea, the average number of fronts is almost the same, 16 HYCOM fronts and 20 altimetric fronts. However, the average R1 and R2 for the western Pacific at 0.71 and 0.78, respectively, are higher than the average R1 and R2 of the Mediterranean Sea at 0.53 and 0.65, respectively. Figure 8 shows that just as in the western Pacific, accuracy increases with the magnitude of the fronts being measured. HYCOM R2 is noticeably higher than all other scores in all categories. HYCOM R1, NCOM R1, and NCOM R2 are generally consistent, although HYCOM R1 does usually exceed NCOM R1. It is also very evident in this figure that for frontal magnitudes of 50 cm or higher, the metrics lose efficacy due to small numbers of fronts. Over the full 365 days of the time series, there are 73 altimetric fronts of more than 40 cm, 45 HYCOM fronts,

FIG. 7. R1 (black) and R2 (red) scores in the Mediterranean Sea for HYCOM (*) and NCOM (○). Time series from September 2016 to September 2017. Scores are shown for thresholds of (a) σ_{gradient} , (b) $1.5\sigma_{\text{gradient}}$, and (c) $2\sigma_{\text{gradient}}$.

and 35 NCOM fronts. These numbers are too small to make meaningful assessments of the results, beyond realizing that such large fronts are very rare in this region.

7. Discussion

It is helpful to compare this metric to other possible metrics of model evaluation to provide some perspective. Root-mean-square difference (RMSD) between the model output and observations provides an estimate of “how different” the model is from the data. For this metric, smaller is better: a “perfect”

model would have RMSD of zero from the data. It is worth noting that for any metric, the dynamics of the model are only properly tested by comparing to data that are not assimilated into the model; otherwise, as in the examples in this manuscript, the test measures the assimilation method rather than the dynamical “perfection” of the model. Regardless, the intent here is only a demonstration of method of comparing model output to observations. RMSD between the gradients of the collocated model and along-track data is calculated for each model in both regions (Table 3). Using this as a metric, the performance of HYCOM in the western Pacific is worse than its performance in the Mediterranean

TABLE 2. Mediterranean Sea R1 and R2.

Model	$\Delta\eta$	R1	R2
HYCOM	$1.0\sigma_{\text{gradient}}$	0.53	0.65
	$1.5\sigma_{\text{gradient}}$	0.43	0.53
	$2.0\sigma_{\text{gradient}}$	0.32	0.41
NCOM	$1.0\sigma_{\text{gradient}}$	0.48	0.55
	$1.5\sigma_{\text{gradient}}$	0.38	0.43
	$2.0\sigma_{\text{gradient}}$	0.29	0.32

Sea. This is likely a function of the mean gradient: gradients are higher in the western Pacific, so the RMSD is also likely to be higher. However, this is at odds with the frontal metric, which indicated that frontal representation in the western Pacific is better than in the Mediterranean Sea. In both cases, HYCOM has slightly lower RMSD than NCOM, which agrees with the result of the frontal metric. It is also the case that RMSD has an unlimited range and thus the “goodness” or “badness” of the metric is less intuitive.

Another possible metric is the normalized standard deviation, where the standard deviation of the model output is divided by the standard deviation of the data. A “perfect” score here would be 1, with the model and data exhibiting equal variability. In this case, the western Pacific scores are better than the Mediterranean Sea scores, in agreement with the frontal metric. In the Mediterranean Sea, HYCOM is better

than NCOM. Both HYCOM and NCOM have more variability than the observations. This is in line with the frontal metric, where low R2 scores indicate large numbers of model fronts not replicated in the observations.

To analyze these differences in a more concrete, specific way, the example shown in Fig. 3b is expanded in Fig. 9 to include NCOM as well as HYCOM. Figure 9a shows SSH along the highlighted section of Fig. 3a, from altimetry, HYCOM, and NCOM. Figure 9b shows the gradients of NCOM, HYCOM, and the altimetry, and Fig. 9c shows the SSH again, with locations identified as fronts highlighted. Fronts where SSH increases along-track are highlighted in pink, and decreasing fronts are highlighted in blue. This shows the horizontal extent of the fronts as determined by the algorithm. In Fig. 9c, the three lines are offset by 0.2 m so differences in frontal location are clear. At the beginning of the track, both models just barely exceed the threshold for a front, while the observed gradient is not steep enough. At around 300 km along-track, NCOM and HYCOM again find an increase sharp enough to be defined as a front, but the observed gradient is below the threshold, as seen in Fig. 9b. A decrease in SSH centered around 425 km is found by both models and the observations. At 600 km, an increase noted in NCOM is not found in either HYCOM or the altimetry, and a decrease in both HYCOM and altimetry at 750 km is not found by NCOM.

FIG. 8. As in Fig. 6, but for the Mediterranean Sea.

TABLE 3. Alternate metrics: RMSD and normalized standard deviation ($\sigma_{\text{model}}/\sigma_{\text{data}}$).

Model	Region	RMSD	$\sigma_{\text{model}}/\sigma_{\text{data}}$
HYCOM	West Pacific	0.097	0.97
	Mediterranean	0.064	1.18
NCOM	West Pacific	0.11	0.99
	Mediterranean	0.077	1.28

Finally, all three find an increase at around 900-km along-track distance. The total number of fronts is five in HYCOM, five in NCOM, and three in the altimetry. For R1, the number of observed fronts found in the model, HYCOM found three of the three fronts ($R1 = 3/3 = 1.00$) and NCOM found two ($R1 = 2/3 = 0.67$). Of the five HYCOM fronts, all three are found in altimetry, for an $R2 = 3/5 = 0.6$. Two of the five NCOM fronts are found in altimetry, so $R2 = 2/5 = 0.4$. Both R1 and R2 indicate that HYCOM represents frontal structures for this section of altimetry more accurately than NCOM. Looking at other possible metrics, the RMSD between NCOM and altimetry gradients is 1.16, while that between HYCOM and altimetry is 0.94, which also indicates that HYCOM resembles the observations more closely. Normalized standard deviations indicate that both models are more variable than the observations in this location, with $\sigma_{\text{HYCOM}}/\sigma_{\text{data}} = 1.12$ and $\sigma_{\text{NCOM}}/\sigma_{\text{data}} = 1.15$. HYCOM replicates all observed fronts and creates fewer nonobserved fronts than NCOM, and is the better choice for this situation, as indicated by the frontal metrics.

The spatial variability of this analysis can also be considered. Figure 10 shows the average HYCOM score in the western Pacific in $1^\circ \times 1^\circ$ boxes. The color of the circle indicates score, while the size of the circle indicates the number of dates with fronts in that box; thus, large circles indicate scores based on more fronts and therefore have higher confidence. Also shown is the approximate average location of the Kuroshio (the 75-cm contour in the mean SSH map). Close to the front of the Kuroshio itself, circles are large due to the persistence of the phenomenon, and confidence is high, indicating that the model generally puts the Kuroshio in the right place. Farther from the front, circles are smaller, possibly indicating transient events like the passing of an eddy. Scores are still generally good. Figure 11 shows the same information for the Mediterranean Sea. There is nothing in the Mediterranean Sea analogous to the Kuroshio in the western Pacific. Circles are large, indicating many mesoscale features appear throughout the region. In the central Mediterranean Sea, large circles with low scores (green and blue) are found, indicating high numbers of fronts

FIG. 9. Front-finding for one section for altimetry and both models. (a) SSH for HYCOM, NCOM, and altimetry. (b) Gradient of SSH for HYCOM, NCOM, and altimetry. (c) SSH for HYCOM, NCOM, and altimetry, with a 0.2-m offset to more easily distinguish between the three lines, with locations of fronts highlighted in blue where SSH gradients are negative and in red where SSH gradients are positive.

with low accuracy. In the southwest Mediterranean Sea there are several locations with large red circles, indicating that the model is more successful in this location. The dynamics of the local regions are clearly at play; these issues indicate that further analysis is necessary in these regions to determine what the model is depicting accurately and what it is missing that leads to such a high “miss rate” in certain areas.

While the simplicity of a metric with one value between 0 and 1 was the primary aim of this analysis, the information about frontal characteristics gathered in pursuit of this aim can yield dividends as well. In the process of identifying fronts, the vertical magnitude and horizontal extents of fronts are recorded as well. Histograms of these frontal characteristics in different areas can help illuminate the source of the disparities between modeled and observed results. The histograms of magnitude (Figs. 12a,b) indicate that in both the Mediterranean Sea and the western Pacific, the modeled magnitudes tend to be slightly smaller; the peak in modeled magnitude is lower than the peak in observed magnitude. The histograms of size are broader than the histograms of magnitude, and in distributions of modeled and observed sizes are nearly

FIG. 10. Map of western Pacific, showing the HYCOM score in each location. Color of the circle shows score, and size shows the number of dates when fronts exist for comparison.

identical in both the western Pacific (Fig. 12c), and the Mediterranean Sea (Fig. 12d). Combining these two attributes, Figs. 12e and 12f show the slope of fronts, calculated as magnitude divided by size. This indicates that in both the western Pacific and the Mediterranean Sea, observed fronts are steeper than modeled fronts, using HYCOM or NCOM. It is interesting that these results are consistent between basins, even though the models' ability to replicate fronts varies with location. While a full analysis of this effect is beyond the scope of the present investigation, it provides a direction for future research into understanding the deficiencies in modeling mesoscale ocean features.

The similarity between the HYCOM and altimetric distribution of feature size (the horizontal extent of features, as measured by the distance with a slope higher than the critical value) also provides some context for the choice of 15 points (85 km) as a scale for smoothing. Initially, this value was chosen because 85 km is a reasonable scale at which to elucidate mesoscale features. Two other smoothing scales were also tested, 9 points (or about 50 km) and 21 points (about 120 km). The effects on the number of features and the metric scores were as expected. When the shorter smoothing scale was used, more fronts were found (because the inputs were noisier) and scores were lower. When the longer smoothing scale was used, fewer fronts were found and scores were higher. A more interesting comparison can

FIG. 11. As in Fig. 10, but for the Mediterranean Sea.

be made between the histograms of size for these three smoothing scales. While the absolute numbers of features found changed, for the HYCOM features, the normalized size distribution peaked at about 160 km for a smoothing scale of 50 km, with a slightly flatter peak of 160–200 km for both longer smoothing scales of 85 and 120 km (Fig. 13a). For the altimetric features, the size distribution for 50 km was sharp and peaked at 120 km. The distribution for the 85-km smoothing scale most resembled that from HYCOM, with a slightly flatter peak around 160–200 km. The peak for the longer smoothing scale was the broadest, with a maximum at 200 km but a distribution clearly shifted toward longer lengths (Fig. 13b). Since the situation calls for determining which features in observations are mirrored in the model, it seems appropriate to use a smoothing scale that produces features with roughly the same distribution of sizes.

8. Conclusions

We have defined a metric for evaluation of fronts in SSH in a model product as compared to those measured in satellite altimetry. This metric can be applied to any region (globally or locally) and provides estimates of both the representativeness and the reliability of the model, so that the user knows whether observed fronts are likely to be represented correctly in the model and whether model-predicted fronts are likely

FIG. 12. (top) Histograms of magnitudes of observed and modeled fronts in the (a) western Pacific and (b) Mediterranean Sea. (middle) Histograms of sizes of observed and modeled fronts in the (c) western Pacific and (d) Mediterranean Sea. (bottom) Histograms of slopes of observed and modeled fronts in the (e) western Pacific and (f) Mediterranean Sea.

to be present in the operational environment. Examples in the western Pacific and the Mediterranean Sea demonstrate the uses of the metric, as well as its shortcomings.

The intent of this manuscript is not to compare NCOM with HYCOM and declare a “winner.” Comparisons between the two are an illustration of the metric and its attributes. Moreover, for the most part NCOM and HYCOM are functionally equivalent in this regard; the results show slight differences, but these are unlikely to be significant. On the other hand, what can be said is that attention should be paid to the inability of

either model to properly describe small magnitude, transient features that make up most of the mesoscale environment in the Mediterranean Sea. It is also important to note that the data to which the model output was compared was assimilated by both models, and thus the metric does not truly test the skill of the model dynamics. Such a test would require data that were not assimilated; either some altimetry could be withheld from the assimilation, or a different (independent) dataset could be used. An assessment such as this is beyond the scope of the present manuscript, as the intent here is to present a method for comparing the

FIG. 13. Histograms of sizes for different smoothing scales in (a) HYCOM and (b) altimetry.

frontal features on two maps, not to fully assess or compare NCOM or HYCOM dynamics.

There are certainly limitations to this metric. To start with, the identification of fronts is limited by their location relative to the satellite track. For example, fronts are identified only if they are normal to the satellite track. If the satellite track were to be parallel to a front, it could have no signature at all; in any case, the assessment of the abruptness and steepness of a front will be mitigated by its angle relative to the satellite track. In many cases, the spacing of along-track data means some fronts will be missed entirely; the coverage does not allow us to “see” every front on every day. Additionally, as presently defined, this metric does not give any information regarding the comparative magnitude of observed and modeled fronts; it is limited to determining whether fronts are present or not present. This metric presents a simple, straightforward way to assess whether or not the mesoscale features in a model are in the same place as those that are observed, on a day to day basis. In our example, it is clear that both models do a relatively good job of placing the Kuroshio in the right place, but that the transient features that compose the mesoscale field in the Mediterranean Sea are a more difficult task. This provides both useful information to the user and a point to focus on for the model developer, and can lead to advancements for both in the future.

Acknowledgments. This work was funded by the Naval Oceanographic Office Oceanographic Department.

We thank two anonymous reviewers for comments which improved the manuscript.

REFERENCES

- Barron, C. N., A. B. Kara, P. J. Martin, R. C. Rhodes, and L. F. Smedstad, 2006: Formulation, implementation and examination of vertical coordinate choices in the Global Navy Coastal Ocean Model (NCOM). *Ocean Modell.*, **11**, 347–375, <https://doi.org/10.1016/j.ocemod.2005.01.004>.
- Belkin, I. M., 2009: Observational studies of oceanic fronts. *J. Mar. Syst.*, **78**, 317–318, <https://doi.org/10.1016/j.jmarsys.2008.10.016>.
- Beron-Vera, F. J., Y. Wang, M. J. Olascoaga, G. J. Goni, and G. Haller, 2013: Objective detection of oceanic eddies and the Agulhas leakage. *J. Phys. Oceanogr.*, **43**, 1426–1438, <https://doi.org/10.1175/JPO-D-12-0171.1>.
- Cayula, J. F., and P. Cornillon, 1992: Edge-detection algorithm for SST images. *J. Atmos. Oceanic Technol.*, **9**, 67–80, [https://doi.org/10.1175/1520-0426\(1992\)009<0067:EDAFSI>2.0.CO;2](https://doi.org/10.1175/1520-0426(1992)009<0067:EDAFSI>2.0.CO;2).
- , and —, 1995: Multi-image edge detection for SST images. *J. Atmos. Oceanic Technol.*, **12**, 821–829, [https://doi.org/10.1175/1520-0426\(1995\)012<0821:MIEDFS>2.0.CO;2](https://doi.org/10.1175/1520-0426(1995)012<0821:MIEDFS>2.0.CO;2).
- Chaigneau, A., A. Gizolme, and C. Grados, 2008: Mesoscale eddies off Peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns. *Prog. Oceanogr.*, **79**, 106–119, <https://doi.org/10.1016/j.pocean.2008.10.013>.
- Chelton, D. B., M. G. Schlax, R. M. Samelson, and R. A. de Szoeke, 2007: Global observations of large oceanic eddies. *Geophys. Res. Lett.*, **34**, L15606, <https://doi.org/10.1029/2007GL030812>.
- , —, and —, 2011: Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.*, **91**, 167–216, <https://doi.org/10.1016/j.pocean.2011.01.002>.
- Cummings, J., 2011: Ocean data quality control. *Operational Oceanography in the 21st Century*, A. Schiller and G. B. Brassington, Eds., Springer, 91–121.
- , and O. M. Smedstad, 2013: Variational data assimilation for the global ocean. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, Vol. II, S. K. Park and L. Xu, Eds., Springer, 303–343.
- Dong, S., J. Sprintall, and S. Gille, 2006: Location of the Antarctic polar front from AMSR-E satellite sea surface temperature measurements. *J. Phys. Oceanogr.*, **36**, 2075–2089, <https://doi.org/10.1175/JPO2973.1>.
- Kara, A. B., and H. E. Hurlburt, 2006: Daily inter-annual simulations of SST and MLD using atmospherically forced OGCMs: Model evaluation in comparison to buoy time series. *J. Mar. Syst.*, **62**, 95–119, <https://doi.org/10.1016/j.jmarsys.2006.04.004>.
- Kazmin, A. S., and M. M. Rienecker, 1996: Variability and frontogenesis in the large-scale oceanic frontal zones. *J. Geophys. Res.*, **101**, 907–921, <https://doi.org/10.1029/95JC02992>.
- Kelly, K. A., 1991: The meandering Gulf Stream as seen by the Geosat altimeter: Surface transport, position, and velocity variance from 73° to 46°W. *J. Geophys. Res.*, **96**, 16 721–16 738, <https://doi.org/10.1029/91JC01380>.
- Martin, P. J., 2000: Description of the Navy Coastal Ocean Model Version 1.0. Tech. Rep. NRL/FR/7322–00-9962, 42 pp., <https://apps.dtic.mil/dtic/tr/fulltext/u2/a387444.pdf>.
- Metzger, E. J., and Coauthors, 2014: US Navy operational global ocean and Arctic ice prediction systems. *Oceanography*, **27**, 32–43, <https://doi.org/10.5670/oceanog.2014.66>.
- Qiu, B., 1994: Determining the mean Gulf Stream and its recirculations through combining hydrographic and altimetric data. *J. Geophys. Res.*, **99**, 951–962, <https://doi.org/10.1029/93JC03033>.

- , 2002: The Kuroshio Extension system: Its large-scale variability and role in the midlatitude ocean-atmosphere interaction. *J. Oceanogr.*, **58**, 57–75, <https://doi.org/10.1023/A:1015824717293>.
- Rowley, C., and A. Mask, 2014: Regional and coastal prediction with the relocatable ocean nowcast/forecast system. *Oceanography*, **27**, 44–55, <https://doi.org/10.5670/oceanog.2014.67>.
- Sokolov, S., and S. R. Rintoul, 2009: Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 1. Mean circumpolar paths. *J. Geophys. Res.*, **114**, C11018, <https://doi.org/10.1029/2008JC005108>.
- Stopa, J. E., and K. F. Cheung, 2014: Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis. *Ocean Modell.*, **75**, 65–83, <https://doi.org/10.1016/j.ocemod.2013.12.006>.
- Yu, Z., and Coauthors, 2015: Seasonal cycle of volume transport through Kerama Gap revealed by a 20-year global HYbrid Coordinate Ocean Model reanalysis. *Ocean Modell.*, **96**, 203–213, <https://doi.org/10.1016/j.ocemod.2015.10.012>.
- Zhu, X., and Coauthors, 2016: Comparison and validation of global and regional ocean forecasting systems for the South China Sea. *Nat. Hazards Earth Syst. Sci.*, **16**, 1639–1655, <https://doi.org/10.5194/nhess-16-1639-2016>.
- Ziegeler, S. B., J. D. Dykes, and J. F. Shriver, 2012: Spatial error metrics for oceanographic model verification. *J. Atmos. Oceanic Technol.*, **29**, 260–266, <https://doi.org/10.1175/JTECH-D-11-00109.1>.