

## Evaluation of regional numerical weather prediction model surface fields over the Middle Atlantic Bight

Mark F. Baumgartner<sup>1</sup> and Steven P. Anderson

Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

**Abstract.** Coastal ocean models often rely on the surface fields from numerical weather prediction (NWP) models for realistic surface boundary conditions, but the errors in these fields are poorly understood. We evaluate the surface meteorological and flux fields provided by three of the regional NWP models in operation during 1996 and 1997 at the U.S. National Centers for Environmental Prediction (NCEP): the Eta-48, Eta-29, and Rapid Update Cycle (RUC-1) models. These model fields are compared to in situ measurements made from an air-sea interaction buoy deployed from July 1996 to June 1997 at a midshelf location in the Middle Atlantic Bight during the Coastal Mixing and Optics experiment. In addition, data from six National Data Buoy Center buoys are used to evaluate spatial errors in the model fields. The Eta-29 and RUC-1 models overestimate the net ocean-to-atmosphere heat flux by an average 83 and 74  $\text{W m}^{-2}$ , respectively, with notable errors in each of the individual heat flux components. The poorly resolved sea surface temperature fields used in the 1996–1997 regional NWP models lead to significant errors in the latent and sensible heat fluxes over the continental shelf and slope. Moreover, wind speeds are slightly overestimated in the Eta-48 and Eta-29 models while the RUC-1 model underestimates them by more than  $1 \text{ m s}^{-1}$ . All of the models have mean wind direction errors of  $7^\circ$  to  $13^\circ$  east of north. In light of these evaluations, considerations for improving the accuracy of the surface flux fields for use in future ocean modeling studies are discussed.

### 1. Introduction

Oceanic variability on the continental shelf is closely tied to temporal and spatial variability in atmospheric forcing. To accurately model or forecast the coastal ocean response to this forcing, synoptic estimates of air-sea heat, freshwater, and momentum fluxes are required as surface boundary conditions. No observing system yet exists to provide all of these flux measurements, so ocean modelers often rely on the surface fields from numerical weather prediction (NWP) models as realistic surface forcing input. The U.S. National Centers for Environmental Prediction (NCEP) run both global and regional NWP models from which surface fields are available for use in ocean research. The NCEP NWP regional model fields are attractive for use in coastal ocean modeling because they can provide near-real-time gridded estimates of the air-sea fluxes at high spatial and temporal resolutions over much of the U.S. continental shelf.

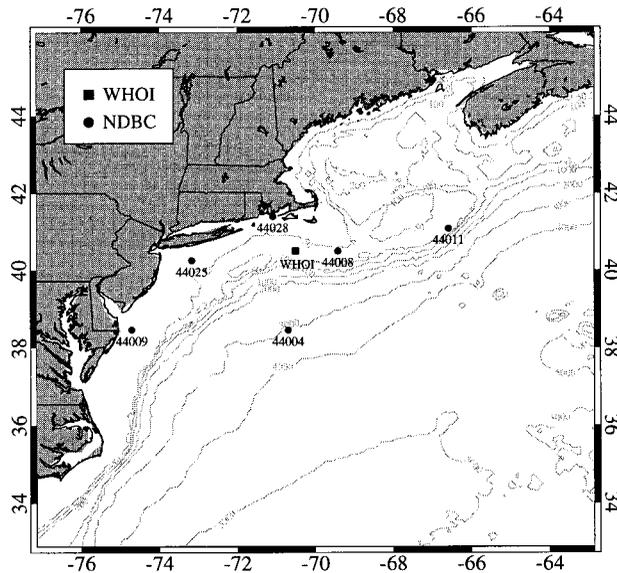
The surface fields from regional NWP models are currently being used as input to the National Oceanic and Atmospheric Administration's (NOAA) Coastal Ocean Forecast System for the U.S. east coast [Aikman *et al.*, 1996; Kelley *et al.*, 1997; Breaker *et al.*, 1999], the Great Lakes Forecasting System at Ohio State University [Bedford and Schwab, 1994; Schwab and Bedford, 1994; Kelley *et al.*, 1998], and the National Ocean Service's forecast system for Chesapeake Bay [Bosley and Hess, 1997]. The use of these regional NWP products for synoptic

meteorological or air-sea flux fields is sure to increase as interest in the coastal zone continues to grow. While the regional NWP model surface fields have great potential for ocean research, use of these fields must be accompanied by an understanding of their errors. Evaluation of these fields has been limited, however, by the relative paucity of high-quality, in situ air-sea flux estimates with which the model fields can be compared.

Such air-sea flux estimates were made during the recent Coastal Mixing and Optics (CMO) experiment, which took place from July 1996 to June 1997 at a midshelf location south of Cape Cod, Massachusetts, in the Middle Atlantic Bight (Figure 1). The scientific objective of the moored array component of CMO is to identify and understand the oceanic mixing processes influencing the evolution of the vertical temperature stratification on the continental shelf. To address this objective, an air-sea interaction surface buoy was deployed for 11 months to characterize the local atmospheric forcing of the ocean. This deployment spanned the four seasons and yielded a continuous, complete, and accurate record of surface conditions and air-sea fluxes in the Middle Atlantic Bight.

While these measurements adequately describe local atmospheric forcing and oceanic response, a complete understanding of the observed variability in the ocean at the CMO site requires the characterization of remote atmospheric forcing. This remote forcing can directly influence mixed layer depth via Ekman pumping, set up pressure gradients and their associated flow through horizontal variability in the atmospheric forcing, or establish ocean conditions elsewhere which are ultimately advected to the observation site. Our long-term goal in acquiring and evaluating the NWP models is to characterize this remote atmospheric forcing and to create accurate air-sea flux fields that can be used in future CMO ocean modeling efforts.

<sup>1</sup>Now at College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis.



**Figure 1.** Coastal Mixing and Optics (CMO) and National Data Buoy Center (NDBC) buoys in the Middle Atlantic Bight. Isobaths are shown in meters. WHOI is Woods Hole Oceanographic Institution.

Moving toward that goal, we present here our evaluations of surface fields from the 1996–1997 versions of three NCEP regional NWP models: the Eta-48, the Eta-29, and the Rapid Update Cycle (RUC-1) models. The high-quality CMO data are used to evaluate the model surface meteorology and air-sea flux fields. Data from six National Data Buoy Center (NDBC) buoys are also used to evaluate spatial errors in the model meteorological products, but since none of these buoys carried moisture, radiation, or precipitation sensors, no flux comparisons are possible. On the basis of these evaluations, we also present some deliberation on generating improved surface forcing fields from the model output for use in future ocean model hindcast studies. While this study is focused on the CMO observation period and the Middle Atlantic Bight, the results of the evaluations are an important step toward documenting the accuracy of NWP fields that, until the summer of 1998, were being used for operational ocean nowcasts and forecasts. Note, however, that these models have since been replaced by the new 32 km Eta and 40 km RUC models. By evaluating the models in operation during the CMO experiment, we are not only seeking to improve the air-sea flux fields for CMO ocean modeling efforts, but we also hope to provide a baseline against which the accuracy of future generations of regional NWP models can be compared.

The acquisition and processing of the NWP model surface products are described in section 2. Details regarding the buoy measurements and air-sea flux estimates are also provided in section 2. Evaluations of both the meteorological and flux fields from each of the models are presented in section 3. Considerations for improving the accuracy of the model fields in light of the evaluations are discussed in section 4. Summary comments are provided in section 5.

## 2. NWP Models and in Situ Data

### 2.1. NWP Model Data

The spatial domain of the Eta-48 model consisted of a 38-level, eta-coordinate vertical grid and a 48 km, rotated latitude/

longitude, semistaggered Arakawa E grid [Black *et al.*, 1993]. It was run every 12 hours, and the model output was remapped to an 80 km Lambert conformal conic projection prior to distribution over the Internet Data Distribution (IDD) system. We obtained the Eta-48 model surface fields from the IDD using the Unidata Program Center's Local Data Manager software (LDM5 Site Manager's Guide, 1996, University Corporation for Atmospheric Research, available at <http://www.unidata.ucar.edu>). While forecasts from 0 to 48 hours every 6 hours were available, only the 0, 6, 12, 18, and 24 hour surface forecasts were extracted from the IDD data stream and archived. Eta-48 model surface fields that were available from the IDD system included east and north components of the wind, air temperature, relative humidity, mean sea level pressure, and accumulated precipitation. The winds were assumed to be at 10 m, and the air temperature and relative humidity were assumed to be at 2 m (G. DiMego, NCEP, personal communication, 1998). Unfortunately, no sea surface temperature or air-sea heat or momentum fluxes were available for the Eta-48 model from the IDD system.

The Eta-29 model was run twice daily on a 29 km, rotated latitude/longitude, semistaggered Arakawa E grid with 50 levels in the vertical domain [Black, 1994]. We extracted and archived the 0, 3, 6, 9, and 12 hour surface forecasts from the 0300 and 1500 model runs. These data were acquired directly from NCEP. Distributed model fields included 10 m east and north components of the wind, mean sea level pressure, 10 m potential temperature, 10 m specific humidity, precipitation rate, sea surface temperature, sensible and latent heat fluxes, incoming longwave radiation, net shortwave radiation, and wind stress.

The RUC-1 model was run on a 25-level hybrid isentropic sigma-coordinate vertical grid and a 60 km polar stereographic horizontal grid [Benjamin *et al.*, 1994a, b]. The model was run every 3 hours and produced hourly forecasts up to 12 hours from the model run time. We acquired and archived the 1, 2, and 3 hour surface forecasts from the NOAA Information Center (NIC) file transfer protocol (FTP) server. RUC-1 model surface fields available from the NIC included east and north components of the wind, mean sea level pressure, potential temperature, condensation pressure, precipitation rate, sea surface temperature, sensible and latent heat fluxes, and global radiation. The winds were assumed to be at 10 m, and the potential temperature and condensation pressure were assumed to be at 2 m (G. DiMego, NCEP, personal communication, 1998).

The model data required minimal processing before evaluation. The Eta-29 and RUC-1 potential temperatures were converted to air temperatures, while the RUC-1 condensation pressure was converted to specific humidity. An outgoing component of longwave radiation ( $LW \uparrow$ ) was not included in the Eta-29 product suite, so this was estimated as

$$LW \uparrow = e\sigma T^4 + (1 - e)LW \downarrow \quad (1)$$

where  $e$  is the emissivity of the sea surface ( $e = 0.97$ ),  $\sigma$  is the Stefan-Boltzmann constant,  $T$  is the Eta-29 sea surface temperature in degrees Kelvin, and  $LW \downarrow$  is the Eta-29 incoming longwave radiation. By estimating the outgoing longwave radiation, evaluations of the Eta-29 net longwave radiation and net heat flux are possible. Finally, the wind field in each of the models was provided in grid-relative coordinates, so that, for example, a wind direction of  $90^\circ$  (oceanographic convention) did not indicate that the wind was blowing toward the east but,

**Table 1.** Positions and Hull Types of National Data Buoy Center (NDBC) Buoys

Buoy	Longitude, °W	Latitude, °N	Hull Type
44004	70.69	38.46	6 m NOMAD
44008	69.42	40.50	3 m discus
44009	74.70	38.46	3 m discus
44011	66.58	41.08	6 m NOMAD
44025	73.17	40.25	3 m discus
44028	71.09	41.40	USCG large navigation

NOMAD is Navy Oceanographic and Meteorological Automatic Device.

instead, was blowing toward the next grid cell in the  $+i$  direction. The wind field in each model was rotated by a grid-specific, spatially varying offset to convert from the grid-relative to Earth-relative coordinates.

Complete time series of Eta-48, Eta-29, and RUC-1 data were compiled from the archived forecast data. The Eta-48 model time series consisted of the 0 and 6 hour forecasts from every model run. If the 0 hour forecast from model run  $N$  was missing, it was filled in with the 12 hour forecast from model run  $N - 1$ . If the 12 hour forecast from model run  $N - 1$  was missing, then the 24 hour forecast from model run  $N - 2$  was used. Approximately 8.7% and 3.2% of the 0 hour forecasts were filled with 12 and 24 hour forecasts, respectively. When the 6 hour forecast was missing from model run  $N$ , it was filled with the 18 hour forecast from model run  $N - 1$ . Roughly 8.1% of the 6 hour forecasts were filled in this manner. The Eta-29 time series was constructed with the 0, 3, 6, and 9 hour forecasts concatenated together. Since there was no precipitation forecast for the 0 forecast hour, the data from the prior model run's 12 hour forecast were used to replace it. The 1, 2, and 3 hour forecasts that were archived for the RUC-1 model were also concatenated to create a complete time series. Data from each of the models were extracted from the grid point closest to the Woods Hole Oceanographic Institution (WHOI) and NDBC buoys for comparison to the in situ measurements.

Acquisition and archival spanned July 1996 through June 1997. The potential temperature and condensation pressure in the RUC-1 model were not archived until October 25, 1996, and so air temperature and humidities are available for this

model only from late October 1996 to June 1997. Missing data in the model time series were primarily due to network and power outages. When possible, gaps in the Eta-48 data were filled from a 36 hour archive maintained by the University of Illinois, Urbana-Champaign. Data return for all the model products was between 92% and 95%. Only 66% of the July 1996 to June 1997 RUC-1 time series was filled with potential temperature and condensation pressure. Complete details of the acquisition and archival of the NCEP model data via the IDD and NIC server are provided by Baumgartner and Anderson [1997].

Some changes were made to the models during the time of this study, and these changes are noted in the evaluations. Since the 1996–1997 CMO observation period, NCEP has introduced the 32 km Eta [DiMego *et al.*, 1998; Rogers *et al.*, 1998] and 40 km RUC models [Benjamin *et al.*, 1998]. These two new models have replaced the Eta-48, Eta-29, and RUC-1 models as the operational regional NWP models. Please note that the evaluations described herein do not apply to the current operational models. For the sake of brevity, the RUC-1 model will hereafter be referred to simply as the RUC model.

## 2.2. Buoy Data

In situ data were collected from a 3 m toroid buoy moored by WHOI at 70.50°W, 40.50°N between July 30, 1996, and June 13, 1997, and from six NDBC buoys moored at various locations throughout the Middle Atlantic Bight (Table 1 and Figure 1). The WHOI buoy carried two redundant vector-averaging wind recorder (VAWR) meteorological packages [Weller *et al.*, 1990], each of which recorded wind speed and direction, barometric pressure, air temperature, relative humidity, incoming short- and longwave radiation, and sea surface temperature every 15 min. Redundant precipitation measurements were collected from two stand-alone, self-siphoning rain gauges (R. M. Young model 50202) and recorded every 3.75 min. A Gill three-axis ultrasonic anemometer (model 1012R2A) was also deployed on the buoy, which recorded 15 min averages of horizontal and vertical wind speed every 30 min. The sensor instantaneous accuracies, sampling intervals, and nominal heights are provided in Table 2. The accuracies in Table 2 were estimated from multiple sensor comparisons conducted during this and other experiments [Weller and Ander-

**Table 2.** Vector-Averaging Wind Recorder Sensor Specifications

Parameter	Sampling Method	Accuracy	Height
<i>VAWR</i>			
Wind speed	900 s average	$\pm 2\%$ above $0.7 \text{ m s}^{-1}$	3.3
Wind direction	900 s average	$\pm 5.6^\circ$	3.0
Air temperature	225 s average	$\pm 0.2^\circ\text{C}$ when wind $> 5 \text{ m s}^{-1}$	2.6
Sea temperature	225 s average	$\pm 0.005^\circ\text{C}$	-1.5
Barometric pressure	2.6 s sample	$\pm 0.2 \text{ mbar}$ when wind $< 20 \text{ m s}^{-1}$	2.7
Relative humidity	3.5 s sample	$\pm 2\% \text{ RH}$	2.7
Incoming shortwave radiation	900 s average	$\pm 3\%$	3.4
Incoming longwave radiation	900 s average	$\pm 10\%$	3.4
Specific humidity*		estimated at $\pm 0.2 \text{ g kg}^{-1}$	
<i>Stand Alone</i>			
Precipitation	225 s sample	$\pm 10\%$	3.1
Sonic anemometer	1800 s sample	$\pm 1.5\%$	3.3

Heights are reported as meters above the mean water line. Wind speed and direction are vector averaged over the sampling interval.

\*Value is computed from air temperature and relative humidity measurements.

**Table 3.** NDBC Data Acquisition, Control, and Telemetry Payload Sensor Specifications [Meindl and Hamilton, 1992]

Parameter	Sampling Method	Accuracy	Height	
			3 and 6 m Buoys	USCG Large Navigation
Wind speed	480 s average	$\pm 1 \text{ m s}^{-1}$ or $\pm 10\%$	5.0	13.8
Wind direction	480 s average	$\pm 10^\circ$	5.0	13.8
Air temperature	480 s average	$\pm 1^\circ\text{C}$	5.0	12.3
Sea temperature	480 s average	$\pm 1^\circ\text{C}$	-1.0	-1.5
Barometric pressure	480 s average	$\pm 1 \text{ mbar}$	0.0	0.0

Heights are in meters relative to the mean water line. Wind speeds are scalar averaged over the sampling interval. Wind direction is calculated as  $\arctan(u/v)$ , where  $u$  and  $v$  are averaged east and north components, respectively, of a unit vector oriented in the direction of the wind.

son, 1996; Moyer and Weller, 1997; Weller et al., 1998; Galbraith et al., 1997].

The meteorological data from the WHOI buoy were largely taken from one of the two redundant VAWR meteorological packages (VAWR serial number 704). However, both of the VAWR cup and vane anemometers failed for 103 hours near the end of the deployment. Since the agreement between the VAWR winds and those measured from the sonic anemometer was excellent ( $r = 0.999$ ), the latter time series was used during this period. Each wind direction measurement was adjusted to account for the local magnetic deviation. All of the relative humidity sensors failed by May 8, 1997. No measurements of outgoing shortwave or longwave radiation were collected, so these quantities were estimated. Outgoing shortwave radiation was estimated as 5.5% of the incoming shortwave radiation. Outgoing longwave radiation was estimated from the measured incoming longwave radiation and the sea surface temperature measured at 1.46 m using (1) and a sea surface emissivity of 0.97.

The NDBC buoys each carried a Data Acquisition, Control and Telemetry (DACT) sensor payload that recorded wind speed and direction, wind gust speed, significant wave height, wave period, air and sea surface temperature, and barometric pressure once every hour. Sensor accuracies, sampling intervals, and nominal heights are provided in Table 3 [Meindl and Hamilton, 1992].

The WHOI VAWR data were decimated to 1, 3, and 6 hour time series to match the RUC, Eta-29, and Eta-48 model time series, respectively. Air-sea fluxes of sensible and latent heat and momentum at the WHOI site were estimated from the decimated meteorological measurements using version 2.5 of the bulk flux algorithm developed for the Tropical Ocean-Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA COARE) [Fairall et al., 1996]. Bulk heat and momentum flux estimates computed with the TOGA COARE algorithm have been shown to agree very well with direct covariance estimates of these fluxes in the Middle Atlantic Bight and elsewhere [Martin, 1998; Edson and Fairall, 1998]. Sensible, latent, and radiative heat fluxes and freshwater fluxes could not be estimated at the NDBC buoys, since no moisture, radiation, or precipitation sensors were included in the DACT meteorological system.

Two and 10 m values for the wind, air temperature, and humidity were estimated from the meteorological measurements at the WHOI buoy for compatibility with the model values at these standard heights. These were derived using the boundary layer profiles computed with the TOGA-COARE

bulk flux algorithm. Ten meter wind speeds were also estimated at the NDBC buoys using an assumption of neutrally stable conditions. The true boundary layer profiles at these sites could not be determined using the bulk flux algorithm without an accompanying humidity measurement.

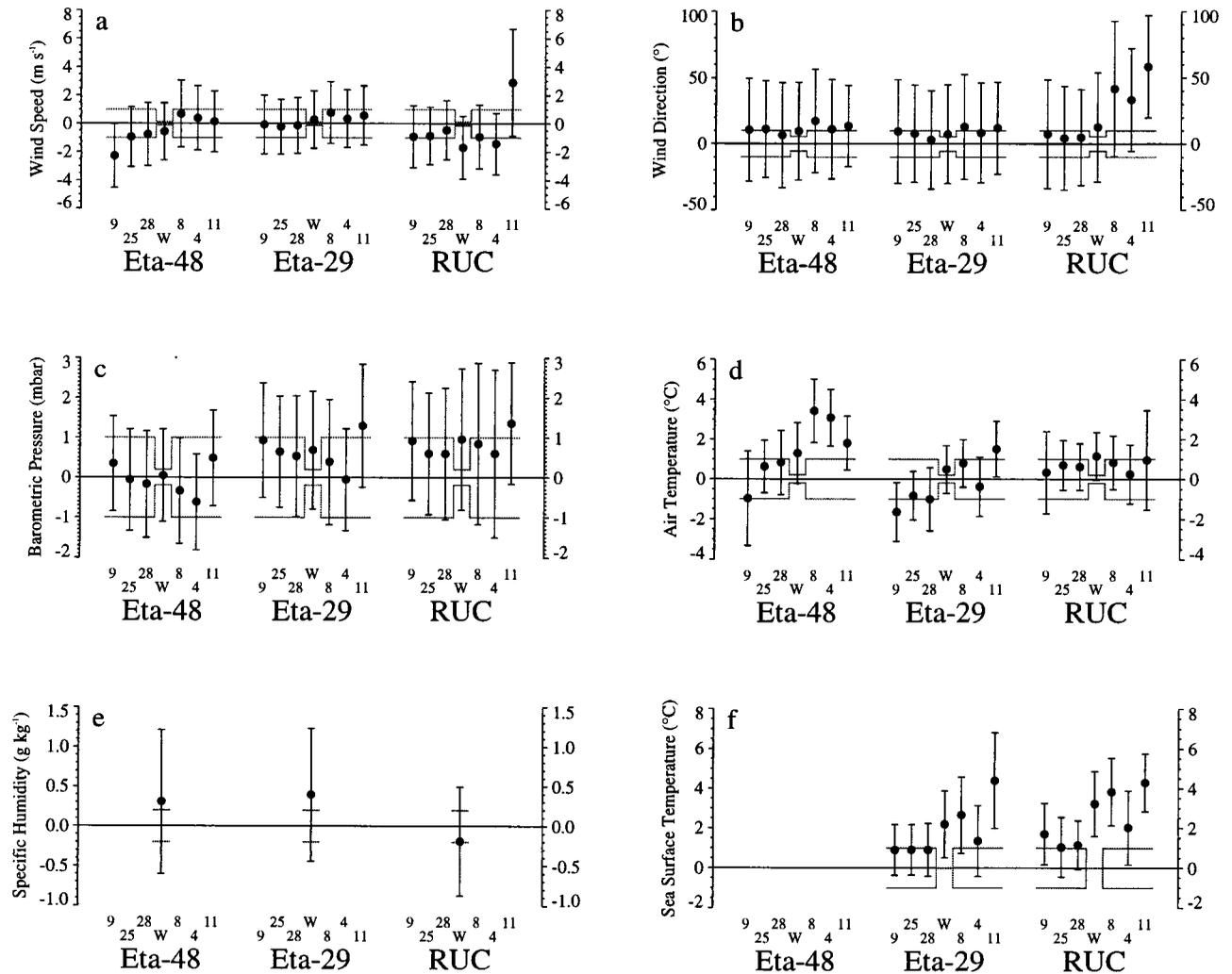
### 3. Evaluation

#### 3.1. Meteorological Observations

The Eta-48, Eta-29, and RUC model fields are evaluated at each buoy location by subtracting the buoy time series from that of the model and computing the statistics of this difference. In all of the model versus buoy comparisons, positive differences indicate that the model observation is higher than the corresponding in situ measurement. The accuracy of the moored instrument is used as the threshold for determining if the model and buoy are in agreement. The mean and standard deviation of each comparison are shown in Figure 2, a monthly mean time series of each variable is provided in Figure 3, a monthly mean difference time series is provided in Figure 4, and the statistics of the differences at the WHOI buoy are presented in Tables 4, 5, and 6. Although the NDBC buoy comparisons provide important spatial information about the accuracy of the model surface fields, the WHOI buoy measurements are of a much higher quality, and so only these statistics are presented in tabular form. We have adopted the oceanographic convention that wind direction indicates the direction toward which the wind is blowing.

**3.1.1. Wind speed and direction.** The mean difference between the Eta-48 and WHOI wind speeds of  $-0.57 \text{ m s}^{-1}$  is lower than the accuracy of the VAWR anemometer at the average wind speed over the entire deployment. The average of the mean NDBC and WHOI biases is  $-0.48 \text{ m s}^{-1}$ , however, there is evidence that the model errors in wind speed vary spatially over the Middle Atlantic Bight (Figure 2a). Note that the NDBC buoy comparisons indicate slightly higher than observed model wind speeds at the offshore sites and lower than observed at the inshore sites. The Eta-48 wind direction is consistently rotated east of north relative to all of the buoy measurements. The average rotation at the WHOI buoy is  $9.5^\circ$ , while the mean offset for all the buoys combined is  $11.3^\circ$ .

The Eta-29 model wind speeds agree very well with the observed buoy wind speeds. The mean difference between the 10 m model wind speeds and those at the WHOI buoy was  $0.28 \text{ m s}^{-1}$ . Compared to the Eta-48 analysis, there is little evidence for an inshore to offshore amplification in the model wind speed errors. The Eta-29 wind directions are also rotated



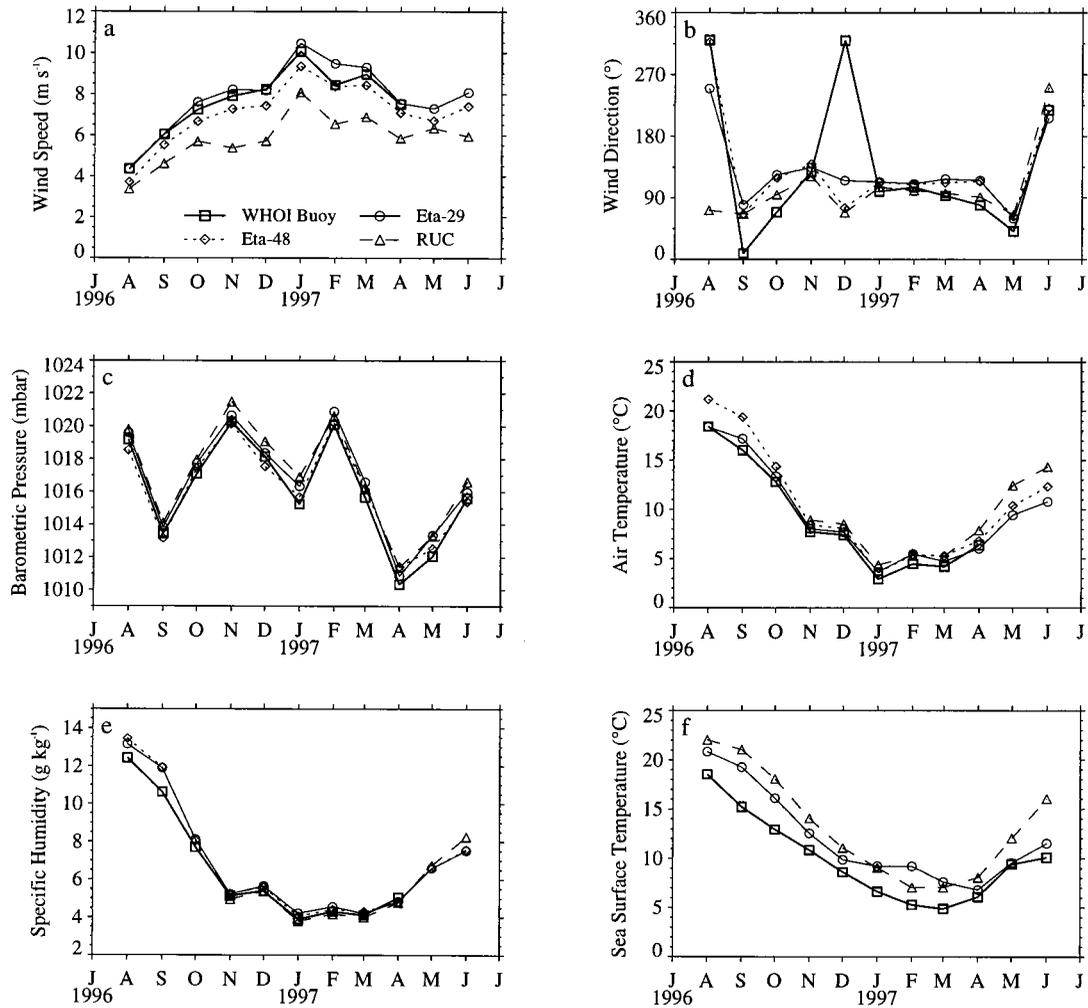
**Figure 2.** Mean biases of model fields minus in situ measurements at WHOI (W) and NDBC buoys (buoy numbers have been truncated to the last one or two digits) for (a) wind speed, (b) wind direction, (c) barometric pressure, (d) air temperature, (e) specific humidity, and (f) sea surface temperature. Error bars represent standard deviations of the differences. Positive differences indicate that the model is higher than the buoy observations. The accuracy of the buoy sensors is indicated by the stippled line. Inshore NDBC buoys are grouped to the left of the “W,” and the offshore NDBC buoys are to the right of it.

an average  $7.3^{\circ}$  east of north relative to the WHOI buoy measurements. This bias seems consistent both in time (Figure 4b) and space (Figure 2b).

The RUC model winds are consistently lower than the buoy observations in all but the NDBC 44011 comparison. The RUC model winds at the WHOI site are, on average,  $1.69 \text{ m s}^{-1}$  lower than the buoy observations, and the mean error at all of the buoys except NDBC buoy 44011 is  $-1.05 \text{ m s}^{-1}$ . Wind directions in the RUC model are generally rotated east of north relative to the buoy measurements, but there is considerable spatial variability in these errors. The errors are greatest near the RUC domain boundary and decrease with distance away from it (Table 7). These boundary effects are manifested in both wind speed and direction, but the effects on wind speed are confined to grid points at or very near the boundary, whereas the errors in wind direction remain large farther away from the domain boundary. The data at grid points within 300 km of the RUC model boundary are relaxed to the output of the NCEP Nested Grid Model (S. Benjamin, NOAA Forecast

Systems Laboratory, personal communication, 1998), and the accuracy of the model fields at the offshore NDBC buoys seems to be influenced by this relaxation.

**3.1.2. Barometric pressure.** The Eta-48 surface barometric pressure agrees very well with in situ measurements. All of the comparisons yield biases that are small relative to the instrument accuracy, and the average of the mean biases for both the WHOI and NDBC buoy comparisons is  $0.0 \text{ mbar}$ . The Eta-29 model surface pressure is higher than observed at the WHOI buoy by an average  $0.7 \text{ mbar}$ . All of the NDBC buoy comparisons, except one, reveal consistently high Eta-29 pressures, but these are generally within the accuracy of the NDBC sensor. The RUC surface pressures are  $1.0 \text{ mbar}$  too high when compared to the WHOI barometric pressure measurements. The RUC pressures at the NDBC sites are also consistently higher than the in situ observations. The average of the mean biases for all the platforms is  $0.8 \text{ mbar}$ . These results suggest that the errors in the pressure field for all three models are quite consistent over both space (Figure 2c) and time (Figure 4c).



**Figure 3.** Monthly means of meteorological observations of (a) wind speed, (b) wind direction, (c) barometric pressure, (d) air temperature, (e) specific humidity, and (f) sea surface temperature. Note that Eta-29 10 m wind speed, air temperature, and specific humidity are shown with 3 m buoy values. Wind direction is in oceanographic convention (e.g., a wind direction of  $90^\circ$  indicates that the wind is blowing toward the east).

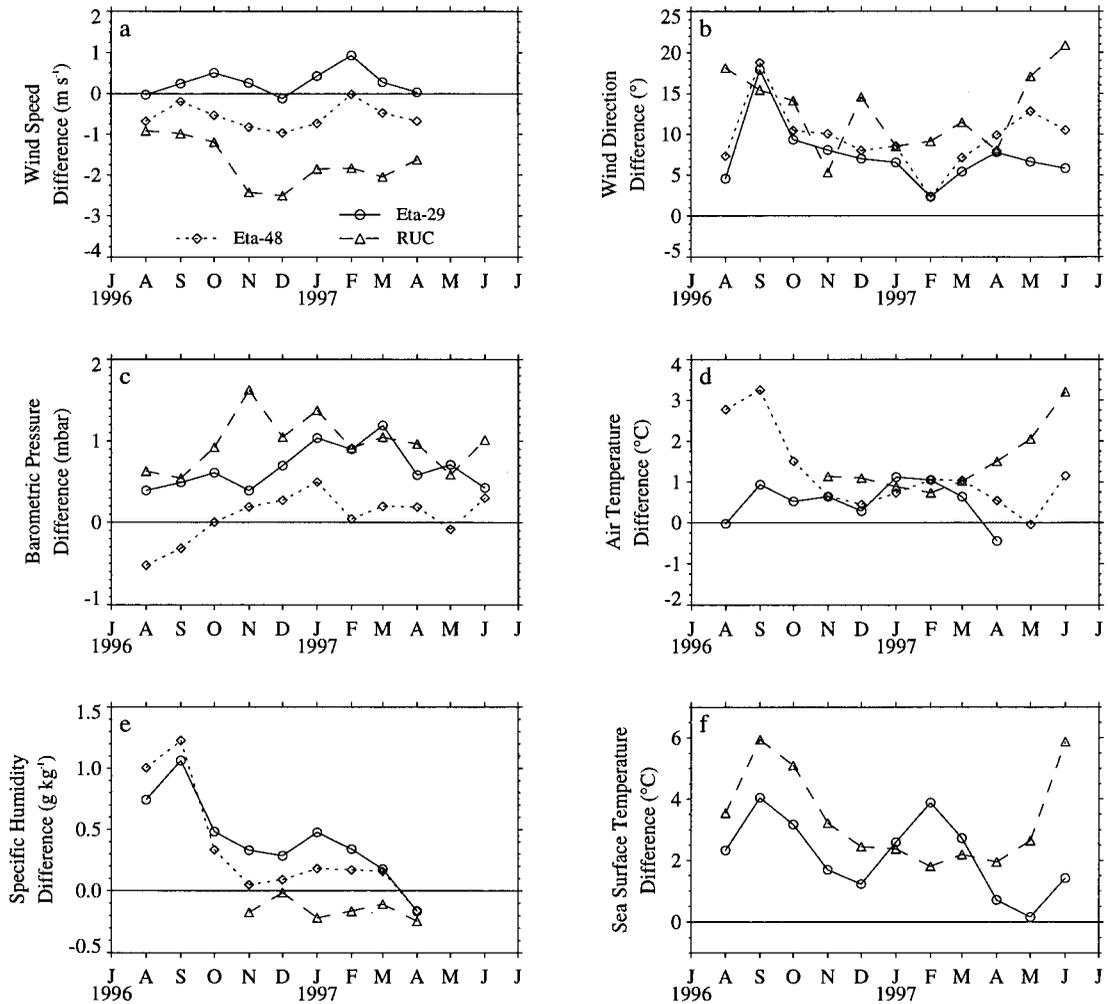
**3.1.3. Air temperature.** The mean bias in the Eta-48 air temperature at the WHOI site is  $1.30^\circ\text{C}$ , however, the errors in this field are time dependent (Figure 4d). Prior to November 1, 1996, the mean bias is  $2.51^\circ\text{C}$ , and afterward, the bias is reduced to  $0.70^\circ\text{C}$ . The errors at the offshore NDBC buoys are also significant, but only the closest NDBC buoy, 44008, shows the same time-dependent error. The errors at the inshore buoys are all within the accuracy of the NDBC air temperature sensor. Note that the errors at NDBC buoy 44009 are much more variable than the other platforms as indicated by the higher standard deviation (Figure 2d). The Eta-48 time series at this buoy contains a pronounced diurnal cycle of daytime heating and nighttime cooling that is more characteristic of a terrestrial atmospheric boundary layer than a marine one. Although the model grid point chosen for this comparison is over the ocean in the 80 km Lambert conformal conic domain, it is probably influenced by grid points over land in the model's native 48 km domain.

The Eta-29 model air temperature is only  $0.49^\circ\text{C}$  higher than observed at the WHOI buoy, and this bias appears to be consistent over time (Figure 4d). However, like the Eta-48, the Eta-29 has a spatial trend in the air temperature biases. The

comparisons at the inshore NDBC sites show lower than observed temperatures, while the Eta-29 model air temperatures at all of the offshore buoys, except 44004, are higher than observed. Unlike the Eta-48 and RUC models, the Eta-29 model air temperature at NDBC 44009 does not have a strong diurnal cycle.

The mean bias in the RUC air temperature at the WHOI site is  $1.16^\circ\text{C}$ . The mean errors at the NDBC buoys are all within the accuracy of the NDBC sensor, and no spatial patterns in the errors are evident. A temporal trend in the errors is present, however, in the time series of the monthly biases at the WHOI buoy (Figure 4d). RUC model air temperatures are higher than observed in the spring of 1997, averaging  $1.00^\circ$  and  $1.79^\circ\text{C}$  before and after April 1, respectively. The errors at NDBC buoy 44011 vary more than the other platforms, most likely because of its proximity to the edge of the RUC domain. The RUC air temperatures at NDBC 44009 also contain a diurnal cycle similar to the Eta-48 air temperatures at the same buoy, but these errors are not related to any data processing issues since the RUC was distributed on its native grid.

**3.1.4. Specific humidity.** The humidity fields for all of the regional models are only evaluated at the WHOI buoy, since



**Figure 4.** Monthly means of model minus WHOI buoy differences of (a) wind speed, (b) wind direction, (c) barometric pressure, (d) air temperature, (e) specific humidity, and (f) sea surface temperature. Positive differences indicate the model is higher than the in situ measurements. Note that the Eta-29 comparisons in Figures 4a, 4d, and 4e end in April 1997. The 10 m wind speed, air temperature, and specific humidity could not be estimated from the WHOI buoy measurements using the Tropical Ocean–Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TC) algorithm after April 1997 since the buoy humidity sensors failed by early May.

the NDBC buoys did not carry any moisture sensor. The Eta-48 model specific humidity bias is  $0.31 \text{ g kg}^{-1}$ , which is largely influenced by humidity errors prior to November 1 (Figure 4e). Mean differences before and after this date aver-

**Table 4.** Statistics of the Eta-48 Model Errors at the Woods Hole Oceanographic Institution (WHOI) Buoy

Variable	<i>N</i>	Bias	Standard Deviation	rmse	<i>r</i>
Wind speed, $\text{m s}^{-1}$	1043	-0.57	2.02	2.09	0.863
Wind direction, deg	1191	9.5	37.0	38.2	0.943
Barometric pressure, mbar	1183	0.1	1.2	1.2	0.992
Air temperature, $^{\circ}\text{C}$	1045	1.30	1.54	2.01	0.980
Specific humidity, $\text{g kg}^{-1}$	1042	0.31	0.91	0.96	0.976

*N* is the number of samples in the comparison, bias is the mean difference, standard deviation is the standard deviation of the differences, rmse is the root-mean-square error, and *r* is the correlation coefficient. Positive biases indicate that the model observations are higher than the in situ buoy measurements.

age  $0.83$  and  $0.05 \text{ g kg}^{-1}$ , respectively. The Eta-29 model errors are similar to those of the Eta-48. The mean bias for this model is  $0.40 \text{ g kg}^{-1}$ , while the differences before and after November 1 average  $0.76$  and  $0.22 \text{ g kg}^{-1}$ , respectively. The RUC model specific humidity is, on average,  $0.19 \text{ g kg}^{-1}$  drier than observed. Since archival of the RUC air temperatures and hu-

**Table 5.** Statistics of the Eta-29 Model Errors at the WHOI Buoy

Variable	<i>N</i>	Bias	Standard Deviation	rmse	<i>r</i>
Wind speed, $\text{m s}^{-1}$	2131	0.28	2.04	2.05	0.863
Wind direction, deg	2446	7.3	37.2	37.9	0.942
Barometric pressure, mbar	2426	0.7	1.5	1.6	0.987
Air temperature, $^{\circ}\text{C}$	2131	0.49	1.20	1.30	0.980
Specific humidity, $\text{g kg}^{-1}$	2131	0.40	0.84	0.93	0.976
Sea surface temperature, $^{\circ}\text{C}$	2473	2.19	1.68	2.76	0.936

See footnotes in Table 4.

**Table 6.** Statistics of the Rapid Update Cycle (RUC) Model Errors at the WHOI Buoy

Variable	<i>N</i>	Bias	Standard Deviation	rmse	<i>r</i>
Wind speed, m s <sup>-1</sup>	6067	-1.69	2.22	2.80	0.815
Wind direction, deg	7000	12.6	41.3	43.2	0.918
Barometric pressure, mbar	6936	1.0	1.8	2.0	0.984
Air temperature, °C	4262	1.16	1.21	1.67	0.958
Specific humidity, g kg <sup>-1</sup>	4262	-0.19	0.68	0.71	0.947
Sea surface temperature, °C	7058	3.23	1.64	3.62	0.961

See footnotes in Table 4.

midities began in mid-October 1996 and all of the WHOI humidity sensors failed by early May, only 6 months of data in the late fall, winter, and early spring are used in this comparison.

**3.1.5. Sea surface temperature.** Sea surface temperature (SST) comparisons are only possible for the RUC and Eta-29 models, since no surface temperatures were available for the Eta-48 model acquired from the IDD. Both the Eta-29 and RUC model surface temperatures are very high when compared to the in situ measurements at the WHOI buoy, averaging 2.19° and 3.23°C above observed, respectively.

The Eta-29 model SST is not computed by the model but is, instead, externally supplied as a boundary condition to the model. It is derived from an operational analysis of surface temperature produced by the National Environmental Satellite, Data and Information Service (NESDIS) from the advanced very high resolution radiometer (AVHRR) carried aboard the NOAA polar orbiting environmental satellites (G. DiMego, NCEP, personal communication, 1998). The gridded analysis product has a resolution of 50 km. This resolution is too coarse, however, to capture large horizontal gradients in surface temperature found on smaller spatial scales over the continental shelf and slope. NESDIS also generates a gridded analysis at 14 km resolution, and on the basis of a qualitative analysis of this product, the errors observed in the Eta-29 model surface temperature appear to be due to the inclusion of warmer waters south of the CMO site in the 50 km grid cell in which the WHOI buoy falls. These warmer surface temperatures cause a warm bias in the 50 km spatially averaged SST when compared to the in situ measurement at the CMO site.

The RUC model SST is consistently warm relative to the in situ measurements in both space and time. This model's surface temperature is taken from a 1° by 1° climatology [Reynolds and Smith, 1995], which is incorporated in the model as a monthly step function. While interannual variability would produce differences between the observed and climatological SST, the magnitudes of those differences are unexpectedly large. To assess the potential errors in this climatology, the RUC SST is compared to monthly mean surface temperatures computed from hourly observations of SST at NDBC buoy 44008 over a 14.6 year period from August 1982 to April 1997 (Figure 5). For 9 of the 11 months, the RUC model SST is higher than the maximum monthly mean ever measured at NDBC 44008, and for 5 of the 11 months, the RUC SST is at least 1°C warmer than the maximum monthly mean at the NDBC buoy. The coarse 1° by 1° climatology from which the RUC SST is derived is also susceptible to the same warm bias found in the Eta-29 SST due to large horizontal gradients in surface temperature. Hence the large differences between the

RUC SST and that measured at the WHOI buoy appear to be related not only to the accuracy of the underlying climatology but also its resolution.

**3.1.6. Precipitation.** Because of high spatial and temporal variability, precipitation from the regional models is compared to the WHOI measured precipitation in monthly accumulations only (Figures 6a and 8a). The R. M. Young rain gauges deployed on the WHOI buoy were not heated, so rainfall rates during winter may be misrepresented. Air temperatures are below 0°C at the CMO site for a total of 306 hours or 7% of the time between November 1, 1996, and May 1, 1997. The Eta-48 and Eta-29 models are remarkably similar throughout most of the time series, but they differ slightly in September 1996. This difference is due to less rainfall in the Eta-29 model during Hurricane Edouard on September 2 than in the Eta-48 model. Monthly rainfall accumulations are high by about 50 mm during the winter in both Eta models but improve by early summer.

The proximity of the CMO site to Woods Hole was problematic in one respect. Hurricane Edouard passed so close to Massachusetts that it interrupted power in Woods Hole and no RUC data were retrieved for that event. Since the hurricane was a significant rainfall event in September, the RUC monthly accumulation for that month is omitted in Figures 6a and 8a. However, during other events in September, the RUC model consistently forecasts much less rain than observed. This supports the conclusion that the RUC model precipitation is too low in the fall. Throughout the rest of the year, the RUC forecast agrees favorably with the observed rainfall, averaging only 8 mm per month less than observed from October 1996 to June 1997.

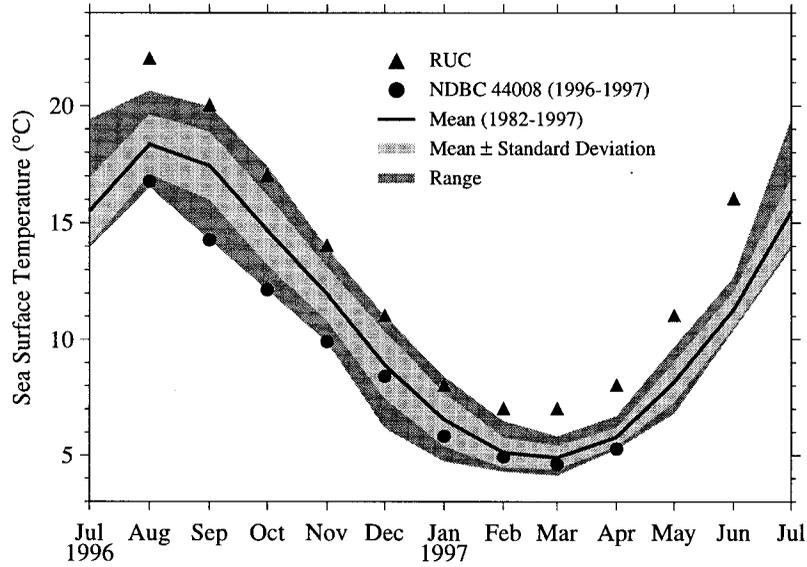
### 3.2. Air-Sea Fluxes of Heat and Momentum

Air-sea fluxes of heat were distributed with the Eta-29 and RUC models, while only the Eta-29 model included wind stress (distributed as friction velocity). These fields are compared to the estimates of the air-sea fluxes at the WHOI buoy computed using the TOGA COARE (TC) bulk flux algorithm [Fairall *et al.*, 1996]. The differences between the model fluxes and those estimated at the buoy are due to not only the errors in the meteorology and sea surface temperatures discussed above but also the differences in the boundary layer parameterizations used in the model and TC bulk flux algorithms. To eliminate the differences between the flux algorithms, new sensible and latent heat fluxes and wind stresses were estimated from the model meteorology and sea surface temperature using the TC algorithm.

The newly constructed fluxes and the buoy estimates are directly comparable, since the differences between the two are due only to the differences in the surface meteorology. Radi-

**Table 7.** Biases in the RUC Model Wind Speed and Direction at the WHOI and NDBC Buoys

Buoy	Distance to Domain Boundary, km	Wind Speed Bias, m s <sup>-1</sup>	Wind Direction Bias, deg
NDBC 44009	412	-0.92	7.4
NDBC 44025	358	-0.86	4.3
NDBC 44028	301	-0.46	4.9
WHOI	240	-1.69	12.6
NDBC 44008	120	-0.93	41.7
NDBC 44004	118	-1.43	33.4
NDBC 44011	0	2.90	58.4

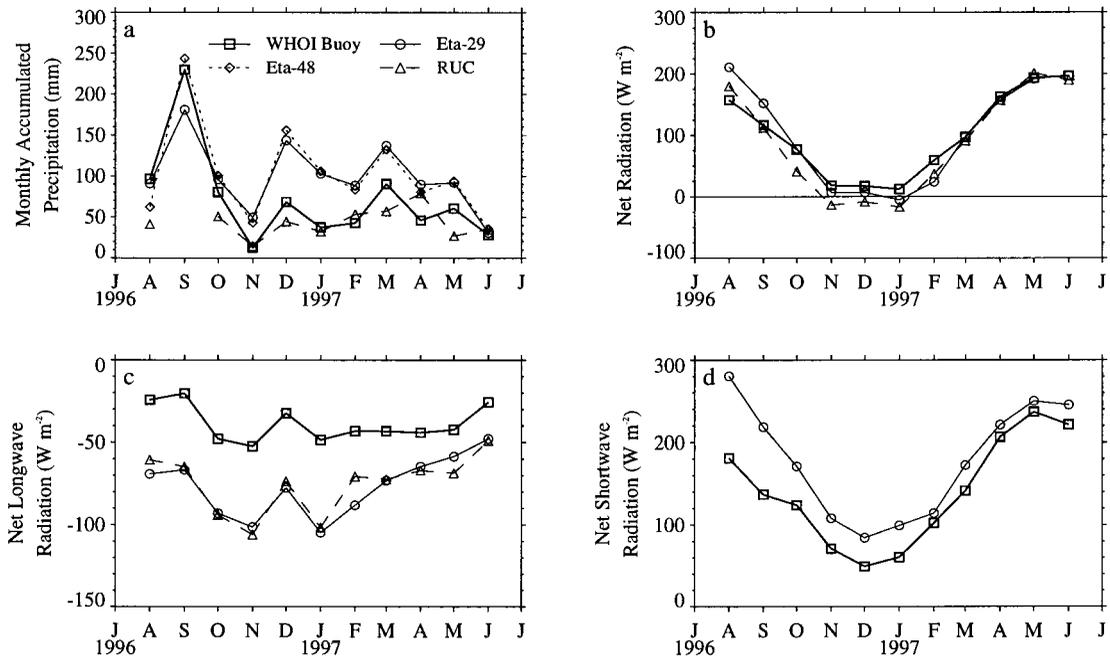


**Figure 5.** Rapid Update Cycle (RUC) model sea surface temperature compared to a monthly climatology from NDBC buoy 44008 (1982–1997). RUC SST data are taken from the model grid point closest to NDBC buoy 44008. Monthly mean surface temperatures from the NDBC buoy for 1996–1997 are also shown. This buoy failed in April 1997.

ative heat fluxes are not affected by the boundary layer parameterization, so no adjustment to these fluxes was necessary for the comparisons. No air-sea fluxes were provided with the IDD distribution of the Eta-48 model, nor was a sea surface temperature included from which estimates of the sensible and latent heat fluxes or wind stress could be derived using the TC algorithm. As such, no Eta-48 air-sea flux comparisons are included here. The differences between the model and WHOI buoy estimates of the fluxes are provided in Tables 8 and 9 for the Eta-29 and RUC models, respectively. Monthly mean time

series of the fluxes and the model minus buoy differences are provided in Figures 6, 7, 8, and 9. We have adopted the oceanographic convention that positive heat fluxes indicate ocean heating and atmospheric cooling.

**3.2.1. Sensible heat flux.** The Eta-29 model sensible heat flux is  $27.2 \text{ W m}^{-2}$  lower, on average, than the buoy estimate, indicating that the Eta-29 model overestimates oceanic heat loss (i.e., atmospheric heat gain). The mean difference between the buoy and the sensible heat fluxes derived from the model meteorology using the TC algorithm is reduced, and the



**Figure 6.** Monthly means of (a) precipitation and (b) net, (c) net longwave, and (d) net shortwave radiative heat fluxes.

**Table 8.** Statistics of the Eta-29 Model Air-Sea Flux Errors at the WHOI Buoy

Variable	<i>N</i>	Bias	Standard Deviation	rmse	<i>r</i>
<i>Model Fields</i>					
Sensible heat flux, $\text{W m}^{-2}$	2131	-27.2	50.5	57.4	0.895
Latent heat flux, $\text{W m}^{-2}$	2131	-64.0	75.4	98.9	0.845
Incoming longwave radiation, $\text{W m}^{-2}$	2419	-28.9	27.6	40.0	0.840
Net longwave radiation, $\text{W m}^{-2}$	2419	-39.3	28.1	48.3	0.743
Net shortwave radiation, $\text{W m}^{-2}$	2333	42.8	116.3	123.9	0.890
Net radiation, $\text{W m}^{-2}$	2326	3.7	109.0	109.1	0.906
Net heat flux, $\text{W m}^{-2}$	2038	-83.4	164.2	184.1	0.888
Wind stress magnitude, $\text{N m}^{-2}$	2131	0.029	0.090	0.095	0.844
<i>Estimates From TC Algorithm</i>					
Sensible heat flux, $\text{W m}^{-2}$	2124	-22.7	23.8	32.8	0.894
Latent heat flux, $\text{W m}^{-2}$	2124	-34.5	41.1	53.7	0.854
Net heat flux, $\text{W m}^{-2}$	2031	-49.7	123.4	133.0	0.908
Wind stress magnitude, $\text{N m}^{-2}$	2124	0.017	0.080	0.082	0.860

See footnotes in Table 4. TC is Tropical Ocean-Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE).

variability in the errors is reduced significantly (Table 8). Since the Eta-29 model wind speed and air temperature are relatively accurate, the errors in the sensible heat flux are mostly attributable to the large warm bias in the Eta-29 SST. Indeed, if the model SST is replaced with the in situ SST from the WHOI buoy and the fluxes are recomputed with the TC algorithm, the error in the sensible heat flux falls to a mean  $6.2 \text{ W m}^{-2}$  (standard deviation of  $16.3 \text{ W m}^{-2}$  and root-mean-square error of  $17.4 \text{ W m}^{-2}$ ).

The RUC model sensible heat flux is  $39.3 \text{ W m}^{-2}$  lower, on average, than the buoy estimate. The largest monthly mean differences occur in the winter, which suggests that there may be a seasonal cycle in the RUC sensible heat flux errors (Figure 9a). The errors in the sensible heat fluxes derived from the TC algorithm and the RUC meteorology are much lower, averaging only  $-12.2 \text{ W m}^{-2}$ , and the variability in these errors is reduced by a factor of 2.5. The remaining errors are due to the interaction of warmer than observed sea surface and air temperatures and lower than observed wind speeds. Since the sensible heat flux is roughly proportional to the air-sea temperature difference, the effect of the large warm bias in SST on the sensible heat flux is mitigated by the warm bias in air temperature.

**3.2.2. Latent heat flux.** The Eta-29 model overestimates the oceanic latent heat loss by an average  $64.0 \text{ W m}^{-2}$ . Latent

heat flux errors are generally larger in winter, with the exception of December 1996 (Figure 9c). These errors are reduced by computing new latent heat fluxes using the Eta-29 meteorology in the TC algorithm. The mean bias in the TC latent heat flux is  $-34.5 \text{ W m}^{-2}$ , and the variability in the errors measured by the standard deviation of the differences is reduced from  $75.4$  to  $41.1 \text{ W m}^{-2}$ . As is the case for the RUC sensible heat flux, the effect of the warm bias in the Eta-29 SST on the latent heat flux is mitigated by the moist bias in the model-specific humidity. Because both the surface saturation humidity and the specific humidity are too moist, the errors in the air-sea humidity gradient (and hence latent heat flux) are not as large as they would be if one of these two variables was more accurate.

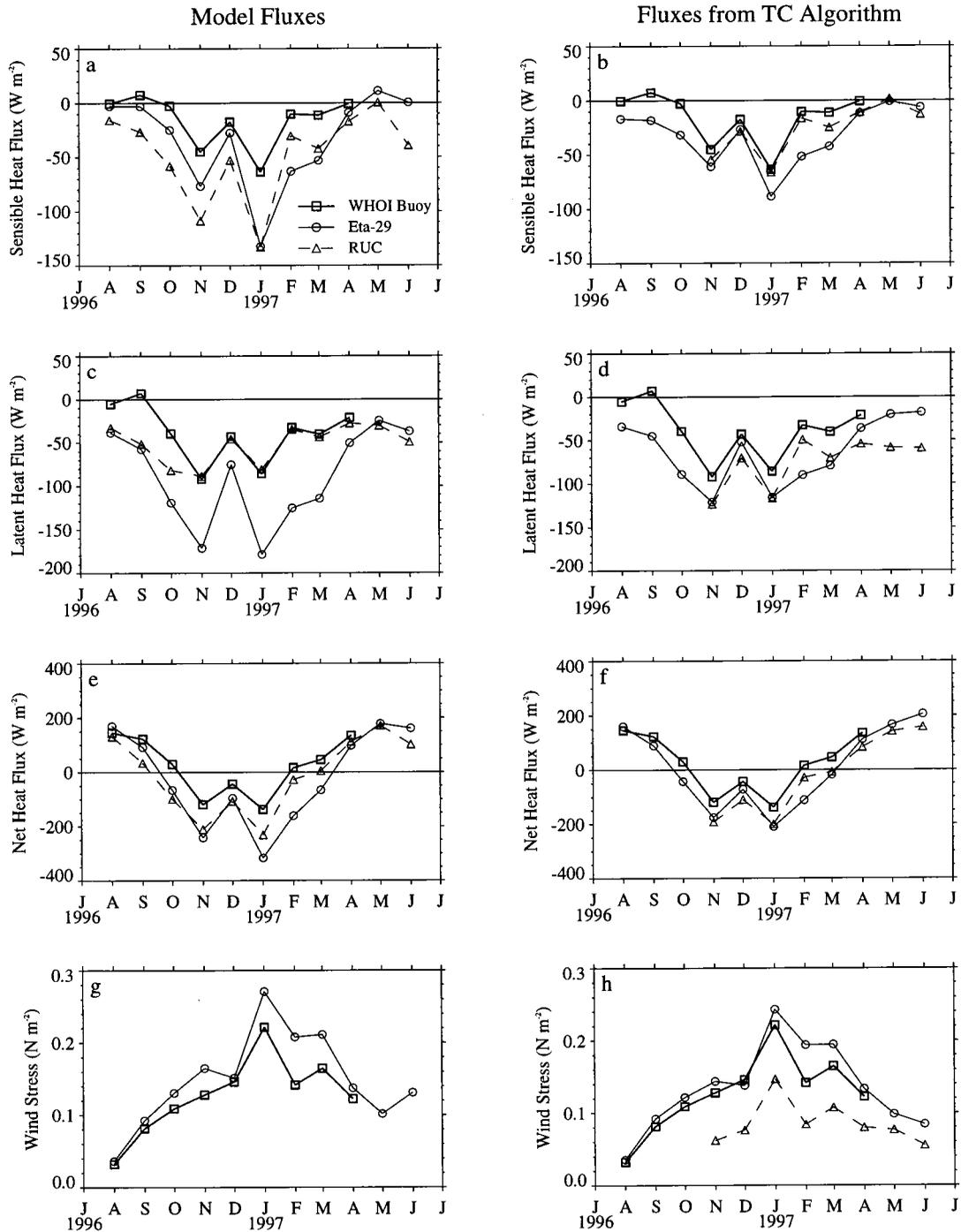
The latent heat fluxes in the RUC model prescribe too much ocean cooling from August to October 1996, but the monthly means are within  $6 \text{ W m}^{-2}$  of the buoy estimate after November 1996 (Figure 9c). This is a curious result, considering that the model SST is much warmer than observed and the RUC specific humidity is slightly dry. Unlike the Eta-29 comparisons of both sensible and latent heat flux and the evaluation of the RUC sensible heat flux, the latent heat flux derived from the RUC meteorology using the TC algorithm does not compare more favorably with the WHOI buoy estimates than the original RUC latent heat flux. The mean error increases from

**Table 9.** Statistics of the RUC Model Air-Sea Flux Errors at the WHOI Buoy

Variable	<i>N</i>	Bias	Standard Deviation	rmse	<i>r</i>
<i>Model Fields</i>					
Sensible heat flux, $\text{W m}^{-2}$	6051	-39.3	48.9	62.7	0.900
Latent heat flux, $\text{W m}^{-2}$	6051	-17.4	35.7	39.7	0.800
Net longwave radiation, $\text{W m}^{-2}$ *	3545	-39.7	30.8	50.2	0.773
Net radiation, $\text{W m}^{-2}$	6901	-13.5	89.0	90.0	0.927
Net heat flux, $\text{W m}^{-2}$	6051	-73.5	111.4	133.5	0.920
<i>Estimates From TC Algorithm</i>					
Sensible heat flux, $\text{W m}^{-2}$	4103	-12.2	19.7	23.2	0.923
Latent heat flux, $\text{W m}^{-2}$	4103	-34.6	40.1	52.9	0.858
Net heat flux, $\text{W m}^{-2}$	4103	-72.0	94.2	118.5	0.929
Wind stress magnitude, $\text{N m}^{-2}$	4103	-0.059	0.091	0.108	0.803

See footnotes in Table 4.

\*Values are estimated from nighttime values of net radiation.



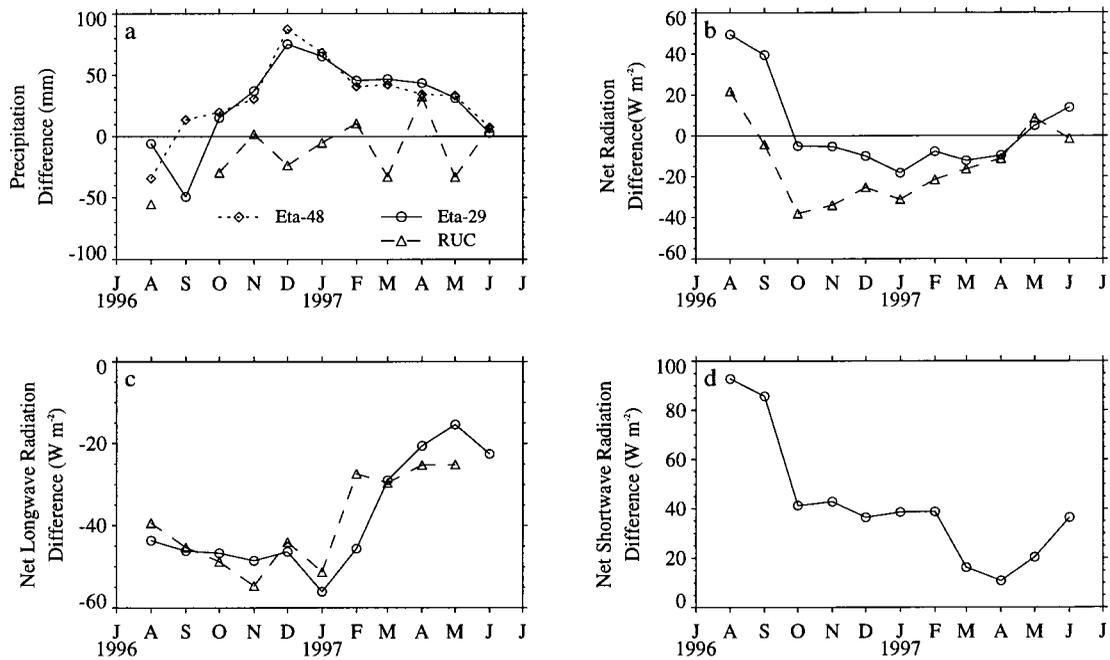
**Figure 7.** Monthly means of (a, b) sensible, (c, d) latent, and (e, f) net heat fluxes and (g, h) wind stress magnitude. (left) Monthly means of the model fluxes versus the WHOI estimates. (right) Monthly means of the fluxes derived from the model meteorology using the TC algorithm versus the WHOI estimates.

$-17.4 \text{ W m}^{-2}$  for the RUC latent heat flux to  $-34.6 \text{ W m}^{-2}$  for the TC latent heat flux. Moreover, the standard deviation of the error increases from  $35.7$  to  $40.1 \text{ W m}^{-2}$ , and the root-mean-square error increases by 33%. These results suggest that the latent heat transfer coefficient may be tuned to account for the large warm bias in the RUC SST.

**3.2.3. Radiative heat fluxes.** The Eta-29 model includes the net shortwave and incoming longwave radiation, while the RUC model supplies only the net radiation. The Eta-29 mean

bias in net radiation is  $3.7 \text{ W m}^{-2}$ , but the monthly mean time series shows biases of  $40 \text{ W m}^{-2}$  or more during August and September 1996 (Figure 8b). The mean bias in the RUC net radiation is  $-13.5 \text{ W m}^{-2}$ , which indicates that the model flux from the ocean to the atmosphere is too large. The errors are not constant over time, however, and the greatest deviations from the observed net radiation occur in the late fall and early winter.

The Eta-29 consistently overestimates the net longwave heat



**Figure 8.** Monthly means of model minus WHOI buoy (a) precipitation and (b) net, (c) net longwave, and (d) net shortwave radiative heat fluxes. Positive differences indicate the model is higher than the in situ measurements. Positive flux differences indicate the model is overestimating the atmosphere to ocean heat flux.

flux from the ocean to the atmosphere by a mean  $39.3 \text{ W m}^{-2}$ . The greatest errors occur in the fall and early winter, but these errors improve after changes were made to the operational Eta-29 radiation package in February 1997 (Figure 8c). The bias in the incoming longwave is the largest contributor to this mean error (Table 8), but the estimated outgoing longwave radiation also contains errors due to the bias in the incoming longwave and the large warm bias in the Eta-29 SST [see (1)]. The mean bias in the Eta-29 outgoing longwave radiation due to reflection of the incoming longwave radiation is only  $-0.9 \text{ W m}^{-2}$ , while the mean error in the gray body radiation attributable to the warm bias in the Eta-29 SST is  $11.2 \text{ W m}^{-2}$ . These two component errors combine to produce a mean bias of  $10.4 \text{ W m}^{-2}$  in the estimated Eta-29 outgoing longwave radiation. The RUC net longwave radiation is estimated as the nighttime values of the net radiation. Like the Eta-29 model, the RUC model net longwave radiation prescribes too much ocean cooling by an average  $39.7 \text{ W m}^{-2}$ . The errors are largest in the fall and early winter but improve slightly in late winter and spring.

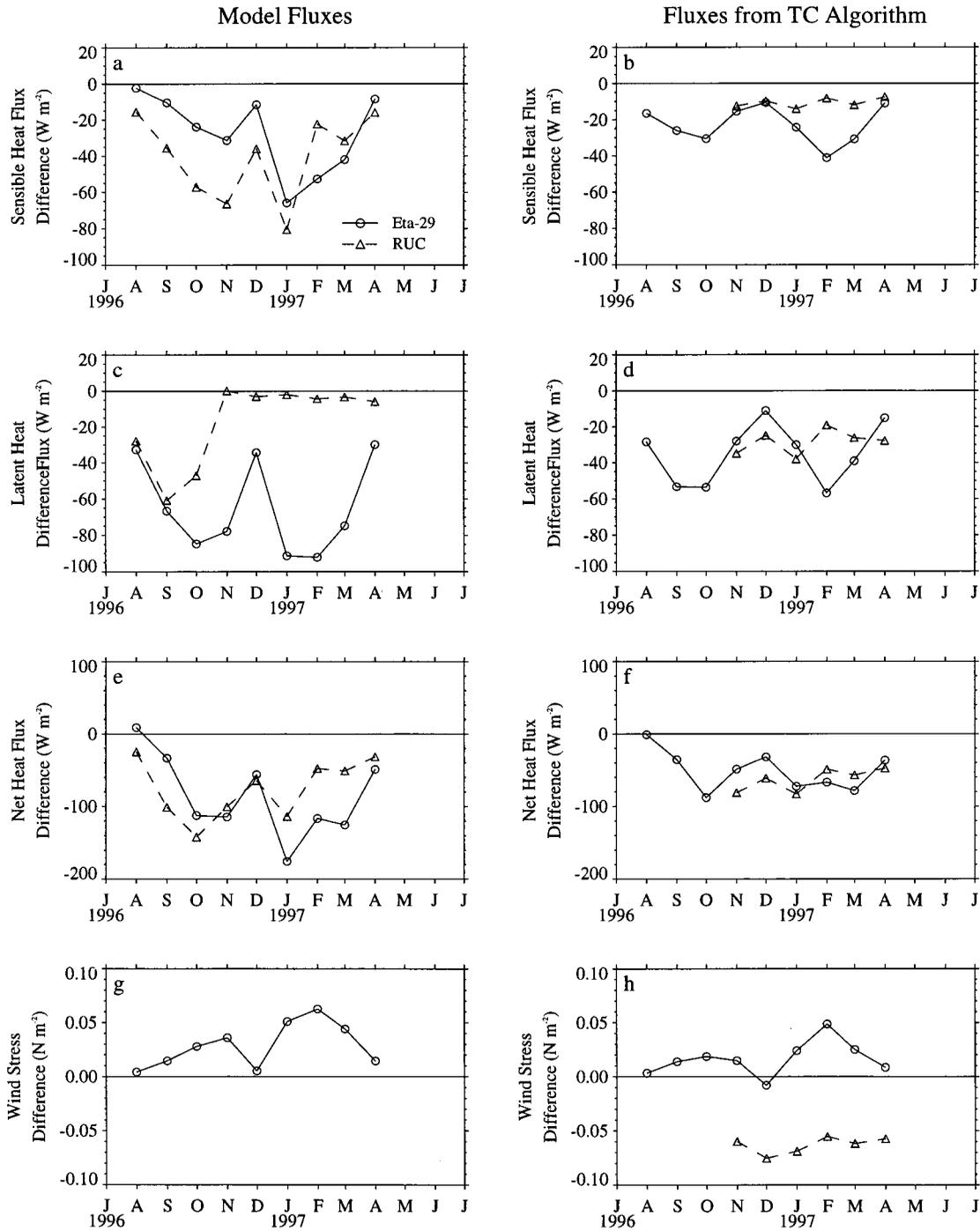
The Eta-29 net shortwave is much higher than observed in the early fall, and the mean bias for the entire time series is  $42.8 \text{ W m}^{-2}$ . The changes in the Eta-29 radiation package in February 1997 are apparent in the reduction of monthly mean biases after February 1997.

**3.2.4. Net heat flux.** The errors described above for the Eta-29 and RUC models contribute to the error in the net heat flux, and it is this total error that will ultimately determine if the model heat flux can be considered accurate enough to produce realistic results in an ocean model. The component biases and the resulting error in the net heat flux for both models are reported in Tables 8 and 9. The net heat flux from the ocean to the atmosphere in the Eta-29 model is overestimated by a mean  $83.4 \text{ W m}^{-2}$  (i.e., too much ocean cooling).

The positive bias in the net shortwave radiation and the negative bias in net longwave roughly balance within  $20 \text{ W m}^{-2}$  after October 1996 (Figure 8b), so the errors in the net heat flux are largely attributable to the errors in latent and sensible heat flux during this time (Figure 9e). The net heat flux computed from the sum of the model net radiative flux and the sensible and latent heat fluxes derived from the TC algorithm is improved because of the reduction in latent and sensible heat flux errors.

Each of the supplied RUC heat flux components overestimates the transfer of heat from the ocean to the atmosphere, resulting in an average error in the net heat flux of  $-73.5 \text{ W m}^{-2}$ . Since the model latent heat flux agrees with the in situ estimate after October 1996, the errors in the net heat flux are attributable to the errors in the sensible heat flux and in net radiation during this time. The error in the net heat flux computed from the sum of the net radiative flux and the sensible and latent heat fluxes from the TC algorithm is comparable to the error in the model net heat flux. Despite an improvement in the sensible heat flux from the TC algorithm, the error in the net heat flux remains virtually the same because of an apparent overestimation in the ocean-to-atmosphere latent heat flux by the TC algorithm relative to the model latent heat flux. As mentioned earlier, the difference between the model latent heat fluxes and those from the TC algorithm is most likely due to tuning of the RUC exchange coefficient for latent heat to account for a much warmer than observed sea surface temperature.

**3.2.5. Wind stress.** The Eta-29 model includes a friction velocity that was converted to wind stress for comparison to the WHOI buoy flux estimates ( $|\tau| = \rho_a u_*^2$ , where  $\tau$  is the wind stress,  $\rho_a$  is the density of air, and  $u_*$  is the friction velocity). The mean bias in the Eta-29 wind stress is  $0.029 \text{ N m}^{-2}$ , and the monthly mean biases are provided in Figure 9g. This bias



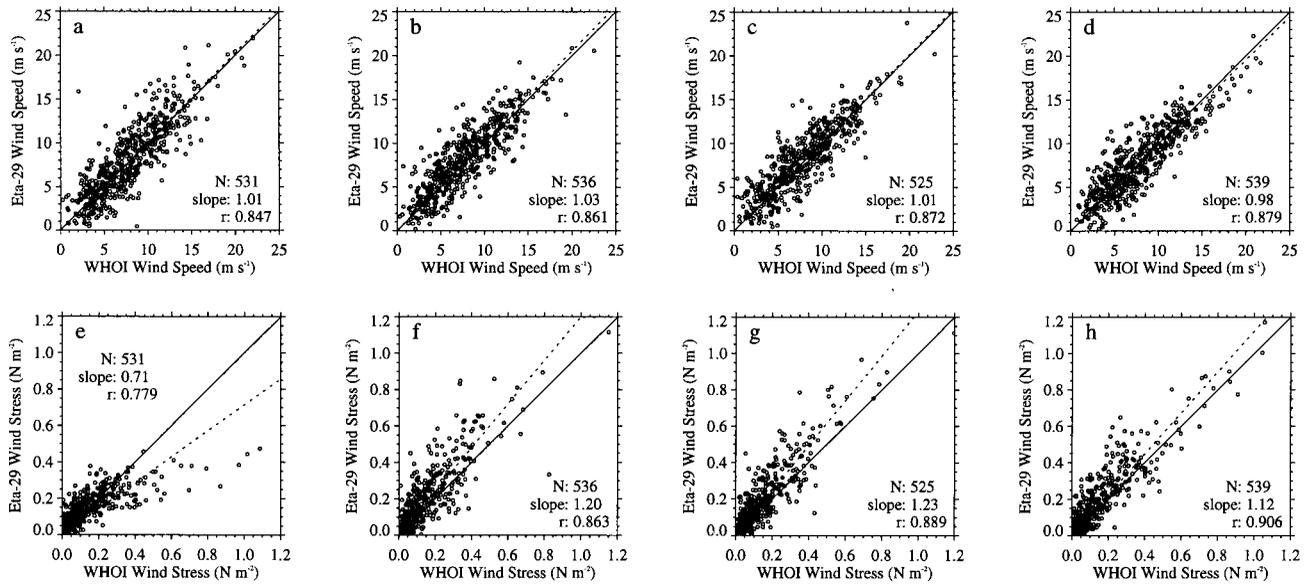
**Figure 9.** Monthly means of model minus WHOI buoy air-sea (a, b) sensible, (c, d) latent, and (e, f) net heat fluxes and (g, h) wind stress. Positive flux differences indicate the model is overestimating the atmosphere to ocean heat flux. (left) Monthly mean errors in the model fluxes. (right) Monthly mean errors in the fluxes derived from the model meteorology using the TC algorithm.

is not only due to differences between the model and in situ meteorology but also to the differences in surface layer parameterizations in the Eta-29 and TC bulk flux algorithms. To account for the latter differences, the wind stress was also estimated from the Eta-29 meteorology using the TC algorithm. The mean error is reduced from 0.029 to 0.017  $N m^{-2}$  or from 25 to 14% of the mean observed wind stress of 0.118  $N m^{-2}$ .

No wind stress was distributed with the RUC model, but a

wind stress can be estimated using the RUC meteorology and the TC bulk flux algorithm. The error in the wind stress thus computed averages  $-0.059 N m^{-2}$ . This lower than observed wind stress is consistent with the lower than observed RUC wind speed.

**3.2.6. Eta-29 air-sea flux errors by forecast.** The evaluations of the Eta-29 model sensible and latent heat fluxes and wind stress are influenced by large differences between the 0,



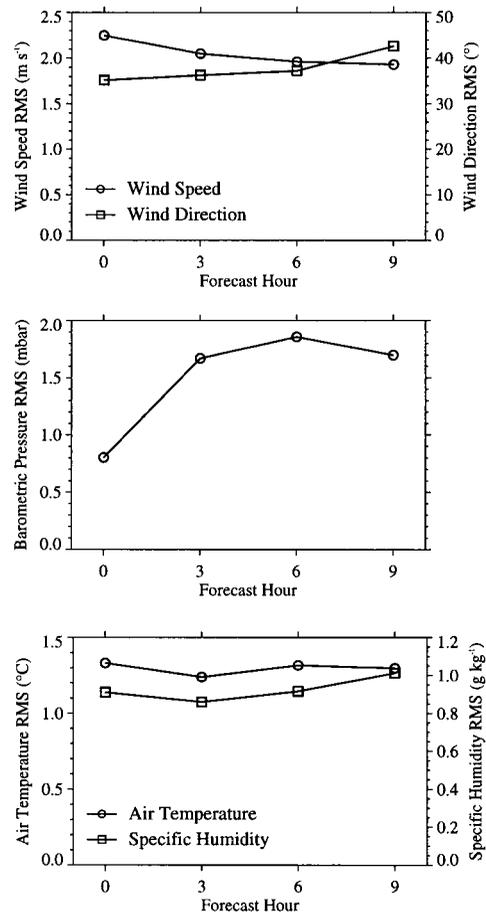
**Figure 10.** Comparisons of the Eta-29 and WHOI buoy wind speed for forecast hours (a) 0, (b) 3, (c) 6, and (d) 9. Similar wind stress comparisons are shown for forecast hours (e) 0, (f) 3, (g) 6, and (h) 9. Dashed lines represent a least squares linear regression with the intercept forced to equal zero. The sample size (N), slope from the linear regression forced through zero, and correlation coefficient  $r$  are reported for each comparison.

3, 6, and 9 hour forecasts from each model run. Figure 10 shows comparisons between the Eta-29 and WHOI buoy wind speed and wind stress for forecast hours 0, 3, 6, and 9 of each model run. At moderate to high wind speeds, the Eta-29 model grossly underestimates the wind stress in the 0 hour forecast (Figure 10e) but is only 12–23% higher than the WHOI values for forecast hours 3, 6, and 9 (Figures 10f, 10g, and 10h). The corresponding wind speed forecasts do not show any evidence of this large between-forecast variation. These results indicate that the Eta-29 0 hour wind stress is unreliable and that these errors are not related to inaccuracies in the 0 hour wind field.

Similar differences between the 0 hour and following forecasts exist in the Eta-29 model latent and sensible heat fluxes. During high-wind events, the Eta-29 model substantially overestimates the sensible and latent heat loss by the ocean for forecast hours 3, 6, and 9. The errors in the 0 hour heat fluxes are considerably smaller. These results suggest that the 0 hour sensible and latent heat fluxes are more reliable during high wind events than successive forecasts.

Despite the between-forecast variability in the sensible and latent heat fluxes and wind stress, no significant differences were detected between forecasts for the Eta-29 meteorological fields. Figure 11 shows the root-mean-square errors during the 0, 3, 6, and 9 hour forecasts for each of the Eta-29 meteorological variables. Contrary to the generally accepted notion that the regional NWP model accuracy suffers during the model spin-up (i.e., for the first 6 hours), there is no significant increase in the accuracy of the meteorological fields between the 0 and 9 hour forecasts. Instead, for wind direction, barometric pressure, and specific humidity, the root-mean-square error actually increases over the four forecasts. The only variable that clearly increases in accuracy is wind speed, but the change between the 0 and 9 hour forecasts is small (16%).

**3.2.7. Implications of SST errors.** For both the Eta-29 and RUC models, the errors in SST are significant. These



**Figure 11.** Root-mean-square errors computed at forecast hours 0, 3, 6, and 9 for the Eta-29 (top) wind speed, wind direction, (middle) barometric pressure, (bottom) air temperature, and specific humidity.

errors are responsible for the large biases in the model estimates of sensible and latent heat fluxes (with the exception of the RUC latent heat flux after October 1996). As a result, the model forecasts of surface air temperature and humidity may also contain errors related to the problems in SST. Indeed, there is evidence that the errors in air temperature and specific humidity are related to errors in SST. The monthly mean errors in the Eta-29 air temperature and SST (Figures 4d and 4f) are highly correlated ( $r = 0.91$ ) as are the errors in the RUC air temperature and SST ( $r = 0.86$ ). The monthly mean errors in the Eta-29 specific humidity and SST (Figures 4e and 4f) are also correlated ( $r = 0.68$ ), but the correlation between the RUC specific humidity and SST is low ( $r = 0.22$ ). The low correlation of the latter comparison may be due to the good agreement between the RUC model latent heat flux and the in situ estimates after October 1996. If the latent heat flux transfer coefficient is tuned to account for large errors in SST, then no relationship would be expected between the errors in specific humidity and SST.

#### 4. Considerations for Improvement

The errors in the model meteorology and radiative fluxes generate annual mean biases in the net ocean-to-atmosphere heat flux as large as  $83.4 \text{ W m}^{-2}$ . The Eta-29 model also overestimates both the wind stress and precipitation, while the RUC model underestimates the wind stress but has relatively accurate precipitation. The magnitudes of these errors are unacceptably large for the purposes of using these fields as surface forcing for an ocean model. In the context of the CMO experiment, the Eta-29 model flux fields, in particular, would produce too much convective overturning and mechanical mixing in an ocean model because of the overestimation of both ocean cooling and wind stress, respectively. Ocean physics over the continental shelf are complex, and the effects of these errors on other important processes such as bottom boundary layer mixing and across-shelf transport are not immediately apparent. Because of these uncertainties, adjustments to the model fields are required to improve accuracy. A simple approach would be to apply adjustments (e.g., bias removal or linear regression) based on analyses of the model errors from in situ comparisons similar to those presented here. For some of the model fields, however, more accurate products from other sources exist. Higher-resolution, remotely sensed, synoptic observations of sea surface temperature are readily available to replace the model SST fields. Similar satellite estimates of downwelling shortwave radiation with high spatial and temporal resolutions are becoming available as well. These products are examined for their potential to improve the model estimates of the air-sea heat flux.

##### 4.1. SST Improvements

A suitable replacement for the model SST is the 14 km gridded AVHRR SST analysis produced by NESDIS. This 3–4 day product is available operationally or from National Climatic Data Center (NCDC) archives. The temporal resolution of the 14 km analysis is a vast improvement over the RUC model monthly climatology. The spatial resolution is also an improvement over the RUC climatology and the 50 km analysis used in the Eta-29 model because the influence of large horizontal gradients over the continental shelf and slope on SST errors is reduced significantly. This product compared very favorably with the in situ measurement of SST at the

WHOI buoy; the 14 km SST was only  $0.57^\circ\text{C}$  warmer than the measured SST, on average, the standard deviation of the differences was  $0.75^\circ\text{C}$ , the root-mean-square error was  $0.94^\circ\text{C}$ , and the correlation coefficient between the two was 0.986.

Acquisition and use of the archived 14 km AVHRR analysis would also improve accuracy by better matching satellite SST observations with the model meteorology in time. The Eta-29 model uses the operational, 3–4 day, 50 km AVHRR analysis product which is only available when the analysis is complete. Hence the sea surface temperatures in the model represent conditions, at worst, 3 to 8 days prior (e.g., the Eta-29 model SST does not show any evidence of the passage of Hurricane Edouard at the CMO site on September 1, 1996, until 5 days later). The 14 km AVHRR analysis data from NCDC archives can be aligned with the model meteorology in time so that it represents conditions, at worst, 2 days before or after a particular time. This would significantly improve sensible and latent heat flux estimates on hourly to daily timescales.

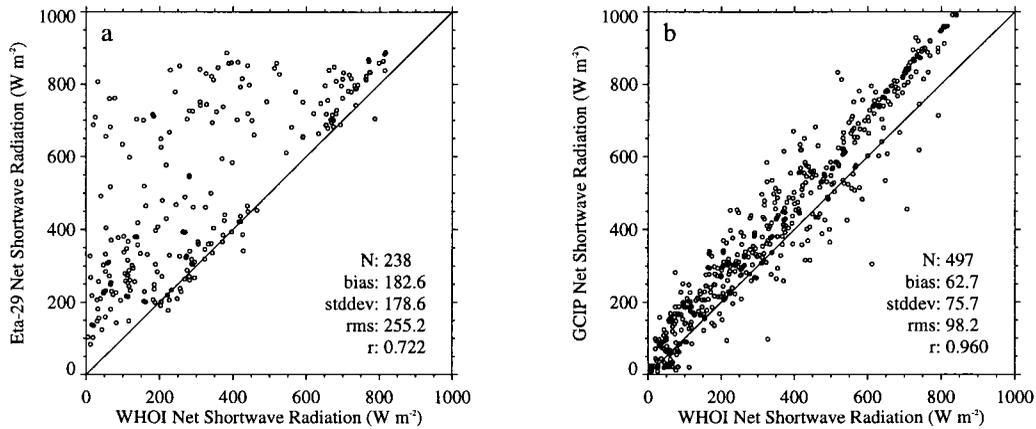
While replacement of the SST field will undoubtedly improve the sensible and latent heat fluxes, the influence of the air temperature and humidity errors on these fluxes should not be ignored. The errors in these fields are correlated with errors in the original model SST fields, so it is likely that a significant component of the air temperature and humidity errors are due to the use of a less accurate SST field when the model was run. These errors must be taken into account to further improve the accuracy of the stability-dependent heat fluxes. For example, a positive error in SST will cause the model to predict more sensible and latent heat loss during unstable conditions, which, in turn, will produce higher than expected air temperatures and humidities. In the case where the error in the model SST and air temperature happen to be equal (i.e., no error in the air-sea temperature difference), replacing the model SST with a more accurate product will actually produce a less accurate sensible heat flux. In the case of the Eta-29, however, the errors in the model SST are considerably larger than the errors in the model air temperature, and so replacement with the 14 km AVHRR analysis product will still produce improved sensible heat fluxes. The same is true of the latent heat flux.

##### 4.2. Simple Adjustments and Air-Sea Flux Estimation

To improve the accuracy of the model air-sea flux fields, adjustments must first be made to each of the meteorological fields. We recommend that the model sensible and latent heat fluxes and wind stresses be abandoned in favor of recomputed fields derived from the adjusted meteorological data using a state-of-the-art bulk flux algorithm. The errors in these model flux fields are large and variable (Tables 8 and 9), which makes the application of a simple adjustment scheme problematic. Moreover, the Eta-29 model contains significant between-forecast errors that make the 0 hour forecast of wind stress unreliable and the 3, 6, and 9 hour forecasts of sensible and latent heat fluxes suspect at high wind speeds. By first reducing the errors in the meteorological data, interactions between error terms in the bulk flux algorithm will be significantly reduced as well. It is these interaction terms that probably cause the errors in the models' stability-dependent fluxes to be so variable.

##### 4.3. Shortwave Radiation

In middle to low latitudes, the net shortwave radiation is almost always the largest daytime component of the net air-sea heat flux. Because the errors in the net shortwave radiation will



**Figure 12.** Comparison between the (a) Eta-29 and 3 hour decimated WHOI net shortwave and (b) Global Energy and Water Cycle Experiment Continental-Scale International Project (GCIP) and 1 hour decimated WHOI net shortwave. Both comparisons include only daytime data from deployment to September 30, 1996.

also be a large component of the errors in the net heat flux, closer examination of these errors is warranted. The bias in the Eta-29 net shortwave is  $42.8 \text{ W m}^{-2}$  (Table 8), but this mean difference includes nighttime values when the in situ and model values agree at  $0 \text{ W m}^{-2}$ . Using only daytime values of net shortwave, the bias more than doubles to  $89.4 \text{ W m}^{-2}$  (standard deviation of  $154.2 \text{ W m}^{-2}$ , root-mean-square error of  $178.2 \text{ W m}^{-2}$ ). The implications of this large error are obscured in the total net heat flux bias reported in Table 8 because of the large negative biases in the Eta-29 sensible and latent heat fluxes and net longwave radiation. The influence of the net shortwave radiation errors on the net heat flux is more easily seen in the root-mean-square errors, since the net shortwave radiation errors are also highly variable.

Because of the high spatial and temporal variability of cloud cover, the Eta-29 model also performs poorly in replicating observed net shortwave radiation at shorter timescales. There is no way to adjust the existing model net shortwave fields to improve accuracy at these timescales, so these fields should be abandoned if more accurate products are available elsewhere. One such replacement product is the downwelling shortwave radiation estimates derived from GOES imagery using the surface radiation budget model from *Pinker and Laszlo* [1992]. This product is currently available only through the Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP) over the continental United States. The temporal and spatial resolution of the GCIP shortwave radiation product match that of the Eta-29 model very well; the GCIP domain has a spatial resolution of  $0.5^\circ$  by  $0.5^\circ$  and produces hourly surface shortwave radiation fields. The CMO site is at the eastern edge of the GCIP study area, and data for July to September 1996 (the latest available data to date) centered at  $40.5^\circ\text{N}$ ,  $70.5^\circ\text{W}$  are used to investigate the accuracy of this product relative to that of the Eta-29 model.

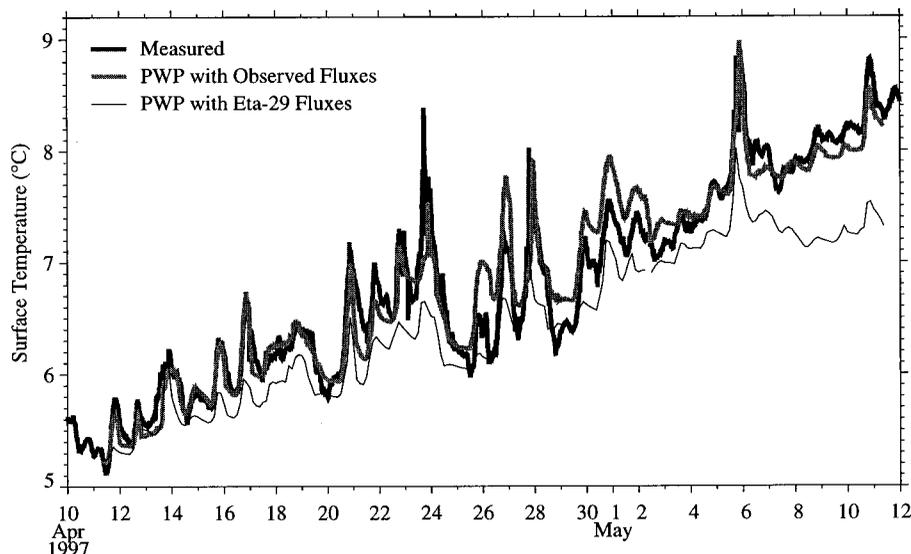
Net shortwave radiation was estimated from the GCIP downwelling shortwave radiation using a sea surface albedo of 5.5% (the same used to estimate net shortwave radiation from the WHOI buoy measurements). Scatterplots of daytime net shortwave radiation from both the Eta-29 model and GCIP are shown against daytime WHOI net shortwave radiation in Figure 12. The GCIP net shortwave radiation is highly correlated

with the in situ estimates ( $r = 0.960$ ), indicating good agreement. This is in stark contrast to the Eta-29 model, which overestimates the net shortwave radiation in most cases. The GCIP net shortwave radiation estimates are 15.4% too high when compared to the WHOI estimates, but this seems related to the choice of an atmospheric transmission coefficient over the ocean that is too high. Because of its higher accuracy and high temporal and spatial resolution, the GCIP product would be an appropriate replacement for the Eta-29 net shortwave field. Unfortunately, the easternmost extent of the GCIP domain extends only to  $70^\circ\text{W}$ , so a large portion of the Middle Atlantic Bight is not covered by this product and reprocessing of the GOES imagery for this area is not planned at this time.

## 5. Conclusion

Although the temporal and spatial resolutions of the regional NWP model surface fields make them fine candidates for use as surface boundary conditions for coastal ocean models, the raw surface fields should be used with some caution. As an example, a one-dimensional mixed layer model [*Price et al.*, 1986] (hereafter referred to as PWP) was used to hindcast the start of the spring restratification over the continental shelf at the CMO site. While the PWP model was developed for open ocean mixed layers and does not account for either bottom boundary layer mixing or horizontal advection, the restratification begins when the shelf waters are both vertically and horizontally homogenous. After the initial surface warming, the shallow surface mixed layer does not interact with the bottom boundary layer. Even in the presence of net along-shelf transport during this period, the thermal gradients, and hence the heat advection, are likely small. The model was initialized with the temperature and salinity profile from the moored CMO data from April 11, 1997.

The model was run twice with surface forcing from (1) the local heat and momentum fluxes estimated from the WHOI buoy meteorological measurements and (2) the Eta-29 air-sea fluxes from the nearest grid point. Despite the simplifications inherent in using the PWP model over the continental shelf, it successfully replicates the net SST change over a 30 day period and much of the diurnal and synoptic (2–10 day) variability when forced with the buoy air-sea fluxes (Figure 13). When



**Figure 13.** Comparison between the measured sea surface temperature (SST) (thick solid curve) and modeled SST during the spring restratification at the CMO site. The PWP model [Price *et al.*, 1986] was run with forcing from the observed WHOI buoy fluxes (stippled curve) and the Eta-29 fluxes (thin solid line). The PWP model was initialized in both cases with temperature and salinity observations from the CMO mooring on April 11, 1997.

forced with the Eta-29 fluxes, however, the net SST change is underestimated and much of the diurnal and synoptic variability is damped. Note that the Eta-29 net heat flux bias over this period is relatively small ( $-28 \text{ W m}^{-2}$ ). Consider the case of trying to model the annual cycle of SST on the continental shelf. In the absence of advection, the  $83 \text{ W m}^{-2}$  bias in the Eta-29 net heat flux, applied vertically over 50 m of water for 1 year, would lead to a  $13^\circ\text{C}$  year-end anomaly in upper ocean temperatures. The consequences of applying such inaccurate buoyancy forcing on the ocean model transports of heat and water mass are uncertain but likely dramatic.

Caution also seems warranted for ocean modeling studies at diurnal and synoptic scales, since the root-mean-square errors of the surface fluxes are large. The largest contributor to the variability in the errors is the model shortwave fields. By replacing the incoming shortwave radiation of the Eta-29 model with the GCIP shortwave radiation product, the standard deviation of the errors in net heat flux from late July through September 1996 drops 46% from  $157.6$  to  $85.7 \text{ W m}^{-2}$ . Moreover, the accuracy at hourly timescales would be improved considerably; the correlation coefficients for the net heat flux with and without the GCIP shortwave product over the same period are 0.929 and 0.863, respectively. The mean bias in the net heat flux is not improved with the addition of the GCIP product during this period, however, because the errors in the ocean cooling heat flux terms (sensible and latent heat and net longwave radiation) are coincidentally balanced by the large errors in the Eta-29 shortwave radiation. With appropriate adjustments, though, the errors in the ocean cooling terms can be reduced as well, which will result in a much improved net heat flux field.

The NCEP regional NWP models continue to change over time, and in the spring of 1998, the RUC-2 replaced the RUC model and the Eta-29 and Eta-48 models were combined into one operational product, the new 32 km Eta. The RUC-2 represents a significant improvement over the RUC-1 with the

inclusion of a larger domain that extends farther east of the CMO site, improved boundary layer physics [Burk and Thompson, 1989], and an SST derived from the 50 km AVHRR analysis. As each new model incarnation comes on line, it should be evaluated against in situ data to assess potential errors in the model fields before they are used in an ocean model. This NWP model evolution issue has been addressed with respect to the global models in the NCEP reanalysis program by producing surface fields over a 40 year period using stable versions of the global data assimilation system and NWP model [Kalnay *et al.*, 1996]. The resulting global data set is of great value to large-scale ocean modelers but does not address the needs of the regional modeling community. Without a regional NWP reanalysis, continued evaluations such as those presented here are needed to increase confidence in the model fields and promote their continued use in ocean modeling efforts.

While this study has examined only the surface fluxes from the point of view of forcing an ocean model, the results do prompt another question: would the weather forecasting skill of the regional NWP models be measurably improved with a better resolved SST boundary condition and state-of-the-art air-sea flux parameterizations? Although we cannot imagine how the NWP forecast skill would be decreased by improved air-sea fluxes, one would likely have to look at particular atmospheric events as case studies to address this question. While we may not be in a position to do this ourselves, the tools and data are available to pursue such a study.

**Acknowledgments.** S. Benjamin (NOAA FSL), D. Fulker (UCAR), S. Lord (NCEP), L. Morone (NCEP), and J. Thiebaux (NCEP) provided encouragement and guidance in our acquisition and interpretation of the NCEP NWP fields. The CMO moored array was conceived and designed by coprincipal investigators S. Anderson, J. Edson, S. Lentz, and A. Plueddemann. We thank W. Ostrom, R. Trask, B. Way, and the WHOI Upper Ocean Processes Group for

their vital contributions to the CMO field operations. The mooring operations were assisted by the fine crew of the R/V *Oceanus* under the direction of Captain P. Howland. J. Edson provided the sonic anemometer data, and R. Pinker (University of Maryland) provided the GCIP radiation fields. The staff of the NDBC is gratefully acknowledged for their data collection, quality control, and distribution efforts. We are indebted to S. Benjamin, J. Gerrity, G. DiMego, J. Kelley, L. Breaker, J. Thiebaut, A. Plueddemann, and two anonymous reviewers for their timely and thoughtful comments on earlier drafts of this manuscript. Funding for this study was provided by the Office of Naval Research under contract N00014-95-1-0339. Woods Hole Oceanographic Institution contribution 9765.

## References

- Aikman, F., III, G. L. Mellor, T. Ezer, D. Sheinin, P. Chen, L. Breaker, K. Bosley, and D. B. Rao, Towards an operational nowcast/forecast system for the U.S. East Coast, in *Modern Approaches to Data Assimilation in Ocean Modeling*, Elsevier Oceanogr. Ser., vol. 61, edited by P. Malanotte-Rizzoli, pp. 347–376, Elsevier Sci., New York, 1996.
- Baumgartner, M. F., and S. P. Anderson, Acquisition, description and evaluation of atmospheric model products for the Coastal Mixing and Optics Experiment, *Tech. Rep. WHOI-97-02*, 196 pp., Woods Hole Oceanogr. Inst., Woods Hole, Mass., 1997.
- Bedford, K. W., and D. J. Schwab, The Great Lakes Forecasting System: An overview, paper presented at the National Conference on Hydraulic Engineering, Am. Soc. of Civ. Eng., Buffalo, N. Y., Aug. 1–5, 1994.
- Benjamin, S. G., K. J. Brundage, and L. L. Morone, The Rapid Update Cycle, I, Analysis/model description, *NOAA/NWS Tech. Procedures Bull. 416*, 16 pp., Natl. Weather Serv., Off. of Meteorol., Silver Spring, Md., 1994a.
- Benjamin, S. G., K. J. Brundage, P. A. Miller, T. L. Smith, G. A. Grell, D. Kim, J. M. Brown, and T. W. Schlatter, The Rapid Update Cycle at NMC, paper presented at the Tenth Conference on Numerical Weather Prediction, Am. Meteorol. Soc., Portland, Ore., July 18–22, 1994b.
- Benjamin, S. G., J. M. Brown, K. J. Brundage, B. E. Schwartz, T. G. Smirnova, and T. L. Smith, The operational RUC-2, paper presented at the 16th Conference on Weather Analysis and Forecasting, Am. Meteorol. Soc., Phoenix, Ariz., Jan. 11–16, 1998.
- Black, T. L., The new NMC mesoscale Eta model: Description and forecast examples, *Weather Forecasting*, 9, 265–278, 1994.
- Black, T. L., D. G. Deaven, and G. DiMego, The step-mountain eta coordinate model: 80 km 'Early' version and objective verifications, *NOAA/NWS Tech. Procedures Bull. 412*, 31 pp., Natl. Weather Serv., Off. of Meteorol., Silver Spring, Md., 1993.
- Bosley, K. T., and K. W. Hess, Development of an experimental nowcast-forecast system for the Chesapeake Bay, paper presented at Fifth International Conference on Estuarine and Coastal Modeling, Am. Soc. of Civ. Eng., Alexandria, Va., Oct. 22–24, 1997.
- Breaker, L. C., H. J. Thiebaut, and J. G. W. Kelley, NOAA's Coastal Ocean Forecast System, *Mar. Weather Log*, in press, 1999.
- Burk, S. D., and W. T. Thompson, A vertically nested regional numerical prediction model with second-order closure physics, *Mon. Weather Rev.*, 117, 2305–2324, 1989.
- DiMego, G. J., M. E. Baldwin, T. L. Black, F. Chen, F. Mesinger, K. E. Mitchell, D. F. Parrish, E. Rogers, and Q. Zhao, Enhancements to the operational "Early" Eta analysis and forecast system at the National Centers for Environmental Prediction, paper presented at the 12th Conference on Numerical Weather Prediction, Am. Meteorol. Soc., Phoenix, Ariz., Jan. 11–16, 1998.
- Edson, J. B., and C. W. Fairall, Similarity relationships in the marine atmospheric surface layer for terms in the TKE and scalar variance budgets, *J. Atmos. Sci.*, 55, 2311–2328, 1998.
- Fairall, C. W., E. F. Bradley, D. P. Rogers, J. B. Edson, and G. S. Young, Bulk parameterization of air-sea fluxes for TOGA COARE, *J. Geophys. Res.*, 101, 3747–3764, 1996.
- Galbraith, N., W. M. Ostrom, B. Way, S. Lentz, S. P. Anderson, M. F. Baumgartner, A. J. Plueddemann, and J. B. Edson, Coastal Mixing and Optics Experiment Mooring Deployment Cruise Report, R/V *Oceanus* Cruise Number 284, 31 July–11 August 1996, *Tech. Rep. WHOI-97-13*, 92 pp., Woods Hole Oceanogr. Inst., Woods Hole, Mass., 1997.
- Kalnay, E., et al., The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77, 437–471, 1996.
- Kelley, J. G. W., F. Aikman III, L. C. Breaker, and G. L. Mellor, Coastal ocean forecasts, *Sea Technol.*, 38, 10–17, 1997.
- Kelley, J. G. W., J. S. Johnson, K. W. Bedford, and D. J. Schwab, Generation of three-dimensional lake model forecasts for Lake Erie, *Weather Forecasting*, 13, 659–687, 1998.
- Martin, M. J., An investigation of momentum exchange parameterizations and atmospheric forcing for the Coastal Mixing and Optics Program, M.S. thesis, Mass. Inst. of Technol. and Woods Hole Oceanogr. Inst. Joint Program, Cambridge, Mass., 1998.
- Meindl, E. A., and G. D. Hamilton, Programs of the National Data Buoy Center, *Bull. Am. Meteorol. Soc.*, 73, 985–993, 1992.
- Moyer, K. A., and R. A. Weller, Observations of surface forcing from the Subduction experiment: A comparison with global model products and climatological datasets, *J. Clim.*, 10, 2725–2742, 1997.
- Pinker, R. T., and I. Laszlo, Modeling surface solar irradiance for satellite applications on a global scale, *J. Appl. Meteorol.*, 31, 194–211, 1992.
- Price, J. F., R. A. Weller, and R. Pinkel, Diurnal cycling: Observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing, *J. Geophys. Res.*, 91, 8411–8427, 1986.
- Reynolds, R. W., and T. M. Smith, A high resolution global sea surface temperature climatology, *J. Clim.*, 8, 1571–1583, 1995.
- Rogers, E., et al., Changes to the NCEP operational "Early" Eta analysis/forecast system, *NOAA/NWS Tech. Procedures Bull. 447*, Natl. Weather Serv., Off. of Meteorol., Silver Spring, Md., 1998.
- Schwab, D. J., and K. W. Bedford, Initial implementation of the Great Lakes Forecasting System: A real-time system for predicting lake circulation and thermal structure, *Water Pollut. Res. J. Can.*, 29, 203–220, 1994.
- Weller, R. A., and S. P. Anderson, Surface meteorology and air-sea fluxes in the western equatorial Pacific warm pool during the TOGA Coupled Ocean-Atmosphere Response Experiment, *J. Clim.*, 9, 1959–1990, 1996.
- Weller, R. A., D. L. Rudnick, R. E. Payne, J. P. Dean, N. J. Pennington, and R. P. Trask, Measuring near-surface meteorology over the ocean from an array of surface moorings in the subtropical convergence zone, *J. Atmos. Oceanic Technol.*, 7, 85–103, 1990.
- Weller, R. A., M. F. Baumgartner, S. A. Josey, A. S. Fischer, and J. C. Kindle, Atmospheric forcing in the Arabian Sea during 1994–1995: Observations and comparisons with climatology and models, *Deep Sea Res., Part II*, 45, 1961–1999, 1998.
- S. P. Anderson, Department of Physical Oceanography, Woods Hole Oceanographic Institution, Woods Hole, MA 02543.
- M. F. Baumgartner, College of Oceanic and Atmospheric Sciences, Oregon State University, 104 Ocean Administration Building, Corvallis, OR 97331. (mbaumgar@oce.orst.edu)

(Received June 12, 1998; revised January 29, 1999; accepted April 16, 1999.)