Available online at www.sciencedirect.com
ScienceDirect
Journal of Marine Systems 62 (2006) 95–119
JOURNAL OF MARINE SYSTEMS
www.elsevier.com/locate/jmarsys

Daily inter-annual simulations of SST and MLD using atmospherically forced OGCMs: Model evaluation in comparison to buoy time series

A. Birol Kara *, Harley E. Hurlburt

Oceanography Division, Naval Research Laboratory, Code 7320, Bldg. 1009, Stennis Space Center, MS, USA

Received 18 January 2005; received in revised form 26 August 2005; accepted 26 March 2006
Available online 9 June 2006

Abstract

A systematic methodology for model–data comparisons of sea surface temperature (SST) and mixed layer depth (MLD) is presented using inter-annual simulations from an ocean general circulation model (OGCM), the Naval Research Laboratory (NRL) Layered Ocean Model (NLOM), over 1980–1998. The model–data comparisons performed here are applicable to other OGCMs and are sufficiently detailed to allow easy comparison of results from other models with NLOM, including statistics at specific buoy locations in specific years. There are three model configurations: coarse resolution (1/2°global) and fine resolution (1/8° global and 1/16° Pacific). These are used to assess the sensitivity to OGCM resolution, including the impact of increasing non-determinicity due to flow instabilities in the 1/8° marginally eddy-resolving and 1/1°6 fully eddy-resolving simulations. In addition, sensitivity to the choice of atmospheric forcing product is investigated. For all three models the atmospheric forcing is from the European Centre for Medium-Range Weather Forecasts (ECMWF) during 1980–1998, and for the 1/16° Pacific configuration only the forcing from the Fleet Numerical Meteorology and Oceanography Center (FNMOC) Navy Operational Global Atmospheric Prediction System (NOGAPS) during 1990–1998 is also used. Availability of NOGAPS thermal forcing begins in 1998. Daily averaged buoy time series from the National Data Buoy Center (NDBC) and the Tropical Atmosphere Ocean (TAO) array are used for inter-annual evaluation of these atmospherically forced OGCMs with no assimilation of SST data and no date-specific assimilation of any data type. For the purpose of model evaluation several statistical metrics are calculated comparing buoy and model time series: mean error (ME), root-mean-square difference (RMSD), correlation coefficient (R), non-dimensional skill score (SS), and normalized RMSD (NRMSD). SST comparisons to the 340 yearlong daily buoy SST time series spanning 1980–1998 gave median values of −0.09 °C for ME, 0.82 °C for RMSD, 0.92 for R, and 0.73 for SS for the 1/2° global model. Positive SS values, an indication of model success, are found for 286 out of 340 buoys (≈84%). An advantage of using a floating mixed layer approach in the global model is demonstrated for simulation of deep and shallow MLD at a buoy location in the Arabian Sea during the 1994–1995 Monsoon period. Model–data comparisons for 1998, when the equatorial Pacific ocean experienced a sharp transition from a strong El Niño to a strong La Niña event, revealed almost no sensitivity to the choice of thermal forcing product. In general, the sensitivity of SST accuracy to model resolution or atmospheric forcing choice was also low based on comparisons to 194 yearlong daily SST time series during 1990–1998. The median RMSD values are 0.68°, 0.61°, 0.84° and 0.84 °C for 1/2°, 1/8°, 1/16° ECMWF-forced simulations and the 1/16° with NOGAPS wind forcing and ECMWF thermal forcing. However, some regional differences exist, e.g., median RMSD values of 0.85 °C for the ECMWF-forced vs. 0.88 °C for the NOGAPS-forced simulations against the 127 TAO buoys, and 0.76 °C for NOGAPS-forced vs. 0.84 °C for ECMWF-forced simulations against the 67 NDBC buoys, all outside the equatorial region. Overall, the results of this paper revealed that 6 hourly wind and thermal forcing

* Corresponding author.
E-mail addresses: birol.kara@nrlssc.navy.mil (A. Birol Kara), harley.hurlburt@nrlssc.navy.mil (H.E. Hurlburt).

0924-7963/$ - see front matter
doi:10.1016/j.jmarsys.2006.04.004

products exist that are sufficiently accurate to allow simulated SSTs and MLDs in an OGCM that are typically accurate to within <1 °C in comparison to daily buoy time series, when there is no assimilation of SST data and when model SST is used in the calculation of latent and sensible heat fluxes. These results indicate that the atmospheric forcing products and the OGCM are suitable for assimilation and forecasting of SST over the global ocean.

Keywords: OGCM evaluation; Daily buoy data; Skill score; SST; MLD; Model resolution

1. Introduction and motivation

Evaluation of atmospherically forced ocean models through quantitative model–data comparison studies is a necessary pre-requisite for any quantitative ocean prediction. In that regard, we investigate whether (a) available atmospheric forcing products, especially thermal forcing (e.g., near-surface air temperature, net shortwave and longwave radiation at the sea surface), and (b) an ocean general circulation model (OGCM) with an embedded mixed layer sub-model are able to simulate sufficiently accurate sea surface temperature (SST) and mixed layer depth (MLD) for that purpose, when (c) there is no assimilation of (or relaxation to) any SST data in the model. These are essential capabilities for SST assimilation using the OGCM as a first guess, and for SST forecasting (Hurlburt et al., 2002).

In general, it is important that an OGCM makes reasonably accurate predictions, so that it can provide a source of information about the effects of oceanic changes. The major difficulty in an OGCM validation has been a lack of global oceanic data sets of sufficient quality and duration to characterize the error statistics. The reason is that highly accurate in situ observational data are relatively sparse in the global ocean, especially in the form of long duration time series with sub-daily sampling, which are critical for evaluation of quantities like SST, sea level and MLD. High vertical resolution is also needed for the latter.

Although the spatial sampling is sparse, taken together, there are many moored buoys located in different regions of the global ocean. For example, the Tropical Atmosphere Ocean (TAO) array (McPhaden, 1995) has reported high-resolution (e.g., hourly) SST and subsurface temperature time series at various depths since 1986, and the National Data Buoy Center (NDBC) has provided sub-daily SST time series at coastal and open ocean locations since before the 1980s. The latter are available from the National Oceanographic Data Center (NODC). Although there have been many high temporal resolution, quality-controlled time series available in many different regions of the global ocean, to our knowledge, inter-annual global ocean simulations predicting SST and other upper ocean quantities have not been verified against long-term daily SST and MLD time series from these buoys in a systematic way.

In the literature, simulations from OGCMs have been evaluated against observational data sets (e.g., Schopf and Loughe, 1995; Chassignet et al., 1996; Hurlburt et al., 1996; Hu and Chao, 1999; Hurlburt and Hogan, 2000), but long-term time series from moored buoys were not used as part of the validation process. While it is obvious that data sets from buoys (for example, SST time series) are some of the most independent and accurate sources for model validation of upper ocean quantities, one-dimensional mixed layer models (e.g., Martin, 1985; Price et al., 1986; Kantha and Clayson, 1994; Large et al., 1994) were usually evaluated at one or a few particular locations, e.g., Ocean Weather Station (OWS) Papa and OWS November where subsurface data were collected. From a global ocean modeling point of view, validating model results at a single location cannot provide sufficient information about the ocean model's successes or deficiencies.

With the availability of high temporal resolution moored buoy measurements from TAO and NDBC, model–data and model–model comparisons can now be made by using time series data from many buoy locations along with detailed statistical metrics. While the focus of this paper is to evaluate a specific OGCM, the Naval Research Laboratory (NRL) Layered Ocean Model (NLOM), and to identify its strengths and weaknesses, the validation procedure and data sets are applicable for use in any other OGCM. In addition, sufficient information for exact model inter-comparisons with NLOM results is included.

The global NLOM is used in the world's first operational eddy-resolving (1/16°) global ocean now-cast/forecast system. That system has been running in real time since 18 Oct 2000, and became an operational product of the U.S. Navy on 27 Sep 2001 (Smedstad et al., 2003). The model parameterizations, tuning and baseline evaluation of the bulk type mixed layer

submodel embedded in NLOM is discussed by Wallcraft et al. (2003). The model performance in predicting climatological SST and MLD using climatologically forced simulations is presented in Kara et al. (2003a). Those studies showed that the model can be successfully used for simulating the climatological annual mean and seasonal cycle of SST and MLD over the global ocean, providing a strong justification to perform simulations on inter-annual time scales with daily resolution.

This paper now discusses SST and MLD results from the inter-annual simulations using high temporal resolution, i.e., 6 hourly (hereinafter referred to as 6 h), atmospheric forcing to investigate the simulation skill of NLOM. The verification procedure for analyzing the model output is specifically designed for application to any OGCM. For model–data comparisons we make extensive use of buoy time series over the time frame 1980–1998 and establish a set of statistical metrics. In addition, statistical model–data comparisons are presented for each year at each individual buoy location separately, so that any other numerical ocean model that predicts upper ocean quantities (e.g., SST and MLD) can be compared with NLOM results.

The sensitivity of the model response to different atmospheric forcing products is also investigated, since this choice can have a significant impact on various aspects of OGCM response (e.g., Samuel et al., 1999; Metzger, 2003; Halliwell, 2004; Kara et al., 2005; Hogan and Hurlburt, 2005). For example, differences in model simulations may arise because atmospheric products have their unique biases over the global ocean (e.g., Rienecker et al., 1996; Trenberth and Caron, 2001). The impact of grid resolution on simulations is also another important topic in model evaluation (Hogan and Hurlburt, 2000, 2005; Hurlburt and Hogan, 2000) as one of the challenges in ocean model studies is to develop more accurate and computationally efficient global eddy-resolving OGCMs for research and operational use. This requires a rigorous validation of results from coarse- vs. fine-resolution model configurations, a topic investigated here in relation to accuracy of SST simulation.

This study is organized as follows: A brief description of NLOM with an embedded mixed layer is given in Section 2. Statistical metrics used for the mixed layer model evaluation are given in Section 3, followed in Section 4 by coarse-resolution (1/2°) model–data comparisons for SST and MLD. In Section 5, a monthly mean model SST formed over the years 1979–1998 from the inter-annual simulations is validated against a climatological data set over the global ocean. The sensitivity of simulated SST accuracy to model resolution (1/2° vs. 1/8° vs. 1/16°) and to the choice of atmospheric forcing product is discussed in Section 6. A summary and conclusions are presented in Section 7.

2. Ocean model

The numerical ocean model used here, NLOM, is a primitive equation layered formulation where the equations have been vertically integrated through each layer. It is a descendent of the model by Hurlburt and Thompson (1980). Since then, it has been greatly enhanced, and a bulk type mixed layer model has been embedded in the dynamical model (Wallcraft et al., 2003; Kara et al., 2003a).

2.1. NLOM overview

The main focus of this paper is evaluation of SST and MLD, so we only give a brief description of the embedded mixed layer submodel parameterizations. The embedded mixed layer model employed here carries prognostic equations for the SST (T_m) and MLD (h_m) as follows:

$$\frac{\partial T_m}{\partial t} + \mathbf{v}_1 \cdot \nabla T_m = -\frac{\max(0,\omega_m)}{h_m}(T_m - \Delta T_m - T_b) + \frac{Q_a - Q_P e^{-h_m/h_P}}{\rho_0 C_{pa} h_m}, \quad (1)$$

$$\frac{\partial (h_m)}{\partial t} \nabla \cdot (h_m \mathbf{v}_1) = \omega_m. \quad (2)$$

A brief description of variables appearing in (1) and (2) is provided in Table 1. Major free parameters in these equations are net surface heat flux (Q_a), temperature shear at the base of the mixed layer [$\Delta T_b = (T_m - \Delta T_m) - T_b$], and vertical mixing velocity (ω_m). These free parameters are obtained from the surface energy budget, a continuous model temperature profile, and a modified version of the Kraus and Turner (1967) model, respectively. There is no direct SST relaxation term in the model equations, but entrainment at the base of the mixed layer allows the dynamical layer density relaxation to influence SST. The mixed layer is not confined to lie within the upper dynamical layer (i.e., it is a floating mixed layer). To some extent, it is independent of the dynamical layers; however, it is not entirely passive as described in Wallcraft et al. (2003). In particular, (1) a deep mixed layer distributes surface forcing across the multiple dynamical layers, (2) thermal expansion is based on the mixed layer temperature

Table 1
Model variables along with their units in the SI system

Symbol	Description of the variable used in model formulation
C_{pa}	Specific heat of air (1004.5 J kg^{-1} K^{-1})
h_m	Mixed layer depth (m)
h_P	Solar radiation absorption length scale (m)
Q_a	Net heat flux at the ocean surface (W m^{-2})
Q_P	Penetrating solar radiation (W m^{-2})
t	Time (s)
T_b	Temperature just below the base of the mixed layer, i.e., $T_b = T_m - \Delta T_m - \Delta T_b$ (°C)
T_m	Sea surface temperature (°C)
$\mathbf{v}_1$	Layer 1 velocity (m s^{-1})
ΔT_b	Temperature difference at the base of the mixed layer (°C)
ΔT_m	Temperature change across the mixed layer (°C)
ω_m	Vertical mixing velocity at the base of the mixed layer (m s^{-1})
ρ_0	Reference density (1000 kg m^{-3})

It is noted that NLOM can be run in hydrodynamic mode (spatially and temporally constant density within each layer) or in thermodynamic mode (spatially and temporally varying density within each layer), i.e., density is a prognostic variable. In thermodynamic mode, it can be run with or without an embedded mixed layer. Only the latter is used here. NLOM has a single scalable and portable computer code (Wallcraft and Moore, 1997) which can be run in a wide variety of configurations, including reduced gravity where the lowest layer is infinitely deep and at rest, and finite depth which allows vertically compressed but otherwise realistic bottom topography confined to the lowest layer. The reader is referred to Wallcraft et al. (2003) for a full description of the model formulations and Kara et al. (2002a) for the parameterizations used in calculating atmospheric forcing fields. Further details about high-resolution model configuration (e.g., 1/16° Pacific) may be found in Metzger and Hurlburt (2001).

rather than layer 1 temperature, allowing the mixed layer to affect the surface dynamic height, and (3) surface heat and momentum fluxes depend on mixed layer temperature through stability dependent air–sea exchange coefficients of Kara et al. (2002a).

The van Leer monotonic scheme (Carpenter et al., 1990; Lin et al., 1994), used for advection of scalars, contributes sufficient non-linear diffusion for stability with a very thin mixed layer. Although this scheme allows NLOM to be run using a very shallow MLD, a minimum value of 10 m is imposed here (see (2)) because the mixed layer formulation is not accurate for very shallow mixed layers. The temperature change across the mixed layer is ΔT_m, which is specified as a function of latitude based on the NRL Mixed Layer Depth (NMLD) climatology with values ranging between 0.1 °C and 1.5 °C (Kara et al., 2003b). Further details about the NMLD climatology can be found in Kara et al. (2002b) and online at http://www7320.nrlssc.navy.mil/nmld/nmld.html.

The net surface heat flux that has been absorbed (or lost) by the upper ocean to depth z, is parameterized as the sum of the downward surface solar irradiance, upward longwave radiation, and downward latent and sensible heat fluxes. Efficient and computationally inexpensive bulk formulae (Kara et al., 2000a) that include the effects of dynamic stability with updated exchange coefficients (Kara et al., 2002a) are used to calculate latent and sensible heat fluxes (based on model SST) at every model time step. The net surface heat flux forcing includes the attenuation of the shortwave radiation with depth (Kara et al., 2004). The surface solar irradiance is decomposed into its infrared (Q_{IR}) and penetrating radiation, so-called Photosynthetically Active irRadiance (PAR). The model reads in monthly attenuation of PAR (k_{PAR}) based on the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) measurements (McClain et al., 1998), so that effects of water turbidity can be taken into account in the model. Note that the radiation absorption length scale in (1) is expressed as $h_P = 1/k_{PAR}$.

Climatologies for density relaxation within each layer are monthly for layer 1 and layer 2 and annual for the other layers. The climatologies for each layer are derived from the bimonthly MODAS climatology (Fox et al., 2002). Since most of the information about circulation anomalies resides in the layer thickness anomalies, the relaxation does not significantly damp anomalies, even on decadal time scales (Jacobs et al., 1994), but it does prevent drift in the temperature just below the mixed layer, a quantity provided to the mixed layer submodel. When combined with calculating the latent and sensible heat flux using model SST, this is generally sufficient to keep SST approximately on track without any explicit relaxation to observed SST (Wallcraft et al., 2003; Kara et al., 2003a). Including air temperature and model SST in the bulk formulae for sensible and latent heat fluxes automatically provides a physically realistic tendency towards the correct SST.

2.2. NLOM simulations

Three different model configurations are used in this paper: 1/2° global, 1/8° global, and 1/16° Pacific. Both the coarse- and fine-resolution global models extend from 72°S to 65°N, and more precisely the grid resolutions are 0.5° in latitude and 0.703125° in longitude, and 0.125° in latitude and 0.17578125° in longitude. The fine-resolution Pacific model (0.0625° in latitude and 0.087890625° in longitude) covers the Pacific Ocean from 20°S to 62°N and 109.125°E to 77.21°W. All models have 6 dynamical layers plus the mixed layer, and vertically compressed but otherwise realistic bottom topography confined to the lowest

layer in the simulations reported here. Shallow sill depths are maintained by constraining the flow to small values below that depth in straits. In all models the lateral boundaries follow the 200-m isobath with a few exceptions. The model boundary conditions are kinematic and no slip.

None of the simulations discussed in this paper include any date-specific oceanic data assimilation. All models were first run using climatological atmospheric forcing until statistical equilibrium was achieved. For climatological wind stress forcing, we use 6-h sub-monthly anomalies from European Centre for Medium-Range Weather Forecasts (ECMWF) operational weather forecast model (ECMWF, 1995) in combination with the monthly mean wind stress climatology of Hellerman and Rosenstein (HR) (1983) because of mixed layer sensitivity to high frequency forcing (e.g., Wallcraft et al., 2003). Monthly mean thermal forcing from the Comprehensive Ocean–Atmosphere Data Set (COADS) described in da Silva et al. (1994) was used to spin-up all models. Fields required from the atmospheric forcing product for use in thermal forcing include the sum of net shortwave and longwave radiation at the sea surface, and air temperature and air mixing ratio at 10 m above the sea surface.

After reaching statistical equilibrium, all model simulations were extended inter-annually using 6-h wind and thermal forcing from ECMWF or the Fleet Numerical Meteorology and Oceanography Center (FNMOC) Navy Operational Global Atmospheric Prediction System (NOGAPS). Further details about the NOGAPS system can be found in Rosmond et al. (2002). The reason for using wind and thermal forcing parameters from both ECMWF and NOGAPS is to investigate the impact of atmospheric forcing choice on the inter-annual model SST simulations (see Section 6). Thermal forcing is based on 6-h data, and the hybrid wind forcing is used. Specifically, for the hybrid wind approach, the long-term ECMWF or NOGAPS mean is subtracted out and replaced by the HR annual mean (denoted as ECMWF/HR and NOGAPS/HR, respectively). As a result, the long-term temporal mean is largely driven by the HR winds, while the seasonal and inter-annual variability are driven by the ECMWF or NOGAPS component. The use of hybrid winds is compatible with the spin-up simulation which was forced by the HR winds. Note that grid resolution for the ECMWF data is $1.125°\times1.125°$, while that for the NOGAPS data is $1.25°\times1.25°$. These were interpolated to each model grid.

NLOM simulations were performed using forcing from the ECMWF Re-Analysis (ERA-15) product (Gibson et al., 1999) during 1979–1993 and archived ECMWF operational data during 1994–1998. For NLOM simulations forced with wind and thermal forcing parameters from NOGAPS, the inter-annual simulations were performed from 1990 to 1998 because NOGAPS forcing was unavailable prior to 1990, and NOGAPS thermal forcing was unavailable prior to 1998. Thus, during 1990–1997 wind forcing was used with thermal forcing from ECMWF except that NOGAPS winds were used in the calculation of sensible and latent heat fluxes.

3. Statistical metrics

The statistical metrics used for comparing SST and MLD time series predicted by the model (NLOM) and those from the buoys are mean error (ME), root-mean-square difference (RMSD), correlation coefficient (R) and non-dimensional skill score (SS). Let X_i ($i=1, 2, \ldots, n$) be the set of n buoy (reference) values, and let Y_i ($i=1, 2, \ldots, n$) be the set of corresponding NLOM estimates. Also let $\overline{X}(\overline{Y})$ and $\sigma_X(\sigma_Y)$ be the mean and standard deviation (S.D.) of the reference (estimate) values, respectively. Following Murphy (1995), the preceding statistical measures can be expressed as follows:

$$\mathrm{ME} = \overline{Y}-\overline{X}, \tag{3}$$

$$\mathrm{RMSD} = \left[\frac{1}{n}\sum_{i=1}^{n}(Y_i-X_i)^2\right]^{1/2}, \tag{4}$$

$$R = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})/(\sigma_X\sigma_Y), \tag{5}$$

$$\mathrm{SS} = R^2-\underbrace{[R-(\sigma_Y/\sigma_X)]^2}_{B_{\mathrm{cond}}}-\underbrace{[(\overline{Y}-\overline{X})/\sigma_X]^2}_{B_{\mathrm{uncond}}}. \tag{6}$$

In the time series comparisons n is equal to 365 (or 366 for leap years), i.e., for a given buoy location and year. ME is the difference between the mean NLOM and the mean buoy values over the time series. RMSD can be considered as an absolute measure of the distance between buoy and NLOM time series and a useful absolute measure of NLOM accuracy. The R value is a measure of the degree of linear association between buoy and NLOM time series. The SS in (6) is the fraction of variance explained by NLOM minus two non-dimensional biases (conditional bias, B_{cond}, and unconditional bias, B_{uncond}) which are not taken into account in the R formulation (see (5)). In brief, B_{uncond}

(also called systematic bias) is a non-dimensional measure of the difference between the means of the buoy and NLOM SST time series, while B_{cond} is a measure of the relative amplitude of the variability in the two data sets. Note that R^2 is equal to SS only when B_{cond} and B_{uncond} are zero. Because these two biases are never negative, the R value can be considered to be a measure of "potential" skill (Stewart, 1990), i.e., the skill that one can obtain by eliminating all bias from NLOM. Note the SS is 1.0 for perfect NLOM predictions, and is negative for $B_{cond}+B_{uncond}>R^2$.

It is important to consider that statistical measures of performance usually focus on a specific aspect of quality, such as linear association in phase in the case of R, or represent a particular composite of several aspects of quality, such as linear association, B_{cond} and B_{uncond} in the case of SS based on RMSD (e.g., Murphy, 1988). Positive skill is usually the criterion for a minimal level of acceptable performance. While it has become common to refer to the aspect of performance measured by the R value as "skill" (Murphy and Winkler, 1987), it more appropriately viewed as a measure of potential skill rather than actual skill (e.g., Murphy and Epstein, 1989).

4. Model–data comparisons

For model–data comparisons we have used observational data from a total of 40 buoys located in different regions over the global ocean (Fig. 1). Half of these buoys are from the TAO array (McPhaden, 1995) located in the equatorial Pacific Ocean and the other half are from NDBC buoys located in the Gulf of Mexico and some distance off the U.S. coasts, including Hawaii and Alaska. Both the NDBC and the TAO buoys include different buoy observables during different time intervals. The buoy observables (or calculated quantities) used here are SST and MLD, the latter from TAO only.

One challenge was how best to compare intermittent time series of different lengths and covering different time intervals, while allowing inter-annual comparison of verification statistics at the same location and comparison of statistics at different locations over the same time interval. As a result, the time series were divided into 1-year segments with daily averaged quantities. This approach also facilitates later inter-model comparisons. Sufficiently detailed results are included here to let NLOM to be an initial benchmark for such comparisons.

The largest number of complete time series obtained from TAO and NDBC buoys are for SST. Over the time frame 1980–1998 there are 196 yearlong daily SST time series available from the NDBC buoys and 144 from the TAO buoys (Tables 2 and 3, respectively). Each buoy location has 5 to 16 yearlong daily SST time series within this time frame. Thus, there are ample SST time series to evaluate daily NLOM SST in different regions of the global ocean during this 19-year period. In

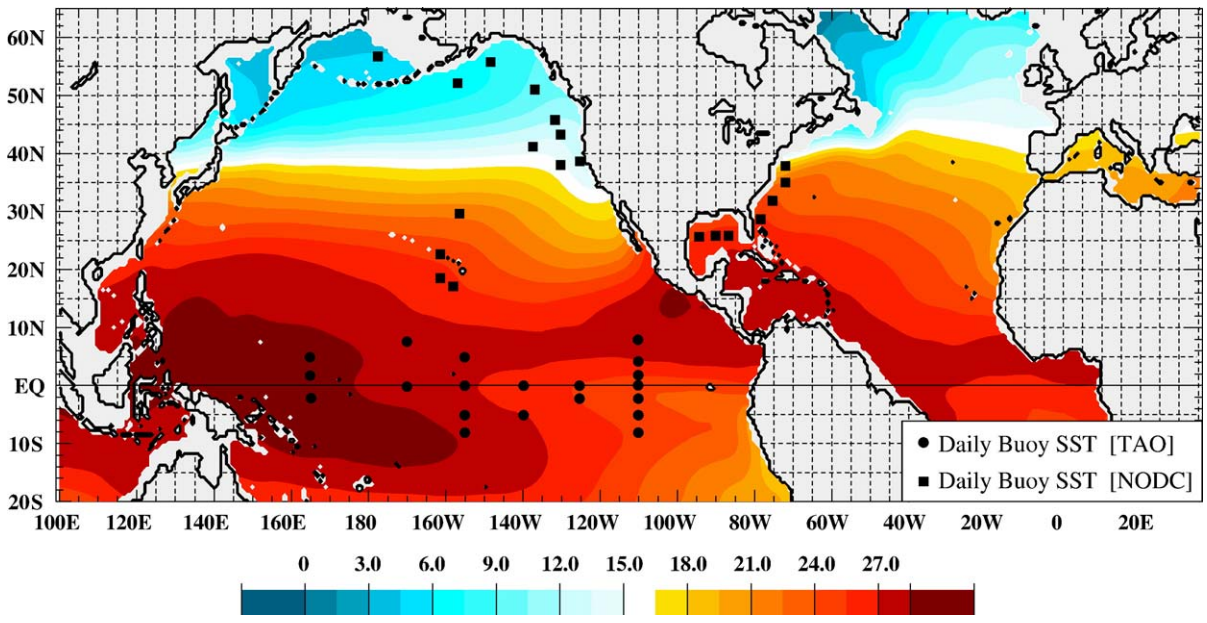

Fig. 1. Buoy locations superimposed on the annual mean of climatological sea surface temperature (SST) from the Coupled Ocean Atmosphere Data Set (COADS). The TAO buoys span the equatorial Pacific Ocean.

Table 2
Availability of the daily sea surface temperature (SST) data from the National Data Buoy Center (NDBC) used for verifying NLOM SST over the time frame 1980–1998

NDBC SST	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98
(17°N, 153°W)	.	.	.	.	.	+	.	.	.	+	+	.	+	+	+	+	.	.	.
(17°N, 158°W)	.	.	.	.	.	+	.	+	.	.	.	.	.	.	+	+	+	+	+
(19°N, 161°W)	.	.	.	.	.	+	+	+	.	.	+	.	+	+	+	+	+	+	.
(23°N, 162°W)	.	.	.	.	+	+	.	+	.	.	+	.	+	+	.	+	.	+	+
(26°N, 086°W)	.	+	+	+	.	+	+	.	.	+	+	+	.	+	.	+	+	.	+
(26°N, 090°W)	.	+	+	+	+	.	+	+	.	.	+	.	.	+	+	+	.	+	.
(26°N, 094°W)	.	+	.	.	+	.	.	+	+	+	+	+	.	+	+	+	+	.	+
(29°N, 079°W)	.	.	.	.	.	.	.	.	.	+	+	+	+	+	+	+	+	+	+
(32°N, 075°W)	.	.	+	+	+	.	.	.	+	.	.	+	.	+	+	+	.	+	.
(35°N, 073°W)	+	.	.	+	.	.	.	.	.	+	+	.	+	.	.	.	.	+	+
(36°N, 122°W)	.	.	.	.	+	.	+	+	+	.	.	.	.	.	.	+	+	.	.
(38°N, 071°W)	.	.	.	+	.	.	+	+	+	.	+	.	+	+	.	+	+	+	+
(38°N, 130°W)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	+	+	+	+
(41°N, 137°W)	.	.	+	+	.	.	.	+	.	+	+	.	.	+	+	+	+	.	+
(43°N, 130°W)	+	+	+	+	.	+	+	.	.	+	.	.	+	+	+	+	+	+	.
(46°N, 131°W)	+	+	+	+	+	.	+	.	+	+	.	+	+	.	+	+	+	+	+
(51°N, 136°W)	+	+	+	+	+	.	.	+	.	.	.	.	.	.	.	.	.	.	.
(52°N, 156°W)	.	.	+	+	.	+	+	+	+	.	+	+	+	.	+	.	+	+	.
(56°N, 148°W)	+	+	+	+	.	+	+	+	+	+	.	+	+	+	.	+	+	+	+
(57°N, 178°W)	.	.	.	.	.	.	+	.	+	+	.	.	+	+	+	+	.	+	+
	5	7	9	11	7	8	10	11	8	10	11	7	11	13	12	17	13	14	12

In the table, a '+' sign indicates that daily SST from 1 January to 31 December (a total of 365 days or 366 days for leap years) is available for that particular year, and they are compared with the daily SST values obtained from the model simulations for each year. The total number of yearlong SST time series available for each year and for each buoy are given at the end of each column and each row. SST data collected by the National Oceanic Atmospheric Administration (NOAA) National Data Buoy Center (NDBC) can be found in the NOAA Marine Environmental Buoy Database (http://www.nodc.noaa.gov/BUOY/buoy.html) operated by NODC.

Table 3
Same as Table 2 buoys but for daily SST availability from the Tropical Atmosphere Ocean (TAO) Array (http://www.pmel.noaa.gov/tao/index.shtml) over the time frame 1986–1998

TAO SST	86	87	88	89	90	91	92	93	94	95	96	97	98	
(0°N, 11°0W)	+	+	+	+	+	+	+	+	.	+	.	+	.	10
(0°N, 125°W)	.	.	+	+	.	.	+	+	+	+	.	+	+	8
(0°N, 140°W)	+	.	+	+	+	.	+	+	+	.	.	+	+	9
(0°N, 155°W)	.	.	.	.	.	.	+	+	+	+	+	+	+	7
(0°N, 170°W)	.	.	.	.	.	.	.	+	+	+	+	+	.	5
(2°N, 110°W)	+	.	+	.	+	.	+	+	+	.	.	+	.	7
(2°N, 165°E)	.	.	.	+	.	.	+	+	+	+	+	+	.	7
(5°N, 110°W)	.	.	.	+	+	+	+	+	+	.	+	.	.	7
(5°N, 155°W)	.	.	.	.	.	.	+	+	+	+	+	+	+	7
(5°N, 165°E)	.	.	.	+	+	+	+	+	+	+	+	+	+	10
(8°N, 110°W)	.	.	.	.	.	.	+	+	+	+	+	.	.	5
(8°N, 170°W)	.	.	.	.	.	.	.	+	+	+	+	+	+	6
(2°S, 110°W)	.	.	+	+	+	+	+	+	+	+	+	.	+	10
(2°S, 125°W)	.	.	.	.	.	.	+	+	+	+	+	+	+	7
(2°S, 165°E)	.	.	.	.	.	.	+	+	+	+	+	+	+	6
(5°S, 110°W)	.	.	.	+	.	+	+	+	+	.	.	+	.	6
(5°S, 140°W)	.	.	.	.	.	+	+	+	+	+	+	+	+	8
(5°S, 155°W)	.	.	.	.	.	.	+	+	+	+	+	+	+	7
(8°S, 110°W)	.	.	.	.	.	.	.	+	+	+	+	+	+	6
(8°S, 155°W)	.	.	.	.	.	.	.	+	+	+	+	+	+	6
	3	1	5	8	6	6	16	20	19	16	15	17	12	144

An assessment of instrumental accuracies indicates errors of about 0.03 °C for SST (Freitag et al., 1994).

particular, the daily SST and MLD time series give information on a whole range of time scales from >1 day to inter-annual. It should be noted that both TAO and NDBC buoys report hourly SST measured at a depth of 1 m below the sea surface. For model–data comparisons we constructed daily averaged SSTs from the buoys and applied no smoothing to the original buoy values but small data gaps are filled by cubic interpolation. Time series with more than a few small gaps (>1 month) are excluded. The model (NLOM) used in this paper does not simulate the diurnal cycle and no date-specific data are assimilated (just climatological density relaxation below the mixed layer, see Section 2). Thus, there is no need to form a daily average of model SST.

TAO buoys do not directly report MLD but they provide subsurface temperatures measured at different depths down to 500 m. Thus, we calculated these quantities using daily subsurface temperatures for model–data comparisons as will be explained in Section 4.2. There were no yearlong subsurface temperature time series available from the NDBC buoys. Therefore, the NDBC buoys are not used for model–data comparisons of MLD.

4.1. Sea surface temperature (SST)

As examples to illustrate the model assessment procedure used throughout the paper, detailed model–data comparison results are shown for two different

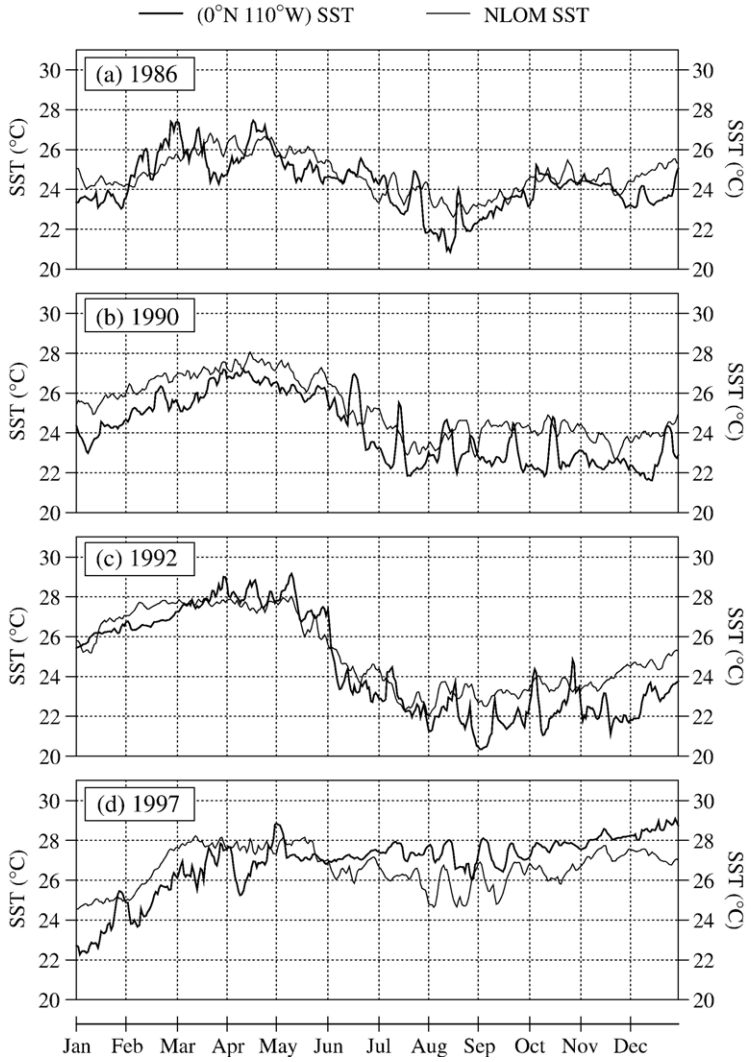

Fig. 2. Daily averaged SST from a Tropical Atmosphere Ocean (TAO) buoy at (0°N, 110°W) located in the eastern equatorial Pacific (thick line) and modeled SST from the 1/2° global NLOM simulation (thin line) forced with ECMWF wind/thermal fluxes during 1979–1998: (a) 1986, (b) 1990, (c) 1992, and (d) 1997. Note that NLOM included no assimilation of SST data. The *x*-axis is labeled starting from the beginning of each month.

buoy locations. One of these buoys, a TAO buoy, is located in the eastern equatorial Pacific Ocean (0°N, 110°W), and the other one, a NDBC buoy, is located off the Alaska coast (52°N, 156°W). Yearlong SST time series comparisons performed in 1986, 1990, 1992, and 1997 are shown in Fig. 2 for (0°N, 110W) and in Fig. 3 for (52°N, 156°W). Overall, atmospherically forced NLOM with an embedded mixed layer is able to simulate daily SST, including its seasonal and inter-annual variations quite well for both cases in all years but with a 10- to 15-day lead in the spring–summer rise in SST at (52°N, 156°W).

Table 4 shows statistical model–data comparisons between the yearlong NLOM and buoy SST time series at (0°N, 110°W) for the years available (see Table 3). The same type of model–data comparison is shown in Table 4 for (52°N, 156°W). Annual mean SST biases (i.e., ME) between the NLOM predicted SSTs and buoy SSTs are generally small at both locations. The model is able to the capture the phase of SST variability quite well because the *R* values are generally high. The model success is also evident from mostly large and positive SS values, indicating that NLOM is able to simulate SST with acceptable accuracy for nearly all years. There is only one year (1989) that NLOM did not predict SST with positive SS at (0°N, 110°W). In this case, although the *R* value is large, 0.84, the SS value is negative because of large ME (1.37 °C) that year.

Here it is worthwhile to note that using only one statistical metric is not sufficient to determine whether or not the model performed well. For example, at (0°N, 110°W) the RMSD value of 1.52 °C in 1988 is larger

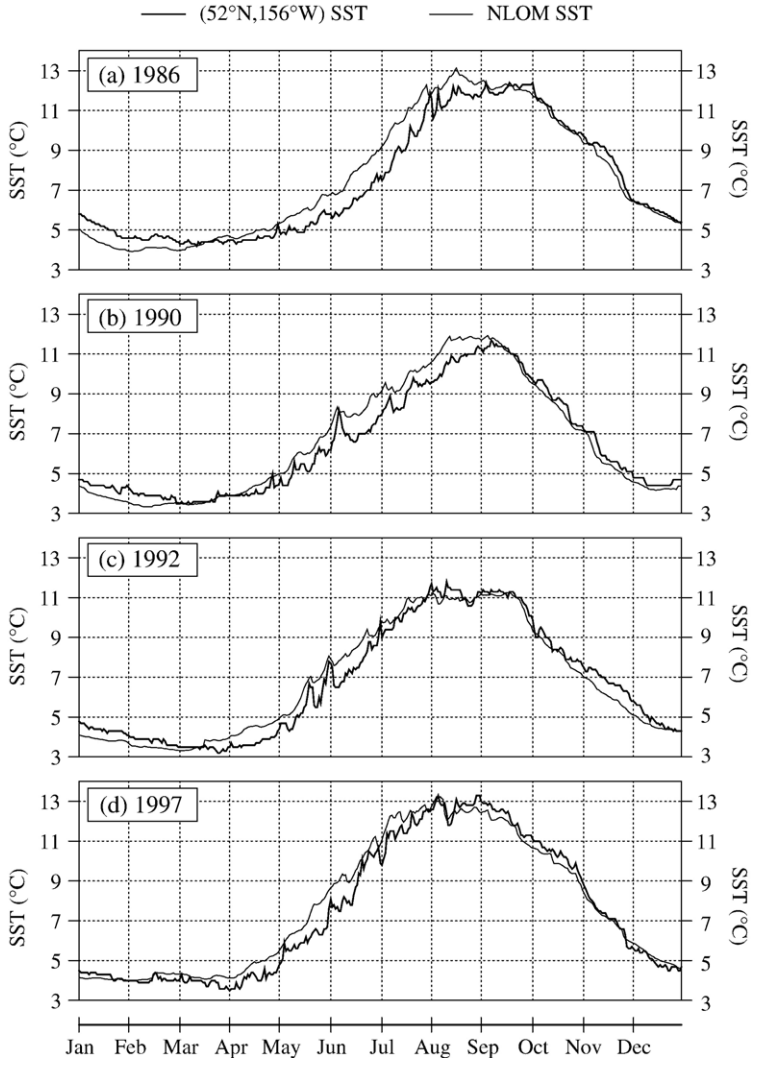

Fig. 3. Same as Fig. 2 but from a National Data Buoy Center (NDBC) buoy at (52°N, 156°W) located near the Alaska coast.

Table 4
Statistical verification of daily SST simulated by NLOM against the TAO buoy at (0°N, 110°W) in the equatorial Pacific Ocean; similarly for NLOM versus the NDBC buoy at (52°N, 156°W) off the Alaska Coast

(0°N, 110°W) SST	Annual mean		S.D.		Statistical analysis (n=1 year)			
	Buoy (°C)	NLOM (°C)	Buoy (°C)	NLOM (°C)	ME (°C)	RMSD (°C)	R	SS
1986	24.30	24.73	1.31	0.97	0.43	0.96	0.75	0.46
1987	25.66	25.89	1.57	1.16	0.23	0.99	0.79	0.60
1988	23.88	24.61	2.20	1.59	0.73	1.52	0.80	0.52
1989	23.82	25.19	1.56	1.24	1.37	1.61	0.84	−0.07
1990	24.16	25.24	1.65	1.44	1.08	1.35	0.87	0.34
1991	24.96	25.42	1.49	1.56	0.46	1.03	0.82	0.52
1992	24.53	25.17	2.52	1.88	0.64	1.20	0.93	0.77
1993	24.63	25.38	1.86	1.61	0.75	1.18	0.87	0.59
1995	23.34	24.65	1.92	1.95	1.31	1.63	0.87	0.28
1997	26.89	26.63	1.45	0.99	−0.26	1.31	0.50	0.18
(52°N, 156°W) SST	Annual mean		S.D.		Statistical analysis (n=1 year)			
	Buoy (°C)	NLOM (°C)	Buoy (°C)	NLOM (°C)	ME (°C)	RMSD (°C)	R	SS
1982	6.95	6.95	2.47	2.66	0.00	0.73	0.96	0.91
1983	7.50	7.32	2.92	3.12	−0.18	0.39	0.99	0.98
1985	6.77	7.30	2.74	2.82	0.53	0.84	0.97	0.91
1986	7.21	7.63	2.85	3.08	0.42	0.84	0.97	0.91
1987	6.31	7.28	2.43	2.91	0.97	1.24	0.97	0.74
1988	5.76	7.15	2.71	2.86	1.39	1.51	0.98	0.69
1990	6.53	7.18	2.60	2.89	0.66	0.90	0.98	0.88
1991	6.63	7.30	2.67	2.86	0.67	0.92	0.98	0.88
1992	6.77	7.29	2.84	2.82	0.52	0.76	0.98	0.93
1994	6.69	6.76	2.73	2.85	0.07	0.46	0.99	0.97
1996	6.88	7.29	2.66	2.87	0.41	0.78	0.97	0.91
1997	7.53	7.74	3.37	3.23	0.21	0.57	0.99	0.97

Comparisons are made for each year separately, using a variety of statistical metrics as described in the text. Note that there is no assimilation of SST in the NLOM simulations. A SS value <0 indicates a poor prediction from NLOM. The statistical results are calculated based on 365 daily (or 366 for leap years, i.e., 1988, 1992 and 1996) values. S.D. denotes standard deviation.

than that of 0.96° C in 1986. This may indicate that the NLOM SST prediction in 1988 is worse than in 1986. However, the non-dimensional SS value in 1988 (0.52) is higher than that in 1986 (0.46). This means NLOM SST prediction in 1988 shows slightly better skill than in 1986. This is because the standard deviation of buoy SST (1.31 °C) in 1986 is smaller than that of buoy SST (2.20 °C) in 1988. Thus, the non-dimensional SS analysis provided a distinction between the two cases. Such a methodology is especially useful when assessing global model performance at different locations, e.g., those which have little seasonal cycle and a small standard deviation value over a yearlong period (e.g., SST in the equatorial Pacific warm pool) and others which have a large seasonal cycle (e.g., SST off the Alaska coast) or other large variability.

The model–data comparisons like those performed at (0°N, 110°W) and (52°N, 156°W) (Table 4) were applied to all the TAO and NDBC buoys. RMSD values based on yearlong time series at each buoy location are shown in Appendix A. Similar tables for ME, R and SS were also obtained (not shown). Overall, NLOM performance in predicting SST is assessed over the period 1980–1998 using a total of 340 yearlong sufficiently complete daily time series, considering each TAO and NDBC buoy for each year marked by a "+" on Tables 2 and 3. Combining the values for each statistic, we then calculate cumulative frequency of ME, RMSD, R and SS to obtain modal (median) statistics (Fig. 4). Error statistics for 340 yearlong daily time series over the time frame 1980–1998 give median ME of −0.09 °C, RMSD of 0.82 °C, R value of 0.92 and SS of 0.73.

As a final assessment, yearly mean SST and the standard deviation of SST for each year are calculated for both NLOM and buoy values using the yearlong time series at each buoy location, yielding 340 values for each. In general, yearly mean buoy SSTs are well predicted by the yearly mean model SSTs as is evident from the low scatter between the two (Fig. 5). The R values between buoy and NLOM SSTs are 0.99 and 0.94 for the yearly means and standard deviations,

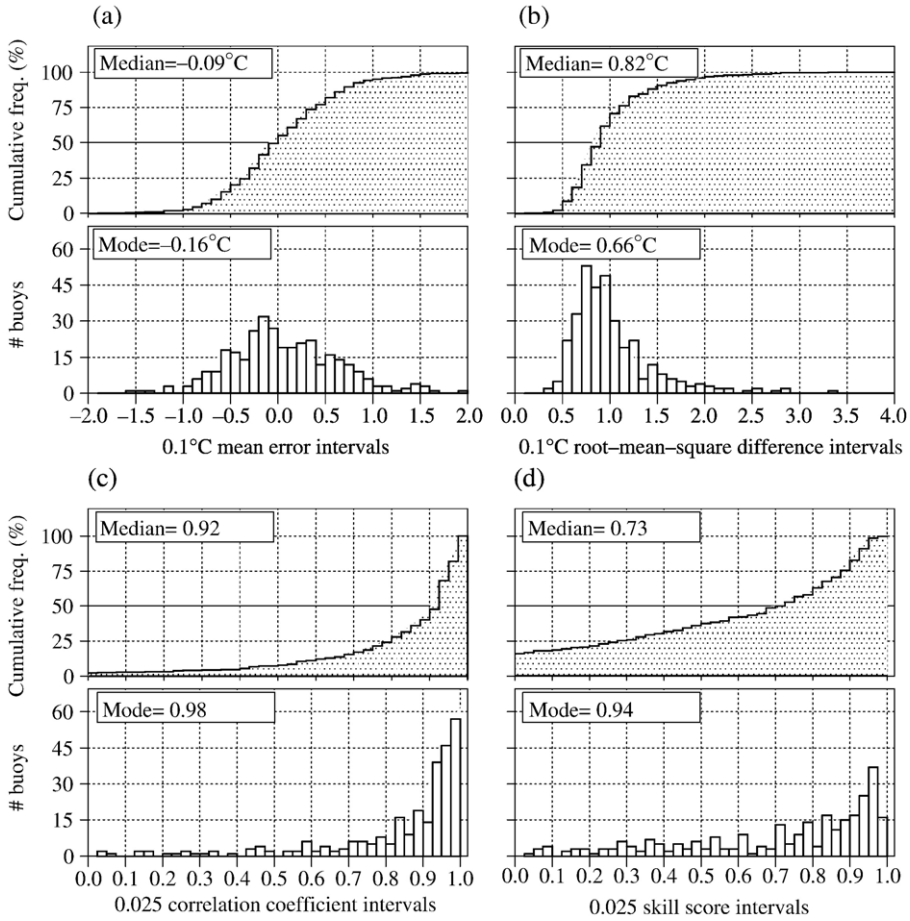

Fig. 4. Cumulative frequency and total number of cases between buoy SST and NLOM SST for the 340-year-long daily time series: (a) ME, (b) RMS difference, (c) *R*, and (d) SS. See text for description of each statistical metric. Median value corresponds to 50%. Note that the *x*-axis labels in all panels are at boundaries between histogram and cumulation intervals. Negative SS values indicate unsuccessful NLOM predictions. While they are not shown in the histograms; they are evident in cumulative frequency plots.

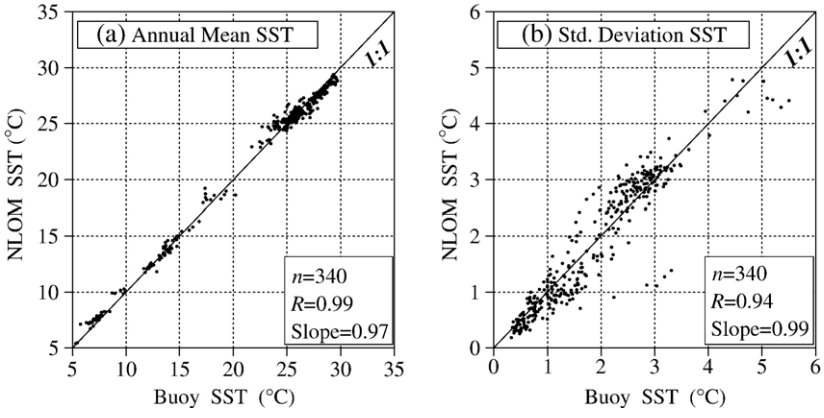

Fig. 5. Scatter plots of annual mean (S.D.) of NLOM SST vs. reference mean (S.D.) of buoy SST values. Annual means are calculated for each TAO and NDBC buoy (see Tables 2 and 3). Total number of cases (*n*), correlation coefficient (*R*) and slope of the least square line are also given.

respectively. Nearly perfect linear correspondence is evident from slope values (regression coefficient) close to 1, and this is true for both yearly mean and standard deviation values.

4.2. Mixed layer depth (MLD)

Before performing any model–data comparisons of MLD, two commonly used ocean surface layer depth definitions are introduced: (1) an isothermal layer depth (ILD), and (2) an isopycnal layer depth (MLD). The methodology for determining the ILD and MLD is fully described by Kara et al. (2000b), showing that MLD is best described using a variable density definition with $\Delta T = 0.8$ °C. In brief, the ILD [MLD] can be summarized in its simplest form as being the depth at the base of an isothermal [isopycnal] layer, where the temperature [density] has changed based on a fixed value of ΔT from the temperature [density] at a reference depth of 10 m. Note here that a fixed $\Delta\sigma_t$ is not used for MLD determination. For the isopycnal layer this means $\Delta\sigma_t = \sigma_t(T+\Delta T, S, P)\ \sigma_t(T, S, P)$ where P is set to zero. In the case of buoys with subsurface salinity as well as temperature, subsurface temperature and salinity profiles are used for calculating MLD, while in the case of NLOM, the prognostic Eq. (2) determines model MLD.

The sensitivity of the model MLD performance to the definitions of ILD and MLD is first examined using a buoy located at (15.5°N, 61.5°E). This buoy was deployed in the Arabian Sea where there was strong seasonal forcing over the time period 1994–1995,

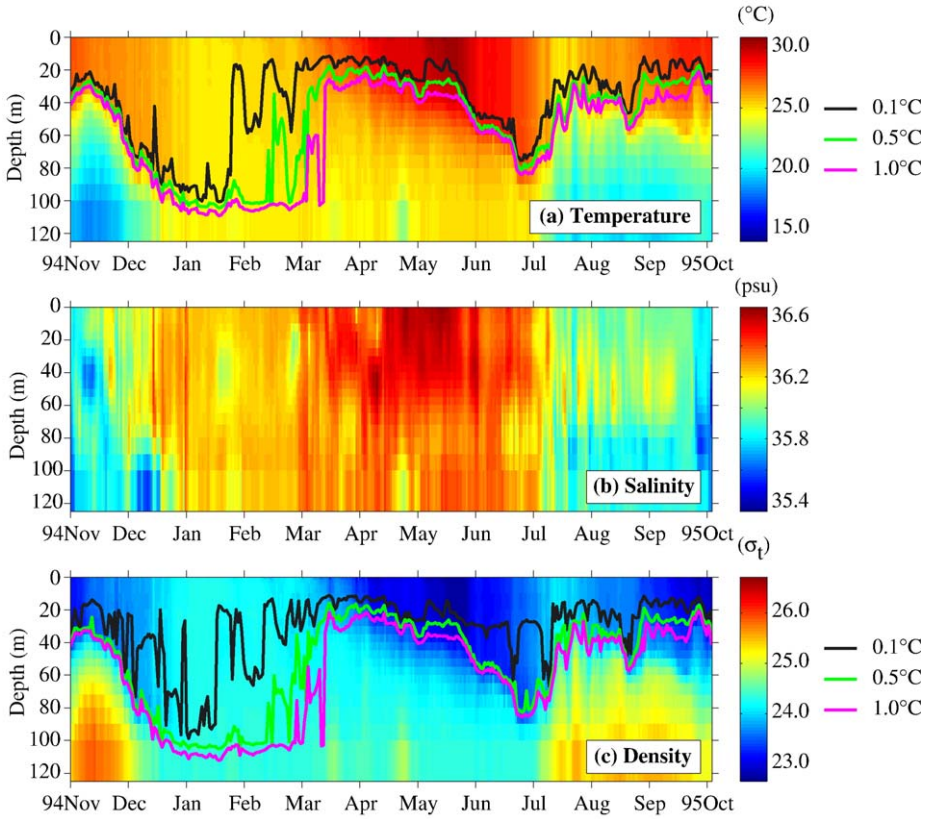

Fig. 6. Daily averaged times series of subsurface temperature, salinity and calculated density down to 150 m at the buoy located at (15.5°N, 61.5°E) during the 1994–1995 Monsoon period. Water temperature was measured from 0.17 m to 300 m depth with vertical resolution that varied from <0.5 m (in the upper 2 m) to 25 m (between 100 m and 300 m depth). Salinity was measured from 1.4 m to 250 m with resolution between 10 m and 50 m. Also shown are isothermal layer depth (ILD) overlain on the subsurface temperatures and mixed layer depth (MLD) overlain on the subsurface densities. Both ILD and MLD are defined by a ΔT criterion, with a $\Delta\sigma_t$ derived from a ΔT in the case of MLD. We have limited consideration to three ILD (MLD) obtained using ΔT criteria of 0.1, 0.5 and 1.0 °C. Note that the salinity stratification is not as important here as in the equatorial warm pool, and strong seasonal variation in ILD and MLD occurs because of strong and seasonally reversing monsoon winds. The ILD usually follows the MLD but there are significant differences among the time series of the three different layer depths calculated using ILD or MLD definitions. This is especially true during the northeast monsoon. Details of subsurface temperature and salinity measurements from this Woods Hole Oceanographic Institute (WHOI) buoy can be found in Weller et al. (1998).

including the northeast and southwest monsoons. The northeast monsoon in 1994 was characterized by steady but moderate winds, while the southwest monsoon in 1995 was characterized by strong winds (e.g., Weller et al., 1998), allowing one to examine the possibility of shallow and deep mixed layer formation. OGCM prediction of MLD is therefore a great challenge at this location. In addition, the mooring record has a continuous time series of subsurface temperature and salinities at a 7.5-min sampling interval or less. A data set like this provides an excellent opportunity to properly define MLD using a density-based criterion which depends on temperature and salinity. For simplicity, daily averages of subsurface temperature and salinities were formed from the original 7.5-min data.

The buoy ILD and MLD at (15.5°N, 61.5°E) are calculated from daily averaged subsurface temperatures and salinities following the definition of Kara et al. (2000b), as mentioned earlier. Fig. 6 includes measurements of daily averaged subsurface temperatures, salinity, and calculated incompressible density at the buoy location. The ILD and MLD are overlain on the temperature and density plots, respectively. A markedly deeper MLD is observed using the larger temperature difference criteria, a feature consistent with MLD results in other regions, e.g., in the North Pacific (Kara et al., 2000c). This indicates that choosing an incorrect layer definition to determine buoy MLD may result in serious errors in the model evaluation. Kara et al. (2003b) found that the optimal MLD in the Arabian Sea was obtained using a density-based criterion ($\Delta\sigma_t$) with ΔT=0.8 °C. Daily MLD calculated from buoy data using this definition and that predicted by NLOM using the prognostic equation (see (2)) demonstrate that they agree with each other reasonably well. The model reproduces deep MLD during the monsoon periods and very shallow buoy MLD during the inter-monsoon periods. The buoy MLD is shallow during the second half of the southwest monsoon, and the model also shows a similar feature during this period.

In a previous study (Rochford et al., 2000), daily MLD time series were obtained from the thermodynamic reduced-gravity version of NLOM configured for the Indian Ocean basin. It had three active layers and an embedded mixed layer that was confined to lie within the uppermost dynamical layer when this model was compared to daily MLD time series at the buoy location above (see Fig. 7). It failed to shallow sufficiently during the spring and fall inter-monsoon periods because a minimum MLD of 20 m was imposed for numerical stability. Since then, substantial improvements (e.g., a new advection scheme, better heat flux parameterizations, and a floating mixed layer, including a stable depth) in the NLOM embedded mixed layer have been made (Wallcraft et al., 2003). In the global NLOM mixed layers as shallow as 10 m are permitted because an improved "van Leer" advection scheme (Lin et al., 1994) was implemented. A smaller RMSD

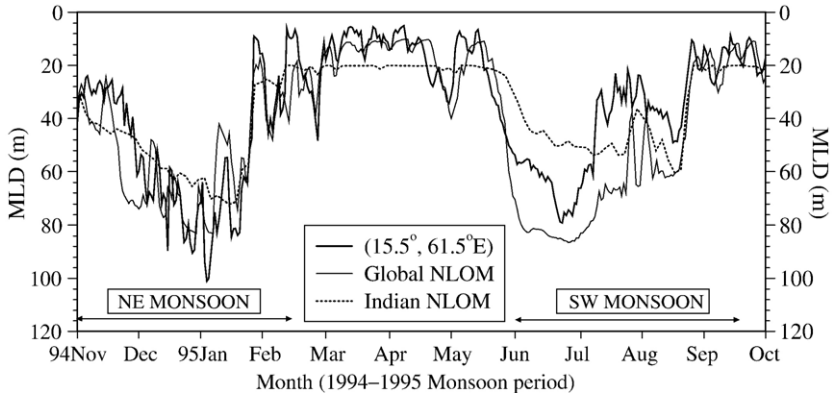

Fig. 7. Daily MLD time series from the buoy and NLOM simulations at (15.5°N, 61.5°E) in the Arabian Sea. The northeast monsoon (1 Nov 1994 to 15 Feb 1995) begins with steady northeast winds in November, lasting until February, with the winds being less intense than the southwest monsoon (1 Jun 1995 to 15 Sep 1995). The mooring provides a high-frequency set of meteorological and oceanographic observables. For example, the mooring record has continuous time series of subsurface temperature and subsurface salinity at a 7.5-min sampling interval or less, which were used to compute daily averaged buoy MLD. For the observation period, from 1 Nov 1994 to 1 Oct 1995, the ME and RMSD between daily buoy and global (Indian) model MLD values are only 8.7 m (10.4 m) and 15.4 m (19.6 m), respectively. The mean and S.D. of the observed MLD are 35.8 m and 23.1 m, respectively. The global NLOM (Indian NLOM) is able to reproduce the seasonal cycle of MLD well with a R value of 0.81 (0.80). In addition to differences in mixed layer submodel formulation, some of the differences in the results from the two models are due to slight differences in the atmospheric forcing used in the simulations.

(15.4 m vs. 19.6 m for global NLOM vs. Indian NLOM) between model and buoy MLD is obtained as a consequence. The global NLOM outperforms the Indian Ocean NLOM in reproducing the deep MLD during the northeast monsoon because the embedded mixed layer in the global model can extend below the top dynamical layer, unlike the Indian Ocean precursor.

Finally, the model performance in predicting MLD is also examined using TAO buoys. The buoys reporting subsurface temperatures can be considered under two categories based on the water depths at which subsurface measurements are taken (Table 5). There are 16 buoys from which daily subsurface temperatures are available for most of the years from 1990 to 1998. They yield 104 yearlong daily subsurface temperature time series. Daily salinity measurements were not available from the TAO buoys, not enabling us to determine true surface layer depth based on the MLD definition. Thus, it is worthwhile to ask what ΔT gives an ILD that corresponds best to the optimal definition of MLD at each buoy location (Kara et al., 2003b). Using the global climatological monthly MLD and ILD (isothermal layer depth) fields from the NMLD climatologies, Kara et al. (2002b) presented annual and monthly mean ΔT fields. These fields provide ΔT values for determining an ILD that is approximately equivalent to the optimal MLD at each buoy location. Although ΔT values can be quite variable from month to month (not shown), using a location-specific ΔT difference value is a good approximation to use in calculating an ILD-defined MLD at each TAO buoy (Table 5). One reason for using the annual mean is that the monthly mean ΔT values have small standard deviation at each location over the annual cycle.

Table 5
A list of daily subsurface temperature availability from TAO buoys used for validating NLOM subsurface temperatures from 1990 to 1998

TAO buoy	Set I	Set II	Mean (°C)	Min. (°C)	Max. (°C)	S.D. (°C)
(0°N, 110°W)	+	.	0.63	0.54	0.78	0.07
(0°N, 125°W)	+	.	0.62	0.47	0.83	0.10
(0°N, 155°W)	.	+	0.68	0.61	0.75	0.04
(0°N, 170°W)	.	+	0.72	0.62	0.89	0.08
(2°N, 110°W)	+	.	0.56	0.47	0.69	0.07
(2°S, 110°W)	+	.	0.68	0.54	0.82	0.10
(2°S, 125°W)	+	.	0.57	0.38	0.74	0.11
(5°N, 155°W)	.	+	0.62	0.46	0.75	0.11
(5°N, 165°E)	.	+	0.32	0.19	0.44	0.08
(5°S, 110°W)	+	.	0.65	0.35	0.97	0.18
(5°S, 140°W)	+	.	0.44	0.13	0.69	0.16
(5°S, 155°W)	.	+	0.63	0.42	0.79	0.12
(8°N, 110°W)	+	.	0.42	0.20	0.67	0.14
(8°N, 170°W)	.	+	0.43	0.22	0.58	0.12
(8°S, 110°W)	+	.	0.55	0.13	1.04	0.25
(8°S, 155°W)	.	+	0.53	0.34	0.71	0.11

The water depths at which subsurface measurements were taken are 1, 20, 40, 60, 80, 100, 120, 140, 180, 300 and 500 m for set I and 1, 25, 50, 75, 100, 125, 150, 200, 250, 300 and 500 m for set II.
In the table, a '+' sign indicates that daily subsurface temperatures from set I or set II buoys are available for that particular buoy location, and they are used for model–data comparisons of MLD. Also given are annual mean ΔT values based on the Naval Research Laboratory Mixed Layer Depth (NMLD) climatologies (Kara et al., 2003a,b), minimum and maximum ΔT values during the 12-month period along with standard deviation of ΔT. Annual mean ΔT values shown in this table are used for calculating ILD defined MLD from subsurface temperatures at each buoy location. It should be noted that an assessment of instrumental accuracies indicates about 0.1 °C for subsurface temperature (Freitag et al., 1994). Similar to SST, a few missing subsurface temperature values were filled using a cubic interpolation before performing any model–data comparisons.

As an example, the optimal MLD at (5°N, 165°E) is obtained using an annual mean ΔT value of 0.32°C. Fig. 8 shows daily optimal MLD calculated from subsurface temperatures as well as the MLD from NLOM from 1995 to 1998. The MLD can vary considerably from day to day, and inter-annual variability is clearly evident at this location. There is no significant underestimation or overestimation of buoy MLD by the model. In addition to the classical statistical metrics introduced in Section 4, we also use normalized RMSD (NRMSD) for model–data comparisons of MLD to further examine the model performance. It is defined as $\mathrm{NRMSD}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(Y_i - X_i)}{X_i}\right]^2$. The NRMSD replaces SS because the former reduces skewness in the distribution of the errors, i.e., the standard deviation for mixed layer too shallow is much less than the standard deviation for mixed layer too deep without normalization. For example, winter MLDs are usually deeper and have larger standard deviation than summer MLDs (Kara et al., 2003a). It is noted here that the SS depends on the standard deviation of the reference data set, while NRMSD depends on the corresponding mean. Percentage error in MLD expressed by NRMSD is also a meaningful quantity that is easily understood. Essentially, at (5°N, 165°E), NRMSD values are 0.29, 0.21, 0.17 and 0.27 from 1995 to 1998, respectively, showing that MLD predictions from NLOM deviate 17% to 29% from the buoy values (see Fig. 8).

Based on the 104 yearlong daily subsurface temperature time series from 16 buoys, median error statistics yielded a ME of −13.8 m, a RMSD of 18.2 m, and a NRMSD of 0.29. The correlation between the 104 pairs of annual mean MLD from the

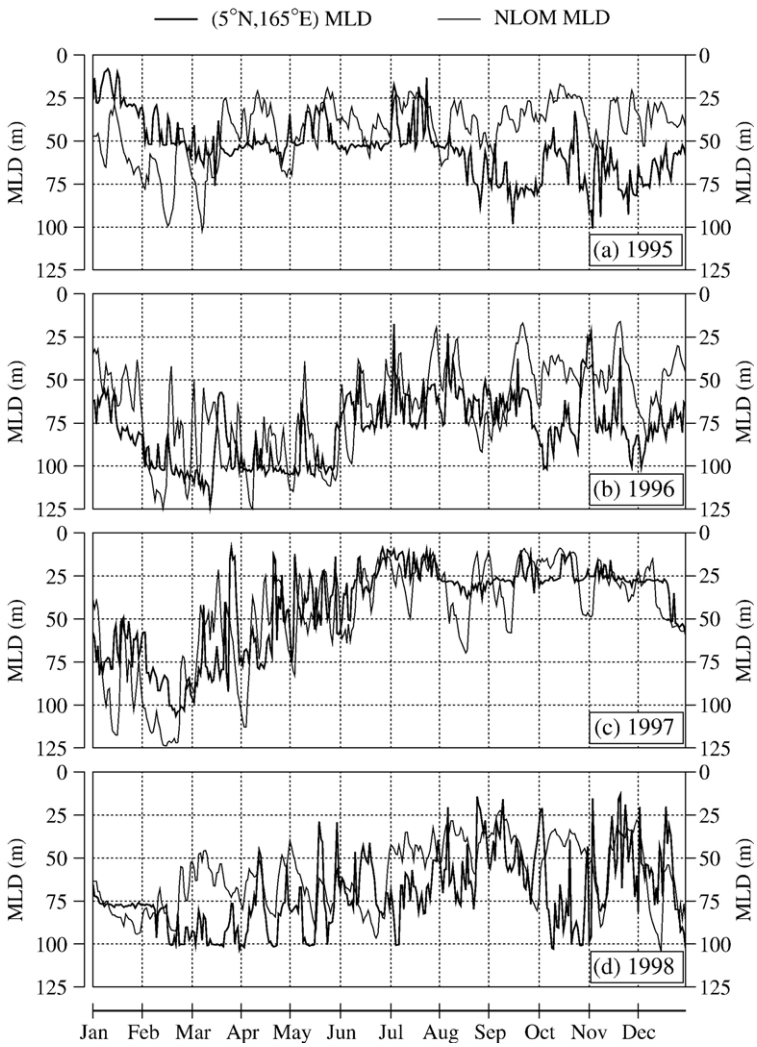

Fig. 8. Daily MLD at 5°N, 165°E (thick line) and modeled MLD from the inter-annual 1/2° global NLOM simulation (thin line) with ECMWF forcing in (a) 1995, (b) 1996, (c) 1997 and (d) 1998. Note that buoy MLD is calculated from subsurface temperatures using an ILD definition with $\Delta T = 0.32$ °C at this particular location, as explained in the text. ME values are −10.6, −13.1, 7.9, and −16.9 m from 1995 to 1998, respectively. In the figure, the *x*-axis is labeled starting from the beginning of each month.

model and the TAO buoys is 0.76, and the corresponding correlation between the MLD standard deviations is 0.59 (Fig. 9). The results indicate that NLOM provides reasonable MLD simulations in the regions tested.

5. Model validation with climatology

A monthly mean model SST climatology is formed using the inter-annual model simulations from 1979 to 1998. This allows additional spatial and temporal model–data comparisons over the global ocean. For this purpose the climatological monthly mean NLOM SSTs are compared to climatological monthly mean COADS SST (da Silva et al., 1994). The January mean SST from NLOM is obtained by averaging all January values from 1979 to 1998 at each model grid point over the global ocean. This process is repeated for each month, yielding time series of monthly mean climatological SST values at each model grid point. Similarly, time series of the monthly mean climatological SST is obtained from the COADS data set and interpolated onto the NLOM grid. The annual mean and seasonal cycle of NLOM SST are evaluated by comparing 12 monthly NLOM SST with COADS SST at each model grid point using various statistical metrics (see Section 3). Global maps for these statistical metrics (Fig. 10) are discussed in this section.

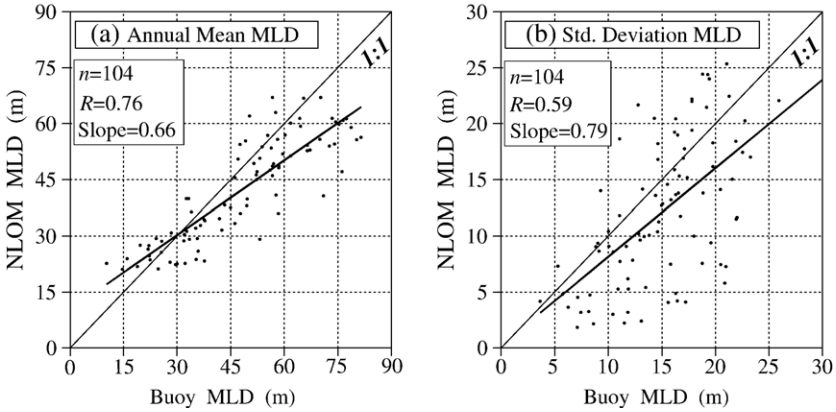

Fig. 9. Scatter plots of annual mean (standard deviation) of NLOM MLD versus reference mean (S.D.) from buoy MLD values. The least squares t is the dark solid line. Annual means are calculated for each TAO buoy and for each year, separately (see Tables 2 and 3). The total number of cases (n), correlation coefficient (R) and slope of the least square line are also given.

In general, ME between NLOM and COADS SST over the 12 months shows that model errors in predicting the SST in the open oceans are usually small, with bias values smaller than 0.5 °C over most of the global ocean (Fig. 10a). The annual mean of the model SST is cooler by about 1 °C along the Kuroshio and Gulf Stream pathways. While resolution of 1/2° is not adequate for realistic simulation of these current systems, leading to pathway errors and advection that is too weak (Hurlburt et al., 1996; Hurlburt and Metzger, 1998; Hurlburt and Hogan, 2000), NLOM with an embedded mixed layer is still able to simulate SST in these regions fairly well. The annual mean of SST in the equatorial warm pool region is well predicted by NLOM. In this region SST bias is less than 0.5 °C. RMSD values between NLOM and COADS SSTs are usually less than 1 °C, but large RMSD values are noted between ≈40° and ≈60°N (Fig. 10b).

The SST seasonal cycle is well predicted by the model because R values are close to 1 over the most of the global ocean (Fig. 10c). The model had difficulty in obtaining a correct seasonal cycle in some places south of ≈40°S and in the western equatorial Pacific warm pool with non-significant R values <0.6. Because SST variability (i.e., small standard deviation) and bias in these regions is so small the RMSD values are small. The non-dimensional SS map (see Fig. 10d), which accounts for both B_{cond} values (Fig. 10e) and B_{uncond} values (Fig. 10f), clearly shows that NLOM is able to simulate SST well over most of the global ocean because the SS is high and the two non-dimensional biases are usually <0.1. The SST skill in some parts of the Southern Hemisphere and the western equatorial Pacific warm pool are <0.1 and even <0, indicating poor model skill in simulating the variability in these regions, although RMSD values are small.

The model performance in predicting SST is also determined using the zonally averaged values of the statistical maps (Fig. 10). Zonal averaging for each statistical metric is performed along 1° latitude belts from 65°N to 72°S. ME values of <0.5 °C and RMSD values of <1.0 °C are evident at nearly all latitude belts (Fig. 11). The model yields R values >0.8 and positive SS values in nearly all latitude belts. The largest B_{cond} and B_{uncond} values are seen between 40°S and 50°S. While this bias is not reflected in R, it causes some SS values <0.1 at these latitudes.

When the model was run climatologically using wind forcing from the Hellerman and Rosenstein (1983) monthly mean wind stress climatology with 6-h ECMWF variability added and monthly thermal forcing from COADS, similar results were obtained (Kara et al., 2003a), except that NLOM SST was slightly warmer (≈0.5 °C) than COADS SST at latitude belts north of 40°N. Smaller positive ME values at these latitude belts are seen in the results presented here, where the annual mean from the Hellerman and Rosenstein (1983) replaced the long-term mean from ECMWF, and thermal forcing from ECMWF was used over the time frame 1979–1998. Thus, differences in statistical results between these two cases are due primarily to the differences in atmospheric thermal forcing for the model. The model simulations in this paper used high-resolution wind and thermal forcing and a model climatology was created for only 20 years, while the original climatological simulations used monthly thermal forcing and the monthly climatological Hellerman and Rosenstein (1983) wind stress with

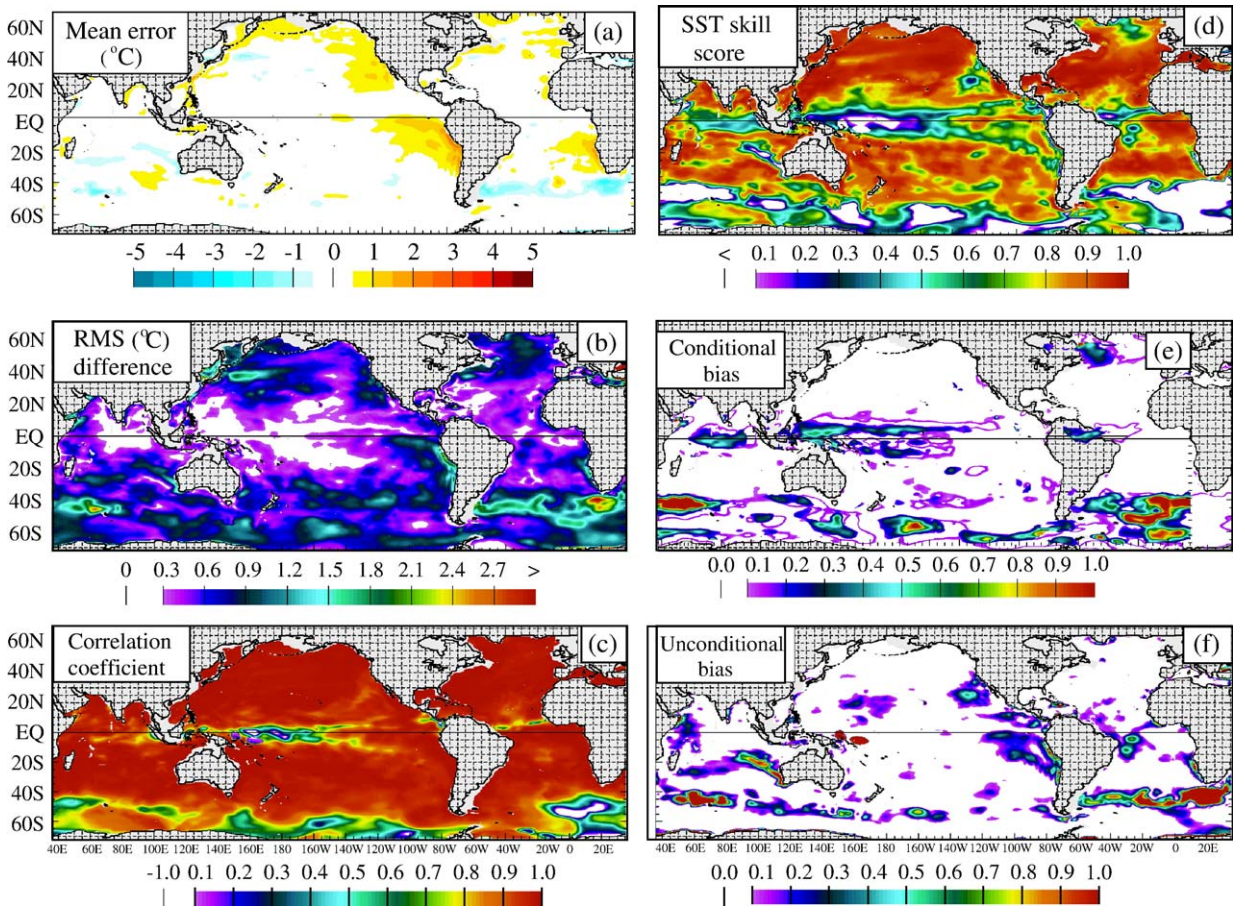

Fig. 10. Comparison of the monthly mean SST between the COADS climatology (1949–1995) and NLOM monthly mean climatology over the global ocean during 1979–1998. Left panels: (a) Mean error (ME) with a contour interval of 0.5 °C, (b) root-mean-square difference (RMSD) with a contour interval of 0.03 °C and (c) correlation coefficient (R) with a contour interval of 0.01. Right panels: (d) non-dimensional skill score (SS) with a contour interval of 0.01, (e) non-dimensional conditional bias (B_{cond}) with a contour interval of 0.01, and (f) non-dimensional unconditional bias (B_{uncond}) with a contour interval of 0.01. The fact that R ignores biases in means and amplitudes suggests that it may overestimate the performance of a model. Thus, SS is also used for model evaluation. It should be noted that color bars for the B_{cond} and B_{uncond} maps start from 0.0 because these two statistical metrics can never be negative as explained in the text. SST RMSD is small in the warm pool and Antarctic; however, a combination of small R and large B_{cond} results in relatively low SS values. The global averages of ME and RMSD are −0.11 °C and 0.56 °C, respectively. The global average of R value is 0.90. Similarly, the global averages of SS, B_{uncond} and B_{cond} are 0.65, 0.09 and 0.08, respectively.

the addition of 6 h to sub-monthly wind stress anomalies.

6. The impact of OGCM resolution and atmospheric forcing products on model Simulated inter-annual SST accuracy

The purpose of this section is to investigate two issues: (1) the sensitivity of simulated daily inter-annual SSTs to NLOM resolution, particularly eddy-resolving vs. non-eddy resolving, and (2) the sensitivity to choice of atmospheric forcing product. Regarding the first issue, the determinicity of ocean model responses to atmospheric forcing is reduced in an eddy-resolving model due to flow instabilities. Ocean eddies along with meandering currents and fronts are ubiquitous in the world ocean (see http://www.ocean.nrlssc.navy.mil/global_nlom; Le Traon et al., 1998; Hurlburt and Hogan, 2000; Smedstad et al., 2003; Shriver et al., in press). Regarding the second issue, the sensitivity of SST to the choice of wind or thermal forcing could be used as a discriminator for the quality of the atmospheric forcing product as well as OGCM performance. Thus, NLOM skill in predicting SST is further examined using fine-resolution model simulations, including comparisons to the 1/2° global model simulation.

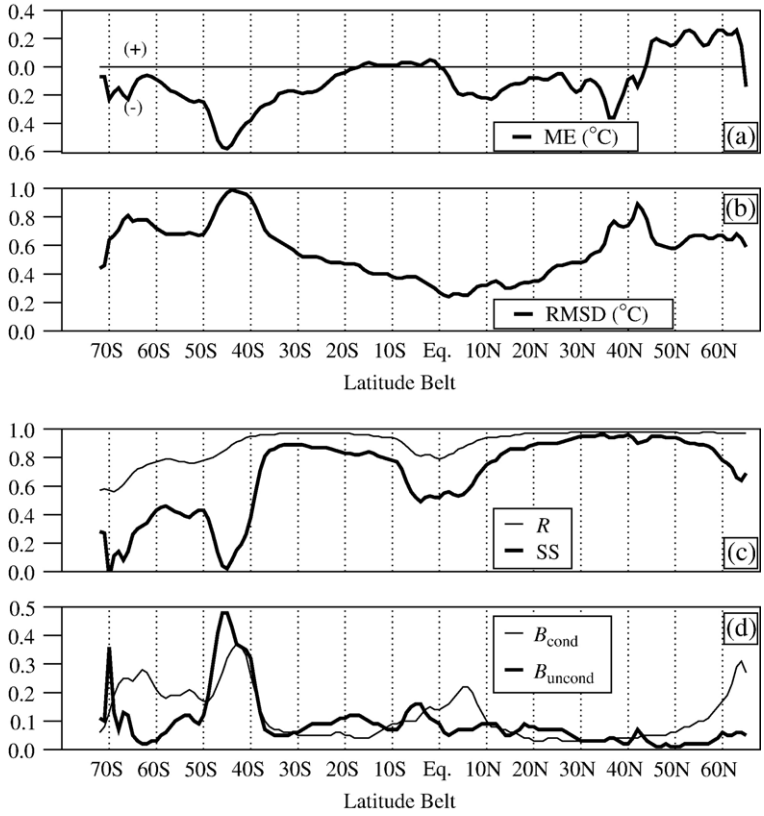

Fig. 11. Zonally averaged statistical comparisons between the monthly mean NLOM and COADS SSTs (see Fig. 10) when NLOM was run using 6-h wind and thermal forcing from ECMWF over the time frame 1979–1998: (a) ME, (b) RMSD, (c) R and SS (dark solid line), and (d) B_{cond} and B_{uncond} (dark solid line). As explained in the text, monthly mean NLOM SST was calculated over 20 years, and compared with the monthly mean climatological COADS SST. Zonal averaging is performed for each 1° latitude belt from 72°S to 65°N over the global ocean. Zonally averaged statistics further confirm model success with $R>0.8$ and SS>0 at almost all latitude belts.

The finer resolution models are a 1/8° global model and a 1/16° Pacific model (north of 20°S) as described in Section 2.2. The latter model is forced using two atmospheric products, ECMWF and NOGAPS. There is no assimilation of SST data in any of the model simulations, i.e., all models are atmospherically forced. The hybrid wind approach described in Section 2.2 is used in all simulations. Initializing inter-annual simulations from a spun-up climatologically forced simulation with the same long-term mean in the wind forcing (1) reduces the adjustment to the inter-annual forcing product, (2) reduces differences between the long-term temporal mean circulation in climatologically and inter-annually forced simulations, and (3) allows initialization of simulations with different inter-annual forcing products from the same climatological spin-up. Because the NOGAPS forcing is only available since 1990, the period of comparison here is 1990–1998.

First we examine the sensitivity of basin-wide model SST to input wind and thermal forcing. Using different combinations of wind and thermal forcing from ECMWF and NOGAPS in forcing the 1/16° Pacific model, three inter-annual model simulations were performed: expt 1 used wind and thermal forcing from ECMWF (1990–1998); expt 2 used wind forcing from NOGAPS and thermal forcing from ECMWF (1990–1998), and in expt 3 both wind and thermal forcing were from NOGAPS, but the simulation was run only for 1998 because thermal forcing from NOGAPS is not available for earlier years. In the case of NOGAPS wind forcing and ECMWF thermal forcing, note that NOGAPS wind speed is used in the bulk formulae used for calculation of latent and sensible heat fluxes (Kara et al., 2002a), but all the other fields used in the thermal forcing are from ECMWF (see Section 2).

The only common year among the three simulations is the El Niño to La Niña transition year (1998), which is characterized by an abrupt transition from unusually large warm to cold ocean temperature anomalies in the equatorial Pacific. The 1997–1998 ENSO event was

unusually strong (e.g., McPhaden and Yu, 1999; Enfield, 2001; Harrison and Vecchi, 2001) with the eastern Pacific much cooler than normal, and with the cool water extending farther westward than usual during the strong La Niña 1998 portion of the event (e.g., McPhaden, 1999; Vialard et al., 2001). Therefore, predicting SST during this event is a challenge for an OGCM, especially for an atmospherically forced OGCM which is predicting a direct model response to the atmospheric forcing without assimilation of SST data.

Annual mean SST difference maps for the year 1998 from expt 2 (Fig. 12b and e) and expt 3 (Fig. 12c and f) can be compared to examine the effects of thermal forcing on SST simulation when wind forcing is the same for both experiments. Effects of both wind and

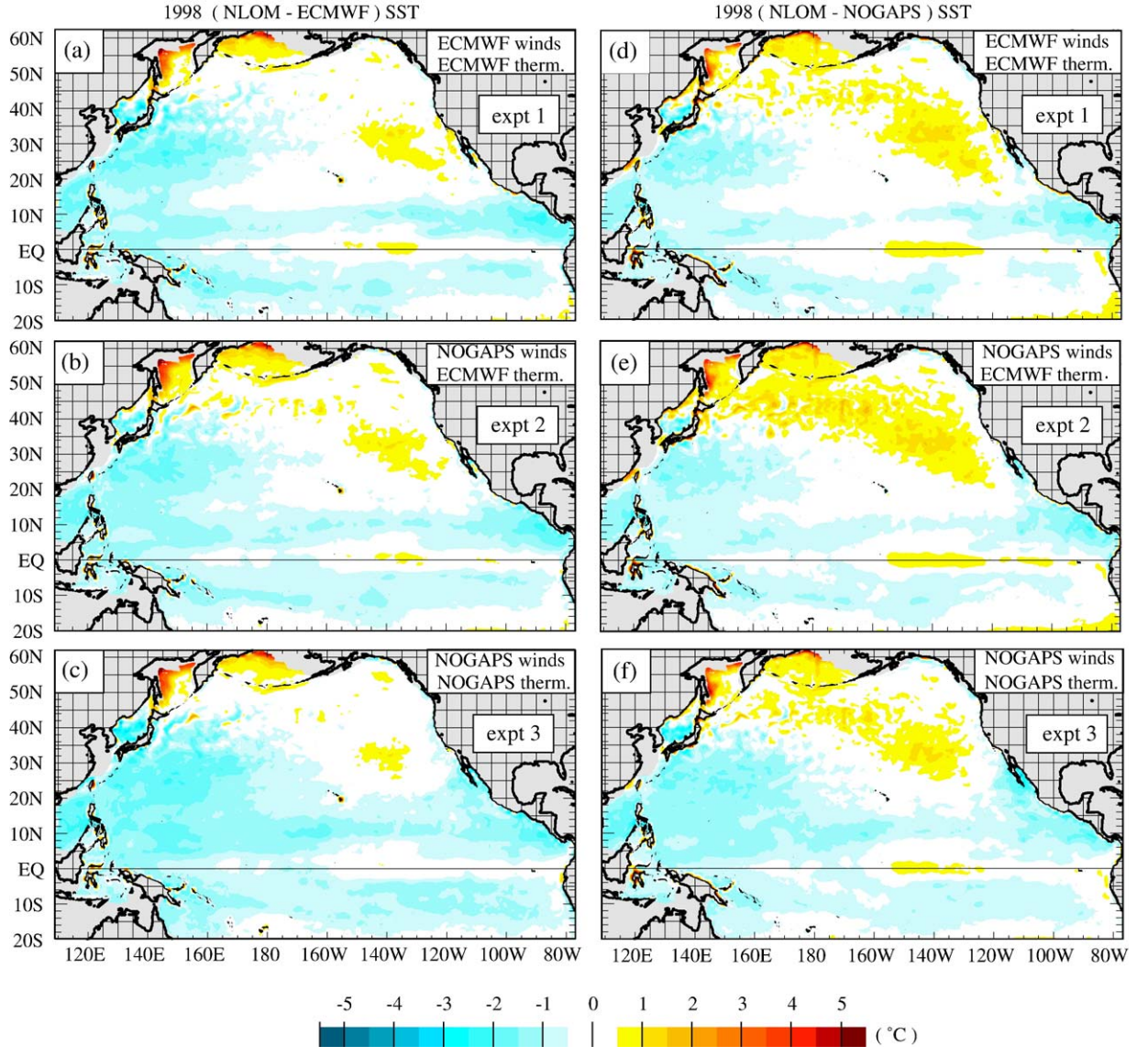

Fig. 12. Annual mean SST difference between 1/16° Pacific NLOM and ECMWF (panels (a) to (c)) and between NLOM and NOGAPS (panels (d) to (f)) during 1998. The model was driven by (a) and (d) ECMWF wind and thermal forcing, (b) and (e) NOGAPS winds and ECMWF thermal forcing, and (c) and (f) NOGAPS wind and thermal forcing. Note that wind stress was constructed from 6 hourly 10-m winds for ECMWF and that the NOGAPS surface stress product was used. There is no assimilation of SST data by the ocean model. The domain-wide RMS of the 1998 mean SST differences are 0.68°, 0.71°, and 0.75 °C for (a), (b), and (c), respectively; and they are 0.74°, 0.79°, and 0.76 °C for (d), (e), and (f), respectively. Thus, the RMSD statistics for the annual mean bias in 1998 show only minor differences.

thermal forcing can be seen in Fig. 12a and c (NLOM–ECMWF SSTs), and Fig. 12d and f (NLOM–NOGAPS SSTs). There are only minor differences from 30° to 50°N, and the spatial structure of the SST bias field is essentially the same. This is true regardless whether the model is validated against ECMWF SST or NOGAPS SST.

Similar to the analysis presented for the 1/2° global model (Section 4.1), model–data comparisons are further performed using daily buoy SST time series for all simulations presented in this paper, i.e., the 1/2° global model, the 1/8° global model, the 1/16° Pacific model (with ECMWF wind and thermal forcing), and the 1/°16 Pacific model (with NOGAPS wind and ECMWF thermal forcing) during 1990–1998, which is the common time period covered by the simulations. A total of 127 yearlong daily SST time series from 20 TAO buoys and 67 yearlong daily SST time series from 13 NDBC buoys were used for model–data comparisons. When the 1/16° NLOM was run using ECMWF/HR and NOGAPS/HR wind forcing (see Section 2.2) and the same thermal forcing (ECMWF), SST errors with respect to the buoy SSTs did not differ significantly. In particular, ≈30% of ME values are between −0.6° and −0.3 °C and ≈50% of RMS differences are between 0.6° and 0.9 °C for both simulations during 1990–1998 (Fig. 13).

Finally, median statistics for TAO, NDBC and all buoys (TAO and NDBC) are summarized in Table 6 to compare performance of the model when using various model configurations (1/2° global, 1/8° global, and 1/16° Pacific) and various atmospheric forcing products (ECMWF and NOGAPS) as well. Overall, the finer resolution (1/16°) NLOM performs slightly better at NDBC buoy locations than at TAO buoy locations. When the model uses wind forcing from ECMWF (NOGAPS) but the same thermal forcing (ECMWF in both cases), model SST comparisons to the 194 yearlong SST time series gave median values of 0.84 °C (0.84 °C) for RMS difference, −0.35 °C (−0.33 °C) for ME, 0.80 (0.77) for *R*, and 0.46 (0.45) for SS. Specifically, use of wind forcing from ECMWF rather than NOGAPS provided slightly better SST simulation at TAO buoy locations; while, use of wind forcing from NOGAPS

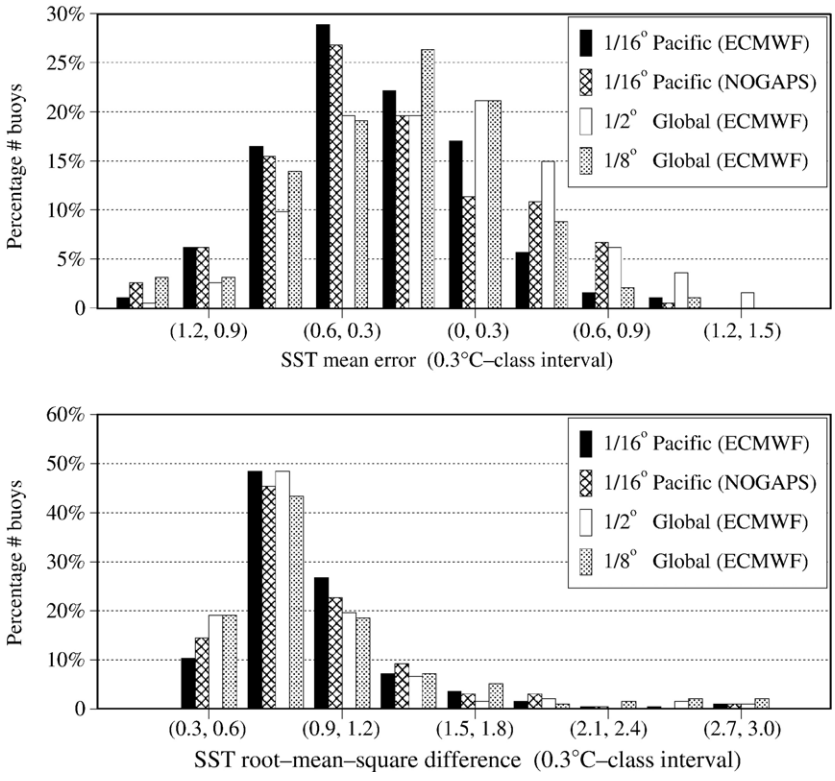

Fig. 13. Percentage histogram of ME and RMS differences between buoy SSTs and SSTs simulated by the 1/16° NLOM forced with ECMWF and NOGAPS hybrid wind stresses. Also shown are results from the 1/2° and 1/8° global models during the same time period. The statistics include all 194-year-long daily time series: (a) ME in 0.3 °C-class intervals, and (b) RMS difference in 0.3 °C-class intervals. Note that a ME class interval of (0,0.3) on the *x*-axis, for example, denotes ME values of ≥0 °C but <0.3 °C between NLOM SST and buoy SST during 1990–1998. See Table 6 for detailed median error statistics.

Table 6
Median error statistics for yearlong daily time series over the time frame 1990–1998 when NLOM with an embedded mixed layer was forced with ECMWF/HR and NOGAPS/HR hybrid wind stresses, separately

NLOM simulation	Wind forcing	Buoy SST	Total buoys	RMS (°C)	ME (°C)	*R*	SS
1/16° Pacific	ECMWF	TAO	127	0.85	−0.39	0.72	0.04
1/16° Pacific	NOGAPS	TAO	127	0.88	−0.50	0.66	0.02
1/2° Global	ECMWF	TAO	127	0.68	−0.21	0.82	0.13
1/8° Global	ECMWF	TAO	127	0.62	−0.16	0.87	0.18
1/16° Pacific	ECMWF	NDBC	67	0.84	−0.11	0.95	0.88
1/16° Pacific	NOGAPS	NDBC	67	0.76	−0.01	0.96	0.88
1/2° Global	ECMWF	NDBC	67	0.66	−0.07	0.97	0.91
1/8° Global	ECMWF	NDBC	67	0.61	−0.24	0.97	0.95
1/16° Pacific	ECMWF	All	194	0.84	−0.35	0.80	0.46
1/16° Pacific	NOGAPS	All	194	0.84	−0.33	0.77	0.45
1/2° Global	ECMWF	All	194	0.68	−0.17	0.86	0.69
1/8° Global	ECMWF	All	194	0.61	−0.20	0.91	0.74

Also shown are results from the 1/2° and 1/8° global model forced with ECMWF/HR winds during the same time period. Thermal forcing is from ECMWF in all cases because the NOGAPS archive does not include heat fluxes prior to 1998. Median values for TAO and NDBC buoys are shown separately. There are 127-year-long daily SST (127×365 days) time series from TAO buoys and 67-year-long daily SST (67×365 days) time series from NDBC buoys used for model–data comparisons during 1990–1998. For leap years the number of days is 366 rather than 365 in a year.

rather than ECMWF provided slightly better SST simulation in comparison to the NDBC buoys. In comparison to the 1/2° global model results, relatively large ME and RMS difference are evident from the high resolution Pacific simulations (see Table 6). As expected, the coarse resolution global model does not resolve eddies well, and the fine resolution Pacific model has substantial non-deterministic eddy activity, resulting in non-deterministic SST variability and relatively large SST prediction errors. However, the most accurate results are from the 1/8° global model, which has substantially weaker non-deterministic variability than the 1/16° Pacific model.

7. Conclusions

Global ocean models are powerful tools for ocean nowcasting and forecasting in addition to exploring past events. Testing model ability to reproduce a wide variety historical variability is essential in evaluating model performance. In this paper, the inter-annual simulations from the atmospherically forced (1/2° global, 1/8° global, and 1/16° Pacific) NLOM with an embedded floating mixed layer are used in demonstrating a systematic methodology for model–data comparisons of sea surface temperature (SST) and mixed layer depth (MLD), which can be applied to any inter-annually forced basin scale or global OGCM. 6-h forcing from ECMWF is used for all three models spanning 1979–1998, and 6-h forcing from NOGAPS is used for the 1/16° Pacific configuration only during 1990–1998. There was no assimilation of any observed buoy or other inter-annual oceanic data in the model simulations.

Detailed comparison between atmospherically forced inter-annual OGCM simulations and many daily buoy time series over many years has not been done before to our knowledge. The information in this paper is sufficient that the results can be used as a benchmark for many exact comparisons between NLOM and other inter-annually forced OGCM simulations of SST. This includes results of comparisons to buoy time series in specific years and at specific locations (see Appendix A).

The validation for the 1/2° NLOM was made for each individual buoy in each year during 1980–1998 where sufficiently complete daily SST and subsurface temperature time series were available. Error statistics with respect to TAO and NDBC buoys for 340 yearlong daily SST time series over the time frame 1980–1998 gave median mean error (ME) of −0.09 °C, RMS SST difference (RMSD) of 0.82 °C, correlation (*R*) of 0.92 and skill score (SS) of 0.73. The model performance in predicting monthly mean SST averaged over 1980–1998 was also investigated by comparing the annual mean and seasonal cycle of SST obtained from the model simulations to those from a SST climatology at each grid point over the global ocean. The results demonstrate that the SST errors in the open ocean are small (<0.5) over most of the global ocean. While RMSDs are small in the western equatorial Pacific warm pool, the correlation and skill score are poor. Because of the very small amplitude of the seasonal cycle and the SST variability in this region, higher accuracy in (a) the atmospheric forcing (including salinity forcing, which is not used here) and in (b) the numerical model are needed to accurately represent this variability than is required in most other regions. High southern latitudes are another region where the model tends to show poor performance in representing the seasonal cycle, i.e., low skill score and relatively high RMSD. Contributors to poor model performance are lack of an ice model and relatively high sensitivity to salinity forcing, which is not used. It is also a region of sparse data for the COADS SST climatology and the atmospheric forcing.

Extensive model–data comparisons were also performed for MLD using 104 yearlong daily subsurface temperature time series from 16 TAO buoys. A

variable temperature difference (ΔT) value was used for determining the optimal MLD at each buoy location. The buoy-specific ΔT values were obtained from global isothermal and mixed layer depth climatologies. Such information allowed us to determine optimal buoy MLD from subsurface temperature data alone by using an equivalent isothermal layer depth. The median NLOM MLD deviation is 29% of the MLD value from the TAO buoys. Median ME and RMSD are −13.8 and 18.2 m, respectively, indicating a shallow bias from the model MLD. In addition, comparisons between daily MLD time series from the 1/2° global NLOM and a buoy in the Arabian Sea clearly revealed the importance of using floating mixed layer model that is allowed to extend through multiple model layers rather than a non-floating mixed layer model restricted to the top model layer.

The SST results were also compared among the 1/2° global, 1/8° global , and 1/16° Pacific models. Since the atmospheric forcing from NOGAPS was not available prior to 1990, these comparisons were performed from 1990 to 1998. Over 1990–1998, the median RMS SST difference is 0.84 °C for the 1/16° Pacific model (with both ECMWF and NOGAPS forcings), 0.68 °C for the 1/2° global model (ECMWF forcing) and 0.61 °C for the 1/8° global model (ECMWF forcing). The finer resolution (1/8°) global model slightly outperforms the coarser resolution (1/2°) global model (skill score of 0.74 vs. 0.69) in comparison to 194 yearlong daily buoy SST time series. To examine the model sensitivity to the atmospheric forcing, the mixed layer model in the 1/16° Pacific model was used for inter-annual simulations with realistic 6-h atmospheric forcing from two operational weather center archives: ECMWF and NOGAPS. Three inter-annual model simulations were performed using different combinations of wind and thermal forcing from ECMWF and NOGAPS during 1998, when the equatorial ocean experienced a sharp transition from a strong El Niño to a strong La Niña event. The 1/16° Pacific model simulates much more non-deterministic variability due to flow instabilities than the 1/2° and 1/8° global models and exhibits slightly larger RMSD in comparisons to the buoy time series. However, ECMWF vs. NOGAPS forcing yielded only minor differences with NOGAPS wind forcing, giving slightly lower RMSD in comparison to mid-latitude buoy time series. Overall, the results are not very sensitive to the model resolution or the choice between the two forcing products.

We presented the accuracy and shortcomings of the SST and MLD simulations from an atmospherically forced OGCM (NLOM) on inter-annual time scales. A key result is that 6-h wind and thermal forcing products exist that have sufficient accuracy to allow realistic simulated SSTs on daily to inter-annual time scales in an OGCM. The SSTs simulated here are typically accurate to within <1 °C RMSD when there is no assimilation of SST data and when the model SST is used in the calculation of latent and sensible heat fluxes. The results of this evaluation study provide an essential indication that NLOM is a suitable platform for SST data assimilation and forecasting, the primary motivation for this study. A 1/16° global NLOM is already the model component of a data-assimilative eddy-resolving global ocean prediction system that has been running in real time since 18 Oct 2000, and which became an operational product of the U.S. Naval Oceanographic Office (NAVOCEANO) on 27 Sep 2001 (Smedstad et al., 2003). Real-time and archived results can be seen on the web at http://www.ocean.nrlssc.navy.mil/global_nlom. An upgrade of this system to 1/32° resolution has been running in near real time since Nov 2003 (with results at the same web address) and is planned to replace the 1/16° system (Shriver et al., in press).

Acknowledgments

We would like to extend our special thanks to A. J. Wallcraft, E. J. Metzger, O. M. Smedstad, L. F. Smedstad and J. F. Shriver of the Naval Research Laboratory (NRL) at the Stennis Space Center for numerous discussions. Additional thanks go to the National Data Buoy Center (NDBC) and M. McPhaden of the Tropical Ocean Array (TAO) project office for making daily SST and subsurface temperature data available. The NDBC data were provided via the National Oceanographic Data Center (NODC). Appreciation is extended to two anonymous reviewers for their constructive suggestions. The numerical simulations were performed under the Department of Defense High Performance Computing Modernization Program, using a SGI Origin 2000 and a Cray T3E at the Naval Oceanographic Office, Stennis Space Center, Mississippi and a Cray T3E at the Arctic Region Supercomputer Center, Fairbanks, Alaska. This work was funded by the Office of Naval Research (ONR), and is a contribution to the 6.2 Basin-Scale Prediction System project under program element 602435N and to the 6.1 Dynamics of Low Latitude Western Boundary Currents project under program element 601153N. This is contribution NRL/JA/7304-03-0051 and has been approved for public release.

Appendix A. Statistical tables

Here we provide statistical evaluation of NLOM SST in comparison to each individual buoy in each year, separately. These tables allow the performance of any other model to be evaluated against the NLOM SST results when using same buoy locations and years. It should be noted that a direct intercomparison between various types of numerical models (e.g., between level and isopycnal ocean models) may not be as simple as it seems, because each OGCM is usually formulated with its own unique configuration, turbulence closure parameterization, and primitive equation simplifications.

One thing to note about the model validation procedure is that positions of moored buoys can change over the course of a few days to a week depending on the current regime by up to ≈3 km. This is a diameter within which the buoy moves. Since each mooring moves in time and space from its deployment position, we calculated average position based on the historical latitude and longitude data for each buoy. Thus, NLOM SST was extracted using the average latitude and longitude values for each buoy. For ease of notation, nearest integer values of average latitude and longitude is used for each buoy in the tables shown here and throughout the text. Although there are other TAO and NDBC buoys that measure SST in addition to those used in this paper, the buoys used here have at least 5 yearlong SST time series during 1980–1998. NDBC buoys located in water shallower than 200 m were excluded because the 200-m isobath is generally the NLOM boundary.

Table H.1
Statistical comparisons of daily SST between NLOM and NDBC buoys

Location	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
(17°N, 153°W)	.	.	.	.	.	0.57	.	.	.	0.45	0.43	.	0.45	0.49	0.74	0.73	.	.	.
(17°N, 158°W)	.	.	.	.	.	0.54	.	0.33	.	.	.	.	.	.	0.84	0.61	0.62	0.60	0.42
(19°N, 161°W)	.	.	.	.	.	0.43	0.71	0.47	.	.	0.28	.	0.48	0.68	0.35	0.84	0.95	0.62	.
(23°N, 162°W)	.	.	.	.	0.60	0.56	.	1.18	.	.	0.79	.	0.79	0.87	.	0.58	.	1.07	1.45
(26°N, 086°W)	.	1.76	1.85	1.65	.	1.47	0.78	.	.	1.18	1.27	1.41	.	1.86	.	1.99	2.03	.	1.45
(26°N, 090°W)	.	0.83	1.03	1.59	1.44	.	1.28	0.57	.	.	0.60	.	.	0.61	0.58	1.09	.	0.72	.
(26°N, 094°W)	.	0.69	.	.	0.68	.	.	1.01	0.65	0.58	0.65	0.64	.	0.67	0.94	0.67	0.58	.	0.92
(29°N, 079°W)	.	.	.	.	.	.	.	.	.	0.85	0.83	0.86	0.75	0.82	0.80	0.79	0.83	0.74	0.82
(32°N, 075°W)	.	.	1.11	1.18	0.80	.	.	.	1.03	.	.	0.99	.	1.36	0.99	1.38	.	1.13	.
(35°N, 073°W)	1.18	.	.	1.33	.	.	.	.	.	0.84	1.32	.	1.31	.	.	.	.	1.30	1.06
(36°N, 122°W)	.	.	.	.	0.97	.	1.34	1.24	1.19	.	.	.	.	.	.	1.67	1.45	.	.
(38°N, 071°W)	.	.	.	1.13	.	.	1.86	2.47	2.70	.	3.28	.	1.93	1.81	.	1.50	2.03	2.05	2.64
(38°N, 130°W)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0.81	0.97	0.97	0.63
(41°N, 137°W)	.	.	0.90	0.96	.	.	.	0.70	.	0.68	0.64	.	.	0.91	0.72	0.68	0.67	.	1.08
(43°N, 130°W)	1.10	0.87	0.85	0.77	.	0.65	1.12	.	.	0.78	.	.	1.08	0.88	0.74	0.73	0.87	0.62	.
(46°N, 131°W)	1.09	1.17	1.36	0.66	0.72	.	0.76	.	0.92	0.88	.	0.86	0.95	.	0.91	0.68	0.83	0.57	1.03
(51°N, 136°W)	1.14	0.83	1.42	0.72	0.64	.	.	0.85	.	.	.	.	.	.	.	.	.	.	.
(52°N, 156°W)	.	.	0.73	0.39	.	0.84	0.84	1.24	1.51	.	0.90	0.92	0.76	.	0.46	.	0.78	0.57	.
(56°N, 148°W)	0.80	0.73	0.84	0.54	.	0.66	0.60	0.51	0.44	0.87	.	0.86	0.48	0.79	.	0.57	0.78	0.53	0.48
(57°N, 178°W)	.	.	.	.	.	.	0.80	.	0.84	0.86	.	.	0.44	0.80	0.66	0.81	.	0.49	0.75

Comparisons are made for each year separately, using root-mean-square difference (RMSD). Note that there is no assimilation of SST in the NLOM simulations.

Table H.2
Statistical comparisons of daily SST between NLOM and TAO buoys

Location	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
(0°N, 110°W)	0.96	0.99	1.52	1.61	1.35	1.03	1.20	1.18	.	1.63	.	1.31	.
(0°N, 125°W)	.	.	2.40	1.62	.	.	0.84	0.99	0.80	1.19	.	1.21	2.49
(0°N, 140°W)	0.73	.	2.13	1.59	1.07	.	1.01	1.11	0.74	.	.	0.92	2.71
(0°N, 155°W)	.	.	.	.	.	.	0.71	0.76	0.61	0.65	0.71	0.77	1.71
(0°N, 170°W)	.	.	.	.	.	.	.	0.58	1.07	0.64	0.83	0.62	.
(2°N, 110°W)	1.19	.	1.51	.	1.10	.	1.30	1.10	1.37	.	.	1.42	.
(2°N, 165°E)	.	.	.	0.52	.	.	0.65	0.57	0.92	0.87	1.02	0.66	.
(5°N, 110°W)	.	.	.	0.62	0.40	0.70	0.58	0.87	0.92	.	0.66	.	.

(continued on next page)

Table H.2 (*continued*)

Location	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
(5°N, 155°W)	.	.	.	.	.	.	0.58	0.78	0.96	0.66	0.59	0.55	0.69
(5°N, 165°E)	.	.	.	0.40	0.32	0.32	0.59	0.47	0.67	0.88	1.12	0.76	1.06
(8°N, 110°W)	.	.	.	.	.	.	0.43	0.70	0.99	0.67	0.64	.	.
(8°N, 170°W)	.	.	.	.	.	.	.	0.97	1.01	0.55	0.71	0.73	0.99
(2°S, 110°W)	.	.	1.76	1.46	0.78	0.88	0.89	0.78	0.77	0.94	1.17	.	2.15
(2°S, 125°W)	.	.	.	.	.	.	0.62	0.79	0.95	0.82	0.85	1.09	1.93
(2°S, 165°E)	.	.	.	.	.	.	0.65	0.67	1.02	0.80	1.08	0.64	.
(5°S, 110°W)	.	.	.	0.82	.	0.86	0.83	0.63	0.56	.	.	0.78	.
(5°S, 140°W)	.	.	.	.	.	0.59	0.57	0.68	0.87	0.68	0.58	0.73	1.25
(5°S, 155°W)	.	.	.	.	.	.	0.66	0.74	0.94	0.47	0.58	0.62	0.77
(8°S, 110°W)	.	.	.	.	.	.	.	0.61	0.53	0.42	0.43	0.71	1.10
(8°S, 155°W)	.	.	.	.	.	.	.	0.65	0.63	0.30	0.40	0.68	0.92

Comparisons are made for each year separately, using root-mean-square difference (RMSD). Note that there is no assimilation of SST in the NLOM simulations.

References

Carpenter Jr., R.L., Drogemeier, K.K., Woodward, P.R., Hane, C.E., 1990. Application of the piecewise parabolic method to meteorological modeling. Mon. Weather Rev. 118, 586–612.

Chassignet, E.P., Smith, L.T., Bleck, R., Bryan, F.O., 1996. A model comparison: numerical simulations of the north and equatorial Atlantic Ocean circulation in depth and isopycnic coordinates. J. Phys. Oceanogr. 26, 1849–1867.

da Silva, A.M., Young, C.C., Levitus, S., 1994. Atlas of Surface Marine Data: Volume 1. Algorithms and Procedures. NOAA Atlas NESDIS, vol. 6. U.S. Dept. of Commerce, Washington, D. C. 83 pp.

Enfield, D.B., 2001. Evolution and historical perspective of the 1997–98 El Niño–Southern Oscillation event. Bull. Mar. Sci. 69, 7–25.

ECMWF, 1995. User Guide to ECMWF Products. Meteorol. Bull. M3.2. 71 pp. Available from ECMWF, Shinfield Park, Reading RG2 9AX, UK.

Fox, D.N., Teague, W.J., Barron, C.N., Carnes, M.R., Lee, C.M., 2002. The Modular Ocean Data Analysis System (MODAS). J. Atmos. Ocean. Technol. 19, 240–252.

Freitag, H.P., Feng, Y., Mangum, L.J., McPhaden, M.J., Neander, J., Stratton, L.D., 1994. Calibration, Procedures and Instrumental Accuracy Estimates of TAO Temperature, Relative Humidity and Radiation Measurements. NOAA Tech. Memo. ERL PMEL-104. 32 pp. Available from PMEL, 7600 Sand Point Way, Seattle, WA 98115, USA.

Gibson, J.K., Kållberg, P., Uppala, S., Hernandez, A., Nomura, A., Serrano, E., 1999. ECMWF Re-Analysis Project Report Series: 1. ERA description (Version 2). 74 pp. Available from ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Halliwell Jr., G.R., 2004. Evaluation of vertical coordinate and vertical mixing algorithms in the Hybrid Coordinate Ocean Model (HYCOM). Ocean Model. 7, 285–322.

Harrison, D.E., Vecchi, G.A., 2001. El Niño and La Niña equatorial Pacific thermocline depth and sea surface temperature anomalies. 1986–1998. Geophys. Res. Lett. 28, 1051–1054.

Hellerman, S., Rosenstein, M., 1983. Normal monthly wind stress over the world ocean with error estimates. J. Phys. Oceanogr. 13, 1093–1104.

Hogan, P.J., Hurlburt, H.E., 2000. Impact of upper ocean–topographic coupling and isopycnal outcropping in Japan/East Sea models with 1/8° to 1/64° resolution. J. Phys. Oceanogr. 30, 2535–2561.

Hogan, P.J., Hurlburt, H.E., 2005. Sensitivity of simulated circulation dynamics to the choice of surface wind forcing in the Japan/East Sea. Deep Sea Res 52, 1464–1489.

Hu, D., Chao, Y., 1999. A global isopycnal OGCM. Validations using observed upper-ocean variabilities during 1992–1993. Mon. Weather Rev. 127, 706–725.

Hurlburt, H.E., Hogan, P.J., 2000. Impact of 1/8° to 1/64° resolution on Gulf Stream model–data comparisons in basin-scale subtropical Atlantic Ocean models. Dyn. Atmos. Ocean. 32, 283–329.

Hurlburt, H.E., Metzger, E.J., 1998. Bifurcation of the Kuroshio Extension at the Shatsky Rise. J. Geophys. Res. 103, 7549–7566.

Hurlburt, H.E., Thompson, J.D., 1980. A numerical study of Loop Current intrusions and eddy shedding. J. Phys. Oceanogr. 10, 1611–1651.

Hurlburt, H.E., Wallcraft, A.J., Schmitz Jr., W.J., Hogan, P.J., Metzger, E.J., 1996. Dynamics of the Kuroshio/Oyashio current system using eddy-resolving models of the North Pacific Ocean. J. Geophys. Res. 101, 941–976.

Hurlburt, H.E., Bell, M.J., Evensen, G., Barron, C.N., Hines, A., Smedstad, O.M., Storkey, D., 2002. Operational global ocean prediction systems. Proceedings of the "En route to GODAE" International Symposium, 13–15 June 2002, Biarritz, France, pp. 97–105. Available at http://www.ocean.nrlssc.navy.mil/global nlom/global_nlom/publist.html.

Jacobs, G.A., Hurlburt, H.E., Kindle, J.C., Metzger, E.J., Mitchell, J. L., Teague, W.J., Wallcraft, A.J., 1994. Decade-scale trans-Pacific propagation and warming effects of an El Niño anomaly. Nature 370, 360–363.

Kantha, L.H., Clayson, C.A., 1994. An improved mixed layer model for geophysical applications. J. Geophys. Res. 99, 25235–25266.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2000a. Efficient and accurate bulk parameterizations of air–sea fluxes for use in general circulation models. J. Atmos. Ocean. Technol. 17, 1421–1438.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2000b. An optimal definition for ocean mixed layer depth. J. Geophys. Res. 105, 16803–16821.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2000c. Mixed layer depth variability and barrier layer formation over the North Pacific Ocean. J. Geophys. Res. 105, 16783–16801.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2002a. Air–sea flux estimates and the 1997–1998 ENSO event. Boundary - Layer Meteorol. 103, 439–458.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2002b. Naval Research Laboratory Mixed Layer Depth (NMLD) Climatologies. NRL Report No. NRL/FR/7330/02/9995. 26 pp. Available from NRL, Code 7323, Bldg 1009, Stennis Space Center, MS 39529–5004, USA.

Kara, A.B., Wallcraft, A.J., Hurlburt, H.E., 2003a. Climatological SST and MLD simulations from NLOM with an embedded mixed layer. J. Atmos. Ocean. Technol. 20, 1616–1632.

Kara, A.B., Rochford, P.A., Hurlburt, H.E., 2003b. Mixed layer depth variability over the global ocean. J. Geophys. Res. 108, 3079. doi:10.1029/2000JC000736.

Kara, A.B., Hurlburt, H.E., Rochford, P.A., O'Brien, J.J., 2004. The impact of water turbidity on the inter-annual sea surface temperature simulations in a layered global ocean model. J. Phys. Oceanogr. 34, 345–359.

Kara, A.B., Hurlburt, H.E., Wallcraft, A.J., Bourassa, M.A., 2005. Black sea mixed layer sensitivity to various wind and thermal forcing products on climatological time scales. J. Climate 18, 5266–5293.

Kraus, E.B., Turner, J.S., 1967. A one-dimensional model of the seasonal thermocline: II. The general theory and its consequences. Tellus 19, 98–106.

Large, W.G., McWilliams, J.C., Doney, S.C., 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. 32, 363–403.

Le Traon, P.Y., Nadal, F., Ducet, N., 1998. An improved mapping method of multisatellite altimeter data. J. Atmos. Ocean. Technol. 15, 522–533.

Lin, S.-J., Chao, W.C., Sud, Y.C., Walker, G.K., 1994. A class of the van Leer-type transport schemes and its application to the moisture transport in a general circulation model. Mon. Weather Rev. 122, 1575–1593.

Martin, P.J., 1985. Simulation of the mixed layer at OWS November and Papa with several models. J. Geophys. Res. 90, 903–916.

McClain, C.R., Cleave, M.L., Feldman, G.C., Gregg, W.W., Hooker, S.B., Kuring, N., 1998. Science quality SeaWiFS data for global biosphere research. Sea Technol. 39, 10–16.

McPhaden, M.J., 1995. The Tropical Atmosphere Ocean (TAO) array is completed. Bull. Am. Meteor. Soc. 76, 739–741.

McPhaden, M.J., 1999. Genesis and evolution of the 1997–98 El Niño. Science 283, 950–954.

McPhaden, M.J., Yu, X., 1999. Equatorial waves and the 1997–98 El Niño. Geophys. Res. Lett. 26, 2961–2964.

Metzger, E.J., 2003. Upper ocean sensitivity to wind forcing in the South China Sea. J. Oceanogr. 59, 783–798.

Metzger, E.J., Hurlburt, H.E., 2001. The nondeterministic nature of Kuroshio penetration and eddy shedding in the South China Sea. J. Phys. Oceanogr. 31, 1712–1732.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Mon. Weather Rev. 116, 2417–2424.

Murphy, A.H., 1995. The coefficients of correlation and determination as measures of performance in forecast verification. Weather Forecast. 10, 681–688.

Murphy, A.H., Epstein, E.S., 1989. Skill scores and correlation coefficients in model verification. Mon. Weather Rev. 117, 572–581.

Murphy, A.H., Winkler, R.L., 1987. A general framework for forecast verification. Mon. Weather Rev. 115, 1330–1338.

Price, J.F., Weller, R.A., Pinkel, R., 1986. Diurnal cycling: observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing. J. Geophys. Res. 91, 8411–8427.

Rienecker, M.M., Atlas, R., Schubert, S.D., Willett, C.S., 1996. A comparison of surface wind products over the North Pacific Ocean. J. Geophys. Res. 101, 1011–1023.

Rochford, P.A., Kindle, J.C., Gallacher, P.C., Weller, R.A., 2000. Sensitivity of Arabian Sea mixed layer to 1994–1995 operational wind products. J. Geophys. Res. 105, 14141–14162.

Rosmond, T.E., João, T., Peng, M., Hogan, T.F., Pauley, R., 2002. Navy Operational Global Atmospheric Prediction System (NOGAPS): forcing for ocean models. Oceanography 15, 99–108.

Samuel, S.L., Haines, K., Josey, S.A., Myers, P.G., 1999. Response of the Mediterranean Sea thermohaline circulation to observed changes in the winter wind stress field in the period 1980–93. J. Geophys. Res. 104, 7771–7784.

Schopf, P.S., Loughe, A., 1995. A reduced-gravity isopycnal ocean model: hindcasts of El Niño. Mon. Weather Rev. 123, 2839–2863.

Shriver, J.F., Hurlburt, H.E., Smedstad, O.M., Wallcraft, A.J., Rhodes, R.C., in press. 1/32° real-time global ocean prediction and value-added over 1/16° resolution. J. Mar. Sys.

Smedstad, O.M., Hurlburt, H.E., Metzger, E.J., Rhodes, R.C., Shriver, J.F., Wallcraft, A.J., Kara, A.B., 2003. An operational eddy-resolving 1/16° global ocean nowcast/forecast system. J. Mar. Syst. 40–41, 341–361.

Stewart, T.R., 1990. A decomposition of the correlation coefficient and its use in analyzing forecasting skill. Weather Forecast. 5, 661–666.

Trenberth, K.E., Caron, J.M., 2001. Estimates of meridional atmosphere and ocean heat transports. J. Climate 14, 3433–3443.

Vialard, J., Menkes, C., Boulanger, J.-P., Delecluse, P., Guilyardi, E., McPhaden, M.J., Madec, G., 2001. A model study of oceanic mechanisms affecting equatorial Pacific sea surface temperature during the 1997–98 El Niño. J. Phys. Oceanogr. 31, 1649–1675.

Wallcraft, A.J., Moore, D.R., 1997. The NRL Layered Ocean Model. Parallel Comput. 23, 2227–2242.

Wallcraft, A.J., Kara, A.B., Hurlburt, H.E., Rochford, P.A., 2003. The NRL Layered Global Ocean Model (NLOM) with an embedded mixed layer submodel: formulation and tuning. J. Atmos. Ocean. Technol. 20, 1601–1615.

Weller, R.A, . Baumgartner, M.F., Josey, S.A., Fischer, A.S., Kindle, J. C., 1998. Atmospheric forcing in the Arabian Sea during 1994–1995: observations and comparisons with climatology models. Deep-Sea Res. 45, 1961–1999.