Ocean Modelling 34 (2010) 166–184

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

Evaluations of threshold and curvature mixed layer depths by various mixing schemes in the Mediterranean Sea

A. Birol Kara [1], Robert W. Helber *, Alan J. Wallcraft [2]

Naval Research Laboratory, Ocean Dynamics and Prediction Branch, Stennis Space Center, MS 39529, USA

ARTICLE INFO

Article history:
Received 10 July 2009
Received in revised form 26 May 2010
Accepted 27 May 2010
Available online 4 June 2010

Keywords:
Mixed layer depth
Ocean mixing model
Mediterranean Sea

ABSTRACT

Predictive ability of five different embedded turbulent mixing models that range from second-order turbulent closure to bulk mixing parameterization is examined in the Mediterranean Sea. Each is embedded in the HYbrid Coordinate Ocean Model (HYCOM). Mixed layer depth (MLD), which is one of the most important upper ocean variables, is used to evaluate the treatment of turbulent processes in each model run. In addition to overall spatial and temporal variability, analyses of MLD are presented using an extensive set (3976) of temperature and salinity profiles from various data sources during 2003–2006. Results obtained from simulations (with no data assimilation and relaxation only to salinity) for the five mixing models are compared with observed MLDs obtained from in situ temperature and salinity profile observations. To ensure the robustness of the validation statistics MLD is computed using both curvature and threshold based methodologies. Results indicate that while all mixing schemes represent the MLD well, the bulk mixing models have substantial accuracy deficiencies relative to the higher order mixing models. The modeled MLDs are slightly deeper than observed MLDs with the mean bias error ~10 m for the higher order mixing models while the bulk mixing model bias error is 15 m or more. The RMS error for the higher order mixing models is ~40 m while it is ~50 m for the bulk mixing models. The bulk mixing models had substantially larger errors particularly for the curvature MLD definition.

Published by Elsevier Ltd.

1. Introduction

The Mediterranean Sea is a relatively large semi-enclosed basin whose bottom depth is quite variable (Fig. 1). Water mass formation in the region is primarily related to upper ocean mixing through mixed layer depth (MLD) (e.g., Astraldi et al., 2002). In particular, the Modified Atlantic Water (MAW) and the Levantine Intermediate Water (LIW) contribute to the entire surface layer of the western Mediterranean (Mertens and Schott, 1998) while convective mixing occurs in the Gulf of Lyons (Marshall and Schott, 1999), which creates very deep MLDs.

While the knowledge about the major features of upper ocean mixing and MLD is essential in the Mediterranean Sea, inter-annual variability of subsurface temperature and salinity in situ data is rarely available at fine spatial scales in the basin. Such insufficient information limits our ability in determining the seasonal variability of MLD. In this paper, an eddy-resolving ocean model including five different embedded turbulent mixing models is used to compensate for sparse data coverage, and temporal variability of mixing processes along with MLD is examined at fine spatial scales (e.g., 3–4 km resolution). The mixing models vary from simple bulk models (Kraus and Turner, 1967) to more computationally expensive turbulent closure models (Mellor and Yamada, 1982). Each model type has advantages and disadvantages, and choices are dictated by needs. When comparing mixing models for predicting MLD, a few important questions arise: (1) which mixing model is suitable for simulating MLD in the Mediterranean Sea? (2) Are there substantial differences among the models? (3) What are the errors in MLD associated with each model? The Mediterranean Sea provides a good test bed region to seek possible answers to these questions because many mid-latitude ocean processes are found in this basin. In addition, evaluation studies for different mixing models are very limited in this basin (Fernãndez et al., 2006). Compared to the global ocean, setting up an ocean model for the Mediterranean Sea is computationally less expensive, allowing for the evaluation of five different mixing models.

We perform multi-year simulations to investigate performances of a set of five mixing models in predicting MLD in the Mediterranean Sea. The evaluation of the performance of these models has not been done before in this region. Thus, the major purpose of this paper is to determine predictive ability of each model and find differences in MLD among them.

* Corresponding author. Tel.: +1 228 688 5430; fax: +1 228 688 4759.
E-mail addresses: robert.helber@nrlssc.navy.mil (R.W. Helber), alan.wallcraft@nrlssc.navy.mil (A.J. Wallcraft).
[1] Deceased author.
[2] Tel.: +1 228 688 4813; fax: +1 228 688 4759.

1463-5003/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.ocemod.2010.05.006

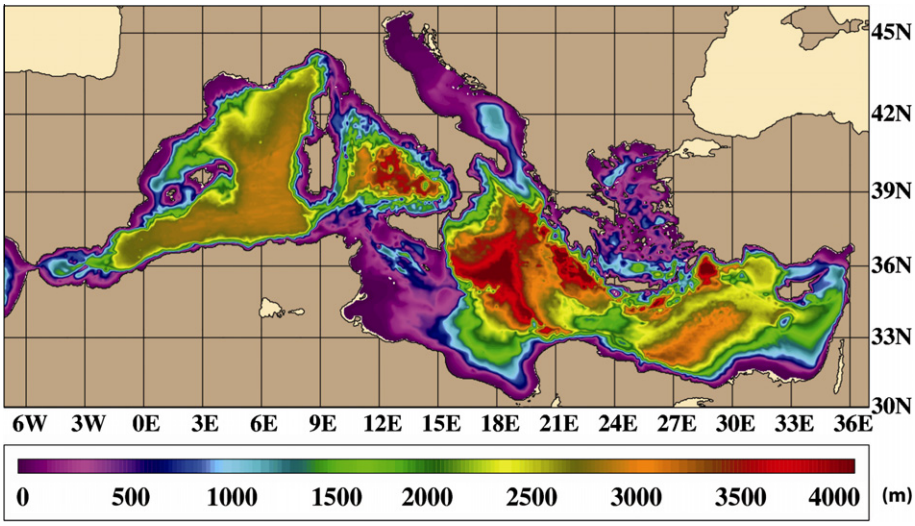

Fig. 1. Bottom depth (m) in the Mediterranean Sea as constructed from the 1 min resolution General Bathymetric Chart of the Oceans (GEBCO) which is available online at http://www.gebco.net/. There are large regions where the depths are shallow (e.g., <500 m), such as most of the Adriatic and Aegean Seas. The land-sea boundary (in light tan) is set at 5 m, i.e., water depths <5 m are excluded and considered as land for the ocean model simulations. We use this domain and drop the latitude and longitude labels from similar figures for simplicity.

2. Ocean model description

The HYbrid Coordinate Ocean Model (HYCOM) used in this paper is a community ocean model (http://oceanmodeling.rsmas.miami.edu/hycom/). It is a generalized hybrid isopycnal (σ) and terrain-following (z-level) coordinate primitive equation model with the original design features described in Bleck (2002). HYCOM uses the layered continuity equation to make a dynamically smooth transition to z-levels (fixed-depth coordinates) in the unstratified surface mixed layer and σ-levels (terrain-following coordinates) in shallow water. The optimal coordinate is chosen every time step using a hybrid coordinate generator.

2.1. Mediterranean sea model

The Mediterranean Sea HYCOM was configured on a Mercator grid with a resolution of 1/25° cos (lat) × 1/25° (latitude × longitude). Thus, the model has a resolution of 3.8 km at the southern regions (at approximately 32°N) and 3.3 km at the northern latitudes (at approximately 42°N). Zonal and meridional array sizes are 1235 and 549, respectively.

The model has 20 hybrid layers. The target density values (in σ_t) for layers 1–20 are 19.50, 21.00, 22.50, 24.00, 25.50, 26.50, 27.25, 27.75, 28.15, 28.40, 28.60, 28.75, 28.90, 29.00, 29.05, 29.08, 29.11, 29.16, 29.18 and 29.22. The density difference values were chosen so that the layers tend to become thicker with increasing depth, with the lowest abyssal layer being the thickest. The top four layers are in z-level coordinates at all times. The minimum thickness of layer 1 is 3 m. The model is initialized based on the 1/8° climatological monthly mean temperature and salinity fields from the Generalized Digital Environmental Model (GDEM) climatology (NAVOCEANO, 2003) which has 78 levels in the vertical.

2.2. Mixing submodels

There are five turbulent mixing submodels embedded within HYCOM. The name, abbreviation, and reference for each mixing model are given in Table 1.

2.2.1. K-profile parameterization

KPP is a 1st order turbulence closure model that is intermediate in computational complexity between bulk mixing models and 2nd-order turbulence closures. It is currently the standard mixing model for HYCOM because it is relatively insensitive to low vertical resolution, and the hybrid coordinate approach tends to require fewer layers/levels than fixed vertical coordinate approaches. The KPP scheme provides mixing from surface to bottom, matching the large surface boundary layer diffusivity/viscosity profiles to weak diapycnal diffusivity/viscosity profiles in the interior ocean.

2.2.2. Goddard Institute for Space Studies

GISS is a level-2 Reynolds-stress model, in which diffusivity profiles for viscosity, temperature and salinity are parameterized as functions of the Brunt–Väisälä frequency, the gradient Richardson number and the turbulent kinetic energy dissipation rate. The model formulation is valid within the mixed layer and below and includes salt fingering and diffusive double diffusion. Unlike KPP, nonlocal effects are not taken into account. The model equations are solved within two different regimes, depending on whether resolved or unresolved shear is the dominant influence on the stability. The former regime represents the intense mixing of the surface boundary layer, while the latter represents the comparatively quiescent ocean interior.

2.2.3. Mellor–Yamada level-2.5

MY is a level-2.5 Reynolds-stress model that solves the equations for turbulent kinetic energy (TKE) and TKE times the turbulence length scale to estimate the viscosity and diffusivity coefficient profiles. The MY model is the only vertical mixing algorithm in HYCOM that accounts for the advection and diffusion of

Table 1
Abbreviation and references for the mixing models used in the Mediterranean Sea HYCOM, which is forced with 3 hourly wind and thermal atmospheric variables from NOGAPS during 2003–2006. The model spin-up was first run using 3-hourly atmospheric forcing from NOGAPS for the given mixing model during 2001–2003.

Acronym	Mixing models in HYCOM	Reference
KPP	K-Profile Parameterization	Large et al. (1997)
GISS	Goddard Institute for Space Studies	Canuto et al. (2002)
MY	Mellor–Yamada level-2.5	Mellor and Yamada (1982)
KT	Kraus and Turner	Kraus and Turner (1967)
PWP	Price–Weller–Pinkel	Price et al. (1986)

turbulence. It is almost 1.5 times more computationally expensive than the other mixing models.

2.2.4. Kraus and Turner

KT assumes that all properties are homogeneous within the mixed layer. The mixed layer properties are changed through surface fluxes (wind-stress and buoyancy loss/gain) and also through entrainment from below. The mixed layer deepens in response to wind mixing and buoyancy loss at a rate determined by energetics. A balance between production and dissipation of energy integrated over the whole mixed layer is assumed.

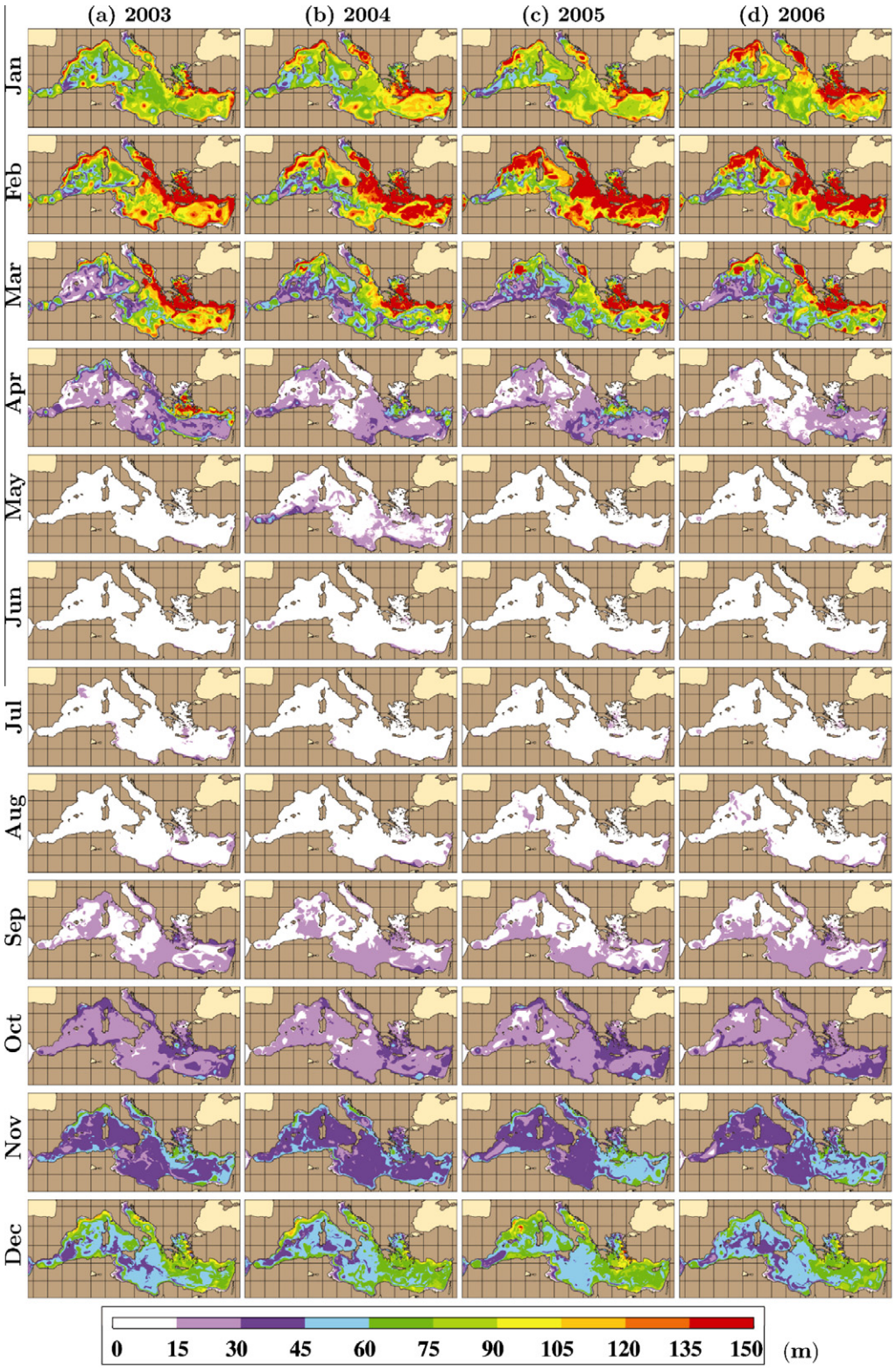

Fig. 2. Monthly mean threshold MLD fields in the Mediterranean Sea from 2003 to 2006 from the MY mixing model simulation.

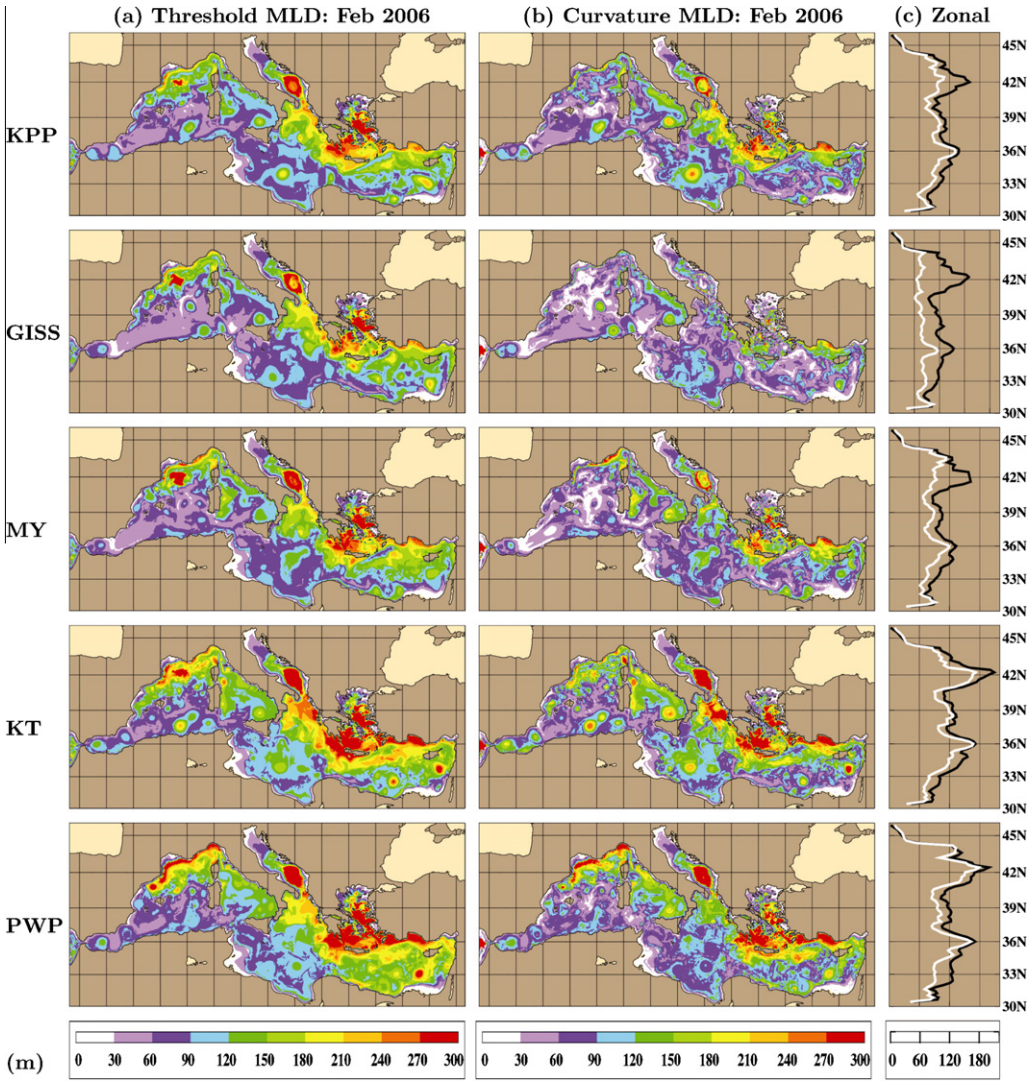

Fig. 3. Spatial variations of mean MLD predicted by five mixing models in February 2006. MLD is computed based on two methodologies as described in the text: the (a) threshold and (b) curvature MLD definition. Also included in (c) are the zonal averages of MLD (the threshold MLD in black and the curvature MLD in white).

2.2.5. Price–Weller–Pinkel

PWP uses the Richardson-number criteria to determine the depth of the mixed layer instead of calculating it prognostically from the energy balance. Vertical mixing at each grid point is performed in three steps. First, static instability is relieved in the upper ocean mixed layer if it exists. Second, mixed layer entrainment is performed based on a bulk Richardson-number criterion. Third, shear-instability mixing between adjacent model layers is calculated based a gradient Richardson-number criterion. PWP provides for shear-instability mixing beneath the surface mixed layer, but it does not provide for background mixing due to other processes such as internal wave breaking.

2.3. Atmosphere forcing and model simulations

HYCOM reads in the following time-varying atmospheric fields: for the momentum equation forcing (zonal and meridional components of wind stress) and for the thermal forcing (air temperature, air mixing ratio, and wind speed at 10 m above the sea surface; precipitation, net shortwave radiation, and net longwave radiation at the sea surface). They are all obtained from 0.5°-resolution Navy Operational Global Atmospheric Prediction System (NOGAPS) archives at 3 hourly intervals (Hogan and Rosmond, 1991).

The creeping sea-fill methodology is applied to all the thermal forcing variables to reduce the land contamination in the atmospheric forcing variables near the coastal boundaries before using them for the model simulations (Kara et al., 2008). Latent and sensible heat fluxes are calculated using the model's top layer (3 m) temperature at each model time step with bulk formulations (Kara et al., 2005). Similarly, wind stresses are computed based on 10 m winds from NOGAPS using stability-dependent exchange coefficients.

Five HYCOM simulations are performed using each of the mixing models given in Table 1. The model configuration, atmospheric forcing, bottom topography, etc., for all the simulations are identical except for the mixing model used. Each simulation was then extended beyond the initial spin-up from 1st January 2003 to 31st December 2006. The model is a stand-alone ocean model with no data assimilation. There is only initialization (temperature and salinity) from climatology with relaxation to surface salinity.

3. MLD determination

We will evaluate performance of each mixing model during 2003–2006 using two methodologies for determining MLD. Both (1) threshold and (2) curvature MLD definitions are described

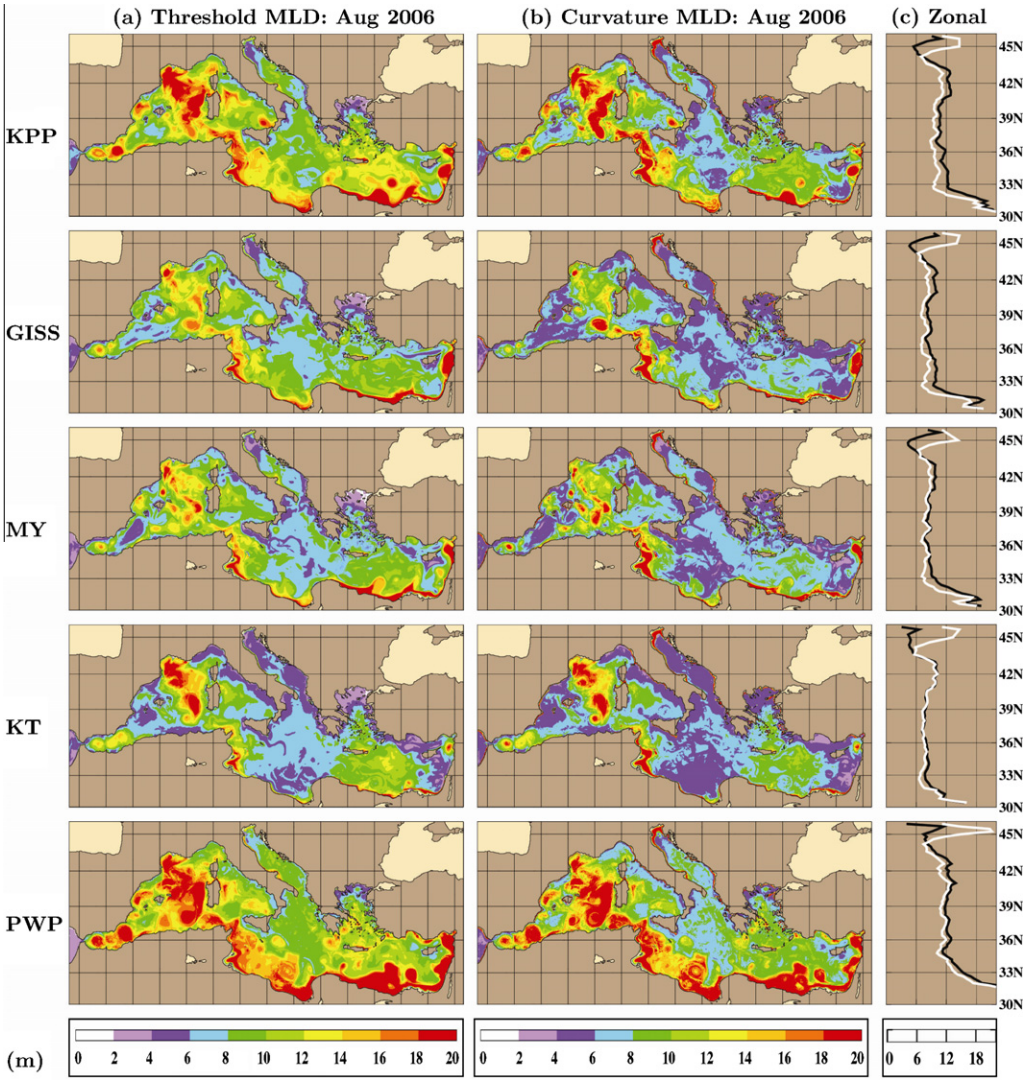

Fig. 4. The same as Fig. 3 but in August 2006. Note that the color palette range is different to better demonstrate spatial variability of summer MLD.

below and are used throughout the paper to determine the MLD of model output as well as observational data. By using two different definitions our goal is to confirm the validity of results and identify prediction characteristics differences highlighted by comparing results from the two MLD definitions. Throughout the paper, notations of T, S and σ_θ denote potential temperature, salinity and potential density, respectively. Density is expressed as $\sigma_\theta = \rho_\theta - 1000\ \mathrm{kg\ m^{-3}}$, where ρ_θ is the seawater potential density. The potential density is based on T and S values at given depths computed from potential temperature and the equation of state following (Fofonoff and Millard, 1983).

We follow the threshold definition of Kara et al. (2000). The algorithm can be obtained online at http://www7320.nrlssc.navy.mil/nmld/nmld.html. MLD is described as the base of an isopycnal layer, where density has changed based on a value of $\Delta\sigma_\theta$ from the surface density. The $\Delta\sigma_\theta$ value is variable in space and time computed by $\Delta\sigma_\theta = \sigma_\theta(T + \Delta T, S, P) - \sigma_\theta(T, S, P)$ where T and S are the surface values of temperature and salinity, P is zero, and $\Delta T = 0.8\ ^\circ\mathrm{C}$. Thus, the fixed ΔT gives rise to a variable $\Delta\sigma_\theta$ based on surface conditions. Note that in a threshold methodology, MLD can be sensitive to the threshold value (de Boyer Montégut et al., 2004). However, this methodology can easily be applied to density profiles which have either fine or coarse vertical resolutions (Ohno et al., 2004).

The curvature definition of Lorbacher et al. (2006) is also applied in detecting MLD. The algorithm is available at http://www.ifm-geomar.de/index.php?id=mixed-layer-depth. Unlike the threshold definition, the curvature method examines the given density profile and determines MLD based on the depth of maximum curvature. This definition determines the shallowest curvature peak of a density profile near the sea surface and represents the depth of the most recent turbulence penetration depth whereas the threshold MLD method tends to represent the seasonal MLD (e.g., Helber et al., 2008).

4. Evaluation of mixing models in determining MLD

Atmospherically-forced HYCOM simulation using each one of the mixing models (i.e., KPP, GISS, MY, KT and PWP) produces daily T and S at a grid resolution of approximately 3.5 km from the sea surface to the bottom of the ocean in the Mediterranean Sea. The result is spatially-varying daily T and S fields for each mixing model from January 2003 to December 2006. The model is sampled everywhere once a day at 00Z (midnight UTC). Since the thermal atmospheric forcing applied to the Mediterranean Sea HYCOM has a one day running mean, diurnal effects are minimized in the model and sub-daily sampling is not needed. Since both

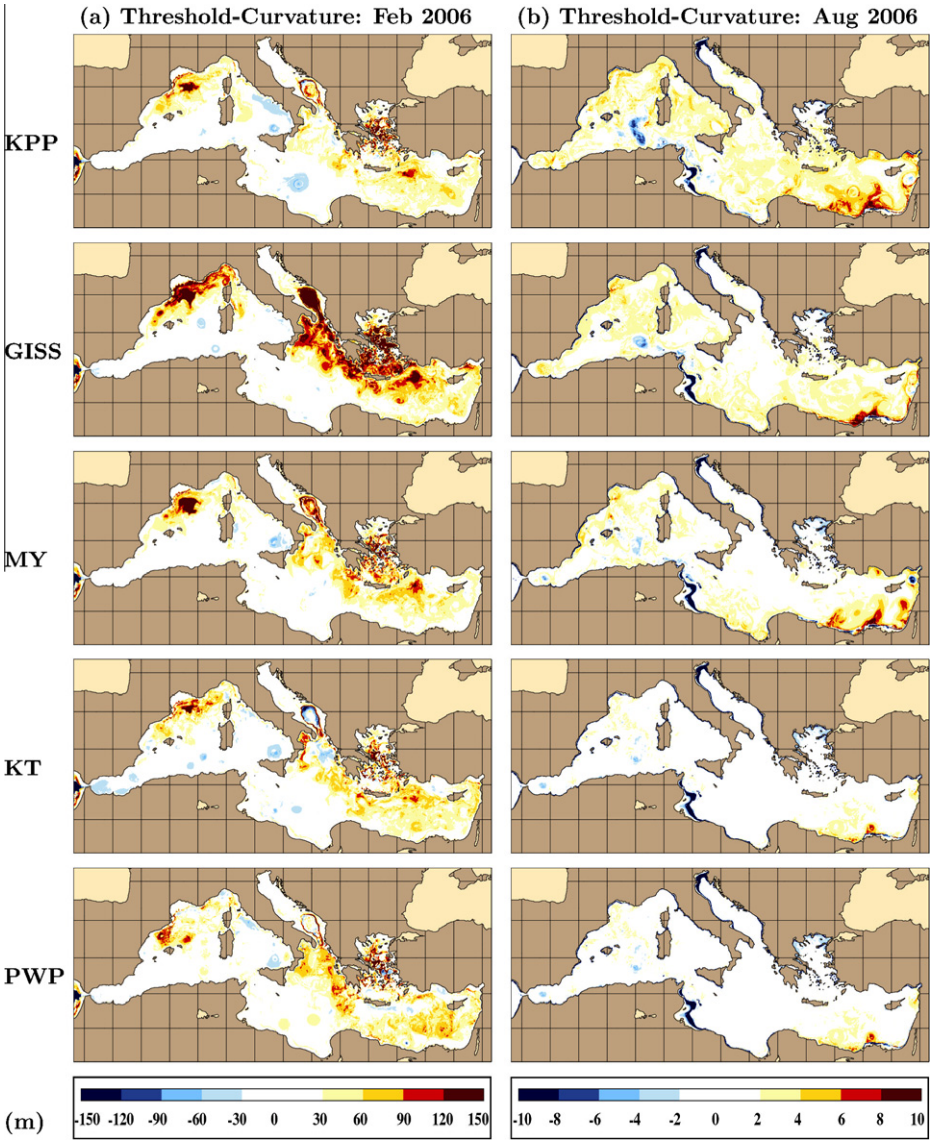

Fig. 5. Spatial variability of threshold minus curvature MLD differences for (a) February and (b) August 2006.

threshold- and curvature-based MLD algorithms are based on density profiles, the daily density fields are computed from T and S fields using the equation of state.

Monthly mean threshold MLD fields (computed using daily MY output) in the Mediterranean Sea from 2003 to 2006 are shown to have spatial and temporal variations (Fig. 2). The daily density fields from the model outputs of T and S were created first, and then the daily MLD is computed. We obtain monthly means from the daily fields for each month. We do not form monthly density fields from T and S and then compute the monthly MLD from the mean field because this would produce a less accurate MLD.

The resulting MLD fields from the MY simulation clearly reveal seasonal variability in the Mediterranean Sea during 2003–2006 (Fig. 2). We follow Boyer et al. (2006) in our definition of the seasons: January, February and March (winter); April, May and June (spring); July, August and September (summer); October, November and December (fall). Deep MLDs (>150 m) are generally evident in the eastern part of basin in January, February and March. The mixed layer is very shallow (<15 m) almost everywhere from May to August in all years. The shoaling of MLD is followed by the deepening period starting in September and MLD gradually increases each month through December over the entire basin.

A striking features of MLD in the Mediterranean Sea is the inter-annual variability during 2003–2006 (Figs. 2a–d). For example, MLDs in January reveal similar features for all years, although they are slightly deeper (by ≈15 m) in some parts of the eastern part of the region and the Adriatic Sea in 2004 and 2006. Although the 4-year time period considered here is not long enough to rule out strong inter-annual variability, monthly MLDs clearly reveal similar magnitudes for any given month from one year to another.

4.1. Spatial variations of mld by five mixing models

The MLD fields provided in Fig. 2 are from the MY mixing model using the threshold MLD definition. In addition to the three questions listed in the introduction, we also wish to answer an additional question (4): do results change when the simulations are evaluated with different MLD definitions (threshold versus threshold)? To answer these questions monthly MLDs from all simulations are examined for February 2006 when the MLDs tend to be deeper, which can better highlight differences (Fig. 3). Remember

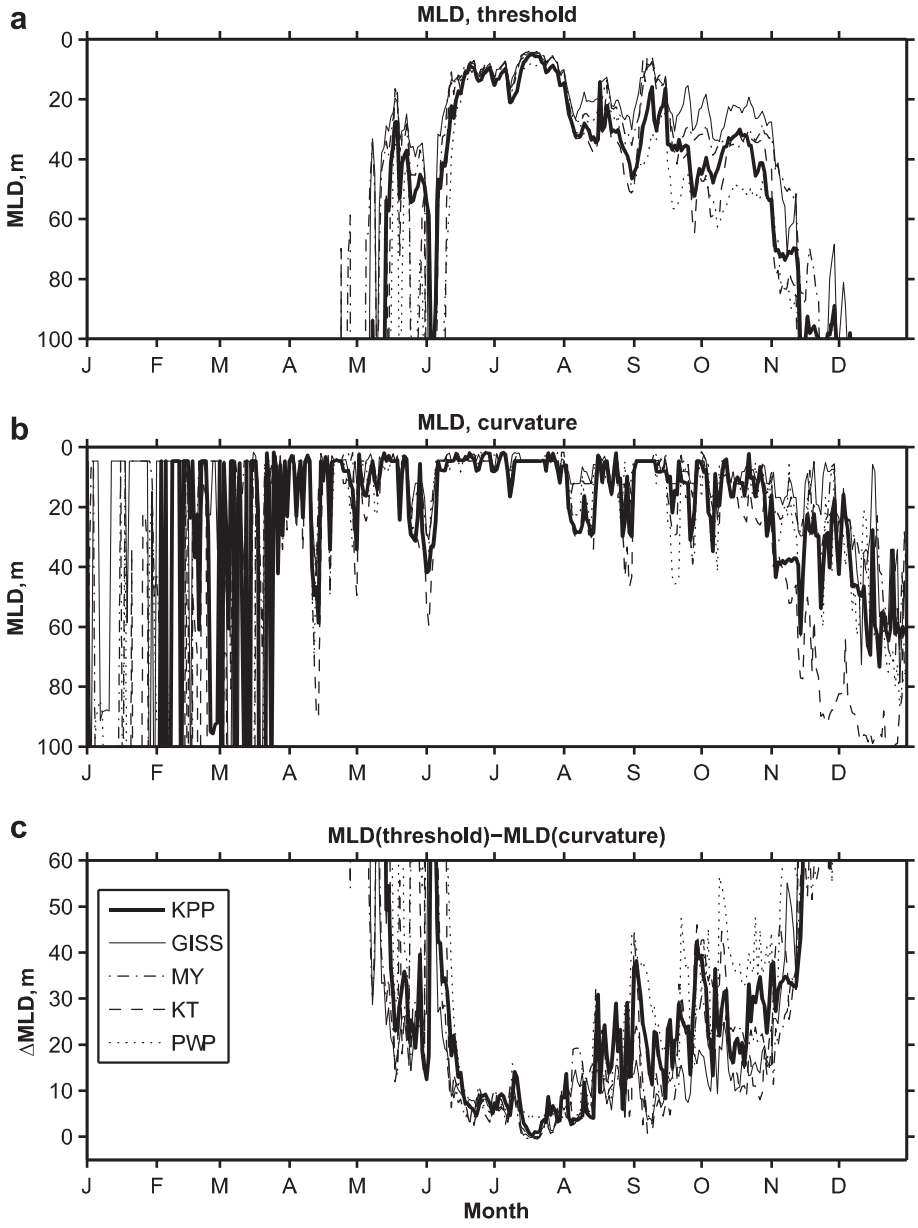

Fig. 6. Times-series of (a) threshold MLD, (b) curvature MLD, and (c) threshold minus curvature MLD difference from a point in the gulf of Lyons (42.09°N, 5°E). The line-styles listed in the legend are the results from the five mixing models.

that the atmospheric forcing for all model runs is identical as obtained from NOGAPS (see Section 2).

Monthly MLDs from KPP, GISS, MY, KT and PWP reveal similar values in the Mediterranean Sea when using either the threshold definition (Fig. 3a) or curvature definition (Fig. 3b). Deep MLDs (>300 m) are produced by all models in the northeastern part including the Aegean Sea and the South Adriatic Sea. MLDs computed from the curvature of the density profiles also demonstrate high values in these two regions for KT and PWP but not for KPP, GISS and MY (Fig. 3b). The MLDs based on the curvature definition tend to be shallower than those based on the threshold definition but such differences are more evident when MLD is relatively deep, e.g., in the eastern basin. In general, the curvature MLD algorithm detects the first change in density relative to a very uniform mixed layer, without any threshold of leeway below the well mixed layer. The threshold MLD method generally returns deeper values. This effect is amplified in regions of relatively small vertical density change such as during winter in deep convection regions such as the Gulf of Lyons. Relatively deep (shallow) MLDs exist in the eastern (western) region when using either one of the methodologies.

Differences in summer MLDs obtained from KPP, GISS, MY, KT and PWP are also examined in August 2006 (Fig. 4) when summer stratification results in shallower MLDs. Similar to February 2006, all mixing models generally yield similar MLD values over the basin when using either MLD methodologies. For the threshold MLD definition, the simulation with KT results in MLD which is typically deeper by approx 2–3 m basin-wide in comparison to other mixing models on August of 2006 (Fig. 4a). With the curvature MLD definition, similar features exist with minor differences (Fig. 4b). For example, the existence of very thin mixed layers of 2–4 m is more evident when the curvature-based MLD definition was applied.

Maps of the difference between the threshold and curvature MLD methods for each mixed layer model are shown in Fig. 5. Since in winter the MLD is generally deeper than in summer, and since the difference between threshold and curvature MLD methods are largest for deep MLDs, there are larger differences in Fig. 5

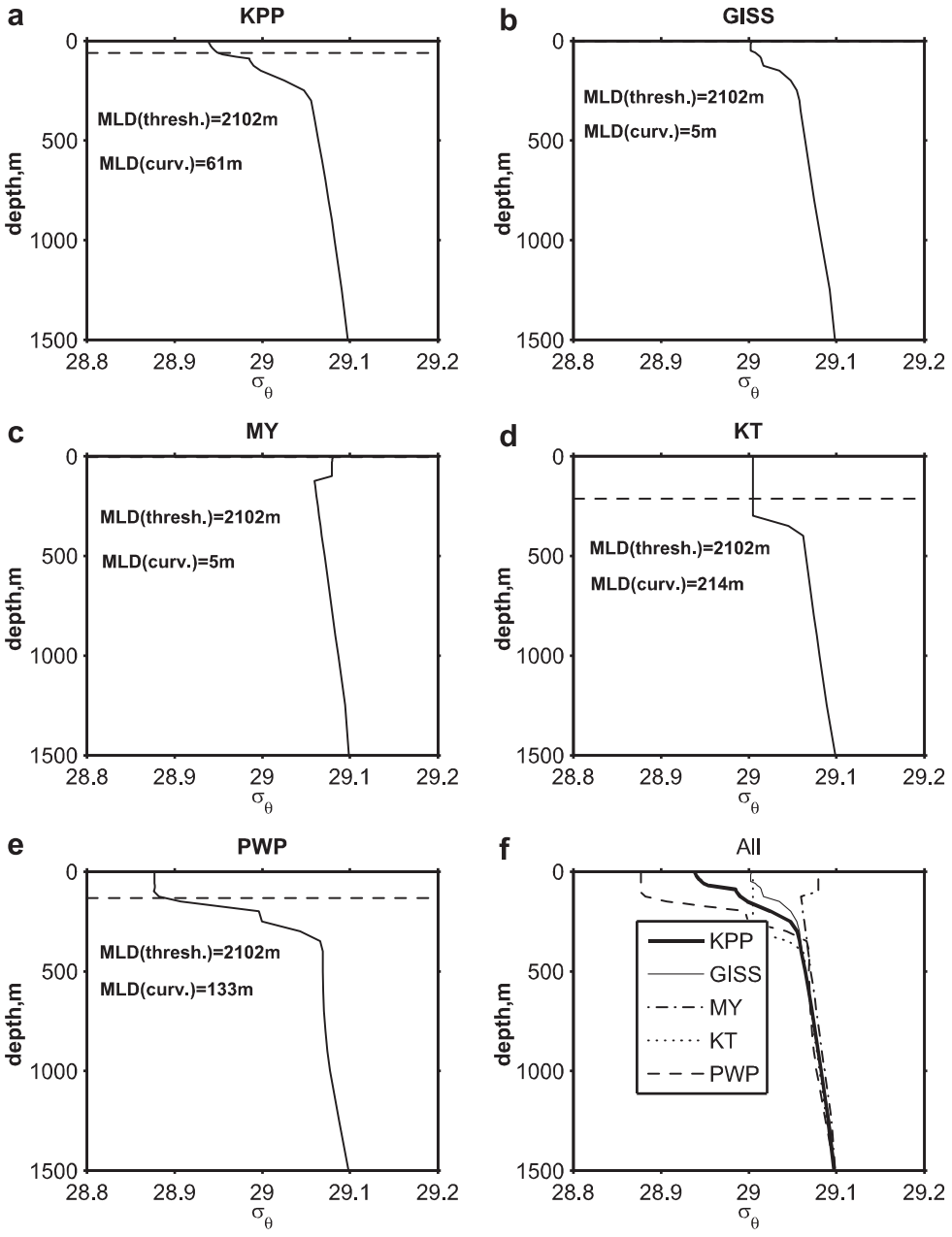

Fig. 7. An example winter profile from Gulf of Lyons (42.09°N, 5°E) during 20th February 2006 from the (a) KPP, (b) GISS, (c) MY, (d) KT, and (e) PWP mixing model runs. Plot (f) shows profiles from all mixing models for comparison. The dashed lines in (a)–(e) represent the curvature MLD while the threshold MLD is at the bottom of the profile as indicated inside each plot.

for February 2006. Over most of the Mediterranean, however, the differences are within 30 m. The outlier mixing model for February 2006 is the GISS model since it has the largest differences between the MLD methods. This suggests that the GISS mixing model generated a near surface curvature peak earlier in the season than the other mixing models in the red regions of Fig. 5a (for GISS). In the Gulf of Lyons, the Adriatic Sea and in the Rhodes Gyre region, the bulk mixing models KT and PWP tend to differ from the higher order schemes KPP, GISS, and MY. The MLD methods differ more for the higher order mixing schemes in those regions than do the bulk mixing models (Fig. 5a).

The nature of the threshold versus curvature differences can be understood by looking at the time series for 2006 from the Gulf of Lyons shown in Fig. 6. At this location (42.09°N, 5°E) the threshold MLD is deeper than 100 m until March whereas the curvature MLD varies substantially. The noise in the estimates of MLD in the winter and early spring is due to the nearly uniform density over the upper ocean at that location. When the ocean is very well mixed, all methods for determining MLD become noisy. For this reason, we focus on the upper 100 m in Fig. 6 to highlight the differences in MLD during spring, summer, and fall. The curvature method is more prone for having results with a large variance under the condition of a well mixed water column, since it must determine the MLD based on very small fluctuations in curvature. In contrast, during the summer when the upper ocean is well stratified and there are large gradients at the base of the MLD, both threshold and curvature methods agree. During summer the mixing models also have a greater agreement (Fig. 6). During the fall, when cooling occurs and the MLDs begin deepening, the mixing models begin to differ more substantially.

Fig. 7 shows an example profile from the Gulf of Lyons during the winter (20th February 2006) when the curvature versus the threshold MLD definitions differed drastically. The threshold MLD occurred at the bottom of the profile while the curvature method identified a MLD above 215 m for all mixing models. The threshold MLD definition returned very deep MLDs because the variance of the entire water column was below the threshold density value. Fig. 8 shows how the threshold and curvature MLD methods can differ on more typical profiles during the fall (27th October 2006) in the Gulf of Lyons. Note that the threshold MLD method identifies the depth of the seasonal MLD, while the curvature MLD method identifies the penetration depth of very recent mixing, for all mixing models except KT. The KT model is a bulk model and does not produce the finer scale structures found in the other models.

A quantitative analysis of MLD differences for each model is performed by taking the GISS mixing model run as the reference at each grid point and averaging over the Mediterranean Sea (Table 2). The reason for considering the GISS model as a reference is that it results in relatively small bias in comparison to MLDs from observed profiles (see Section 5). In general, MLDs from KPP and MY agree better with those from GISS with bias values of 7 m in February 2006 when using the threshold definition. The same is also true when applying the curvature definition but the mean biases are much larger with values of 25 m for KPP–GISS and 17 m for MY–GISS.

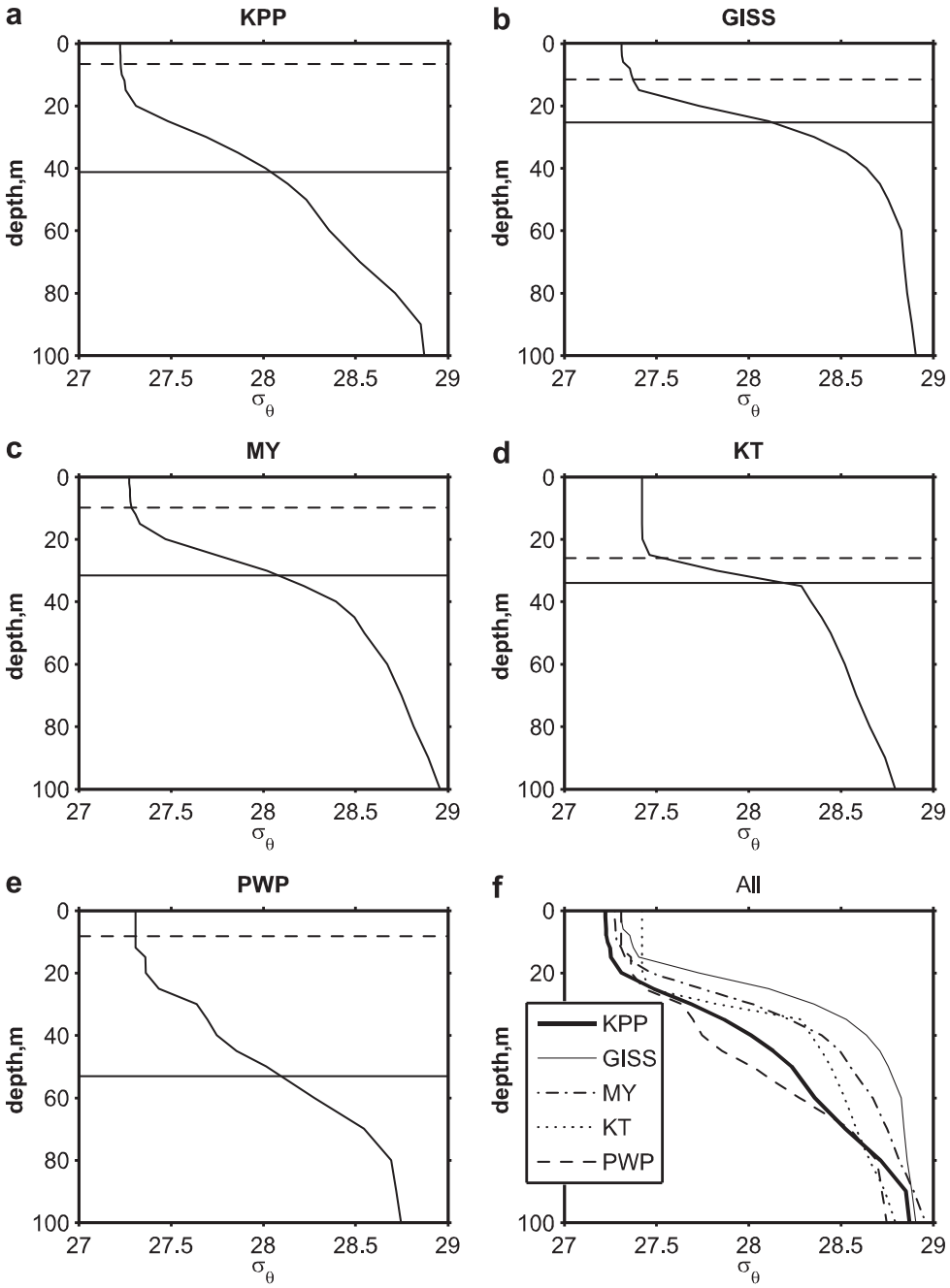

Fig. 8. An example summer profile from Gulf of Lyons (42.09°N, 5°E) during 27 October 2006 from the (a) KPP, (b) GISS, (c) MY, (d) KT, and (e) PWP mixing model runs. Plot (f) shows profiles from all mixing models for comparison. The solid (dashed) lines in (a)–(e) represent the threshold (curvature) MLD.

Table 2
Basin-averages of MLD differences and ratios and their standard deviations for mixing model HYCOM runs described in the text and Table 1. Differences in meters and ratio values are computed with respect to MLDs from the GISS simulation. Basin-averaged values of difference and ratio standard deviations are given in parentheses. Note that unlike February 2006, summer MLDs from each model reveal almost no biases in comparison to GISS.

	February 2006		August 2006	
	Threshold	Curvature	Threshold	Curvature
MLD difference (m)				
KPP–GISS	7 (39)	25 (44)	2 (3)	2 (3)
MY–GISS	7 (61)	17 (43)	0 (2)	0 (3)
KT–GISS	29 (50)	47 (64)	−1 (3)	0 (3)
PWP–GISS	27 (51)	41 (63)	4 (4)	4 (4)
MLD ratio				
KPP/GISS	1.1 (0.4)	1.6 (1.2)	1.3 (0.3)	1.3 (0.4)
MY/GISS	1.1 (0.4)	1.4 (1.0)	1.0 (0.2)	1.1 (0.3)
KT/GISS	1.3 (0.5)	2.0 (1.8)	1.1 (0.3)	1.0 (0.3)
PWP/GISS	1.3 (0.5)	1.9 (1.8)	1.5 (0.4)	1.6 (0.6)

During February 2006, when the MLDs are deeper, the KPP and MY have smaller average and standard deviation differences with GISS than do the bulk parameterization mixing models KT and PWP. The difference between the bulk mixing models (KT and PWP) and the higher order mixing models (KPP and MY) is larger for the curvature MLD definition. The average threshold difference goes from 7 m for KPP–GISS to 29 for KY–GISS a 120% increase while the average curvature difference goes from 25 m for KPP–GISS to 47 for KY–GISS a 200% increase. This is because the bulk mixing models do not produce the correct vertical gradients that the higher order mixing models have. The difference in profile shape, between the higher order mixing models and the bulk models, can be seen in Fig. 8f.

Because winter MLDs are much deeper than summer MLDs, for a fair comparison basin-averaged MLD ratios with respect to GISS are calculated in addition to mean biases (Table 2). MLD ratios are generally similar during February and August. The smallest ratios are found for KPP and MY in February, but MLDs from MY and KT are closest to those from GISS in August as evident from ratio values of unity. Thus, the answer to the question (1) in the introduction is that while all mixing models clearly show similar MLD patterns in the Mediterranean Sea, differences among the models can be large in some regions, with MY generally being very close to GISS. In addition, changing the definition in determining MLD can alter the predictive skill of a given mixing model especially in February 2006, which is an answer to the question (4) listed in this section.

4.2. Mixing model difference case studies

To demonstrate mixing model differences we consider two locations in the Mediterranean Sea. The first is in the Gulf of Lyons at 42.09°N and 5°E and the second is located in the Rhodes Gyre region at 36.1°N and 28.68°E. Both regions produce deep mixing convection and the mixing models have MLD differences as seen in Fig. 5. Fig. 9 shows a time series of total surface heat flux (Q_{tot}) in W m^{-2} with the convention that positive is a downward heat flux. Plotted below the heat flux is subsurface temperature difference (KPP–KT) over the upper 150 m. The general trend is that the subsurface temperature differences between KPP and KT model runs are small in the winter (Figs. 9 and 12). The exception is in the Rhodes Gyre during January 2003 (Fig. 12). In both the Gulf of Lyons and the Rhodes gyre regions, subsurface differences between the mixing models begins to increase after cooling events in the summer. The differences increase in vertical depth range, getting deeper through the fall as deep convection begins to occur. After the winter time deepening of the mixed layer, the mixing model results do not differ much until the spring warming occurs. At the beginning of September 2003 there was a strong cooling event that resulted in the KT run becoming cooler in the near surface and warmer below by mid September. In effect, the KPP cooled the upper ocean more slowly that the KT run (Fig. 9).

Over a five day period at the start of September 2003, Fig. 10 shows the evolution of each mixing model for a profile in the Gulf of Lyons. During this strong cooling event all models behaved similarly. During 2006 also in early September, warming conditions with the KT mixing model behaved substantially differently, resulting in a large cool anomaly below the mixed layer (Fig. 11).

In the Rhodes Gyre region, 36.1°N, 28.68°E, the seasonal differences between mixing models is different. In Fig. 12, we see that the mixed layer is cooler in the KPP mixing model than the KT model in the summer while below the MLD KPP is warmer extending into the fall. Fig. 13 shows in detail a warming event at the beginning of August where the KPP mixing model showed large shoaling of the mixed layer not found in the other mixing model results. For the rapid cooling event in mid October 2006 (Fig. 14), all models behaved similarly except for the PWP mixing model that cooled more strongly in the near surface and below the MLD.

5. MLD comparisons with observation profiles

HYCOM allows one to examine temporal variability of MLD at approximately 3.5 km resolution in the Mediterranean Sea. In these analyses, the performance of five mixing models in the eddy-resolving model was quantified when the model used the same atmospheric forcing for all simulations. As expected, systematic errors in determining the MLD may occur due to several reasons, including inaccuracies in the atmospheric forcing. This could limit fair evaluation of a given mixing model. In addition, no independent data set was used for validating MLD from each mixing model in Section 4. Therefore, we will further present a validation study to determine the predictive capability of KPP, GISS, MY, KT and PWP in determining MLD. This is accomplished using many individual T and S profile observations made in the Mediterranean Sea during 2003–2006.

5.1. Profile data and quality control

In situ T and S profiles were acquired from three data sources: (1) Argo float data (Gould et al., 2004), (2) the US Navy's Master Oceanographic Observation Data Set (MOODS) (Teague et al., 1990), and (3) the World Ocean Database 2005 (WOD05) (Boyer et al., 2006). A breakdown of the number of profiles by month and year is given in Table 3.

For the analysis, we combine 4 years of data from 2003 to 2006 and compute MLDs from each individual T and S profile. For this study the WOD05 data ended in January 2005 while the MOODS data continued through 2006. As a result, more than twice the T and S profiles are from MOODS than WOD05 (1025 for MOODS versus 480 for WOD05). In summary, there are a total of 3976 profiles from which MLD is computed and analyzed.

While the profile data we use in this paper are quality controlled as obtained from their original sources, errors still exist. In addition, since a major goal is to examine the performance of each mixing model in predicting MLD, there are vertical sampling requirements that not all profiles meet. For these reasons, additional procedures are performed to edit the T and S profiles that will be used for validation. Our procedures are designed to identify problems that may compromise the integrity of the comparisons.

All profiles must pass the following tests: the first depth level in a given profile must be less than 10 m. This is required for

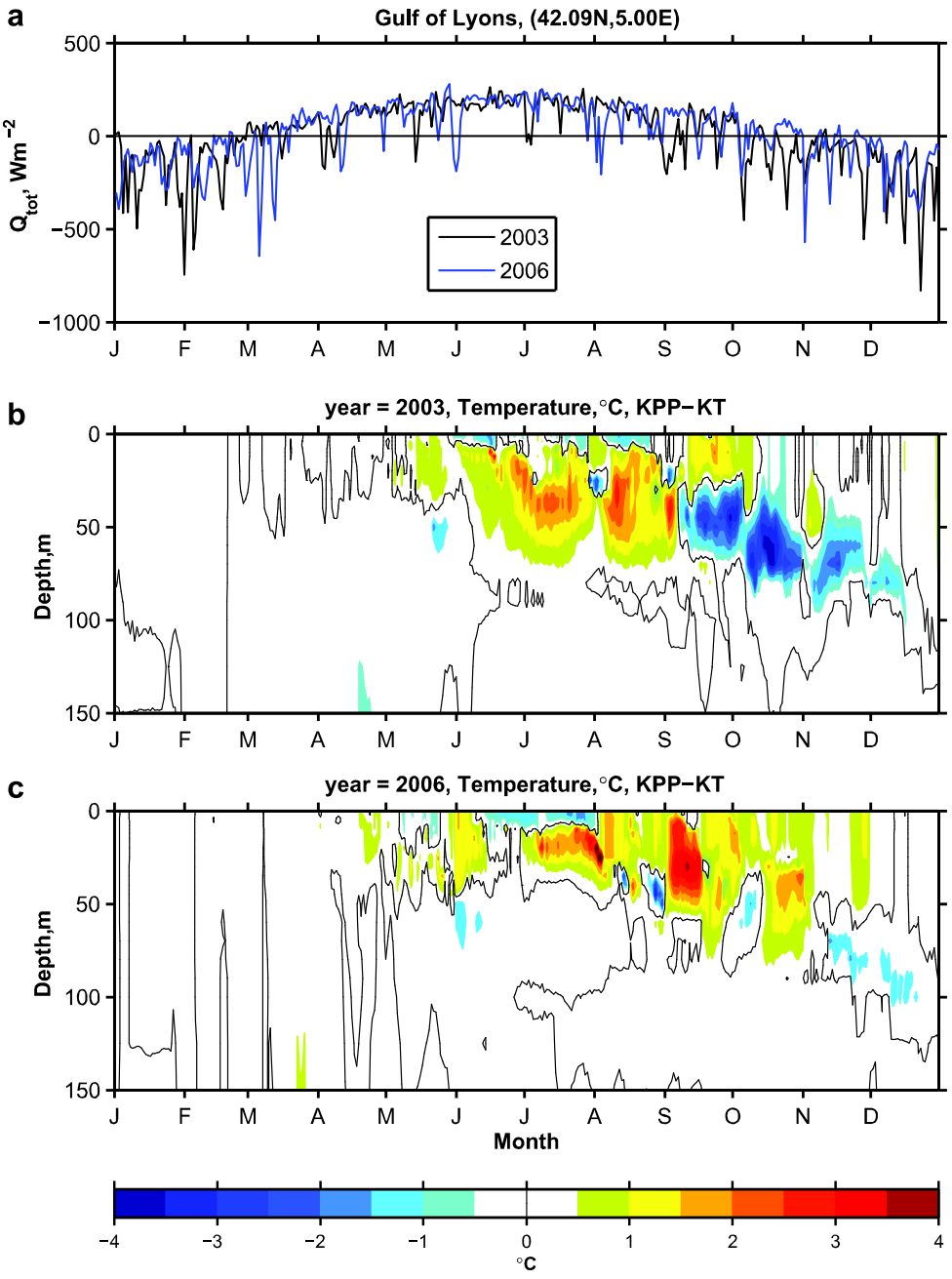

Fig. 9. One year time series for a point in the Gulf of Lyons located at 42.09°N, 5°E for 2003 and 2006 of (a) total surface flux (Q_{tot}) positive into the ocean in W m^{-2} and subsurface temperature differences for KPP–KT versus depth in °C for (b) 2003 and (c) 2006. In (a) the black line is Q_{tot} for 2003 and the blue line is Q_{tot} for 2006.

computation of MLD because if a profile starts too deep, the MLD may be missed entirely. The maximum first depth, however, needs to be deeper than 5 m because that is the starting depth of most Argo profiles. To avoid very near surface heating effects, which tend to be independent of the MLD, the MLD algorithms start looking for the MLD at the first profile depth below 5 m. As a result, the minimum MLD for a given profile can be anywhere from 5 m to 10 m, depending on the depth sampling of the profile. Profiles must have at least three depth levels. In very rare cases where the water depth is very shallow three depth levels may be able to identify the existence of a MLD. Profiles must have depth levels that are monotonically increasing. Profiles must be in the ocean. In a few cases, the profile location is on land according to the bathymetry of the model at the nearest grid point. The land-sea boundary in HYCOM is set at 5 m, i.e., water depths <5 m are excluded and considered as land.

In our analysis observed profiles must not have depth sampling gaps >40 m for the depth range $\geqslant$0 and <150 m, 80 m for the depth range $\geqslant$150 m and <300 m, and >200 m for the depth range >300. The allowable gaps increase with depth because some MOODS profiles are sub-sampled at deep levels to save memory, resulting in coarse sampling at depth. The upper level limit of 150 m is chosen because most XBTs reach at least this depth. Many XBTs are designed for 200 m but some do not make it that deep.

It is also important to emphasize that a given density profile can end at a depth of, say 200 m, but MLD can reach a depth >200 m. For this case, any given MLD criterion can mistakenly detect MLD as the lowest bottom depth level, i.e., 200 m. Thus, in our procedures, if the MLD computed using threshold and curvature methodologies from observed profiles occurs within 10 m of the last depth sample and the bottom depth of the water is more than ± 50 m away, the MLD value for that profile is not considered.

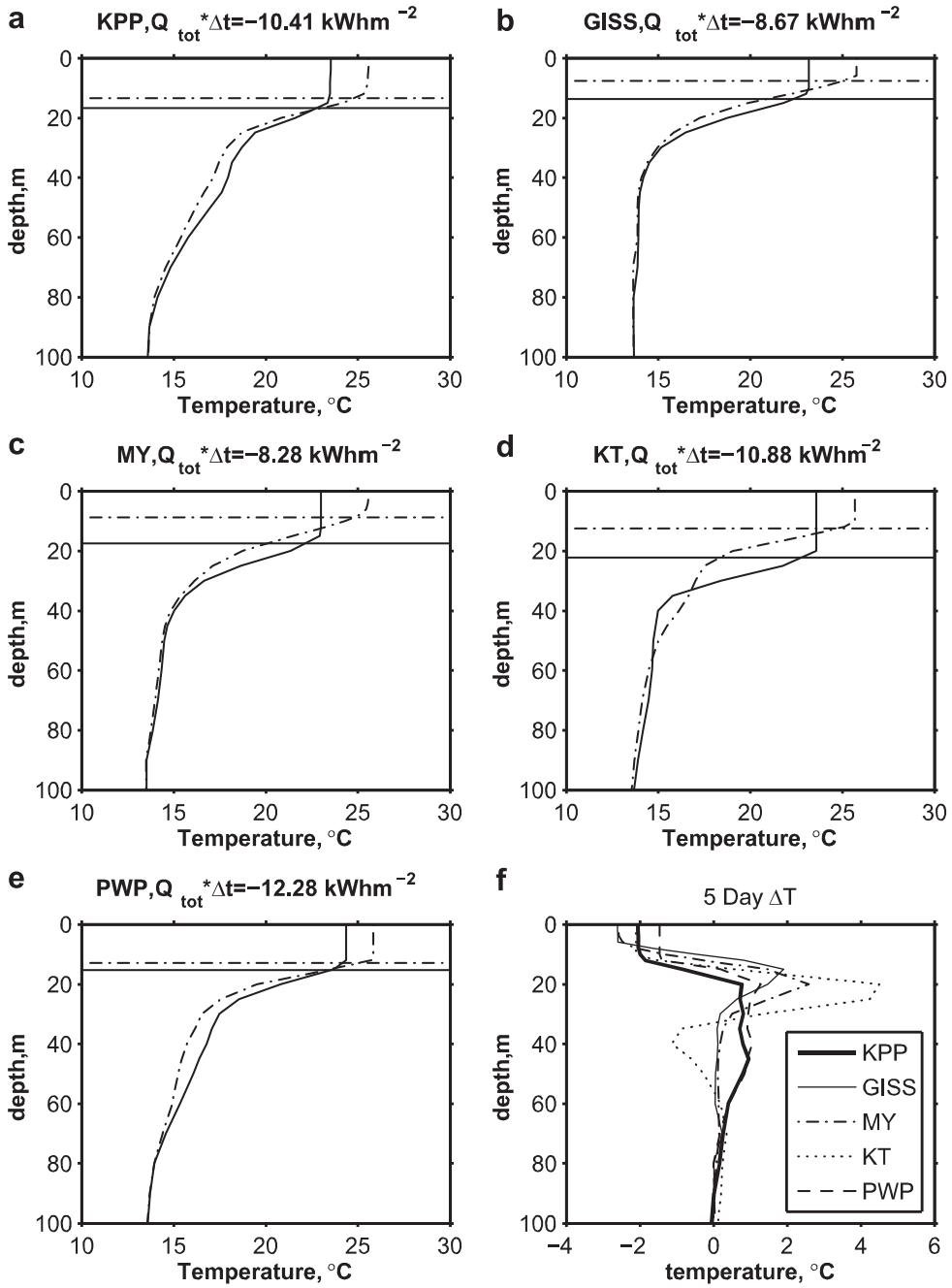

Fig. 10. An example profile located in the Gulf of Lyons in the same location as in Fig. 9 for a five day interval starting 31st August 2003 for (a) KPP, (b) GISS, (c) MY, (d) KT, and (e) PWP mixing models. The dashed-dot vertical curve represents the model profile at 31st August and the solid vertical curve is for 4th September. In (f) the difference between the beginning and ending profiles are shown for the five mixing models denoted in the legend. In (a)–(e) the horizontal dashed-dot line represents the curvature MLD while the horizontal solid line represents the threshold MLD. The number with units of kW h m^{-2} in (a)–(e) represents the integral of the net heat flux, Q_{tot}, for the five day interval from the corresponding mixing model. The negative values indicate that the sea surface is cooling during this time.

This is only applied to the observation profiles from MOODS, ARGO and WOD05.

5.2. MLD validation for models

We compare MLDs obtained from observed potential density profiles with those simulated by each mixing model (KPP, GISS, MY, KT and PWP). First, the observed and quality controlled *T* and *S* profiles are matched in space and time to the *T* and *S* profiles from each model case interpolated to the observation depth levels. Potential density is then computed from all *T* and *S* profile pairs. Finally, both the threshold and curvature MLDs are computed from both the observation and model potential density profiles.

In Section 2, it was indicated that there are a total of 20 hybrid layers in the model, and these vertical layers in HYCOM move in time and space. Vertical sampling of density values from the model can affect MLD from each mixing model since an interpolation is performed between the two layers in determining the depth of mixed layer. However, the layers in HYCOM are 3–10 m thick near the surface and track the density in the ocean interior so the MLD calculation should be relatively accurate. The situation can be much worse for observational profiles, where the sampling pattern is not physically-based. Thus, we also apply another flag which

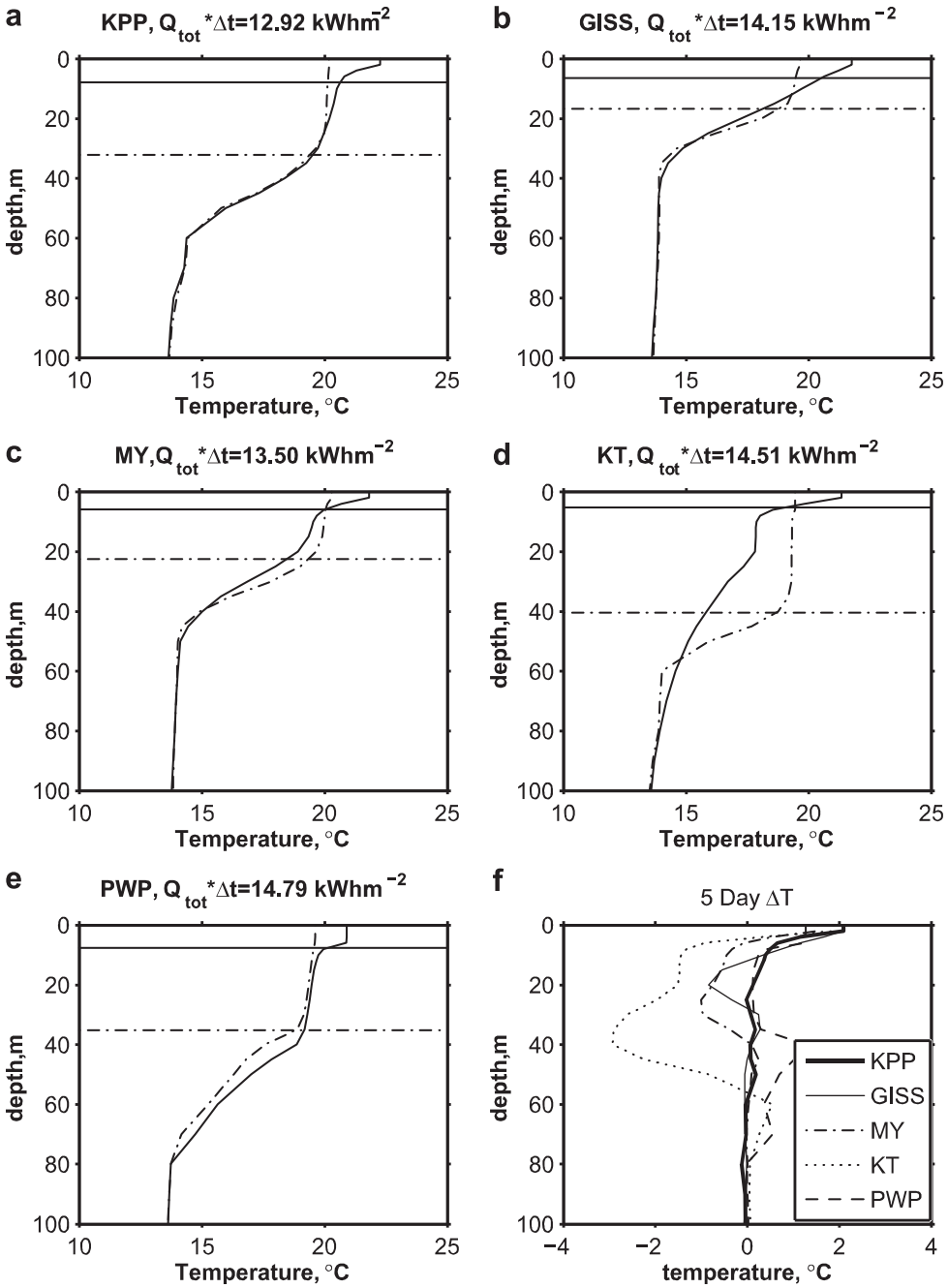

Fig. 11. An example profile located in the Gulf of Lyons in the same location as in Fig. 10. Each panel is also in the same format as in Fig. 10 except that the five day interval starts on 1st September and ends on 5th September 2006 and is during a warming period.

identifies the quality of the profiles based on the vertical resolution, so that model MLD errors can be quantified based on the vertical depth sampling quality.

A total of four categories for the vertical depth sampling quality are given in Table 4. Level 1 is the strictest category, with levels 2, 3, and 4 having increasingly weaker depth sampling requirements. For the validations of the mixing models, we first classify each density profile obtained from ARGO, MOODS and WOD05 data sets based on the vertical depth sampling qualities of levels 1–4. We then sample the model at the same depths which are present in the observational profiles and compute the MLD from the sampled profile.

Evaluations of each mixing model are performed on all vertical resolution levels shown in Table 4 to examine whether or not the validation statistics change depending on how fine/coarse the profile resolution is when determining the MLD. The comparisons are performed using the following statistical metrics: mean error (ME), root-mean-square (RMS) difference, correlation coefficient (R) and normalized RMS (NRMS). These metrics are computed based on the time series of MLD between observations and models as follows:

$$\mathrm{ME} = \overline{Y} - \overline{X}, \tag{1}$$

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

$$R = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \Big/ (\sigma_X \sigma_Y), \tag{3}$$

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

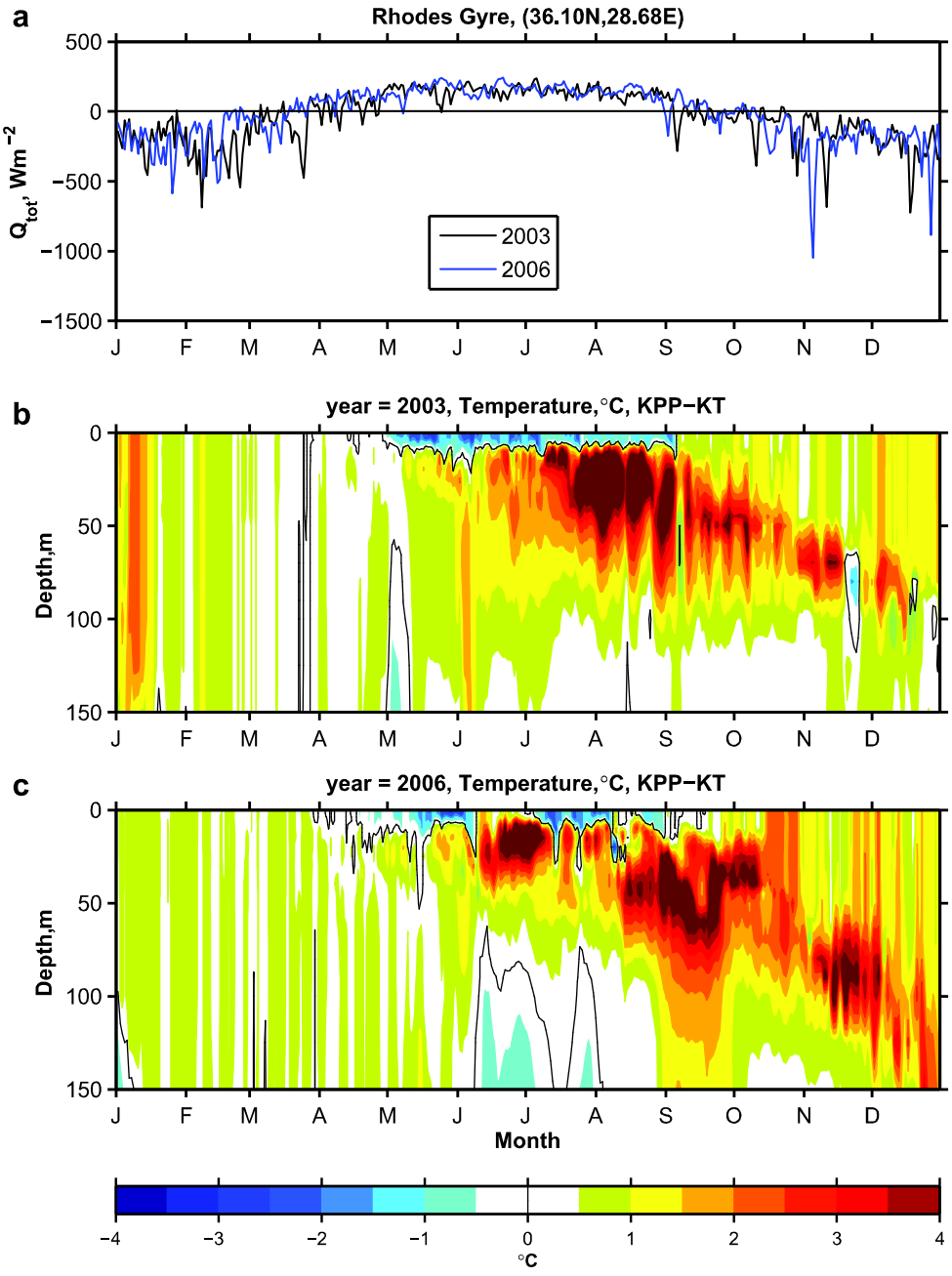

Fig. 12. One year time series at a point in the Rhodes gyre region located at 36.1°N, 28.69°E for 2003 and 2006 of (a) total surface flux (Q_{tot}) positive into the ocean in W m^{-2} and subsurface temperature differences for KPP–KT versus depth in °C for (b) 2003 and (c) 2006. In (a) the black line is Q_{tot} for 2003 and the blue line is Q_{tot} for 2006.

where X and Y denote observed and simulated MLDs, respectively. The NRMS is used in addition to the other traditional statistical metrics since it reduces skewness in the distribution of the errors. In other words, 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 deviations than summer MLDs.

The resulting statistical values between observed and simulated MLDs are provided in Fig. 15 for the threshold MLD and in Fig. 16 for the curvature MLD definition. Computations are performed for resolution levels 2–4, separately since the number of profiles is different for each one. Level 1 edited profiles are not used since only 235 profiles passed that level of resolution quality, while levels 2, 3 and 4 had 453, 3492, and 3652 profiles, respectively.

The results shown in Figs. 15 and 16 indicate that the GISS mixing model performs the best relative to observation profiles and has the lowest RMS, ME, and NRMS for both threshold and curvature MLD definitions. Both KPP and MY mixing models have slightly larger errors. The bulk mixing models, KT and PWP have substantially larger errors. This increase in error is even greater for the curvature MLD definition. For both MLD definitions, mean errors between observed and simulated MLDs are small with values of <10 m, in general.

Mean error and RMS differences for the MLD increase when coarser vertical depth sampling qualities (e.g., levels 3 and 4) are applied. This indicates that specifying the vertical depth sampling quality of profiles can have an impact on determining the MLD. In addition, the use of a curvature methodology rather than the threshold methodology can also alter the validation results as

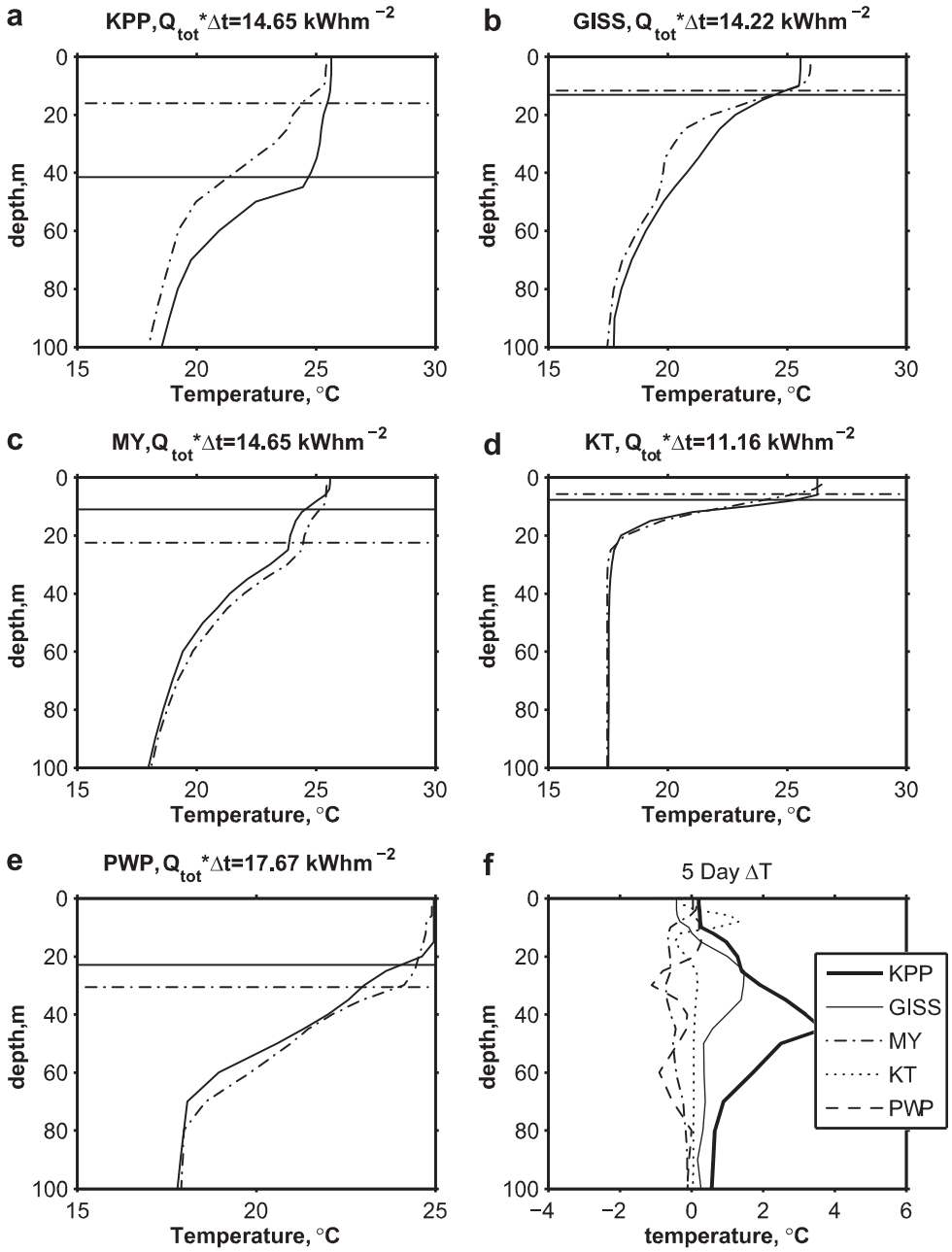

Fig. 13. An example profile located in the Rhodes gyre region in the same location as in Fig. 12 for a five day interval starting 1st August 2003 for (a) KPP, (b) GISS, (c) MY, (d) KT, and (e) PWP mixing models. The dashed-dot vertical curve represents the model profile on 1st August and the solid vertical curve is for 5th August. In (f) the difference between the beginning and ending profiles are shown for the five mixing models denoted in the legend. In (a)–(e) the horizontal dashed-dot line represents the curvature MLD while the horizontal solid line represents the threshold MLD. The number with units of kW h m^{-2} in (a)–(e) represents the integral of the net heat flux, Q_{tot}, for the five day interval from the corresponding mixing model. Positive values indicate warming at the sea surface.

evident from ME, *R* and NRMS values. The performance of the mixing models, however, is similar for each MLD methodology. In the case of level 4, the lowest threshold RMS values are 44 m and 46 m for GISS and MY, respectively. The same is true for the curvature RMS with values of 36 m and 39 m. The RMS error values for the curvature MLD are smaller than those for the threshold MLD.

Example observation profiles for the Gulf of Lyons are shown in Fig. 17. During February (Fig. 17a) there is a drastic difference between the observed threshold and curvature MLD. This is because the whole water column is well mixed and the MLD definitions are unable to detect the MLD, though the threshold MLD is more representative. During October (Fig. 17b) the observation has a sharp gradient at the base of the mixed layer while only the KPP profile has a sharp gradient but at the wrong depth. The example in Fig. 17b is representative of the fall season in the Gulf of Lyons, in that cooling conditions that drive the thermocline deeper are more difficult for HYCOM to reproduce and the mixing models have the largest differences.

6. Summary and conclusions

Through the use of five mixing models (KPP, GISS, MY, KT and PWP), we examined the variability of MLD in the Mediterranean Sea from 2003 to 2006. All models were run using the same high temporal resolution (3 h) atmospheric forcing. The resulting

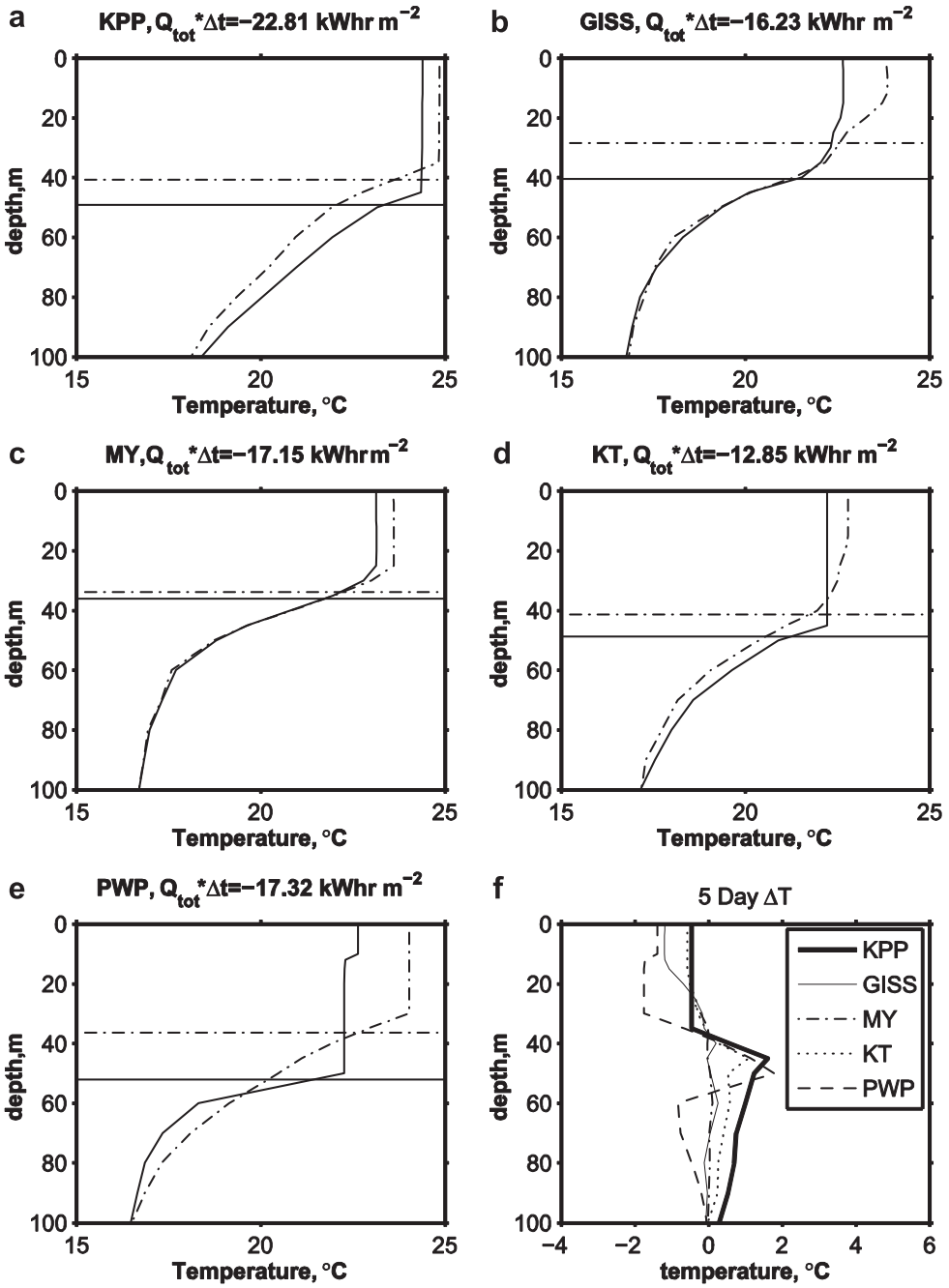

Fig. 14. An example profile located in the Rhodes Gyre region in the same location as in Fig. 12. Each panel is also in the same format as in Fig. 13 except that the five day interval starts on 16th October and ends on 20th October 2006 and is during a surface cooling period.

Table 3
Total number of *T*-only and *T* and *S* profiles for each month for 2003–2006.

Month	Type	January	February	March	April	May	June	July	August	September	October	November	December	Total
2003	*T*	0	3	3	7	29	41	40	40	119	83	72	5	442
2003	*T* and *S*	38	41	44	29	43	31	47	41	51	76	65	63	569
2004	*T*	14	20	16	15	39	39	12	69	164	421	322	276	1407
2004	*T* and *S*	71	54	73	62	53	58	60	52	114	124	80	86	887
2005	*T*	314	212	113	67	175	95	0	4	27	81	193	76	1357
2005	*T* and *S*	87	94	119	105	94	86	99	81	65	96	106	110	1142
2006	*T*	3	57	14	81	3	14	22	32	24	117	190	30	587
2006	*T* and *S*	97	113	119	135	153	125	108	90	93	114	119	112	1378
Total	*T*	331	292	146	170	246	189	74	145	334	702	777	387	3793
Total	*T* and *S*	293	302	355	331	343	300	314	264	323	410	370	371	3976

Table 4
Vertical depth sampling quality used for determining the MLD. The maximum allowable vertical distance between two data levels, over three depth ranges are shown for four level categories. For example, level 1 profiles have vertical resolution of (i) $\leqslant$5 m from the surface to 150 m, (ii) $\leqslant$10 m between 150 m and 300 m, and (iii) $\leqslant$20 m for the rest of the profile.

Depth range	Level 1 (m)	Level 2 (m)	Level 3 (m)	Level 4 (m)
0–150 m	5	10	25	40
150–300 m	10	20	50	80
Above 300 m	20	40	100	200

subsurface temperature and salinity fields were then processed to obtain MLD. Because MLD determination can vary depending on the definition used, we applied two different methodologies. To investigate the sensitivity of model-data comparisons to depth sampling, the vertical sampling quality of the observation profiles was varied. Additional observation selection procedures were also applied to ensure robustness of the results.

A goal of this paper is to answer four questions, three listed in the introduction and one listed in Section 4:

(1) Which mixing model is suitable for simulating MLD in the Mediterranean Sea? While all five models (KPP, GISS, MY, KT, and PWP) performed well, the GISS mixing model as implemented in NYCOM for the Mediterranean Sea had the lowest errors when evaluate against observations using both threshold and curvature MLD definitions.

(2) Are there substantial differences among the mixing models? While statistical values for the validation are quite similar, GISS and MY slightly outperform others. The bulk mixing models (KT and PWP) have substantial accuracy deficiencies relative to the higher order mixing models (KPP, GISS, and MY). The accuracy differences between the higher order models are considerably smaller. The added computational expense of MY mixing model does not seem to be justified based on the results of this experiment.

(3) What are the errors in MLD associated with each model? The modeled MLDs are slightly deeper than observed ones, which may be due to submesoscale processes not represented in any of the mixing models. The mean bias error (ME) tended to be less than 10 m for the higher order mixing models (KPP, GISS, and MY) while the bulk mixing model ME is 15 m or more. The RMS error for the higher order mixing models is $\sim$40 m while it is $\sim$50 m for the bulk mixing models.

(4) Do results change when the simulations are evaluated with different MLD definitions (threshold versus curvature)? The bulk mixing models (KT and PWP) had substantially larger

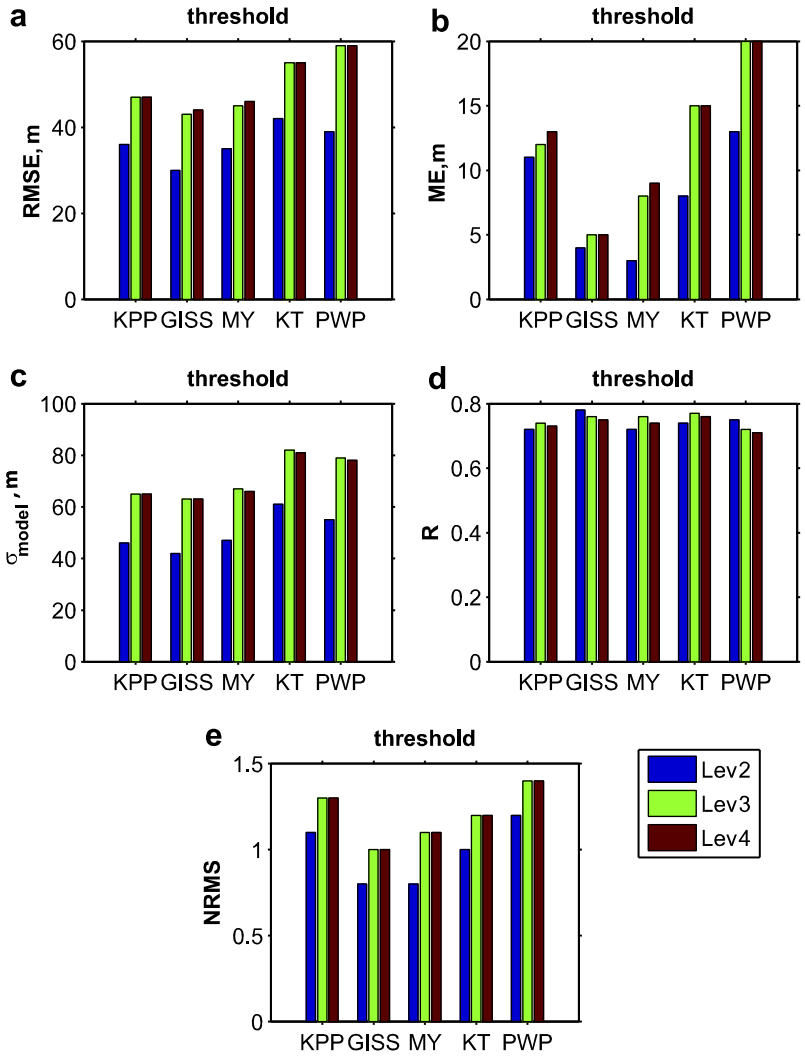

Fig. 15. Validation result using the threshold MLD definition for (a) RMSE, (b) ME, (c) model standard deviation (σ_{model}), (d) R and (e) NRMS for each mixing model indicated at the bottom of each plot. The bar graph colors represent data editing levels 2, 3, and 4 as indicated in the legend.

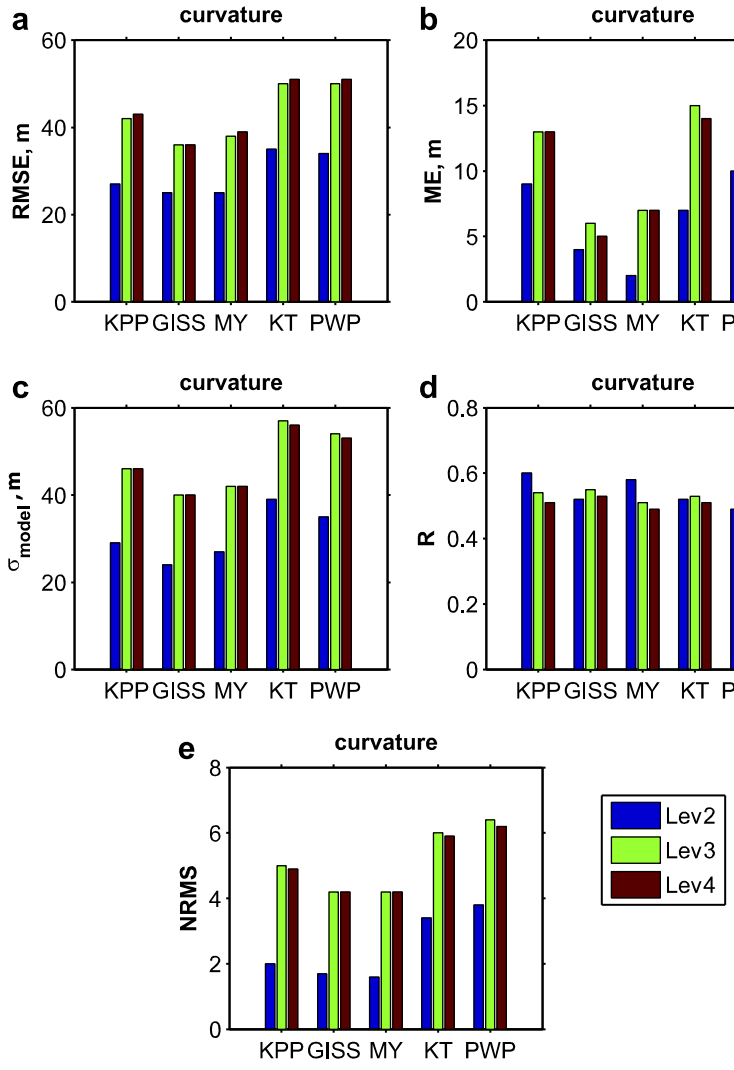

Fig. 16. Validation result using the curvature MLD definition for (a) RMSE, (b) ME, (c) model standard deviation (σ_{model}), (d) R and (e) NRMS for each mixing model indicated at the bottom of each plot. The bar graph colors represent data editing levels 2, 3, and 4 as indicated in the legend.

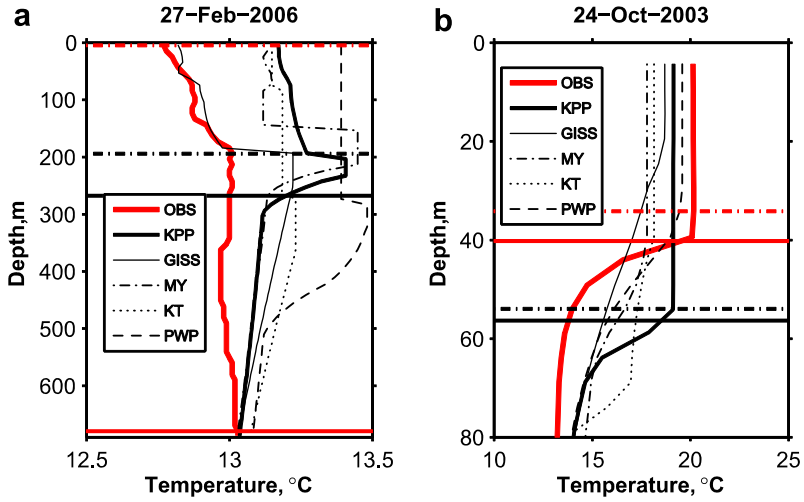

Fig. 17. Example observation profiles (red curves) from the Gulf of Lyons region on (a) 27th February 2006 and (b) 24th October 2003. The black curves are model results for the same time and location as the observation for the five mixing model case shown in the legend, which identifies the line-style. The red horizontal lines are the threshold MLD (solid red) and the curvature MLD (red dashed-dot) from the observation profile. The black horizontal lines are the threshold MLD (solid black) and the curvature MLD (black dashed-dot) from the KPP mixing model case.

errors particularly for the curvature MLD definition. This suggests that the bulk mixing models do a poorer job of capturing the vertical gradient of the observed profiles.

We also demonstrate that care must be taken in determining the allowable vertical gaps in observation profiles before performing model-data comparisons. Large vertical data gaps can cause misleading results when used in model validation. This was demonstrated with the use of four different vertical sampling quality levels.

With regard to vertical gradient and shape characteristics, the higher order mixing models (KPP, GISS, and MY) tended to outperform the bulk formulation mixing models (KT and PWP). Deficiencies in profile shape have a bigger impact when using the curvature MLD definition.

Acknowledgments

We would like to thank D. Franklin for processing the atmospheric forcing fields, P. Thoppi and J. Richman for helpful discussions, and the helpful comments of two anonymous reviewers. This research is funded by the Office of Naval Research (ONR) under program element 601153N as part of the NRL 6.1 Global Remote Littoral Forcing via Deep Water Pathways Project. With sadness, the co-authors report that the lead author, Dr. A. Birol Kara, passed away on 14th September 2009 at the age of 39 after a valiant battle against cancer. This paper is contribution NRL/JA/7320/09/9104 and has been approved for public release.

References

Astraldi, M., Gasparini, G.P., Vetrano, A., Vignudelli, S., 2002. Hydrographic characteristics and interannual variability of water masses in the central Mediterranean: a sensitivity test for long-term changes in the Mediterranean Sea. Deep Sea Res. Part I 49, 661–680.

Bleck, R., 2002. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. Ocean Modell. 4, 55–88.

Boyer, T.P., Antonov, J.I., Garcia, H.E., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., et al., 2006. World Ocean Database 2005 DVDs. In: Levitus, S. (Ed.), NOAA Atlas NES DIS, vol. 60. US Government Printing Office, Washington, DC, USA, p. 190.

Canuto, V.M., Howard, A., Cheng, Y., Dubovikov, M.S., 2002. Ocean turbulence, Part II: Vertical diffusivities of momentum, heat, salt, mass, and passive scalars. J. Phys. Oceanogr. 32, 240–264.

de Boyer Montégut, C., Madec, G., Fischer, A.S., Lazar, A., Iudicone, D., 2004. Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. J. Geophys. Res. 109, C12003. doi:10.1029/2004JC002378.

Fernãndez, V., Umlauf, L., Dobricic, S., Burchard, H., Pinardi, N., 2006. Validation and intercomparison of two vertical-mixing schemes in the Mediterranean Sea. Ocean Sci. Discuss. 3, 1945–1976.

Fofonoff, N.P., Millard, R.C. Jr., 1983. Algorithms for computations of fundamental properties of seawater. UNESCO Technical Papers in Marine Science, No. 44, p. 53.

Gould, J., Roemmich, D., Wijffels, S., Freeland, H., Ignaszewsky, M., Jainping, X., et al., 2004. ARGO profiling floats bring new era of in situ ocean observation. EOS, Trans. AGU 85 (179), 190–191.

Helber, R.W., Barron, C.N., Carnes, M.R., Zingarelli, R.A., 2008. Evaluating the sonic layer depth relative to the mixed layer depth. J. Geophys. Res. 113, C07033. doi:10.1029/2007JC004595.

Hogan, T.F., Rosmond, T.E., 1991. The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. Mon. Weather Rev. 119, 1786–1825.

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

Kara, A.B., Hurlburt, H.E., Wallcraft, A.J., 2005. Stability-dependent exchange coefficients for air–sea fluxes. J. Atmos. Ocean. Technol. 22, 1080–1094.

Kara, A.B., Wallcraft, A.J., Barron, C.N., Hurlburt, H.E., Bourassa, M.A., 2008. Accuracy of 10 m winds from satellites and NWP products near land-sea boundaries. J. Geophys. Res. 113, C10020. doi:10.1029/2007JC004516.

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

Large, W.G., Danabasoglu, G., Doney, S.C., McWilliams, J.C., 1997. Sensitivity to surface forcing and boundary layer mixing in a global ocean model: annual-mean climatology. J. Phys. Oceanogr. 27, 2418–2447.

Lorbacher, K., Dommenget, D., Niiler, P.P., Köhl, A., 2006. Ocean mixed layer depth: a subsurface proxy of ocean-atmosphere variability. J. Geophys. Res. 111. doi:10.1029/2003JC002157.

Marshall, J., Schott, F., 1999. Open-ocean convection: observations, theory, and models. Rev. Geophys. 37, 1–64.

Mellor, G.L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys. Space Phys. 20, 851–875.

Mertens, C., Schott, F., 1998. Interannual variability of deep-water formation in the Northwestern Mediterranean. J. Phys. Oceanogr. 28, 1410–1424.

NAVOCEANO, 2003. Database description for the generalized digital environmental model (GDEM-V) Version 3.0. OAML-DBD-72, 34 pp. (Available from NAVOCEANO Oceanographic Data Bases Division, Stennis Space Center, MS 39522—5001).

Ohno, Y., Kobayashi, T., Iwasaka, N., Suga, T., 2004. The mixed layer depth in the North Pacific as detected by the Argo floats. Geophys. Res. Lett. 31, L11306. doi:10.1029/2004GL019576.

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.

Teague, W.J., Carron, M.J., Hogan, P.J., 1990. A comparison between the Generalized Digital Environmental Model and Levitus climatologies. J. Geophys. Res. 95, 7167–7183.