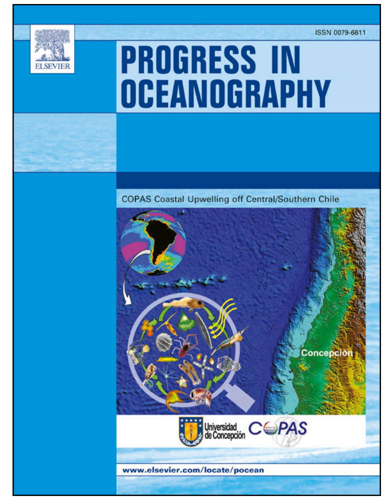
Journal Pre-proofs



A comparison of five surface mixed layer models with a year of observations in the North Atlantic

Gillian M. Damerell, Karen J. Heywood, Daley Calvert, Alan L. M. Grant, Michael J. Bell, Stephen E. Belcher

PII: S0079-6611(20)30055-0
DOI: https://doi.org/10.1016/j.pocean.2020.102316
Reference: PROOCE 102316

To appear in: *Progress in Oceanography*

Received Date: 2 May 2019
Revised Date: 12 December 2019
Accepted Date: 18 March 2020

Please cite this article as: Damerell, G.M., Heywood, K.J., Calvert, D., M. Grant, A.L., Bell, M.J., Belcher, S.E., A comparison of five surface mixed layer models with a year of observations in the North Atlantic, *Progress in Oceanography* (2020), doi: https://doi.org/10.1016/j.pocean.2020.102316

A comparison of five surface mixed layer models with a year of observations in the North Atlantic

Gillian M. Damerell[a,*], Karen J. Heywood[a], Daley Calvert[b], Alan L. M. Grant[c], Michael J. Bell[b], Stephen E. Belcher[b]

[a] *Centre for Ocean and Atmospheric Sciences, School of Environmental Science, University of East Anglia, Norwich, United Kingdom*
[b] *Met Office, Hadley Centre, FitzRoy Road, Exeter, Devon, United Kingdom*
[c] *Department of Meteorology, University of Reading, Reading, United Kingdom*

Abstract

Five upper ocean mixed layer models driven by ERA-Interim surface forcing are compared with a year of hydrographic observations of the upper 1000 m, taken at the Porcupine Abyssal Plain observatory site using profiling gliders. All the models reproduce sea surface temperature (SST) fairly well, with annual mean warm biases of 0.11°C (PWP model), 0.24°C (GLS), 0.31°C (TKE), 0.91°C (KPP) and 0.36°C (OSMOSIS). The main exception is that the KPP model has summer SSTs which are higher than the observations by nearly 3°. Mixed layer salinity (MLS) is not reproduced well by the models and the biases are large enough to produce a non-trivial density bias in the Eastern North Atlantic Central Water which forms in this region in winter.

All the models develop mixed layers which are too deep in winter, with average winter mixed layer depth (MLD) biases between 160 and 228 m. The high variability in winter MLD is reproduced more successfully by model estimates of the depth of active mixing and/or boundary layer depth than by model MLD based on water column properties. After the spring restratification event, biases in MLD are small and do not appear to be related to the preceding winter biases.

There is a very clear relationship between MLD and local wind stress in all models and in the observations during spring and summer, with increased wind speeds leading to deepening mixed layers, but this relationship is not

*Corresponding author
Email address: g.damerell@uea.ac.uk (Gillian M. Damerell)

Preprint submitted to Progress in Oceanography *December 12, 2019*

present during autumn and winter. We hypothesize that the deepening of the MLD in autumn is so strongly driven by the annual cycle in surface heat flux that the winds are less significant in the autumn. The surface heat flux drives a diurnal cycle in MLD and SST from March onwards, though this effect is much more significant in the models than in the observations.

We are unable to identify one model as definitely better than the others. The only clear differences between the models are KPP's inability to accurately reproduce summer SSTs, and the OSMOSIS model's more accurate reproduction of MLS.

Keywords: Ocean models, surface mixed layer, ocean gliders

1. Introduction

Climate models are important tools for understanding the climate and its response to various forcings (Flato et al., 2013). The surface mixed layer forms the boundary between the ocean and atmosphere, and regulates exchanges of heat, momentum and trace gases. The ability of the oceans to buffer atmospheric climate change by absorbing and then storing heat and radiatively important trace gases relies heavily upon the exchanges in the surface mixed layer (Belcher et al., 2012). Thus surface mixed layer parameterisations which accurately reproduce observed behaviour are a vital tool in developing climate models which can make reasonable predictions of the future response to anthropogenic activity.

Here we compare various 1D mixed layer models with observations (Damerell et al., 2016) of mixed layer properties taken over a full year in the Northeast Atlantic using profiling gliders, as part of the Ocean Surface Mixing, Ocean Submesoscale Interaction Study (OSMOSIS). Various properties are considered to compare the performance of the various models. First and foremost is the ability of the models to reproduce the observed sea surface temperature (SST), since this is of considerable importance for the exchange of heat with the atmosphere. Unlike at Ocean Station Papa (OSP), where Large et al. (1994) find that model/observation SST comparisons are only reliable from April to October because of the relative importance of net surface heat fluxes and advective fluxes at different times of year, Lazarevich et al. (2004) found that in the North Atlantic a modified form of the Price-Weller-Pinkel mixed layer model, using NCEP-derived surface forcing, accurately reproduced float-observed temperatures and meteorological-based SSTs to within

1°C for an entire year. Moreover, Damerell et al. (2016), using the same observational dataset as used here, found that the mixed layer temperature is strongly correlated ($r = 0.87$) with the cumulative net surface heat flux from ECMWF ERA-Interim reanalysis data (Dee et al., 2011). The main differences were during the autumn, when cooler water from below is entrained into the mixed layer, and late summer, when the very shallow mixed layer depth (MLD) means that some of the absorption of solar radiation will occur below the mixed layer. These processes (entrainment of water from below and penetration of solar radiation) are represented in the models used here, so we expect the models to reproduce observed SST reasonably well for the whole year.

Mixed layer salinity (MLS) is discussed because of its impact on mixed layer density and MLD. (It was not practical to compare sea surface salinity as the nature of glider data collection means there are gaps in the surface salinity data after quality control.) Unlike SST, Damerell et al. (2016) find that the MLS of this dataset is not correlated with the surface freshwater fluxes from ERA-Interim though it is weakly correlated with the currents ($r = 0.4$). They conclude that the changes in MLS must be influenced by advection into the area of water masses of different salinity and/or vertical mixing with waters of different salinity from the ocean interior, and while the latter may be reproduced in 1D models, the former is not. Hence we do not necessarily expect the MLS of the models to agree with the observations particularly well. We also compare the MLD, since this is an important factor in the development of the surface mixed layer and interaction with the ocean interior.

We discuss the coherence between observations and model output, and coherence with surface forcing. Note that we use potential temperature and practical salinity throughout, and all densities are potential density anomalies (σ_θ) relative to the surface and will be given without units.

Many other authors have compared 1D models to ocean observations, e.g., Large et al. (1994); Kantha and Clayson (1994); Burchard and Bolding (2001); Lazarevich et al. (2004); Acreman and Jeffery (2007); Pookkandy et al. (2016). However, this has generally been done using observations from moorings (such as OSP) where the limited vertical resolution will affect measurement of the MLD, or observations from floats which may have limited vertical and/or temporal resolution, or from ship CTDs which will not provide long time series of profiles in one location. The profiling gliders used here provide profiles to 1000 m with a vertical resolution of 2 m, at approximately

2-hourly intervals for a whole year. Thus the observational data is particularly well suited to comparisons with model output. The good temporal resolution also allows the application of wavelet coherence methods (section 4) to this question.

Section 2 describes the observational data set with which the models will be compared. Section 3 summarizes the key features of each of the models and describes the model setup. The Price-Weller-Pinkel (PWP), K-Profile-Parameterisation (KPP), Generic Length Scale (GLS) and Turbulent Kinetic Energy (TKE) models are described extensively elsewhere (e.g., Price et al., 1986; Lazarevich and Stoermer, 2001; Large et al., 1994; Gaspar et al., 1990; Rodi, 1987) so we give only brief descriptions here. We include a more complete description of the recently developed OSMOSIS model. Section 4 describes the wavelet analysis methods used to investigate the periodic behaviour of the data and models. Section 5 presents the results and compares the model and observed behaviour, and section 6 contains the conclusions.

2. Ocean glider observations of upper ocean hydrography

The OSMOSIS project incorporated a year-long observational programme centred 41 km to the southeast of the Porcupine Abyssal Plain sustained observatory (PAP-SO; Lampitt et al., 2010a), with observations collected within a 15 km radius of 48.7° N, 16.2° W (figure 1). This location is considered remote from the topographic complexities of the continental slope and the Mid-Atlantic Ridge (Hartman et al., 2012), and thus remote from places where strong internal tides might be generated. It is located in the inter-gyre region between the North Atlantic subpolar and subtropical gyres where the mean flow is relatively weak and eddy kinetic energy is moderate. The variability in physical properties is likely to be representative of large areas of the mid-latitude gyres.

As part of the OSMOSIS field campaign, profiling ocean gliders (Seagliders) were deployed for periods varying between two and five months, between them covering an entire year from 4th September 2012 to 7th September 2013. The Seaglider dataset consists of 3785 profiles at approximately 2-hourly intervals of temperature and salinity to 1000 m, with a vertical resolution of 2 m after gridding. Details of the sensors, data processing, quality control and calibration are given by Damerell et al. (2016). Temperature and salinity are considered accurate to 0.01°C and 0.01 respectively. The 15 km radius within which the observations were collected is comparable to the spacing

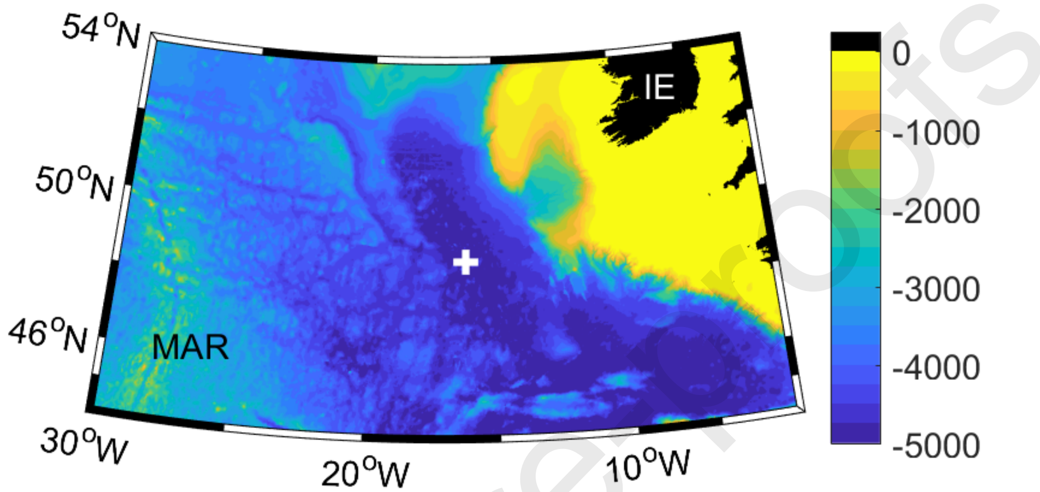


Figure 1: Bathymetry of the north-east Atlantic basin. The white cross marks the location of the OSMOSIS field campaign. MAR=mid-Atlantic Ridge. IE=Ireland.

between CTD locations of a typical ship-based hydrographic survey, and for the purposes of this paper, we treat the data as if they had all been obtained at the same location. There is an implicit linkage between spatial and temporal variability in glider observations, and here we choose to treat it as purely temporal variability.

The depth of the surface mixed layer is calculated using a threshold value of temperature or density from a near-surface value at 5 m depth ($\Delta T = 0.2$°C or $\Delta\sigma_\theta$ =0.03), whichever is the shallower (de Boyer Montegut et al., 2004). (MLD is calculated in the same manner for each model, see section 3.2.) Thus, we aim to find the MLD even in cases where temperature and salinity vary with depth in a density-compensating manner, as well as cases where density varies with depth due to changes in salinity rather than temperature. In 67% of the record the MLD is set by the density threshold, 19% by the temperature threshold, and in 13% of the record the two thresholds give the same MLD. There is no clear seasonal pattern in which threshold sets the MLD. We chose 5 m as the reference depth because above that there are too many gaps in the observational data due to the removal of salinity spikes during quality control. Spiking in the near-surface is unfortunately common in glider observations due to surface manoeuvres altering the flow of water past the sensors, cooling or warming while at the surface and air bubbles and particulates in/on sensors when leaving the surface. Note that this means that MLDs shallower than 5 m cannot be identified.

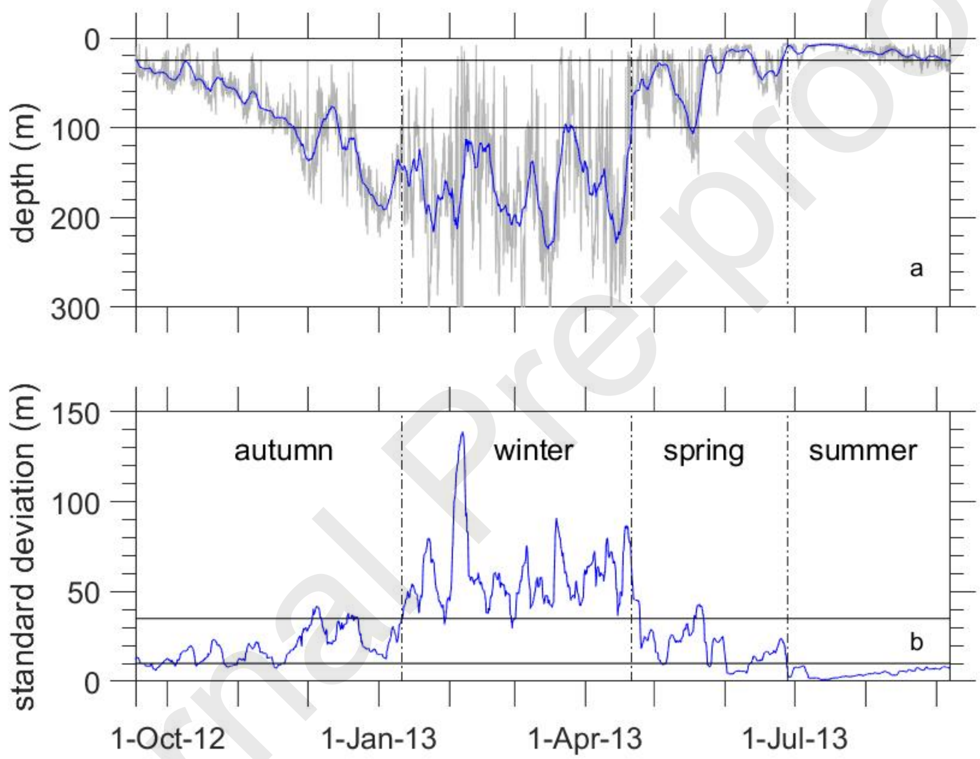


Figure 2: Definition of seasons as used in this paper. a) MLD calculated from the observations (gray), and running mean MLD (blue) calculated at each observation time over a 5-day window (i.e., with a window extending from 2.5 days before that observation time to 2.5 days after that observation time). Black horizontal lines are at 25 and 100 m. b) standard deviation of the observed MLD, calculated over a 5-day window as for the running mean MLD. This will be referred to as the running standard deviation of MLD. Black horizontal lines are at 10 and 35 m. Black vertical dotted lines on both panels show the dates which divide the year into seasons, as labeled on b).

We divide the year into four seasons based on the behaviour of the observed MLD. The start of winter is deemed to be the day when the running mean MLD, calculated over a 5 day window, is deeper than 100 m and the running standard deviation of MLD (calculated over the same 5 day window) is greater than 35 m (figure 2), and these criteria are fulfilled for a period of at least 5 days. In other words, winter is the period when the MLD is consistently deeper than 100m but is also quite variable due to the lack of a strong pycnocline within the upper water column (see below). The start of spring is deemed to be the day when the running mean MLD is shallower than 100 m and remains so for a period of at least a week, consistent with previous definitions used in this area (Lampitt et al., 2010b). Summer is deemed to be the period when the running mean MLD is shallower than 25 m, and the running standard deviation of MLD is less than 10 m, i.e., the MLD is consistently shallow and shows low variability due to the presence of a strong pycnocline. Using these definitions, autumn is the period from the start of the time series on 24 September 2012 to 10 January 2013, winter is from 11 January to 20 April 2013, spring is from 21 April to 27 June 2013, and summer from 28 June to the end of the time series on 7 September 2013.

A strong, stable pycnocline forms in summer, then gradually erodes during the autumn, until during winter there is very weak stratification to considerable depth. Erickson and Thompson (2018), using the same dataset, found that this definition of MLD still retained credibility in winter as chlorophyll values become near-zero at approximately the same depth (their figure 5). However, the winter MLD is sensitive to the precise thresholds used and it may be more accurate to say that the base of the mixed layer is no longer very well defined because of the lack of a strong pycnocline within the upper water column.

3. Models

3.1. Model selection

Although 1D models do not include full ocean physics and in particular the many lateral processes, this can allow for a cleaner inter-comparison of those processes which are included. The topics studied using 1D models vary widely. Some examples include: studies of the effect of new model processes (Chen et al., 1994) which is easier to do in a 1D model before integration into a full ocean model; studies of the effect of model resolution and tuning (Acreman and Jeffery, 2007); understanding physical processes varying from the role of local atmospheric forcing on mixed layer depth (Pookkandy et al., 2016), to tidally driven controls on the location of mixing fronts (Sheehan et al., 2018), to glacial meltwater fractions in the polar oceans (Biddle et al., 2017); investigating net community production (Martz et al., 2008; Yang et al., 2017); understanding spring bloom dynamics (Sharples et al., 2006).

Models were chosen for this study to include commonly used examples of the range of approaches used to parameterise the surface mixed layer (see, for example, Burchard et al. (2008) for a discussion of the different approaches to this question). These models assume the turbulent mixing is dominated by vertical fluxes, and varying degrees of complexity are used to parameterise these fluxes. Perhaps the simplest approach is that of bulk boundary layers where ocean properties are assumed to be vertically uniform in the mixed layer. PWP (Price et al., 1986) is an example of this type: a computationally efficient bulk mixed layer model which has been used for many years to study ocean physics and biogeochemistry (e.g., Lazarevich et al., 2004; Frants et al., 2013; Viglione et al., 2018; Farahat and Abuelgasim, 2019) due to its simplicity and ease of use (further details in section 3.3).

Another widely used approach is that of turbulent kinetic energy closure (TC), where the profiles of eddy diffusivity and viscosity are dependent on the local turbulent kinetic energy, which is prognostic (e.g., Mellor and Yamada, 1982; Kantha and Clayson, 1994; Harcourt, 2015). The properties of the turbulent flow are modelled directly by solving the Reynolds budgets for the second-order moments. The GLS and TKE mixed layer models used here are examples of 'one-' and 'two-equation' TC schemes (see further details in section 3.5). GLS and TKE are implemented in the NEMO ocean modelling framework (Madec, 2008) which is widely used for climate modelling (see, for example, list of publications at https://www.nemo-ocean.eu/).

K-profile parameterisation models aim to fill the middle ground between bulk mixed layer models and TC schemes by allowing for vertical property variations in the mixed layer via a specified vertical shape function (Large et al., 1994). Vertical turbulent fluxes in the absence of vertical gradients of ocean properties are permitted through a non-local transport parameterisation (Burchard et al., 2008; Van Roekel et al., 2018). The version used here is a single column of the Multi-Column K Profile Parameterisation mixed layer model (Hirons et al., 2015), which is used as a relatively computationally efficient alternative to a full ocean model in coupled atmosphere-ocean climate simulations and process studies (e.g., Lee and Klingaman, 2018; Hirons et al., 2018) (further details in section 3.4). Modifications to the KPP scheme to represent Langmuir turbulence (which arises through the interaction of ocean surface waves and the currents (McWilliams et al., 1997)), have been described by Li et al. (2016) and Li and Fox-Kemper (2017). However, for this study only the standard version of the KPP model is considered.

Finally, the OSMOSIS mixed layer model is a new boundary layer model developed as part of OSMOSIS, and currently undergoing implementation in NEMO (further details in section 3.6). Like the KPP scheme, turbulent transports in the OSMOSIS scheme are parameterised using non-local flux-gradient relationships which are related to the Reynolds budgets for the turbulent fluxes (Holtslag and Moeng, 1991; Abdella and McFarlane, 1997) obtained from large-eddy simulation. In the OSMOSIS scheme non-local flux-gradient relationships are used for both unstable and stable boundary layers. Unlike the KPP version used here, the OSMOSIS model has been designed to represent Langmuir turbulence, which has been advocated for in second-moment closures (e.g., Harcourt, 2013, 2015). The OSMOSIS scheme does not contain a parameterisation for the effects of shear across the base of the pycnocline, and there is no contribution of shear-driven mixing in either the mixed layer or the interior.

3.2. Model initiation and setup

All the models are forced at the surface with ECMWF ERA-Interim reanalysis data (Dee et al., 2011) listed in table 1 and shown in figure 3. ERA-Interim has a horizontal resolution of 0.75°, or approximately 80 km. We use data from the closest grid point (48.75° N, 16.5° W), 23 km from the centre of the OSMOSIS observations (48.7° N, 16.2° W). The time resolution of the surface fluxes is three hours. All models use a 10 minute time step, and the surface forcing data were linearly interpolated to the same 10 minute intervals to avoid any differences in how the models treat forcing data which are more sparse than the model time step.

Model performance has been shown to depend on vertical resolution (e.g., Large et al., 1994; Acreman and Jeffery, 2007), so here we use a fairly high vertical resolution of 1 m in every model. The models were all initialised with the same observed profiles of temperature and salinity collected by glider SG566 on 24th September 2012, interpolated to the 1 m grid (figure 4). The models are run from 24th September 2012 to 7th September 2013 (the end of the observational period) and output variables every hour.

All models use Jerlov water type 1B, which is considered to be an appropriate water type for the open Atlantic (Simonot and Le Treut, 1986; Stips, 2011). Jerlov water type refers to a set of coefficients that define the double

Parameter	Units
Surface thermal radiation	W m^{-2}
Surface solar radiation	W m^{-2}
Surface sensible heat flux	W m^{-2}
Surface latent heat flux	W m^{-2}
Precipitation	m of water
Wind components at 10 m$^+$	m s^{-1}
Coefficient of drag with waves$^+$	
2D wave spectra	m^2 s $radians^{-1}$
Surface Stokes drift components*	m s^{-1}

* Obtained from 2D wave spectra

$^+$ Surface stress calculated using drag coefficient and wind components

Table 1: Surface forcing parameters from ECMWF ERA-Interim reanalysis data.

exponential profile for shortwave radiation absorption (Paulson and Simpson, 1977). In using the same water type for the whole year we are ignoring the effect of changes in the optical properties of the water column due to, for example, phytoplankton growth. While this may increase differences between each model's output and the observations (Large et al., 1994), this will affect all the models similarly so should not invalidate comparisons between models. Since not all the models incorporate background diffusion, this is set to zero in those models which do include it. All model parameters (except background diffusion and Jerlov water type) are set to the default values for that model as described in the cited literature. This amounts to a particular choice of parameter values for each model and the results might differ for other choices, however investigation of the effect of parameter values is beyond the scope of this study.

SST for each model is the temperature at the first model grid depth, i.e., 1 m, comparable to the SST for the glider data which is the median value in the uppermost 2 m bin. We calculate MLD for each model based on the output profiles of temperature and salinity in exactly the same way in which MLD is calculated for the observations, so that we will be comparing like with like. However, each model also provides an estimate of the depth of active mixing or boundary layer depth, which are described below for each model. These will be referred to as the model's 'internal' mixing layer depth (IMLD), but note that this is not the same parameter for each model. For the TKE and GLS models this is diagnosed from the vertical eddy diffusivity and

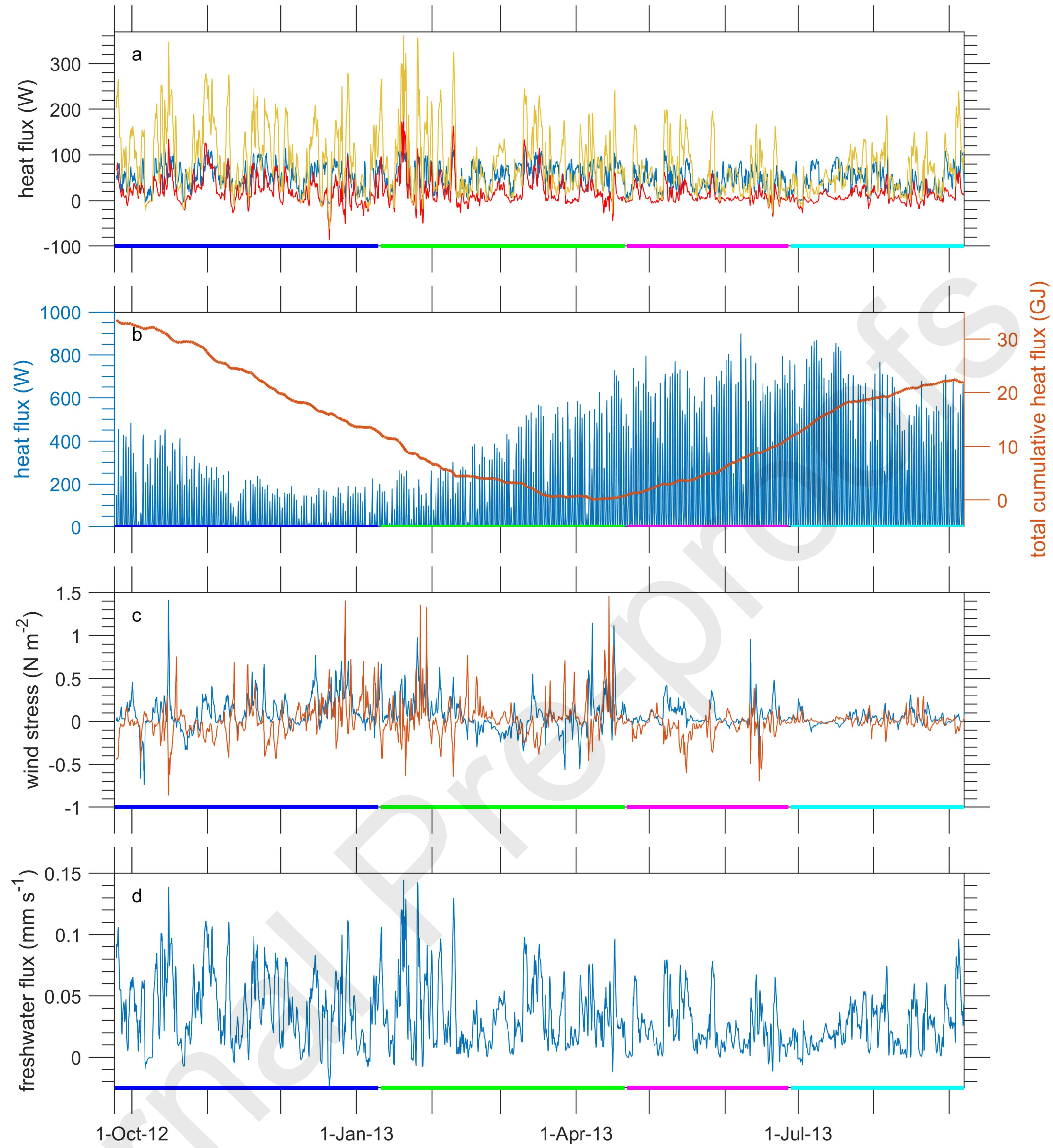


Figure 3: Surface forcing used to drive the models. a) Outgoing surface heat flux, positive upwards. Blue = longwave radiation, red = sensible heat, orange = latent heat. b) Blue = incoming shortwave radiation, positive downwards, red = total cumulative surface heat flux, positive downwards. c) Wind stress. Blue = zonal component, red = meridional component. d) Freshwater flux, i.e., precipitation minus evaporation, positive downwards. The coloured bars at the base of the panels mark the seasons: blue = autumn; green = winter; magenta = spring; cyan = summer.

has no impact on the vertical mixing scheme itself, but for the PWP, KPP and OSMOSIS models these are length scales that have actual numerical

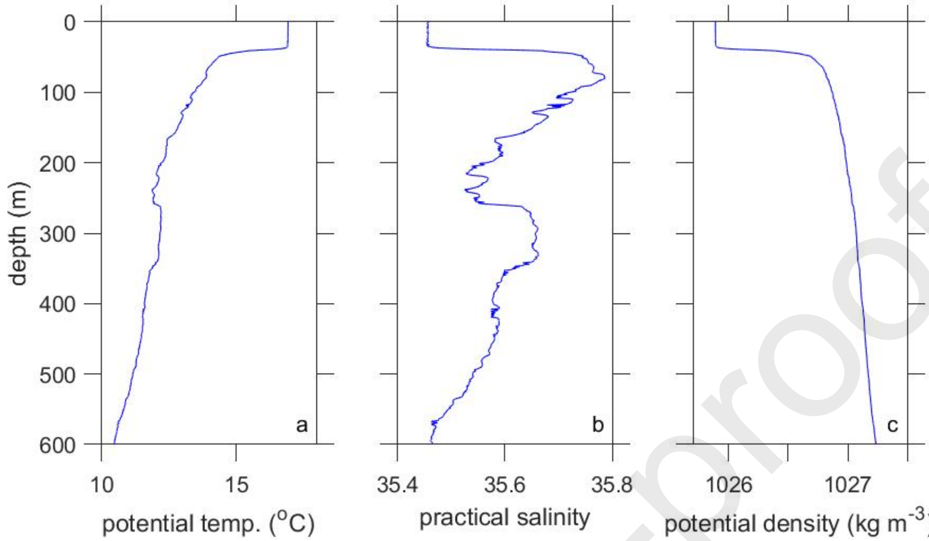


Figure 4: Profiles used to initialize the models: a) potential temperature, b) practical salinity, c) potential density.

impacts. All MLDs and IMLDs will be shown as positive downwards.

The observational dataset does not include estimates of the depth of active mixing, so we are unable to make direct comparisons between an observed depth of active mixing and the models' IMLDs. However, one would always expect the MLD in the ocean to be greater than or equal to the depth of active mixing because properties will be homogeneous at the depths where mixing is occurring plus there may be remnant homogeneous layers beneath from previous mixing episodes.

In model studies the relationship between MLD and IMLD can depend on the definition of IMLD used in that model, and on the definition of MLD with which it is compared. For example, Large et al. (1994) found boundary layer depths (IMLDs in our terminology) in large eddy simulations around 10% deeper than the mixed layer depth definition they were using (their figure 1). However, in the simulations discussed here, each model's IMLD was shallower than that model's MLD at all time steps. In other words, there is no prima facie reason to expect model IMLD to be deeper than the observed MLD. Hence, if a model's IMLD is deeper than the observed MLD we can deduce that it must be deeper than the depth of active mixing in the real ocean by at least as much as the difference between the model's IMLD and the observed MLD. If the model's IMLD is shallower than the observed

MLD we do not know how it differs from the depth of active mixing in the real ocean.

3.3. PWP

The PWP model (Price et al., 1986) was developed to investigate mixed layer processes in tropical oceans. It is a bulk mixed layer model, which means that it considers the main driving equations over the entire mixed layer, and averages the ocean properties (temperature, salinity, and meridional and zonal current velocities) over that layer. The focus is on the parameterisation of shear production of turbulent kinetic energy across the base of the mixed layer and in the pycnocline, which is parameterised through gradient Richardson number calculations. (Richardson number is a measure of the relative importance of stratification to destabilizing shear. "Bulk" Richardson number is a term used when the Richardson number is calculated over a slab containing several depth bins, whereas "gradient" Richardson number is not defined in the mixed layer itself but is calculated in the stratified region below the mixed layer.) The IMLD is found as the minimum depth required to keep a bulk Richardson number (Ri_b) of a well-mixed layer greater than a prescribed critical value, $Ri_b > 0.65$. This value was determined from field and laboratory experiments (Price et al., 1978). The model implementation used originates from Lazarevich and Stoermer (2001), which is a translation of the original PWP Fortran implementation into Matlab code.

3.4. KPP

The KPP mixed layer model is a turbulence closure scheme model which uses eddy diffusivity to parameterise small-scale turbulence within the mixed layer (Large et al., 1994). The model was developed from atmospheric boundary layer models that incorporated nonlocal transport terms in their mixing parameterisations. The diffusivity is formulated to agree with similarity theory of turbulence in the surface layer and is subject to the conditions that both it and its vertical gradient match the interior values at the base of the boundary layer. The diffusivities of the interior mixing processes (internal waves, shear instability, and double diffusion) are modeled as constants, functions of a gradient Richardson number, and functions of the double-diffusion density ratio. The IMLD is the minimum of three mixed layer depth definitions: the Ekman depth, the Monin-Obukhov length, and the depth where the bulk Richardson number exceeds the threshold $Ri_b > 0.3$ (Large et al., 1994). An important feature of this model is that the boundary layer allows entrainment into stable stratification below the mixed layer and can produce realistic exchanges of properties between the mixed layer and thermocline. The model script used is a single column of the Multi-Column K Profile Parameterisation ocean model (Hirons et al., 2015), developed by the National Centre for Atmospheric Science at the University of Reading (see https://puma.nerc.ac.uk/trac/KPP_ocean).

3.5. TKE and GLS

The TKE and GLS models refer to the 'TKE' and 'GLS' vertical mixing schemes implemented in the NEMO model (Madec, 2008). These schemes are based on the Turbulent Kinetic Energy scheme of Gaspar et al. (1990) and the Generic Length Scale framework of Umlauf and Burchard (2003) respectively, which both belong to the so-called 'Algebraic Stress Model' class of vertical mixing parameterisation (Burchard et al., 2008). This type of parameterisation approximates the turbulent fluxes using the eddy viscosity principle:

$$\begin{aligned} \overline{w'U'} &= -K_M \partial_z \overline{U} \\ \overline{w'T'} &= -K_H \partial_z \overline{T} \end{aligned} \tag{1}$$

where U is a horizontal velocity component, w is the vertical velocity component (positive upwards), T is a tracer, and K_M and K_H are respectively the eddy viscosity and eddy diffusivity. The prime and overbar notations represent the fluctuating and time-average components of the quantity respectively (i.e. Reynolds decomposition). K_M and K_H have the form:

$$\begin{aligned} K_M &= c_k l_k \sqrt{k} \\ K_H &= c_k^H l_k \sqrt{k} \end{aligned} \tag{2}$$

where c_k and c_k^H are dimensionless coefficients or stability functions, l_k is a mixing length and k is the turbulent kinetic energy. The calculation of c_k, c_k^H, l_k and k depends on the choice of turbulence closure. In the TKE scheme c_k and c_k^H are constant coefficients, and k is calculated using a prognostic budget equation. In stable stratification l_k is calculated using the simplified algebraic form suggested by Blanke and Delecluse (1993) where $l_k \propto N^{-1}$ (N is the buoyancy frequency), and l_k is bounded by the distance to the nearest physical boundaries (sea surface and bottom). In unstable stratification where $N^2 < 0$, l_k is the distance to the nearest physical boundary (sea surface/bottom) or layer of stable stratification. In the GLS framework c_k and c_k^H are complex nonlinear stability functions, and both l_k and k are calculated using prognostic budget equations. The GLS framework encompasses several well known closures for l_k and k, including '$k-kl$' (Mellor and Yamada, 1982), '$k-\epsilon$' (Rodi, 1987) and '$k-\omega$' (Wilcox, 1988). Due to the number of prognostic equations solved, the TKE scheme and GLS framework are examples of 'one-' and 'two-equation' closures respectively.

Reffray et al. (2015) explore the performance of the NEMO TKE and GLS vertical mixing schemes in a 1D column model case study at Ocean Station PAPA. Of the various closures implemented in the GLS framework, they find that the '$k-\epsilon$' model gives the best results in terms of temperature and salinity biases. Furthermore, they find that the TKE scheme significantly understates vertical mixing in the boundary layer and show that an ad-hoc parameterisation representing unresolved vertical mixing processes (Rodgers et al., 2014) is able to alleviate this. This parameterisation is implemented as an additional source of TKE that decays exponentially with depth. Reffray et al. (2015) show the TKE scheme to be highly sensitive to the choice of e-folding length scale and find that a 10 m length scale (their 'TKE_10m' experiment) gives the best results.

We use the 'TKE_10m' and '$k-\epsilon$' configurations of Reffray et al. (2015) as the basis for our TKE ('NEMO TKE') and GLS ('NEMO GLS') simulations respectively. The reader is referred to Reffray et al. (2015) for more details but should note that our simulations use a more recent version of NEMO (3.6), although this should have a negligible impact on the results. Additionally, K_M and K_H are set to an arbitrarily large value wherever static instabilities occur to ensure that these are homogenised within a time step. This has the effect of reducing the winter MLD by $\mathcal{O}$(10m).

For both NEMO simulations the IMLD is taken as the turbocline depth, which is the shallowest model depth where $K_H < 5 \times 10^{-4}$ m^2 s^{-1}.

3.6. OSMOSIS model

The OSMOSIS scheme combines a bulk model of the surface boundary layer (e.g. Kraus and Turner, 1967), which is coupled to a turbulence model based on non-local flux-gradient relationships (e.g. Large et al., 1994). The bulk model is used to determine the evolution of the depth of the boundary layer, and the turbulence model determines the mean profiles within the boundary layer, which are represented on a finite difference grid.

In unstable conditions the boundary layer is assumed to deepen through entrainment. The energy needed to entrain denser water from below the boundary layer is assumed to be supplied by a combination of Langmuir turbulence (McWilliams et al., 1997) and convective turbulence. The equation for the depth of the boundary layer is

$$\frac{\partial h_{\mathrm{bl}}}{\partial t} = -\frac{\overline{w'b'}_{ent}}{\Delta B} + \overline{w} \tag{3}$$

where h_{bl} is the boundary depth, $\overline{w'b'}_{ent}$ is the buoyancy flux associated with entrainment, ΔB is the difference between the buoyancy averaged over the depth of the boundary layer and the buoyancy just below the base of the boundary layer, and $\overline{w}$ is the large-scale vertical velocity, which is assumed to be zero in the integrations presented here. The layer averaged buoyancy is obtained by averaging the buoyancies on the model levels, which provides the coupling between the bulk and turbulence components of the OSMOSIS scheme.

The buoyancy flux associated with entrainment is parameterised as

$$\overline{w'b'}_{ent} = -0.03\frac{w_{*L}^3}{h_{\mathrm{bl}}} - 0.2\overline{w'b'}_0 \tag{4}$$

where w_{*L} is the velocity scale for Langmuir turbulence (Grant and Belcher, 2009) and $\overline{w'b'}_0$ is the surface buoyancy flux. The parameterisation of the contribution made by Langmuir turbulence to $\overline{w'b'}_{ent}$ is taken from Grant and Belcher (2009).

$$w_{*L} = (u_*^2 u_{s0})^{1/3} \tag{5}$$

where u_* is the surface friction velocity and u_{s0} is the surface Stokes drift.

For stable conditions the equation for the depth of the boundary layer is

$$\Delta\widetilde{B}\frac{\partial h_{\mathrm{bl}}}{\partial t} = \left(0.06 + 0.52\frac{h_{\mathrm{bl}}}{L_L}\right)\frac{w_{*L}^3}{h_{\mathrm{bl}}} + \overline{w'b'}_L \tag{6}$$

where $\overline{w'b'}_L$ is the buoyancy flux averaged over the depth of the boundary layer and L_L is analogous to the Obukhov length (Pearson et al., 2015), and is defined as $L_L = -w_{*L}^3/2\overline{w'b'}_L$. The definition of $\Delta\widetilde{B}$ depends on whether the depth of the boundary layer is increasing or decreasing. When h_{bl} is increasing, $\Delta\widetilde{B} = \Delta B$, and when h_{bl} is decreasing, $\Delta\widetilde{B} = w_{*L}^2/h_{\mathrm{bl}}$. The choice for $\Delta\widetilde{B}$ when h_{bl} is decreasing limits the rate at which the depth of the boundary layer can decrease.

The layer average buoyancy flux, $\overline{w'b'}_L$, is estimated by assuming that the sum of the turbulent and radiative heating rates is constant over the depth of the boundary layer (Kim, 1976), which gives

$$\overline{w'b'}_L = \frac{1}{2}\overline{w'b'}_0 + g\alpha_E \left(\langle I \rangle - \frac{1}{2}\left(I_0 + I_h\right) \right) \tag{7}$$

where α_E is the thermal expansion coefficient of sea water, $\langle I \rangle$ is the solar irradiance averaged over the depth of the boundary layer, I_0 is the solar irradiance at the surface and I_h is the solar irradiance at the base of the boundary layer.

A more complete description of the OSMOSIS scheme can be found at https://forge.ipsl.jussieu.fr/nemo/chrome/site/doc/NEMO/manual/pdf/NEMO_manual.pdf.

4. Wavelet analysis methods

To investigate variations in the spectral properties of the data, we use the wavelet analysis method of Torrence and Compo (1998). Given the number of factors which can affect mixed layer properties it was deemed important to use an analysis method which could pick out significant periodicities which are only present for a portion of the total record, because such periodicities might not be identified in power spectra of the whole time series.

The time series of observed SST and MLD were first linearly interpolated to regular 4-hourly intervals, and the output from each model was sub-sampled to the same 4-hourly intervals. (This sub-sampling does not make a significant difference to the results presented.) We chose to use 4-hourly intervals because although the gliders obtain profiles roughly every 2 hours, they are not regularly spaced in time. Due to the "V" shape of glider movement, each upcast and next downcast are separated by only a few minutes near the surface, with a wait of nearly 4 hours until the next pair. Similarly near the bottom of the profile each downcast and next upcast are closely spaced in time with a wait of nearly 4 hours until the next pair. It is only around the middle of the profiling depth that data is obtained at approximately regular 2-hourly intervals. Hence 4 hours was considered a more appropriate interpolation interval.

Because the distributions of SST and MLD are distinctly non-normal, we transform all the time series into records of percentiles (in terms of their cumulative distribution function), thus forcing the probability density functions to be rectangular (Grinsted et al., 2004). The resulting time series are padded with zeros to avoid wraparound effects and the wavelet power spectra calculated using a Morlet wavelet. Significance is determined by comparison with a theoretical red-noise spectrum calculated from the lag-1 autocorrelation coefficient for each time series. The null hypothesis is defined for the wavelet power spectrum as follows (Torrence and Compo, 1998): it is assumed that the time series has a mean power spectrum (the theoretical red-noise spectrum, given in equation 8); if a peak in the wavelet power spectrum is significantly above this background spectrum, then it can be assumed to be a true feature with a certain percent confidence.

$$P_k = \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi k/N)} \tag{8}$$

where P_k is the mean power spectrum, $k = 0, 1$ ... $N/2$ is the frequency index, α is the lag 1 autocorrelation coefficient, and N is the number of values in the time series. Wavelet spectra of the total surface heat flux and wind speed were calculated in the same way, except that the time series were not transformed into records of percentiles because the distribution of these variables was approximately normal.

To further investigate the relationships between different time series, we calculate wavelet coherence following the methods of Torrence and Webster (1999), using the code made available by Grinsted et al. (2004). Wavelet coherence can be thought of as the localized correlation coefficient in time frequency space; it shows whether non-stationary time series are co-varying at a particular frequency (but not at other frequencies) and at a particular time (but not throughout the entire record). This analysis method was chosen because simple correlations or coherence tests over the entire time series might not identify the relationships which the wavelet coherence method exposes. Significance is determined using Monte Carlo methods as detailed by Grinsted et al. (2004). Note that the annual relationship between surface forcing and mixed layer properties (cooling and deepening in autumn, warming and shoaling in spring) will not appear significant because the time series are too short. Hence the strong correlation between SST and cumulative net surface heat flux found by Damerell et al. (2016) will not be apparent because it was largely a consequence of the strong annual cycle. Multi-year time series would be required for the annual cycle to appear significant in this wavelet analysis.

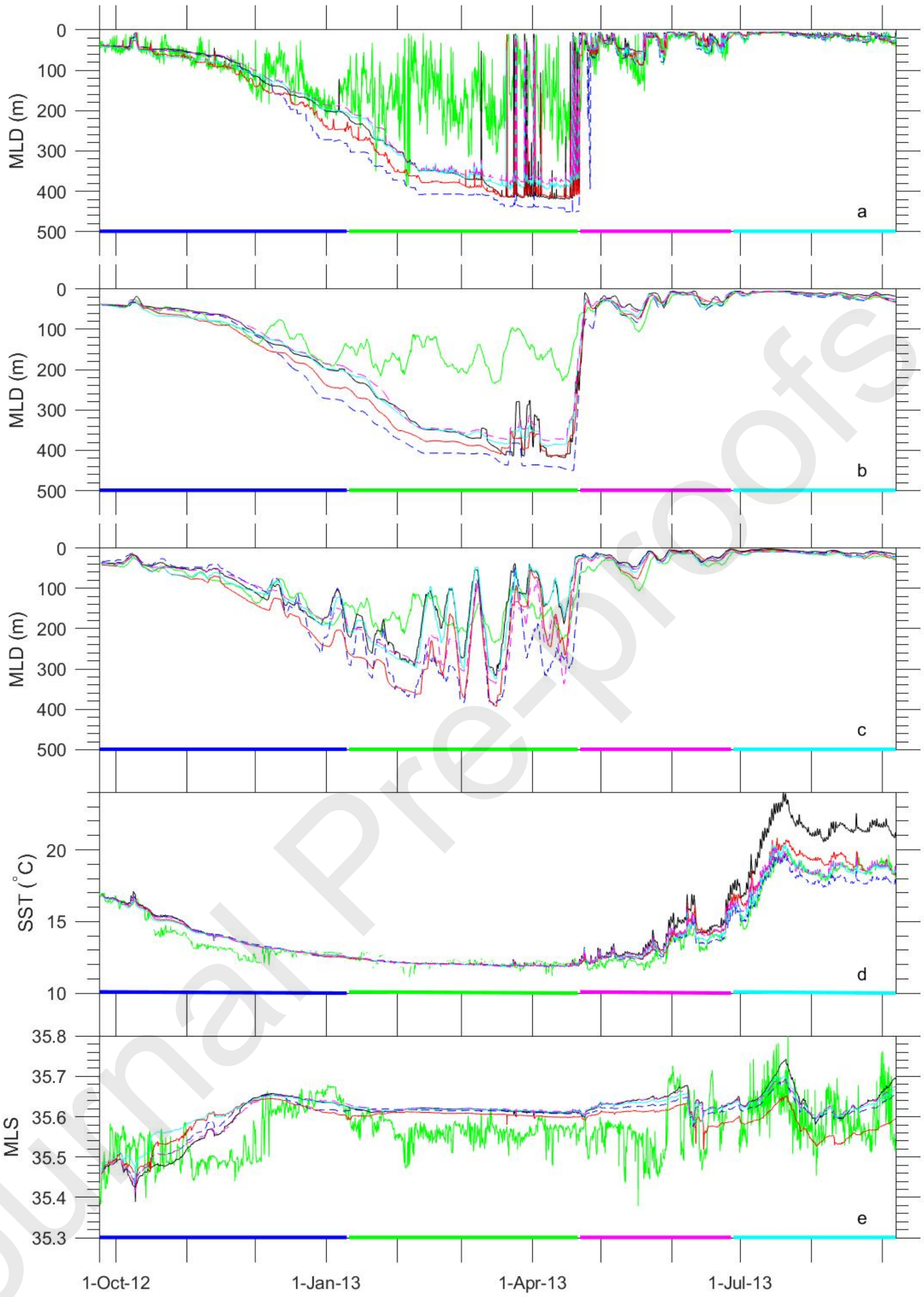


Figure 5: Mixed layer and sea surface properties over the year of the OSMOSIS field campaign, from both the glider observations and the models. a) MLD. b) MLD smoothed by applying a 5-day running mean as in section 2; this is shown for clarity only and is not used in the analysis. c) Observed MLD and models' internal MLDs, all smoothed by applying a 5-day running mean as in panel b). d) SST. e) MLS. In all panels, green line = glider observations, dashed blue line = PWP, cyan line = NEMO GLS, dashed magenta line = NEMO TKE, black line = KPP, red line = OSMOSIS model. The coloured bars at the base of the panels mark the seasons: blue = autumn; green = winter; magenta = spring; cyan = summer.

5. Results and Discussion

5.1. SST overview

All the models output SSTs which are broadly representative of the observed time series (figure 5d). The annual cycle of cooling during autumn, a fairly constant temperature over the winter, then warming to a peak in July is seen clearly in all the models. Seasonal mean biases in each model are less than 1°C (table 2), similar to the model/observation differences found by Lazarevich et al. (2004), except that KPP is considerably warmer than the observations in summer. This suggests that the drivers of SST variability in this region are largely 1-dimensional, unlike at OSP where advective effects are considered important in the winter (Large et al., 1994).

	observed	model bias				
model	SST	PWP	NEMO GLS	NEMO TKE	KPP	OSMOSIS
Autumn	13.75	0.48	0.37	0.45	0.47	0.39
Winter	12.12	-0.05	-0.07	-0.06	-0.06	-0.07
Spring	12.87	0.29	0.51	0.67	0.97	0.58
Summer	18.15	-0.40	0.22	0.30	2.91	0.74
whole year bias		0.11	0.24	0.31	0.91	0.36
rms difference		0.57	0.52	0.60	1.48	0.66

Table 2: Seasonal mean observed SST, and seasonal biases between each model and observed SST (°C). Positive bias = model SST warmer than observed SST.

The distribution of observed SST is bimodal (figure 6a) with a large peak at a temperature of 12°C. This is due to the period from early February until late May when the SST remains nearly constant at around 12°C. The average winter SST of 12.12°C (table 2) is slightly cooler than the winter SSTs of 12.14°C (2003), 12.25°C (2004) and 12.61°C (2005) found by Hartman et al. (2010) at the PAP-SO. None of the models reproduce the coldest SSTs seen in the observations, which reach a minimum of 11.1°C. GLS, TKE and KPP reach a minimum temperature of 11.8°C and PWP and OSMOSIS reach a minimum of 11.9°C. However, it is clear (figure 5d) that this is because the models show less variability in winter SSTs than the observations. The average winter SST is in fact slightly cooler in each model (table 2) than in the observations (between 0.05 and 0.07°C cooler).

The second, smaller peak of the bimodal distribution (figure 6 and figure 5d) is due to the period in late July and August when the SST again remains

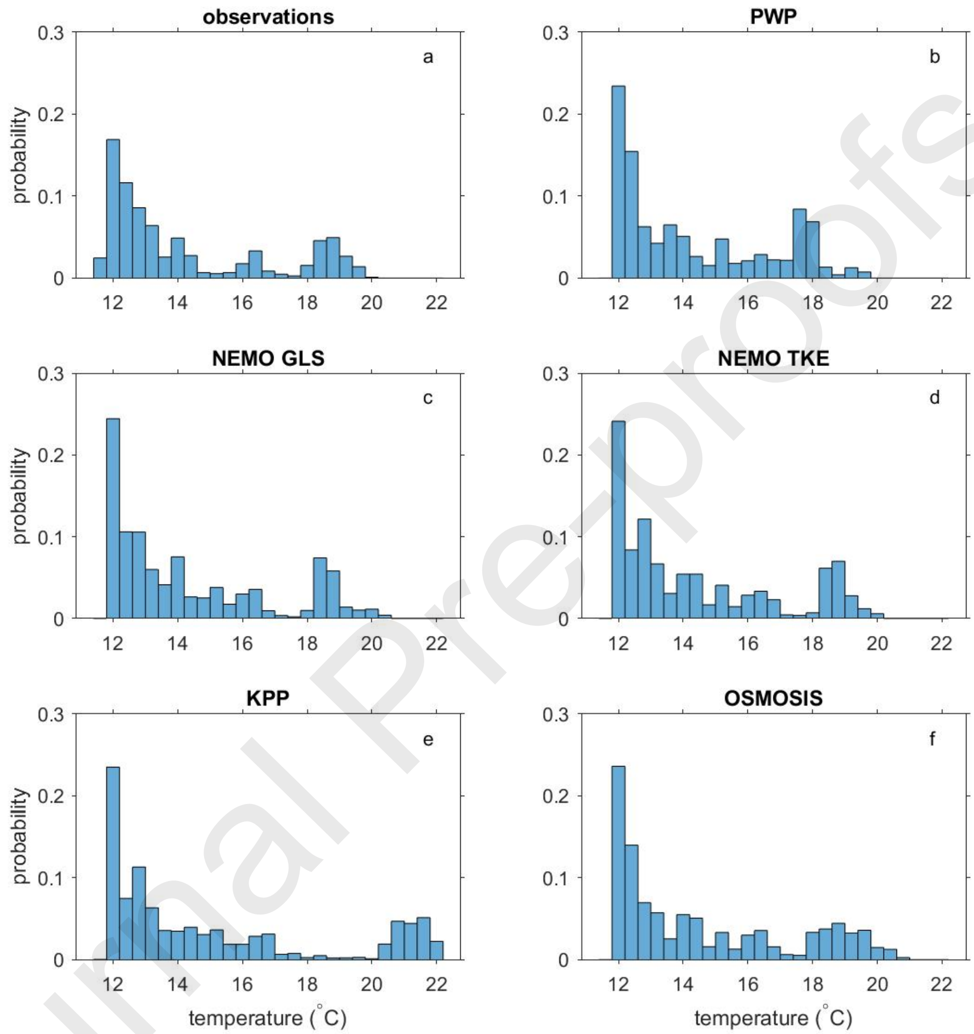


Figure 6: Histograms of SST for the observations and for each model. These are shown as probabilities, i.e., the height of the bar equals the number of counts in that bin divided by the total number of data points for that variable.

nearly constant around 18 - 19°C, consistent with the summer SSTs reported by Hartman et al. (2010). PWP, GLS, TKE and OSMOSIS have summer temperature biases between -0.40 and 0.74°C, but it is only KPP which really differs from the observations, with a mean bias in the summer of 2.9°C (table

2), similar to the summer SST bias in KPP seen by Acreman and Jeffery (2007). KPP also has the largest warm bias in spring (0.97°C). We postulate that this is related to differences in MLD/IMLD: KPP has the shallowest MLD and IMLD in the spring and summer (table 4) which will tend to trap heat in the mixed layer. TKE, in particular, has similar MLD biases to KPP (though not quite as shallow in spring and summer), but KPP's IMLD is considerably shallower than TKE's in spring and summer. Unlike TKE, where the IMLD is purely diagnostic, KPP's IMLD has an impact on model physics so may be a factor in KPP's SST bias in spring and summer.

Burchard and Bolding (2001) compared two 1D TC schemes with observations at OSP and found a shallow MLD bias in summer, which we estimate to be around 10 m from their figure 18. They attribute this to either erroneous surface fluxes or strong advective effects. However, they also comment that one model's predicted MLD is shallower than the other's, leading to warmer summer SSTs in that model. We estimate from their figure 18 that the difference in MLD is perhaps around 2 m, and the difference in SST around 0.3°C. This illustrates that during the summer when the mixed layer is shallow, relatively small differences in MLD can produce quite significant differences in SST.

PWP is unusual in exhibiting a cold bias in the summer. Archer et al. (1993) compared PWP simulations with observations at OSP over a 6-year period and also found cold biases in model summer SSTs of a similar magnitude to those seen here, as did Lazarevich et al. (2004) in their comparisons of PWP with float-observed temperatures and NCEP reanalysis SSTs in the north Atlantic. Archer et al. (1993) suggest that this may be due to small inaccuracies in the surface heat fluxes, but that seems unlikely here since the other models all have warm SST biases in summer.

5.2. Mixed layer salinity

The models do not do a very good job of reproducing the observed MLS (figure 5e and figure 7), though this is not entirely unexpected (section 1). In particular, they fail to capture the short term variability over periods of hours to days. Only some large-scale changes are captured, notably the increase in MLS in mid-July when the mixed layer is extremely shallow and high temperatures are leading to large surface evaporation (see also the large latent heat flux in July despite low wind speeds in figure 3). The distribution of the observed MLS (figure 7a) is approximately a wide Gaussian, with a mean of 35.57 and a large standard deviation of 0.06. The distributions are much narrower for all the models and are shifted towards higher salinities, with only a small tail of values at the lower end. OSMOSIS has a mean MLS of 35.59 (closest to the observations), PWP, KPP and TKE have a mean MLS of 35.61 and GLS a mean of 35.62, considerably higher than the observed mean of 35.57. OSMOSIS has the smallest bias in all seasons except the autumn (table 3). However, it is worth noting that the lower annual bias

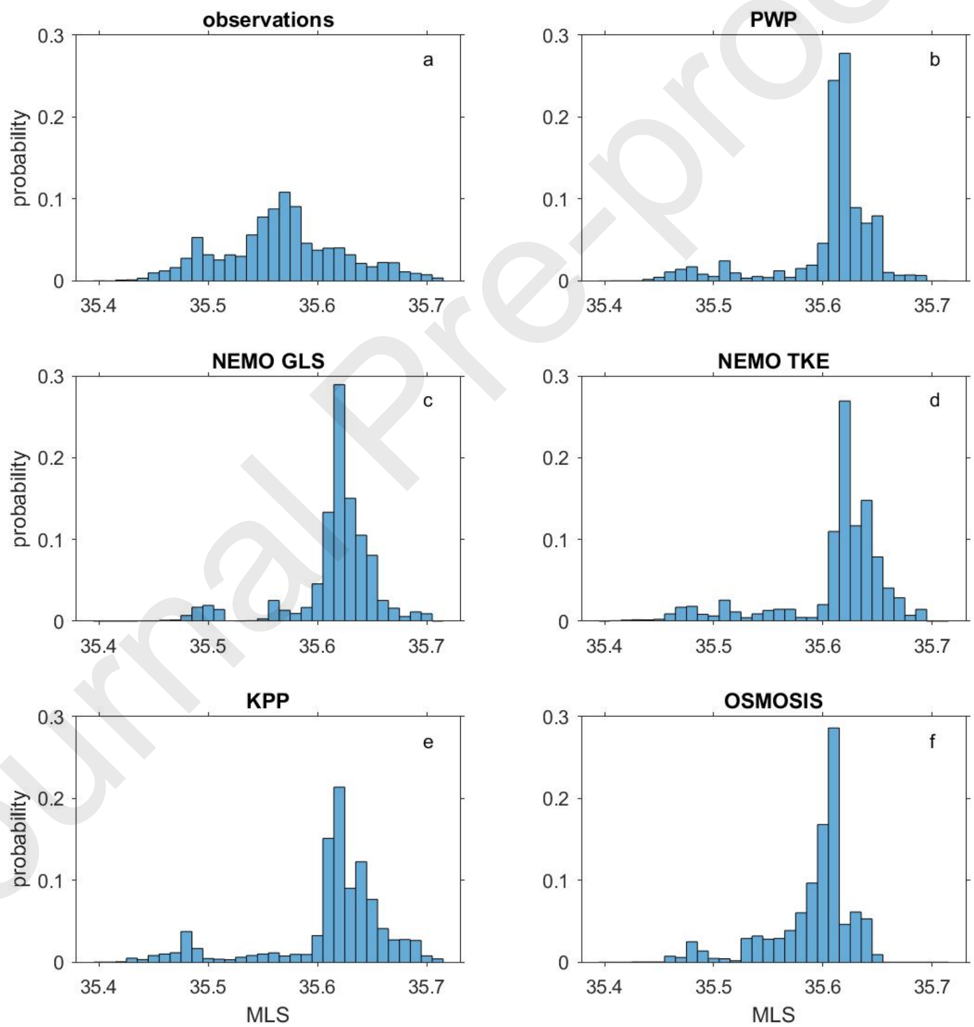


Figure 7: Histograms of MLS for the observations and for each model, shown as probabilities as in figure 6.

achieved by OSMOSIS is largely because it has both positive and negative biases which cancel out to some extent; the rms difference between OSMOSIS and the observations is only slightly smaller than for the other models. The annual average biases in MLS (table 3) of 0.02 to 0.05 represent 6-14% of the range in observed MLS over this year. When comparing the model-observation agreement of MLS and SST (e.g., figure 5), it is worth bearing in mind that the range in SST is determined by a very large scale process, i.e., the annual cycle in surface heat flux. Without a similar driver of large annual change in MLS, small variations can appear more significant than they really are. However, as will be discussed in section 5.3, the salinity biases here are large enough to produce significant density biases.

	observed			model bias		
	MLS	PWP	NEMO GLS	NEMO TKE	KPP	OSMOSIS
Autumn	35.55	0.02	0.04	0.02	0.02	0.03
Winter	35.57	0.05	0.05	0.05	0.05	0.04
Spring	35.57	0.05	0.06	0.07	0.07	0.04
Summer	35.60	0.03	0.04	0.04	0.05	-0.02
whole year bias		0.04	0.05	0.04	0.04	0.02
rms difference		0.07	0.07	0.07	0.07	0.06

Table 3: Seasonal mean observed MLS, and seasonal biases between model and observed MLS (psu). Positive bias = model MLS greater than observed MLS.

Model MLS is dependent on the surface fluxes of precipitation and latent heat (from which evaporation is calculated). These fluxes can be very localized and are difficult to measure and model consistently (e.g., difficulties in modelling cloud cover), so it is not surprising that MLS derived from reanalysis surface flux products is not very similar to observed values. Moreover the localized nature of these fluxes means MLS can vary considerably in the horizontal, leading to variability in observed MLS due to advection which is obviously not present in a 1D model.

It is worth noting, however, that local differences in MLS in this region are unlikely to have a large influence on large scale climate modelling because MLS does not directly affect the atmosphere in the same way that SST does. Biases in MLS over a wide area and long time scales might be important since these would affect water mass formation and circulation, but that is beyond the scope of this paper.

5.3. Mixed layer density

The study region is a region where Eastern North Atlantic Central Water (ENACW) forms during the winter (Pollard and Pu, 1985; Pollard et al., 1996). The slightly cooler SSTs and slightly higher MLSs (tables 2 and 3) in winter would lead to the formation of a higher density water mass than that found in the real ocean, which could have implications for the wider circulation.

We estimate equivalent density biases by calculating density for the observed average winter temperature of 12.12°C and salinity of 35.57, then subtracting(adding) the winter temperature(salinity) bias for each model and recalculating density. The winter temperature biases in table 2 lead to density biases of approximately +0.01, and the winter salinity biases in table 3 lead to density biases of approximately +0.04. The combined biases (i.e., calculating density using both the temperature and salinity biases) amount to an increase in density of approximately 0.05. Since the ENACW of subtropical origin found beneath the surface mixed layer in this region (Damerell et al., 2016) is found at σ_θ in the range 27 to 27.2 (Harvey, 1982), a density bias of 0.05 is not insignificant. However, as will be discussed in section 5.4, the wintertime density biases do not seem to impact negatively on the spring restratification and subsequent development of the MLD and SST.

5.4. MLD overview

The observed MLDs (figure 5a, b) are broadly consistent with previous observations in this area (e.g., Hartman et al., 2015; Henson et al., 2012; Martin et al., 2010; Steinhoff et al., 2010; Hartman et al., 2010), taking into account the varying MLD definitions used in different studies. Henson et al. (2012) consider differences in monthly mean MLD in years with positive or negative North Atlantic Oscillation (NAO) index in winter. They used the Hadley Centre's EN3 objectively analysed temperature and salinity data from 1959 to 2009, and calculated MLD as the depth at which a density difference of 0.03 kg m^{-3} from the surface was observed. The composite MLD for positive NAO years reached a maximum of 280 m in March, whereas in negative NAO years it reached only 170 m. They relate this to the greater wind stress in positive NAO years, resulting in increased mechanical mixing. Our results are in agreement, with an average winter mixed layer depth of 165 m (table 4) and weakly negative winter NAO index in 2013.

Winter MLD has been shown to be an important driver of nitrate flux into the surface mixed layer (Hartman et al., 2010; Steinhoff et al., 2010).

Temporary shoaling of the MLD during winter and spring may therefore influence nutrient fluxes. In this region, the winter shoaling of the MLD appears to be linked to sporadic short-lived chlorophyll blooms observed during OSMOSIS in winter, well before the main spring bloom event in June (Erickson and Thompson, 2018; Binetti et al., this issue; Rumyantseva et al., this issue). Previous studies have used data from Argo floats, XBTs, CTDs and moorings over a wide area (45°N to 52°N and 26.08°W to 8.92°W, excluding the shelf area) around the PAP-SO to estimate MLDs (Hartman et al., 2010, 2015). In all the years considered, those estimates showed MLDs increasing fairly smoothly from September to the time of maximum depth (which varied from January to March), and then decreasing again to the summer minimum. This differs from the pattern observed here where the mean MLD remained approximately constant over the winter (167, 161 and 163 m in January, February and March respectively) but with high variability. (For example, compare our figure 5b with Hartman et al. (2010) figure 4b and Hartman et al. (2015) figure 3b.) The winter time range of MLD observed by the gliders was 11 m to 378 m. This high variability in MLD is likely to be significant for nutrient fluxes and winter blooms (Hartman et al., 2010; Steinhoff et al., 2010; Erickson and Thompson, 2018; Binetti et al., this issue; Rumyantseva et al., this issue).

Model MLDs are broadly representative of the observed MLDs (figure 5a and b) except in winter when the model MLDs are too deep, with winter average biases between 160 and 228 m (table 4), and do not exhibit the same variability as the observations. This can be partially explained by the fact that in this region there is considerable submesoscale activity in winter, which

	observed	model bias				
model	MLD	PWP	NEMO GLS	NEMO TKE	KPP	OSMOSIS
Autumn	91	25(-3)	12(1)	2(-6)	7(-5)	25(23)
Winter	165	228(104)	169(16)	160(59)	173(7)	198(82)
Spring	42	-3(-21)	-11(-15)	-16(-15)	-17(-21)	-10(-15)
Summer	15	1(-5)	0(-1)	-3(-1)	-5(-7)	0(-2)
whole year bias		74(24)	51(2)	44(12)	48(-5)	64(28)
rms difference		137(106)	105(73)	102(74)	110(79)	121(90)

Table 4: Seasonal mean observed MLD, and seasonal biases between model and observed MLD (m). Figures in brackets are mean differences between each model's IMLD and the observed MLD. Positive bias = model MLD/IMLD deeper than observed MLD.

will tend to restratify the mixed layer (Thompson et al., 2016). This submesoscale activity is not present in these one dimensional models. Viglione et al. (2018) find a similar result when comparing MLDs from a 1D PWP model with observations in Drake Passage: the lack of submesoscale instabilities in the model results in MLDs which are too deep and insufficiently variable. The models' IMLDs are also deeper than the observed MLD in winter, indicating that they are likely to be deeper than the depth of active mixing in the real ocean. The winter-time difference between model IMLD and observed MLD is smallest for KPP and GLS (table 4) but this is largely because they are too deep at the start of winter and become shallower than the observed MLD towards the end of winter (figure 5c) and these differences cancel out, whereas for PWP and OSMOSIS, the winter-time IMLDs remain consistently too deep giving a greater average difference with the observations.

It is noticeable, however, that all the model IMLDs reproduce the observed wintertime shoaling and deepening of the MLD much better than the model MLDs (figure 5b, c), as well as having smaller average differences in winter (table 4). As discussed above, this temporary shoaling may be significant for fluxes of nitrates into the mixed layer in winter, so model IMLD may be more useful for understanding winter bloom dynamics than MLD calculated in the manner used for observations.

The general pattern is that in autumn and winter, model MLDs are deeper than the observed MLDs, whereas in spring and summer the model MLDs are shallower than observed MLDs. The shallow biases in spring and summer will result in a 'trapping' of surface forcing effects, i.e., the effects of the surface forcing will tend not to reach as deep as they should. This will affect the ability of these models to reproduce summer water mass formation, air-sea fluxes, and bloom dynamics through the interaction between mixed layer depth and nutrient fluxes.

All the models reproduce the observed spring restratification, though one or two days later than in the observations (table 5). One would generally expect the depth of active mixing to shoal before the mixed layer depth, and indeed each model's IMLD shoals several days earlier than that model's MLD. However, we could not find any observations in this region in the literature which indicate how much earlier one would expect the depth of active mixing to shoal than the MLD, so we are unable to comment on which model's IMLD behaves most like the real ocean.

It is noticeable that the biases in MLD are fairly small in spring and summer despite the preceding large biases in winter MLD and the winter

	observations	PWP	NEMO GLS	NEMO TKE	KPP	OSMOSIS
MLD	21	23	22	22	23	23
IMLD		21	17	19	17	19

Table 5: Date (in April 2013) of spring restratification of the MLD for the observations, and date of spring restratification of the MLD and IMLD for each model. Model dates are calculated in the same way as for the observations, as described in section 2.

mixed layer density biases (which are largely due to biases in MLS (section 5.2)). Large et al. (1994) compared KPP with observations at OSP, and also found that the spring restratification reduced biases in MLD. However, their simulation was initialised on 15th March, only about a month before the spring restratification, and the initial MLD bias was only about 15 m. Here the spring restratification removes much larger MLD biases.

The spring and summer MLD biases are not correlated with the winter MLD or MLS biases (tables 3 and 4). Similarly, the spring and summer SST biases are not correlated with the winter MLD or MLS biases (tables 2, 3 and 4). The surface forcing generating the spring restratification appears to be a sufficiently dominant process that preceding biases are unimportant. This suggests that when using a 1D model in a similar ocean environment (mid-latitudes away from topography) it may be acceptable to initialize the model using a relatively low resolution profile (e.g., from an Argo float) in late winter when stratification is low, and simply allow the model to generate the spring stratification, rather than requiring a higher resolution profile (capable of resolving a steep pycnocline) suitable for initializing during the spring or summer.

5.5. Diurnal cycles

All the models show some evidence of a diurnal cycle in SST (figure 8), significant at the 95% confidence level, starting in March and continuing to the beginning of September. The surface forcing which drives the models also shows a significant diurnal cycle in total surface heat flux from March to September (figure 9a), and all the model SSTs show evidence of a coherent relationship with the cumulative total surface heat flux at a diurnal timescale for much of the year (figure 10), though this is more obvious from mid-February onwards than in the autumn and early winter.

This diurnal cycle is not, however, as significant in the observed SSTs as in the model SSTs (figure 8), and the observations also show much less

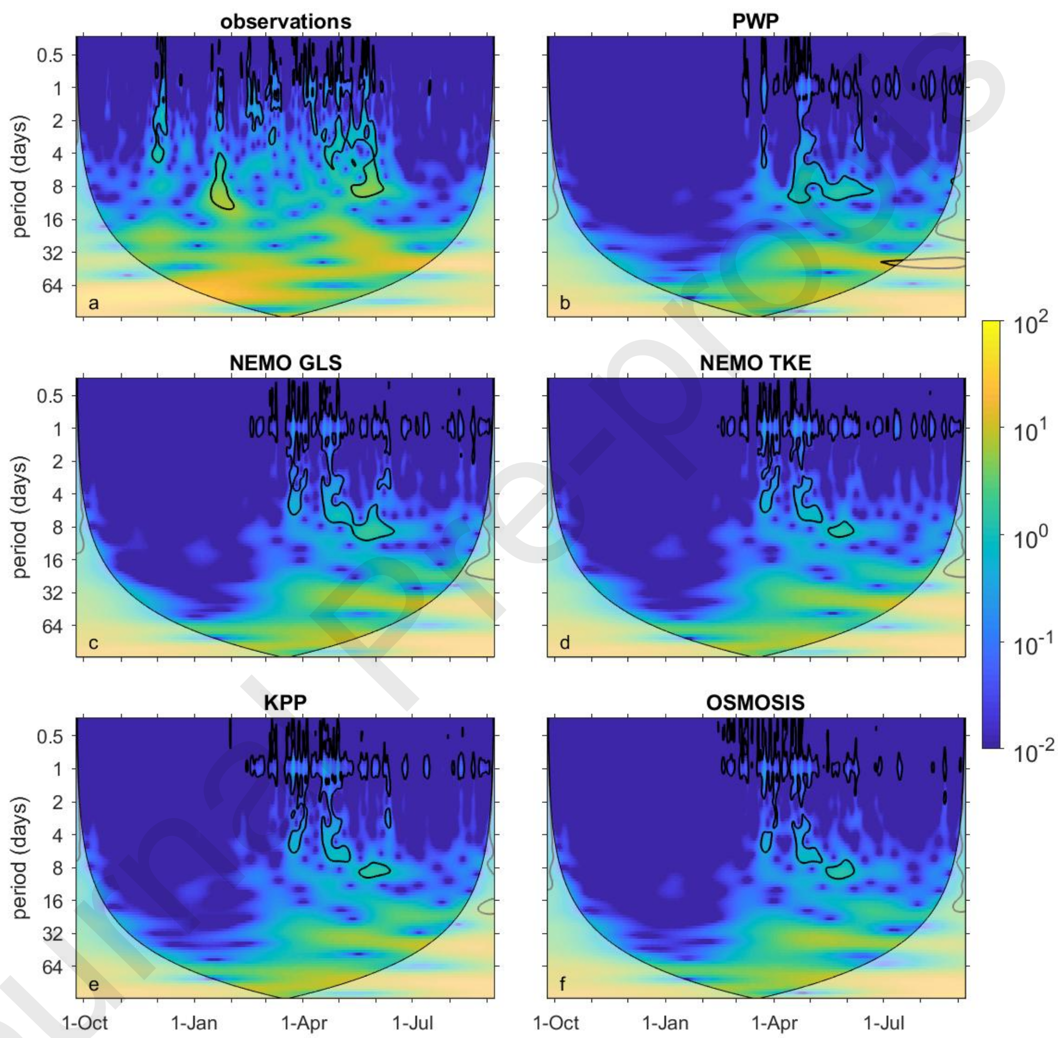


Figure 8: Wavelet spectra of SST for the observations and for each model. In each panel, the black contours enclose regions of greater than 95% confidence level calculated using the corresponding red-noise spectrum as the null hypothesis (see text). The shaded regions on either end indicate the cone of influence, where edge effects become important and results should be viewed with caution.

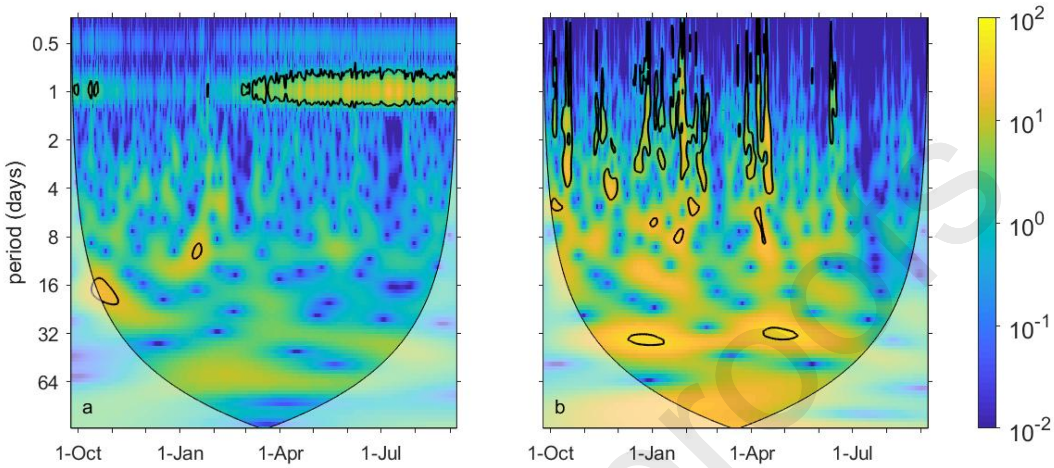


Figure 9: Wavelet spectra of the surface forcing: a) total surface heat flux, b) wind stress. Otherwise as figure 8.

coherence with the surface heat flux (figure 10). In the real ocean the diurnal cycle may be masked by noise from other ocean processes not present in the models, such as the influence of advection, internal waves and submesoscale processes, and from the fact that the glider is not sampling in one location. Biases or missing processes in the surface forcing may also lead to discrepancies between the observed and modelled SSTs. For example, Giglio et al. (2017) have recently demonstrated the significance of wind gusts in regulating how fast surface water is mixed to greater depths when daily mean winds are weak, and the reanalysis wind stress used to drive the models will not include wind gusts in a realistic fashion. Moreover, cloud cover is known to be difficult to model and this will lead to discrepancies between the reanalysis surface heat flux driving the models and the surface heat flux affecting the real ocean (Taylor , ed.; Large and Yeager, 2009). For example, reduced cloud cover during the spring and summer will tend to lead to increased heat flux into the ocean during the day, and increased heat flux out of the ocean at night. This would increase the magnitude of the diurnal cycle of SST in the models as compared with the observations. All these factors could lead to a much reduced diurnal cycle in the observations compared with the models.

As with SST, we again see a significant relationship between MLD and the cumulative surface heat flux at diurnal time scales (figure 11), though this is not as pronounced as for SST. This relationship is again considerably

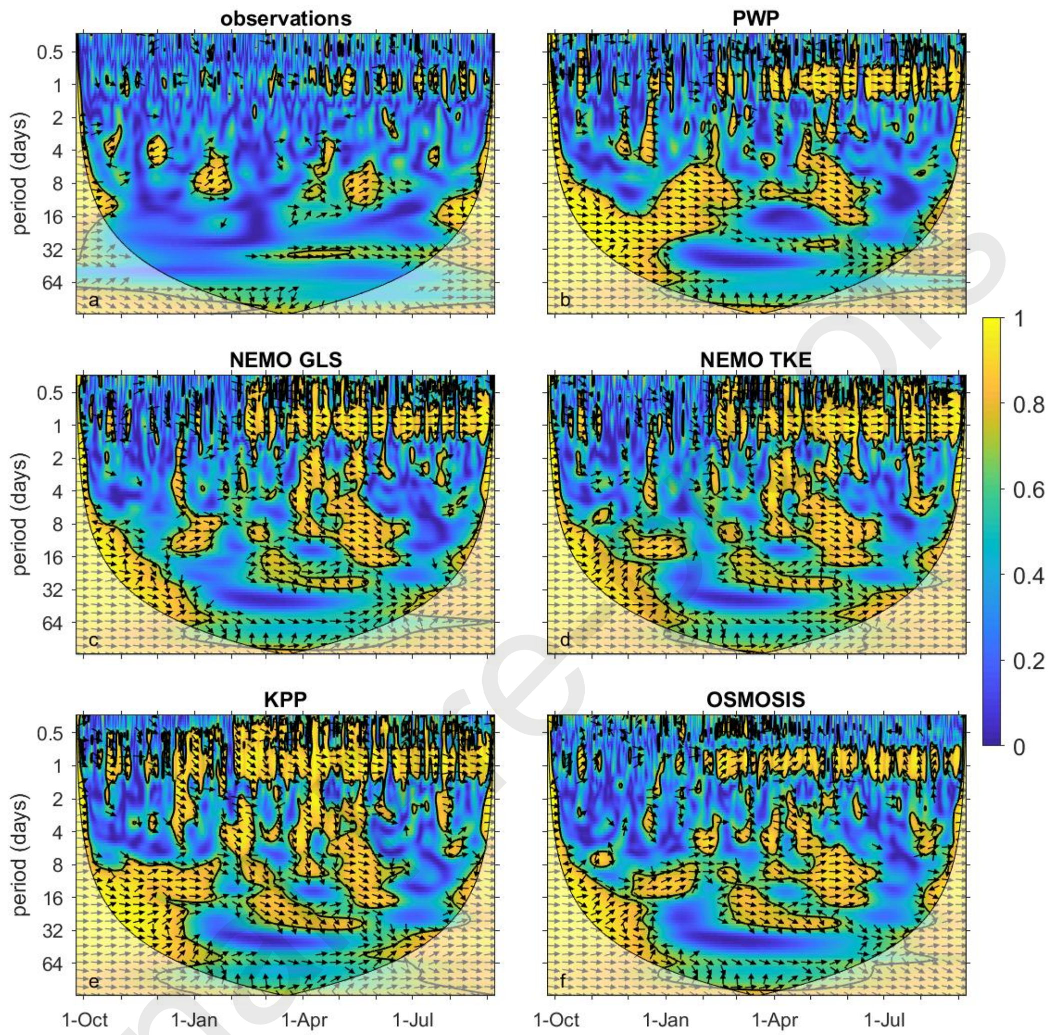


Figure 10: Wavelet coherence for SST and cumulative total surface heat flux. In each panel, the black contours enclose regions of greater than 95% confidence level calculated using Monte Carlo simulations (see text). The shaded regions on either end are as in figure 8. The arrows represent the relative phase - arrows pointing to the right imply the time series are in phase, arrows pointing left imply anti-phase, arrows pointing straight up imply the surface heat flux leads SST by a quarter of a cycle. Note that this indication of lag in all wavelet coherence figures is relative to the length of the cycle. For example, an arrow pointing up and right at an angle of 45° refers to a lag of an eight of a cycle - e.g., arrows pointing up and right at 45° in this figure mean SST lags the surface heat flux by one day for a cycle with an 8-day period but by 4 days for a cycle with a 32-day period.

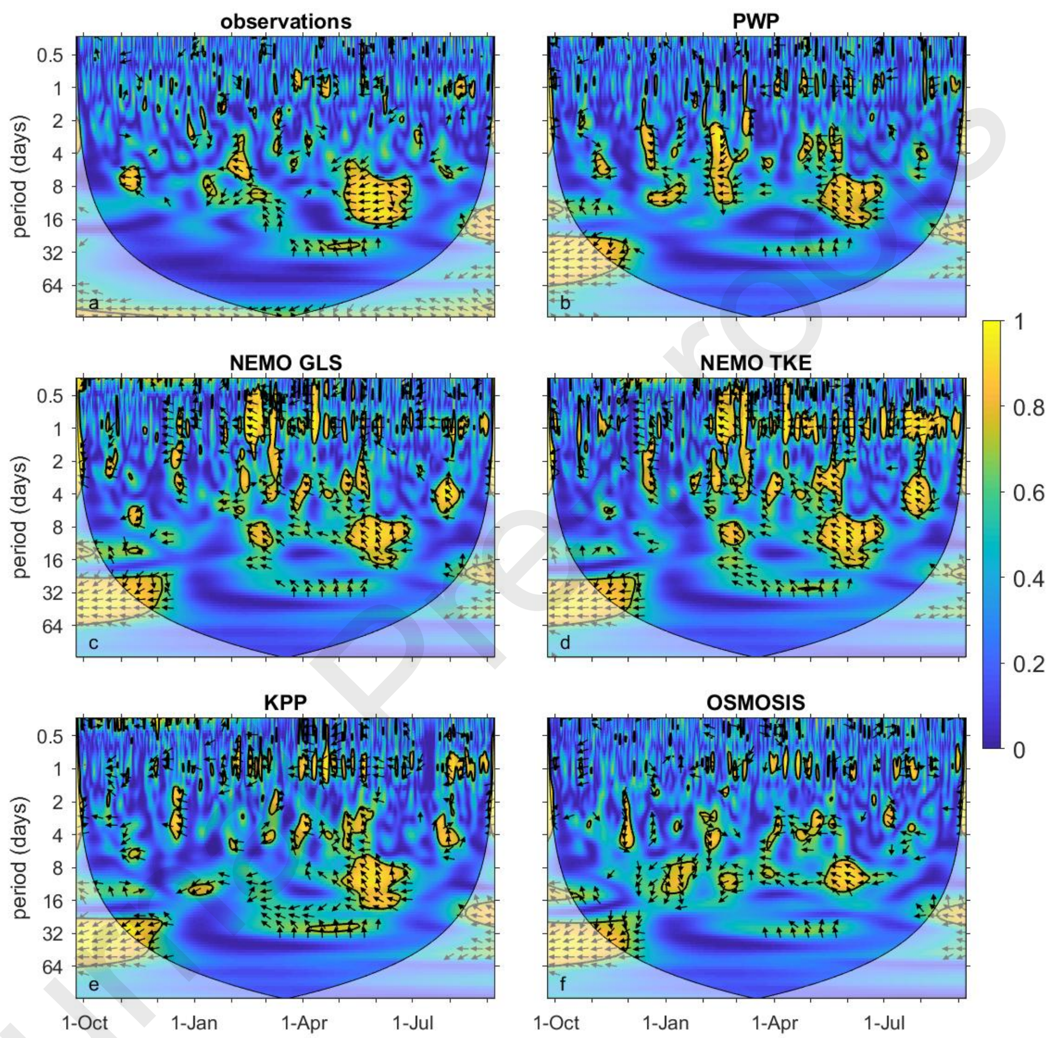


Figure 11: Wavelet coherence for MLD and cumulative surface heat flux. Otherwise as in figure 10.

more present in the models than in the observations. The MLD and surface heat flux are in approximate anti-phase, as one would expect (i.e., surface heat flux increases, MLD shoals). With solar radiation incoming during the day the SST warms and the mixed layer shoals due to thermal stratification. At night the ocean loses heat to the atmosphere, convection occurs, the SST decreases and the MLD increases. But the relationship with surface heat flux is less pronounced for MLD than for SST because the MLD is also influenced by the stratification in the profile below the mixed layer, and also is more directly affected by wind driven mixing.

5.6. Longer time scales

In May and June, at periods between approximately 4 and 20 days, the cumulative total surface heat flux is in anti-phase with the observed MLD (figure 11), and approximately in phase with the observed SST (figure 10), i.e., surface heat flux increases, MLD shoals, SST increases. These can be seen as the main warming events in SST, clearly related to large changes in MLD in the spring (figure 5). All the models exhibit similar behaviour. This timescale is typical for the passage of weather regimes. Wind stress is also a factor in these events both through the effect of wind driven mixing on the MLD and through the effect of wind speed on the latent and sensible heat fluxes.

There is clear evidence of a coherent relationship between wind stress and MLD for all models and the observations from late March onwards (figure 12) at periods between 4 and 60 days. MLD and wind stress are approximately in phase (i.e., wind stress increases, mixed layer deepens), though with the MLD lagging the wind stress by around an eighth of a cycle. This highlights the significance of local wind events in the spring, which can temporarily deepen the mixed layer. During the year observed, such spring deepening events reached as much as 100 m which is likely to be significant for spring bloom dynamics (Erickson and Thompson, 2018). No such relationship with local wind events is seen earlier in the year, despite the generally stronger wind stress in autumn and winter than spring and summer. We hypothesize that the deepening of the mixed layer seen in the autumn is so strongly driven by the annual cycle in surface heat flux that the additional effect of the winds at this time of year is less significant.

There is also some evidence of a coherent relationship between SST in the models and wind stress (figure 13), from March onwards. This is a lagged anti-phase relationship, i.e., as wind stress decreases, SST increases but with a lag of approximately an eighth of a cycle or less. This is due to the shoaling of the mixed layer as wind stress decreases: a shallower mixed layer will mean the effect of the surface heat flux will be concentrated in a thinner band of water, and in the spring the surface heat flux will tend to warm the ocean. Hence SST increases as the wind stress decreases. The relationship between wind stress and observed SST is much more tenuous than with the model SSTs, due to the processes in the real ocean and atmosphere not present in

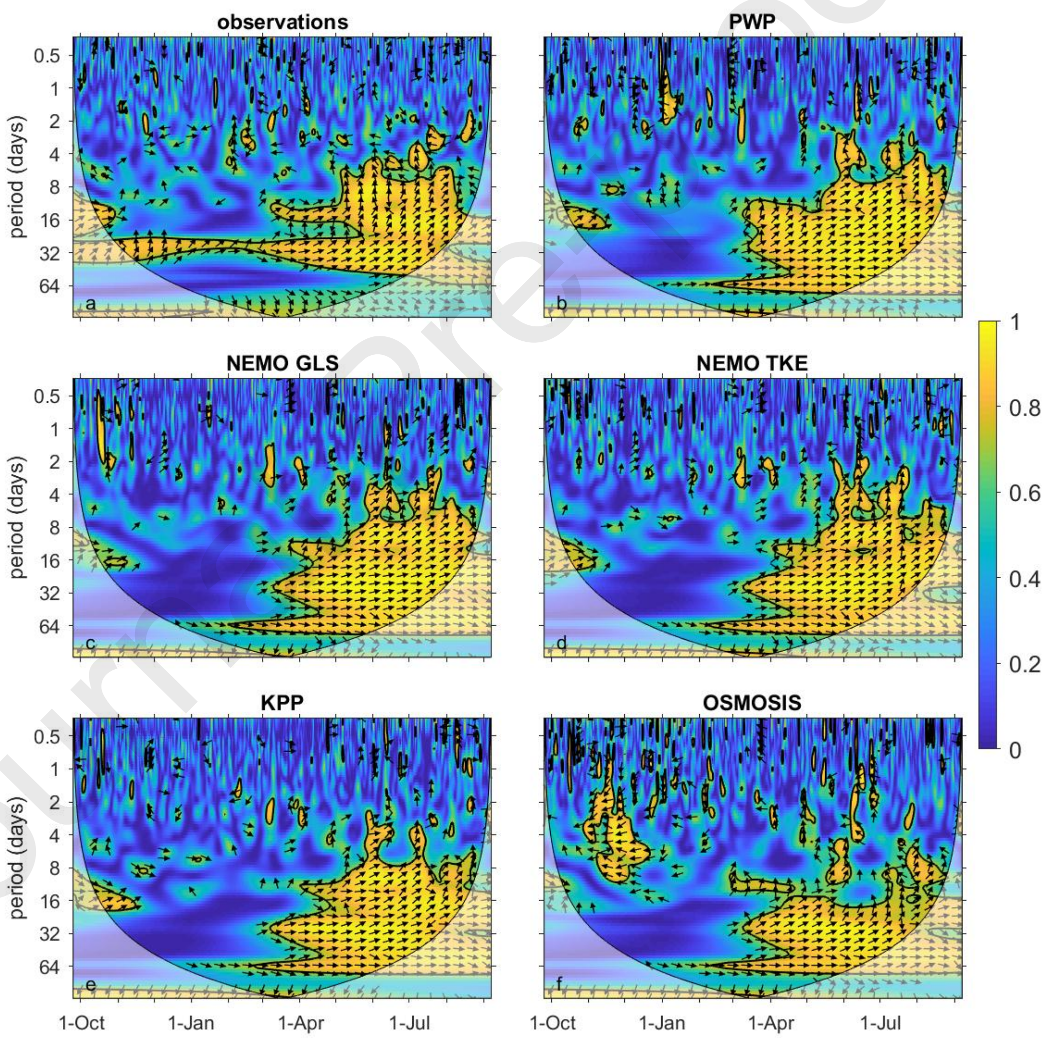


Figure 12: Wavelet coherence for MLD and wind stress. Otherwise as in figure 10.

the models nor in the reanalysis surface forcing.

6. Conclusions

Five mixed layer models driven by ERA-Interim surface forcing have been compared with a year of observations in the North Atlantic. All the mod-

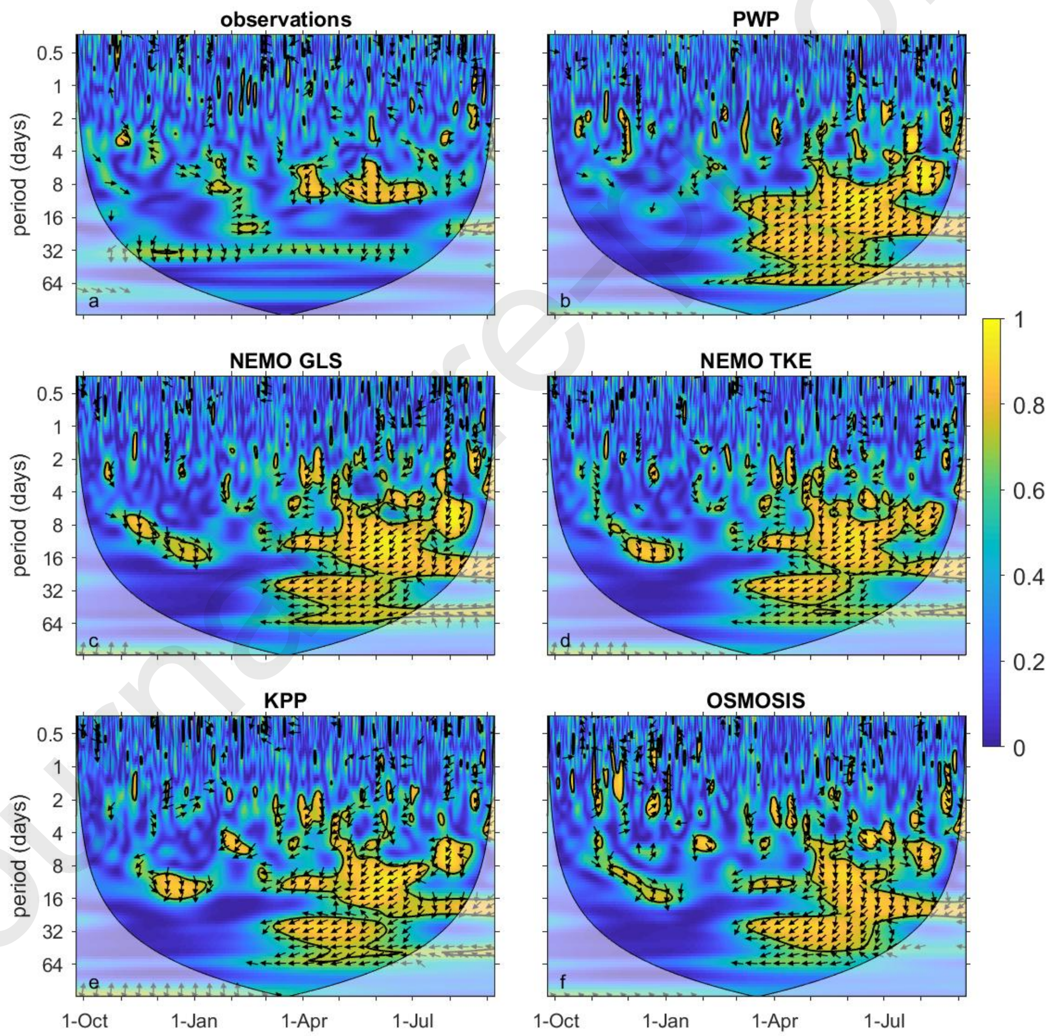


Figure 13: Wavelet coherence for SST and wind stress. Otherwise as in figure 10.

els reproduce SST fairly well in terms of the annual cycle, except that the KPP model has summer SSTs which are approximately 3°C warmer than the observations. Short timescale variability in SST is not predicted well by the models, likely due to the many sources of variability in SST not present in a 1-D model. The models do not reproduce the observed MLS well, but this is not unexpected as advection is expected to play a role in MLS in this region, and because precipitation biases are not uncommon in reanalysis surface forcing data. The biases are large enough to produce a non-trivial density bias. In particular, the slightly cooler temperatures and higher salinities in the winter in all models would lead to the formation of ENACW of greater density than that in the real ocean, which could have related effects on ocean circulation. However, this does not seem to affect the subsequent spring restratification and evolution of the MLD and SST.

Both the wind stress and surface heat flux are involved in driving periods of temporary deepening and shoaling of the MLD through the spring, though the effects of wind stress are felt throughout spring and summer whereas the surface heat flux is only a factor in May and June. Wind stress is not related to MLD during the autumn despite the high wind stresses in autumn. We hypothesize that the deepening of the MLD in autumn is so strongly driven by the annual cycle in surface heat flux that the winds are less significant in the autumn.

The surface heat flux also drives a diurnal cycle in MLD and SST from March onwards, though this effect is much clearer in the models than in the observations. We believe this is because the models and reanalysis forcing data do not include a number of processes which complicate the observed SST and MLD, so the diurnal cycle is less apparent in the observations.

We are not able to say that one model is 'better' than the others, they all have strengths and weaknesses. PWP has the lowest bias in spring MLD, second lowest in summer MLD, but it has the largest biases in autumn and winter MLD. Similarly it has the lowest biases in winter and spring SST, but fairly large SST biases in autumn and summer. KPP's IMLD has by far the smallest deep bias in winter, but KPP also has by far the largest bias in SST. TKE has the smallest annual mean bias in MLD but the second largest bias in spring SST. GLS has the second smallest annual mean bias and smallest rms difference in SST, but the largest bias in annual mean MLS and largest rms difference for MLS. OSMOSIS has the smallest bias in annual mean MLS, but the second largest bias in annual mean MLD and SST.

It is noticeable that all models had low biases in MLD in spring and summer despite the MLS and MLD biases in the preceding winter. This suggests that initializing these models using a relatively low resolution profile (e.g., from an Argo float) in late winter when stratification is low may give a quite reasonable spring stratification, which could be useful in regions where higher resolution profiles capable of resolving a steep pycnocline are not available. The variability in winter time MLD, which may be of significance for nutrient fluxes and winter bloom dynamics, is reproduced much better by model IMLDs than model MLDs.

Given the lack of differences between them, any of these models would give similar results when used for modelling in seasonal areas similar to the OSMOSIS site, i.e., at mid latitudes away from topography.

7. Acknowledgments

We thank the scientists, technicians, officers and crew of *RRS Discovery* cruise D381, *RV Celtic Explorer* cruise CE13001 and *RRS James Cook* cruises JC085, JC087 and JC090, especially those who went out of their way to accommodate our glider deployments. We thank the many people involved in ship-based and glider data collection, processing, sampling and calibration. The observational data used in this study were the result of five research cruises and a full year of glider piloting and involved many contributions not represented in the author list. This work was supported by NERC grants NE/I019905/1, NE/I020083/1 and NSF award OCE 1155676. GMD and KJH were supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 741120). We thank Louise Biddle and Nicholas Klingaman for useful discussions on running PWP and KPP respectively.

The observational data are held at the British Oceanographic Data Centre and can be accessed at https://doi.org/10/cqc6. The ERA-Interim data were obtained from the European Centre for Medium-Range Weather Forecasts (downloaded from http://www.ecmwf.int/en/research/climatereanalysis/era-interim in April 2016). North Atlantic Oscillation index data were provided by the Climate Analysis Section, National Centre for Atmospheric Research, Boulder, Colorado, downloaded from http://climatedataguide.ucar.edu/guidance/hurrellnorth-atlantic-oscillation-nao-index-pc-based in October 2014.

8. References

Abdella, K., McFarlane, N., 1997. A new second-order turbulence closure scheme for the Planetary boundary layer. J. Atmos. Sci. 54, 1850–1867. doi:10.1175/1520-0469(1997)054<1850:ANSOTC>2.0.CO;2.

Acreman, D.M., Jeffery, C.D., 2007. The use of Argo for validation and tuning of mixed layer models. Ocean Modell. 19, 53–69. doi:10.1016/j.ocemod.2007.06.005.

Archer, D., Emerson, S., Powell, T., Wong, C.S., 1993. Numerical hindcasting of sea-surface pCO2 at Weathership Station Papa. Prog. Oceanogr. 32, 319–351. doi:10.1016/0079-6611(93)90019-A.

Belcher, S.E., Grant, A.L.M., Hanley, K.E., Fox-Kemper, B., Van Roekel, L., Sullivan, P.P., Large, W.G., Brown, A., Hines, A., Calvert, D., Rutgersson, A., Pettersson, H., Bidlot, J.R., Janssen, P.A.E.M., Polton, J.A., 2012. A global perspective on Langmuir turbulence in the ocean surface boundary layer. Geophys. Res. Lett. 39, L18605. doi:10.1029/2012GL052932.

Biddle, L.C., Heywood, K.J., Kaiser, J., Jenkins, A., 2017. Glacial meltwater identification in the Amundsen Sea. J. Phys. Oceanogr. 47, 933–954. doi:10.1175/JPO-D-16-0221.1.

Binetti, U., Kaiser, J., Damerell, G.M., Rumyantseva, A., Martin, A.P., Henson, S., Heywood, K.J., this issue. Net community oxygen production derived from Seaglider deployments at the Porcupine Abyssal Plain site (PAP; northeast Atlantic) in 2012-13. Prog. Oceanogr. .

Blanke, B., Delecluse, P., 1993. Variability of the tropical Atlantic Ocean simulated by a general-circulation model with two different mixed-layer physics. J. Phys. Oceanogr. 23, 1363–1388. doi:10.1175/1520-0485(1993)023<1363:VOTTAO>2.0.CO;2.

de Boyer Montegut, 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.

Burchard, H., Bolding, K., 2001. Comparative analysis of four second-moment turbulence closure models for the oceanic mixed layer. J. Phys. Oceanogr. 31, 1943–1968. doi:10.1175/1520-0485(2001)031¡1943:CAOFSM¿2.0.CO;2.

Burchard, H., Craig, P.D., Gemmrich, J.R., van Haren, H., Mathieu, P.P., Meier, H.M., Smith, W.A.M.N., Prandke, H., Rippeth, T.P., Skyllingstad, E.D., et al., 2008. Observational and numerical modeling methods for quantifying coastal ocean turbulence and mixing. Prog. Oceanogr. 76, 399–442.

Chen, D., Rothstein, L.M., Busalacchi, A.J., 1994. A hybrid vertical mixing scheme and its application to tropical ocean models. J. Phys. Oceanogr. 24, 2156–2179. doi:10.1175/1520-0485(1994)024<2156:AHVMSA>2.0.CO;2.

Damerell, G.M., Heywood, K.J., Thompson, A.F., Binetti, U., Kaiser, J., 2016. The vertical structure of upper ocean variability at the Porcupine Abyssal Plain during 2012-2013. J. Geophys. Res. 121, 3075–3089. doi:10.1002/2015JC011423.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J..J., Park, B..K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J..N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. Roy. Meteor. Soc. 137, 553–597. doi:10.1002/qj.828.

Erickson, Z.K., Thompson, A.F., 2018. The seasonality of physically-driven export in the northeast Atlantic Ocean. Global Biogeochem. Cycles 32, 1144–1162. doi:10.1029/2018GB005927.

Farahat, A., Abuelgasim, A., 2019. Role of atmospheric nutrient pollution in stimulating phytoplankton growth in small area and shallow depth water bodies: Arabian Gulf and the sea of Oman. Atmos. Environ. , 117045doi:10.1016/j.atmosenv.2019.117045.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Rummukainen, M., 2013. Evaluation of Climate Models. pp. 741–866. doi:10.1017/CBO9781107415324.

Frants, M., Gille, S.T., Hatta, M., Hiscock, W.T., Kahru, M., Measures, C.I., Mitchell, B.G., Zhou, M., 2013. Analysis of horizontal and vertical processes contributing to natural iron supply in the mixed layer in southern Drake Passage. Deep-Sea Res 90, 68–76. doi:10.1016/j.dsr2.2012.06.001.

Gaspar, P., Grégoris, Y., Lefevre, J.M., 1990. A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: Tests at station Papa and long-term upper ocean study site. J. Geophys. Res. 95, 16179–16193.

Giglio, D., Gille, S.T., Subramanian, A.C., Nguyen, S., 2017. The role of wind gusts in upper ocean diurnal variability. J. Geophys. Res. 122, 7751–7764. doi:10.1002/2017JC012794.

Grant, A.L.M., Belcher, S.E., 2009. Characteristics of Langmuir turbulence in the ocean mixed layer. J. Phys. Oceanogr. 39, 1871–1887. doi:10.1175/2009JPO4119.1.

Grinsted, A., Moore, J.C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes Geophys. 11, 561–566. doi:10.5194/npg-11-561-2004.

Harcourt, R.R., 2013. A second-moment closure model of Langmuir turbulence. J. Phys. Oceanogr. 43, 673–697. doi:10.1175/JPO-D-12-0105.1.

Harcourt, R.R., 2015. An improved second-moment closure model of Langmuir turbulence. J. Phys. Oceanogr. 45, 84–103. doi:10.1175/JPO-D-14-0046.1.

Hartman, S.E., Jiang, Z.P., Turk, D., Lampitt, R.S., Frigstad, H., Ostle, C., Schuster, U., 2015. Biogeochemical variations at the Porcupine Abyssal Plain sustained Observatory in the northeast Atlantic Ocean, from weekly to inter-annual timescales. Biogeosciences 12, 845–853. doi:10.5194/bg-12-845-2015.

Hartman, S.E., Lampitt, R.S., Larkin, K.E., Pagnani, M., Campbell, J., Gkritzalis, T., Jiang, Z.P., Pebody, C.A., Ruhl, H.A., Gooday, A.J., Bett, B.J., Billett, D.S.M., Provost, P., McLachlan, R., Turton, J.D., Lankester, S., 2012. The Porcupine Abyssal Plain fixed-point sustained observatory (PAP-SO): variations and trends from the Northeast Atlantic fixed-point time-series. ICES Journal of Marine Science 69, 776–783. doi:10.1093/icesjms/fss077.

Hartman, S.E., Larkin, K.E., Lampitt, R.S., Lankhorst, M., Hydes, D.J., 2010. Seasonal and inter-annual biogeochemical variations in the Porcupine Abyssal Plain 2003-2005 associated with winter mixing and surface circulation. Deep-Sea Res. 57, 1303–1312. doi:10.1016/j.dsr2.2010.01.007.

Harvey, J., 1982. Theta-S relationships and water masses in the eastern North Atlantic. Deep-Sea Res. 29, 1021–1033. doi:10.1016/0198-0149(82)90025-5.

Henson, S., Lampitt, R., Johns, D., 2012. Variability in phytoplankton community structure in response to the North Atlantic Oscillation and implications for organic carbon flux. Limnol. Oceanogr. 57, 1591–1601. doi:10.4319/lo.2012.57.6.1591.

Hirons, L.C., Klingaman, N.P., Woolnough, S.J., 2015. MetUM-GOML1: a near-globally coupled atmosphere ocean-mixed-layer model. Geosci. Model Dev. 8, 363–379. doi:10.5194/gmd-8-363-2015.

Hirons, L.C., Klingaman, N.P., Woolnough, S.J., 2018. The impact of air-sea interactions on the representation of tropical precipitation extremes. J. Adv. Model. Earth Syst. 10, 550–559. doi:10.1002/2017MS001252.

Holtslag, A.A.M., Moeng, C.H., 1991. Eddy diffusivity and countergradient transport in the convective atmospheric boundary layer. J. Atmos. Sci. 48, 1690–1698. doi:10.1175/1520-0469(1991)048<1690:EDACTI>2.0.CO;2.

Kantha, L.H., Clayson, C.A., 1994. An improved mixed-layer model for geophysical applications. J. Geophys. Res. 99, 25235–25266. doi:10.1029/94JC02257.

Kim, J.W., 1976. A generalized bulk model of the oceanic mixed layer. J. Phys. Oceanogr. 6, 686–695. doi:10.1175/1520-0485(1976)006<0686:AGBMOT>2.0.CO;2.

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. doi:10.1111/j.2153-3490.1967.tb01462.x.

Lampitt, R.S., Billett, D.S.M., Martin, A.P., 2010a. The sustained observatory over the Porcupine Abyssal Plain (PAP): Insights from time series observations and process studies. Deep-Sea Res. 57, 1267–1271. doi:10.1016/j.dsr2.2010.01.003.

Lampitt, R.S., Salter, I., de Cuevas, B.A., Hartman, S., Larkin, K.E., Pebody, C.A., 2010b. Long-term variability of downward particle flux in the deep northeast Atlantic: causes and trends. Deep-Sea Res. 57, 1346–1361. doi:10.1016/j.dsr2.2010.01.011.

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. doi:10.1029/94RG01872.

Large, W.G., Yeager, S.G., 2009. The global climatology of an interannually varying air-sea flux data set. Climate Dyn. 33, 341–364. doi:10.1007/s00382-008-0441-3.

Lazarevich, P., Rossby, T., McNeil, C., 2004. Oxygen variability in the near-surface waters of the northern North Atlantic: observations and a model. J. Mar. Res. 62, 663–683. doi:10.1357/0022240042387547.

Lazarevich, P., Stoermer, S., 2001. A Matlab version of the Price, Weller, Pinkel model. URL: http://www.po.gso.uri.edu/rafos/research/pwp/.

Lee, J.C.K., Klingaman, N.P., 2018. The effect of the quasi-biennial oscillation on the Madden-Julian oscillation in the Met Office Unified Model global ocean mixed layer configuration. Atmos. Sci. Lett. 19, UNSP e816. doi:10.1002/asl.816.

Li, Q., Fox-Kemper, B., 2017. Assessing the effects of Langmuir turbulence on the entrainment buoyancy flux in the ocean surface boundary layer. J. Phys. Oceanogr. 47, 2863–2886. doi:10.1175/JPO-D-17-0085.1.

Li, Q., Webb, A., Fox-Kemper, B., Craig, A., Danabasoglu, G., Large, W.G., Vertenstein, M., 2016. Langmuir mixing effects on global climate: WAVEWATCH III in CESM. Ocean Modell. 103, 145–160. doi:10.1016/j.ocemod.2015.07.020.

Madec, G., 2008. NEMO ocean engine: Note du Ple de modlisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619.

Martin, A.P., Lucas, M.I., Painter, S.C., Pidcock, R., Prandke, H., Prandke, H., Stinchcombe, M.C., 2010. The supply of nutrients due to vertical turbulent mixing: A study at the Porcupine Abyssal Plain study site in the northeast Atlantic. Deep-Sea Res. 57, 1293–1302. doi:10.1016/j.dsr2.2010.01.006.

Martz, T.R., Johnson, K.S., Riser, S.C., 2008. Ocean metabolism observed with oxygen sensors on profiling floats in the South Pacific. Limnol. Oceanogr. 53, 2094–2111. doi:10.4319/lo.2008.53.5_part_2.2094.

McWilliams, J.C., Sullivan, P.P., Moeng, C.H., 1997. Langmuir turbulence in the ocean. J. Fluid Mech. 334, 1–30.

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

Paulson, C.A., Simpson, J.J., 1977. Irradiance measurements in the upper ocean. J. Phys. Oceanogr. 7, 952–956. doi:10.1175/1520-0485(1977)007<0952:IMITUO>2.0.CO;2.

Pearson, B.C., Grant, A.L.M., Polton, J.A., Belcher, S.E., 2015. Langmuir turbulence and surface heating in the ocean surface boundary layer. J. Phys. Oceanogr. 45, 2897–2911. doi:10.1175/JPO-D-15-0018.1.

Pollard, R.T., Griffiths, M.J., Cunningham, S.A., Read, J.F., Perez, F.F., Rios, A.F., 1996. Vivaldi 1991-A study of the formation, circulation and ventilation of Eastern North Atlantic Central Water. Prog Oceanogr 37. doi:10.1016/S0079-6611(96)00008-0.

Pollard, R.T., Pu, S., 1985. Structure and circulation of the Upper Atlantic Ocean northeast of the Azores. Prog Oceanogr 14, 443–462. doi:10.1016/0079-6611(85)90022-9.

Pookkandy, B., Dommenget, D., Klingaman, N., Wales, S., Chung, C., Frauen, C., Wolff, H., 2016. The role of local atmospheric forcing on the modulation of the ocean mixed layer depth in reanalyses and a coupled single column ocean model. Climate Dyn. 47, 2991–3010. doi:10.1007/s00382-016-3009-7.

Price, J.F., Mooers, C.N.K., Vanleer, J.C., 1978. Observation and simulation of storm-induced mixed-layer deepening. J. Phys. Oceanogr. 8, 582–599. doi:10.1175/1520-0485(1978)008<0582:OASOSI>2.0.CO;2.

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. doi:10.1029/JC091iC07p08411.

Reffray, G., Bourdalle-Badie, R., Calone, C., 2015. Modelling turbulent vertical mixing sensitivity using a 1-D version of NEMO. Geosci. Model Dev. 8, 69–86.

Rodgers, K.B., Aumont, O., Fletcher, S.M., Plancherel, Y., Bopp, L., Montégut, C.D.B., Iudicone, D., Keeling, R., Madec, G., Wanninkhof, R., 2014. Strong sensitivity of Southern Ocean carbon uptake and nutrient cycling to wind stirring. Biogeosciences 11, 4077–4098.

Rodi, W., 1987. Examples of calculation methods for flow and mixing in stratified fluids. J. Geophys. Res. 92, 5305–5328.

Rumyantseva, A., S., H., Martin, A.P., Thompson, A.F., Damerell, G.M., Kaiser, J., Heywood, K.J., this issue. Phytoplankton spring bloom initiation: The impact of atmospheric forcing and light in the temperate North Atlantic Ocean. Prog. Oceanogr. .

Sharples, J., Ross, O.N., Scott, B.E., Greenstreet, S.P.R., Fraser, H., 2006. Inter-annual variability in the timing of stratification and the spring bloom in the North-western North Sea. Cont. Shelf Res. 26, 733–751. doi:10.1016/j.csr.2006.01.011.

Sheehan, P.M.F., Berx, B., Gallego, A., Hall, R.A., Heywood, K.J., Hughes, S.L., Queste, B.Y., 2018. Shelf sea tidal currents and mixing fronts determined from ocean glider observations. Ocean Sci. 14, 225–236. doi:10.5194/os-14-225-2018.

Simonot, J.Y., Le Treut, H., 1986. A climatological field of mean optical properties of the World Ocean. J. Geophys. Res. 91, 6642–6646. doi:10.1029/JC091iC05p06642.

Steinhoff, T., Friedrich, T., Hartman, S.E., Oschlies, A., Wallace, D.W.R., Kortzinger, A., 2010. Estimating mixed layer nitrate in the North Atlantic Ocean. Biogeosciences 7, 795–807. doi:10.5194/bg-7-795-2010.

Stips, A., 2011. Fitting measured irradiance of Jerlov water types to double exponential functions using R.

Taylor (ed.), P., 2000. Final report of the joint WCRP/SCOR Working Group on airsea fluxes: intercomparison and validation of ocean-atmosphere energy flux fields, WCRP-112, WMO/TD-No.1036, World Climate Research Programme, 303 pp.

Thompson, A.F., Lazar, A., Buckingham, C., Naveira Garabato, A.C., Damerell, G.M., Heywood, K.J., 2016. Open-ocean submesoscale motions: A full seasonal cycle of mixed layer instabilities from gliders. J. Phys. Oceanogr. 46, 1285–1307. doi:10.1175/JPO-D-15-0170.1.

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Amer. Meteor. Soc. 79, 61–78.

Torrence, C., Webster, P., 1999. Interdecadal changes in the ENSO-monsoon system. J. Climate 12, 2679–2690.

Umlauf, L., Burchard, H., 2003. A generic length-scale equation for geophysical turbulence models. J. Mar. Res. 61, 235–265.

Van Roekel, L., Adcroft, A.J., Danabasoglu, G., Griffies, S.M., Kauffman, B., Large, W., Levy, M., Reichl, B.G., Ringler, T., Schmidt, M., 2018. The KPP boundary layer scheme for the ocean: revisiting its formulation and benchmarking one-dimensional simulations relative to LES. J. Adv. Model. Earth Syst. 10, 2647–2685. doi:10.1029/2018MS001336.

Viglione, G.A., Thompson, A.F., Flexas, M.M., Sprintall, J., Swart, S., 2018. Abrupt transitions in submesoscale structure in southern Drake Passage: glider observations and model results. J. Phys. Oceanogr. 48, 2011–2027. doi:10.1175/JPO-D-17-0192.1.

Wilcox, D.C., 1988. Reassessment of the scale-determining equation for advanced turbulence models. AIAA journal 26, 1299–1310.

Yang, B., Emerson, S.R., Bushinsky, S.M., 2017. Annual net community production in the subtropical Pacific Ocean from in situ oxygen measurements on profiling floats. Global Biogeochem. Cycles 31, 728–744. doi:10.1002/2016GB005545.

Highlights

Unique, year-long, high resolution glider dataset compared with 5 mixed layer models.

Model winter mixed layers are too deep, with average biases between 160 and 228 m.

After spring restratification, biases in MLD are small and unrelated to winter biases.

Model biases in mixed layer salinity produce non-trivial density biases, but this does not affect the subsequent spring and summer MLD and SST.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: