Luneva Maria V. (Orcid ID: 0000-0003-3240-5427)

Wakelin Sarah L (Orcid ID: 0000-0002-2081-2693)

Holt Jason T (Orcid ID: 0000-0002-3298-8477)

Inall Mark E (Orcid ID: 0000-0002-1624-4275)

Kozlov Igor E (Orcid ID: 0000-0001-6378-8956)

Palmer Matthew R. (Orcid ID: 0000-0003-3227-042X)

Toberman Matthew (Orcid ID: 0000-0003-1688-1853)

Polton Jeff A. (Orcid ID: 0000-0003-0131-5250)

Challenging vertical turbulence mixing schemes in a tidally energetic environment: Part I. 3D shelf-sea model assessment.

Maria V. Luneva[(1)], Sarah Wakelin[(1)] , Jason T. Holt[(1)], Mark E. Inall [(2)], Igor E. Kozlov[(3,4)] , Matthew R. Palmer[(1)] , Matthew Toberman[(2)], Evgenia V. Zubkova[(3)], Jeff A. Polton[(1)].

(1) National Oceanography Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool, United Kingdom, L3 5DA
(2) Scottish Association for Marine Science.
(3) Satellite Oceanography Laboratory, Russian State Hydrometeorological University, Saint-Petersburg, 195196, Russia.
(4) Marine Hydrophysical Institute of RAS, Sevastopol, Russia

Key points:
Turbulence closures are assessed in 3D simulations in tidally energetic environment;
In the shelf seas, all closures simulate a seasonal pycnocline that is too shallow, bottom temperatures that are too warm, there is no clear winner out of the choice of closures.
Choice of turbulence closure affects the ecosystem behaviour, shifting the timing of peak chlorophyll by one month, and regional chlorophyll differences of order 100%.

.

Abstract:

Mixing in the ocean and shelf seas is critical for the vertical distribution of dynamically active properties, such as density and biogeochemical tracers. Eight different decadal simulations are used to assess the skill of vertical Turbulent Mixing Schemes (TMS) in a 3D regional model of tidally active shelf seas. The TMS differ in the type of stability functions used and in the Ozmidov/ Deardorff /Galperin limiter of the turbulence length scales. We review the dependence of the critical Richardson and Prandtl numbers to define the "diffusiveness" of the TMS.

The skill in representing bias and variability of stratification profiles is assessed with 5 different metrics: surface and bottom temperatures; pycnocline depth, thickness and strength. The assessment is made against hydrography from three datasets (28,000 profiles in total).

Bottom and surface temperatures are found to be as sensitive to TMS choice as to horizontal resolution or heat flux formulation, as reported in other studies. All TMS under-represent the pycnocline depth and benthic temperatures. This suggests physical processes are missing from the model, and these are discussed.

Different TMSs show the best results for different metrics, and there is no outright winner. Simulations coupled with an ecosystem model show the choice of TMS strongly affects the ecosystem behaviour: shifting the timing of peak chlorophyll by one month, showing regional chlorophyll differences of order 100%, and redistributing the production of chorophyll between the pycnocline and mixed layer.

1. Introduction

Seasonal stratification is a ubiquitous feature of the upper ocean. Stable stratification typically results from increased summer heating and reduced wind stress and convection, which suppresses mixing as well as the exchange of heat, momentum and nutrients between the surface layer and deeper ocean. Tidally energetic seas alternate between well-mixed and stratified states, both seasonally and across tidal mixing fronts. Strong summer stratification

isolates the upper layer from the tidally driven benthic boundary layer, affecting ocean–atmosphere exchange (e.g. of carbon and oxygen) and biological activity (Rippeth et al. 2005, 2009, 2014). Indeed, cross-pycnocline mixing becomes the main mechanism to resupply nutrients to the euphotic layer and supports continued summertime primary production. Hence accurately modelling mixing in seasonally stratified seas is critical to modelling ecosystem processes and understanding the environmental impacts of future change (Holt et al., 2016).

Shelf seas modelling requires simulation of the dynamics of the benthic tidally driven layer, the upper, mostly wind and buoyancy driven layer, and the pycnocline. Small scale turbulence and diapycnal mixing in shelf seas vary in time and space, with time scales of tidal cycle and spring neap modulations. Short term (~1-3hours) mixing spikes in the thermocline associated with the alignment of the wind with inertially rotating surface currents or tidal shear have been observed, with temporally averaged nutrient fluxes increasing by between 4 and 20 times during the shear bursts ([Burchard and Rippeth, 2009; Lenn et al., 2011; Williams et al., 2013). Strong tides produce areas mixed from the surface to bottom, forming tidal fronts and associated jet currents. To simulate time and spatially dependent profiles of vertical diffusion/viscosity many of the widely used regional models (for example: Carniel et al., 2009, O'Dea et al., 2017) include a generic length scale "GLS", (Umlauf and Burchard, 2005) module. The GLS module (conveniently deployed as a stand-alone model in the General Ocean Turbulence Model, GOTM, at *gotm.net*) proposes the choice of turbulence model with two prognostic equations along with various closures. The prognostic equations describe the evolution of turbulent kinetic energy (TKE) and turbulent length scale (or their combination) with time.

The closures describe the dependence of diffusivity/viscosity on velocity shear and stratification and are central to the modelling assumptions made in the mixing scheme, particularly under stratified conditions. There has been significant recent progress in closure development (e.g. Canuto et al, (2010)), which now include parameterisations of double diffusion, Langmuir circulation (Harcourt 2015, Kantha and Clayson, 2004) and wave breaking (Gerbi et al, 2009). All these closures are derived from the assumption of equilibrium and locality, which assumes a balance between production, dissipation and transfer of energy between components of Reynolds stresses and heat/salt fluxes, while time derivative terms and turbulent transports are neglected. Existing non-local models (e.g.

Cheng et al., (2005)) are still too complex and to our knowledge are not included in any current ocean circulation model.

Most turbulent closures have been verified through the simulation of mixed layer depth (MLD) and sea surface temperature (SST), with little systematic assessment of their impact on the whole vertical structure of stratification and pycnocline properties in 3D simulations.

This paper aims to fill this gap and assess widely used closures in the context of shelf sea applications. Using the GLS routines in the Nucleus for European Modelling of the Oceans (NEMO) system (Madec and the NEMO team, 2008), comparisons are made between simulated and observed characteristics of the stratification. Seasonally stratified shelf seas often have an approximately two water layer structure. We characterise stratification profiles by five characteristics: sea surface temperature, benthic temperature, depth and thickness of the pycnocline, and the strength of the pycnocline, expressed by the maximum Brunt-Väisäla frequency. To assess model properties in multi-dimentional space of parameters, we employ Taylor (2000) diagrams and also, we introduce a "model skill" measure combining errors in variability and biases of all characteristics in one dimensionless parameter (Holloway et al., 2011).

The pycnocline is a critical biogeochemical interface, and so we also evaluate the effects of closure type on simulated ecosystem properties, using the Earth and Regional Seas Ecosystem Model (ERSEM), (Butenschon et al, 2016) directly coupled to these simulations. To our knowledge, assessment of vertical (diapycnal) mixing schemes with a 3D dynamically active ocean model coupled with ecosystem model has not previously been documented.

The paper is organized as follows: in section 2 we describe the turbulence closures under consideration; section 3 describes the model set up, validation data and methods of comparisons; section 4 contains an inter-comparison between model runs and comparison with hydrography on decadal and shorter time scales; in section 5 we discuss the implication of the results and the effects of turbulence closures on an ecosystem model simulation. Conclusions are presented in section 6, and section 7 describes code availability.

2. Review of vertical turbulence closures.

2.1 General approach. Turbulent coefficients of vertical diffusion of heat and salt k_h, k_S and viscosity k_{mu}, k_{mv} are defined as:

$$\begin{pmatrix} k_{mu} \\ k_{mv} \\ k_h \\ k_s \end{pmatrix} = - \begin{pmatrix} \overline{u'w'}/\partial_z U \\ \overline{v'w'}/\partial_z V \\ \overline{T'w'}/\partial_z T \\ \overline{S'w'}/\partial_z S \end{pmatrix} \quad , \tag{1}$$

where the second order turbulent moments $\overline{u'w'}, \overline{v'w'}, \overline{T'w'}, \overline{S'w'}$ are Reynolds stresses in the east and northwards direction and vertical turbulent heat and salt fluxes, respectively. The partial derivative with respect to depth is denoted by ∂_z. Variables *U, V, T, S* are ensemble means over all sizes of turbulent eddies of velocity components, temperature and salinity. Correspondingly, *u', v', w', T', S'* are fluctuating, turbulent components, overlining in the righthand side of (1) denotes averaging over an ensemble of turbulent eddies. In isotropic fully developed turbulence these coefficients ($k_{mu,mv,T,S}$) are the function of two turbulent scalar characteristics with independent dimensions. Without losing generality, we have chosen turbulent velocity scales, $q = (\overline{u'^2} + \overline{v'^2} + \overline{w'^2})^{1/2}$, and turbulent length scale, l, as the governing parameters. From scaling arguments, following Kolmogorov (1942), one gets:

$$k_{mu,mv,T,S} = q l \; A_{u,v,T,S} \tag{2}$$

where $A_{u,v,T,S}$ are constants in this particular case. More generally, when the turbulence is anisotropic (perhaps due to shear, rotation and/or stratification) $A_{u,v,T,S}$ are not constants and are the so-called stability functions. These stability functions then become dependent on dimensionless parameters that characterise the sources of anisotropy of the flow, such as velocity squared shear ($\Sigma^2 = \partial_z U^2 + \partial_z V^2$), stratification, (defined by $N^2 = g\rho_0{}^{-1} \partial_z \rho$) , or frequency of rotation Ω. Here ρ is density, ρ_0 is the reference density and g is gravitational acceleration. For these particular sources of anisotropy, dimensionless parameters, turbulence scale factors, are

$$G_m = \tau^2 \Sigma^2, G_h = \tau^2 N^2 , G_\Omega = \tau\Omega \tag{3}$$

where $\tau = l/q$ is a turbulent timescale. The stability functions are then

$$A_{u,v,T,S} = A_{u,v,T,S}(G_{m,h,\Omega\ldots}) \tag{4}$$

The most important dimensionless characteristics of turbulent flows are the Richardson number, *Ri*, the flux Richardson number, *Rf*, and the Prandtl number, *Pr*. These are defined as follows:

$$Ri = \frac{N^2}{\Sigma^2} = \frac{G_h}{G_m}; \quad Rf = \frac{g\overline{\rho' w'}}{\overline{u'w'}\partial_z U + \overline{v'w'}\partial_z V} = \frac{Ri}{Pr}; \quad Pr = A_{u/} A_{T.} \tag{5}$$

Linear stability theory (Miles, 1961, Howard 1961) for a flow with constant steady shear predicts the necessary, but not sufficient condition for stability to be:

$$Ri > Ri_{cr} = 1/4 \tag{6}$$

Nonlinear analysis, considered by Abarbanel et al., (1984), provides necessary and sufficient conditions for stability to be:

$$Ri > Ri_{cr} = 1 \tag{7}$$

However, when internal waves are present, the stability condition of plane waves becomes (Orlanski and Brien, 1969):

$$Ri > Ri_{cr} = 1 + |k_x/k_z| > 1 \tag{8}$$

where k_z, k_x are vertical and horizontal components of the wave number. Hence, the stability condition with internal waves present (8) is at least four times weaker than the condition induced by constant shear (6).

To close the problem (2)-(4), one must define the two dimensional independent turbulent scales and find the associated stability functions (4). The hierarchy of turbulent models, based on ideas by Kolmogorov (1942), Prandtl and Wieghart (1945) and Rotta (1951), relies on the principle of separation of variables into turbulent and ensemble mean parts, and further derivation of the set of equations of higher order turbulent moments. This approach was first developed in the late 1960s and early 1970s when two-equation models became a practical tool for calculation of turbulent flows in engineering applications. Launder and Jones (1969) proposed a $(k - \varepsilon)$ model which solves prognostic equations for turbulent kinetic energy $k = q^2/2$ and its dissipation rate ε (units of m^2s^{-3}). Instead of $(k - \varepsilon)$ equations, Mellor and Yamada, (1974, 1982) employed prognostic equations for q^2 , $q^2 l$. The final number of equations depends on the order of anisotropy needed to be achieved: as an example, Cheng et al. (2005), simulating non-local convection properties, stop the chain of the equations at the fourth order moments, finally solving the system of 36 equations. Most geophysical applications, however, employ no more than the 2.5 level closure models (according to the

Mellor and Yamada, (1974) classification). These use two prognostic equations to find the turbulent length scale and the turbulent kinetic energy, and then use simplified algebraic equations for second order moments to describe the anisotropy of the flow and determine the stability functions. Different closures at level 2.5 use from 7 (e.g. Galperin et al., 1988) to 10 simplified algebraic equations (e.g. Canuto et al, 2001). Many of these models have been combined in the General Ocean Turbulence Model ((Burchard and Bolding, 2001, Burchard, et al., 2008, Umlauf and Burchard, 2005)) and are implemented in NEMO as a 'GLS' – general length scale module, which allows employment of prognostic equations for any combination ($q^2, q^n l^m$) models.

The GLS module in NEMO has a variety of 'closures' – i.e. different ways to define the stability functions in terms of the key turbulence scale factors (3): dimensionless shear, G_m, and stratification, G_h. G_Ω is not invoked as Coriolis effects are neglected in these closures. All these closures are 'local', i.e. based on a set of algebraic equations for turbulent buoyancy fluxes and Reynolds stresses in which time derivative and turbulent transports are neglected. The resulting number of stability functions is reduced to two: one for momentum (effecting turbulent viscosity) and one for heat and salt (turbulent vertical diffusivity) such that:

$$A_m = A_u = A_v, \quad A_h = A_S = A_T \tag{9}$$

2.2 Turbulent closures of 2.5 level in GLS module.

In this study we explore the following closures, present in NEMO:

GKHR: The (Galperin et al., 1988) closure is very similar to Mellor and Yamada (1982). That work considered equilibrium regimes of turbulence, where production of turbulent kinetic energy balances dissipation rate. They found upper and lower limits for G_h for various flow regimes: In the convective regime $G_h < 0 \; and \; G_h > -0.029.$, with the upper limit being a corollary to the necessity for length scales to remain finite. In stably stratified flow $G_h > 0$, and a constraint comes from the Ozmidov length scale limit (the size beyond which eddies are unable to overturn) (see section 2.3 below).

KC94: The Kantha and Clayson, (1994) closure is an extension of the GKHR and Mellor and Yamada, (1982) models. Following Andren and Moeng, (1993), they added a shear term and a term proportional to the scalar variance of the pressure-temperature correlation term. The critical flux Richardson number in the equilibrium case, if a limiter is not applied, becomes

$Rf_c = 0.213$ and critical gradient Richardson number $Ri_c = 0.25$, in accordance with the linear stability criteria (6).

K3: the (Kantha, 2003) closure is the same as KC94, but with adjusted constants that allow for the existence of turbulence at $Ri > 0.25$. The critical Richardson number in this closure in a steady state regime, if limiters are not applied, correspond well to the CHCD closure, discussed below.

CHCD: The Canuto et al. (2001) and Cheng et al. (2002) closures add anisotropic shear production (Σ^4 terms), buoyancy, salinity production and the vorticity contribution to pressure-velocity, pressure-density and pressure-salinity correlations. This model produces turbulent regimes for Richardson numbers Ri~1 (as demonstrated with non-linear analysis by Abarbanel et al., (1984)). The critical Richardson numbers are $Ri_{cr} = 0.85, Ri_{cr} = 1.03$ for different sets of constants A and B (see Appendix A and Fig 1a) respectively.

2.3 Properties of the turbulence mixing schemes (TMS) as applied to regional simulations. Effects of Deardorff /Galperin limiter.

All closures considered in this study allow the existence of turbulence at gradient Richardson number, Ri>1, in unsteady, decaying stages of turbulence or where the vertical diffusion of turbulent kinetic energy is important in the balance. Lateral transport of turbulent properties has been neglected throughout this work on the basis that the tidal excursion is at best marginally resolved. In the equilibrium state, where production due to shear balances the dissipation rate and work done opposing stable stratification, the gradient Richardson number depends on the type of stability function and the imposed Deardorff, -Galperin limiter (Kantha, 2003, Chen et al., 2002, 2003), which comes from the Ozmidov length scale limit: $L_o = c_l(\varepsilon N^{-3})^{1/2}$: the size beyond which eddies are able to overturn, so

$$Nlq^{-1} \leq c_l \qquad (10a),$$

The range of c_l according different observational estimates is (0.2 : 0.6) (Deardorff, 1973, Lewellen and Teske, 1973, Galperin et al., 1988). It defines the maximal threshold for stability factor G_h in the wide range of

$$G_h \leq (c_l)^2 \approx [0.039 - 0.36], \qquad (10b)$$

and the critical Richardson number (Appendix 1) in steady state becomes dependant on c_l and type of closure (Fig.1a).

When the Richardson number exceeds the steady state limit, turbulence decays and one should be left with background diffusivity produced by unresolved subgrid processes such as internal waves, near-inertial oscillations, etc. In the NEMO GLS code the background turbulent kinetic energy q^2_{bgr} is prescribed. The turbulent length scale is then set to the upper limit of (10a) i.e. $l = c_l q N^{-1}$ and the background diffusivity becomes inversely proportional to the buoyancy frequency (Holt and Umlauf, 2008), which is consistent with the Gargett and Holloway, (1984) parameterisation of internal waves:

$$k_T = \frac{c_l A_h \, q_{bgr}}{N}, \qquad k_m = \frac{c_l A_m \, q_{bgr}}{N} \ , \tag{11a}$$

and the dissipation rate of TKE is proportional to N and given by

$$\varepsilon = 0.5 \breve{c}_0 q^2_{bgr} N / c_l \tag{11b}$$

where $\breve{c}_0$=0.554[3] for GKHR/KC94/K3 and $\breve{c}_0$= 0.527[3] (CHCD).

The multipliers in (11a), $c_l A_h \, , c_l A_m$ depend on the type of closure and limiter (Fig1b-c) and thus the "background" diffusivity is a closure-dependent. Multipliers grow with closure in the order GKHR→ KC94→KC3→CHCD(A)→CHCD(B). The Prandtl number Pr at $Ri \to Ri_{cr}$ (Fig 10d) grows with limiter c_l for all closures and $Pr > 1$ for Kantha (2003) and CHCD closures (see CHCD Fig 3) at the observed range of $c_l \in [0.2, 0.6]$. GKHR predicts $Pr < 1$ in the full range of c_l and KC94 at ($c_l < 0.527$) (see KC94 Fig1) assuming more effective exchange of heat compared with exchange of momentum at the conditions of strong stratification. The latter is in agreement with a Large Eddy Simulation of a stably stratified atmospheric boundary layer (Zhou and Chow, 2011) and data from the Surface Heat Budget of the Arctic Ocean (SHEBA) field program in the Arctic (Sorbjan and Grachev, 2010).

Cheng et al., (2003) pointed out that length scale limiters are required in regimes of weak stratification (Ri < 0.25), rather than at high Ri. This is supported by large eddy simulations and planetary boundary layer observations. Indeed Holt and Umlauf, (2008) demonstrated, that in tidally active seas, when turbulence is not in equilibrium, a Deardorff / Galperin limiter is required in conditions of weak stratification in order to prevent a strong increase of turbulent length scale near regions of shear-driven convection. They show that in the absence of this limiter, the CHCD (A) closure resulted in a three-fold reduction of stratification in the North-West European shelf seas, in conflict with observation.

2.4 Prognostic equations. Following Carniel et al., (2009), we use the $(k - \varepsilon)$ model with constants consistent with each closure applied (see Umlauf and Burhard, 2005). Turbulent

kinetic energy $k = 0.5q^2$ and dissipation rate ε are highly variable both in the upper boundary layer due to diurnal variability, and in the bottom boundary layer due to tidal friction effects (Simpson et al. 1996). This results in strong temporal variability, non-equilibrium regimes of turbulence both in the boundary layers and the pycnocline. The existence of turbulence at Richardson numbers far above steady state critical values is found in observations (Palmer at al, 2013) and models, even in the absence of (or unresolved) internal waves in the pycnocline. Time dependent turbulence scale factors affect the turbulent diffusivity and viscosity, and feedback on the temporal evolution of k and ε .

Further, for simplicity we will call the combination of prognostic equations, stability function (closure), length scale limiter application and background mixing parameterisation as the Turbulence Mixing Scheme (TMS)

3. Methods

3.1 Numerical model and experiments

We use the North-West European shelf NEMO configuration with ~7km resolution (AMM7; Fig 1) and 51 vertical hybrid terrain-following / geopotential levels, compressed towards the bottom and surface. This is a shelf sea NEMO V3.6 (Madec and NEMO Team, 2016), currently used for operational forecasts/hindcasts by the UK Met Office (and known as CO6). Most of the model details are similar to (O'Dea et al., 2017), which is the V3.4 version, and the validation of that version versus observations and a finer resolution model can be found in (Graham et al., 2018). The model explicitly resolves tides, driven by geopotential forcing, including 13 tidal constituents, present in NEMO code and detected in observations (Davis and Kwong, 2000). Tidal lateral boundary conditions are defined by the (Flather, 1976) algorithm and use sea surface elevations and velocity for 9 harmonics from a tidal model of the North Atlantic (Flather, 1981). Temperature and salinity lateral boundary conditions are prescribed by the Global Seasonal Forecast system version 5 (GLOSEA5) (MacLachlan et al., 2015) using the nominally 1/4 degree global ORCA NEMO model. For surface boundary conditions, we use the CORE formulation (Large and Yeager, 2009) with ERA-interim atmospheric reanalysis variables (Dee et al., 2011). Other details, and the sensitivity to forcing and physics options, can be found in O'Dea et al., (2017). The present model does not use data assimilation.

The standard run of V3.6 corresponding to ST conditions in this study, was initialised in January 1980 by interpolating temperature and salinity fields from the 1/4° ORCA025 hindcast of the standard global ocean configuration GO5.0 (Megann et al., 2014). Most other simulations considered here start from the ST run at January 1996.

This study considers a series of perturbation experiments, which are defined in Table 1. The difference between the runs arise only from the choice of stability functions, the Galperin limiter (c_l) and the way of prescribing surface boundary conditions for TKE and dissipation rate. Five runs have been performed for 1996-2010 (GA, KC, K3, CAA, CAB).

Simulation results have been shown to be very sensitive to the background turbulent kinetic energy (Costa et al., 2017). We use

$$q_{bgr}^2 = 2 \times 10^{-6} m^2 s^{-2} \,, \qquad (12a)$$

which is consistent with the minimal observed level of turbulence at the conditions of strong stratification (Monin and Ozmidov, 1985).

In conditions of very strong stratification, the limiting condition for diffusivity/viscosity is imposed

$$k_{T,m} = \max\left(k_{T,m},\ 10^{-6} m^2 s^{-1}\right) \qquad (12b)$$

The parameters of the "standard" run can be found in the supporting materials (namelist_cfg). Condition "ln_clim=.true." activates background diffusivity option (10a). We performed two additional runs (CAL,KCL), where condition (10a) is not activated. The Galperin-Deandroff limiter is applied only to $G_h = \min(G_h, c_l^2)$, affecting the stability functions, but not turbulence length scale itself. In this case, an additional condition for background dissipation rate is required to define background diffusivity:

$$\varepsilon = \max(\varepsilon, \varepsilon_{bgr}) \qquad k_T = \frac{\widetilde{c_{0l}} A_T (q_{bgr})^4}{4\varepsilon} \,,\ k_m = \frac{\widetilde{c_{0l}} A_m (q_{bgr})^4}{4\varepsilon} \qquad (12c)$$

, the NEMO default value of c_l=0.52 (GKHR) have been employed here. Stability functions A_T , A_m are in the range of (0.04-0.08) and (0.04-0.16) respectively. From (12c) it is evident that setting too small a background dissipation rate, (which is default $10^{-12} m^2 s^{-3}$ in NEMO, in units, normalised by water density 10^3kg/m^3), will result in extremely high background viscosity/diffusivity exceeding $10^{-3}\ m^2 s^{-1}$ relevant to the active mixed layer. We use $\varepsilon_{bgr} = 10^{-9} m^2 s^{-3}$ instead, which is appropriate to the observed background values in this region (Simpson et al., 1996, Simpson and Tinker, 2009, Rippeth et al., 2014). In that case resulting diffusivity/viscosity vary in the range of $(6 \times 10^{-6} - 1.3 \times 10^{-5}) m^2 s^{-1}$ and $(6 \times 10^{-6} - 2.6 \times 10^{-5}) m^2 s^{-1}$ respectively.

There are two options to set the sea surface roughness parameter, z_0 , in NEMO. In the first case, z_0 is set as some portion of wave height, h_s: $z_0 = \alpha h_s$, and h_s is defined by empirical parameterisation following (Racle et al, 2008), where α is a constant, estimated to be in the range 0.5:1.3. In the second case, a Charnock type parameterisation (Charnok, 1955) is used: $z_0 = \alpha_{ch}\, \tau_w/(\rho_o g)$, where τ_w, ρ_o are wind shear stress and water reference density, α_{ch} is a constant in the range of (0.1 : 1.6) x 10^5 (Carniel et al, 2009). In equilibrium wind-wave conditions, both formulations are physically equivalent, and $\alpha = 4.16\, \alpha_{ch}(\frac{\rho_a}{\rho_w})C_d^{-1} \approx 10^{-5}\alpha_{ch}$.

Runs CAA and ST differ only by parameterisations of the surface roughness: ST has a smaller surface roughness (see Table 1) and hence less mechanical mixing due to wave breaking. Run KCL has a smaller critical steady state Richardson number, compared with K3 and CAA, but a weaker limit on the turbulent length scale, which potentially effects the upper limit of vertical diffusivity in the non-equilibrium regime. Note that runs (CAL, KCL) are restarted from conditions at January 1, 1998 from the CAA, KC runs respectively and simulated from 1998-2010.

ERSEM, used to examine the sensitivity of ecosystem variables to different turbulence schemes, is a lower trophic level biogeochemical model of phytoplankton, zooplankton, detritus and bacteria, which explicitly resolves carbon, nitrogen, oxygen, phosphorous and silicon cycles (Blackford et al., 2004, Butenschön et al., 2016). On-line coupling to NEMO provides ERSEM with 3D temperature and salinity, velocity and mixing distributions (and hence water column structure) allowing advection and diffusion of the ecosystem variables. The ecosystem model is as much as 10 times more computationally expensive compared to the only-physics runs. In shelf-seas, the annual biogeochemical cycle is characterised by the summer exhaustion of nutrients and regeneration after winter mixing. For these reasons, AMM7-ERSEM simulations are shorter: three runs (CAAE, CABE, GAE) for 2014-2015 are conducted, restarted on 1st January 2014 and initialised with data from a multi-decadal "standard" AMM7-ERSEM run (ST, using the CHCD stability function), beginning in January 1980 (Butenschön et al., 2016).

3.2 Observational data

(a) Long-term variability. For the assessment of the turbulence closures on decadal timescale we use the Met Office Hadley Centre "EN.4.2.0" data set of quality controlled ocean temperature and salinity profiles (Good et al., 2013). After the model spin up we select all temperature and salinity profiles in the area shallower than 200m during the stratified period (March to October, 1998-2010), with $N^2_{max} > 2\ x\ 10^{-5} s^{-2}$. We filter out profiles with $N^2_{max} > 0.5\ s^{-2}$, associated with river plumes, the exact location of which is not well resolved by this relatively coarse resolution model. Daily mean model variables are interpolated as a nearest- neighbour in horizontal and linearly in vertical onto the coordinates of the observations with the resulting dataset containing 22,000 profiles. Note, that the model has a relatively fine vertical resolution on the shelf, especially near the surface and bottom, where the hybrid coordinates are compressed.

(b) Evaluation of synoptic scale variability. High resolution towed-undulating CTD (SCANFISH) hydrography data are used to compare model skill at fine spatial scales. We use data collected during July-September 1998 in 22 transects (~3,000 profiles) in the Celtic Sea, (Brown et al., 2003), and two surveys in June-July 2001 in the North Sea (Cruise R/V Corystes, Centre for Environment, Fisheries and Aquaculture Science; Badin et al., 2009), with 17 transects (1,400 profiles in total). Both papers give detailed descriptions of the regions and the dominant physical processes there. The Celtic Sea survey region is a very energetic region with internal waves and mesoscale eddies, while the North Sea region is shallower, with weaker internal tides and baroclinic jets dominating the summer currents. Given the spacing of the saw-tooth undulation (0.3-1km) of the SCANFISH is much finer than the model resolution considered here (~7km) we consider each down- and up-pass as a vertical profile at the location of its mid-point. Locations of the SCANFISH transects are shown in Fig 2a-c. The time of each SCANFISH transects are restricted to dates only, without details of timing of individual profiles. For this reason we used model daily mean temperature and salinity profiles for comparison. This filters tidal components and diurnal variability, resulting in weaker variability in model variables compared with unfiltered observations.

3.3 Methods of intercomparison.

The evaluation of model skill is not a trivial task and depends on the metrics, databases chosen, and length of observations (Gleckler et al, 2008, Allen et al., 2007). We

evaluate model skill in simulating pycnocline characteristics and bulk properties related to vertical stratification. Previous modelling studies (O'Dea et al., 2012, 2017, Graham et al. 2018) demonstrate persistent strong warm biases of benthic temperatures in this region, comparing with observations. In our analysis, we examine, how local properties of the pycnocline are related to simulation errors in bulk characteristics. All stratification characteristics are calculated on the same vertical grid, and using the same procedures both in observations and model outputs.

The total stratification of the water column on the shelf can be evaluated using the potential energy anomaly (PEA) :

$$PEA = \int_{-H}^{0} g(\rho - \rho_m)zdz, \quad (13)$$

where $\rho_m = H^{-1}\int_{-H}^{0}\rho dz$ is depth-mean density of the fluid, g is gravity, H=min*(200m, h)*, h is the depth of the sea. We use this characteristic to classify models from less diffusive (high *PEA*) to more diffusive (lower *PEA*).

For model evaluation we also use the mixed layer depth (MLD), which is defined by a density criteria, at the depth where the difference with density at 10m does not exceed 0.01 sigma units.

In seasonal pycnoclines in shelf seas, the structure of stratification (temperature and salinity profiles) is approximately two layered, consisting of upper and bottom weakly stratified layers, separated by a strong pycnocline (Fig 3a). To evaluate the model skill in reproducing stratification characteristics compared with hydrographic data, we schematically describe each profile by the following parameters: (i) sea surface temperature (SST), (ii) near bed temperature (SBT) and local pycnocline characteristics: (iii) maximum of Brunt-Väisäla frequency in the pycnocline N^2_{max}, (iv) depth of pycnocline (Z_d) and (v) thickness of pycnocline (Z_T), given by

$$Z_d = \frac{\int_{-H}^{-\delta} N^2 z dz}{\int_{-H}^{-\delta} N^2 dz}, \quad {Z_T}^2 = \frac{\int_{-H}^{-\delta} N^2 (z - Z_d)^2 dz}{\int_{-H}^{-\delta} N^2 dz}, N^2 = -g\rho_0^{-1}\partial_z\rho. \quad (14)$$

Here δ is a masking depth, corresponding to the depth of the shallow diurnal pycnocline (Fig3a). Typically we take δ = 5m, unless observations start from a deeper levels. The depth

of the pycnocline, Z_d is a 'centre of mass' of Brunt-Väisäla frequency and can be expressed also in terms of mean and boundary density:

$$Z_d = \frac{\int_{-H}^{-\delta} N^2 z dz}{\int_{-H}^{-\delta} N^2 dz} = \frac{(\rho(H) - \rho_m) * H - \rho(\delta)\delta}{\rho(H) - \rho(\delta)} \quad , \tag{14a}$$

As most of the domain is temperature stratified, Z_d is strongly related to the ability of the model to predict mean, surface and bottom temperatures.

To examine the model's skill in reproducing the variability of (i)-(v) (Fig 3a) we employ Taylor diagrams (Taylor, 2001), providing a graphical summary of how closely the model patterns match the observations. The similarity between two patterns is quantified in terms of their correlation: *R*, their centred root-mean-square difference: E', and the amplitude of their variations, which is represented by their standard deviations (here σ_{Tm} and σ_{T0} are modelled and observed standard deviations of variable *T*). Being normalised by their standard deviations, all observed variables correspond to a single point (1,0) on the diagram, each model variable correspond to the vector with a length scale $r = \sigma_{Tm}/\sigma_{T0}$ and angle equal to *arc cos(R (o,m))*, where subscripts o denotes observation and m denotes model, *R(o,m)* is the correlation between observations (o) and modelled (m) variables. The root-mean-square difference E' is equal to the distance between model and observed point (Taylor, 2001), and is given by:

$$\frac{E'^2}{\sigma_{T0}^2} = 1 + \sigma_{Tm}^2/\sigma_{T0}^2 - 2R\sigma_{Tm}/\sigma_{T0} \tag{15}$$

We introduce a "skill measure", following ideas proposed by Holloway et al. (2011), to combine variability and bias errors into a single diagnostic variable:

i={*1:5*} ={ *SST, SBT,*Z_d,Z_T,N^2_{max}}

$$Ski^i = \{(1 - E')(1 - \frac{|M^i - O^i|}{\max(M^i) + \max(O^i) - \min(M^i) - \min(O^i)})\}^{1/2} \tag{16a}$$

$$Ski^i = \{(1 - E')(1 - \frac{|M^i - O^i|}{|M^i| + |O^i|})\}^{1/2} \tag{16b}$$

$$Ski = \{\prod_{i=1}^{5} Ski^i\}^{1/5} \quad . \tag{17}$$

Where "O^i", " M^i" denote variables (i-v), | | denotes spatially and temporary averaging of variables (i.e. over 22,000 data in the EN4 comparison, ~3,000 profiles in the Celtic Sea, and 1,400 in the North Sea SCANFISH data sets). For temperature (and salinity), where mean values are far above unity, we use (16a) (e.g. $M = T_{model}$ and $O = T_{observed}$), so biases are normalised by maximum differences over all the observed period, for other pycnocline characteristics we use (16b). Each multiplier in (16) or (17) is less than unity. The square root in (16) and 5_{th} root in (17) is taken to avoid the metric vanishing with increasing number of parameters under consideration. Perfect agreement between modelled and observed parameters occurs when every term in the product is unity. We subjectively consider model skills as high with *Ski* >0.7; satisfactory if 0.5 < *Ski* <0.7; moderate low if 0.4<*Ski* <0.5; low if *Ski* <0.4 and absent of skill when *Ski* is imaginary, i.e. when errors are larger than 100% for one of the variables (i)-(v).

4. Results

4.1 Turbulent mixing schemes (TMS) gradation. Area-mean differences in PEA, simulated with different TMSs are not strong (Fig 3b) in the Celtic Sea, with the maximum difference reached in late summer – early autumn (about 4%). In the North Sea differences are much stronger (25%) (see Fig 3c). While in Fig 3(a,b) 2001 has been chosen in the context of the SCANFISH survey in the North Sea, when high resolution data are available, this behaviour repeats from year to year. Locally, differences in total stratification on the shelf (taken here to be shallower than 200m) reach 50-100% in the North Sea (Fig 2), in the location close to the SCANFISH survey in 2001. Higher domain-mean PEA corresponds to stronger shelf-wide stratification. Using the domain-mean PEA criteria, the TMSs are sorted from the less diffusive (more stratified) to more diffusive (less stratified) as:

$$\text{GA} \rightarrow \text{KC} \rightarrow \text{K3} \rightarrow \{\text{KCL, ST, CAA}\} \rightarrow \text{CAB} \rightarrow \text{CAL} \qquad (18)$$

In general, this classification is in agreement with the critical Richardson number gradation, increasing from GA (Ri_{cL} = 0.12) to CAL (Ri_{cL} =0.49); see Table 1. However, despite the higher critical steady state Richardson number in the K3 run (Ri_{cL} =0.231) compared with KCL run (Ri_{cL} =0.193), the latter TMS is more diffusive, pointing to the importance of the length scale limiter. In the above, KCL, ST, CAA are combined in (18) into a "moderately diffusive group" with relatively similar properties, despite the differences in critical Richardson numbers and surface boundary conditions.

4.2 Comparison of pycnocline characteristics on decadal timescales.

The SBT is affected by the history of horizontal and vertical heat fluxes on seasonal and inter-annual time scales (Holt et al., 2012), without being directly constrained by surface boundary conditions in the summer. Similarly, differences in temperature over the water column, *DT =SST-SBT*, are influenced by integral characteristics of shelf sea dynamics; i.e. 'leakage' of heat through the pycnocline and lateral transport on seasonal time scales. The depth and thickness of the seasonal pycnocline, as well as N^2_{max}, depend more on local pycnocline properties controlled by short-term variability at synoptic, semidiurnal, diurnal and inertial timescales, for example internal waves (e.g. Palmer et al 2013b) and small scale eddies (Badin et al 2009).

Both modelled and observed local pycnocline properties, such as N^2_{max}, demonstrate positive correlations with bulk DT/Z_T, but exhibit strong scatter (Fig 4a and b). Different models show similar patterns, in figure 4 we show an example for the 'moderately' diffusive K3 run. Modelled and observed $N^2_{\max}$ are positively correlated, also with strong scatter (Fig 4c). This scatter suggests a contribution from processes other than vertical mixing, i.e. advection or lateral mixing. These processes may be related to the presence of multiple small scale tidal fronts within the domain (see figure 5 from Graham et. al, 2018). Indeed, when we consider the northern part of the North Sea in isolation, where small scale tidal fronts are less prevalent, stronger similarity between observed and modelled properties in the pycnocline becomes evident (Fig 4 g-i). Generally SST and SBT are well predicted by all TMSs, with very high correlations (close to 0.99; Fig 5a), and positive biases in SBT (model too warm– indicative of a general over-supply of heat from surface to depth). The model error decreases by approximately 10% for both bottom temperature and $N^2_{\max}$, from more diffusive to less diffusive TMSs.

In the majority of the simulations, modelled surface temperatures exhibit positive biases with respect to the observations, with stronger errors in SST (0.13-0.18 ^{o} C) in less diffusive GA-KC models, and small negative biases (<0.02 ^{o}C) in CAL-CAB runs (Fig 4b). Positive biases in the benthic temperature increase with more diffusive models: from 0.03 ^{o}C in the GA case to 0.39 ^{o}C in the CAL case.

The choice of turbulent mixing scheme is shown to have a larger impact on the model properties than the choice of surface boundary conditions: the ST and CAA runs differ only in the surface boundary conditions, and result in changes in SBT biases of approximately

0.05 ^{o}C, while differences in choice of stability functions produce changes in biases of 0.1-0.4^{o}C (Fig 5b).

While in the majority of runs both surface and bottom temperatures exhibit positive biases, the mean temperature (over all observed profiles, showing changes in heat content) has a negative bias in all runs except CAL. Biases change (Fig 5d), from positive 0.11 ^{o}C (CAL) to increasingly negative towards less diffusive mixing schemes (-0.31^{o}C ,GA), which is about a 12% difference in bias, normalised by the standard deviation of 3.22 ^{o}C. This is due to an underestimation of pycnocline depth in all runs: from 3.2m in CAL to 8m in GA. These differences arise due to the feedback between SST and surface heat fluxes, presumably the sensible heat flux, dominating during night time and autumn cooling. SSTs also affect the latent heat flux and thus evaporation, although this is a smaller effect than the sensible heat flux, as demonstrated by small differences in mean salinity biases: ~0.017 PSU between runs (from 0.07 PSU in CAB run to 0.087 PSU in GA), which is only 2.4% of the salinity standard deviation of 0.69PSU. Mean temperatures and so water column heat content on the shelf are better simulated by the CAB and CAA models, while N^2_{max} is better predicted by the low-diffusive TMS runs GA, KC, K3 (Fig. 5d).

Model skill, defined by (16), is very high for SST and SBT in all cases (above 0.8), satisfactory (in the range 0.5 to 0.6) for pycnocline depth and thickness and varies from low (from 0.3-0.33) to moderately low (0.43-0.46) for N^2_{max} in the pycnocline (Fig 5e). Combined skills, evaluated using statistics of all five variables (17) are satisfactory (0.58-0.63) with a slight growth from CAB to GA models.

Surprisingly, KCL and ST runs have very similar characteristics, despite the difference in closures, critical Richardson numbers, surface boundary conditions and limiters applied. This similarity persists in further analysis on the shorter timescale.

4.3. Simulating synoptic variability: comparison with SCANFISH data

The SCANFISH transects, performed in the Celtic and southern part of the North Seas have much higher spatial resolution (~0.3-1km) than the model, ~7km. The Celtic Sea survey covers a region of the shelf that is approximately 100m deep and reveals intense mesoscale and submesoscale activity. Typical eddy length scales are of the order 10km (see Fig 6a; and Holt and James, 2006) and sharper, small-scale variations in the thermocline are observed that are indicative of internal waves. Neither of these baroclinic processes are resolved by the hydrodynamic model. As the exact time of individual profiles is not known, for comparison

we used daily averaged model data. For these reasons, as expected, the simulated thermocline is a comparatively flat and smooth (Fig 6b).

The North Sea SCANFISH survey is conducted in significantly shallower water of around 40-70m at the southern end of the transects and deeper (~100m) at the northern end (Fig 6c-e, Fig 2c). Most of the transects cross a thermal front, which is most pronounced at the east part of domain. The models accurately reproduce the locations of lateral thermal fronts (Fig 6c-e); however these are much thicker and smoother in more diffusive TMSs (Fig 6e), inducing stronger warm biases of benthic temperatures (Fig 6 g-f). Observed pycnocline features appear less dynamic, with generally less vertical variability along isopycnals compared with those in the Celtic Sea (Fig 6a).

The terrain following model coordinates (O'Dea et al. 2017) produce a vertical resolution of about 1-2m over these shallow depths, which is close to the resolution of the SCANFISH survey (1m, in the data format received). Correlations between simulated N^2_{max} in the pycnocline and simulated bulk gradient of temperature T/Z_t , are very high for both regions, $R(N^2_{max,}DT/Z_t) = 0.85, 0.9$ in the North and Celtic Seas respectively. This exceeds that found in the observations (R=0.6 in the North Sea and R=0.85 in the Celtic), which are nonetheless still high. This shows the dominance of vertical heat transport in the simulated and observed stratification. Smaller correlations in the North Sea reflect the importance of larger scale horizontal advection of heat and salt here. In both seas, the variance in the simulated N^2_{max} is 2-3 times smaller than that observed, which is σ^2_T =3.6 x 10^{-3} s^{-4} for the Celtic Sea, based on 3140 'profiles' and 3.4 x 10^{-3} s^{-4}, based on 1398 'profiles' in the North Sea. Modelled variances are σ^2_T =1.6 x 10^{-3} s^{-4} and 1.1 x 10^{-3} s^{-4} respectively. Differences in N^2_{max}variance can partly be explained by filtering of diurnal and tidal variability by the 25h averaging of the model data. The variability of N^2_{max} is approximately the same in the Celtic and North Seas, but is determined by different processes: by eddies and internal waves in the Celtic Sea and by the presence of small scale tidal fronts in the southern part of the North Sea (Fig 6a, 6c).

Considering the statistics of the differences between observations and each model (Fig 7), the SST variability and biases (-0.1 : 0.3°C) are the best simulated of all variables considered (i-v) (see Fig 7a,b,c). Given uncertainties of the heat flux from ERA-INTERIM (resolution about 0.7°~60-80km) and the strong variability of hydrophysical variables in shelf seas, we conclude that SSTs are very well predicted by all models. The most poorly modelled variables are pycnocline thickness and N^2_{max} with errors of about 80% in the North Sea and

90-100% in the Celtic Sea (Fig 7a,b). This can be attributed to high frequency variability in the data, which is smoothed or missing in the models' process representation, and also to inadequate horizontal resolution.

For both SCANFISH surveys, the models' skills in simulating variability of SST, SBT, Z_d, Z_t and N^2_{max} correlate with the diffusivity gradation of the TMSs (18), (see Taylor diagrams (Fig. 7a-b) and bias bar charts (Fig. 7c-f)). However, the direction of 'improvement' of model skill is different for the two seas.

In the North Sea, the models' ability to reproduce variability improves from 'more diffusive' (CAL, CAB) to 'less diffusive' TMSs (KC, GA) for all variables, resulting in a decrease of errors by 10-25 % (Fig 7a). Less diffusive TMS runs show nearly half the negative biases in N^2_{max} (Fig 7f, -40-50% of bias for GA-KC) compared with (-80% -90%) for the CAL-CAB case. However, CAB-CAL show reduced bias in pycnocline depth in this location (-1:-2 m for CAB-CAL versus -4:-5m for K3-GA). All models predict the variability of benthic temperature reasonably well, with high correlations (0.8-0.9) and variability errors (Eqn 15) reducing from (0.6 to 0.4) from CAL to GA TMS. Benthic temperature is positively biased (1.4-2.9 °C), with the worst SBT simulated by more "diffusive" TMS. Thus, excessive "diffusiveness" of CAL-CAA TMS runs results in a deeper pycnocline with weaker stratification, which allows stronger diffusive flux of heat from the surface to the bottom layer and overheats it. In shallow locations, corresponding to the areas of maximum *PEA* differences between the models (Fig 2), the surface mixed layer is highly coupled with the bottom mixed layer, resulting in a shift of tidal mixing fronts and mixing from surface to the bottom in CAB-CAL runs compared with GA (see Fig 5c-e).

The situation is different in the Celtic Sea, where the simulation of variability of some characteristics slightly improves from less diffusive to more diffusive TMSs: for pycnocline depth by 5% and for SST by 10% (Fig 7b). Biases are small for all variables, except the pycnocline depth, for which biases are negative everywhere and better predicted by the CAA and CAB models. K3-GA models strongly underestimate Z_d with a deficit of 4-7m. We hypothesise, that sporadic mixing, produced by mesoscale and submesoscale events, shear instability or breaking of internal tidal waves deepen and broaden the pycnocline in the Celtic Sea, and their effects are better represented by the more diffusive closures. In other words, use of more diffusive turbulence closures is probably compensating for the lack of unresolved processes or their parameterizations in moderate resolution 3d models.

Considering the skill scores for the comparison with SCANFISH data (Fig 8), all models have high skills in the prediction of surface and benthic temperatures, with slightly better

properties of CAB-CAA models in the Celtic Sea and GA-KC in the North Sea (Fig 8ab), consistent with Fig. 7. Similar trends are valid for pycnocline characteristic (Fig 8 c-e), with much stronger preference for KC-GA in the North Sea. Despite stronger biases in the prediction of the pycnocline depth, GA-KC TMSs have greater skill than CAA-CAB for this particular characteristic. Integral model skill (Fig 8f) shows the same tendency: slightly better skills of CAL-CAB runs in the Celtic Sea, and much better prediction by K3, KC and GA TMSs in the North Sea.

In general, model skill to predict pycnocline characteristics on the synoptic scale is satisfactory (>0.5) only for GA-KC models in the North Sea. Combining skill metrics in both regions as $Ski^i = \{Ski^i{}_{NS} \cdot Ski^i{}_{CS}\}^{1/2}$ for each variable and for total skills, the preference for low diffusive (GA, KC) TMSs for SBT and $N^2{}_{\max}$ becomes evident. Equilibrium TMSs (ST, K3, KCL) win in the prediction of pycnocline depth only. Resulting skills (Fig 8e) grow from 0.39 to 0.48 from high to low "diffusive" TMS.

KCL and ST models show very similar results despite the differences in the formulation of the closure, the critical Richardson numbers in steady state and the limiting length scale parameter. Increasing length scale limiter and type of limiting (from KC to KCL and from CAA to CAL) results in a reduction of skills in some parameters (bottom temperature and N^2_{max}) or nearly neutral in others. Similarly, an increase in surface roughness and of the effects of breaking waves (from ST run to CAA) does not improve model skill.

5. Discussion

5.1 Comparison with other sensitivity studies in the North-West European shelf

Here we demonstrate that the differences induced by changing the stability functions, as discussed in this work, produce model biases of the same order of magnitude (~0.2-0.4 °C) as differences arising from changes in heat flux or from improvements in horizontal resolution, as reported in other studies in this region. For example, O'Dea et al., (2017) used different light attenuation schemes in 30 year simulations of AMM7 and found summer month differences in SST and bottom temperature to be 0.2-0.3 °C in the seasonally stratified part of the shelf (their Fig. 15a). Similarly, Graham et al. (2018) compared 20 year runs of AMM15 (1.5 km resolution) and an AMM7 run, which corresponds to the ST run in this study. They report that the summer months (July – August, JJA) bottom temperature biases (model minus EN4) varied from $0.47 \pm 1.19°C$ to $0.50 \pm 1.31°C$ for the different configurations. An SST

comparison with EN4 gives the following biases and standard deviations: $0.18 \pm 0.97°C$ and $0.12 \pm 0.99°C$ for AMM15 and AMM7 correspondingly. These values are not identically comparable with our study, as the simulations length differed (20 years and 14 years correspondingly) and analyses were performed for different periods of time (JJA versus March –October). However, the model appears to be more sensitive to differences in TMSs (Fig 5b), than to resolution: a noteworthy result. SST biases increase from $0.07 \pm 0.74°C$ in CAB run to $0.18 \pm 0.78°C$ in GA run, while SBT biases decrease from $0.33 \pm 1.01°C$ in CAB to $0.01 \pm 0.84°C$ in GA run. ST run has $0.09 \pm 0.74°C$ in SST and $0.18 \pm 0.92°C$ in SBT. The smaller biases in SBT compared with Graham, 2018 paper is mostly due to different averaging over seasons, as SBT differences are much smaller during March-May (not shown here).

5.2 Implications for ecosystem modelling

Here we consider the degree to which the differences in stratification and mixing, caused by the different mixing schemes, impact on the simulation of the shelf sea ecosystem. We performed three runs with the most (CAB), the least (GA) and moderately (ST) diffusive TMSs (but with standard limiter parameters) using the ERSEM ecosystem model. These runs start from the same initial condition in January 1, 2014 from the multidecadal "standard" AMM7-ERSEM run, which corresponds to ST case.

Ecosystem validation is not a trivial matter, and requires a multi-variable and multi-metric approach (Allen et al, 2007, Edwards et al, 2012). We leave a more detailed analysis to future work and for brevity focus on chlorophyll as a combined indicator of phytoplankton biomass and primary production. Figure 9(a,d,g) shows the water column integrated and area mean chlorophyll concentration in 2015 for the Celtic and North Seas and for the whole domain. These shelf sea regions have two distinct phytoplankton bloom peaks in April-May and June, whereas the whole domain has a peak in June (Fig 9g). The initial bloom starts earlier in the GA run, with about a 20 day delay in the timing of the bloom peak in the CAB run, which is later and stronger. We relate the earlier bloom in the GA run to an earlier on-set of stratification in spring, which retains phytoplankton in the euphotic zone allowing earlier bloom initiation. The secondary peak is much stronger in the North Sea in the CAB run.

After a strong bloom, caused by the excess of nutrients in the CAB run built up before July, waters are then depleted of nutrients with a consequent strong reduction in chlorophyll compared to the GA run. The earlier bloom initiation in lower seasonal light conditions in the

GA case results in a weaker and longer bloom (Fig 9d). Over the whole domain, depth integrated temporal differences in chlorophyll reach 20-30%. Depth integrated area and annual mean chlorophyll increases with the diffusiveness of the model, GA: 39.3+/- 33.6 mg m^{-2}, ST: 41.5+/- 37.9 mg m^{-2} and CAB: 43.4+/- 42.7 mg m^{-2}.

The first strong bloom peak is associated with processes in the mixed layer (Fig 9b,e,h). Chlorophyll, integrated both over the mixed layer and below the pycnocline, exhibits very strong temporal synoptic scale variability. This is related to changes in the MLD with wind (or night convection), destratification and restratification due to solar radiation and has a variability of 20-40 mg m^{-2} compared to 80-120 mg m^{-2} mean peak magnitudes. Time mean chlorophyll is almost equally distributed between the upper mixed layer and subsurface, pycnocline GA: 17.5+/- 17.3 mg m^{-2} in ML versus 22.8 +/- 22.5 mg m^{-2} below ML (pycnocline), ST: 19.9 +/- 20.7 mg m^{-2} in ML versus 22.7+/- 24.2 mg m^{-2} below, CAB: 23.6 +/- 25.2 mg m^{-2} in ML versus 21.0+/- 25.1 mg m^{-2} in pycnocline. Thus the portion of subsurface chlorophyll reduces from 62% to 44% with increasing diffusiveness of the model. The nearly equal distribution of annual mean chlorophyll between ML and pycnocline is consistent with (Rippeth et al., 2014) estimate of up to 50% of annual carbon fixation in North –West European Shelf being related to pycnocline mixing.

After the strong bloom in spring (80-120 mg m^{-2}) ML chlorophyll reduces to 10-20 mg m^{-2} in strongly stratified conditions in summer (days 160-240). On the contrary, sub ML level of chlorophyll is stabilised at the much higher level of 40-60 mg m^{-2} (Fig 9c,f). On the shelf, in winter, chlorophyll levels nearly vanish due to strong cooling in well mixed seas and reduced light levels.

Edwards et al. (2012) performed a detailed chlorophyll assessment of the NEMO-ERSEM operational model for the northwest European shelf versus satellite-derived chlorophyll and four “Smart Buoys”, located in relatively shallow regions. Seasonal mean variability (their figure 9) in NEMO-ERSEM and satellite-derived surface chlorophyll is ~1-10 mg m^{-3} during spring and summer, of the same order as the model results presented here (Fig 10).

Despite similar patterns in surface chlorophyll, depth integrated relative differences between CAB-GA are large (Fig 10, bottom panel), exceeding 100% in different seasons. During April-May, depth integrated chlorophyll values are higher by as much as 80% in the GA run in the Central and Northern parts of the North Sea (Fig 10d), and slightly higher in the CAB run in the Celtic Sea. In June, CAB chlorophyll exceeds that of the GA run almost everywhere on the shelf by 40-60%.

In accordance with figures 9 (b, e, h) in July surface chlorophyll is reduced (Fig 10 i,j,k), while subsurface maxima below the mixed layer exceed surface values by 5 times in some locations (Fig 10m, n, q). Differences between models become more prominent in the shelf regions (Fig 10m,q) with three times larger subsurface concentrations in GA than in CAB. .

Thus, the ecosystem characteristics are highly sensitive indicators of the quality of the turbulence closure scheme and in future could be used for the assessment of these schemes using, for example, ocean colour data from satellites.

5.3 Which closure works best?

There is no clear answer to this question. We considered here three different turbulent closures, based on works of GKHR, KC94 and CHCD with different physics and different critical steady state Richardson numbers. Two simulations considered here (KC and K3), correspond to different constants in KC94 and (Kantha, 2003), and two (CAA, CAB) in CHCD. Another variation arises from the parameter value and way the Deardorff /Galperin limiters are applied, which effects the steady state critical Richardson numbers and background diffusivity (Fig. 1). This variation of parameters allows the TMS to be graduated from low diffusive to high diffusive with steady state Richardson numbers in the range from 0.12 (GA) to 0.492 (CAL).

The limiting parameter is critical: the KCL case, using the KC94 closure, appears to be slightly more diffusive than the ST run, which uses the CHCD closure, but in most characteristics their skill is indistinguishable (Fig. 5, 7, 8). Application of limiters is important: the weak limiting case (CAL versus CAA, Fig.4a) results in extra heating of the bottom temperature (0.4 °C compared to 0.2 °C) and reverses the sign of heat content biases, when compared with the EN4 profile dataset.

In all comparisons, variability (except SST in the Celtic Sea, Fig 7b) is better predicted by less diffusive models, with differences of about 10-20%.

At multi-annual timescales more diffusive models, CAL-CAA, better control the heat content, but result in an excessively smoothed pycnocline, compared to both the EN4 dataset and the SCANFISH surveys. Conversely, less diffusive (KC-GA) TMS predict more precisely the strength of the pycnocline and bottom temperature, but have strong negative biases in pycnocline depth (up to 8m) and in heat content (up to 0.3 °C).

In these experiments the surface layer thickness largely controls the SST and so the sensible heat flux and total heat content. Indeed, having the same shortwave radiation and advective heat flux, less diffusive models simulate a shallower pycnocline, higher SST and hence less heat content in the domain compared with observations, with average temperatures lower by 0.2-0.3 °C. With a shallower pycnocline, it is easier to gain and lose heat in the mixed layer during the diurnal cycle.

One of the most diffusive closures, CAB, produces negative biases in heat content and pycnocline depth, but also has negative biases in the strength of stratification. We conclude that all the models have a general deficit of mechanically driven mixing, presumably due to the absence of parameterisations of missing physical processes on the shelf (Inall et al., 2011, Rippeth et al., 2009, 2014, Palmer et al, 2013, 2015, Green et al., 2008, Gerkema and Shrira, 2005).

5.4 What processes and parameterisations may be missing in the models?

It is clear from the analysis presented here that missing physical processes play a key role in the model-observation mis-match. Some of these processes are on their way to being resolved or already included in other models (e.g. long internal tides and the effects of gravity waves), the other group of processes are non-hydrostatic and still require reliable parameterisations. Candidate 'missing processes', worthy of further consideration are:

(a) Shear instability of long internal tides, which is resolved in fine-resolution kilometre scale models (Guihou et al, 2018, Graham et al, 2018), but not in the current study. Graham et al, (2018) shows that "over the continental shelf break, there are still warm biases compared with observations, however it was reduced compared with AMM7". Interaction of increased tidal shear with near-inertial oscillations, induced by wind, or wind-driven Ekman shear stresses (Burhard and Rippeth, 2009, Lenn et al, 2011) is a good candidate, which would lead to deepening of the mixed layer and pycnocline.

(b) Effects of gravity waves. The strongest impact on turbulence and shear stresses is produced by young waves (Brown and Wolf, 2009), that are sharper and steeper, and so have a surface roughness parameter three times greater than for well-developed, old waves (Janssen at al., 2002) with stronger wave dissipation. A coupled ocean-wave-atmosphere system for the NWES region (Lewis et al, 2018a,b) includes feedbacks on the turbulence due to shear stresses between ocean-waves and the atmosphere and the effects of waves on ocean roughness. Those studies show a reduction of SST biases versus observations compared with

the ocean-only model only, presumably due to an increase in mixed layer and pycnocline depth.

(c) Langmuir circulations (LCs) are another feature, induced by waves that is not present in the current parameterisation. LCs are organised structures, similar to convective rolls, induced by the combined effects of current shear and wave Stoke's drift. Despite recent progress in the extension of turbulence closures to include LCs (Kantha, & Clayson 2004, Harcourt, 2015, Kukulka and Harcourt, 2017), the role of interaction between LCs, wave breaking and strong stratification is not well understood. Most parameterizations are based on Large-Eddy Simulations that do not have the effects of wave breaking and only a few experiments account for strong stratification. Craik-Leibovich instability theories (Craik and Leibovich, 1976, Leibovich, 1983) do not predict the development of instabilities, generating of LCs at the conditions of strong stratification or a high viscosity environment (the latter is the case of convection or when strong wave breaking happens).

Another restriction from Craik - Leibovich theory is that LC instability is unlikely under conditions of short wavelength dominant waves (Leibovich, 1983). The importance of swell waves was highlighted in McWilliams et al, (2015) in a LES simulation study, which demonstrated a strong effect of swells/wind alignment on the depth and intensity of mixed layer. Thus the presence of swell waves aligned with wind shear in the weakly stratified ocean is the preferred niche of LCs. LCs turbulent closures (Kantha & Clayson 2004, Harcourt, 2015) developed for the $(q^2, q^2 l)$ equations, could in principle be adapted to the GLS scheme (Kantha, 2004).

(d) Direct effects of the horizontal (northward) component of the Coriolis force are usually neglected in turbulence closures. These terms act in a similar way to stratification, depending on the sign of the zonal velocity shear. If this sign is positive, rotation effects enhance mixing and vice versa. Galperin et al. (1989), using their extension of the GKHR model, found that in an unstratified flow these effects can enhance/suppress the depth of the Ekman mixing layer by 50%, with weaker (5-10%) effects in a stably stratified environment. Zikanov et al, (2003) examined the structure of the Ekman layer in unstratified flows in low and moderate latitudes and found a similar estimate of 50% in enhanced/supressed mixing by the effects of the horizontal component of the Coriolis force.

(e) Near inertial oscillations associated with the horizontal component of the Coriolis force (Shrira and Forget, 2015) provide another source of mixing. NIOs of this type cannot penetrate through the seasonal pycnocline and concentrate strong shear at the top of the pycnocline, and so deepen it. Extra, unresolved mechanical mixing, if present, will deepen the summer pycnocline and decrease night and seasonal autumn heat loss from the sea surface, reduce water column heat content loss .

(f) Another source of turbulence in the tidally active seas is from sharp and steep nonlinear internal (solitary) waves. These are widely present in observations (Holt and Thorpe, 1997, Inall at al., 2000, Rippeth and Inall 2002, Palmer et al, 2015) but absent in the AMM models, both due to resolution and the use of the hydrostatic assumption. Observations in the South China Sea (Liu and Lozovatsky, 2012) demonstrated that the averaged dissipation rate of turbulent kinetic energy in the upper pycnocline is approximately linearly related with the available potential energy of the internal waves. Simulation of these waves requires both a horizontal resolution as fine as 50m (Vlasenko et al, 2014) and the lifting of the hydrostatic assumption, and so will not be realisable for a least the next 10 years at a shelf scale.

Thus, models, used for forecasting and hindcasting require new parameterizations of these processes based on in-situ and satellite observations, theory and Large Eddy Simulations. MacKinnon et al, 2017 review in detail the results of the program "Climate Process Team on internal wave-driven ocean mixing", aiming to improve parameterizations of internal wave driven ocean mixing in the deep ocean, however shelf-sea processes are not included in that study

6. 6. Summary

In decadal scale simulations of the dynamics of the tidally active North-West European shelf seas, we assess the properties of eight vertical (diapycnal) turbulence mixing schemes (TMS), present in the NEMO model. Each scheme (Table 1 and Fig. 1) is a combination of a turbulent closure, i.e. (Galperin et al,1988), (Kantha and Clayson, 1994) , Kantha (2003) or (Canuto et al, 2001), different sets of constants, used in these closures and the application of the Deandorf/Galperin limiter c_l.

Ozmidov (Deardorff, /Galperin) limiting condition, derived from a scaling analysis, implies the upper limit on the turbulent length scale: at strong stratification, eddies exceeding

the Ozmidov scale are unable to overturn. If limiting conditions are applied, critical gradient Richardson number in steady state becomes dependent on the limiter value and type of closure, summarised in (Fig 1a). The formulation of background mixing is part of the mixing scheme (10a), and inversely proportional to stratification, with limiter dependent coefficients (Fig1b-d).

Using the Potential Energy Anomaly as a measure of total stratification, TMSs are graduated from "less diffusive" to "more diffusive" (Eqn 15): GA (Galperin et al., 1988) →KC (Kantha and Clayson, 1994) → K3 (Kantha, 2003)→CAA, CAB (Canuto et al., 2001), in agreement with the steady state critical Richardson numbers, Ri_{cl}, which varies from 0.12 to 0.492 in the different schemes (see Fig 1a and Table 1).

We characterise the density profiles in the shelf seas by 5 parameters (Fig 3a): Surface and benthic temperatures, pycnocline depth, thickness and strength (Fig 2d) To asses model results versus observation we propose a measure, "model skills", which combines errors in variability in biases for all considered characteristics in a single parameter (eqs 16-17) As smaller model's errors, the closer 'model skills' to unity.

For the chosen pycnocline characteristics we find "model skill" to be moderate to high (>0.6) for moderate and low "diffusive" TMSs (ST, K3, KC, GA) on decadal timescale (Fig 5e).

In the simulations of 'sea weather', comparing seasonal simulations with SCANFISH data, only the low "diffusive" models (KC, GA) have satisfactory skills (> 0.6) in the North Sea (Fig 8f). There is a strong area dependence in the model biases (Fig.7): in the synoptic scale assessment of the North Sea simulations bottom temperature are too warm (+3°C) and N^2_{max} in the pycnocline is too low (-100%) in the more diffusive models, with biases twice as large as in less "diffusive" TMSs (+1.44 °C and -50%). In the tidally energetic Celtic Sea, low diffusive TMSs fail to simulate sufficiently deep pycnoclines, with biases reaching 8m, twice as large as biases in the North Sea. Thus there is not a "winner" between these turbulence mixing schemes.

In all simulations, the strength of pycnocline is too week and pycnocline depth is too shallow. Analysing the behaviour of different metrics, we conclude, that some processes, responsible for the mechanical mixing are not included into the models (section 5.4). That might be effects of young waves under conditions of growing winds, Langmuir circulations

caused by winds and swells (old waves), non-traditional near-inertial oscillations, generated as winds drop and strong shallow seasonal pycnocline and others. Each of the "missing processes" occupies their own niche in physical parameter space.

We found that the sensitivity of models to mixing schemes and choice of limiters is comparable to the sensitivity to horizontal resolution (Graham et al, 2018) and surface flux formulation (O'Dea et al, 2018).

Finally, we find that simulations with the coupled ecosystem model show a strong effect of the choice of mixing scheme on the timing of phytoplankton blooms, on the spatial patterns of chlorophyll (Figs 9, 10) and partitioning of chlorophyll production between mixed layer and pycnocline. We found that, in accordance with previous observation-based estimates, nearly half of the yearly chlorophyll production occurs in the pycnocline, and this dominates in the strongly stratified period of the year (July-August). In less "diffusive" TMS (GA) mixed layer spring blooming starts earlier and peak is weaker, compared with "moderate" ST and " diffusive" CAB runs. The pycnocline bloom is longer lasting, with 20-30% greater intensity in the GA run. The portion of subsurface chlorophyll reduces from 62% to 44% with increasing diffusiveness of the TMS. This sensitivity motivates an assessment against high frequency/resolution ecosystem and biogeochemical observations, but this is left for future work.

7. Code and data availability.

AMM7 is a regional configuration of NEMO (Nucleus for European Models of the Ocean), at version 3.6 stable (Madec,2016). Model code is available from the NEMO website (www.nemo-ocean.eu). ERSEM code (ERSEM16.06 commit e5272b7f) is available for download from https://www.pml.ac.uk/Modelling_at_PML/Models/ERSEM. EN4 database (version EN.4.2.1) is available from Met Office website https://www.metoffice.gov.uk/hadobs/en4/. SCANFISH data and model outputs are available by the request to authors (mane1@noc.ac.uk). NEMO-GLS section of namelist and compilation keys is given in Appendix B.

Appendix A. Critical Richardson number in equilibrium regimes and the Deardorff-Galperin limiter.

The critical gradient Richardson number (Kantha, 2003, Chen, 2003) is defined by quadratic equation

$$(c_1 g_h^2 + c_4\, g_h + 2)Ri^2 + (c_5 g_h - c_2 g_h^2)Ri + c_3 g_h^2 = 0; \qquad \text{(A1)}$$

Where $g_h = (c_l{}^2 B_1^2)$ is the upper limit of $g_h = G_h B_1^2$. The list of constants, $c_1, c_2, c_{3\,,} c_{4\,,} c_5$, B_1 for GA, K3 case can be found in Kantha, 2003. For CHCD cases we calculate relevant constants using equations (18d, 23c) from (Cheng et al, 2002):

$$(c_1 \times 10^{-3},\ c_2 \times 10^{-3}, c_{3\,,} \times 10^{-5}, c_{4\,,} c_5 \times 10^{-5},\ B_1) =$$

$$\begin{Bmatrix} CHCD(A): 1.60, & 1.40\,, & 1.09\,, & 0.184, & -3.91, & 19.3 \\ CHCD(B): 1.30, & 1.30\,, & 1.50\,, & 0.158, & -4.77, & 16.6 \\ K3 \quad\ : 2.87, & 2.48\,, & -0.18\,, & 0.259 & -4.73, & 16.6 \\ GKHR \ : 6.89, & 1.33\,, & 3.55\,, & 0.0\ \ , & -4.72, & 16.6 \end{Bmatrix} \qquad \text{(A2)}$$

The KC94 case has been taken from Umlauf and Burchard, (2005, Fig.3), as constants (A2) are not found in the published literature.

$$Ri_{cL} = \begin{Bmatrix} c_l: & 0.267, & 0.53, & +\infty \\ GA: & 0.124, & 0.165, & 0.194 \\ KC: & 0.143, & 0.193, & 0.2416 \\ K3: & 0.231, & 0.479, & 0.894 \\ CAA: & 0.2497, & 0.492, & 0.849 \\ CAB: & 0.2543, & 0.521\,, & 1.023 \end{Bmatrix} \qquad \text{(A3)}$$

Appendix B. Details of NEMO set up

Model has been compiled using the following fppkeys:

key_dynspg_ts key_ldfslp key_zdfgls key_mpp_mpi key_netcdf4 keyy_nosignedzero key_traldf_c2d key_dynldf_c2d key_bdy key_tide key_vvl key_iomput.

Full namelist_ref and namelist_cfg could be found in supplemental matherials. We varied the following options in namelist (shown in bold)

!---
&namzdf_gls ! GLS vertical diffusion ("key_zdfgls")
!---
ln_length_lim = .true. (for all cases except CAL, KCL)
ln_sigpsi = .true.
nn_bc_bot = 1
nn_bc_surf = 0
nn_clos = 1
nn_stab_func = 0 (varied from 0- to 4)
! 0-GKHR, 1=KC94 , 2=CanutoA, 3=Canuto B, 4=K3, added option
nn_z0_met = 1 (for ST case and =2 for others)

```
   rn_charn        = 100000.0
   rn_clim_galp    = 0.267
   rn_crban        = 100.0
   rn_emin         = 1.0e-6
   rn_epsmin       = 1.0e-9
   rn_frac_hs      = 1.3
   rn_hsro         = 0.003
/
```


Acknowledgements

The authors acknowledge the National Capability funding in Ocean Modelling provided by the UK Natural Environment Research Council (NERC), NERC funded research grants PycnMix (grant number NE/ L003325/1) and Shelf Seas Biogeochemistry (NE/K001698/1). IK and EZ acknowledge the support by RFBR under Grant No. 16-35-60072 mol_a_dk and by the Russian Science Foundation under Grant No. 17-77-30019. We are grateful to two anonymous reviewers, whose valuable comments allowed us to significantly improve the manuscript. We also acknowledge use of the UK National High Performance Computing resource (http://www.archer.ac.uk).

Literature.

Abarbanel, H. D., D. D. Holm, J. E. Marsden, and T. Ratiu (1984), Richardson number criterion for the non-linear stability of 3D stratified flow. *Phys. Res. Lett.,* **52,** 2352–2355.

Allen J.I., J.T. Holt, J. Blackford, R Proctor (2007), Error quantification of a high-resolution coupled hydrodynamic ecosystem coastal-ocean model: Part 2. Chlorophyll-a, nutrients and SPM, J Marine Sys. 68, oi:10.1016/j.jmarsys.2007.01.005

Andren, A., and C.-H Moeng,(1993), Single point closures in a neutrally stratified boundary layer, J Atmos.Sci.,5 0, 3366-3379.

Badin, G., R.G. Williams, J.T. Holt, L.J. Fernand (2009), Are mesoscale eddies in shelf seas formed by baroclinic instability of tidal fronts?, J. Geophys. Res.-Oceans, 114, C10021.

Belcher, S. E., et al. (2012), A global perspective on Langmuir turbulence in the ocean surface boundary layer, Geophys. Res. Lett., 39, L18605, doi:10.1029/2012GL052932

Blackford, J. C., Allen, J. I., and Gilbert, F. J. (2004), Ecosystem dynamics at six contrasting sites: a generic modelling study, J. Marine Syst., 52, 191–215.

Brown, J., L. Carrillo, L. Fernand, K.J. Horsburgh, K.J., A.E. Hill, E.F. Young, K.J. Medler (2003), Observations of the physical structure and seasonal jet-like circulation of the Celtic Sea and St. eorge's Channel of the Irish Sea. Continental Shelf Research, 23, 533-561.

Burchard, H. and K. Bolding (2001), Comparative analysis of four second-moment turbulence closure models for the oceanic mixed layer, J. Phys. Oceanogr. 31.

Burchard, H. and K. Bolding (2002). GETM - A General Estuarine Transport Model, Scientific Documentation, European Commission**:** 157pp.

Burchard, H., P.D. Craig, J.R. Gemmrich, H. van Haren, P.-P. Mathieu, H.E.M. Meier, W. A.M. Nimmo Smith, H. Prandke, T.P. Rippeth, E.D. Skyllingstad, W.D. Smyth, D.J.S. Welsh, and H.W. Wijesekera (2008), Observational and numerical modeling methods for quantifying coastal ocean turbulence and mixing, *Progr. Oceanogr.*, 76, 399-442.

Burchard, H. and T.P. Rippeth (2009), Generation of bulk shear spikes in a shallow stratified tidal seas. J. Phys. Oceanogr, 39, 969-985.

Butenschön, M., J. Clark , J.N. Aldridge, J.I. Allen, J.I., Y. Artioli, J. Blackford, J. Bruggeman, P. Cazenave, S. Ciavatta, S. Kay, G. Lessin, S. van Leeuwen, J. van der Molen, L. de Mora, L. Polimene, S. Sailley, N. Stephens, R. Torres, R. (2016) ERSEM 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels. *Geosci. Model Dev.*, 9, 1293-1339.

Canuto, V.M., A. Howard, J. Cheng, M.S. Dubovikov, M.S. (2001), Ocean turbulence. Part I. One-point closure model-momentum and heat vertical diffusivities. J. Phys. Oceanogr. 31 (6), 1413–1426.

Canuto, V. M., A.M. Howard, J. Cheng, C.J. Muller, A. Lebbosier, A. and S.R. Jayne (2010), Ocean turbulence, III: New GISS vertical mixing scheme, Ocean Modell., 34, 70–91. doi:10.1016/j.ocemod.2010.04.006.

Carniel S., J. C. Warner, J. M. Chiggiato (2009), Investigating the impact of surface wave breaking on modeling the trajectories of drifters in the northern Adriatic Sea during a wind-storm event, Ocean Modelling, 30 , 225–239.

Charnok, H. (1955), Wind stress on a water surface, Quart. J. R. Meteorol. Soc. 81, 639–640.

Cheng, Y., V.M. Canuto, A..M., Howard (2002), An improved model for the turbulent PBL. J. Atmosph. Sci, 59, 1550-1565.

Cheng, Y., V.M. Canuto, A.M. Howard (2003), Comments on "On an improved model for the Turbulent PBL" An improved model for the turbulent PBL". J. Atmosph. Sci, 60, 3034-3046.

Cheng, Y., V.M. Canuto, A.M. Howard (2005), Non-local convective PBL model based on new third and fourth order moments: J. Atmosph. Sci 62, 2189–2204.

Costa, A., A.M. Doglioli, P. Marsaleix, A.A. Petrenko (2017), Comparison of in situ microstructure measurements to different turbulence closure schemes in a 3-D numerical circulation ocean model. Ocean Modelling, 120, 1-17
http://dx.doi.org/10.1016/j.oceanmod.2017.10.002

Craik, A. D. D. and S. Leibovich (1976), A rational model for Langmuir circulations, J. Fluid Mech. 73:401-26..

Deardorff, J.W. (1973), The use of subgrid transport equation in a three-dimensional model of atmospheric turbulence. ASME paper N73-FE21.

Dee, D., S. Uppala, A. Simmons (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597.

Edwards K.P., R.Barciela and M. Butenson (2012). Validation of NEMO-ERSEM operational ecosystem model for the North West European Continental Shelf (2012), Ocean Sci., 8, 983–1000, 2012 www.ocean-sci.net/8/983/2012/ doi:10.5194/os-8-983-2012

Flather RA. (1976). A tidal model of the northwest European continental shelf. Mem. Soc. R. Sci. Liege 10(6): 141–164.

Flather, R.A. (1981), Results from a model of northeast Atlantic relating to the Norwegian coastal current, in: The Norwegian Coastal Current, edited by Saetre, R. and Mork, M. 427-458.

Galperin, B., L.H. Kantha, S. Hassid, A. Rosati (1988), A quasi-equilibrium turbulent energy model for geophysical flows. J. Atmos. Sci. 45, 55–62.
Galperin, B., A. Rosati, L.H. Kantha, G.L. Mellor (1989), Modelling rotating stratified flows with application to oceanic mixed layers, J. Phys. Oceanogr. 19, 902-916.

Gargett, A.E and G. Holloway (1984), Dissipation and diffusion by internal wave breaking, J. Marine Sci., 42, 1, 15-27, https://doi.org/10.1357/002224084788506158

Gerbi, G.P., J.T. Trowbridge, E.A. Terray, E.A., A.J. Plueddemann, T. Kukulka (2009), Observations of turbulence in the ocean surface boundary layer: energetics and transport, J. Phys.Oceanogr, 39, 1077-1096, DOI: 10.1175/2008JPO4044.

Gerkema, T., and V.I. Shrira (2005), Near inertial waves in the ocean: Beyond the ‘‘traditional approximation.’’ J. Fluid Mech., 529, 195–219, doi:10.1017/S0022112005003411.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux (2008), Performance metrics for climate models, J. Geophys. Res. 113 D06104, doi:10.1029/2007JD008972.

Good, S. A., M.J. Martin, and N.A. Rayner (2013), EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res. Oceans, 118, 6704–6716, doi:10.1002/2013JC009067.

Graham, J. A., E. O'Dea, J.T. Holt, J. Polton, H.T. Hewitt, R. Furner, K. Guihou, A. Brereton, A. Arnold, S. Wakelin, J.M. Castillo Sanchez, and C.G. Mayorga Adame (2018), AMM15: a new high-resolution NEMO configuration for operational simulation of the European north-west shelf, Geosci. Model Dev., 11, 681-696, https://doi.org/10.5194/gmd-11-681-2018.

Green, J.M., J.H. Simpson, S. Legg, S. and M.R. (2008), Internal waves, baroclinic energy fluxes and mixing at the European shelf edge. Continental Shelf Research, 28(7), pp.937-950.

Harcourt R.R (2015), An improved Second-Moment Closure Model of Langmuir Turbulence. J. Phys. Oceanogr. 45, 84 -103. doi: *doi.org/10.1175/JPO-D-14-0046.1*

Holloway, G., A. Nguyen, Z. and Wang (2011), Oceans and ocean models as seen by current meters, J. Geophys. Res., 116, C00D08, doi:10.1029/2011JC007044.

Holt, J.T., S.A. Thorpe. The propagation of high frequency internal waves in the Celtic Sea (1997), Deep Sea Res., 44, pp2087-2161.

Holt, J. T. and I. D. James (2006). "An assessment of the fine scale-eddies in a high resolution model of the shelf seas west of Great Britain." Ocean Modelling **13**: 271-291.

Holt, J., and L. Umlauf (2008), Modelling the tidal mixing fronts of the northwest European continental shelf. *Cont. Shelf. Res.*, 28, 887-903.

Holt, J., Hughes, S., Hopkins, J., Wakelin, S. L., Holliday, N. P., Dye, S., Gonzalez-Pola, C., Hjollo, S. S., Mork, K. A., Nolan, G., Proctor, R., Read, J., Shammon, T., Sherwin, T., Smyth, T., Tattersall, G., Ward, B., and Wiltshire, K. H.(2012). Multi-decadal variability and trends in the temperature of the northwest European continental shelf: A model-data synthesis, Progress in Oceanography, 106, 96–117, doi:10.1016/j.pocean.2012.08.001, 2012.

Holt, J., C. Schrum, et al. (2016). Potential impacts of climate change on the primary production of regional seas: a comparative analysis of five European seas, Progress in Oceanography, 140, 91-115.

Inall, M.E., T.P. Rippeth, T.J. Sherwin (2000), Impact of nonlinear waves on the dissipation of internal tidal energy at a shelf break. J. Geophys. Res, 105, C4, 8687-8705

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

Kantha, L. H.: On an improved model for the turbulent PBL,(2003). J Atmos. Sci., 60, 2239-2246

Kantha, L.H. and Clayson, C.A. (2004), On the effect of surface gravity waves on mixing in the oceanic mixed layer. Ocean Modell. 6, 101–124.

Kolmogorov, A.N. (1942) Equations of turbulent motion in an incompressible fluid. Izv. Akad. Nauk. SSSR ser. Fiz. 6, 56-58.

Kukulka T. and R.R Harcourt (2017), Influence of Stokes Drift decay scale on Langmuir turbulence, J.Phys . Oceanogr., 47, 1637-1656. *doi.org/10.1175/JPO-D-14-0046.1*

Large, W. and S. Yeager (2009), The global climatology of an interannually varying air–sea flux data set, Clim. Dynam., 33, 341–364.

Launder, B.E. and W.P. Jones (1969), Sink flow turbulent boundary layers, J. Fluid Mech. 38, 817–831.

Liu Z. Y. and I. Lozovatsky (2012), Upper pycnocline turbulence in the northern South China Sea. Chin Sci Bull, , 57: 23022306, doi: 10.1007/s11434- 012-5137-8.

Leibovich S., 1983, The form and dynamics of Langmuir circulations. Ann. Rev. Fluid Mech. 1983. 15:391-427

Lenn, Y. D., T. P. Rippeth, C. P. Old, S. Bacon, I. Polyakov, V. Ivanov, and J. H€olemann (2011), Intermittent intense turbulent mixing under ice in the Laptev Sea continental shelf, J. Phys. Oceanogr., 41(3), 531–547

Lewis H.W., J.M. Castillo Sanchez, A. Arnold, J. Fallmann1, A. Saulter, J. Graham, M. Bush, J. Siddorn, T. Palmer, A. Lock, J. Edwards, L. Bricheno, A. M. de la Torre, J. Clark, (2018) The UKC3 regional coupled environmental prediction system. Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-245

Lewellen, W.S and M. Teske, (1973), Prediction of the Monin-Obukhov similarity functions from an invariant model of turbulence, J.Atmos. Sci., 30, 1340-1345.

Liu Z. Y. and I. Lozovatsky (2012), Upper pycnocline turbulence in the northern South China Sea. Chin Sci Bull, , 57: 23022306, doi: 10.1007/s11434- 012-5137-8.

MacKinnon et al., (2017), Climate Process Team on internal wave-driven ocean mixing. BAMS, November 2017, 2428-2454, *https://doi.org/10.1175/BAMS-D-16-0030.1*

MacLachlan, C., A. Arribas, K.A. Peterson, A. Maidens, D. Fereday, A.A. Scaife, M. Gordon, M., Vellinga, A. Williams, R.E. Comer, J. Camp, P. Xavier, and G. Madec (2015), Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, Q. J. Roy. Meteor. Soc., 141, 1072–1084, https://doi.org/10.1002/qj.2396.

MacWilliams, J.C., E. Hucle, J. Liang, P.P. Sullivan (2014), Langmuir turbulence in Swell, J.Phys Oceanorg, 44, 870-890, doi: 10.1175/JPO-D-13-0122.1

Madec, G. and NEMO Team (2016) , NEMO reference manual 3_6_STABLE: "NEMO ocean engine" Note du Pole de mod_elisation, Inst. Pierre-Simon Laplace (IPSL), France, No 27 ISSN 1288-1619.

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J., and Sinha, B.: GO5.0: the joint NERC–Met Office NEMO global ocean model for use in coupled and forced applications, (2014), Geosci. Model Dev., 7, 1069–1092, https://doi.org/10.5194/gmd-7-1069-2014.

Mellor G.L. and T. Yamada (1974), A hierarchy of turbulence closure models for planetary boundary layers. J. of Atmos. Science, 31, 1791-180.

Mellor, G.L. and T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problems. Reviews of Geophysics and Space Physics 20, 851–875.

Monin, A.S and Ozmidov, R.V (1985) Turbulence in the Ocean, trans from Russsian edited by H. Tennekes, D. Reidel Publishing Co., Dordreht/Boston/Lancaster,

O'Dea, E., R. Furner, S. Wakelin, J. Siddorn, J. While, P. Sykes, R. King, J. Holt and H. Hewitt (2017), The CO5 configuration of the 7 km Atlantic Margin Model: large-scale biases and sensitivity to forcing, physics options and vertical resolution, Geosci. Model Dev., 10, 2947-2969, https://doi.org/10.5194/gmd-10-2947-2017..

Orlanski I.and K. Bryan (1969), Formation of the thermohaline step structure by large-amplitude internal waves, J. Geophys. Res, 74, 6975-6983.

Palmer, M.R., G.R. Stephenson, M.E. Inall, C. Balfour, A. Dusterhus, J.A.M. Green (2015), Turbulence and mixing by internal waves in the Celtic Sea determined from ocean glider microstructure measurements, J. of Marine Systems, 144. 57-69, 10.1016/j.jmarsys.2014.11.005.

Prandtl, L and K. Wieghart (1945) Yeber ein Neues Formel System fur die ausgebieldeto Turbulenz. Nachr. Akad. Wiss., Gottingen, Mathematik-Phys. , Kl. 6-19.

Rascle, N., F. Ardhuin, P. Queffeulou, D. Croizé-Fillon (2008), A global wave parameter database for geophysical applications. Part 1: Wave–current– turbulence interaction parameters for the open ocean based on traditional parameterizations. Ocean Modell. 25, 154–171. http://dx.doi.org/10.1016/ j.ocemod.2008.07.006.

Rippeth T.P., M.E. Inall (2002). Observations of the internal tide and associated mixing across the Malin Shelf, J. Geophys. Res. 107, C3, 3032, doi:101028/2000JC000761.

Rippeth, T. P, M. R. Palmer, J. H. Simpson, and N. R. Fisher (2005),Thermocline mixing in summer stratified continental shelf seas. Geophys. Res. Lett., 32, L05602, doi:10.1029/2004GL022104.

Rippeth, T. P., P. Wiles, M. R. Palmer, J. Sharples, and J. Tweddle (2009), The diapycnal nutrient flux and shear-induced diapycnal mixing in the seasonally stratified western Irish Sea, Cont. Shelf Res., 29, 1580-1587

Rippeth, T. P., B. J. Lincoln, H. A. Kennedy, M. R. Palmer, J. Sharples, and C. A. J. Williams (2014), Impact of vertical mixing on sea surface pCO2 in temperate seasonally stratified shelf seas, J. Geophys. Res. Oceans, 119, 3868–3882, doi:10.1002/2014JC010089

Rotta, J. C. (1951) Statistische theorie nichthomogener Turbulenz. *Z. Phys.*, **129,** 547–572.

Shrira, V.I. and P. Forget (2015), On the Nature of Near-Inertial Oscillations in the Uppermost Part of the Ocean and a Possible Route toward HF Radar Probing of Stratification, J. Phys. Oceanogr., 45, 2660-2678.

Simpson J.H., W.R. Crawford, T.P. Rippeth, A.R. Campbell, J.V. Cheok (1996), The vertical structure of turbulent dissipation in shelf seas, J. Phys. Oceanogr. 26 (8), 1579–1590.

Simpson, J. H., and J. P. Tinker (2009), A test of the influence of tidal stream polarity on the structure of turbulent dissipation, Cont. Shelf Res, 29, 320-332.
Res., 29, 320–332.

Sorbian, Z and Grachev, A.A (2010) An evaluation of the flux –gradient Richardson relationship in the stable boundary layer, Biundary Layer meteorology, 385-405.

Taylor, K.E. (2001), Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**: 7183–7192. doi:10.1029/2000JD900719.

Thomas, H., Y. Bozec, K., Elkalay, H.J. De Baar (2004), Enhanced open ocean storage of CO2 from shelf sea pumping. Science 304 (5673), 1005–1008.

Vlasenko, V., N. Stashchuk, M.E. Inall, J.E., Hopkins (2014), Tidal energy conversion in a global hot spot: On the 3D dynamics of baroclinic tides at the Celtic Sea shelf break, J. Geophys. Res. Oceans 119, 3249-3265.

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

Umlauf, L. and H. Burchard (2005), Second-order turbulence closure models for geophysical boundary layers. A review of recent work, Cont. Shelf Res. 25, 795–827, 2005.

Williams, C. A. J., J. Sharples, M. Green, C. Mahaffey, and T. P. Rippeth (2013a), The maintenance of the subsurface chlorophyll maximum in the western Irish Sea, Limnol. Oceanogr. Methods, 3, 61–73.

Williams, C. A. J., J. Sharples, C. Mahaffey, and T. P. Rippeth (2013b), Wind driven pulses of nutrients to phytoplankton in stratified shelf seas, Geophys. Res. Lett., 40, 5467–5472, doi:10.1002/2013GL058171.

Zhou, B. and Chow, F.K. (2012). Turbulence modelling for the stable atmospheric boundary layer and implication for wind energy, Flow Turbulence Combust (2012) 88:255–277 DOI 10.1007/s10494-011-9359-7

Zikanov, O., D.N. Slinn , M.H. Dhanak (2003), Large-eddy simulations of the wind-induced turbulent Ekman layer, J. Fluid Mech., 495, 343–368. DOI: 10.1017/S0022112003006244.

Table 1. Numerical experiments, the third column describes the closure, and limiting condition (see Appendix A, B), $Ri_{cL(c_l)}$ is the actual steady state Richardson number for the value of limiter applied. The sixth column shows constant used for the roughness parameter.

Run name	Duration years	Closure and $/c_l/$ (Appendix B)	$Ri_{cL(c_l)}$	ERSEM	Surface boundary
GA	1996-2010	GKHR (B6):0.267	**0**.124	no	$\alpha = 1.3$
KC	1996-2010	KC94 (B6):0.267	0.143	no	$\alpha = 1.3$
K3	1996-2010	Kantha,2003 (B6):0.267	0.231	no	$\alpha = 1.3$
CAA	1996-2010	CHCD(A) (B6):0.267	0.25	no	$\alpha = 1.3$
CAB	1996-2010	CHCD(B) (B6):0.267	0.254	no	$\alpha = 1.3$
ST	1990-2015	CHCD(A) (B6):0.267	0.25	no	$\alpha_{ch} = 100000$, ($\alpha \cong 1$)
KCL	1998-2010	KC (B5):0.53	0.193	no	$\alpha = 1.3$
CAL	1998-2010	CHCD(A) (B5):0.53	0.429	no	$\alpha = 1.3$
CAAE	2014-2015	CHCD(B) (B6): 0.267	0.25	yes	$\alpha_{ch} = 100000$, ($\alpha \cong 0.5$)

CABE	2014-2015	CHCD(B) (B6):0.267	0.254	Yes	$\alpha = 1.3$
GAE	2014-2015	GKHR (B6):0.267	0.124	Yes	$\alpha = 1.3$

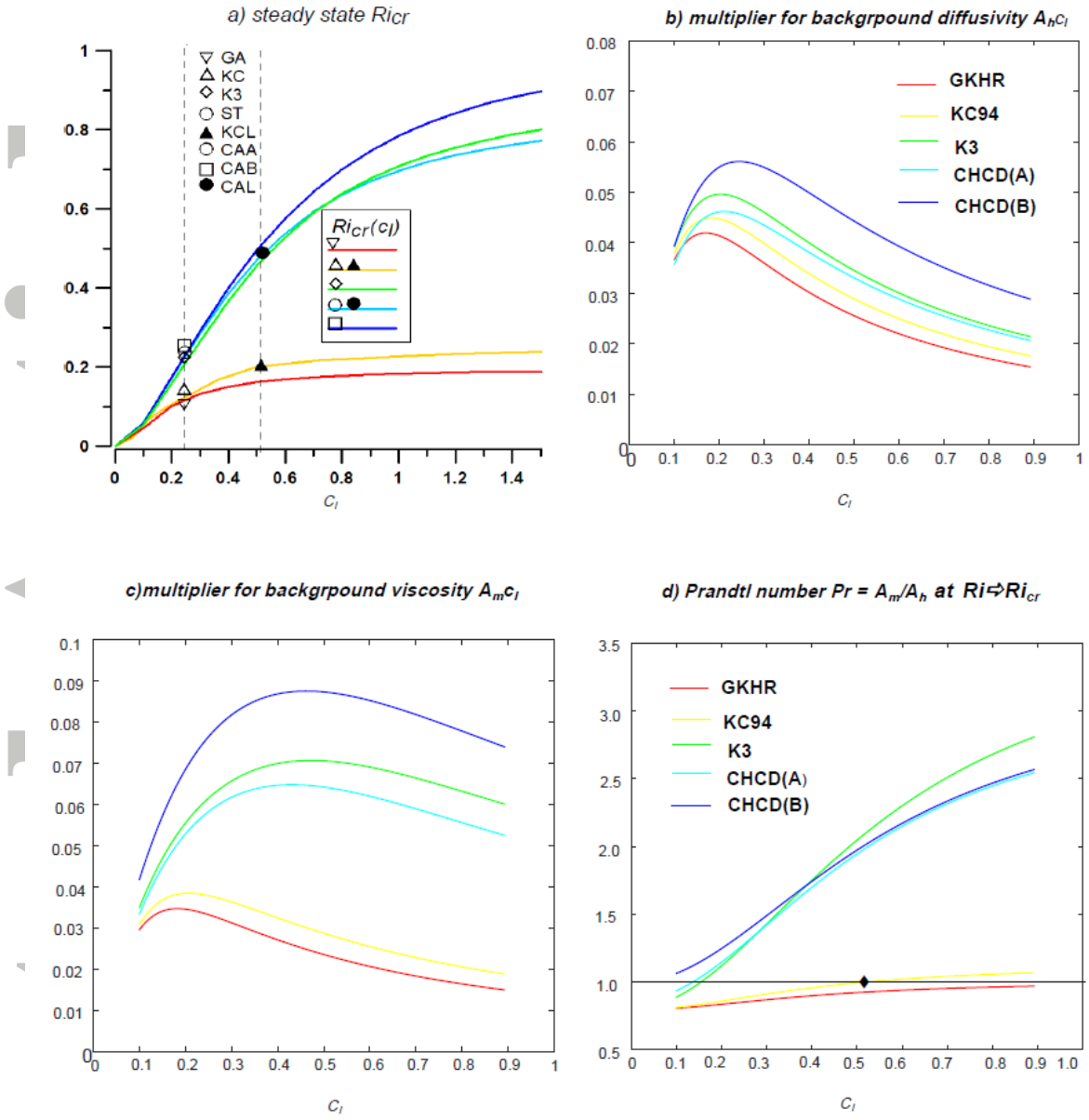

Figure 1. a) Dependence of steady state critical gradient Richardson number Ri_{cr} on closure and Galperin/ Deardorff limiter, b-c) Multipliers for background diffusivity $SA_h c_l$ and viscosity $A_m c_l$ in (10a) for different closures, d) dependence of background Prandtl number on limiter c_l in conditions of $Ri \geq Ri_{cr}$ and steady state. Blue-CHCD (B), magenta CHCD (B), green K3, yellow-KC94 and red GKHR cases.

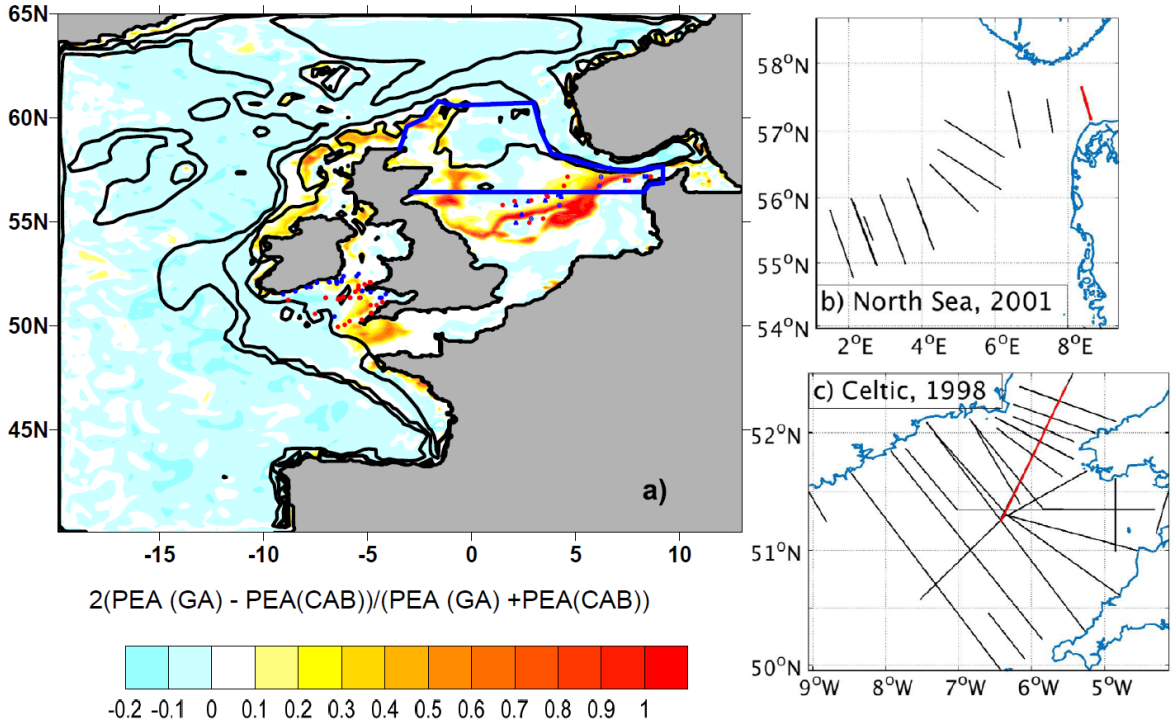

Figure 2. (a) Model domain and locations of SCANFISH surveys. Contours show 100 m, 300 m and 1000 m isobaths. Origins and final locations of the SCANFISH data are shown by blue and red dots. Colours shows the relative difference in potential energy anomaly in July 2001 between the most and least diffusive runs, discussed in section 3. Northern part of the North Sea, discussed in Fig. 3 is shown by blue contour; (b,c) zoomed SCANFISH transects in the North and Celtic Seas. Sections in red indicate the locations of transects, shown in Fig 5. d) Typical stratification profile and its schematics using pycnocline characteristics (i-v) from section 3.3

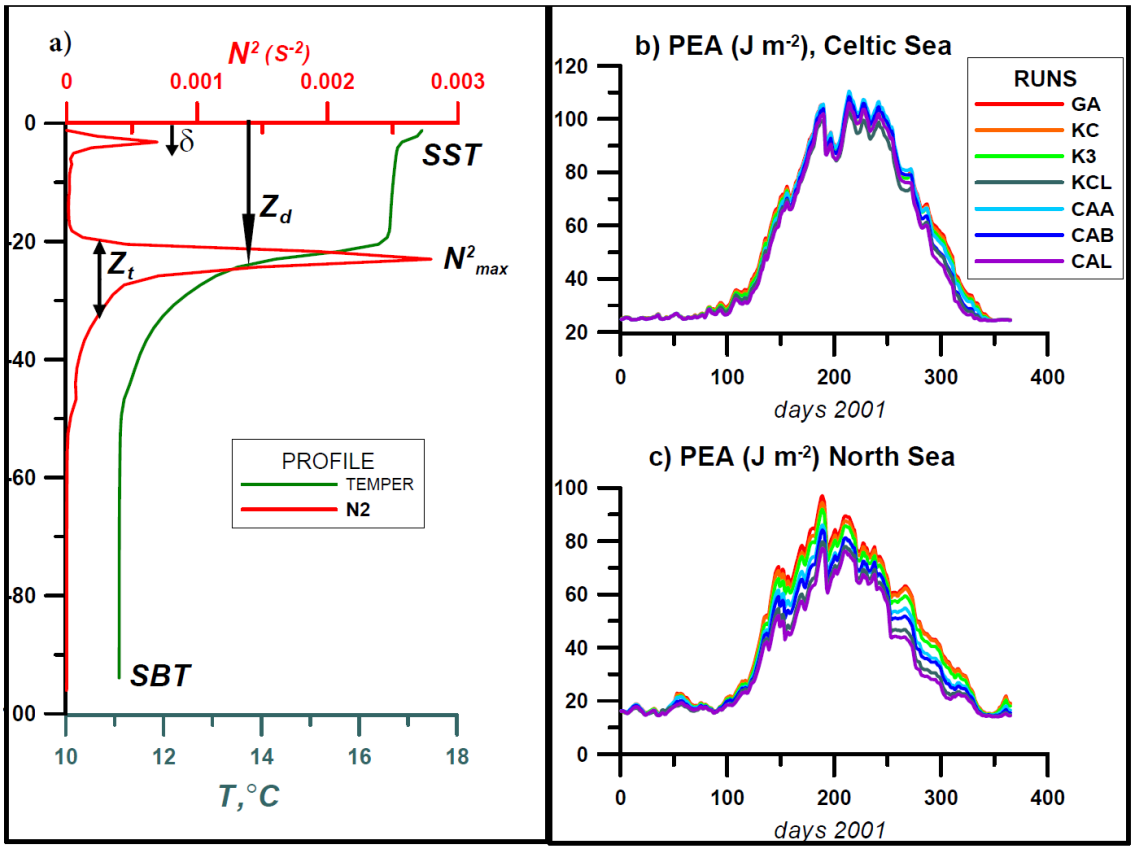

Figure 3. Example of modelled temperature and N^2 profiles in the Celtic Sea and 5 pycnocline characteristics of stratification (section 3.3), used in model assessment, δ is a 'masking' depth used in (14). (b-c) Evolution of shelf-wide mean PEA in 2001 for representative models from Table 1: b) the Celtic Sea and c) the North Sea

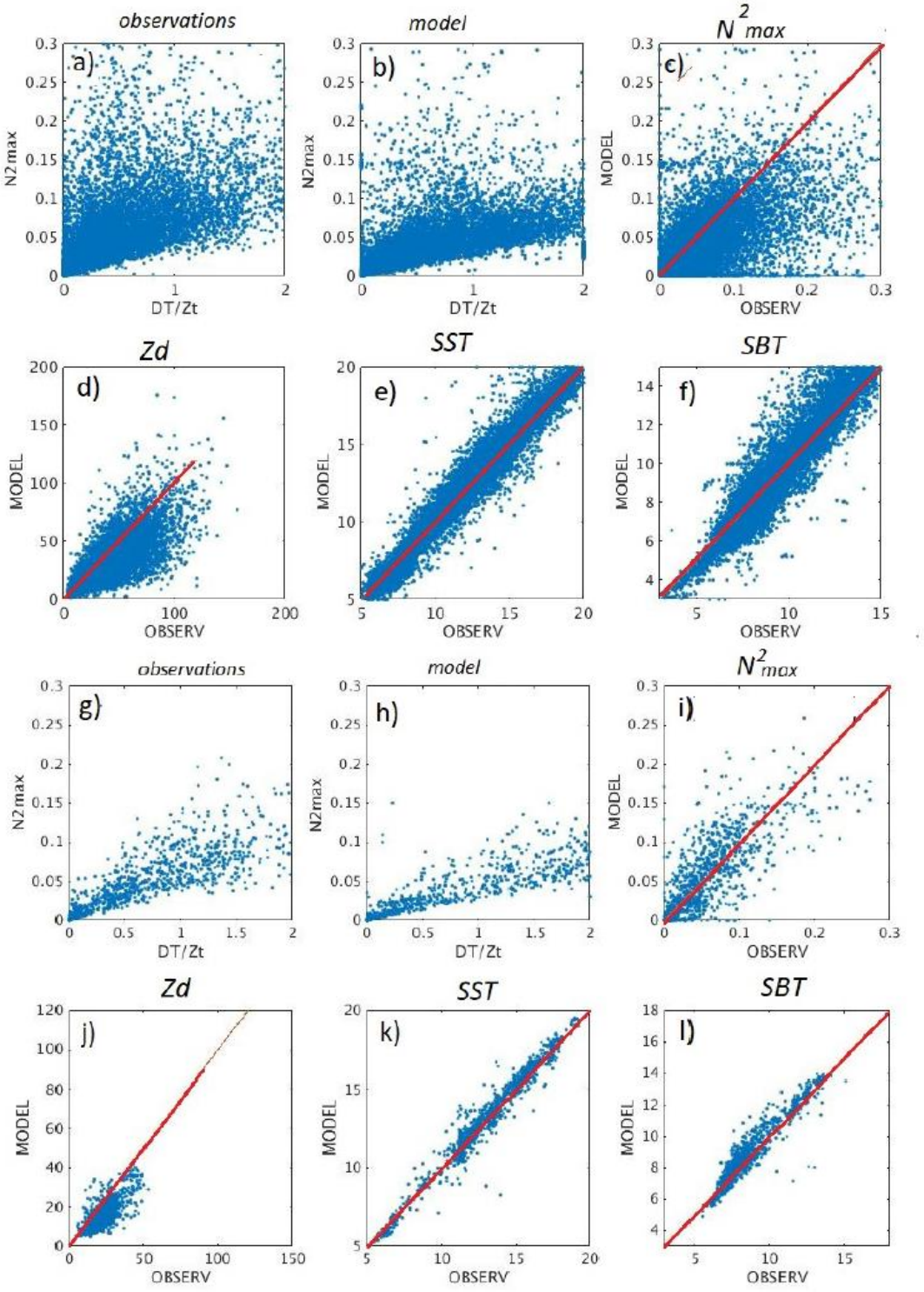

Figure 4. Scatterplots of model variables (i)-(v) versus EN4 observations for K3 closure during 1998-2012, (a)-(f) in whole shelf domain; (g-l) for the Northern part of the North Sea, shown by blue contour in Figure 1.

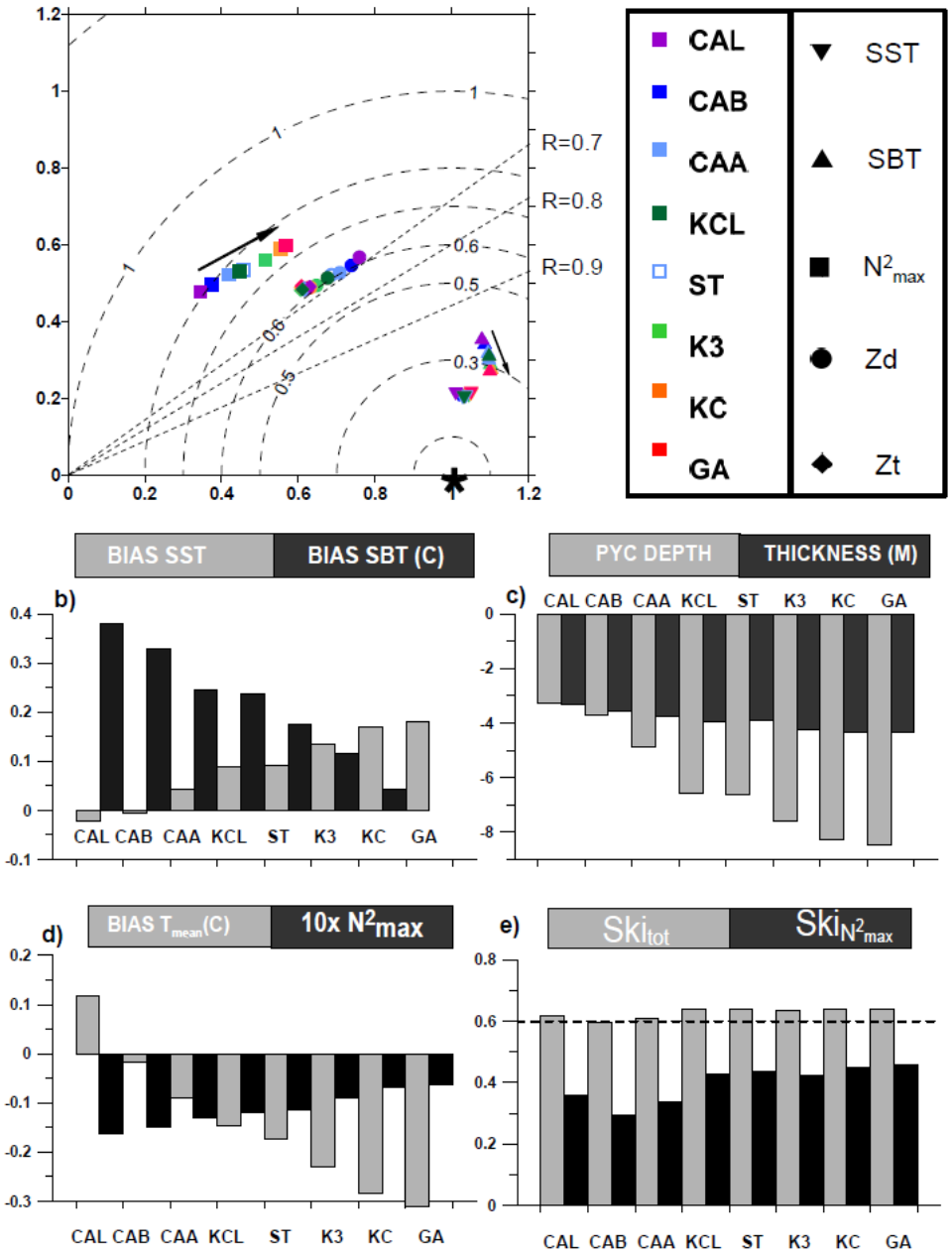

Figure 5. (a) Taylor diagram (after Taylor, (2001)) summarising the relative skill of model variables relative to observations in 1998-2012. Assessed variables (i)-(v) are denoted by varying symbol shapes. Assessed closure models are denoted by symbol colours. For each variable / model pairing the simulated standard deviation (normalised by the observed standard deviation) and the correlation between the model and observations are plotted as a radius and a phase (clockwise from the y-axis) such that the star denotes a perfect simulation with a correlation R = 1, and standard deviation of one, and distance from the star (dashed lines) denotes the model error; Biases in (b) SST and SBT; (c) pycnocline depth Z_d and thickness Z_T; (d) shelf mean temperature and N^2_{max} at pycnocline; (e) model skill evaluated by (12) for N^2_{max} and total skill. Models are order in increasing overall diffusivity, as characterised by the PEA assessment (18).

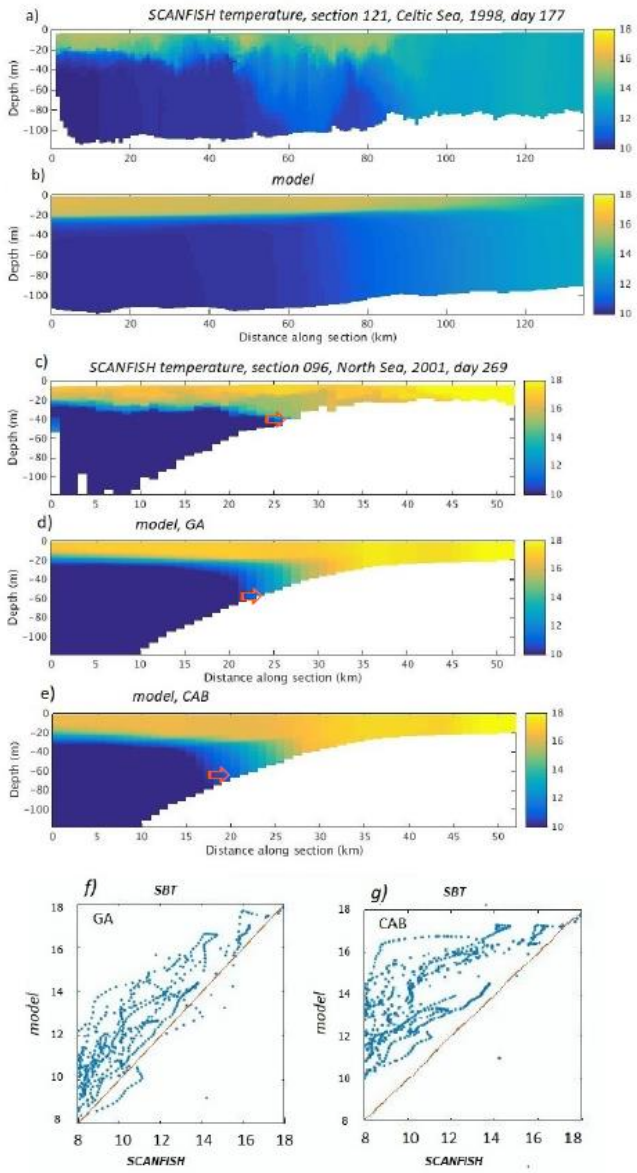

Figure 6. (a-b) Examples of SCANFISH section observations versus the model in the Celtic Sea during July-September 1998. Model data are daily averaged and each datum is taken at the specific date of observation at each location. Temperature across section 121 in observations and in the GA model; (c-e) Examples of SCANFISH section observations versus model in the North Sea in June-August 2001Temperature across section N96 in the North Sea for (c) observations, (d) GA and (e) CAB runs ,(f, g) scatterplots of benthic temperature at both runs in the North Sea (all transects). Red arrows in the figures (c,d) show the width of thermal fronts

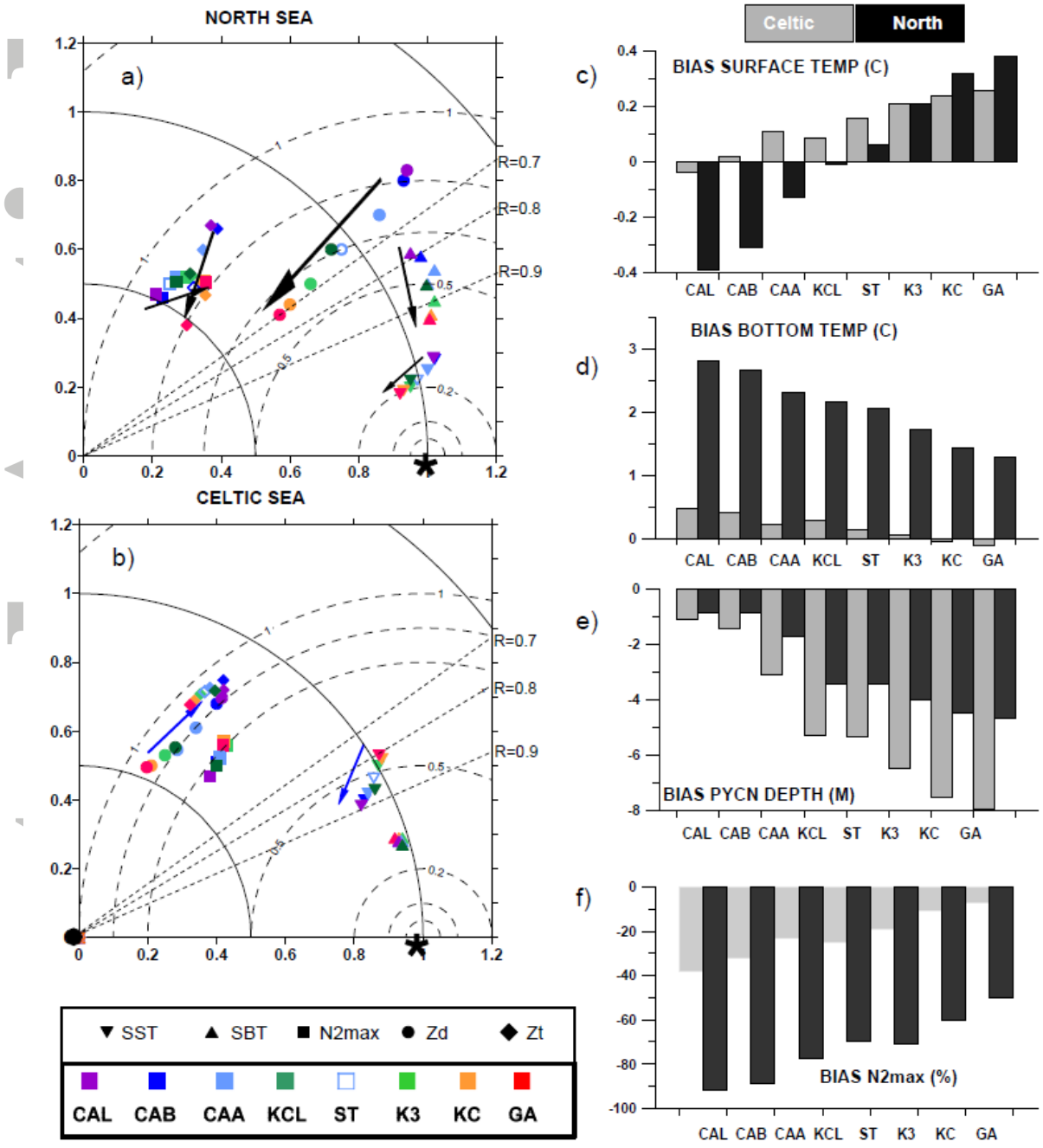

Figure 7. Panels (a-b) show Taylor diagrams for simulated pycnocline characteristics versus SCANFISH surveys (i-v) : Arrows show improvement from model to model. Panels (c-f) show biases for SST, SBT, Z_d and N^2_{max}, with the latter being calculated as a percentage. Grey and black bars denote the Celtic Sea and North Sea, respectively.

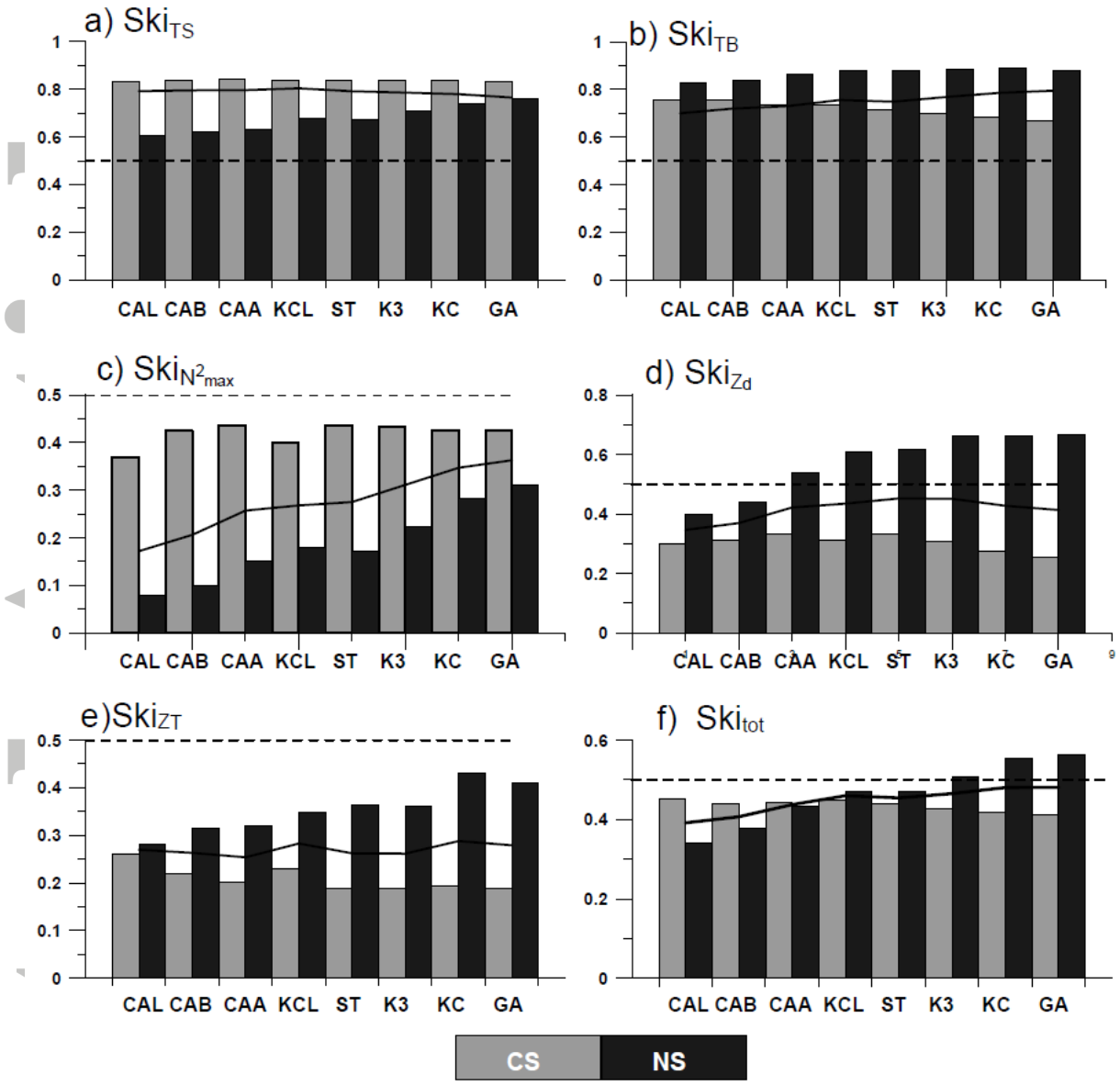

Figure 8. Model skill for assessment for (a) SST, (b) SBT, (c) N^2_{max}, (d) Z_d , (e) Z_T and (f) total skill, based on these 5 variables. Grey- in the Celtic Sea, black- in the North Sea, contour solid line shows combined skills for each SCANFISH survey, dashed line denotes skill threshold of 0.5.

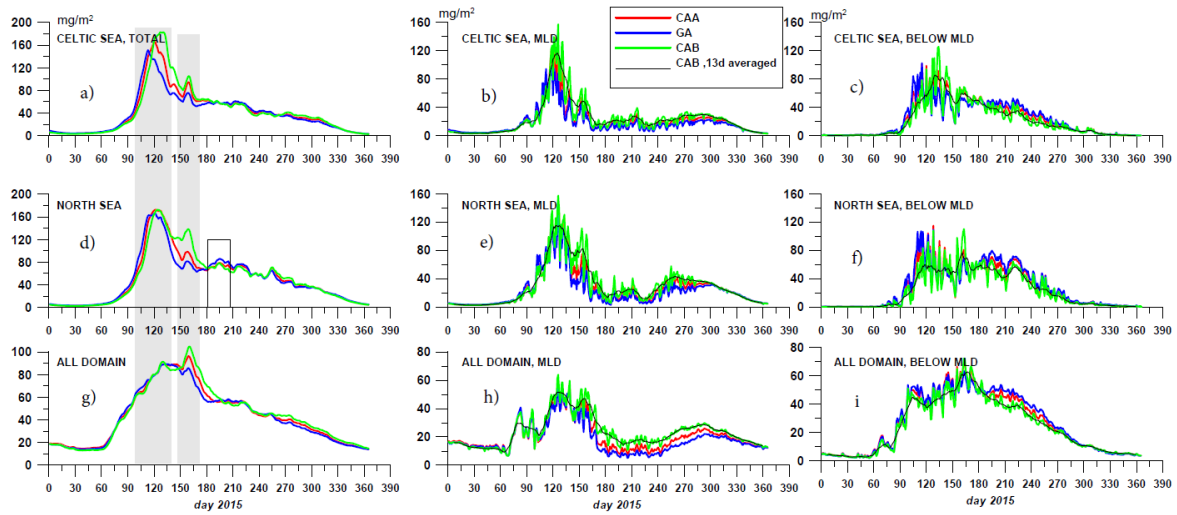

Figure 9. Depth integrated and area mean chlorophyll in 2015. (a,b,c) the Celtic Sea integrated over water column, surface mixed layer only and below mixed layer depth, (d-f) – as (a-c), but for the North Sea, (g-i) for basin mean. Grey colour marks periods of time averaging of chlorophyll fields, shown in Figure 10. Blue: GA, red: ST, green: CAB runs. Black curves in central and right panels are 13 days running means for the CAB run.

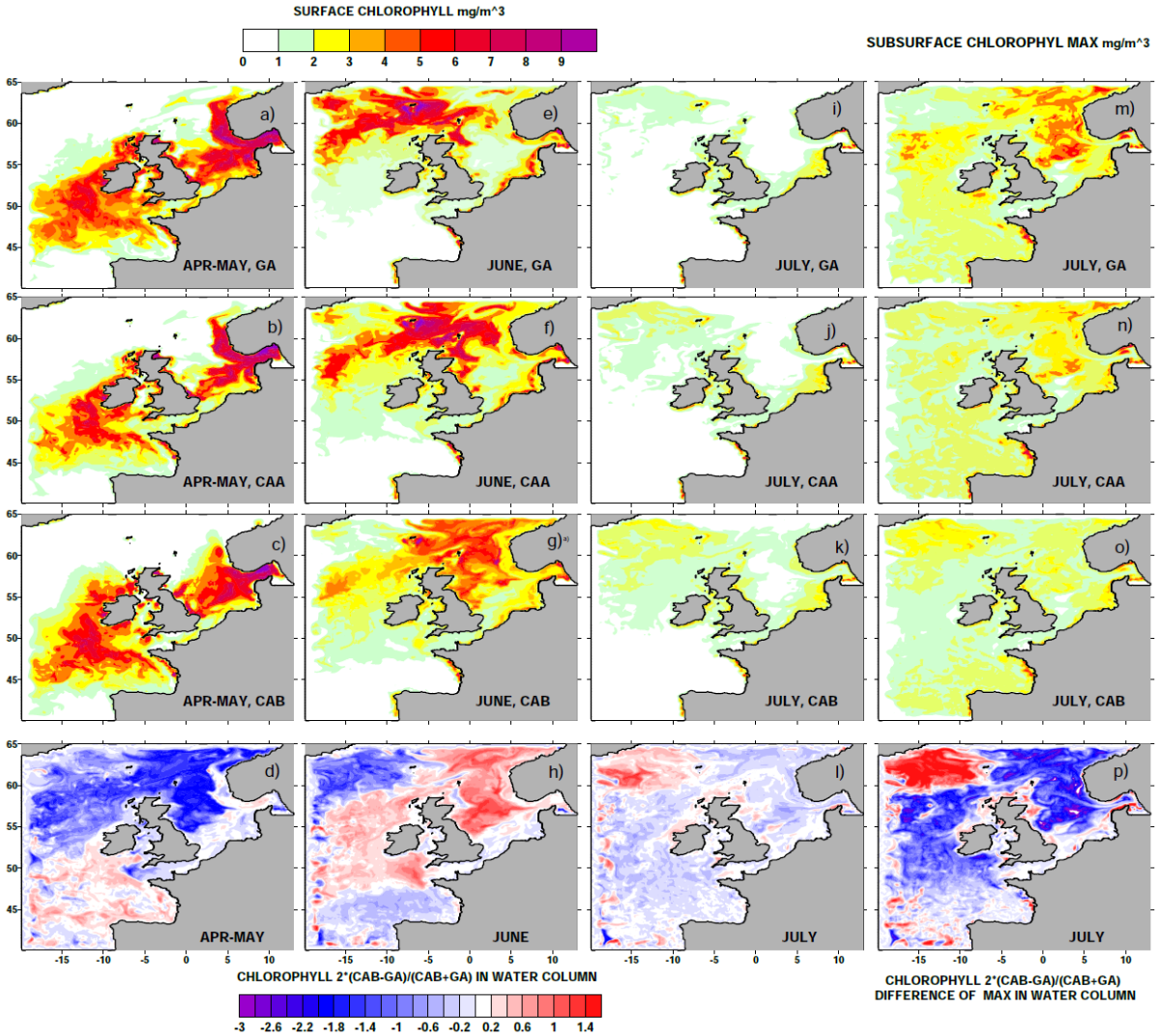

Figure 10. (a-c, e-g, i-k) Surface chlorophyll at the time periods shown in figure 8:April-May, June, July), (d,h),depth integrated chlorophyll differences between CAB and GA for the same period, (m,n,o) maximum of chlorophyll below mixed layer in July, p – relative difference between maximums of subsurface (below ML) chlorophyll in CAB and GA runs .