CrossMark

J̶M̶ PAPERS

Goldilocks mixing in oceanic shear-induced turbulent overturns

A. Mashayek[1], C.P. Caulfield[2,3] and M.H. Alford[4,†]

[1]Department of Civil and Environmental Engineering, Imperial College London, London SW7 2BX, UK

[2]BP Institute for Multiphase Flow, University of Cambridge, Madingley Rise, Madingley Road, Cambridge CB3 0EZ, UK

[3]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, UK

[4]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92037 USA

We present a new, simple and physically motivated parameterization, based on the ratio of Thorpe and Ozmidov scales, for the irreversible turbulent flux coefficient $\Gamma_{\mathcal{M}} = \mathcal{M}/\epsilon$, i.e. the ratio of the irreversible rate $\mathcal{M}$ at which the background potential energy increases in a stratified flow due to macroscopic motions to the dissipation rate of turbulent kinetic energy ϵ. Our parameterization covers all three key phases (crucially, in time) of a shear-induced stratified turbulence life cycle: the initial, 'hot' growing phase, the intermediate energetically forced phase and the final 'cold' fossilization decaying phase. Covering all three phases allows us to highlight the importance of the intermediate one, to which we refer as the 'Goldilocks' phase due to its apparently optimal (and so neither too hot nor too cold, but just right) balance, in which energy transfer from background shear to the turbulent mixing is most efficient. The value of $\Gamma_{\mathcal{M}}$ is close to 1/3 during this phase, which we demonstrate appears to be related to an adjustment towards a critical or marginal Richardson number for sustained turbulence ~ 0.2–0.25. Importantly, although buoyancy effects are still significant at leading order for the turbulent dynamics during this intermediate phase, the marginal balance in the flow ensures that the turbulent mixing of the (density) scalar is nevertheless effectively 'locked' to the turbulent mixing of momentum. We present supporting evidence for our parameterization through comparison with six oceanographic datasets that span various turbulence generation regimes and a wide range of geographical location and depth. Using these observations, we highlight the significance of parameterizing an inherently variable flux coefficient for capturing the turbulent flux associated with rare energetic, yet fundamentally shear-driven (and so not strongly stratified) overturns that make a disproportionate contribution to the total mixing. We also highlight the importance of representation of young turbulent patches in the parameterization for connecting the small scale physics to larger scale applications of

† Email address for correspondence: mashayek@ic.ac.uk

mixing such as ocean circulation and tracer budgets. Shear-induced turbulence is therefore central to irreversible mixing in the world's oceans, apparently even close to the seafloor, and it is critically important to appreciate the inherent time dependence and evolution of mixing events: history matters to mixing.

Key words: stratified flows, mixing and dispersion, ocean processes

1. Introduction

The efficiency of shear-induced turbulent mixing has been widely studied over the past several decades using various observational, experimental, theoretical and computational approaches (Peltier & Caulfield 2003; Ivey, Winters & Koseff 2008; Gregg *et al.* 2018; Caulfield 2020, 2021). Many works have relied on parameterizing the turbulent flux coefficient, Γ, (loosely the ratio of the mixing rate, with various definitions, to the turbulent kinetic energy dissipation rate ϵ) in terms of various non-dimensional parameters such as the buoyancy Reynolds number, $Re_b = \epsilon/(\nu N^2)$, where ν is the kinematic viscosity and N^2 is an appropriate buoyancy frequency, the turbulent Froude number $Fr = \epsilon/[Nk]$, where k is the turbulent kinetic energy (density) or (for sheared turbulence) the Richardson number, $Ri = N^2/S^2$, where S is the vertical shear of some 'background' streamwise velocity. Indeed, some of these approaches are dimensionally insufficient to represent mixing as previously discussed in the literature (e.g. see Ivey & Imberger 1991; Shih *et al.* 2005; Mashayek & Peltier 2011*b*, 2013; Mater & Venayagamoorthy 2014). More specifically, it has been shown that despite an empirically observed emergence of power-law dependence of Γ on Re_b over various turbulent regimes, when observational, experimental and numerical data of Γ are plotted against Re_b, they typically do not overlap, but rather display a wide scatter (Bouffard & Boegman 2013; Mashayek *et al.* 2017*c*; Monismith, Koseff & White 2018). It has been demonstrated that such scatter has leading-order implications for the role of mixing in sustaining the deep branch of ocean circulation (De Lavergne *et al.* 2016; Mashayek *et al.* 2017*c*; Cimoli *et al.* 2019). In fact, recent observational work of Ijichi *et al.* (2020) casts further doubt on parameterizing Γ in terms of Re_b alone.

Of course, such non-dimensional parameters can be interpreted as ratios of length scales, and an alternative but thus clearly related approach to the parameterization of Γ, first proposed by Ivey & Imberger (1991), has been to consider the ratio of the Ozmidov and Thorpe scales, R_{OT}, defined as

$$R_{OT} \equiv L_O/L_T, \qquad (1.1)$$

where $L_O \equiv (\epsilon/N^3)^{1/2}$ and the Thorpe scale L_T is the root mean square (r.m.s.) of the displacements required to reorder a particular profile of density measurements into a monotonically decreasing profile. Use of R_{OT} has been suggested by some as an indication of 'overturn age' and therefore a suitable basis for a parameterization of Γ (Gibson 1987; Ivey & Imberger 1991; Smyth, Moum & Caldwell 2001; Mashayek, Caulfield & Peltier 2017*a*; Ijichi & Hibiya 2018). However, for practical reasons, in physical oceanography it is often operationally assumed that R_{OT} is equal or close to one to infer estimates of rate of dissipation of kinetic energy.

At its heart, this assumption relies on the combination of the ideas that L_T may be thought of as the characteristic height scale of an overturning turbulent event, while L_O is the largest (vertical) scale that is not strongly or (perhaps more accurately) dominantly affected by the background stratification, and that these scales should be similar for a

vigorous turbulent patch. However, this view does not capture the importance of the inherent time dependence of mixing events, and in particular that mixing events evolve through a life cycle, with subsequent phases retaining an imprint of previous stages in the flow evolution. As Villermaux (2019) noted for passive scalar mixing, followed up by Caulfield (2021) in the density-stratified context, history really does matter for mixing, and so it is important to parameterize mixing events throughout their entire life cycle.

In this work we build on more fundamental theoretical grounds, specifically the assumption of the existence of an inertial subrange, mixing length theory and criteria for shear instability, and propose a parameterization for Γ in terms of R_{OT}. Our approach thus has two central pillars. First, as we suppose shear instability drives the mixing, the flows always have a stratification that is not too strong to suppress the growth of an instability to a significant (overturning) amplitude. Here, we refer to such a stratification as subcritical, in the specific sense that the key parameter quantifying the relative strength of the stratification to the shear, the Richardson number (defined more precisely below), is sufficiently small to allow the growth of such a shear-driven overturning instability. It is always important to remember that buoyancy can still play a leading-order role in such subcritical flows, and as we discuss further below, there is accumulating evidence that flows generically adjust so that characteristic values of the Richardson number become close to the critical or marginal value at which instability can (just) occur. Second, the inherent time dependence of such mixing is a distinguishing characteristic which must always be captured by our modelling and/or parameterization.

In the limit $R_{OT} \gg 1$, our parameterization reduces to a scaling relation previously offered by others, however, our inherently time-dependent yet subcritically stratified interpretation relies on understanding this limit as being associated with the late-time decay of an 'old' shear-driven overturning mixing event. In the $R_{OT} \ll 1$ limit, however, the parameterization reduces to a new scaling for 'young' turbulent patches. We argue that neither of the two limits' scaling is correct in isolation, but that their merged form is appropriate since ocean data are known to be distributed near $R_{OT} \sim 1$ (Dillon 1982; Ferron *et al.* 1998; Thorpe 2005; also shown in § 4). Vitally, this intermediate value should still be associated with a subcritical, shear-driven overturning mixing event at an intermediate, and energetic stage in its temporal evolution. We also determine the value of the sole coefficient in the parameterization based on the further, physically and empirically motivated assumptions of turbulent Prandtl number ∼1 and a critical Richardson number around 1/4. These two assumptions are of course at their heart assumptions that the stratification is not too strong to dominate all aspects of the dynamics: the turbulent Prandtl number being around one implies that scalar mixing is (at least approximately) locked to the mixing of momentum; while the very presence of shear-driven instabilities requires a sufficiently low value of the Richardson number.

To support this proposed parameterization, we demonstrate good agreement with several oceanic datasets corresponding to various turbulence types and different depth ranges. The value inferred for the coefficient of the parameterization based on data regression closely matches the theoretical prediction. Thus, our primary finding is that a parameterization based entirely on physical grounds (even up to its sole coefficient) appears to explain the efficiency of mixing of the observed turbulent patches for a range of oceanic turbulent processes, as long as, crucially, the source of the turbulence is background shear, and the conceptual picture is that the turbulence is time developing, evolving through various phases of shear-driven mixing events, inevitably associated with stratification not being too strong.

A. Mashayek, C.P. Caulfield & M.H. Alford

Our theory and observations suggest $\Gamma \sim 1/4 - 1/3$ for most of such shear-driven turbulent data when R_{OT} is within a factor of 3 of unity. Significantly, the mixing coefficient Γ, is predicted to be slightly larger, yet still quite comparable to, the classical value of 0.2. This somewhat more efficient mixing is associated with a phase of energetic turbulence in which the stratification, not yet overly eroded, gives rise to a rich cascade of hydrodynamic instabilities that efficiently channel energy from the parent overturn and the background shear through shear production. Through analogy with the common usage in astrobiology of 'Goldilocks zone' for the circumstellar habitable zone, we refer to this phase as Goldilocks turbulence, as it is neither too hot, nor too cold, but just right.

Less informally, it demonstrates the vital importance of appreciating the time history of shear-driven mixing events, specifically the ensuing efficient mixing occurring after the break down of a relatively large primary shear-driven overturning, which once again, can only arise if the stratification may be considered to be below some critical strength. We also show evidence, based on numerical simulations, that for energetic ocean turbulence (i.e. at sufficiently high Reynolds numbers and sufficiently small Richardson numbers) most of the overturn evolution life cycle corresponds to the Goldilocks phase ($R_{OT} \sim 1$), thereby giving credence to arguments of Gregg (1987) and Caldwell (1983) as opposed to Gibson (1987), who argued that most observations were what he referred to as fossilized turbulence, i.e. the turbulence from late in the life cycle of a mixing event.

Finally, although it is undoubtedly tempting to use an averaged value of Γ in models, we demonstrate that doing so has the potential to lead to large inaccuracies, due perhaps unsurprisingly to a range of inherent nonlinearities in the system. For example, we show that the total buoyancy flux obtained through summation of the contribution of individual patches in various datasets is strongly dominated by those patches with the largest values of the rate of dissipation of kinetic energy (as one would expect) and so it is important to capture as accurate a value as possible of Γ for such patches, rather than using a mean value extracted from data associated with all patches. We also highlight the significance of the large Γ associated with young turbulence for quantification of bulk mixing in a grid cell of a coarse resolution climate model.

To present our key points, the rest of the paper is organized as follows. In § 2, we present definitions of the various important length scales and parameters. We then describe the key aspects of the phenomenology of a shear-driven mixing event in § 3, extracted from a numerical simulation. In § 4, we briefly describe six oceanic datasets, which we then use to verify our proposed parameterization for Γ in terms of R_{OT} in § 5. We also compare and contrast this parameterization with previous studies, in particular those of Maffioli, Brethouwer & Lindborg (2019) and Garanaik & Venayagamoorthy (2019) based around the use of the turbulent Froude number. Specifically, we highlight the interesting fact that similar scalings can arise based on conceptually different physical interpretations, not relying on the inherent time dependence and not particularly strong stratification at the heart of our interpretation of shear-driven overturning mixing events. We highlight three key implications of our parameterization in § 6. In § 7 we argue that, while observationally desirable, parameterization of Γ based on Re_b or (Re_b, Ri) might be impractical. We draw our conclusions in § 8, and suggest some future avenues of research.

2. Basic definitions

The basic turbulence scales that we employ herein are the Kolmogorov scale, L_K, representing the scale below which viscous dissipation takes kinetic energy out of the system, the Ozmidov scale, L_O, the maximum (vertical) scale that is not strongly affected by stratification, the Corrsin scale, L_C, the maximum scale that is not strongly

affected by the background shear, and finally the Thorpe scale, L_T, a geometrical vertical scale characteristic of displacement of notional fluid parcels within an overturning turbulent patch. The first three may be defined as

$$L_K = \left(\frac{\nu^3}{\epsilon}\right)^{1/4}, \quad L_O = \left(\frac{\epsilon}{N^3}\right)^{1/2}, \quad L_C = \left(\frac{\epsilon}{S^3}\right)^{1/2}. \qquad (2.1a\text{--}c)$$

It is important to remember that the calculation of characteristic values of ϵ, N and S for a given flow must always involve some averaging, whether over ensembles, space or time.

Note also that while we conveniently will refer to L_T as a turbulence length scale, it is more appropriate to think of it as a geometric, and somewhat subjective property of an identified 'patch' of turbulence. (Dillon 1982; Thorpe 2005; Chalamalla & Sarkar 2015; Mater *et al.* 2015; Mashayek *et al.* 2017*a*). Nevertheless, it has proved to be an important measure which can be readily inferred from observations. Once such length scales have been defined, their ratios lead naturally to various non-dimensional parameters. In particular, definitions of a Richardson number and a buoyancy Reynolds number can be written as

$$Ri = \left(\frac{L_C}{L_O}\right)^{2/3} = \frac{N^2}{S^2}, \quad Re_b = \left(\frac{L_O}{L_K}\right)^{4/3} = \frac{\epsilon}{\nu N^2}, \qquad (2.2a,b)$$

where here we further assume that the background shear S is the vertical shear of horizontal velocity.

In this paper, we are interested in parameterizing various properties of turbulent mixing, in particular its 'efficiency'. Here, we choose to define the instantaneous efficiency of turbulent mixing in terms of a ratio of conversion rates: i.e. the 'mixing' rate $\mathcal{M}$ at which the minimal background potential energy is irreversibly increasing due to macroscopic fluid motions (Winters *et al.* 1995; Caulfield & Peltier 2000) divided by $\mathcal{M} + \epsilon$, the rate at which kinetic energy is being irreversibly lost

$$\mathcal{E}_i = \frac{\mathcal{M}}{(\mathcal{M} + \epsilon)}. \qquad (2.3)$$

As expected for an 'efficiency', $\mathcal{E}_i \leqslant 1$ strictly. A corresponding flux coefficient (often somewhat confusingly referred to as an efficiency, although in principle $\mathcal{M} > \epsilon$ is possible) may then be defined as

$$\Gamma_{\mathcal{M}} = \frac{\mathcal{E}_i}{1 - \mathcal{E}_i} = \frac{\mathcal{M}}{\epsilon}, \qquad (2.4)$$

where the subscript $\mathcal{M}$ makes explicit that this definition of the turbulent flux coefficient is in terms of the irreversible mixing rate. (As discussed in recent reviews by Gregg *et al.* (2018) and Caulfield (2020, 2021), there are several different possible definitions for this flux coefficient, and it is important to be clear which particular definition is being utilized when comparing different data sources.)

Generically, there is no reason to suppose that $\Gamma_{\mathcal{M}}$ is constant. As originally argued by Osborn (1980), an understanding of the properties of $\Gamma_{\mathcal{M}}$ can lead to a parameterization for the (vertical) eddy or turbulent diffusivity of buoyancy κ_T, defined as

$$\kappa_T \equiv -\frac{\mathcal{B}}{N^2} = -\frac{\langle w'b' \rangle}{N^2}, \qquad (2.5)$$

where w' and b' are turbulent velocity and buoyancy perturbations, and hence $\mathcal{B}$ is an (appropriately averaged) vertical buoyancy flux. Assuming that transport terms, reversible

processes and spatio-temporal variability in general can be ignored (see, for example, Mashayek, Caulfield & Peltier (2013) for further discussion) $\mathcal{B}$ can be approximated by $\mathcal{M}$ in many situations of interest, and so the enhanced turbulent diffusivity is given by

$$\frac{\kappa_T}{\kappa} \equiv -\frac{\mathcal{B}}{\kappa N^2} \approx \Gamma_{\mathcal{M}} \left(\frac{\nu}{\kappa}\right) \left(\frac{\epsilon}{\nu N^2}\right) = \Gamma_{\mathcal{M}} Pr Re_b, \tag{2.6}$$

where κ is the (molecular) diffusivity of the buoyancy field, and so Pr is a (molecular) Prandtl number.

3. Phenomenology of shear-driven mixing events

For the construction of a physically motivated mixing parameterization, it is useful to consider the time evolution of various length scales and the flux coefficient during the turbulence life cycle of a transient, shear-driven mixing event, as a particularly simple model of a breaking wave. To this end, in figure 1 we describe the evolution of turbulence due to the breakdown of the canonical shear instability referred to as the Kelvin–Helmholtz instability (KHI). The KHI paradigm has been commonly suggested to be relevant to oceanic turbulent overturning mixing events (Smyth *et al.* 2001; Mater *et al.* 2015); we will provide further support towards this idea throughout this paper.

Figure 1(*a*) illustrates growth of a Kelvin–Helmholtz 'billow', its turbulence transition, breakdown and relaminarization, from a simulation of a flow with initial hyperbolic tangent velocity and buoyancy distributions

$$\boldsymbol{u} = U(\tilde{z})\hat{\boldsymbol{x}}; \quad U(\tilde{z}) = \frac{U_0}{\tanh}\left(\frac{z}{d_0}\right); \quad b = b_0 \tanh\left(\frac{z}{\delta_0}\right), \tag{3.1a-c}$$

$$Re_0 \equiv \frac{U_0 d_0}{\nu}; \quad Ri_b \equiv \frac{b_0 d_0}{U_0^2}; \quad R \equiv \frac{d_0}{\delta_0}, \tag{3.2a-c}$$

defining the appropriate flow Reynolds number Re_0, 'bulk' Richardson number Ri_b and scale ratio R; $\hat{\boldsymbol{x}}$ is the unit vector in the streamwise x-direction, and tildes denote dimensionless quantities. For this particular simulation (see Mashayek *et al.* (2013) for details) $Re_0 = 6000$, $Ri_b = 0.16$, $R = 1$ and $Pr = 1$. In the sense discussed in § 1, this stratification is definitely subcritical, in that shear-driven overturning instabilities are able to grow to finite amplitude, and indeed to break down to turbulence.

Physical and quantitative interpretation of the associated mixing may be achieved through consideration of the various turbulence scales shown in figure 1(*b*). There are three key phases which occur sequentially during the life cycle of a generic shear-driven mixing event: an early growing young or 'hot' phase; an intermediate, energetic and efficient mixing phase; and an ultimate decaying or 'cold' phase. As we demonstrate, the intermediate phase appears to lead to 'optimal' mixing, neither too hot nor too cold but 'just right', so as already mentioned in § 1, we refer to this as the Goldilocks mixing phase, inspired by the fairy tale. In what follows we will describe these three phases by discussing the evolution of the relevant above-defined turbulent length scales. The specific methodology for calculating the scales from direct numerical simulations (DNS) were discussed in Mashayek *et al.* (2017*a*).

By dimensionless time $\tilde{t} = 79$ (scaled with the advective time scale d_0/U_0), the primary overturn has fully grown, as is marked by the peaking of L_T. This represents an accumulation of available potential energy (APE) from the kinetic energy (KE) reservoir, originally stored in the background shear. Upon the saturation of the primary billow, small scale turbulence grows, effectively through feeding on the APE. This leads to a significant
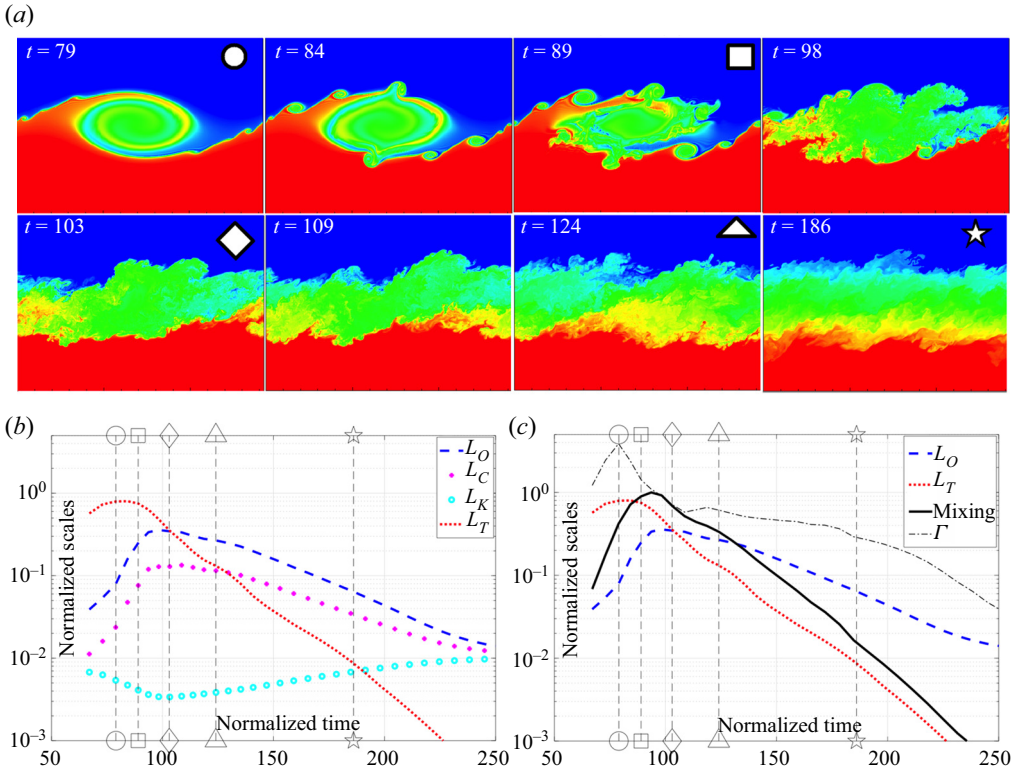
(*a*)

(*b*) (*c*)

Figure 1. (*a*) Turbulent life cycle of a shear-driven KHI. Colours represent density, with red, green and blue representing dense, intermediate and light waters, respectively. (*b*) Evolution of the various (non-dimensional) length scales defined in § 2, scaled with the time-evolving half-depth of the shear layer (following Mashayek *et al.* 2017*a*). (*c*) Evolution of L_T and L_O as well as the non-dimensional irreversible mixing rate $\tilde{\mathcal{M}}$ (scaled with U_0^3/d_0) and turbulent flux coefficient $\Gamma_{\mathcal{M}}$. The vertical dashed lines in (*b,c*) mark various characteristic times that correspond to different images in (*a*) with similar symbol markings. These plots are reproduced from the data associated with a simulation with $Re_0 = 6000$, $Ri_b = 0.16$, $Pr = 1$, $R = 1$ from Mashayek *et al.* (2013).

increase in the rate of dissipation of turbulent kinetic energy (TKE), leading to a relatively rapid growth of the Ozmidov and Corrsin scales and a decrease in the Kolmogorov scale, together representing a widening of the inertial subrange within which energy can be transferred from the injection scale to dissipation scale through a cascade of eddies, largely unaffected by either stratification or shear.

A critical time in the turbulence life cycle occurs when the characteristic scales of the turbulence approach those associated with the initial vertical scale of the primary overturn, i.e. when $L_O \sim L_T$. Under the (reasonable) assumption that it is appropriate to think of L_T as the injection scale for the turbulent motions, due to the implied conversion of APE to KE at this scale, then this critical time marks the instant at which the broadest possible range of scales largely unaffected by stratification occurs, and so there is an opportunity for efficient extraction of energy from the background (see Mashayek *et al.* (2017*a*), and references therein for a discussion). Unsurprisingly, $L_C < L_O$, as it is to be expected that the characteristic Richardson number, defined in (2.2*a,b*) in terms of the length scales, $Ri < 1$. Therefore, although the scales below L_O are largely unaffected by the stratification, there is typically a range of scales still strongly affected by the background shear,

928 A1-7

which nevertheless appear to still allow efficient mixing processes to occur where it is appropriate to think of the (dynamical) influence of the stratification as being perhaps important, but definitely not dominant.

After this critical time, APE is depleted rapidly, as manifested through the relatively sharp drop in L_T; L_O and L_C also decay, albeit markedly more slowly than L_T. We can interpret this behaviour through remembering that L_O and L_C represent the largest scales that the turbulent eddies notionally could have accessed had they been sufficiently 'fed' energetically. However, since the turbulence has decayed so much, the overturning (and presumably injection) scale L_T has become smaller than both L_O and L_C. This empirically observed decay lag between L_O and L_T is of significance for the theoretical arguments that we make in the following sections. Crucially, this decay should not be interpreted as anisotropic turbulence collapse due to 'strong' stratification, but rather primarily due to kinetic energetic losses to viscous dissipation, and, to a lesser extent, to irreversible mixing increasing the potential energy of the system. Ultimately, the flow enters a (molecular) diffusion-dominated regime, as diffusivity drops back to molecular values and so $L_O \rightarrow L_K$, as is shown as $\tilde{t} \rightarrow 250$ in (*b*).

Figure 1(*c*) shows the evolution of the turbulent flux coefficient, $\Gamma_{\mathcal{M}}$, and the (non-dimensional) irreversible mixing rate $\tilde{\mathcal{M}}$, i.e. scaled with the characteristic scales U_0^3/d_0. As would be expected based on the evolution of scales in (*b*), mixing is most efficient between the peak of L_T (and so when APE is largest) and the 'critical' time when $L_T \approx L_O$. This reflects the richness of the zoo of hydrodynamic instabilities within the broad inertial subrange and their ability to stir and hence irreversibly mix density (and tracer) gradients efficiently (see, for example, Mashayek & Peltier (2012*a*,*b*) for a detailed analysis of the link between the cascade of hydrodynamic instabilities that facilitate turbulence breakdown and the characteristics of turbulent mixing). The subsequent decay of the mixing rate is highly correlated with the decay of its (driving) source of energy (i.e. APE), and so is highly correlated with the decay of L_T. From the time APE saturates onwards, L_T/L_O decays essentially monotonically since L_T consistently decays (relatively rapidly) and L_O first grows and then decays at a slower rate. For this reason, and as we will discuss further, this ratio has been suggested to be a good proxy for evolution of mixing during turbulent life cycles in shear flows (Smyth *et al.* 2001; Mashayek *et al.* 2017*a*). This is a reasonable suggestion, as is apparent from the apparent correlation between the ratio L_T/L_O and $\Gamma_{\mathcal{M}}$.

4. Data

We now consider six oceanic datasets to constrain and inform our proposed parameterizations for Γ, principally in terms of R_{OT}. All six were collected from free-falling microstructure profilers, which sample shear and temperature with sub-centimetre resolution in order to estimate the dissipation rate of KE, ϵ, and the thermal variance dissipation rate, χ. Each also carries a conductivity–temperature–depth instrument in order to sample the profiles of temperature and salinity needed to estimate potential density and buoyancy frequency N. Ozmidov scales are then directly calculated from ϵ and N, while Thorpe scales are calculated following Dillon (1982) by re-ordering the measured potential density, and evaluating the r.m.s. of the required displacements of individual measurements, which is essentially the one-dimensional observational analogue of the approach used by Winters *et al.* (1995).

Following the methodology proposed by Moum (1996), a turbulent flux coefficient Γ_{χ} is computed from ϵ and an appropriately scaled version of the thermal dissipation rate,

χ, corresponding to the destruction rate of buoyancy variance, or equivalently, under the further assumption that the characteristic buoyancy frequency is constant, the destruction rate of available potential energy

$$\Gamma_\chi \equiv \frac{\chi}{\epsilon}. \tag{4.1}$$

As discussed further in recent reviews (Gregg *et al.* 2018; Caulfield 2020, 2021), under certain circumstances it is reasonable to suppose that $\Gamma_\mathcal{M} \simeq \Gamma_\chi$, an assumption we make here when comparing observational and numerical simulation data.

The Tropical Instability Wave Experiment (TIWE) dataset includes turbulent patches sampled at the equator at 140°W in the shear-dominated upper-equatorial thermocline, between 60 m and 200 m depths, spanning both the upper and lower flanks of the Pacific Equatorial Undercurrent (Lien *et al.* 1995; Smyth *et al.* 2001). The FLUX STAT (FLX91) experiment sampled turbulence at the thermocline ($\sim$350–500 m depth), in part generated through shear arising from downward-propagating near-inertial waves, approximately 1000 km off the coast of northern California (Moum 1996; Smyth *et al.* 2001). The IH18 experiment measured full-depth turbulence (up to $\sim$5300 m deep) primarily generated by tidal flow over the Izu-Ogasawara Ridge (western Pacific, south of Japan), a prominent generation site of the semidiurnal internal tide that spans the critical latitude of 28.88N for parametric subharmonic instability (Ijichi & Hibiya 2018). The Samoan Passage data are measurements of abyssal turbulence generated by hydraulically controlled flow over sills in the depth range 4500–5500 m in the Samoan Passage, an important topographic constriction in the deep limb of the Pacific Meridional Overturning Circulation (see Alford *et al.* (2013) and Carter *et al.* (2019), we use data from the latter). The BBTRE data are from turbulence induced by internal tide shear in the deep Brazil Basin ($\sim$2500–5000 m depth) and were acquired as a part of the original Brazil Basin Tracer Release Expermient (BBTRE; Polzin *et al.* 1997), recently re-analysed by Ijichi *et al.* (2020). Also re-analysed by Ijichi *et al.* (2020), we use the data from DoMORE which focused on flow over a sill on a canyon floor in the Brazil Basin (Clément, Thurnherr & St. Laurent 2017; Ijichi *et al.* 2020). In total, these datasets comprise nearly 50 000 patches and six different turbulence generation regimes over a wide range of depth and geographical locations.

5. Parameterizing $\Gamma_\mathcal{M}$ as a function of L_O/L_T

Building on the phenomenology of the overturning life cycle illustrated in figure 1, in this section we propose a functional dependence of $\Gamma_\mathcal{M}$ on the key ratio of length scales, which has historically been defined as $R_{OT} \equiv L_O/L_T$, as in (1.1). Although the canonical example shown in figure 1 was of the breakdown of a KHI, for our following arguments to hold, all that is needed is the notion of an overturning-induced turbulence (not necessarily a classical shear instability) that comprises a growing phase, an energetically mixing phase and a decaying phase. As we argue in more detail below, this decaying phase can be considered, at least in some sense, as a 'fossilizing' phase of turbulence, with some points of analogy with the classical arguments of Gibson (1987). Importantly, underlying this approach is the concept that the stratification is not too strong, so that in can be considered to be subcritical, in the very specific sense that the developing turbulence is not itself thoroughly dominated by the stratification, but rather that vigorous overturnings have managed to develop. This concept can be supported by identification of spatio-temporally localized turbulent patches within flows where the stratification, although possibly dynamically important, is not, at least locally, dominant, while in bulk

terms, the larger scale flow can be classified as being strongly stratified, as demonstrated by Portwood *et al.* (2016).

5.1. *Length scale ratio for $\Gamma_{\mathcal{M}}$*

First, we assume that the largest energy containing eddies of length L_i inject energy into the system and that such an input is ultimately balanced by an energy sink at the viscous dissipation scale after a (net) forward cascade. This leads to the classic turbulent (and inherently unstratified) integral scale,

$$L_i \propto \frac{k^{3/2}}{\epsilon}, \tag{5.1}$$

where k is the TKE. This scaling has at its heart that the turbulence being considered is not strongly affected by stratification.

Next, we invoke classic mixing length theory (Taylor 1915; Prandtl 1925), which gives the turbulent diffusivity in terms of a mixing length L_m and k

$$\kappa_T \propto L_m \, k^{1/2}. \tag{5.2}$$

Here, the (possibly significant) dynamical influence of ambient stratification may be embedded within the mixing length, but it also contains an implicit assumption that the mixing of the buoyancy is assumed to be set by (and so in a sense locked to) the mixing of momentum, as quantified by the right-hand side of the expression. Equivalently, it is assumed that the turbulent Prandtl number $Pr_T \sim O(1)$, defined as

$$Pr_T \equiv \frac{\nu_T}{\kappa_T}; \quad \nu_T \equiv \frac{\langle u'w' \rangle}{S} \simeq \frac{P}{S^2}, \tag{5.3a,b}$$

where ν_T is the eddy diffusivity (of momentum), and we assume that the turbulent production P is dominated by Reynolds stress extraction from the mean shear S. Therefore, it is most definitely not appropriate to think of the flow as 'strongly' stratified in any meaningful sense.

Also, by its very nature 'overturning' mixing must be occurring in (at least locally) a subcritical stratification as observed by Alford & Pinkel (2000). Such overturning mixing is qualitatively different from the 'scouring' mixing in the vicinity of 'sharp' and strong density interfaces, a classification distinction drawn by Woods *et al.* (2010) (see Caulfield (2020, 2021) for more discussion).

Finally, we invoke the Osborn balance presented previously in (2.6):

$$\kappa_T \approx \Gamma_{\mathcal{M}} \frac{\epsilon}{N^2}. \tag{5.4}$$

These three ingredients can form the basis for a simple parameterization that fits all the data introduced earlier, provided various data are thought of as being sampled at different stages in the time evolution of a particular mixing event. To achieve this, however, it is also important to distinguish between the two characteristic length scales defined in (5.1) and (5.2). Remembering that ϵ and N can in turn be related to the Ozmidov length scale since $L_O^2 = \epsilon/N^3$, it is possible to use these equations to arrive at

$$\Gamma_{\mathcal{M}} \propto \left[\frac{L_i L_m^3}{L_O^4} \right]^{1/3}. \tag{5.5}$$

To develop a parameterization, it is now necessary to identify appropriate candidate lengths for L_m and L_i that can actually be quantified. Since here we are interested in

928 A1-10

overturns far from boundaries, we expect that the distance to the boundary is not a relevant candidate. Motivated by the phenomenology discussed in § 3, we argue that there are different appropriate candidates for these scales at different stages of the flow evolution of an overturn in a sufficiently weakly stratified flow.

5.2. *'Young' turbulence scaling for $\Gamma_{\mathcal{M}}$*

During the first growing phase of the turbulence, $L_T \gg L_O$. During that phase, it is appropriate to identify the mixing length $L_m \sim L_T$, as the Thorpe scale is the relevant scale over which the fluid gets stirred and mixed against the background stratification. On the other hand, the actual turbulent motions (and the associated inertial range for the forward cascade of energy) are expected to be constrained to lengths bounded above by L_O, and so a reasonable estimate for L_i should be L_O. Therefore, (5.5) reduces to

$$\Gamma_{\mathcal{M}} \propto R_{OT}^{-1} \tag{5.6}$$

for this first 'young' turbulence phase.

5.3. *'Fossilization' scaling for $\Gamma_{\mathcal{M}}$*

Conversely, during the final decaying or 'fossilization' phase when $L_T \ll L_O$, it seems reasonable that the mixing length can still be identified with the overturning scale, so $L_m \sim L_T$. In this stage, L_T can also be closely related to the integral scale of the turbulence, L_i, as the energy for turbulence and mixing is injected by the remaining, and smaller scale than earlier in the flow evolution, overturnings; L_O may now be interpreted as the largest eddies which could in principle be present within the given ambient stratification. However, for such eddies actually to arise, energy has to be injected in to the system at a sufficient rate, and since L_T has dropped below L_O, there is no longer a mechanism by which that injection can occur. Crucially, this should not be interpreted as 'strong' stratification suppressing (in particular) vertical motions inducing strongly stratified anisotropic flow, but rather that turbulent dissipation has converted a large proportion of the KE, and there is (at the moment) no forcing or energy injection mechanism occurring. Such a gap between the actual eddy length scale (scaling with L_T) and the notionally possible eddy length scale (scaling with L_O, if only the flow were energetic enough) is one of the hallmarks of 'fossil turbulence' (Gibson 1987). It is important to remember that both L_T and L_O decrease with time, but L_T decreases more rapidly, and so the L_O is 'left behind' by the more rapid decay of the energy injection scale associated with L_T.

Thus, in the final fossilization phase, (5.5) reduces to

$$\Gamma_{\mathcal{M}} \propto R_{OT}^{-4/3}, \tag{5.7}$$

where once again it is important to remember that this argument is fundamentally associated with the stratification always being below some critical, threshold strength, not least because of the implicit assumption that buoyancy and momentum are mixed on the same scales and by the same processes. In summary, we thus suppose that (5.6) holds in the $R_{OT} \ll 1$ limit, corresponding to the first growing 'young turbulence phase', while (5.7) holds in the $R_{OT} \gg 1$ limit, corresponding to the final decaying or 'fossilization phase', with the variation in the dependence of the flux coefficient with R_{OT} being associated with different (temporal) phases in the evolution of the overturning stratified mixing event, where buoyancy forces never dominate the dynamics.

5.4. *Goldilocks scaling for $\Gamma_{\mathcal{M}}$*

As previously reported in observations (e.g. see Dillon 1982; Ferron *et al.* 1998), and further confirmed by our analyses of these oceanographic datasets, most of the observed overturns actually exist around $R_{OT} \sim 1$, which, from § 3, occurs during the intermediate, energetically mixing phase. As pointed out in Mashayek *et al.* (2017*a*), this also actually turns out to be a regime of efficient (indeed apparently optimal) mixing, which we refer to as Goldilocks mixing. To achieve a parameterization for this regime, we simply combine (5.6) and (5.7) using the following formula:

$$\Gamma_{\mathcal{M}} = A \frac{R_{OT}^{-1}}{1 + R_{OT}^{1/3}}, \tag{5.8}$$

which is constructed to have the correct asymptotic properties at small and large R_{OT}.

Before discussing the physical interpretation of (5.8) and attempting to determine the scaling coefficient 'A' on physical grounds, it is very important to appreciate that many of the arguments presented above are not new. In fact, employing R_{OT} as a turbulence age proxy goes back several decades as reviewed by Smyth & Moum (2000), who also proposed that it be used to quantify $\Gamma_{\mathcal{M}}$. Numerical simulations and observations have also shown clear dependence of mixing on R_{OT} (Smyth *et al.* 2001; Mashayek *et al.* 2017*a*; Ijichi & Hibiya 2018; Ijichi *et al.* 2020; Smith 2020). Equation (5.7), in particular, has been previously proposed by many authors (Weinstock 1992; Schumann & Gerz 1995; Baumert & Peters 2000; Ijichi & Hibiya 2018).

Indeed, these previous authors typically invoked turbulence properties largely unaffected by 'weak' stratification. We go a step further by arguing that (5.7) is appropriate only late in the life cycle of a time-varying overturning mixing event, where the stratification is only required not to be so strong that it always remains sub-dominant in the flow evolution. Furthermore, L_T was argued by Ivey & Imberger (1991) as representative for the actual mixing length scales (once again without explicit consideration of time dependence), an idea which was verified later via numerical simulations (see Shih *et al.* (2005), for example). However, we are not aware of a previous proposal of (5.6) for 'young', yet vigorously overturning mixing events, and as we shall show, its combination with (5.7) leads to an accurate parametric modelling of $\Gamma_{\mathcal{M}}$ through (5.8) (or indeed alternative definitions like Γ_{χ}) in a way that would not be possible merely based on (5.7).

5.5. *Context in terms of previous studies*

It should also be noted that while we use the paradigm of a single (with three temporal phases) mixing life cycle for arguments herein, our arguments also hold in a system composed of a multitude of interacting, co-existing overturns, at different stages of their individual life cycle. In such systems, the young phase simply refers to the phase where large quasi-laminar and relatively recently formed overturns exist and provide APE to small scale overturns to grow, while the fossilization phase corresponds to when the APE source has been largely depleted and the actual still-occurring overturns are smaller than the notionally possible size scale as marked by L_O.

Mater *et al.* (2015) showed, using several oceanic datasets, that the evolution of R_{OT} in data does in fact resemble that based on the shear-driven KHI induced mixing flow evolution shown in figure 1. In a companion paper, Scotti (2015) argued that in shear-driven mixing $L_T \sim L_O$, whereas in convectively driven mixing (where the underlying source of TKE is APE), L_T can actually overestimate mixing. However, we would argue such a clear and binary distinction between the two types of mixing

is not entirely appropriate. A simple mixing event due to a shear instability can be dominated by a cascade of 'secondary' shear instabilities growing on shear instabilities, for example as in an energetic estuary as shown in Geyer *et al.* (2010), or it can be more appropriately characterized as being convectively driven once the primary billow has rolled up, as observed in the thermocline by Woods (1968). We argue that if a distinction is allowed in principle between the mixing length and the turbulent integral scale, as done here, at least some 'convective' mixing can fit within the same parameterization as 'shear-driven' overturnings; oceanic datasets appear to support this argument. Needless to say, our argument does not extend to truly convection-driven mixing such as that in deep convection zones, but rather to patches of mixing where it is reasonable to think of relatively isolated time-varying overturning mixing events where the stratification is sufficiently subcritical so as not to suppress such overturning.

Caldwell (1983) provided a nice description of the observed data on R_{OT} on physical grounds and summarized the opposing views that became a source of controversy: on one hand, Gibson (1987) argued that most observations were what he referred to as fossilized turbulence, while Gregg (1987) argued that they were of active turbulence. Caldwell proposed that the largest eddies (of scale L_O) feed on the original overturn (of scale L_T). He provided scaling arguments, observational support, and other references, to show that $L_O \sim L_T$ in this growth phase. This was also pointed out by Itsweire *et al.* (1993). This is consistent with the idealized picture presented in figure 1, as well as with simulations of other non-shear-instability-like turbulent overturns (e.g. see Fritts *et al.* 2003; Chalamalla & Sarkar 2015). Caldwell argued that while the fossilization phase is also captured in the data, most of the observed overturns are in a state of 'continuous production'.

This continuous production corresponds to the intermediate energetically mixing phase identified in § 3. As the mixing during this phase seems both to be most efficient and also to be associated with $L_O \sim L_T$, we refer to this phase as being the Goldilocks mixing phase, where the turbulent overturning is neither too 'hot' ($L_T \gg L_O$) nor too 'cold' ($L_T \ll L_O$) but 'just right' ($L_O \sim L_T$). As argued in Mashayek *et al.* (2017*a*), in this Goldilocks phase, the energy is most efficiently supplied to turbulence as the external driving (associated with the primary KHI billow roll up) is at a scale which essentially matches the upper bound of an assumed inertial subrange largely unaffected by the stratification for which the (dynamic) scalar mixing of buoyancy is subordinate to the turbulence cascade.

5.6. *Quasi-equilibrium and self-organized criticality*

It has been suggested (Turner 1973; Sherman, Imberger & Corcos 1978; Salehipour, Peltier & Caulfield 2018; Smyth, Nash & Moum 2019; Smyth 2020) that various kinds of shear-induced geophysically relevant turbulence are in a state of self-organized criticality (SOC) in which the system is marginally unstable and overturns spontaneously emerge when the flow properties intermittently drop below a stability criterion with respect to a critical Richardson number, i.e. in the sense discussed in § 1, the stratification intermittently becomes (just) subcritical, self-organizing and relaxing back towards criticality. Although Howland, Taylor & Caulfield (2018) demonstrated that KHI are exceptionally weak when the minimum Richardson number approaches the classical Miles–Howard (Howard 1961; Miles 1961) criterion of 1/4, calling into question the appropriateness of describing an energetically turbulent flow as being 'marginally unstable' with respect to that specific value, as proposed by Thorpe & Liu (2009), a body of evidence suggests that stratified turbulent flows can indeed adjust towards a state where characteristic values of $Ri \simeq 0.17$–0.25 (Zhou, Taylor & Caulfield 2017; Portwood, de Bruyn Kops & Caulfield 2019; van Reeuwijk, Holzner & Caulfield 2019). As originally

argued by Turner (1973) and Sherman *et al.* (1978), this evidence is at least suggestive that shear-driven turbulent flows adjust to a 'kind of equilibrium' with $Pr_T \sim O(1)$, and characteristic values of Ri close to the linear stability critical value, perhaps fortuitously.

We combine these ideas to propose an appropriate value for the scaling coefficient 'A' in (5.8). Keeping in mind that the $R_{OT} \ll 1$ and $R_{OT} \gg 1$ limits of (5.8) represent the young and decaying or fossilizing limiting phases of a turbulent mixing event, $R_{OT} \sim 1$ thus represents the intermediate Goldilocks mixing phase. Let us assume that such a mixing phase is in the 'kind of equilibrium' proposed by Turner, which in the modern language of physics is essentially equivalent to a state of SOC. However, it might be described, emergence of the balance R_{OT} on physical grounds is consistent with the empirical fact that the majority of observed overturns appear to be in this phase of efficient, and indeed 'optimal' mixing, as has been discussed previously (Dillon 1982; Ferron *et al.* 1998; Mater *et al.* 2015; Mashayek *et al.* 2017*a*).

Therefore, we assume that the flow has a characteristic critical Richardson number Ri_{cr} close to, but perhaps somewhat less than 1/4, i.e. that the stratification of the flow is just slightly subcritical. Although it is perhaps stating the obvious, it important to appreciate that such a Richardson number should be considered as being significantly less than one, and to be associated with a flow where it is not to be expected that the turbulence will be dominated by the stratification. For such a subcritically stratified flow at $R_{OT} \approx 1$, (5.8) reduces to

$$\Gamma_{gold} = \frac{A}{2}. \tag{5.9}$$

Combining (2.5) and (5.3*a*,*b*),

$$Pr_T = \frac{P}{[-B]} \frac{N^2}{S^2} = \frac{Ri}{Ri_f}, \tag{5.10}$$

defining the 'flux' Richardson number Ri_f (Turner 1973). Over sufficiently long averages and ignoring transport terms, we can assume Ri_f is equivalent to a particular definition of a mixing efficiency, since $P \simeq \mathcal{M} + \epsilon$ and $-B \simeq \mathcal{M}$. Therefore

$$Ri_f \simeq \frac{\Gamma_{\mathcal{M}}}{1 + \Gamma_{\mathcal{M}}}, \tag{5.11}$$

(see Ivey & Imberger 1991; Shih *et al.* 2005; Salehipour & Peltier 2015) Therefore,

$$\Gamma_{\mathcal{M}} = \frac{\dfrac{Ri}{Pr_T}}{1 - \dfrac{Ri}{Pr_T}}. \tag{5.12}$$

Since we are always assuming that the mixing of scalar is locked to the mixing of momentum so that $Pr_T \sim 1$, $Ri_f \sim Ri \lesssim 1/4$, and so $\Gamma_{\mathcal{M}} \lesssim 1/3$. As we shall see, the underlying assumption that the flow is 'subcritically stratified' but time evolving, leading to these various scalings, must always be remembered when comparing with previous studies, particularly those of Maffioli *et al.* (2019) and Garanaik & Venayagamoorthy (2019) based around the turbulent Froude number $Fr_T = \epsilon/[Nk]$.

If we further assume that the Goldilocks optimally efficient mixing occurs at some specific Ri_{cr} rather than over some range bounded above by Ri_{cr}, consistently with the flow

being in a self-organized critical equilibrium, we can then express the scaling parameter 'A' as

$$A = \frac{2\dfrac{Ri_{cr}}{Pr_t}}{1 - \dfrac{Ri_{cr}}{Pr_t}}. \tag{5.13}$$

Evidence from a range of flows suggests that vigorous turbulence in a subcritically stratified flow (i.e. in the specific sense that the flow has sufficiently small Ri that the turbulence, characterized by overturnings, can be sustained throughout the flow) has $Pr_T \simeq 1$ (Zhou *et al.* 2017; Portwood *et al.* 2019; van Reeuwijk *et al.* 2019). There is also evidence that appropriate values of Ri_{cr} range from 0.16 (as reported in Portwood *et al.* 2019) through 0.21 (Zhou *et al.* 2017; van Reeuwijk *et al.* 2019) to the canonical 'marginally stable' value of $1/4$ (Salehipour *et al.* 2018; Smyth *et al.* 2019; Smyth 2020). Indeed, due to a variety of reasons, in particular associated with the presence of ambient turbulence, the critical value of the Richardson number associated with linear instability may be expected to be less than this canonical value of $1/4$ (Smyth *et al.* 2001; Thorpe, Smyth & Li 2013; Kaminski & Smyth 2019).

Using these expressions $Pr_T \simeq 1$ and $1/6 \lesssim Ri_{cr} \lesssim 1/4$ suggests that $2/5 \lesssim A \lesssim 2/3$, i.e. $1/5 \lesssim \Gamma_{\mathcal{M}} \lesssim 1/3$, with the canonical lower value being associated with $Ri_{cr} = 1/6$ and $Pr_T = 1$, consistently with the direct calculations of Portwood *et al.* (2019) in a continuously forced flow, and also the (apparently) most efficient mixing in an evolving Kelvin–Helmholtz unstable shear flow (Mashayek *et al.* 2013). Alternatively $\Gamma_{\mathcal{M}} = 0.2$ can also be identified with $Ri_{cr} = 1/4$, $Pr_T = 5/4$, which is also largely consistent with available data, as one might argue that a continuously forced flow should correspond to the intermediate mixing phase of an evolving overturning event. As we will soon show, for the data considered in this work, 'A' lies within $2/5 \lesssim A \lesssim 2/3$ for individual and combined datasets. This implies that the above assumptions about Pr_T, Ri_{cr} are reasonable.

At the risk of belabouring the point, it is also important to appreciate two fundamental aspects of our argument concerning this optimal mixing phase with intermediate values of $R_{OT} \sim O(1)$ between the two asymptotic regimes of 'young' turbulence and the decaying 'fossilization' phase. First, although this regime is indeed intermediate, it still occurs when the flow has relatively small characteristic Richardson numbers significantly less than one, with crucially $Pr_T \simeq 1$, and so also $Ri_f \simeq Ri$. Therefore, the mixing of buoyancy is set by the mixing of momentum, and the density is mixed essentially as a passive scalar, with buoyancy playing an at most sub-dominant dynamical role. Specifically, and critically for our modelling, the properties of the turbulence itself, in particular characteristic length scales and time scales, are largely unaffected by stratification in this intermediate mixing phase within our framework. Second, the time dependence of the flow evolution is central to our argument, and this phase naturally follows the 'young' (and crucially vigorously overturning) phase of the flow evolution, and precedes the later decaying or fossilizing phase. Fundamentally, at no stage in the flow evolution do buoyancy effects dominate any aspect of the flow evolution.

To illustrate our proposed parameterization, figure 2 shows the parameterization in (5.8) plotted for $A = 2/3$ along with the asymptotic limits given by (5.6) and (5.7). Also shown is the naturally equivalent definition for an (instantaneous) mixing efficiency, i.e.

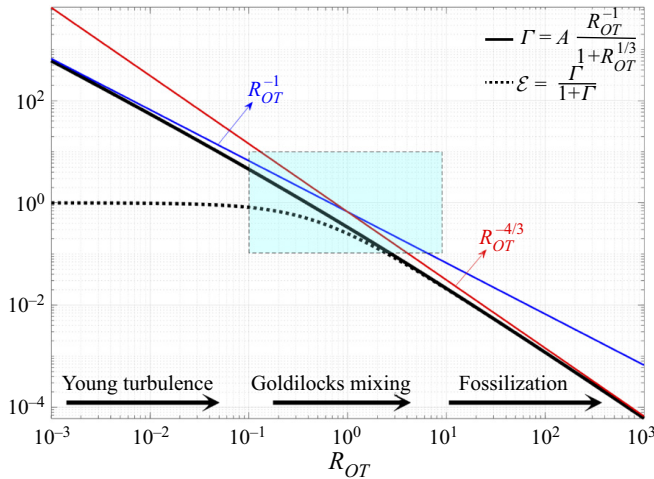$$\mathcal{E}_i \equiv \frac{\Gamma_{\mathcal{M}}}{1 + \Gamma_{\mathcal{M}}}. \tag{5.14}$$

Figure 2. Variation with R_{OT} of the parameterization (5.8) (with $A = 2/3$, implying $\Gamma_{\mathcal{M}} = 1/3$ and $\mathcal{E}_i = 1/4$ at $R_{OT} = 1$). Also shown are the two asymptotic scalings for growing turbulence (5.6) and for decaying turbulence (5.7) with thin grey lines, and the corresponding mixing efficiency, with a dashed red line. The light blue box in the middle of the plot represents the bounding box within which the ocean turbulence data considered in this work lie, as will be shown in figure 3. Conceptually, time for a particular shear-driven mixing event runs from left to right as R_{OT} increases.

This definition of mixing efficiency tends to one in the limit of a non-turbulent flow due to non-zero molecular diffusion (at $Pr = 1$) but the complete absence of turbulent dissipation, hence yielding a perfectly efficient (laminar) mixing, but importantly, no turbulent mixing. Yet again, it is important to remember that the flows of interest should never be thought of as strongly stratified, and so $R_{OT} \to 0$ because $\epsilon \to 0$ in the numerator of the Ozmidov length L_O as there is no turbulence at all, not because N is very large in the denominator, thus suppressing the turbulence due to buoyancy forces. In the limit of $R_{OT} \to \infty$, $\mathcal{E}_i$ tends to zero as would be expected in decaying turbulence, and $R_{OT} \to \infty$ since the Thorpe length L_T is decaying more rapidly than L_O. Finally, at the heart of the intermediate Goldilocks phase, $R_{OT} = 1$, $\mathcal{E}_i = A/(2 + A)$.

Figure 3(a–f) shows an excellent agreement between the parameterized form of $\Gamma_{param} = \Gamma_{\mathcal{M}}$ using (5.8) and measurements of R_{OT} and the observationally inferred $\Gamma_{data} = \Gamma_\chi = \chi/\epsilon$ constructed from measurements of χ and ϵ 'directly' from the six datasets introduced earlier. Each panel also includes the $\Gamma = A\,R_{OT}^{-1}$ and $\Gamma = A\,R_{OT}^{-4/3}$ limits. These two are expected to be relevant only in the limits of $R_{OT} \ll 1$ and $R_{OT} \gg 1$, as was shown in figure 2. However, the datasets in figure 3 lie near the Goldilocks limit $R_{OT} \sim 1$ (the blue box in the middle of figure 2). Thus, the two limiting cases are, perhaps unsurprisingly, inaccurate in the range relevant to the data whereas their joint scaling, as formulated by (5.8), much better represents the data. This is particularly clear when all the data are combined in (g) where a total of ∼50 000 turbulent patches provide sufficient statistics to provide a fair comparison between the data and the parameterization (note that at least a few tens of thousands of samples are needed to capture the distribution statistically significantly, as discussed in Cael & Mashayek (2020)).

As discussed earlier in connection with (5.13), for $Pr_T \simeq 1$ and $1/6 \lesssim Ri_{cr} \lesssim 1/4$, we obtain $2/5 \lesssim A \lesssim 2/3$. As shown in the captions of individual panels in figure 3, the range of values for 'A' obtained based on regression of individual datasets to the parameterization (5.8) falls within this range except for TIWE which is data-limited
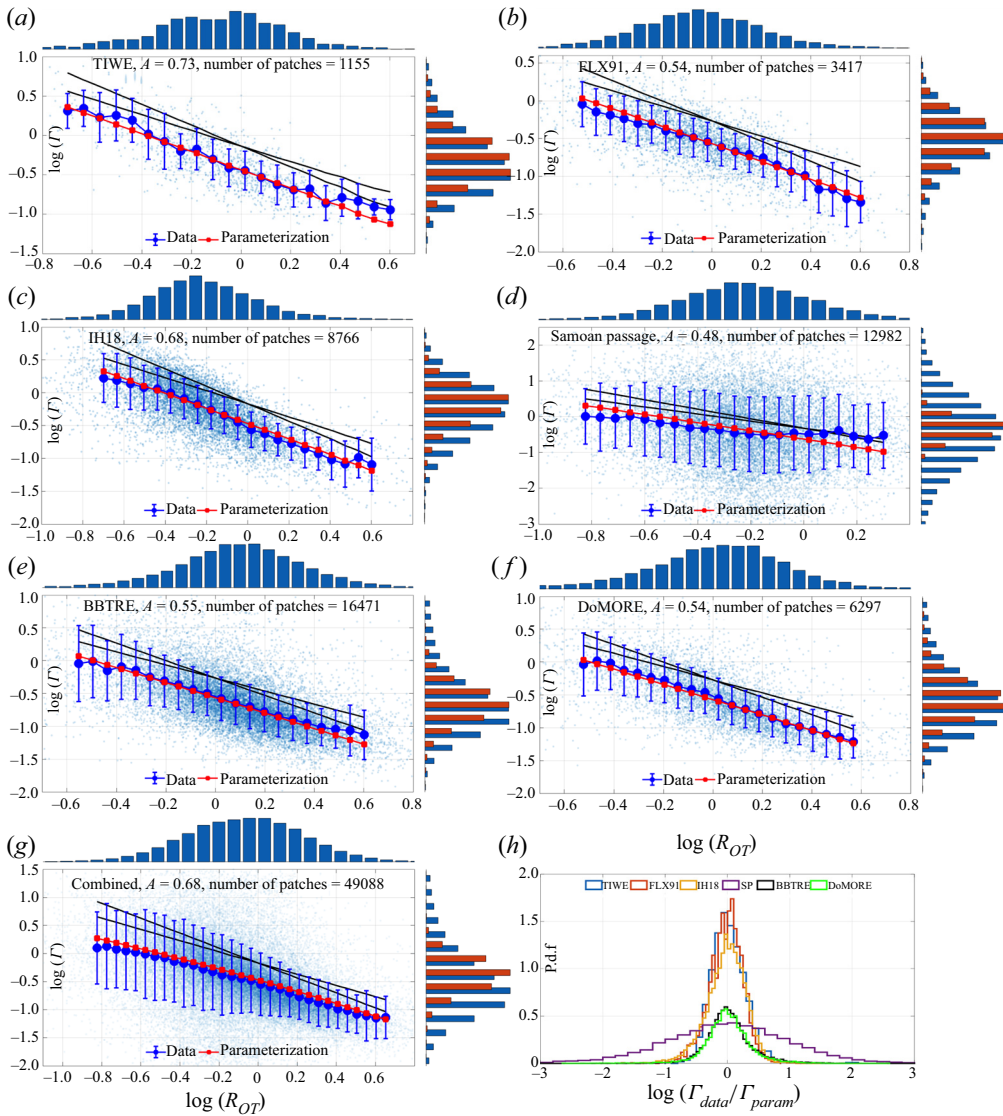
Figure 3. (*a–g*) Observationally inferred turbulent flux coefficient $\Gamma_{data} = \Gamma_\chi$ (constructed from measurements of χ and ϵ) as a function of R_{OT} for six oceanic datasets introduced in §4, as well as for the combined data. In each panel, blue scattered dots in the background represent actual patches, the solid blue line with filled circles and error bars represents the data binned (in $\log(\Gamma_\chi)$ space) and the solid red line with square symbols represents the parameterized $\Gamma_{param} = \Gamma_\mathcal{M}$ using (5.8). Error bars represent ± 1 standard deviation. The value of 'A' for each panel is inferred based on linear regression of the corresponding data to log (5.8). The right side inset compares probability density functions (p.d.f.s) of Γ_{data} vs Γ_{param} and the top inset shows the p.d.f. of data in $\log(R_{OT})$. The solid straight black lines represent $\Gamma = A\,R_{OT}^{-1}$ and $\Gamma = A\,R_{OT}^{-4/3}$. Panel (*h*) shows p.d.f.s of $\Gamma_{data}/\Gamma_{param}$ for the various datasets.

from a statistical standpoint but nevertheless gives an 'A' very close to the upper bound. The corresponding 'A' for the combined data is almost exactly $2/3$ which corresponds to $Pr_T \simeq 1$ and $Ri_{cr} \simeq 1/4$. Of course, this close agreement might well be somewhat fortuitous considering we only use six oceanic datasets. But given the wide range of

processes, geographical locations and oceanic depths spanned by these datasets, it is pleasing that their mixing can be captured so well using a parameterization entirely based on physical grounds: (5.8) depends on the Kolmogorov theory, the mixing length theory and the TKE budget, while '*A*' is determined based on a classical theoretical prediction for the critical Richardson number necessary for shear-induced instability. No empiricism, tuning, or filtering/modification of the data was required to obtain the agreement. Indeed, allowing some empiricism associated with numerical simulations such as those reported in Deusebio, Caulfield & Taylor (2015), Zhou *et al.* (2017), Portwood *et al.* (2019) and van Reeuwijk *et al.* (2019) so that the critical Richardson number $Ri_{cr} \sim 0.2$ would nevertheless only modify the result slightly.

Crucially, this agreement also suggests that ocean mixing likely occurs at scales sufficiently small that the local gradient Richardson number falls below a critical value so that mixing in such a state remains close to and perhaps slightly below marginally unstable, and fundamentally, scalings for the properties of the turbulence are never dominated by the ambient stratification. The turbulence itself may be generated at a scale close to the shear instability, through for example mixing in the thermocline through the breakdown of KHI. Alternatively, the initial generation may be at a much larger scale through (for example) tidal generation, which through linear and nonlinear processes, leads to generation of shear at sufficiently small scales at which local *Ri* can become sufficiently small (Nikurashin & Legg 2011). In other words, we hypothesize that the classic paradigm of a parallel shear flow, as idealized as it undoubtedly is, may plausibly lie at the heart of ocean interior mixing.

To compare the data and parameterization more quantitatively, figure 3(*h*) shows the histograms of the ratio of Γ from data and the parameterized Γ. The agreement is remarkable given the range of dominant driving mechanisms for the different datasets: shear instability triggered by tropical instability waves in the equatorial undercurrent (TIWE), near-inertial-shear (FLX91), internal tides (BBTRE) and hydraulically controlled flow over abyssal canyons and sills (DoMORE and Samoan Passage). For the latter two cases, the agreement is somewhat surprising considering that the measured turbulence is close to the seafloor and possibly partially convectively driven, although even such convectively breaking overturns are shear driven in the sense that the background shear is a main source of energy. Nevertheless, it is reasonable to expect an influence of another length scale, which is absent in the construction of (5.8) for near boundary turbulence. Perhaps this explains the larger spread in the data for the Samoan Passage than predicted by (5.8). Nevertheless, the agreement remains good at the peak of the p.d.f. of the data where $R_{OT} \sim 1$.

5.7. *Comparison to Fr_T-based parameterizations*

Here, we have been focussed on mixing events which can be related to shear-driven processes with subcritical stratification in the specific sense that vigorous overturning events can be observed. As cogently argued by Maffioli *et al.* (2019), building on the insights of Ivey & Imberger (1991), there are several reasons why it is appropriate to attempt to construct parameterizations in terms of the turbulent Froude number Fr_T defined (using our notation) as

$$Fr_T \equiv \frac{\epsilon}{Nk}. \tag{5.15}$$

They observed (from the results of body-forced numerical simulations) that $\Gamma \propto Fr_T^{-2}$ for large Fr_T, which is also consistent with gravity current observations as discussed in Wells, Cenedese & Caulfield (2010).

Of course, large Fr_T corresponds to 'weak' stratification, and so it is entirely expected that the unstratified scaling (5.1) applies. Therefore, remembering also the definition of L_O (2.1*a–c*), appropriately large Fr_T is proportional to

$$Fr_T \propto \left(\frac{L_O}{L_i}\right)^{2/3}. \tag{5.16}$$

Considering this weakly stratified class of flows, Maffioli *et al.* (2019) (effectively) assumed that $Pr_T \sim 1$, $\Gamma_\chi = \Gamma_\mathcal{M}$ and that the mixing length $L_m \sim L_i$ (using our notation), so that (2.1*a–c*) becomes

$$\Gamma_\mathcal{M} \propto \left(\frac{L_i}{L_O}\right)^{4/3} \propto Fr_T^{-2}. \tag{5.17}$$

Garanaik & Venayagamoorthy (2019) extended the discussion of Maffioli *et al.* (2019), constructing and interpreting various Froude number scalings in terms of time scales and length scales, motivated by constructing a parameterization that would be practically useful in terms of oceanic measurements. In the high Fr_T limit, they also argued (in our notation) that $L_i \sim L_T$, and so that $\Gamma_\mathcal{M} \propto Fr_T^{-2} \propto R_{OT}^{-4/3}$. This corresponds to the final decaying fossilization phase described above, though neither Maffioli *et al.* (2019) nor Garanaik & Venayagamoorthy (2019) considered time dependence of mixing events (by design). In contrast to our inherently time-dependent viewpoint, their argument was that $L_T < L_O$ was a generic characteristic of weakly stratified turbulent flows, whereas we postulate that this scaling only occurs late in the flow evolution of a shear-driven overturning mixing event, and in point of fact, our arguments only require the stratification to be subcritical (and thus allow shear-driven overturnings) with sufficiently small but not necessarily negligible characteristic Richardson number.

Furthermore, although not couched in terms of shear flows by Maffioli *et al.* (2019), the Fr_T^{-2} scaling can also be interpreted in terms of (sufficiently small) Richardson numbers, as the natural equivalent scaling is that

$$Fr_T \sim \frac{U_0}{L_i N} \propto Ri^{-1/2}, \tag{5.18}$$

and so

$$\Gamma_\mathcal{M} \propto Fr_T^{-2} \propto Ri. \tag{5.19}$$

This is of course entirely as expected for flows with $Pr_T \sim 1$, as then $\Gamma_\mathcal{M} \lesssim Ri_f \simeq Ri \lesssim 1/4$.

Interestingly, Garanaik & Venayagamoorthy (2019) also presented scaling arguments for two other (effectively assumed to be quasi-steady) classes of flows, neither of which should be construed as being equivalent to the (inherently time-dependent) phases of a subcritically stratified flow as considered here. Building on the observations of Maffioli *et al.* (2019), which in turn were motivated by experimental observations dating back (at least) to Kato & Phillips (1969), (see also Park, Whitehead & Gnanadeskian 1994; Oglethorpe, Caulfield & Woods 2013; Olsthoorn & Dalziel 2015) Garanaik & Venayagamoorthy (2019) presented a scaling argument for the observed constant flux coefficient $\Gamma_\chi \propto Fr_T^0$ in (at least apparently) strongly stratified flows as $Fr_T \to 0$. They argued that the strong stratification completely dominates the dynamics, so that both

the turbulent dissipation rate ϵ and the buoyancy flux B (and hence χ as the flow is quasi-steady) scale with w^2N, where w is the (perturbation) vertical velocity. Therefore Γ_χ is constant in very strong stratification. Clearly, this class of flows is not considered accessible by the (vertical) shear-driven mixing events with subcritical stratification allowing overturnings to reach significant amplitude on which we are focusing in this paper.

Furthermore, as argued in more detail in Caulfield (2021), care must be taken in inferring the existence of this class of strongly stratified and yet vigorously turbulent flows. The motivating experimental data were typically extracted from flows with emergent density staircases, where relatively well-mixed and deep 'layers' are separated by relatively thin 'interfaces' of enhanced density gradient. Indeed, as demonstrated by Portwood *et al.* (2016), flows with low Fr_T (when averaged over relatively large scales) typically exhibit significant spatio-temporal variability, with vigorous turbulence (characterized by large local values of ϵ) being restricted to patches of locally significantly weaker stratification. It is also necessary to be cautious when comparing scaling arguments from different studies, as there are often significant differences in the way in which superficially equivalent non-dimensional parameters are defined.

In particular, as discussed in detail in Brethouwer *et al.* (2007), very strongly stratified turbulent flows at sufficiently high Reynolds number can be profoundly anisotropic, with characteristic horizontal length scales significantly larger than characteristic vertical scales of turbulent patches, which in turn are larger than the Ozmidov scale (see also Caulfield (2020) for further discussion). Therefore, the particular length and velocity scales which are used to construct Froude numbers must be chosen with care and consistency before direct comparisons between different flows (and indeed scaling arguments) are made.

Connecting this (assumed) strongly stratified class of flows as $Fr_T \to 0$ and the weakly stratified class with $Fr_T \gg 1$, Garanaik & Venayagamoorthy (2019) argued for the existence of an intermediate 'moderately stratified' class of flows, where the properties are effectively hybrid between the two end members. They argued that the buoyancy variance destruction rate should scale as in the strongly stratified limit so that $\chi \propto w^2N$, while the turbulent kinetic energy (and its dissipation rate) are both (largely) unaffected by the stratification, so $k \sim w^2$, and so, in this class of flows, they argued that

$$\Gamma_\chi \propto \frac{kN}{\epsilon} = Fr_T^{-1}. \tag{5.20}$$

Furthermore, they argued that this moderately stratified class is also characterized by $Fr_T \sim O(1)$, which finally allowed them to argue that this intermediate class should have the length scaling $\Gamma_\chi \propto R_{OT}^{-1}$ as when $Fr_T \sim O(1)$ it is expected that $R_{OT} \sim O(1)$ as well.

The presented evidence of this intermediate class is perhaps not as convincing as for the two end members. Whatever the quality of the data fit, it is important to appreciate that the arguments leading to this scaling are fundamentally different from our arguments leading to the (superficially same) scaling (5.6) for the early-time 'young' phase of an inherently time-dependent and subcritically stratified shear-driven mixing event. Furthermore, our 'subcritically stratified' intermediate Goldilocks phase is also in marked conceptual contrast to the intermediate, moderately stratified class of flows parameterized by Garanaik & Venayagamoorthy (2019), in the sense that the physical processes and balances should be thought of as arising for fundamentally different reasons.

In their argument, the stratification (and hence the characteristic values of the buoyancy frequency) in their intermediate class has an order-one effect on the mixing. That can be seen perhaps most clearly in terms of the implications of their scaling arguments for Pr_T.

Since they argue that $\chi \propto w^2 N$,

$$\kappa_T \propto \frac{\chi}{N^2} \propto \frac{w^2}{N}. \tag{5.21}$$

On the other hand, since they argue that the TKE $k \propto w^2$ is largely unaffected by stratification, it is natural to suppose that the mixing length for momentum is given by (5.1), and so

$$\nu_T \propto \left(\frac{w^3}{\epsilon} \right) w \rightarrow Pr_T \propto \frac{Nw^2}{\epsilon} \propto Fr_T^{-1}, \tag{5.22}$$

using the scaling (5.21) for κ_T.

It could perhaps be argued that since in this regime $Fr_T \sim 1$, $Pr_T \sim 1$ still using this scaling. Nevertheless, a consistent interpretation of this moderately stratified class of flows leads to the prediction that the dynamics of the mixing of the buoyancy (as described by Pr_T) does depend on the stratification, with in particular Pr_T increasing as the stratification becomes stronger, and hence Fr_T decreases. Importantly such a variation in Fr_T can occur straightforwardly for different quasi-steady forced flows within their assumed class of flows.

This is a qualitatively different mixing behaviour from the mixing behaviour we argue occurs during the intermediate (in time) Goldilocks mixing phase. Our argument implicitly assumes that the stratification is sufficiently subcritical that $Pr_T \simeq 1$ throughout the mixing life cycle, and in particular during the intermediate Goldilocks phase. During this phase, it appears that the stratification is in some sense self-organized close to a critical value of the Richardson number, and so there is not expected to be any parameter dependence of Pr_T: the flow is after all neither too hot nor too cold, but just right. In summary, their argument assumes an intermediate quasi-steady class of flows with hybrid dynamical balance between buoyancy and turbulence, while our intermediate (in time) phase is still (and always) subcritically stratified. We argue that it is the middle phase of an inherently time-dependent flow evolution, that happens after a primary shear-driven overturning breaks down, but before it enters a final fossilization decay phase.

Indeed, caution must always be shown in drawing any connections between descriptions in terms of Fr_T and Ri, particularly as $Ri \simeq 1$ is actually really a quite strongly stratified flow in the context of shear-driven mixing events. Of course, there is no necessity for equality in the various scalings used in the construction of the parameterizations, and the presence of an order-one constant can always allow Ri to be sufficiently small to be subcritical in the sense used in this paper, i.e. for a shear-driven overturning to develop to sufficiently large amplitude to undergo vigorous turbulent breakdown, which we believe is actually the dominant process far from boundaries.

In summary, while (5.7) was proposed previously, to our knowledge (5.5), (5.6) and thereby (5.8) are new. Furthermore, the previously proposed scaling relation only strictly applies for $R_{OT} \gg 1$ which corresponds to a small fraction of the data. Through proposing a scaling for the $R_{OT} \ll 1$ limit and combining the two limits, we have managed to propose a comprehensive parameterization and also to determine the coefficient of proportionality on physical grounds. Note that earlier works used (5.7) with an adjustable coefficient that could be tuned to fit the data better. Figures 2 and 3 suggest that such a tuning is incorrect, and the data 'feel' both (5.6) and (5.7). In the next section, we will further discuss the significance of (5.6) for bulk estimates of mixing on regional and global scales.
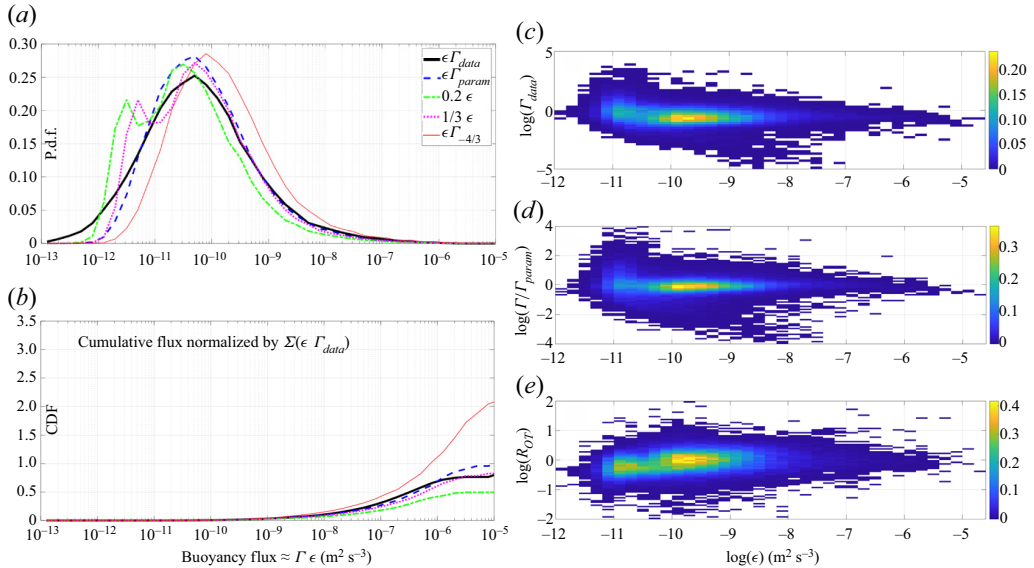
Figure 4. (*a*) Value of $\Gamma \times \epsilon$, considered a proxy for the buoyancy flux through the Osborn relation (2.6), calculated based on Γ_{data}, Γ_{param} using our full parameterization (5.8) and $\Gamma_{-4/3}$ using only the asymptotic scaling (5.7), as well as based on use of constant values $\Gamma = 0.2$ and $\Gamma = 1/3$. The last choice is based on (5.13) if we consider $Ri_{cr} = 1/4$. (*b*) Cumulative flux for the cases of panel (*a*), all normalized by the total $\Gamma_{data}\epsilon_{data}$ both obtained directly from data. (*c–e*) Bivariate histogram plots for the data over $\Gamma - \epsilon$, $\Gamma_{data}/\Gamma_{param} - \epsilon$ and $R_{OT} - \epsilon$ parameter spaces. All plots are made for the data combined for all six experiments in table 1.

6. Three key points

6.1. *The devil is in the tails*

An important question, one mostly left out of the literature discussing properties of various definitions of the turbulent flux coefficient Γ, is the extent to which the nuances of variations in Γ with other parameters matter in practice. There is no simple answer to this question, as it depends on the application for which the rate of mixing is being measured. Furthermore, a key missing piece is how one would connect the fluid-dynamical understanding of mixing of turbulent patches (as studied herein and in articles cited throughout), to an appropriate time- and space-averaged rate of mixing. The latter is out of the scope of our paper (but some discussion is provided in Cael & Mashayek 2020). Since our primary motivation behind this work is to relate efficiency of mixing to transformation of water masses in the ocean interior, which depends on the vertical divergence of the buoyancy flux (see Mashayek *et al.* 2015; Ferrari *et al.* 2016), we attempt to answer the question by simply quantifying the extent to which various measures of Γ will affect the total buoyancy flux obtained by summing the flux due to all the overturns in the datasets that we consider here.

Taking $\Gamma\epsilon$ (for various definitions of Γ) as a proxy for the buoyancy flux κN^2 through the Osborn balance (2.6), in figure 4(*a*) we plot the flux based on $\Gamma_{data} = \Gamma_\chi$ from the data, parameterizations and as constants. While we of course appreciate the many questionable assumptions underlying the approximation of actual flux with $\Gamma\epsilon$, as discussed in Mashayek & Peltier (2013) and Mashayek *et al.* (2013) for example, they do not affect our arguments here in any significant way. As expected based on figure 3, flux parameterized using (5.8) agrees well with the observed flux, while flux parameterized

	$\Sigma(\epsilon\Gamma_{param})/\Sigma(\epsilon\Gamma_{data})$	$\Sigma(0.2\epsilon)/\Sigma(\epsilon\Gamma_{data})$	$\Sigma(\frac{1}{3}\epsilon)/\Sigma(\epsilon\Gamma_{data})$	$\Sigma(\epsilon\Gamma_{-4/3})/\Sigma(\epsilon\Gamma_{data})$
TIWE	0.96	0.53	0.89	2.01
FLX91	0.97	0.72	1.16	2.02
IH18	1.03	0.41	0.67	2.62
Samoan Passage	2.91	1.21	1.99	6.58
BBTRE	0.96	0.52	0.86	2.96
DoMORE	0.91	0.24	0.39	2.27
Combined	1.43	0.63	1.05	3.2

Table 1. Comparison of the inaccuracies between the total buoyancy fluxes using the Osborn approximation $\Gamma\epsilon$ estimated based on constant values and various other parameterizations of Γ.

using (5.7) is biased high. Most importantly, if we consider the total flux due to all overturns, given the several orders of magnitude that ϵ spans, the tail of the histograms makes a significant contribution and so the extent to which Γ is accurately estimated for the tails is key. As shown in panel (*b*), it is the higher ϵ tail that leads to an overestimate when using parameterization (5.7). It is also notable that, while $\Gamma = 0.2$ underestimates the total flux, $\Gamma = 1/3$ actually leads to a much better agreement. Since $\Gamma = 1/3$ is linked (at least within our modelling approach) to assuming that the critical Richardson number $Ri_{cr} \simeq 1/4$, as was discussed above, it seems a more appropriate choice in the simplest approach of using a single constant value of Γ.

Figure 4 plots all six datasets combined. The disagreement between various estimates, however, can be much larger when looked at case by case. To do so, in table 1 we compare the total flux based on various estimates (i.e. the high end of the curves in figure 4*b*) for individual datasets. It is clear that parameterization (5.8) is quite accurate in capturing the total flux for five of the six cases (within 10 %). Even for the Samoan Passage case, which is somewhat of an outlier, it is still within a factor of 3. On the other hand, parameterization (5.7) is less accurate. $\Gamma = 0.2$ underestimates the total flux by as much as a factor of four, and fixing $\Gamma = 1/3$ does a much better job.

It is useful to interpret figure 4(*a*,*b*) and the agreement of data with parameterization (5.8) by looking at the distribution of data over various parameter spaces, as shown in (*c*–*e*) of the figure. While the data are spread around $R_{OT} \sim 1$, the spread is much larger at low ϵ, representing both the early stages of turbulence growth and the later stages of the turbulence after significant decay, as one would expect from consideration of figure 1. For energetic patches that correspond to the optimal Goldilocks mixing phase, R_{OT} nicely collapses around the value $R_{OT} \simeq 1$ as shown in (*e*). For this Goldilocks phase, $\Gamma \sim 1/3$ as shown in (*c*) and in agreement with the assumption of $Ri_{cr} \sim 1/4$, while at early and later stages of turbulence life cycles, Γ can be very large or very small, respectively. Panel (*d*) shows that the parameterization (5.8) manages to collapse this spread on R_{OT}, and that importantly it works well for the tail of the histograms which make the largest contribution to the total flux. In summary, given the accuracy, simplicity, and physically justified nature of the parameterization (5.8), it seems like a natural way to model the turbulent flux coefficient Γ associated with overturns in shear-induced turbulence.

6.2. *Bulk Γ: the importance of young turbulence events*

In ocean and climate general circulation models, small scale wave breaking is not resolved but the combined power available for mixing from tides, winds and other sources is
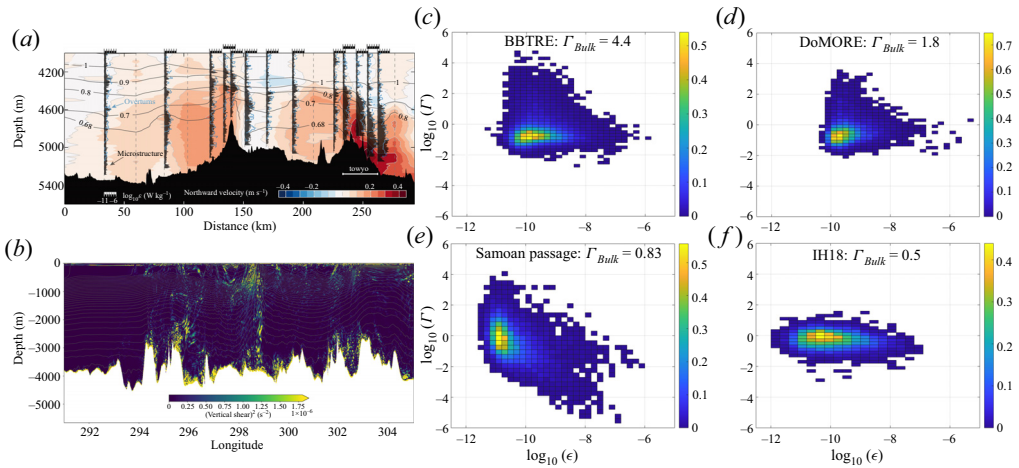
Figure 5. (*a*) An example of an energetic oceanic turbulence zone, in the Samoan Passage, to demonstrate the intermittency of turbulence; reproduced from Alford *et al.* (2013): northward velocity (colours), potential temperature (black contours) and dissipation rate measured by the velocity microstructure profiles (shaded black profiles) and from Thorpe scales (blue profiles). The scale for ϵ is given at the lower left and redrawn above each profile. (*b*) A snapshot of vertical shear from an observationally forced and tuned wave-resolving simulation of the Drake Passage, borrowed from the work of Mashayek *et al.* (2017*b*). (*c*–*f*) Bivariate histograms showing the distribution of patches over the $\epsilon - \Gamma$ parameter space for four of the datasets considered herein. The values of bulk Γ in each panel's title is obtained from (6.1). Note that the data in (*e*) are associated with (*a*).

available as a bulk value per grid cell. For example, an energetic turbulent zone in the ocean, such as that shown in figure 5(*a*) for the Samoan Passage, is roughly represented by 1–2 grid cells (in each direction) in a typical climate model. A bulk Γ is thus required to divide the power available (from tides, winds, currents, etc.) into bulk measures of mixing and dissipation for a given grid cell (see Cimoli *et al.* (2019), for a thorough discussion). In practice, this Γ_{Bulk} is often taken to be 0.2. Even without all the variations and caveats associated with Γ_i for individual turbulent patches, Γ_{Bulk} inevitably depends on the statistics of turbulent patches within each grid cell. Several studies have applied parameterizations for Γ, derived based on patch data, to coarse grid-based calculations to establish the leading-order importance of variations in Γ for the larger scale ocean circulation (De Lavergne *et al.* 2016; Mashayek *et al.* 2017*c*; Cimoli *et al.* 2019). That approach, while illuminating, is also not correct since it applies a patch-based quantity to grid cells of $O(100\ \text{km}) \times O(100\ \text{km}) \times O(100\ \text{m})$ in size. In this subsection we show how Γ_{Bulk} depends on statistics of turbulent patches and that not only does it depend on Goldilocks mixing of energetic patches, but equally on the young patches which possess large Γ and decaying patches with vanishing Γ.

Imagining a typical energetic zone, such as the example in figure 5(*a*), we assume the domain to contain *n* number of turbulent patches (within a grid cell of a coarse resolution model which would represent the region) and over a period associated with a coarse resolution model time stepping. Panel (*b*) shows a snapshot from a wave-resolving high resolution realistic simulation to demonstrate the intermittency of vertical shear generated by surface winds and flow–topography interaction in the very energetic Drake Passage in the Southern Ocean where high flow speeds occur due to the constriction of the Antarctic Circumpolar Currents and the overlying strong westerly winds known as the roaring forties. Even in such an energetic region, the flow is mostly

non-turbulent (from a small scale stratified turbulence perspective – a rich mesoscale and submesoscale turbulence field exists throughout the domain). However, in the 'quiet' regions, background weak mixing almost always exists: 97 % of the data acquired from ∼750 full-depth microstructure profiles from 14 field experiments considered in Cael & Mashayek (2020) possessed a diffusivity larger than 10^{-6} m^2 s^{-1} (also see Waterhouse *et al.* 2014). In the limit of completely laminar flow, $\kappa \to O(10^{-7})$ m^2 s^{-1} and so the flux $\mathcal{M} \approx \kappa N^2$ stays finite while ϵ tends to zero. Thus, while for energetic turbulent patches $\Gamma \approx \Gamma_{gold}$, for quieter regions it tends to $O(10)$ and larger values (see histograms in figure 3).

To show how the energetic, young and decaying turbulent patches collectively set the bulk Γ, we define

$$\Gamma_{Bulk} = \frac{\mathcal{M}_{tot}}{\epsilon_{tot}} \approx \frac{\sum\limits_{i=1}^{n} \Gamma_i \times \epsilon_i}{\sum\limits_{i=1}^{n} \epsilon_i}, \qquad (6.1)$$

where n represents the number of patches in a region of interest. Figure 5(*c–f*) shows the estimated Γ_{Bulk} for four deep energetic turbulent datasets considered herein. As discussed in figure 4, for energetic patches with high ϵ, $\Gamma_i \to \Gamma_{gold}$ while small ϵ corresponds to young and decaying patches which possess large and small Γ_i, respectively. For BBTRE and DoMORE, abundance of young patches result in a large Γ_{Bulk} even though most of the data correspond to Goldilocks mixing. For IH18, a larger fraction of the data correspond to Goldilocks mixing and so does Γ_{Bulk}. A simple thought experiment can help illustrate the impact of young turbulent patches. Imagine a box with three turbulent patches: a young patch ($\epsilon = 0.001$, $\Gamma = 100$), a moderately turbulent patch ($\epsilon = 0.1$, $\Gamma = 10$) and an energetic patch ($\epsilon = 1$, $\Gamma = 1/3$). This results in $\Gamma_{Bulk} = 0.44$.

Our key message is that while most of the collective community efforts have focused on finding the right parameterization for energetic turbulent patches, the net mixing also critically depends on young patches and the large fraction of the water column where only weak background turbulence occurs. This highlights the importance of the new scalings presented in this work in (5.5), (5.6). It also highlights the need for investing efforts in understanding better the statistics of turbulence to quantify its intermittency. Such statistics are key to connecting our understanding of physics of small scale stratified turbulence, as discussed in this work and the ones cited herein, to the larger scale applications. This statistical bridge is arguably highly underdeveloped – see Cael & Mashayek (2020) for a discussion. Finally, while we used the patch data for four experiments in figure 5, it was merely to illustrate the simple point that Γ_{Bulk} can be large even though most patches have $\Gamma \sim \Gamma_{gold}$. We have no reason to believe that the number of patches within each experiment was statistically sufficient to quantify mixing accurately in the geographical region of each experiment. Also, the data we used are only for the already identified patches. The actual microstructure profiles also include regions that are not identified as patches, but should contribute to Γ_{Bulk}.

6.3. *Statistics of L_O/L_T: active vs fossil turbulence*

Thus far we have repeated on several occasions and on empirical grounds (based on the data used in this study and other works cited herein) that 'most of the data are centred around $R_{OT} \sim 1$'. This, however, lies at the heart of a historically significant debate. As discussed in the introduction, Gibson (1987) argued that most observations were of fossilized turbulence, while Gregg (1987) argued that they were of active turbulence.
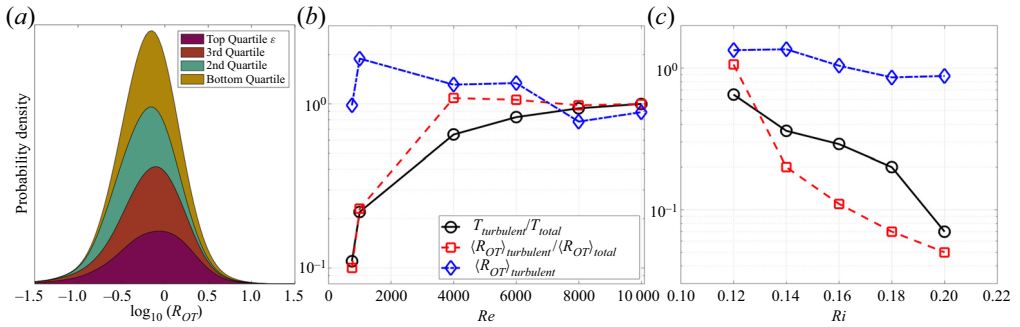
Figure 6. (*a*) Probability density function of R_{OT}, grouped into quartiles of ϵ, from the combined six oceanic datasets considered in this work. (*b,c*) Temporal fraction of turbulence life cycle as well as the ratio of R_{OT} during turbulent phase of the flow to its mean value over the whole life cycle. Each symbol represents a life-cycle-averaged quantity from a direct numerical simulation for a given pair of Reynolds and Richardson numbers. All cases in (*b*) are for $Ri = 0.12$ while all cases in (*c*) are for $Re = 6000$. Turbulent phase of the life cycles are defined as the times when $Re_b > 20$. Panels (*b,c*) are produced from analysis of simulations originally discussed in Mashayek & Peltier (2011*a,b*) and Mashayek & Peltier (2013); Mashayek *et al.* (2013).

The latter, referred to as 'continuous production' by Caldwell (1983) and as Goldilocks mixing herein, is supported by a seemingly universal concentration of data around $R_{OT} \sim 1$.

Figure 6(*a*) shows the histogram of the combined data considered in this work to highlight the clustering around $R_{OT} \sim 1$. Importantly, the clustering improves for more energetic turbulent patches with larger ϵ. Panels (*b,c*) of the figure demonstrate, based on data from direct numerical simulations of shear instabilities, that once the parameter ranges approach those of oceanic turbulent patches (i.e. large Re and subcritical Ri), a larger fraction of the total overturning life cycle corresponds to sustained energetic turbulence (for which $R_{OT} \sim 1$). This explains the statistical distribution of R_{OT} in panel (*a*). In contrast, less energetic turbulence events, such as those created in laboratory or numerical simulations, or those observed in lakes, are expected to have a more broadly distributed (over R_{OT}) distributions. Once again, we argue that $R_{OT} \sim 1$ is associated with an intermediate phase of an energetic time-dependent subcritically stratified mixing event, not an intermediate, hybrid class of moderately stratified quasi-steady mixing events as argued by Garanaik & Venayagamoorthy (2019).

7. A note on Γ as a function of Re_b

Heretofore, we have presented arguments for parameterizing the turbulent flux coefficient in terms of a ratio of length scales. On the other hand, the literature proposing the use of appropriate definitions of a buoyancy Reynolds number Re_b and/or a Richardson number Ri to quantify mixing efficiency is relatively well established (Peltier & Caulfield 2003; Ivey *et al.* 2008; Gregg *et al.* 2018; Caulfield 2021). However, efforts that compared various datasets found them not to overlap when mapped onto these parameters (Bouffard & Boegman 2013; Mashayek *et al.* 2017*c*; Monismith *et al.* 2018). To highlight this for the datasets employed in this article, in figure 7 we plot the totality of the six datasets considered herein in the $\Gamma - Re_b$ space and also include bin-averaged means for each dataset. Put simply, the data are all over the place.

Substantial empirical evidence suggests that in the energetically mixing regime, $\Gamma \propto Re_b^{-1/2}$ (Ivey *et al.* 2008; Bouffard & Boegman 2013; Mashayek *et al.* 2017*c*; Monismith
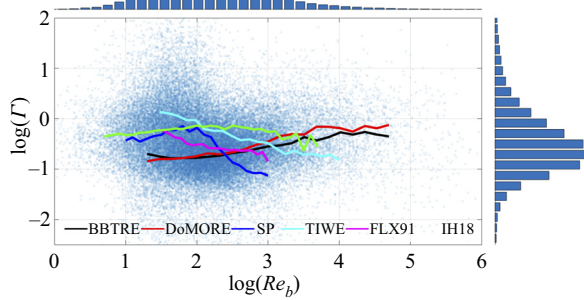
Figure 7. Scatter plot of data for the six oceanic datasets in $Re_b - \Gamma$ space, overlaid by bin-averaged curves for each experiment.

et al. 2018). There is also some evidence from experimental and numerical data (see review in Bouffard & Boegman (2013); also see Mashayek *et al.* 2017c) that for smaller Re_b, $\Gamma \propto Re_b^{1/2}$. Finally, substantial evidence also suggests $\Gamma \propto Ri \sim Fr^{-2}$ in sufficiently weakly stratified flows when $Pr_T \sim O(1)$ (Shih *et al.* 2005; Wells *et al.* 2010; Lozovatsky & Fernando 2013; Deusebio *et al.* 2015; Salehipour & Peltier 2015; Zhou *et al.* 2017; Maffioli *et al.* 2019; Garanaik & Venayagamoorthy 2019). One could argue that each line in figure 7 consists of left and right flanks that represent these limits.

We have tried, unsuccessfully, to interpret these limits as young and decaying turbulence phases and parameterize Γ in a fashion similar to Bouffard & Boegman (2013) and Mashayek *et al.* (2017c) with the addition of Ri to the parameterization. For that to work, most of the data would be expected to be at the transition between the left and right scaling (i.e. the peak of the lines in figure 7) but that did not hold for any of the various data we considered here. From a physical perspective, the most fundamental message of this paper is the need to account for both energy injecting and dissipation scales to quantify mixing; Re_b only conveys information about the latter, and thus is insufficient by design. Our efforts to resort to Ri as a means for including a scale for background shear and reconcile the various data were not fruitful even for the two datasets that possess Ri measurements. Considering Ri is often not measured, and the success of (5.8), we did not pursue $\Gamma = f(Re_b, Ri)$ further here. We also like to point out that while Re_b spans five orders of magnitude in figure 7, the corresponding R_{OT} captures most of the data within two orders of magnitude. So even if a measurable energy injection scale could somehow be combined with Re_b, (5.8) would still be much preferable, as attested to by the nice collapse of data in figure 3(g).

8. Conclusions

In this paper, we have shown that a parameterization of an appropriate definition of the turbulent flux coefficient Γ based on the ratio R_{OT} of Ozmidov and Thorpe scales may be derived using classical ideas of energy transfer within the inertial range, mixing length theory, and criticality conditions for parallel free shear layers. The novelty of this parameterization is that it spans the entire turbulent life cycle of a mixing event, and yet does not involve any empirical tuning coefficients. We have shown that the parameterization agrees well with a host of oceanic observations of turbulent overturns in different geographical, depth, and turbulence regimes. Most energetic turbulent patches appear to correspond to an intermediate phase of turbulence where $R_{OT} \sim 1$, implying an

efficient transfer of energy between the background flow and the hierarchy of eddies that exist in the inertial range. We refer to this efficient and in some sense optimal phase as the Goldilocks mixing phase, and have shown that $\Gamma \simeq 1/3$ in this phase. This value agrees not only with observations but also directly follows from assuming a critical Richardson number $Ri_{cr} \sim 1/4$, suggesting a connection with such flows being in a critical 'kind of equilibrium' state, or close to 'marginal' stability in some sense.

Our work further emphasizes the essential significance of identification and quantification of both the energy injecting and dissipation scales for accurate parameterization of mixing. Of course, this is well known both from observations and simulations. However, it is common to infer turbulent diffusivity from microstructure measurements alone or from Thorpe scale estimates with the assumption of a fixed R_{OT}. We simply further emphasize that both such approaches are fundamentally limited. When ϵ, L_T, and χ are all available, (4.1), (5.8), may be used in conjunction to constrain Γ quite tightly. In the absence of χ, (5.8) can still be used successfully to estimate Γ.

We have also quantified the extent to which variations in Γ matter for the total buoyancy flux based on the sum of contributions of individual patches for the various observational datasets, and have shown that these variations can lead to inaccuracies as high as a factor of 7. The key point appears to be that the statistics of overturn characteristics have a surprisingly long and fat tail in the high ϵ limit. Since ϵ is observed to vary over orders of magnitude, estimating Γ correctly for the tail is crucially important, and a constant value based on the mean of many overturns will inevitably be inaccurate. Our parameterization describes Γ relatively well for such tails and also demonstrates that $1/3$ may be more appropriate than 0.2 as a rule of thumb for a constant Γ.

Finally, we wish to emphasize that everything that we have discussed in this paper is in relation to spatially and temporally isolated turbulent patches. Such patches are identified from profiler data that include quiescent regions. We showed that young turbulent patches with high Γ can bias to high values the bulk value of Γ associated with a coarse resolution climate model grid cell. Thus, not only the high tail of ϵ matters, but so does the portion of the low tail that corresponds to young patches (with the rest of the low tail corresponding to decaying turbulence). Outstanding open questions, pertinent to the quantification of the role of mixing on regional and global scale in the world's oceans, are (i) what is the appropriate underlying distribution of patches in time and space, and (ii) what is the relative importance of the driving mechanisms of ocean turbulence. While the properties of the distribution or 'census' of the turbulent patch population and generating mechanisms are key to any attempt to connect the physics of density-stratified turbulence to most oceanographic applications, both issues remain extremely poorly understood (MacKinnon *et al.* 2017; Cael & Mashayek 2020).

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.

A. Mashayek https://orcid.org/0000-0002-8202-3294;

C.P. Caulfield https://orcid.org/0000-0002-3170-9480;

M.H. Alford https://orcid.org/0000-0002-6318-0737.

REFERENCES

ALFORD, M. & PINKEL, R. 2000 Patterns of turbulent and double-diffusive phenomena: observations from a rapid-profiling microconductivity probe. *J. Phys. Oceanogr.* **30** (5), 833–854.

ALFORD, M.H., GIRTON, J.B., VOET, G., CARTER, G.S., MICKETT, J.B. & KLYMAK, J.M. 2013 Turbulent mixing and hydraulic control of abyssal water in the Samoan Passage. *Geophys. Res. Lett.* **40** (17), 4668–4674.

BAUMERT, H. & PETERS, H. 2000 Second-moment closures and length scales for weakly stratified turbulent shear flows. *J. Geophys. Res.: Oceans* **105**, 6453–6468.

BOUFFARD, D. & BOEGMAN, L. 2013 A diapycnal diffusivity model for stratified environmental flows. *Dyn. Atmos. Oceans*. **61–62**, 14–34.

BRETHOUWER, G., BILLANT, P., LINDBORG, E. & CHOMAZ, J.-M. 2007 Scaling analysis and simulation of strongly stratified turbulent flows. *J. Fluid Mech.* **585**, 343–368.

CAEL, B.B. & MASHAYEK, A. 2020 Log-skew-normality of ocean turbulence. *Phys. Rev. Lett.* **126**, 224502.

CALDWELL, D.R. 1983 Oceanic turbulence: big bangs or continuous creation? *J. Geophys. Res.* **88** (C12), 7543–7550.

CARTER, G.S., VOET, G., ALFORD, M.H., GIRTON, J.B., MICKETT, J.B., KLYMAK, J.M., PRATT, L.J., PEARSON-POTTS, K.A., CUSACK, J.M. & TAN, S. 2019 A spatial geography of abyssal turbulent mixing in the samoan passage. *Oceanography* **32** (4), 194–203.

CAULFIELD, C.P. 2021 Layering, instabilities, and mixing in turbulent stratified flows. *Annu. Rev. Fluid Mech.* **53** (1), 113–145.

CAULFIELD, C.P. & PELTIER, W.R. 2000 Anatomy of the mixing transition in homogeneous and stratified free shear layers. *J. Fluid Mech.* **413**, 1–47.

CAULFIELD, C.P. 2020 Open questions in turbulent stratified mixing: do we even know what we do not know? *Phys. Rev. Fluids* **5**, 110518.

CHALAMALLA, V.K. & SARKAR, S. 2015 Mixing, dissipation rate, and their overturn-based estimates in a near-bottom turbulent flow driven by internal tides. *J. Phys. Oceanogr.* **45** (8), 1969–1987.

CIMOLI, L., CAULFIELD, C.P., JOHNSON, H.L.., MARSHALL, D.P., MASHAYEK, A., NAVEIRA GARABATO, A.C. & VIC, C. 2019 Sensitivity of deep ocean mixing to local internal tide breaking and mixing efficiency. *Geophys. Res. Lett.* **46** (24), 14622–14633.

CLÉMENT, L., THURNHERR, A.M. & ST. LAURENT, L.C. 2017 Turbulent mixing in a deep fracture zone on the Mid-Atlantic Ridge. *J. Phys. Oceanogr.* **47** (8), 1873–1896.

DE LAVERGNE, C., MADEC, G., LE SOMMER, J., NURSER, A.J.G. & GARABATO, A.C.N. 2016 The impact of a variable mixing efficiency on the abyssal overturning. *J. Phys. Oceanogr.* **46** (2), 663–681.

DEUSEBIO, E., CAULFIELD, C.P. & TAYLOR, J.R. 2015 The intermittency boundary in stratified plane Couette flow. *J. Fluid Mech.* **781**, 298–329.

DILLON, T.M. 1982 Vertical overturns: a comparison of Thorpe and Ozmidov length scales. *J. Geophys. Res.* **87** (C12), 9601–9613.

FERRARI, R., MASHAYEK, A., MCDOUGALL, T.J., NIKURASHIN, M. & CAMPIN, J.-M. 2016 Turning ocean mixing upside down. *J. Phys. Oceanogr.* **46** (7), 2239–2261.

FERRON, B., MERCIER, H., SPEER, K., GARGETT, A. & POLZIN, K. 1998 Mixing in the Romanche fracture zone. *J. Phys. Oceanogr.* **28** (10), 1929–1945.

FRITTS, D.C., BIZON, C., WERNE, J.A. & MEYER, C.K. 2003 Layering accompanying turbulence generation due to shear instability and gravity-wave breaking. *J. Geophys. Res.: Atmos. (1984–2012)* **108** (D8), 8452.

GARANAIK, A. & VENAYAGAMOORTHY, S.K. 2019 On the inference of the state of turbulence and mixing efficiency in stably stratified flows. *J. Fluid Mech.* **867**, 323–333.

GEYER, W.R., LAVERY, A.C., SCULLY, M.E. & TROWBRIDGE, J.H. 2010 Mixing by shear instability at high Reynolds number. *Geophys. Res. Lett.* **37**, L22607.

GIBSON, C.H. 1987 Fossil turbulence and intermittency in sampling oceanic mixing processes. *J. Geophys. Res.: Oceans (1978–2012)* **92** (C5), 5383–5404.

GREGG, M.C. 1987 Diapycnal mixing in the thermocline: a review. *J. Geophys. Res.: Oceans* **92** (C5), 5249–5286.

GREGG, M.C., D'ASARO, E.A., RILEY, J.J. & KUNZE, E. 2018 Mixing efficiency in the ocean. *Annu. Rev. Mar. Sci.* **10** (1), 443–473.

HOWARD, L.N. 1961 Note on a paper of John W. Miles. *J. Fluid Mech.* **10**, 509–512.

HOWLAND, C.J., TAYLOR, J.R. & CAULFIELD, C.P. 2018 Testing linear marginal stability in stratified shear layers. *J. Fluid Mech.* **839**, R4.

IJICHI, T. & HIBIYA, T. 2018 Observed variations in turbulent mixing efficiency in the deep ocean. *J. Phys. Oceanogr.* **48** (8), 1815–1830.

IJICHI, T., ST. LAURENT, L., POLZIN, K.L. & TOOLE, J.M. 2020 How variable is mixing efficiency in the Abyss? *Geophys. Res. Lett.* **47** (7), 1–9.

ITSWEIRE, E.C., KOSEFF, J.R., BRIGGS, D.A. & FERZIGER, J.H. 1993 Turbulence in stratified shear flows: implications for interpreting shear-induced mixing in the ocean. *J. Phys. Oceanogr.* **23** (7), 1508–1522.

IVEY, G.N. & IMBERGER, J. 1991 On the nature of turbulence in a stratified fluid. Part I: the energetics of mixing. *J. Phys. Oceanogr.* **21**, 650–658.

IVEY, G.N., WINTERS, K.B. & KOSEFF, J.R. 2008 Density stratification, turbulence, but how much mixing? *Annu. Rev. Fluid Mech.* **40**, 169–184.

KAMINSKI, A.K. & SMYTH, W.D. 2019 Stratified shear instability in a field of pre-existing turbulence. *J. Fluid Mech.* **862**, 639–658.

KATO, H. & PHILLIPS, O.M. 1969 On the penetration of a turbulent layer into stratified fluid. *J. Fluid Mech.* **37** (4), 643–655.

KOLMOGOROV, A.N. 1941 The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *C. R. Acad. Sci. USSR* **30**, 301–305.

LIEN, R.-C., CALDWELL, D.R., GREGG, M.C. & MOUM, J.N. 1995 Turbulence variability at the equator in the central Pacific at the beginning of the 1991–1993 El Nino. *J. Geophys. Res.: Oceans* **100** (C4), 6881–6898.

LOZOVATSKY, I.D. & FERNANDO, H.J.S. 2013 Mixing efficiency in natural flows. *Phil. Trans. R. Soc. Lond.* A **371** (1982), 20120123.

MACKINNON, J.A., *et al.* 2017 Climate process team on internal wave–driven ocean mixing. *Bull. Am. Meteorol. Soc.* **98** (11), 2429–2454.

MAFFIOLI, A., BRETHOUWER, G. & LINDBORG, E. 2019 Mixing efficiency in stratified turbulence. *J. Fluid Mech.* **794**, 3.

MASHAYEK, A., CAULFIELD, C.P. & PELTIER, W.R. 2017*a* Role of overturns in optimal mixing in stratified mixing layers. *J. Fluid Mech.* **826**, 522–552.

MASHAYEK, A., CAULFIELD, C.P. & PELTIER, W.R. 2013 Time-dependent, non-monotonic mixing in stratified turbulent shear flows: implications for oceanographic estimates of buoyancy flux. *J. Fluid Mech.* **736**, 570–593.

MASHAYEK, A., FERRARI, R., MERRIFIELD, S., LEDWELL, J.R., ST. LAURENT, L. & GARABATO, A.N. 2017*b* Topographic enhancement of vertical turbulent mixing in the Southern Ocean. *Nat. Commun.* **8**, 14197.

MASHAYEK, A., FERRARI, R., NIKURASHIN, M. & PELTIER, W.R. 2015 Influence of enhanced abyssal diapycnal mixing on stratification and the ocean overturning circulation. *J. Phys. Oceanogr.* **45** (10), 2580–2597.

MASHAYEK, A. & PELTIER, W.R. 2011*a* Three-dimensionalization of the stratified mixing layer at high Reynolds number. *Phys. Fluids* **23**, 111701.

MASHAYEK, A. & PELTIER, W.R. 2011*b* Turbulence transition in stratified atmospheric and oceanic shear flows: Reynolds and Prandtl number controls upon the mechanism. *Geophys. Res. Lett.* **38** (16), L16612.

MASHAYEK, A. & PELTIER, W.R. 2012*a* The 'zoo' of secondary instabilities precursory to stratified shear flow transition. Part 1. Shear aligned convection, pairing, and braid instabilities. *J. Fluid Mech.* **708**, 5–44.

MASHAYEK, A. & PELTIER, W.R. 2012*b* The 'zoo' of secondary instabilities precursory to stratified shear flow transition. Part 2. The influence of stratification. *J. Fluid Mech.* **708**, 45–70.

MASHAYEK, A. & PELTIER, W.R. 2013 Shear-induced mixing in geophysical flows: does the route to turbulence matter to its efficiency? *J. Fluid Mech.* **725**, 216–261.

MASHAYEK, A., SALEHIPOUR, H., BOUFFARD, D., CAULFIELD, C.P., FERRARI, R., NIKURASHIN, M., PELTIER, W.R. & SMYTH, W.D. 2017*c* Efficiency of turbulent mixing in the Abyssal ocean circulation. *Geophys. Res. Lett.* **44** (12), 6296–6306.

MATER, B.D. & VENAYAGAMOORTHY, S.K. 2014 A unifying framework for parameterizing stably stratified shear-flow turbulence. *Phys. Fluids* **26** (3), 36601.

MATER, B.D., VENAYAGAMOORTHY, S.K., ST. LAURENT, L. & MOUM, J.N. 2015 Biases in thorpe-scale estimates of turbulence dissipation. Part I: assessments from large-scale overturns in oceanographic data. *J. Phys. Oceanogr.* **45** (10), 2497–2521.

MILES, J.W. 1961 On the stability of heterogeneous shear flows. *J. Fluid Mech.* **10**, 496–508.

MONISMITH, S.G., KOSEFF, J.R. & WHITE, B.L. 2018 Mixing efficiency in the presence of stratification: when is it constant? *Geophys. Res. Lett.* **45** (11), 5627–5634.

MOUM, J.N. 1996 Efficiency of mixing in the main thermocline. *J. Geophys. Res.* **101** (C5), 12–57.

NIKURASHIN, M. & LEGG, S. 2011 A mechanism for local dissipation of internal tides generated at rough topography. *J. Phys. Oceanogr.* **41**, 378–395.

OGLETHORPE, R.L.F., CAULFIELD, C.P. & WOODS, A.W. 2013 Spontaneous layering in stratified turbulent Taylor–Couette flow. *J. Fluid Mech.* **721**, R3.

OLSTHOORN, J. & DALZIEL, S.B. 2015 Vortex-ring-induced stratified mixing. *J. Fluid Mech.* **781**, 113–126.

OSBORN, T.R. 1980 Estimates of the local rate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.* **10**, 83–89.

PARK, Y.G., WHITEHEAD, J.A. & GNANADESKIAN, A. 1994 Turbulent mixing in stratified fluids: layer formation and energetics. *J. Fluid Mech.* **279**, 279–311.

PELTIER, W.R. & CAULFIELD, C.P. 2003 Mixing efficiency in stratified shear flows. *Annu. Rev. Fluid Mech.* **35**, 135–167.

POLZIN, K.L., TOOLE, J.M., LEDWELL, J.R. & SCHMITT, R.W. 1997 Spatial variability of turbulent mixing in the Abyssal ocean. *Science* **276** (5309), 93–96.

PORTWOOD, G.D., DE BRUYN KOPS, S.M. & CAULFIELD, C.P. 2019 Asymptotic dynamics of high dynamic range stratified turbulence. *Phys. Rev. Lett.* **122**, 194504.

PORTWOOD, G.D., DE BRUYN KOPS, S.M., TAYLOR, J.R., SALEHIPOUR, H. & CAULFIELD, C.P. 2016 Robust identification of dynamically distinct regions in stratified turbulence. *J. Fluid Mech.* **807**, R2.

VAN REEUWIJK, M., HOLZNER, M. & CAULFIELD, C.P. 2019 Mixing and entrainment are suppressed in inclined gravity currents. *J. Fluid Mech.* **873**, 786–815.

SALEHIPOUR, H. & PELTIER, W.R. 2015 Diapycnal diffusivity, turbulent Prandtl number and mixing efficiency in Boussinesq stratified turbulence. *J. Fluid Mech.* **775**, 464–500.

SALEHIPOUR, H., PELTIER, W.R. & CAULFIELD, C.P. 2018 Self-organised criticality of turbulence in strongly stratified mixing layers. *J. Fluid Mech.* **856**, 228–256.

SCHUMANN, U. & GERZ, T. 1995 Turbulent mixing in stably stratified shear flows. *J. Appl. Meteorol.* **34** (1), 33–48.

SCOTTI, A. 2015 Biases in thorpe-scale estimates of turbulence dissipation. Part II: energetics arguments and turbulence simulations. *J. Phys. Oceanogr.* **45** (10), 2522–2543.

SHERMAN, F.S., IMBERGER, J. & CORCOS, G.M. 1978 Turbulence and mixing in stably stratified waters. *Annu. Rev. Fluid Mech.* **10** (1), 267–288.

SHIH, L.H., KOSEFF, J.R., IVEY, G.N. & FERZIGER, J.H. 2005 Parameterization of turbulent fluxes and scales using homogeneous sheared stably stratified turbulence simulations. *J. Fluid Mech.* **525**, 193–214.

SMITH, J.A. 2020 A comparison of two methods using thorpe sorting to estimate mixing. *J. Atmos. Ocean. Technol.* **37** (1), 3–15.

SMYTH, W.D. 2020 Marginal instability and the efficiency of ocean mixing. *J. Phys. Oceanogr.* **50**, 2141–2150.

SMYTH, W.D., MOUM, J. & CALDWELL, D. 2001 The efficiency of mixing in turbulent patches: inferences from direct simulations and microstructure observations. *J. Phys. Oceanogr.* **31**, 1969–1992.

SMYTH, W.D. & MOUM, J.N. 2000 Length scales of turbulence in stably stratified mixing layers. *Phys. Fluids* **12**, 1327.

SMYTH, W.D., NASH, J.D. & MOUM, J.N. 2019 Self-organized criticality in geophysical turbulence. *Sci. Rep.* **9** (1), 3747.

TAYLOR, G.I. 1935 Statistical theory of turbulence IV-diffusion in a turbulent air stream. *Proc. R. Soc. Lond.* A **151** (873), 465–478.

THORPE, S.A., SMYTH, W.D. & LI, L. 2013 The effect of small viscosity and diffusivity on the marginal stability of stably stratified shear flows. *J. Fluid Mech.* **731**, 461–476.

THORPE, S.A. 2005 *The Turbulent Ocean*. Cambridge University Press.

THORPE, S.A. & LIU, Z. 2009 Marginal instability? *J. Phys. Oceanogr.* **39**, 2373–2381.

TURNER, J.S. 1973 *Buoyancy Effects in Fluids*. Cambridge University Press.

VILLERMAUX, E. 2019 Mixing versus stirring. *Annu. Rev. Fluid Mech.* **51**, 245–273.

WATERHOUSE, A.F., *et al.* 2014 Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate. *J. Phys. Oceanogr.* **44** (7), 1854–1872.

WEINSTOCK, J. 1992 Vertical diffusivity and overturning length in stably stratified turbulence. *J. Geophys. Res.* **97** (21), 653–658.

WELLS, M., CENEDESE, C. & CAULFIELD, C.P. 2010 The relationship between flux coefficient Γ and entrainment ratio E in density currents. *J. Phys. Oceanogr.* **40**, 2713–2727.

WINTERS, K., LOMBARD, P., RILEY, J. & D'ASARO, E.A. 1995 Available potential energy and mixing in density-stratified fluids. *J. Fluid Mech.* **289**, 115–128.

WOODS, A.W., CAULFIELD, C.P., LANDEL, J.R. & KUESTERS, A. 2010 Non-invasive mixing across a density interface in a turbulent Taylor–Couette flow. *J. Fluid Mech.* **663**, 347–357.

WOODS, J.D. 1968 Wave-induced shear instability in the summer thermocline. *J. Fluid Mech.* **32**, 791–800.

ZHOU, Q., TAYLOR, J.R. & CAULFIELD, C.P. 2017 Self-similar mixing in stratified plane Couette flow for varying Prandtl number. *J. Fluid Mech.* **820**, 86–120.