Role of overturns in optimal mixing in stratified mixing layers

A. Mashayek[1], C. P. Caulfield[2,3] and W. R. Peltier[4]

[1]Department of Earth Atmosphere and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, 02139, USA

[2]BP Institute, University of Cambridge, Madingley Road, Cambridge CB3 0EZ, UK

[3]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

[4]Department of Physics, University of Toronto, Ontario, M5S 1A7, Canada

(Received ?? and in revised form ??)

Turbulent mixing plays a major role in enabling the large scale ocean circulation. The accuracy of mixing rates estimated from observations depends on our understanding of basic fluid mechanical processes underlying the nature of turbulence in a stratified fluid. Several of the key assumptions made in conventional mixing parameterizations have been increasingly scrutinized in recent years, primarily on the basis of adequately high resolution numerical simulations. We add to this evidence by compiling results from a suite of numerical simulations of the turbulence generated through stratified shear instability processes. We study the inherently intermittent and time-dependent nature of wave-induced turbulent life cycles and more specifically the tight coupling between inherently anisotropic scales upon which small scale isotropic turbulence grows. The anisotropic scales stir and stretch fluid filaments enhancing irreversible diffusive mixing at smaller scales. We show that the characteristics of turbulent mixing depend on the relative time evolution of the Ozmidov length scale L_O compared to the so-called Thorpe overturning scale L_T which represents the scale containing available potential energy upon which turbulence feeds and grows. We find that when $L_T \sim L_O$, the mixing is most active and efficient since stirring by the largest overturns becomes 'optimal' in the sense that it is not suppressed by ambient stratification. We argue that the high mixing efficiency associated with this phase, along with observations of $L_O/L_T \sim 1$ in oceanic turbulent patches, together point to the potential for systematically underestimating mixing in the ocean, if the role of overturns is neglected. This neglect, arising through the assumption of a clear separation of scales between the background mean flow and small scale quasi-isotropic turbulence, leads to the exclusion of an highly efficient mixing phase from conventional parameterizations of the vertical transport of density. Such an exclusion may well be significant if the mechanism of shear-induced turbulence is assumed to be representative of at least some turbulent events in the ocean. While our results are based upon simulations of shear instability, we show that they are potentially more generic by making direct comparisons with L_T−L_O data from ocean and lake observations which represent a much wider range of turbulence-inducing physical processes.

1. Introduction

Diapycnal turbulent mixing plays a primary role in enabling the large scale ocean circulation (Wunsch & Ferrari 2004). Over the past several decades, significant investment has been made in estimating the strength of diapycnal mixing on the basis of observations of ocean turbulence (see e.g. St. Laurent & Simmons 2006; Waterhouse *et al.* 2014, for reviews). Four common assumptions concerning density stratified turbulence, made for practical purposes in conventional methods employed for the estimation of mixing from observations are that the turbulence is (I) fully developed, (II) stationary, (III) and isotropic, and that (IV) there exists a clear separation of scales between the background mean flow and the superposed isotropic turbulence. In recent years, numerical simulations have become just powerful enough to aid in quantification of inaccuracies associated with these assumptions (Ivey *et al.* 2008; Pham & Sarkar 2010; Mashayek & Peltier 2013; Mashayek *et al.* 2013; Salehipour *et al.* 2015; Salehipour & Peltier 2015; Salehipour *et al.* 2016*a*).

A common hypothesis is that shear-driven mixing in the ocean is at least partially induced by the breaking of internal waves excited by tides and geostrophic motions in the deep ocean or by winds at the surface (Garrett 2003; Nikurashin & Ferrari 2011; Alford & Pinkel 2000). Such mixing comprises many individual breaking events each of which is non-stationary in time. It is at least plausible that some of these breaking events may be considered to be generated by shear instabilities on scales small compared to the internal waves. Such shear instability generated mixing may be characterised by a multi-stage life-cycle. A preparatory period of growth of the internal wave amplitude leads to an initial period of shear instability growth, break down through secondary instabilities triggering a transition to turbulence. This initial period is followed by an intermediate period of what might be considered to be fully-developed turbulence, followed ultimately by a final decay period. Contrary to common assumptions in parameterization schemes (Mashayek & Peltier 2013), in this scenario of shear instability generated mixing the contribution of the intermediate 'fully-developed' period does not necessarily dominate the net vertical cross-density flux of mass and tracers, even at very high flow Reynolds numbers. Furthermore, even in the most turbulent intermediate period, turbulence can be highly non-stationary and anisotropic comprising a range of scales between that of small scale quasi-isotropic turbulence and that of the background mean flow, particularly when there is a dominant shear direction imposed by some 'external process', for example through the intensification of an appreciably larger scale internal wave (Fritts *et al.* 2003; Ivey *et al.* 2008; Mashayek & Peltier 2013; Mashayek *et al.* 2013). Figure 1, produced from results of a numerical simulation to be discussed in detail later, illustrates the cascade of instabilities which form upon a shear instability overturn and which eventually destroy billow coherence. As we will discuss in the paper, this anisotropic highly time-dependent turbulence transition phase of flow makes a major contribution to the net vertical mixing of mass over the entire life cycle of this type of turbulence.

Recently, Mashayek & Peltier (2013) (hereafter MP13) and Mashayek *et al.* (2013) (hereafter MCP13) presented computation-based evidence for breakdown of assumptions I-III when the turbulence is triggered by a initial shear instability. In two important papers (Smyth & Moum 2000*b*,*a*), Smyth & Moum effectively addressed assumptions III and IV (though they did not couch the discussion in precisely those terms) Crucially, their simulations were at signficiantly lower Reynolds number than is now achievable, and thus in particular the shear instabilities they simulated were not prone to the full 'zoo' of secondary instabilities identified in Mashayek & Peltier (2012*a*) and Mashayek & Peltier (2012*b*), and so the subsequent analysis of the turbulence properties is inevitably affected by the absence of physical processes present in geophysically relevant higher Reynolds number flows. In this study, we build on the work of Smyth & Moum (2000*b*) (hereafter SM00) to focus on assumption IV. analyzing data from a more complete set of numerical simulations at substantially higher Reynolds number closer to values representative of energetic ocean mixing zones. In particular we will extend their analysis of scales of turbulence. Through this analysis, we demonstrate that assumption IV may at

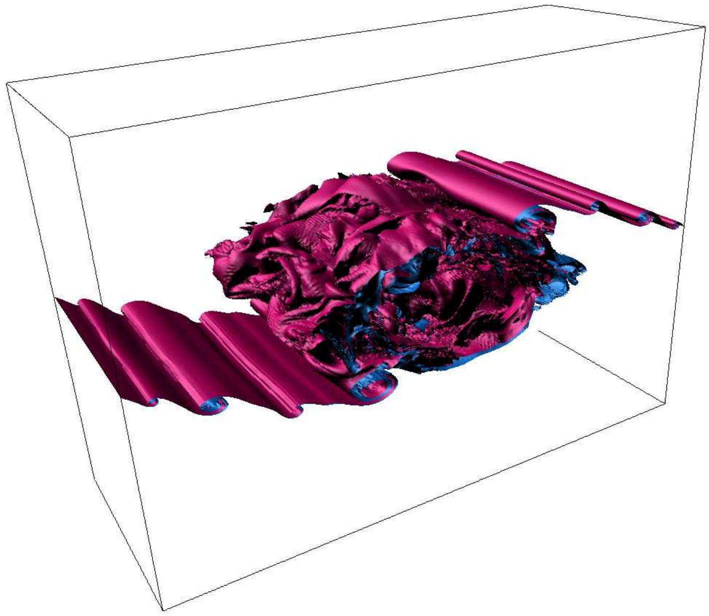

FIGURE 1. Snapshot of turbulence breakdown and mixing due to breaking of an overturning by shear instability in a stably stratified layer (case 12 in table 1). Purple and blue represent light and heavy density iso-surfaces, respectively. The snapshot corresponds to time $t = 80h/\Delta u$, where h is half the initial shear layer depth and Δu is half the total velocity difference.

best hold in only a rather narrow part of the lifecycle for rather special shear instabilities, implying that extending a model based fundamentally on this assumption over the whole turbulence life cycle may well introduce large uncertainty and/or inaccuracy in estimates of net turbulent mixing over the life cycle of an individual wave breaking event, if that wave breaking is generated by the onset of shear instabilities. Wave-induced turbulence in energetic oceanic regions is determined by the combination of many individual breaking events, both essentially isolated in space and time and yet dynamically coupled in some way. Therefore, there is no a-priori basis upon which it can be assumed that the inaccuracies we discuss in this work will have negligible effect in the much more complex real ocean. Of course, it is always important to remember that our results are based on modelling individual wave breaking events in the highly idealized configuration that the vertical shear and density distribution induced by the intensification of the internal waves may be taken to be at least quasi-steady on the time scale of the development of shear instabilities on those distributions.

There has been an increasing recent interest in description of shear induced density stratified turbulent mixing in terms of key physical length scales (see e.g. Mater *et al.* 2013; Scotti 2015), and we will focus herein on the critical importance of the time dependence of characteristic length scales for mixing in a stratified shear flow. Understanding the relative time dependence of length scales within the flow is of general interest, as estimates of diapycnal mixing are often constructed from instantaneous measurements of specific length scales (see Thorpe 2005, for an overview).

Employing shear instability as a canonical mixing agent, our focus will be upon the lasting effect of the primary 'overturning' associated with the primary shear instability which leads to 'efficient' (in a way we define precisely in section 6) irreversible mixing. An important implication of our analyses is that mixing efficiencies may be under-estimated in regions of the ocean in which large overturns are expected since they provide a significant reservoir of energy upon which a broad inertial subrange of turbulence may draw so as to support efficient irreversible mixing. The most or 'optimal' efficient mixing will be shown to occur at the instant during flow evolution when the scale at which energy is injected, through overturning into the turbulence cascade at the upper bound of the inertial subrange becomes sufficiently small to avoid suppression by the ambient stratification. This core idea (as we discuss further below) is consistent with the arguments presented by Ivey & Imberger (1991), though for our flows, the associated value of the mixing efficiency in this 'optimal' situation is found to be higher. Of course it will remain an important issue as to whether the specific model of shear instability generated turbulence that we will employ as basis for our analyses, relying upon the classical Kelvin-Helmholtz instability (KHI), may be considered sufficiently representative of spatio-temporally intermittent, relatively large scale wave breaking processes in general to enable our results to stand without caveat. For example, one key issue is the role of ambient, larger-scale background stratification in the development and break down of shear instabilities. There does exist evidence, however, in support of the relevance of KHI-based analysis for the understanding of stratified turbulence in general (Smyth *et al.* 2001; Bouffard & Boegman 2013; Scotti 2015). We will provide some of the evidence of the generality of the utility of this model of stratified turbulent processes by comparing results from direct numerical simulations with observations.

This paper is organized as follows. In section 2 we briefly describe the suite of turbulence simulations upon which our analyses will be based. Section 3 will provide definitions of the important length scales that may be employed to characterize shear-driven stratified mixing events. Section 4 presents a detailed discussion of the time dependence of the evolution of these scales, focusing especially on what may be considered their generic behaviour in stratified shear-driven mixing at sufficiently high Reynolds number. In section 5 we discuss the importance of the relative evolution of the Ozmidov and Thorpe length scales for quantification of the age of turbulence. In section 6 we briefly discuss the implications of our results and in particular discuss in section 5 the quantitative representation of mixing in geophysically relevant circumstances. Conclusions are offered in the final section 7.

2. Primary shear instability

In this section we discuss the numerical datasets that will be employed to study turbulence transition of primary shear instabilities as well as the bulk dimensionless parameters which characterize them.

2.1. *Numerical simulations*

We employ a suite of high resolution direct numerical simulations (DNS) of the turbulence life cycle of finite-amplitude Kelvin-Helmholtz (KH) billows in stratified shear layers, a common mechanism leading to turbulence transition in the ocean (Smyth & Moum. 2012; Mashayek 2013). The data to be employed are summarized in table 1 and consist of the same set of numerical simulations as were previously analyzed in MP13 and MCP13 for other purposes, augmented by three new simulations, as noted in the table. Each of these simulations describes the three dimensional temporal evolution of a horizontally periodic stably stratified shear layer with the initial background velocity profile $\bar{u}(z)$ and Boussinesq density profile $\bar{\rho}(z)$ defined as

$$\bar{u}(z) = \Delta u\ \tanh\left(\frac{z}{h}\right); \quad \bar{\rho}(z) = \rho_a - \Delta\rho\ \tanh\left(\frac{z}{h}\right), \tag{2.1}$$

case	Re_0	Ri_0	Re_t	Re_b	η_c^{3D}	pairing allowed	source
1	750	0.04	5200	998	0.24	yes	this study
2	4000	0.04	27500	7012	0.44	yes	MP13
3	10000	0.04	68750	12261	0.62	yes	MP13
4	750	0.12	1700	135	0.18	yes	this study
5	1000	0.12	2300	180	0.22	yes	MP13
6	2000	0.12	4600	300	0.32	yes	this study
7	4000	0.12	9200	640	0.32	yes	MP13
8	6000	0.12	13750	704	0.36	yes	MP13
9	8000	0.12	18350	817	0.40	yes	MP13
10	10000	0.12	22900	1012	0.42	yes	MP13
11	6000	0.14	11800	614	0.30	no	MCP13
12	6000	0.16	10300	586	0.29	no	MCP13
13	6000	0.18	9200	413	0.28	no	MCP13
14	6000	0.20	8250	131	0.23	no	MCP13

TABLE 1. Parameter values for the numerical simulations analyzed in this paper. $Pr = 1$ for all cases. The initial Reynolds number Re_0, the initial minimum Richardson number at $z = 0$ Ri_0, the effective Reynolds number Re_t at the start of the fully-developed turbulent period t_{3D}^S, and the cumulative turbulent mixing efficiency η_c^{3D} are all defined in the text.

where Δu and $\Delta\rho$ are half the velocity and density variation, h is half the shear layer thickness, and $\rho_a \gg \Delta\rho$ is the reference density. As reviewed in MP13, this configuration has come to be seen as a the standard model problem for the study of mixing induced by large-scale, overturning shear instabilities. As noted in the introduction, there is an underlying assumption that this background flow distribution may be taken to be steady, and so if it is induced by the intensification of an even larger-scale internal wave, the evolution of that wave occurs on time scales which are long compared to the time scales of the evolution of the primary shear instability of this flow distribution.

2.2. *Governing dimensionless parameters*

Three nondimensional numbers characterize the flow for each case, namely an appropriate Reynolds number Re, quantifying the ratio of inertial to viscous forces, an appropriate Richardson number Ri, quantifying the ratio of buoyancy to inertial forces and the Prandtl number $Pr = \nu/\kappa_m$, the ratio of molecular kinematic viscosity to molecular thermal diffusivity. The initial Reynolds number $Re_0 = \Delta u\ h/\nu$ for each of the simulations of turbulent collapse to be analyzed is listed in Table 1, and is defined based on a length scale that is half the shear layer thickness and a velocity scale that is half the velocity difference across the initial density inversion upon which the shear is imposed prior to its evolution through primary instability into the classical Kelvin-Helmholtz billow form. Indeed, since we are primarily interested in the turbulent phase of flow evolution, the nonlinear Kelvin-Helmholtz billow itself being an essentially laminar structure, a more relevant definition of the Reynolds number might be one based upon a length scale determined by the half shear layer thickness at the onset of turbulence (to be defined in (3.2)), which is larger than the initial layer's half thickness. This modified Reynolds number is denoted by $Re_t > Re_0$ in the table and might usefully be viewed as the relevant parameter for comparison with shear instabilities observed in nature.

Turbulent mixing events associated with the evolution of a Kelvin-Helmholtz billows are strongly time-dependent and transient. Therefore, it is appropriate to define a criterion to identify the time of onset of turbulence which may be considered to be 'fully-developed'. Following Caulfield & Peltier (2000) and MP13, we monitor the inherently three-dimensional turbulent kinetic energy at scales smaller than the Ozmidov scale (representing the size of the largest eddies not suppressed by stratification; to be defined in the next section). Generically, this scale-selected turbulent kinetic energy reaches a maximum magnitude (with respect to time) following a rapid growth during turbulence transition associated with the break down of the primary Kelvin-Helmholtz billow. We identify the onset of what we refer to as fully-developed turbulence with this time of maximum magnitude, which time was named t^S_{3D} (or t_{3D} when context allowed) in Mashayek *et al.* (2013), and Re_t is also evaluated at this time.

It is important to remember that our convention for the definition of Re_0 is different from that used by SM00, which used the total shear layer depth and the total velocity difference. Using our convention, their simulations had $340 < Re_0 < 1250$, with the majority of the simulations being conducted at $Re_0 \simeq 500$. As we demonstrate further below, the absence of the full 'zoo' of instabilities discussed in Mashayek & Peltier (2012*a*) and Mashayek & Peltier (2012*b*). means the properties of flows with such Reynolds numbers are qualitatively different from flows with $Re_0 \gtrsim 4000$ in this 'fully-developed' turbulence stage of flow evolution, and so it is of value to revisit and extend their analyses at such larger Re_0.

The (minimum) bulk Richardson number, $Ri_0 = g\Delta\rho\ h/(\rho_a(\Delta u)^2)$, which applies initially at the midpoint of the shear layer, is also listed in the table. To keep the problem tractable, for practical reasons we avoid varying the Prandtl number and set $Pr = \nu/\kappa_m = 1$. It is important, however, to appreciate that there is recent evidence that the small-scale characteristics of turbulent mixing are affected by larger, more physically relevant values of Pr (Klaassen & Peltier (1985*a*), SM00, Mashayek & Peltier (2011); Bouffard & Boegman (2013); Salehipour *et al.* (2015); Salehipour & Peltier (2015)) even at relatively high values of the Reynolds number. A further important nondimensional parameter, insofar as the characteristics of stratified turbulent mixing are concerned, is the so-called buoyancy Reynolds number Re_b:

$$Re_b = \mathcal{E}/(\nu N^2), \tag{2.2}$$

where here we defin this parameter in terms of an appropriately externally-determined buoyancy frequency 'N' and the (total) kinetic energy dissipation rate $\mathcal{E}$, defined as

$$\mathcal{E} = \frac{\nu}{2V}\int\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)^2 dV, \tag{2.3}$$

where V is the volume of the part of the domain that encompasses the mixing layer (to be defined in the next section), and the Einstein summation convention has been employed. Consistent with the scaling arguments originally presented by Gibson (1980) in support of his concept of 'fossil turbulence', energetic stratified turbulence can be maintained in a form not substantially affected by viscosity for $Re_b \sim \mathcal{O}(10^2)$ or higher with viscous suppression occurring once Re_b falls below $\sim \mathcal{O}(10)$ (Ivey & Imberger (1991), SM00, Thorpe (2005); Ivey *et al.* (2008)). While $\mathcal{O}(10^2) < Re_b < \mathcal{O}(10^3)$ is estimated to be relevant to mixing events in the thermocline and upper (pelagic) ocean, values of $Re_b \sim \mathcal{O}(10^3)$ and larger have been reported in the energetic abyssal oceans where mixing plays a key role in maintaining the ocean meridional overturning circulation (Gargett *et al.* 1984; Itsweire *et al.* 1993; Smyth & Moum 2001; Thorpe 2005; Mashayek *et al.* 2017).

Despite many attempts to characterize stratified turbulence in terms of Re_b alone, it is well-known that on the basis of both dimensional argument and physical understanding it is not sufficient(Mater & Venayagamoorthy 2014; Mashayek 2013; Salehipour *et al.* 2016*b*). A key issue concerning the use of Re_b alone to classify and parametrize turbulence properties in a stratified flow is the time-dependence of the dissipation rate $\mathcal{E}$ (and indeed the spatial dependence of dissipation when not spatially averaged), making it problematic to identify a particular value of Re_b with a specific mixing event. Indeed, for shear-driven turbulence, the dissipation rate $\mathcal{E}$ varies strongly with time, and does not actually exhibit any period when it is not varying strongly. Therefore, it is appropriate to think of a particular mixing event as sampling a range of Re_b, typically growing to a maximum value rapidly as the flow undergoes the transition to turbulence, before decaying with time as the flow relaminarises.

Finally, the range of Ri_0 considered in this study is $0.04 < Ri_0 < 0.2$. For the particular velocity and density profiles defined in (2.1), Ri_0 is the minimum initial value of the (local) gradient Richardson number $Ri_g(z,t)$ defined as

$$Ri_g(z,t) = \frac{-\frac{g}{\rho_a}\frac{\partial\langle\rho\rangle}{\partial z}}{\left(\frac{\partial\langle u\rangle}{\partial z}\right)^2}, \tag{2.4}$$

where angle brackets denote horizontal averaging. The bound $Ri_0 = 0.2$ is chosen to be below the classical value of 1/4 for the global minimum value of Ri_g associated with linear stability of stratified shear flows, according to the Miles-Howard criterion (Miles 1961; Howard 1961). Ri_0 represents the minimum Richardson number in the preturbulent shear layer and so cannot be directly compared to observation-based local estimates of Ri_0, since such observation-based estimates are inevitably bulk estimates, due tothe lack of resolution in the measurement of background shear. An effective bulk measure of the Richardson number Ri based on velocity and density jumps across the entire vertical extent of the mixing region in our simulations is typically $\sim \mathcal{O}(1)$ throughout the turbulent phase of flow evolution.

Cases in Table 1 are divided into two categories with respect to the possibility of an upscale cascade through pairing instability. The simulations previously reported in MP13 extended over two wavelengths of the primary shear instability in the streamwise direction, thus allowing for pairing to occur. However, it was shown in MP13 (for $Pr = 1$) and Salehipour *et al.* (2015) (for $Pr > 1$) that the pairing instability is suppressed as the Reynolds number increases, and that for $Pr = 1$, it becomes significantly diminished for $Re_0 \geqslant 6000$. Thus, the simulations in MCP13 (which were all for $Re_0 = 6000$) imposed streamwise periodicity over only one wavelength of the primary instability. However, as we discuss below in more detail, the degree to which pairing is diminished at high Re influences the properties of turbulence sufficiently to bring previously suggested parameterizations of turbulence into question. Therefore, we have included both types of simulations here, clearly marking those simulations for which pairing is allowed and recognizing that if these simulations were to be repeated at even higher relevant Reynolds number the residual influence of an upscale component of the turbulent cascade could be further mitigated, if not completely eliminated.

It is important to note that in the limit of extremely small Richardson number corresponding to effectively unstratified shear layers, the transition to turbulence may be dominated by vortices which grow on the braid of KH billows rather than in the 'eyelids'. Such braid-centred vortices have a much longer spanwise length scale than the core-centered convective or shear instabilities (Klaassen & Peltier 1985*b*; Caulfield & Peltier 1994; Smyth & Peltier 1994; Potylitsin & Peltier 1999, 1998; Caulfield & Peltier 2000). The spanwise extent of the computational domains were selected according to their corresponding Richardson number in such a way as to resolve the expected developing secondary perturbations.

3. Definition of length scales of turbulence

In this section we introduce various length scales which we invoke to characterize certain aspects of shear-driven stratified mixing events. As discussed in SM00, a natural way to compare length scales for shear flows with different initial minimum Richardson numbers is to nondimensionalise with the (constant for a particular simulation) length scale L_{sc} defined as

$$L_{sc} = \rho_a \Delta u^2/(4g\Delta\rho) = h/(4Ri_0), \tag{3.1}$$

i.e. the notional length scale expressed in terms of the initial velocity difference and density difference which amounts to an initial (bulk) Richardson number with the Miles-Howard marginal value of 1/4.

We consider four dynamically determined and, crucially, inherently time-dependent characteristic length scales, namely the Kolmogorov (L_K), Ozmidov (L_O), Corrsin (L_C) and Thorpe (L_T) scales. All of these scales typically vary significantly during the three distinct periods of the turbulence life cycle discussed in the introduction: an initial or early period of transition to turbulence in which energy is transferred from the background kinetic energy into turbulent kinetic energy (TKE) due to the 'break down' of the organized flow; an intermediate period of sustained energetic stratified turbulence; and a final or late period during which this turbulence decays and the flow relaminarises. We note that while L_K, L_O and L_C are most relevant during the fully turbulent phase of the flow, their formal consideration in earlier phases is helpful for the purposes of the discussions to follow.

To define these characteristic scales in an internally consistent way, it is necessary to obtain an estimate of evolution of the thickness of the initial shear and density layers upon which turbulence grows. Following SM00, we define two integral scales I_ρ and I_u which track the evolution of both thicknesses during the three periods of the shear layer's turbulent evolution:

$$I_\rho(t) = \int_{-L_z/2}^{L_z/2} \left[1 - \left(2\frac{\langle\rho\rangle}{\Delta\rho}\right)^2\right] dz, \qquad I_u(t) = \int_{-L_z/2}^{L_z/2} \left[1 - \left(2\frac{\langle u\rangle}{\Delta u}\right)^2\right] dz. \tag{3.2}$$

where angle brackets denote horizontal averaging. Both scales are defined to have the same thickness as the initial density and shear layers at the onset of the flow evolution, and will vary with time as a consequence of turbulent mixing. Since in our study the Prandtl number is 1, the ratio of these two scales is close to 1. In all definitions and analysis to be provided from this point on, spatial and volume averages are limited in the vertical to the mixing layer as defined by the above-defined time-dependent length scale $I_u(t)$, i.e. over the interval $[-I_u/2, I_u/2]$. In particular Re_t in table 1 is defined using $I_u/2$ at the time when the inherently three-dimensional turbulent kinetic energy reaches its maximum value (i.e. t^S_{3D} as discussed in more detail in Mashayek *et al.* (2013)).

Using these integral scales, the instantaneous representations of background velocity shear, background buoyancy frequency, and Richardson number become:

$$S_b(t) = \frac{\Delta u}{I_u(t)}, \qquad N_b(t) = \sqrt{\frac{g\Delta\rho}{I_\rho(t)}}, \qquad Ri(t) = \frac{g\Delta\rho/I_\rho(t)}{(\Delta u/I_u(t))^2} = \frac{N_b^2}{S_b^2}. \tag{3.3}$$

3.1. *Thorpe scale L_T*

The first of the four scales we discuss is the so-called 'Thorpe scale' L_T, which is a measure of net vertical parcel displacements associated with turbulent mixing. The Thorpe scale calculated from the 3D numerical simulations (L_T^{3D}) is determined by a sorting of the density field $\rho(x, y, z, t)$ into a temporally evolving statically stable staircase of fluid parcels. L_{3D}^T is then the rms of the vertical displacement of the particles from their actual position to the vertical position in the sorted density field. This approach follows previous studies (Winters *et al.* 1995; Caulfield & Peltier 2000). During the sorting process, the horizontal area of each fluid parcel in the mesh in terms of which the numerical simulation is described is set to that of the full domain, and its vertical thickness is adjusted so as to conserve mass. This method leads to a statically stable vertical distribution of density within the domain with the same volume (and hence mass due to the Boussinesq approximation) as the unsorted domain, but one which possesses the minimum potential energy that any adiabatic re-ordering of the discrete fluid particles in the domain could achieve at a given time during flow evolution. The rms of the vertical displacement that each fluid parcel experiences in this sorting procedure is by definition the 3D Thorpe scale L_T^{3D}. As discussed in SM00, this estimate will differ from the Thorpe scale calculated by sorting entire individual water columns, but typically that difference is found to be relatively small. More specifically, the column wise estimate is a measure of overturnings in the flow, whereas the 3D Thorpe scale is a more general representation of density displacements and is meaningful even in the absence of overturnings or when recognizably large scale overturnings have collapsed into fine scale turbulence. Hereafter we will choose L_T^{3D} to be the appropriate time-dependent characteristic measure of overturning and will simply refer to it as L_T. This is a different convention from that employed in SM00, who used L_T to refer to the column-wise estimate, which must be distinguished from our full 3D estimate L_T^{3D}. In Appendix II we discuss differences between the two and their implications for the relevance of our work to oceanographic estimates of the Thorpe scale based on column sorting.

In so far as evolution of L_T in shear instabilities of KH type is concerned, L_T is expected to grow during the initial growth of the primary billows (either precursory to or concurrent with turbulence transition) and it is expected to decrease as the flow mixes thoroughly and relaminarises. As will be discussed in what follows, the evolution of L_T also depends on whether vortex pairing occurs or not. Thus, our simulations differ from those in SM00 since their simulations were initiated with an eigenmode of pairing instability. In the subset of our simulations in which the domain is sufficiently large to house vortex pairing, pairing occurs at low Reynolds number but its onset is a function of Richardson number and pairing also gets increasingly suppressed at higher Reynolds numbers. These subtle differences between the various cases discussed herein and in SM00 (independently of the wide differences in Re_0) have implications for L_T evolution and the relevance of L_T/L_O as a proxy for turbulence age. We return to this in section 5.

3.2. *Ozmidov length L_O, Corrsin length L_C & Kolmogorov length L_K*

The Thorpe scale L_T is a purely geometrical construct, and is defined in terms of properties of the evolving density field alone, with no explicit dependence on the flow velocity field, with the connection being entirely implicit due to the evolving flow dynamics. To characterize turbulence, it is helpful to resort to length scales constructed based on both intrinsic properties of turbulence such as the spatially averaged total kinetic energy dissipation rate $\mathcal{E}$ and bulk external properties such as the background density gradient and velocity shear. Ozmidov and Corrsin scales are defined in terms of such quantities. The (total) dissipation rate has dimensions L^2T^{-3}, and so we define L_O and L_T as the two natural length scales relating the dissipation rate to the background buoyancy frequency $N_b(t)$ and the background shear $S_b(t)$ given in (3.3) through

$$L_O(t) = \left(\frac{\mathcal{E}}{N_b^3}\right)^{1/2};\ L_C(t) = \left(\frac{\mathcal{E}}{S_b^3}\right)^{1/2} \rightarrow Ri(t) = \left(\frac{L_C}{L_O}\right)^{2/3}. \tag{3.4}$$

Physically, for vertical scales larger than both Ozmidov and Corrsin scales, turbulence with sufficiently elevated values of the dissipation rate noticeably 'feels' the influence of stratification and shear.

As discussed in SM00, the temporal evolution of L_C and L_O are broadly similar, although in general $L_C < L_O$, unsurprisingly due to the relationship to $Ri(t)$ as defined in (3.4). In a shear layer of the kind considered here, both N_b and S_b decrease with time, due to the thickening of the mixing layer captured by the increases in the integral length scales I_ρ and I_u respectively. Therefore, the time evolution of both L_O and L_C is dominated by the time dependence of the (total) dissipation rate $\mathcal{E}$, as defined in (3.4) with both reaching their peak values during the most energetic intermediate period of turbulence in which the flow is replete with secondary and higher order instabilities. Similarly to L_T, we also expect L_O (and L_C) to decay as the turbulence decays, as $\mathcal{E}$ markedly decreases from its peak value.

The total dissipation rate may also be used to define a further natural length scale, namely the Kolmogorov dissipation scale L_K, where

$$L_K = \left(\frac{\nu^3}{\mathcal{E}}\right)^{1/4}, \tag{3.5}$$

and represents the scale below which the smallest eddies in the momentum field are viscously dissipated. Since in our cases$Pr = 1$, this is also the scale at which diffusion completely homogenizes the density field (i.e. $L_K = L_B = (\nu\kappa^2/\mathcal{E})^{1/4}$ where the latter is the Batchelor scale). Unlike L_O and L_C, L_K reaches its minimum value during the intermediate period when the turbulence is most energetic and hence the dissipation rate is largest. Before the flow is turbulent, or during the late turbulent decay period of the flow, L_K tends to an asymptotic value set by the small finite rate of dissipation of kinetic energy associated with the laminar shear layer, since here we choose to define L_K using the total dissipation rate $\mathcal{E}$, which does not tend to zero when the flow is laminar. Similarly, L_O and L_C are also defined using $\mathcal{E}$, and so these length scales are still well-defined during the stage of flow evolution when the transition to turbulence is occuring.

3.3. *Relative magnitudes of the scales*

Consistently with the results of SM00 for flows with substantially smaller Re_0, early in the flow evolution, L_T can be substantially larger than L_O, even when L_O is defined using the total dissipation rate. We investigate this scale separation in the next section. The turbulent dynamics at this early stage are highly anisotropic due to the influence of shear and stratification on scales above the Ozmidov scale, and the properties of the turbulence can be changing rapidly. The scales between L_O and L_C are still anisotropic, but largely influenced by shear alone, while the scales between L_C and L_K may be considered to exhibit nearly isotropic three-dimensional turbulence, provided of course that there is sufficient scale separation between L_C and L_K to allow for an inertial cascade. Indeed, since we expect $L_C \lesssim L_O$, this requirement for sufficient scale separation to allow for an inertial cascade of isotropic turbulence is typically unaffected by the background stratification. L_T, L_O, L_C and L_K are all strongly dependent on Re_0 and Ri_0, as well as typically strongly time-dependent. In section 6 we will show that the extent to which these various sub-ranges vary, and indeed even exist in any meaningful sense, has important implications for the irreversible mixing properties of the flow.

It is important to note that while L_C, L_K and L_O are mathematically well defined even in the laminar state of the flow, they only become dynamically relevant when the total dissipation is dominated by turbulent dissipation rather than the laminar phase which is

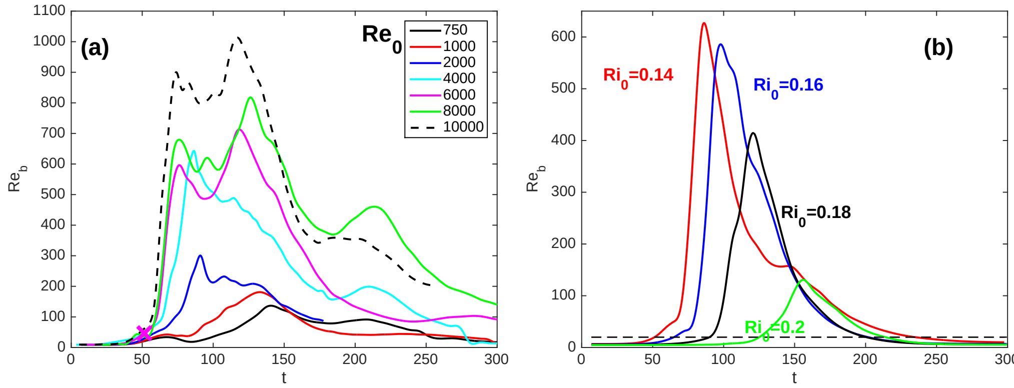

FIGURE 2. Time variation of buoyancy Reynolds number Re_b for: a) cases 4-10 of table 1, showing the variation with Re_0 for $Ri_0 = 0.12$, all with vortex pairing allowed (noting that pairing is increasingly suppressed as Re_0 increases); b) cases 11-14 of table 1, showing the variation with Ri_0 for $Re_0 = 6000$ for simulations with vortex pairing prohibited by design. Time is non-dimensionalised by the eddy turnover timescales $h/\Delta u$ where Δu and h are characteristic scales of the shear flow as defined in (2.1). The onset of fully-developed turbulence for each case corresponds to the time t^S_{3D} when the inherently three-dimensional turbulent kinetic energy peaks following a rapid growth during the transition to turbulence (see Caulfield & Peltier (2000) and Mashayek *et al.* (2013) for details). This time approximately coincides with first peak of L_O and also of Re_b as defined here. The dashed line in the second panel marks $Re_b = 20$ which nominally marks the lower bound of stratified turbulence, even if not truly fully-developed (see SM00 for a further discussion).

only weakly dissipative. As we will show, the sharp increase in the total kinetic energy dissipation rate $\mathcal{E}$ during the rapid transition to turbulence marks sharp changes in these scales in a way which will allow us to employ their evolution through the transition process to understand the mixing properties of the flow better.

3.4. *The buoyancy Reynolds number in terms of length scales*

It is instructive to note that the buoyancy Reynolds number can now be naturally interpreted as a ratio of length scales. If we choose to use N_b as defined in (3.3) as the appropriate choice for the buoyancy frequency in the definition for the buoyancy Reynolds number Re_b as defined in (2.2), we obtain

$$Re_b = \left(\frac{L_O}{L_K}\right)^{4/3}. \tag{3.6}$$

Therefore, the already noted observation that $Re_b \gtrsim \mathcal{O}(100)$ is required for stratified turbulence to be sufficiently vigorous to be largely unaffected by viscosity is equivalent to the requirement that there is a sufficiently wide range of turbulent scales unaffected by both viscosity and stratification (Gargett *et al.* 1984; Thorpe 2005; Bartello & Tobias 2013). As discussed in detail in Salehipour *et al.* (2016*a*), there are a variety of different ways in which a buoyancy Reynolds number may be defined, depending on the specific choice of the dissipation rate, and in particular the buoyancy frequency. Therefore, specific numerical comparisons of Re_b between different studies must be treated with caution.

4. Time evolution of length scales in direct numerical simulations

In this section we consider the temporal evolutions of Re_b and the various length scales defined above. We consider these evolutions in our series of DNS simulations, covering a range of Richardson and Reynolds numbers.

4.1. *Time evolution of* Re_b

Figure 2 illustrates the time evolution of Re_b for simulations with different Re_0 at $Ri_0 = 0.12$ (panel a) and for simulations with different Ri_0 at $Re_0 = 6000$ (panel b). The non-stationary nature of intermittent mixing by shear instability is clearly shown in the figure through the non-monotonic temporal evolution of Re_b.

Figure 2(a) shows a qualitative change in the evolution of Re_b for sufficiently large $Re_0 \gtrsim 4000$. At this intermediate Ri_0, energetic time-dependent turbulence (i.e. with $Re_b > 200$) is maintained over a considerable fraction of the intermediate phase of the turbulence life cycle only for $Re_0 = 4000$ and larger. This is a critical difference from the simulations reported in SM00. It is apparent that any extrapolation on the basis of the results of lower Re_0 experiments or simulations (such as those reported in SM00) to geophysical flows which occur at much larger Re must be treated with caution. Quantitatively, while Re_b (defined in the fashion we use here) never exceeds 150 for $Re_0 = 750$, (typical of the simulations reported in SM00) Re_b remains above 200 for $\sim 75\%$ of the turbulence life cycle for $Re_0 = 6000$, when $Ri_0 = 0.12$. The structure of the time evolution of Re_b also exhibits qualitative differences between the simulations with lower Re_0 and higher $Re_0 \gtrsim 4000$. This observation is consistent with our hypothesis that a rich 'zoo' of secondary instabilities (only present at sufficiently high Re_0) qualitatively modifies the subsequent turbulent evolution once those instabilities have broken down.

We now turn our attention to the dependence on Ri_0 of the behaviour of the flow at such sufficiently high Re_0 to sustain vigorous turbulence. We consider a range of Ri_0 for that turbulence to be non-trivially affected by stratification. As shown in figure 2(b), it is clear that this 'energetic' turbulence (i.e. with $Re_b > 200$) remains long-lived (i.e. spans a significant portion of the turbulence life cycle) for all Ri_0 except $Ri_0 = 0.2$. At this stage it is not clear why this qualitatively different behaviour occurs. One possibility is that the behaviour is associated with the Reynolds number being too small for this particular choice of Ri_0, associated as it is with a primary instability with a growth rate so small that it may be adversely affected by the diffusion of the mean profiles, even at these Reynolds numbers. Alternatively, the behaviour may be due to the fact that the Richardson number is so close to the critical value of 0.25 that the saturation amplitude of the nonlinear billow may so small that it leads to a qualitative change in the flow dynamics. Observational evidence (see for example the recent discussion of turbulence in the eastern equatorial Pacific by Smyth & Moum (2013) and in the Romanche Fracture zone by Van Haren *et al.* (2014)) suggests that at the very large Re_0 characteristic of geophysical situations, instability and the ensuing turbulence onset soon after the Richardson number drops below 0.25, although it is extremely difficult to trace the dynamics precisely at the critical value, and so further investigation of shear instability for high Re_0, and Ri_0 'close' in some sense to the critical value of 1/4 is warranted.

Indeed, when considering geophysical relevance, it may be necessary to treat with caution the dynamics of flows with initially small values of Ri_0, as it is not at all clear how such shear instability would be realizable in reality, as discussed above. And as mentioned earlier, the treatment of such low Ri_0 cases numerically requires particular care in terms of the choice of the spanwise extent of the domain to accommodate the braid instabilities which dominate turbulence transition in the limit of vanishing stratification. The importance of this issue is clearly connected to the rate at which the shear is diminished in a region of fixed background density stratification. If this time scale is sufficiently short, it is certainly at least plausible that a low Richardson number regime would be relevant.

4.2. *Influence of Richardson number*

The fundamental requirement that Re_0 be sufficiently large and (perhaps also that Ri_0 is a range where the flow is non-trivially affected by stratification) to capture geophysically realistic turbulent dynamics can also be observed in the way that the time evolution of the various length scales defined above vary in time for our different simulations. We are particularly interested in identifying what we believe should be 'generic' behaviour for high $Re_0 - Ri_0$ flow, and what is affected by either Re_0 or Ri_0 being too 'small' in some sense. We showed in MCP13 that, in agreement with earlier theoretical predictions, for $Ri_0 = 0.16 \sim 1/6$ mixing is most 'efficient' at sufficiently high Reynolds number. Here, efficiency is the fraction of energy available to turbulence that irreversibly increases the potential energy of the system. (We define efficiency precisely, and discuss this issue further in section 6.) This efficient mixing (at $Ri_0 = 0.16$, $Re_0 = 6000$, $Pr = 1$ in MCP13) is due to an optimal excitation of secondary instabilities. $Ri = 0.16$ is sufficiently high to induce a large number of baroclinically-induced secondary instabilities yet it is not too high to suppress the turbulence. Therefore, here we choose to consider that flow simulation as the 'canonical' case.

In figure 3b, we plot the various length scales defined above for this simulation (case 12 in table 1). For completeness, we have also included the cases with $Ri_0 = 0.14$, $Ri_0 = 0.18$ and $Ri_0 = 0.20$. Similarly to figure 2(b), the evolution of the flow with $Ri_0 = 0.2$ is qualitatively different from the other three simulations.

Focusing on figure 3(b) for the simulation with $Ri_0 = 0.16$, certain generic characteristics are as expected. Firstly the Kolmogorov length scale L_K (plotted with a dotted line) decreases rapidly at turbulence onset, and then recovers relatively slowly towards its laminar value as the turbulence decays after the turbulent kinetic energy saturates (i.e. peaks for the first time). Similarly, both the Ozmidov scale L_O (plotted with a solid line) and the Corrsin scale L_C (plotted with a dashed line) rapidly increase at transition, and then decay slowly towards their initial laminar values. Remembering that for clarity we are plotting $10L_K$ and $2L_C$, it is clear that there is a wide scale separation between L_O and L_K as expected throughout the period (up to approximately $t \simeq 125$) when $Re_b > 200$, demonstrating that there appears to be the possibility for a range of the turbulent length scales which are unaffected by both viscosity and stratification.

Perhaps more surprising is the evolution of the Thorpe scale L_T (plotted with a dashed-dotted line). L_T grows during the initial roll-up of the primary billow, and it grows substantially before turbulent motions onset, signaled by the marked drop of L_K. After reaching a peak before the transition to turbulence, L_T actually decreases rapidly during the period of most intense turbulent motion, indicative of vigorous irreversible, and inherently small-scale mixing, associated with the rich 'zoo' of secondary instabilities discussed in detail in Mashayek & Peltier (2012*a*,*b*). We observe that $L_T > L_O$ during the transition to the turbulent phase of flow evolution while $L_T < L_O$ beyond the point of most intense turbulence (i.e. the time t^S_{3D} with largest L_O and smallest L_K). Consistently with the recent detailed analysis of Mater & Venayagamoorthy (2014), this demonstrates that it is by no means appropriate to assume that L_O is 'the limiting size' of overturns in strongly stratified turbulence during the turbulence growth phase. That $L_T > L_O$ in this phase actually suggests that the shear-driven turbulent mixing events considered here may be a candidate for creating the canonical layered structures within the previously

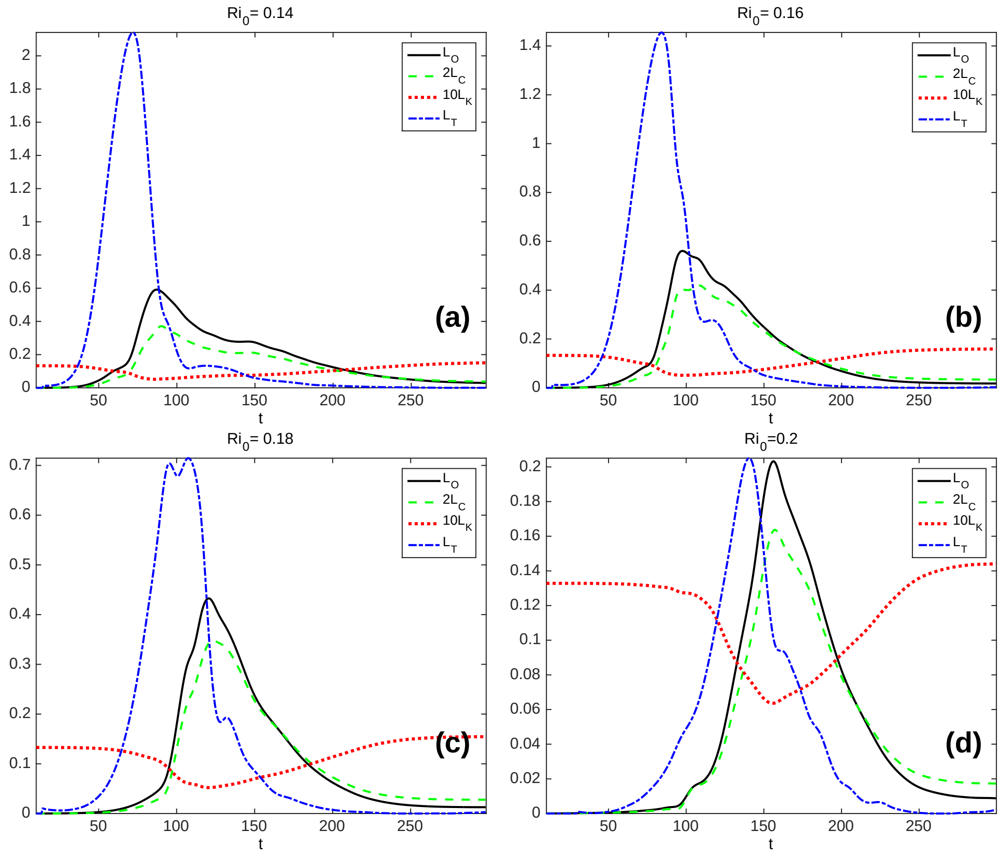

FIGURE 3. Time variation of the various turbulent length scales (normalized by L_{SC} as defined in (3.1)) for a) $Ri_0 = 0.14$ (case 11 of table 1), b) $Ri_0 = 0.16$ (case 12), c) $Ri_0 = 0.18$ (case 13), and d) $Ri_0 = 0.2$ (case 14), all cases for $Re_0 = 6000$.

proposed 'strongly' stratified turbulence scaling regime (see for example Brethouwer *et al.* (2007)).

Furthermore, the relative time dependence of the Thorpe scale and the Ozmidov scale is also of interest. Typically at these Reynolds numbers and Richardson numbers, L_T 'flares', in that it increases rapidly and in turn decreases rapidly before undergoing a slower decay once it has reached very small values. L_O also increases rapidly, but effectively only when L_T has reached its maximum. Interestingly, it appears that L_O reaches its maximum (when the turbulence is most intense, in that $\mathcal{E}$ is largest) very close to the time when $L_O \approx L_T$. Subsequently, L_O 'burns', in that it decreases at a noticeably slower rate than L_T, suggesting a much more extended period of strong turbulence as opposed to strong overturning. We will further discuss the importance of evolution of L_T relative to L_O in section 5.

Figure 3(d) shows that the behaviour is qualitatively different when Ri_0 is increased to 0.2 (noting the dramatic reduction in the extent of the vertical axis with increase in Ri_0). The turbulence is undoubtedly much less intense, with the Ozmidov scale peaking at a markedly reduced maximum value as Ri_0 increases. The relative time dependence of L_O and L_T is also qualitatively different. For $Ri_0 = 0.2$, the Thorpe scale similarly peaks later and at lower values, and decays more slowly. These properties are indicative of a reduction in amplitude and delay and slowing of the primary overturns upon which

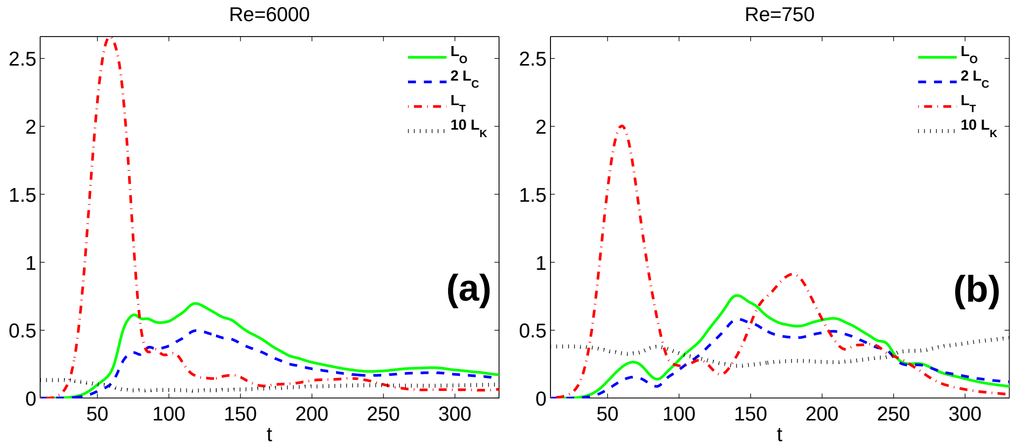

FIGURE 4. Time variation of the various turbulent length scales (normalized by L_{SC} as defined in (3.1)) for: a) case 8 of table 1, with $Re_0 = 6000$ and $Ri_0 = 0.12$; b) case 4 of table 1, with $Re_0 = 750$ and $Ri_0 = 0.12$.

turbulence grows and decays. They also imply a qualitatively different mixing dynamics from the other three cases shown. Indeed, unlike the $Ri_0 = 0.16$ case, at $Ri_0 = 0.2$ the time scale over which L_O increases and that over which it subsequently decays are similar. Furthermore, the dissipation rate does not grow as much above its laminar value in this simulation compared to the $Ri_0 = 0.16$ simulation, and so there is not such a wide length scale separation between the Kolmogorov scale L_K and the Ozmidov scale L_O, indicating that both stratification and viscosity are likely to be modifying the turbulence dynamics substantially. This is all constitutes evidence that the transition to turbulence is relatively weak in this flow, and so may well not be typical of the behaviour of intense geophysical turbulence at very high Reynolds number.

4.3. *Influence of Reynolds number*

We now investigate how the generic behaviour for the time dependence of the various length scales shown by the simulation with $Ri_0 = 0.16$ and $Re_0 = 6000$ in figure 3(b) is affected by variations in Re_0 and Ri_0. Considering the effect of variations in Re_0 first, in figure 4 we plot the time evolution of the various length scales for simulations with $Re_0 = 6000$ and 750 both with $Ri_0 = 0.12$. The time dependence of the various length scales for the higher Re_0 is generally similar to the $Ri_0 = 0.16$ case shown in figure 3(b). There is once again a 'flare' in L_T which appears to trigger a rapid increase in L_O (and L_C) followed by a slower decay towards laminar values. Indeed for this value of Ri_0, there is essentially a period of relatively constant L_O, indicative of sustained turbulence, and there is only a local (as opposed to global) maximum in L_O as L_T drops steeply indicating the break down of the primary billow related overturning.

Clearly, the lower Reynolds number simulation with $Re_0 = 750$ (of the same order as in the flows described in SM00) shown in figure 4(b) is qualitatively different. There is a substantially smaller scale separation between L_O and L_K. Perhaps even more significantly, the temporal evolution of the Thorpe scale L_T, both taken in isolation and relative to the time evolution of L_O is also qualitatively different. The initial rapid decrease in L_T is not associated with a peak in L_O, with the most active turbulence occurring substantially later, principally because of the absence, at this Reynolds number of the 'zoo' of secondary instabilities which affects the simulations shown in figure 3. This is yet more data demonstrating that the evolution of length scales in a stratified shear flow changes

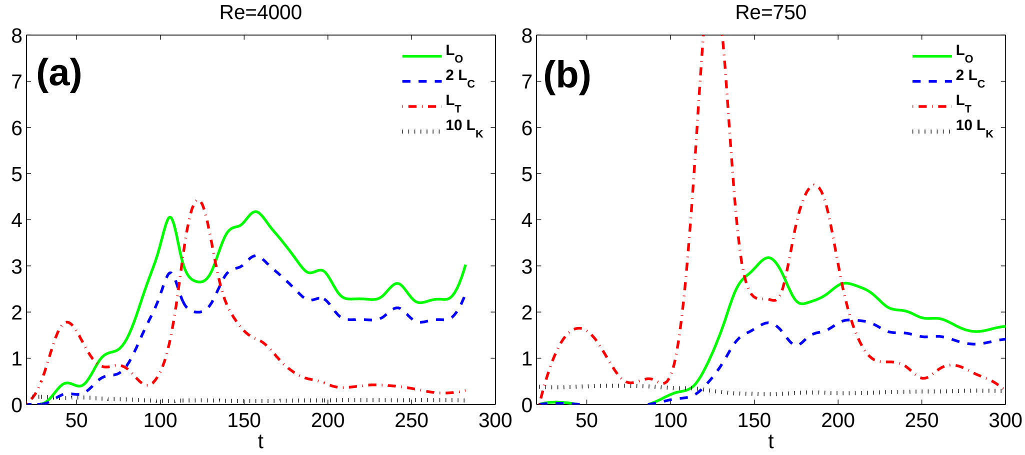

FIGURE 5. Time variation of the various turbulent length scales (normalized by L_{SC} as defined in (3.1)) for: a) case 2 of table 1, with $Re_0 = 4000$ and $Ri_0 = 0.04$; b) case 1 of table 1, with $Re_0 = 750$ and $Ri_0 = 0.04$.

markedly as Re_0 becomes sufficiently large. Therefore, we believe it is clearly necessary to consider flows with $Re_0 \gtrsim 4000$ to investigate assumption IV discussed in the introduction, i.e. that there is a clear separation of scales between the background flow and the superposed (assumed) isotropic turbulence.

4.4. *Summary of evolution of various scales*

In summary, we wish to stress three key aspects of the results presented in this section. First, figures 2–5 show that the assumption of stationary stratified isotropic turbulence is very rarely satisfied, at best only in the energetic turbulence phase of flow for Reynolds numbers sufficiently large and close to a Richardson number 'sweet spot' at which mixing is optimal. According to MCP13, this sweet spot value of Ri_0 is defined by two competing effects: Ri_0 is sufficiently small so that turbulence is not completely suppressed by stratification and yet is sufficiently large for the flow to be replete with buoyancy-driven secondary and higher order instabilities, which are only possible at sufficiently high Re_0. From a length scale perspective this regime is characterized by the existence of a sufficiently wide separation between L_O and L_K. Importantly, these scales are turbulent length scales, by construction distinct from the length scales of the background mean flow. Second, over the entire parameter space we cover herein, the turbulence growth and decay periods of flow evolution, in which assumptions of isotropy and stationarity are clearly violated (as discussed in Smyth & Moum (2000*a*) and Mashayek & Peltier (2013)), together constitute a large fraction of the typical turbulence life cycle. And finally, at sufficiently high Re_0 and Ri_0 in the correct range, there appears to be a typical or generic coupled time-dependence of L_T and L_O. L_T increases rapidly initially before undergoing a slow decay at very small values. L_O, on the other hand, begins to grow rapidly when L_T starts to decrease. L_O reaches its maximum when it is $\sim L_T$, and then decays noticeably more slowly than L_T in the decay period of turbulence. In the next section we turn our attention to the ratio between these two length scales, in particular when in this apparently generic regime for Ri_0 sufficiently large, but not too large, in flows at high Re_0.

5. L_O/L_T as a proxy for turbulence age & efficiency

As originally argued by Thorpe (1977), (see e.g. Scotti (2015) for a detailed discussion) direct measurements of L_T can be used to infer dissipation **if** L_T can be shown to be a simple function of L_O. In such a case, the dissipation rate can be calculated from the expression

$$\mathcal{E} = R_{OT}^2 L_T^2 N^3, \tag{5.1}$$

where

$$R_{OT} = \frac{L_O}{L_T}. \tag{5.2}$$

Indeed, further progress can be achieved by making the further (though not always justified, see for example MCP13) assumption due to Osborn (1980) that the buoyancy flux $\mathcal{B}$, defined as

$$\mathcal{B} = \frac{1}{V}\int \frac{g}{\rho_r}\rho w dv, \tag{5.3}$$

can be linearly related to the dissipation rate $\mathcal{E}$ through a 'universal' turbulent flux coefficient Γ (sometimes referred to as 'mixing efficiency'). Using this assumption, a measurement of the Thorpe scale L_T along with an appropriate buoyancy frequency N are commonly used in the oceanographic research literature (see e.g. Dillon (1982); Kunze *et al.* (2006); Thorpe (2005)) to estimate diapycnal eddy diffusivity through

$$\kappa_T \equiv \frac{\mathcal{B}}{N^2} = \frac{\mathcal{B}}{\mathcal{E}}\frac{\mathcal{E}}{N^2} = \Gamma R_{OT}^2 L_T^2 N. \tag{5.4}$$

As discussed in detail by Mater *et al.* (2015) and Scotti (2015), estimates of the ratio R_{OT} are very sensitive to the existence of large-scale overturnings within the flow, and since the ratio is squared in (5.4), uncertainty in its value has a marked effect on estimates of diapycnal diffusivity.

Furthermore, the time-dependent properties of the ratio R_{OT} are also very important, as its particular value is often used to infer the 'age' of the turbulence involved in observed mixing events (SM00, Smyth *et al.* (2001); Ivey & Imberger (1991); Bouffard & Boegman (2013)). Based on direct numerical simulations of Kelvin-Helmholtz billows at relatively low Re_0, SM00 reported that R_{OT} was typically observed to increase with time (see for example their figure 15) and argued in favour of the observational and entropy-based arguments of Wijesekera & Dillon (1997), that 'older' overturnings should be characterized by large values of $R_{OT} > 1$. We also observe the same qualitative trend as is shown in figure 6 which shows the time evolution of R_{OT} for the same two groups of cases shown in figure 2. This is consistent with our 'generic' observation that, after its initial flare to very large values, L_T decreases rapidly, to very small values, and in particular to values smaller than the more slowly decaying 'burning' L_O. For the single-wavelength simulations in the right panel, R_{OT} is indeed an increasing function of time. Conversely, for simulations shown in the left panel which include two wavelengths of the primary Kelvin-Helmholtz instability and span an order of magnitude increase in Re, R_{OT} grows rapidly at transition, reaching a maximum around the time t_{3D}^S when the inherently three-dimensional turbulence saturates, and then decays rapidly before showing a second oscillatory growth phase driven by variations in the rate of decay of L_T and L_O, due to the complicating merging dynamics. As already discussed, such merging dynamics are suppressed for flows with higher Re_0, and so we do not believe that dynamics associated with merger of primary KHI billows are characteristic of geophysically relevant flows. This belief is reinforced by the fact that perturbations in real flows are highly unlikely to be 'tuned' to trigger merger events, and are typically much more broad-band and noisy

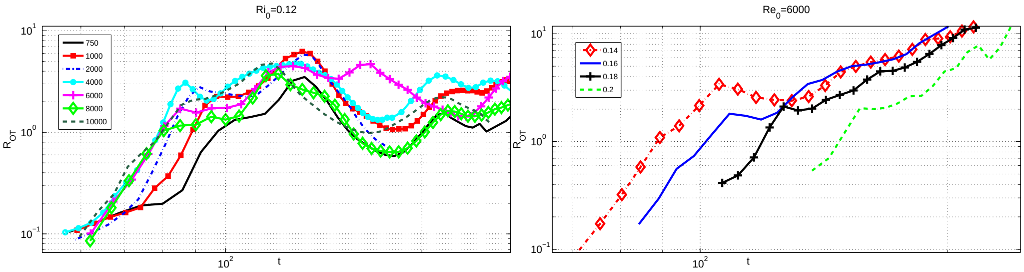

FIGURE 6. .Time evolution of the ratio $R_{OT} = L_O/L_T$ over the turbulent life cycle of flow for the same cases as those shown in Fig. 2: the left panel shows results from simulations with $Ri_0 = 0.12$ and $750 < Re_0 < 10^4$ (with pairing), while the right panel shows results from simulations with $Re = 6000$ and $0.12 < Ri_0 < 0.2$ (with no pairing).

in structure, characterised by low amplitude or residual turbulent motions. Indeed, we are unaware of observations of merging billows in geophysical flows in the atmosphere and ocean, though there is much evidence of observations of long trains of individual billows.

We note that while SM00 simulations were also conducted with streamwise extent which allowed for the development of two wavelengths of the primary instability, similarly to those shown in figure 6(left), their R_{OT} evolution differs qualitatively and quantitatively from our results. They found a more monotonic increase in R_{OT} with time during the decay phase of turbulence. We believe that the difference between their results and ours is due to differences in the simulations' initializations. SM00 initialized their simulations with non-trivial amplitude in the eigenfunction of pairing instability, leading to a relatively rapid pairing of KH billows early in the simulation, which amounts to an initial pre-turbulent significant increase in L_T, and subsequently a marked decrease in L_T in the turbulent phase of flow once vortices have paired. This apparently leads to a monotonic increase in R_{OT} in the turbulent phase of the flow. On the other hand, our two-wavelength simulations are not forced explicitly with the pairing mode eigenfunction and also are conducted at very high Reynolds number. As discussed above, flows with such higher Re_0 are associated both with a significant suppression of the pairing instability, and with fundamentally different character in the transition mechanisms (i.e. the full 'zoo' of secondary instabilities) and the intensity (quantified by the elevated values of Re_b) of the ensuing turbulence. Perhaps unsurprisingly, such differences lead to a characteristically different R_{OT} behaviour during the later stages of flows in which pairing (even if highly suppressed) manifests.

In summary, our results in this section suggest that the evolution of R_{OT} in turbulence life cycles initiated by shear instability is very sensitive to details of the flow evolution such as the existence or lack thereof of an upscale cascade through pairing instability flow initialization. Therefore, it is at least plausible that the time-dependence of R_{OT} is likely to vary according to the degree of ambient or residual turbulence within a flow in which KH billows develop, as is to be expected for a realistic geophysical flow. In spite of the relevance of R_{OT} as a proxy for turbulence age, details of its evolution play an important role in characterizing the properties of the turbulence itself. Essentially, L_T represents the vertical overturning scale of turbulence and so represents the large scale stirring at which energy is being injected into the perturbation fields, while L_O represents the largest eddies which are not strongly influenced by stratification, remembering that eddies smaller than L_O and larger than L_C are still affected by the ambient shear. Therefore, an optimal injection scale for the cascade of energy from larger scale stirring to dissipation is expected. This corresponds to the stirring injection scale (L_T) occurring at the largest scale not suppressed by stratification (L_O), i.e. precisely when $L_O \sim L_T$. Precisely this behaviour was observed by Ivey & Imberger (1991), as this relationship corresponds to the optimal value for mixing of their turbulent Froude number $Fr_T = (L_O/L_T)^{2/3} \simeq 1$. As we see next, in this phase of flow evolution mixing is very efficient. Full discussion of the turbulence cascade and anisotropy in turbulence induced by KH instability over a wide range Reynolds and Richardson numbers is provided in Mashayek & Peltier (2013). Here we have built upon that study to connect it to the turbulent length scales discussed in this section and through that connection to mixing.

6. Implications for the quantification of mixing

The 'mixing efficiency' η is an important quantity which is commonly used for quantifying diapycnal mixing rates from observations of shear-induced turbulence in the ocean and atmosphere. We define η within a Boussinesq framework as the ratio of the kinetic energy converted to potential energy irreversibly via a net irreversible vertical buoyancy flux, to the total irreversible conversion of kinetic energy to both potential energy and internal energy via viscous dissipation. This quantity is sometimes also referred to as the flux Richardson number, although the two quantities are not exactly the same at finite Reynolds numbers, as the denominator of the flux Richardson number is usually defined to be the production of turbulent kinetic energy (see Peltier & Caulfield (2003), MCP13 and Rahmani *et al.* (2014) for more discussion). The mixing efficiency is widely assumed to be $\eta \ \sim 0.15 - 0.2$, equivalent to the canonical model due to Osborn (1980) that the turbulent flux coefficient Γ (as defined in (5.4) is given by $\Gamma \simeq \eta/(1-\eta) \leqslant 0.2$ despite the growing evidence demonstrating that it is highly variable in shear-induced mixing (see the recent results of MCP13 and Rahmani *et al.* (2014)).

As discussed in more detail in Caulfield & Peltier (2000) and Peltier & Caulfield (2003), mixing efficiency can be considered to be a time-dependent quantity, and so it is natural to consider both instantaneous values $\eta_i(t)$, and some appropriate cumulative mixing efficiency η_c for a given mixing event. To calculate η_i from our simulation results, we calculate the net instantaneous irreversible increase in the potential energy of the system, which represents diapycnal mixing $\mathcal{M}$, and then define

$$\eta_i = \frac{\mathcal{M}}{(\mathcal{M} + \mathcal{E})}, \tag{6.1}$$

where $\mathcal{E}$ is the total dissipation rate as defined in (2.3), and $\mathcal{M}$ is determined using the sorting algorithm as initially described by Winters *et al.* (1995) and slightly modified in Caulfield & Peltier (2000). More specifically, $\mathcal{M}$ is defined as the net change in the background potential energy of the system which may be calculated by an adiabatic sorting of the fluid parcels in the whole domain as was described earlier in calculation of the Thorpe scale. Since the background potential energy may only be increased, any change in it will correspond to diapycnal mixing in our setup with periodic boundary conditions. We can also define a cumulative mixing efficiency η_c as is now conventional as

$$\eta_c = \frac{\int_{t_s}^{t_e} \mathcal{M}\ \mathrm{dt}}{\int_{t_s}^{t_e} \mathcal{M}\ \mathrm{dt} + \int_{t_s}^{t_e} \mathcal{E}\ \mathrm{dt}}, \tag{6.2}$$

for appropriately chosen start time t_s and end time t_e. We set $t_s = t_{3D}^S$, and t_e to be the end of the simulation (when the flows have typically relaminarised) to define η_c^{3D}, which we list in table 1 for each of the simulations. (See Mashayek *et al.* (2013) for more discussion.) In what follows we divide the turbulent life cycle of each simulation into a number of intervals and average η over each period to obtain a locally-averaged efficiency η_a for each interval. Each period is set to be of 10 eddy turnover time scales (defined as $h/\Delta u$ where Δu and h are characteristic scales of the shear flow as defined in (2.1)), keeping in mind that the turbulent life cycle of the simulations in table 1 (nominally defined as the period over which $20 < Re_b$) typically extends over 200 to 400 turnover timescales.

In figure 7(a) we show the results of calculations for simulations 6-14 of table 1. This subset includes cases with Re_0 sufficiently high to represent sustained turbulence for a considerable fraction of flow evolution, and with Ri_0 sufficiently large for the behavior to share the key 'generic' characteristics of the simulation with $Ri_0 = 0.16$ as discussed above. To connect the interpretation of evolution of efficiency of mixing in the simulations with the time dependence of the various length scales as described in the previous section, figure 7 shows a scatter plot of L_T vs L_O, with the symbol colours representing η_a. The lines in the figure represent $L_T = L_O$, $L_T = 4 \times L_T$ and $L_T = 0.25 \times L_O$, the latter two providing bounds on the L_O/L_T ratio in observations (see Thorpe 2005, for discussion and references). As discussed earlier, symbols for which $L_T > L_O$ correspond to the period in flow evolution in which eddies (of scales $L \leqslant L_O$) associated with secondary instabilities grow rapidly and efficiently within the primary overturn, while symbols with $L_T < L_O$ correspond to the final period of the flow evolution when the turbulence is decaying and stirring is suppressed by ambient stratification. It is apparent that mixing is most efficient during the earlier period, particularly when $L_T \sim L_O$, (precisely as assumed by Ivey & Imberger (1991)) since the inertial subrange is very efficiently energized at the upper bound (stirring scale) by the available potential energy reservoir stored in the primary overturn. As stirring by large eddies becomes suppressed by stratification in the later period of turbulence, mixing is less efficient. Thus, the high efficiency of mixing at $L_O \sim L_T$ appears to be a direct consequence of the nature of turbulence induced by shear instability at high Reynolds number. Importantly, this violates assumption IV as described in the introduction, because the length scale of the overturning is most definitely **not** widely separated from the important length scales of the turbulent motions.

Furthermore, since this most **efficient** mixing occurs when $L_O \leqslant L_T$, which is also in the build up to the instant when both L_O and Re_b are maximum, the actual total amount of mixing in the build up to $L_O \sim L_T$ is also maximized. In other words, since $\Gamma \simeq \eta_a/((1-\eta_a))$ (for caveats see MP13 and Salehipour & Peltier (2015)), the observation that η_a is maximum when Re_b is maximum strongly suggests that the turbulent diffusivity κ is also maximum at that time, since using (2.2) and (5.4) we have,

$$\kappa_T = \Gamma \frac{\mathcal{E}}{N^2} \simeq \nu \frac{\eta_a}{(1-\eta_a)} Re_b. \tag{6.3}$$

This suggests that the flow at this time is so organised as to maximise the amount of vertical mass flux, because of the combined effects of the turbulence being most intense (i.e. with largest Re_b) and most efficient (i.e. with largest η_a and hence largest Γ).

The above description of the dependence of mixing on the temporal evolution of L_O and L_T was based on simulations of Kelvin-Helmholtz instabilities that form the basis of our work. So, it is legitimate to question their generality insofar as the much more dynamically diverse ocean mixing process is concerned. However, we conjecture that the observation that the existence of distinct overturns provides sufficient available potential energy that can feed efficient turbulent mixing is not a special phenomenon only occuring in KHI flows, but is a more generic property of high Reynolds-number stratified mixing processes, triggered by a wider range of mechanisms, including other shear instabilities, hydraulically controlled flows, or breaking internal waves. Clearly further work is warranted to test this conjecture by investigating the mixing associated with these wider range of mechanisms.

To explore this further, panels (b) through (d) of figure 7 show similar scatter plots to that from our DNS in panel (a). The data for panels (b,c) come from observations made in the thermocline of the ocean while the data for panel (d) come from one of the great lake. Mixing in these natural environments is induced by a mixture of dynamical processes including vertically propagating internal waves and shear instabilities of different types. Panels (b-d) share the same pattern with panel (a) in that mixing efficiency is larger for $L_T > L_O$, further highlighting the role of natural overturns in determining the efficiency of mixing.

We acknowledge that our simulations are highly idealized and that the observational data used in figure 7 are based on a number of crude assumptions made for practical reasons; importantly, the calculation of mixing efficiency from data is difficult and involves large inaccuracies. Furthermore, there seem to be some systematic and as yet unexplained differences between how data are skewed about the $L_O = L_T$ line in the four panels. For example, the lake data in figure 7(d) appear to be more qualitatively similar to the numerical data in figure 7(a) than to the two oceanographic data sets in figures 7(b) and (c). Nevertheless, our main point here is neither dependent on the actual value of mixing efficiency nor is it sensitive to the above-mentioned inaccuracies and idealizations. Essentially, as long as distinct overturns exist throughout turbulence evolution, they play a non-negligible role in determining the efficiency of mixing. This point is one of the main messages of this paper.

We stress that this point is important for two reasons. First, as discussed earlier, conventional parameterization schemes are based on assumptions which are typically better satisfied during the turbulence decay period (i.e. towards the left in each panel). Second, the majority of studies of DNS of shear instabilities have focused on the decay period by filtering the earlier period based on the (at times implied) justification that the early period does not conform to a plausible 'ocean turbulence regime', assumed by (for example) Osborn (1980) to be well-modelled as stationary isotropic turbulence where the steady turbulence production is balanced by an isotropic dissipation rate and a relatively small (positive) buoyancy flux. In combination, these assumptions appear to have led to a circular argument for filtering the part of simulations that does not fit the parameterizations even though the simulations are carried out for the very purpose of improving the parameterizations. It was shown in Mashayek & Peltier (2013) that in direct numerical simulations of shear instabilities, the early period of turbulence makes a non-negligible contribution to the net buoyancy flux over a turbulence life cycle. Furthermore, the analysis of Smyth *et al.* (2001) showed that the $L_O < L_T$ patches in data used in figure 7 make a large contribution to net mixing as well. So, as long as large overturns exist, the contribution of the earlier period of turbulence in which distinct overturns and superimposed turbulence co-exist needs to be taken into account in both parameterization schemes and in analysis of numerical simulations. Of course, it is important to remember that in the observational data there is no 'time-stamp', in that unlike the simulation data there is no way to follow the time evolution of an individual mixing event. However, the observational data are at least consistent with the idea that $L_O < L_T$ patches are associated with vigorous overturnings that will subsequently lead to increased turbulent mixing, and hence L_O remaining larger for a longer time than L_T, i.e. that L_T 'flares' while L_O 'burns', analogously to our simulations.

The contribution of overturns is partially filtered in conventional parameterizations by assumptions of isotropic stationary small scale turbulence existing at a scale distinctly separated from that of the background flow. It has also often been left out of analysis of DNS data for several reasons. Distinct overturns observed in early DNS are often thought to be artifacts of the low Re idealized nature of such simulations, (Peltier & Caulfield 2003), and furthermore, the argument has been advanced that the later-time turbulence is more likely to be representative of stratified turbulence events, not necessarily generated by flows initially strongly unstable to Kelvin-Helmholtz billows (Salehipour *et al.* 2015; Salehipour & Peltier 2015; Salehipour *et al.* 2016*a*).

However, recent direct numerical simulations at high Reynolds number and numerous recent observations of deep ocean turbulence have clearly shown that distinct overturns not only can exist, but in fact are typical in strong mixing zones. It almost appears as if the flow is trying to maximize efficiency of mixing by providing an efficient energy pathway into turbulence by stirring and storage of potential energy through overturns. Recent field experiments focused on abyssal ocean mixing (where mixing plays a key role in closure of abyssal branch of ocean meridional overturning circulation) have all found turbulence to be induced by continuous excitations of large overturns scaling from a few meters up to 500 meters (Ferron *et al.* 1998; Frants *et al.* 2013; Mater *et al.* 2015; Voet *et al.* 2015). Thus, we conjecture that underestimation of mixing due to partial neglect of the role of overturns may well obscure significantly the apparent tendency of turbulence to maximize its mixing efficiency through such overturns.

We think it useful to reiterate our reasoning for not adding data from low Ri_0 cases to figure 7(a). Since the growth rate of the primary Kelvin-Helmholtz instability is a monotonically decreasing function of Ri_0, it is tempting to decrease Ri_0 to reduce computational cost since the simulation will in principle need to be conducted for a shorter time interval for a given computational domain. However, this reduction in computational cost is likely to be swamped by the need to consider larger computational domains, to capture at least some of the merging dynamics, which inevitably introduces large scale stirring. Furthermore, as discussed earlier, the spanwise extent of the domain may possibly need to be expanded to host braid instabilities dominating turbulence transition in the weak stratification limit. Suppressing the stirring associated with such large scale streamwise and spanwise secondary instabilities inevitably reduces the amount of mixing which apparently occurs in a simulation in a smaller domain. Indeed, it is entirely possible that as Re is increased, the relative intensity of secondary instabilities at such smaller Ri_0 may change in as yet not fully understood ways. Since the extent to which such considerations can influence our low Ri direct numerical simulations has not been fully explored due to computational limitations, we refrain from presenting quantitative arguments about mixing properties of such simulations. A detailed discussion of the potentially misleading nature (at least insofar as geophysically relevant mixing is concerned) of low Ri numerical simulations designed to produce high Re_b during the flow evolution is presented in Bartello & Tobias (2013).

In summary, while a number of studies have attempted to parameterize mixing efficiency as a function of Re_b or in terms of L_O/L_T (see Bouffard & Boegman 2013, for a review), we find neither approach to be sufficient. Essentially, Re_b includes information concerning L_O and L_K, while the ratio L_O/L_T clearly lacks explicit information about L_K. As demonstrated here, knowledge of all three scales is needed for characterizing shear-driven stratified turbulent mixing, and so we believe that the large discrepancies between various attempts at parameterizing mixing based on either Re_b or L_O/L_T are due to a lack of such additional knowledge. Despite such discrepancies, we have demonstrated here that the specific role in the efficiency of mixing of the large overturns themselves is significant, corresponding to a non-negligible portion of the turbulence life cycle in which $L_T > L_O$. The role of overturns also appears to be similar for the data from our

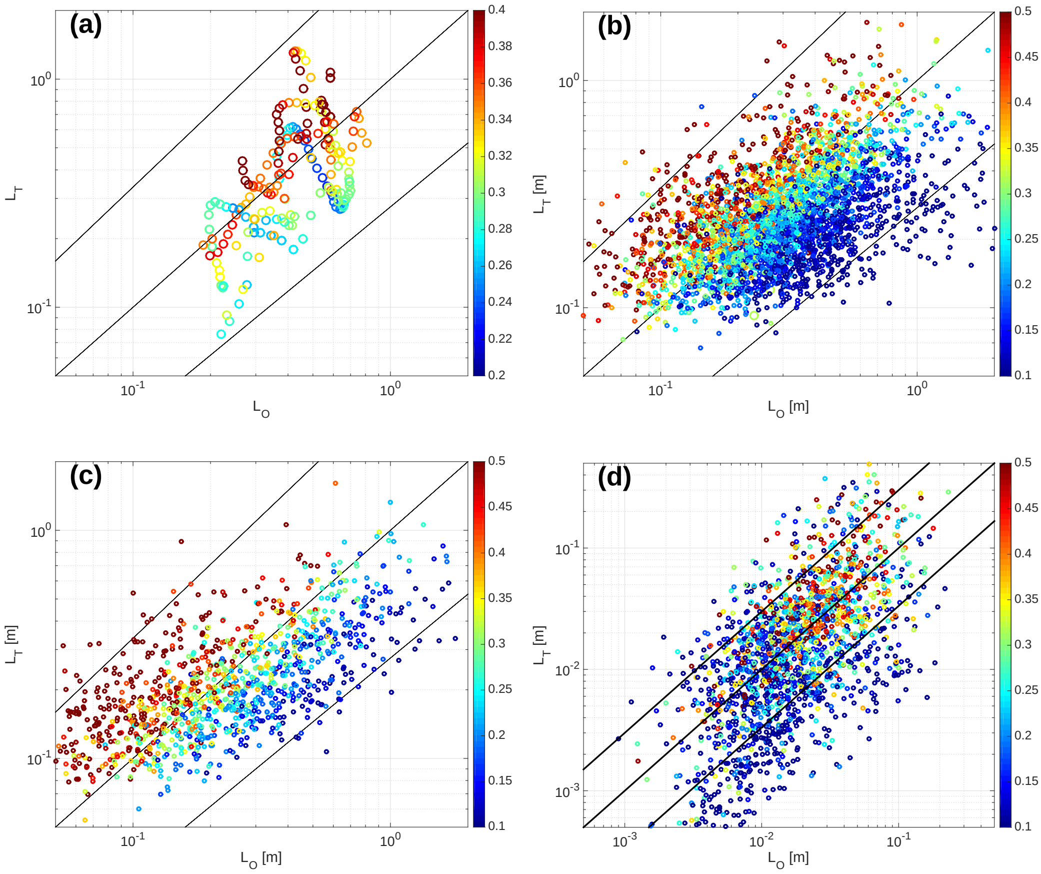

FIGURE 7. Scatter plot of L_T versus L_O from DNS (a) and observations(b,c,d). (a) From DNS cases 6-14 of table 1. (b) From FLX91 oceanic dataset collected ∼1000 km off the coast of northern California. (c) From the TIWE oceanic dataset collected at the equator at 140^{o}W. (d) From lake observations made at thermocline depth in lake Erie in 2008-2009. More information about the sources of these datasets are provided in Appendix I. The lines in each panel represent $L_T = L_O$, $L_T = 4 \times L_O$ and $L_T = 0.25 \times L_O$. Symbol colors and colorbars represent mixing efficiency η_a. Note that the axes in panel (a) are normalized by L_{sc} as was the case throughout this paper while in panels (b-d) they are in units of meters.

simulations and for ocean and lake data. Such efficient mixing is inherently associated with the presence of large-scale overturns. The clearly more efficient mixing associated with such overturns is systematically left out of conventional parameterizations (based around the classical model of Osborn (1980) assuming $\Gamma \leqslant 0.2$) that are used to infer mixing rates from observations.

7. Discussions

We have analyzed a sequence of direct numerical simulations of stratified turbulent mixing events driven by classical shear instability, focusing on a consideration of the relative time-dependence of various natural length scales of turbulence and the implications of aspects of this relative time-dependence for the irreversible vertical mixing of density. Our analyses demonstrate that for ‘small’ stratification, the turbulence and ensuing mixing is dominated by large density overturns and pairing interactions and so any parameterization based on the assumptions of stationary fully developed isotropic turbulence does not hold, in the specific sense that the assumptions I-IV mentioned in the Introduction do not hold. It is important to note, however, that there is evidence that the upscale cascade due to pairing instability may well be suppressed at sufficiently high (perhaps more geophysically relevant) Reynolds numbers, a regime that we have been unable to access with the currently employed suite of direct numerical simulation analyses, although there is always the possibility that other processes may become more important as Re increases markedly.

Conversely, for 'large' stratification with minimum Richardson number sufficiently close to the critical Miles-Howard value of 1/4, mixing is highly time-dependent and a prolonged intermediate period of isotropic stationary turbulence is absent, corresponding to the break down of assumptions II-III mentioned in the Introduction. In this regime, the scale separation between L_O and L_K is relatively narrow and turbulence is greatly influenced by the suppressing influence of stratification. However, it is unclear whether this behaviour is affected by finite Reynolds number effects, as the growth rate of the primary instability is so small that diffusion of the background flow may be affecting adversely the maximum saturated amplitude of the primary instability, in as yet poorly-understood ways.

We argue that the behaviour at slightly smaller intermediate levels of stratification, where pairing events are suppressed, and yet the primary instability is sufficiently vigorous to allow for the onset of a large 'zoo' of secondary instabilities which trigger energetic turbulence leads to a 'generic' shear-driven stratified mixing behaviour. Specifically, this generic behavior exhibits a very efficient turbulence downscale cascade through the inertial subrange when $L_T > L_O$ due to the large pool of potential energy available to sub-L_O eddies due to the large initial overturn, whose vertical scale is characterized by L_T. This translates into high mixing efficiency, which peaks when $L_T \sim L_O$ as at that particular time stirring becomes 'optimal' since it is occurring at the largest energy injection scale possible that is not suppressed by stratification. Although we refer to this behavior as 'generic', it is important to note that the existence of the early $L_T \geqslant L_O$ regime, particularly associated with relatively large-scale overturnings, is not guaranteed in the evolution of all shear-unstable flows, and is likely to be environment-dependent. For example, Kelvin-Helmholtz billows in an energetic estuary have been shown not to evolve distinct vorticity cores which store potential energy with the effective L_T being relatively small(Geyer *et al.* 2010) while other forms of shear instability (such as the Holmboe instability, see Salehipour *et al.* (2016*a*) for further details) are not characterized by large overturns, but rather drive mixing principally through 'scouring' (Woods *et al.* 2010). However, energetic overturning billows similar in structure to those described here have been observed growing on low-frequency internal tides in the abyssal ocean (van Haren & Gostiaux 2010, 2012), in deep ocean fracture zones (Van Haren *et al.* 2014) and in the thermocline (Thorpe 2005). And as we discussed earlier, several deep-ocean field programs have repeatedly shown that abyssal diapycnal mixing is faciliated through large overturns which can range in size from a few to hundreds of meters (Ferron *et al.* 1998; Frants *et al.* 2013; Mater *et al.* 2015; Voet *et al.* 2015).

We show not only that mixing efficiency depends upon L_O/L_T, but that it also depends on the scale separation between L_O and L_K, i.e. the width of the inertial subrange of turbulence, or equivalently the magnitude of the buoyancy Reynolds number Re_b. Fundamentally, the key constituents of efficient and vigorous mixing are that $L_T \gtrsim L_O$ and L_O/L_K is sufficiently large. Therefore, we argue that parameterization of mixing efficiency based on $Re_b = (L_O/L_K)^{4/3}$ alone is insufficient as it misses the important relative properties of L_T and L_O, while parameterization based on L_T/L_O alone is also insufficient as it misses the Re_b contribution. We conjecture that the key physics of both an optimal injection scale and a wide inertial subrange are required. Such flows also violate assumption IV presented in the introduction, as there is not a large scale separation between the external forcing (characterized by L_T) and the turbulence (characterized at the largest scale by L_O).

It is important to note that while some parameterization schemes for inferring mixing from observations assume isotropy of the turbulence (Osborn 1980; Osborn & Cox 1972), a large number of observational studies which measure both L_T and L_O suggest that $L_T \geqslant L_O$, implying non-negligible anisotropy (Dillon 1982; Crawford 1986; Ferron *et al.* 1998; Smyth *et al.* 2001; Mater *et al.* 2015). In fact, $L_O/L_T = 0.8$ is a standard choice made to obtain the dissipation rate $\mathcal{E}$ from L_T calculated based upon finestructure measurements of temperature or salinity and when microstructure estimates are unavailable (see for example Waterhouse *et al.* (2014)). It is particularly important to note that while 0.8 might be a reasonable turbulence lifecyle mean for L_O/L_T, the fact that the ratio is likely much higher during the intermediate period of flow evolution in which buoyancy flux is maximized (as a result of the coexistence of distinct overturns upon which turbulence is superimposed) implies an underestimation of mixing when a constant ratio is used in the finescale parameterization based on the Thorpe scale. Just how large this underestimation is, and how parameterizations may be modified to capture the mixing associated with large-scale overturnings are both topics of ongoing research (see for example Mashayek *et al.* (2017)).

Appendix I: data sets

The first two oceanic datasets employed for construction of panels (b) and (c) in figure 7 were introduced in Smyth *et al.* (2001). Panel (b) corresponds to the FLX91 dataset which was collected during the FLUX STATS cruise in 1991 approximately 1000 km off the coast of northern California (Moum 1996). The dataset used in panel (c) is from the Tropical Instability Wave Experiment (TIWE) and was collected at the equator at 140^oW in 1991 (Lien *et al.* 1995). The dataset used in the construction of panel (d) in figure 7 was introduced in Bouffard & Boegman (2013) and corresponds to observations made at thermocline depth in Lake Erie during the summers of 2008-2009.

Appendix II: L_T^{3D} vs L_T and caveats for oceanographic implications

Our focus in this paper was upon the role of overturns on turbulent mixing in geophysical shear flows, and more specifically a focus on conditions relevant to oceans and lakes. The main message of the paper was based on analysis of energy conversion from the mean kinetic energy (provided by large scale forcing from a variety of sources including estuarine exchanges, low frequency internal wave shear etc.) to available potential energy and from there to a cascade of overturns that take energy down to scales at which diapycnal mixing and viscous dissipation occur. Our main message is that the existence of an intermediate nontrivial overturning scale between the mean background flow and small scale turbulence allows for an efficient energy pathway into diapycnal mixing by providing additional stirring and filamentation, thereby enhancing the efficiency of mixing. To convey this message and its sensitivity to variations in Reynolds and Richardson numbers, we employed a definition of the Thorpe scale, referred to as L_T^{3D}, which is only really practical in three-dimensional numerical modeling. In this appendix we provide a number of caveats highlighting the differences between this measure of overturning and the one-dimensional classical Thorpe scale L_T, which for practical limitations is used to infer mixing rates and is constructed from localized profile measurements in oceanic and lake environments. We emphasize that the main message of our work does not depend on the differences we highlight here. Indeed, the importance of taking into account the existence of such an intermediate overturning scale in parameterization of mixing in the oceanographic context has already been pointed out by Kunze (2014). Our research provides a further fluid mechanical basis for such an argument. Furthermore, we note that while L_T^{3D} cannot be obtained from observations, certain observational techniques such as those employed by Geyer *et al.* (2010) provide a series of parallel profiles measured through turbulent wave trains. Such measurements can provide a means for constructing a L_T^{2D} to fill in the gap between our study and the majority of observational studies based on one-dimensional L_T.

While physically meaningful and suitable for diagnosis from numerical models, the rms three-dimensional Thorpe scale L_T^{3D} obtained in this work by full three-dimensional sorting of the density field has important differences from the one-dimensional L_T. Importantly, while the L_T^{3D} can be nonzero in the presence of a propagating wave without any overturns, or even in the presence of an overturn riding on a background low frequency internal wave, just to take two examples, the one-dimensional L_T is only nonzero in the presence of true overturns. In our study, however, we have only considered flows strongly susceptible to the Kelvin-Helmholtz instability, which overturns upon initiation of (exponential) growth. Thus, this caveat (that L_T^{3D} may return a 'false positive' of overturning) does not concern our specific application and so we are safe in using L_T^{3D} as a surrogate for an overturning scale.

A close comparison of the three-dimensional and one-dimensional Thorpe scales was provided by SM00. They found that the three-dimensional scale exceeds the one-dimensional scale in the decay period of turbulence (induced by shear instability) when the Thorpe scale is small. The generality of this argument in a more complex environment in which vertical displacements are not entirely or even partially driven by overturning instabilities is unclear, especially noting that (as mentioned above) there are scenarios in which the three-dimensional displacement scale might be nonzero while the one-dimensional scale remains zero due to lack of overturning. Nevertheless, this difference is not of central importance in the class of flows which we are considering, since in the case of shear instability both scales are measures of the physical overturning scale, are not too different during the most energetic phase of turbulence over which most of the contribution to the net buoyancy flux is made, and can be employed to provide a measure of the width of the spectral gap between the energy injection scale and the upper bound of the inertial subrange.

However, during the decay period of turbulence, the one-dimensional Thorpe scale is smaller than the three-dimensional Thorpe scale. Therefore, it is to be expected that L_O/L_T grows larger with time than L_O/L_T^{3D}. This has implications for our discussion of figure 6: while L_O/L_T is likely a monotonically increasing function of time and hence might be more naturally treated as a proxy for turbulence age, L_O/L_T^{3D} is not as clear a proxy. From a physical point of view, the difference between L_O/L_T^{3D} and L_O/L_T in the decay period of turbulence in a flow susceptible to Kelvin-Helmholtz instability is testament to the shortcomings of L_T in capturing the totality of the significant flow physics. A close look at figure 6(a) (which represents cases that, unlike those in panel (b), allow for interactions between adjacent billows) reveals that the ratio $R_{OT} = L_O/L_T^{3D}$ remains $\mathcal{O}(1)$ during the decay period of the turbulence. This suggests that as turbulence decays and the energy injecting eddies shrink, so does the Ozmidov scale accordingly. This further suggests that the eddies associated with the dominant energetic injection, which are decaying in amplitude and magnitude since the turbulent kinetic energy and the Thorpe scale are both dropping, may also be thought of as the largest eddies not yet suppressed by turbulence. Conversely, $R_{OT} \equiv L_O/L_T$ (based on the one-dimensional Thorpe scale) suggests that L_T can become much smaller than L_O in this period, which implies that energy injection eddies are much smaller than the maximum size which is not suppressed by stratification, which seems somewhat inconsistent from a physical perspective.

As we discussed above, despite these subtle differences, there are at least two further leading order issues with this proxy. First, it is overly sensitive to the initial conditions of shear instability, in particular whether adjacent billows can interact or merge. Second, it remains to be shown if the evolution of the ratio in observations of more complex nature agrees with that based on shear instability analysis such as ours and that of SM00. While we have provided evidence that scatter plots of L_O versus L_T from observations have certain similarities with our data based on direct numerical simulations, as already noted in section 6, there is no explicit information about time evolution and turbulence age in such observational data. Adding such 'time-stamp' information clearly warrants future study.

Acknowledgements

AM acknowledges support from a David Crighton Fellowship from the Department of Applied Mathematics & Theoretical Physics, University of Cambridge, where this research was initiated and support from an NSERC PDF award. The research of W.R.P. in Toronto is supported by NSERC Discovery Grant A9627. The research activity of C.P.C. is supported by EPSRC Programme Grant EP/K034529/1 entitled 'Mathematical Underpinnings of Stratified Turbulence.' We thank W.D. Smyth and J. N. Moum for sharing the ocean data and D. Bouffard for sharing the lake data used in Fig. 7. We also thank three anonymous reviewers whose insightful and constructive comments have led to a substantial improvement in this paper.

REFERENCES

Alford, M. H & Pinkel, R. 2000 Observations of overturning in the thermocline: The context of ocean mixing. *J. Phys. Oceanogr.* **30** (5), 805–832.

Bartello, P. & Tobias, S. M. 2013 Sensitivity of stratified turbulence to the buoyancy reynolds number. *J. Fluid Mech.* **725**, 1–22.

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.

Caulfield, C. P. & Peltier, W. R. 1994 Three-dimensionalization of the stratified mixing layer. *Phys. Fluids* **6**, 3803–3805.

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.

Crawford, W. R. 1986 A comparison of length scales and decay times of turbulence in stably stratified flows. *J. Phys. Oceanogr.* **16** (11), 1847–1854.

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

Ferron, B., Mercier, H., Speer, K., Gargett, A. & Polzin, K. 1998 Mixing in the Romanche fracture zone. *J. Phys. Oceanogr.* **28** (10), 1929–1945.

Frants, M., Damerell, G. M., Gille, S. T., Heywood, K. J., MacKinnon, J. & Sprintall, J. 2013 An assessment of density-based finescale methods for estimating diapycnal diffusivity in the Southern Ocean. *J. Atmos. Ocean. Tech.* **30** (11), 2647–2661.

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.* **108** (D8), 8452.

Gargett, A. E., Osborn, T. R. & Nasmyth, P. W. 1984 Local isotropy and the decay of turbulence in a stratified fluid. *J. Fluid Mech.* **144** (1), 231–280.

Garrett, C. 2003 Internal tides and ocean mixing. *Science* **301** (5641), 1858–1859.

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 1980 Fossil temperature, salinity, and vorticity turbulence in the ocean. *Elsevier Oceanography Series* **28**, 221–257.

van Haren, H. & Gostiaux, L. 2010 A deep-ocean Kelvin-Helmholtz billow train. *Geophys. Res. Lett.* **37**, L03605.

van Haren, H. & Gostiaux, L. 2012 Detailed internal wave mixing above a deep-ocean slope. *J. Mar. Res.* **70**, 173–197.

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

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.

Klaassen, G.P. & Peltier, W.R. 1985*a* The effect of Prandtl number on the evolution and stability of KelvinHelmholtz billows. *Geophysical and Astrophysical Fluid Dynamics* **32**, 23–60.

Klaassen, G. P. & Peltier, W. R. 1985*b* The onset of turbulence in finite amplitude KelvinHelmholtz billows. *J. Fluid Mech.* **155**, 1–35.

Kunze, E. 2014 The relation between unstable shear layer thicknesses and turbulence lengthscales. *J. Mar. Res.* **72** (2), 95–104.

Kunze, E., Firing, E., Hummon, J. M., Chereskin, T. K. & Thurnherr, A. M. 2006 Global abyssal mixing inferred from lowered ADCP shear and CTD strain profiles. *J. Phys. Oceanogr.* **36** (8), 1553–1576.

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.* **100** (C4), 6881–6898.

Mashayek, A. 2013 Diapycnal mixing in the ocean: From dissipation scale to large scale meridional overturning circulation. PhD thesis, University of Toronto.

Mashayek, A, Caulfield, CP & Peltier, WR 2013 Time-dependent, non-monotonic mixing in stratified turbulent shear flows: implications for oceanographic estimates of buoyancy flux. *Journal of Fluid Mechanics* **736**, 570–593.

Mashayek, A., H., Salehipour, Bouffard, D., Caulfield, C.P., Ferrari, R., Nikurashin, M., Peltier, W. R. & D., Smyth W. 2017 Efficiency of turbulent mixing in the abyssal ocean. *Geophys. Res. Lett.* **in press**.

Mashayek, A. & Peltier, WR 2011 Turbulence transition in stratified atmospheric and oceanic shear flows: Reynolds and Prandtl number controls upon the mechanism. *Geophysical Research Letters* **38** (16), L16612.

Mashayek, A & Peltier, WR 2013 Shear-induced mixing in geophysical flows: does the route to turbulence matter to its efficiency? *Journal of Fluid Mechanics* **725**, 216–261.

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.

Mater, B. D., Schaad, S. M. & Venayagamoorthy, S. K. 2013 Relevance of the Thorpe length scale in stably stratified turbulence. *Phys. Fluids* **25** (7), 076604.

Mater, B. D. & Venayagamoorthy, S. K. 2014 A unifying framework for parameterizing stably stratified shear-flow turbulence. *Phys. Fluids* **26** (3), 036601.

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.

Moum, J. N. 1996 Energy-containing scales of turbulence in the ocean thermocline. *J. of Geophys. Res.* **101** (C6), 14095–14109.

Nikurashin, M. & Ferrari, R. 2011 Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophysical Research Letters* **38** (8), L08610.

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

Osborn, T. R. & Cox, C. S. 1972 Oceanic fine structure. *Geophys. Astrophys. Fluid Dyn.* **3** (1), 321–345.

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

Pham, H. T. & Sarkar, S. 2010 Transport and mixing of density in a continuously stratified shear layer. *J. Turb.* **24**, 1–23.

Potylitsin, P. G. & Peltier, W. R. 1998 Stratification effects on the stability of columnar vortices on the f-plane. *J. Fluid Mech.* **355**, 45–79.

Potylitsin, P. G. & Peltier, W. R. 1999 Three-dimensional destabilization of Stuart vortices: the influence of rotation and ellipticity. *J. Fluid Mech.* **387**, 205–226.

Rahmani, M., Lawrence, G. A. & Seymour, B. R. 2014 The effect of Reynolds number on mixing in Kelvin-Helmholtz billows. *J. Fluid Mech.* **759**, 612–641.

Salehipour, H., Caulfield, C. P. & Peltier, W. R. 2016*a* Turbulent mixing due to the Holmboe wave instability at high Reynolds number. *J. Fluid Mech.* **803**, 591–621.

Salehipour, H. & Peltier, WR 2015 Diapycnal diffusivity, turbulent Prandtl number and mixing efficiency in boussinesq stratified turbulence. *J. Fluid Mech.* **775**, 464–500.

Salehipour, H., Peltier, W.R., C.B., Whalen & MacKinnon, J.A. 2016*b* A new characterization of the turbulent diapycnal diffusivities of mass and momentum in the ocean. *Geophys. Res. Lett.* **43**, 3370–3379.

Salehipour, H., Peltier, W. R. & Mashayek, A. 2015 Turbulent diapycnal mixing in stratified shear flows: the influence of Prandtl number on mixing efficiency and transition at high Reynolds number. *J. Fluid Mech.* **773**, 178–223.

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.

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*a* Anisotropy of turbulence in stably stratified mixing layers. *Phys. Fluids* **12**, 1343–1362.

Smyth, W. D. & Moum, J. N. 2000*b* Length scales of turbulence in stably stratified mixing layers. *Phys. Fluids* **12**, 1327–1342: Herein SM00.

Smyth, William D & Moum, James N 2001 3D turbulence. *Encyclopedia of Ocean Sciences* .

Smyth, W. D. & Moum., J. N. 2012 Ocean mixing by Kelvin-Helmholtz instability. *Oceanography* **25**, 140–149.

Smyth, W. D. & Moum, J. N 2013 Marginal instability and deep cycle turbulence in the eastern equatorial Pacific ocean. *Geophys. Res. Lett.* **40**, 6181–6185.

Smyth, W. D. & Peltier, W. R. 1994 Three-dimensionalization of barotropic vortices on the f-plane. *J. Fluid Mech.* **265**, 25–64.

St. Laurent, L. & Simmons, H. 2006 Estimates of power consumed by mixing in the ocean interior. *J. Climate* **19**, 4877–4890.

Thorpe, S. A. 1977 Turbulence and mixing in a Scottish loch. *Phil. Trans. Roy. Soc. Lond. A* **286** (1334), 125–181.

Thorpe, S. A. 2005 The turbulent ocean. *Cambridge University Press* .

Van Haren, H., Gostiaux, L., Morozov, E. & Tarakanov, R. 2014 Extremely long Kelvin-Helmholtz billow trains in the Romanche Fracture zone. *Geophys. Res. Lett.* **41**, 8445–8451.

Voet, G., Girton, J. B., Alford, M. H., Carter, G. S., Klymak, J. M. & Mickett, J. B. 2015 Pathways, volume transport, and mixing of abyssal water in the samoan passage. *J. Phys. Oceanogr.* **45** (2), 562–588.

Waterhouse, A. F., MacKinnon, J. A., Nash, J. D., Alford, M. H., Kunze, E., Simmons, H. L., Polzin, K. L., St Laurent, L. C., Sun, O., Pinkel, R., Talley, L. D., Whalen, C. B., Huussen, T. N., Carter, G. S., Fer, I., Waterman, S., Garabato, A. C. N., Sanford, T. B. & Lee, C. M. 2014 Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate. *J. Phys. Oceanogr.* **44** (7), 1854–1872.

Wijesekera, H. W & Dillon, T. M. 1997 Shannon entropy as an indicator of age for turbulent overturns in the oceanic thermocline. *Journal of Geophysical Research* **102** (C2), 3279–3291.

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 turbulent mixing across a density interface in a turbulent Taylor-Couette flow. *J. Fluid Mech.* **663**, 343–347.

Wunsch, C. & Ferrari, R. 2004 Vertical mixing, energy, and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314.