FLOW CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

A physical–statistical recipe for representation of small-scale oceanic turbulent mixing in climate models

A. Mashayek[1],* ⓘ, B.B. Cael[2], L. Cimoli[3], M.H. Alford[3] and C.P. Caulfield[4,5] ⓘ

[1]Department of Civil and Environmental Engineering, Imperial College London, London SW7 2BX, UK
[2]National Oceanography Centre, Southampton SO14 3ZH, UK
[3]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA
[4]BP Institute for Multiphase Flow, University of Cambridge, Cambridge CB3 0EZ, UK
[5]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK
*Corresponding author. E-mail: mashayek@ic.ac.uk

Abstract

It is well established that small-scale cross-density (diapycnal) turbulent mixing induced by breaking of overturns in the interior of the ocean plays a significant role in sustaining the deep ocean circulation and in regulating tracer budgets such as those of heat, carbon and nutrients. There has been significant progress in the fluid mechanical understanding of the physics of breaking internal waves. Connection of the microphysics of such turbulence to the larger scale dynamics, however, is significantly underdeveloped. We offer a hybrid theoretical–statistical approach, informed by observations, to make such a link. By doing so, we define a bulk flux coefficient, Γ_B, which represents the partitioning of energy available to an 'ocean box' (such as a grid cell of a coarse resolution climate model), from winds, tides, and other sources, into mixing and dissipation. Here, Γ_B depends on both the statistical distribution of turbulent patches and the flux coefficient associated with individual patches, Γ_i. We rely on recent parametrizations of Γ_i and the seeming universal characteristics of statistics of turbulent patches to infer Γ_B, which is the essential quantity for representation of turbulent diffusivity in climate models. By applying our approach to climatology and global tidal estimates, we show that, on a basin scale, energetic mixing zones exhibit moderately efficient mixing that induces significant vertical density fluxes, while quiet zones (with small background turbulence levels), although highly efficient in mixing, exhibit minimal vertical fluxes. The transition between the less energetic to more energetic zones marks regions of intense upwelling and downwelling of deep waters. We suggest that such upwelling and downwelling may be stronger than previously estimated, which in turn has direct implications for the closure of the deep branch of the ocean meridional overturning circulation as well as for the associated tracer budgets.

Impact Statement

Waves similar to those observed at the beach exist throughout the ocean interior and are induced by tides, winds, currents, eddies, and other processes. Similar to beach waves, internal waves can also roll up and break. Widespread internal-wave breaking helps drive the ocean circulation by upwelling the densest waters that form in polar regions and sink to the ocean abyss. They also play an important role in transport and storage of heat, carbon and nutrients. In this work, we show how well-understood physical understanding can be used in conjunction with statistics of observed ocean turbulence to improve our understanding of the impact of small-scale mixing on the global ocean, and thereby on the climate system.

CrossMark

1. Introduction

Turbulence induced by breaking density overturns in the ocean interior, such as those induced by internal waves or boundary layer turbulence, plays a key role in regulating the global ocean circulation and budgets of climatically important tracers such as heat, carbon and nutrients (Garabato & Meredith, 2022; Talley et al., 2016; Whalen et al., 2020). Such turbulence is primarily excited at the bottom of the ocean through interaction of tides, currents and eddies with bottom topography and at the surface through the wind stress acting on the sea-surface (Alford, 2020; Garabato & Meredith, 2022; Garrett & Kunze, 2007; Legg, 2021; Nikurashin & Ferrari, 2013). As an example, Figure 1 shows an observationally sampled abyssal high mixing zone, the Samoan Passage, where northward flow of Antarctic Bottom Waters through a constriction generates strong turbulence through a generation of wave-induced and hydraulically controlled density overturns (Alford et al., 2013; Carter et al., 2019). Figure 1*b* highlights the intermittency of turbulence, which renders it difficult to sample sufficiently to allow for accurate quantification of properties of interest such as turbulent mixing. Thus, sampling poses a monumental challenge to connecting our understanding of the physics of turbulence, which occur on scales $O(10^{-3}–10^{-1})$ m, to the larger scale regional and global implications, which are relevant on scales $O(10^6–10^7)$ m. In this sense, this problem is analogous to cloud physics: both involve micro-physics with leading order global impacts, both occur on scales much smaller than typical climate model resolutions and thus need to be parametrized, and both have been notoriously difficult to parametrize and contribute significantly to inaccuracies in model solutions. In this paper, we provide a novel methodology, based on recent progress in our understanding of the microphysics of mixing (Mashayek, Caulfield, & Alford, 2021, hereafter MCA21) and statistics of wave-induced turbulence (Cael & Mashayek, 2021, hereafter CM21), to connect the physics of small-scale turbulent mixing to large-scale dynamics. While we make specific choices with regard to the physical parametrization of mixing and statistical distributions of ocean turbulence, our choices may reasonably be thought of as placeholders, based on the best we have today. The machinery to which we refer as a 'recipe', however, is a broader framework, the components of which can improve over time, for linking physics and statistics to represent small-scale mixing in ocean/climate models.
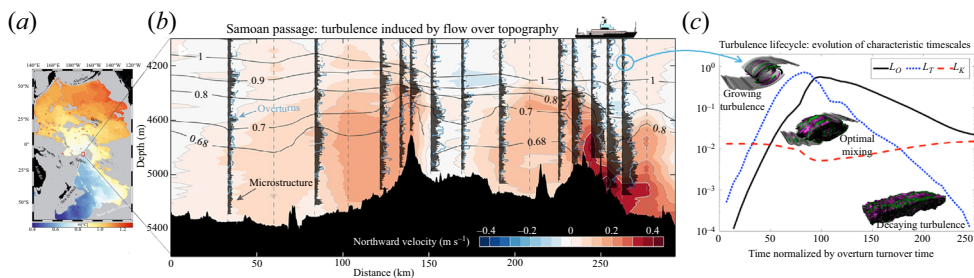

Figure 1. (a) Bottom water temperature in the Samoan Passage, a chokepoint of abyssal ocean circulation. (b) Turbulence in the passage: northward velocity (colours), potential temperature (black contours), and dissipation rate measured by a turbulence microstructure profiler (shaded black profiles) and inferred from Thorpe scales (blue profiles). Panels (a,b) are from Alford et al. (2013). (c) Evolution of the overturn scale, L_T, and upper turbulence bound, L_O, as well as the viscous dissipation scale, L_K, for an archetypal mixing process in the form of a shear instability of Kelvin Helmholtz type (Mashayek, Caulfield, & Peltier, 2013). The insets show the turbulence breakdown of the wave by showing the spanwise (out of the page) vorticity in grey and counter-rotating streamwise (along flow) vorticity iso-surfaces in green and purple, illustrating the hydrodynamic instabilities that facilitate a wave breakdown (Mashayek & Peltier, 2012a, 2012b). Such instabilities create a turbulence cascade that transfers energy from the energy-containing scale (L_T) to the scales where molecular dissipation erodes momentum (L_K).

The paper is organized as follows. In § 2, we review the physics of small-scale turbulent mixing. In § 3, we provide an overview of the statistics of the observed ocean turbulence and argue that the chronic undersampling of turbulence highlights the inevitability of a statistical approach. In § 4.1, we introduce a recipe that can put the physics and statistics together to infer Γ_B. In § 5, we showcase the application of the methodology to an ocean basin. We finish by discussions of the results and their implications in § 7.

2. Physics of wave-induced turbulence

Our understanding of the physics of wave-induced density stratified turbulent mixing has progressed significantly over the past few decades (Caulfield, 2021; Ivey, Winters, & Koseff, 2008; Peltier & Caulfield, 2003). A key question concerns how the total power available to turbulence, $\mathcal{P}$, from winds, tides and other sources, is partitioned into (I) mixing, $\mathcal{M}$, defined as a net vertical irreversible buoyancy flux, and (II) dissipation into heat (due to the seawater viscosity), the rate of which is referred to as ε. The turbulent mixing may be approximated by

$$\mathcal{M} \approx \kappa N^2 \approx \Gamma \varepsilon, \tag{2.1}$$

where $N^2 = -(g/\rho_0)\partial_z \rho$ represents the density stratification, g is the gravitational constant and ρ_0 is a reference density (Osborn, 1980) (note that while the caveats associated with this approximation are important (Mashayek et al., 2013; Mashayek & Peltier, 2013), they do not affect the premise or conclusions of this work in any major way). In physical terms, $\mathcal{M}$ represents a net vertical mass flux as turbulence works against gravity to lift denser waters upward. Such turbulence feeds upon the available potential energy (APE) stored in overturns. Upon mixing, a fraction of the APE gets converted into the background potential energy (i.e. the potential energy that the system would acquire if it were allowed to come to rest adiabatically), with the rest dissipating to heat and increasing the internal energy of the system (which leads to an insignificant temperature rise due to the high heat capacity of seawater; Peltier & Caulfield, 2003). The turbulent diffusivity, κ, is an input parameter in climate models to account for the subgrid-scale unresolved turbulent mixing. The turbulent flux coefficient, $\Gamma = \mathcal{M}/\varepsilon$, is often taken to be a constant value of 0.2 for historical reasons and arguably for the lack of a universally accepted parametrization for it. However, it is well known to be highly variable (Caulfield, 2021; Gregg, D'Asaro, Riley, & Kunze, 2018; Mashayek & Peltier, 2013), with appreciable consequences for ocean circulation (Cimoli et al., 2019; de Lavergne, Madec, Le Sommer, Nurser, & Naveira Garabato, 2015; Mashayek et al., 2017). While $\mathcal{M}$ is often the quantity of interest from a physical perspective, it is difficult to observe directly and is often inferred from ε (via (2.1)) which itself is inferred from microscale shear (i.e. spatial gradients of velocity) measured by microstructure probes on profiling instruments, gliders or moorings. Thus, accurate quantification of Γ is key to inferring ocean mixing from direct observations.

Three basic turbulence scales have proven useful for parametrization of Γ based on observable quantities. To illustrate their relevance, in Figure 1c, we show their evolution over the life cycle of turbulence breakdown of a canonical overturn. The Kolmogorov scale, $L_K = (\nu^3/\varepsilon)^{1/4}$, represents the scale below which viscous dissipation takes kinetic energy out of the system, the Ozmidov scale, $L_O = (\varepsilon/N^3)^{1/2}$, is the largest (vertical) scale that is not strongly affected by stratification, and the Thorpe scale, L_T, is a geometrical scale characteristic of vertical displacement of notional fluid parcels within an overturning turbulent patch (Dillon, 1982; Mashayek, Caulfield, & Peltier, 2017; Smyth & Moum, 2000) – ν is the kinematic viscosity of seawater. The figure shows the initial accumulation of available potential energy (APE) from background shear into the primary billow (large L_T) followed by the growth of smaller eddies within the main billow upon feeding on its APE source (increase in L_O). As turbulence grows, the scale at which energy is taken out of the system, L_K, decreases. The phase where $L_O \sim L_T$ is the most efficient transfer of energy from the overturn to the smaller scales, and thus is the richest dynamic range, marked by the largest gap between L_O and L_K, and thus the most efficient phase of the flow where (an instantaneous) mixing efficiency is defined as $\mathcal{M}/(\mathcal{M} + \varepsilon) = \Gamma/(1 + \Gamma)$ (Peltier & Caulfield, 2003). Beyond this efficient phase, both L_T and L_O decay, as does the turbulence.

The ratio of L_O and L_K, often expressed as the buoyancy Reynolds number $Re_b = (L_O/L_K)^{4/3}$, has been widely used to quantify Γ (Bouffard & Boegman, 2013; Gregg et al., 2018; Mashayek et al., 2017; Monismith, Koseff, & White, 2018) and also to establish the global scale impacts of variations in Γ (Cimoli et al., 2019; de Lavergne et al., 2015; Mashayek et al., 2017). However, some of these efforts, as well as others, have also highlighted the inherent deficiency of Re_b since it only includes instantaneous information on the turbulence scales L_O and L_K while not 'knowing' anything about either the energy containing scale L_T (Gargett & Moum, 1995; Garrett, 2001; Mashayek et al., 2021; Mashayek & Peltier, 2011; Mater & Venayagamoorthy, 2014) or any time dependence of the flow, although evidence is accumulating that 'history matters' in stratified mixing (Caulfield, 2021). An alternate method for inferring mixing (or more accurately inferring ε from which diffusivity or flux may be inferred), employed when direct inference of ε is not available, is to assume that $R_{OT} = L_O/L_T$ is a constant (taken to be between 0.6 and 1; Dillon, 1982; Ferron, Mercier, Speer, Gargett, & Polzin, 1998; Thorpe, 2005) and $\Gamma = 0.2$. This method, too, is inherently deficient (although in practice might be the only available option) since it only 'knows' about the energy containing scale and not the dissipation scale, and so does not 'feel' the width of the dynamic range of the turbulence (the so-called inertial subrange).

Building on an extensive relevant literature, recently, MCA21 extended earlier scaling relations relating R_{OT} to Γ and proposed a simple parametrization for Γ, on basic physical grounds, that agreed well with observational data:

$$\Gamma_i = A \frac{R_{OT}^{-1}}{1 + R_{OT}^{1/3}}, \tag{2.2}$$

where A is a constant ranging between $1/2$ and $2/3$. The subscript i in (2.2) emphasizes the applicability of it to individual turbulent patches as opposed to a bulk region, which typically comprises a multitude of turbulence events separated (in time and space) by 'quiet' zones. Crucially, this parametrization captures the fundamental time-dependent nature of overturning mixing events, allowing for variation in R_{OT} and hence Γ at different stages in the life cycle of a mixing event. Figure 2a, reproduced from MCA21, shows the agreement with a combination of data from ~50 000 turbulent patches gathered from six different field campaigns that sampled turbulence in different geographical locations around the globe, at different depths and turbulence regimes, and from turbulence induced by different processes (see supplementary material available at https://doi.org/10.1017/flo.2022.16 for a brief description of data). Equation (2.2) reduces to A/R_{OT} in the limit of young turbulence ($R_{OT} \ll 1$)) and to $A/R_{OT}^{4/3}$ in the limit of (older) decaying turbulence ($R_{OT} \gg 1$)). While the latter has been suggested by others in the past, as reviewed in MCA21, the former limit and the transition between the two limits were formulated in MCA21.

A nice feature of (2.2) is that it allows for the two ranges to merge smoothly at $R_{OT} \sim 1$. This limit corresponds to efficient mixing when there exists an optimal balance between the stratification and energy available to turbulence. In the young turbulence range, stratification is relatively high and sufficient energy has not yet transferred to turbulence to work against the stratification and mix. So, while Γ_i can be very large, it does not imply much mixing: it is large since ε is very small, not because $\mathcal{M}$ is large. In fact, in the limit of laminar flow, $\Gamma_i \to \infty$. However, in the decaying phase of turbulence, stratification is somewhat eroded and so $\mathcal{M}$ is weaker as there is less to mix, while ε is still finite as even an unstratified flow can have significant ε. Thus, in this limit, $\Gamma_0 \to 0$. It is the $R_{OT} \sim 1$ limit in which the right balance of stratification and power exists and optimal mixing occurs. Since this intermediate phase appears to lead to 'optimal' mixing, neither too hot nor too cold but 'just right', MCA21 referred to this as 'Goldilocks mixing'.

Here we argue that all three (time-dependent) phases of the turbulence life cycle, importantly including the young turbulence limit, are key to connecting the small-scale physics of mixing to the large-scale ocean dynamics. While the Goldilocks mixing phase is when most of the effective turbulent flux occurs, the young and weakly turbulent patches are important since most of the ocean interior is relatively 'quiet' with intermittent bursts of turbulence. We will argue that it is essential to account for the
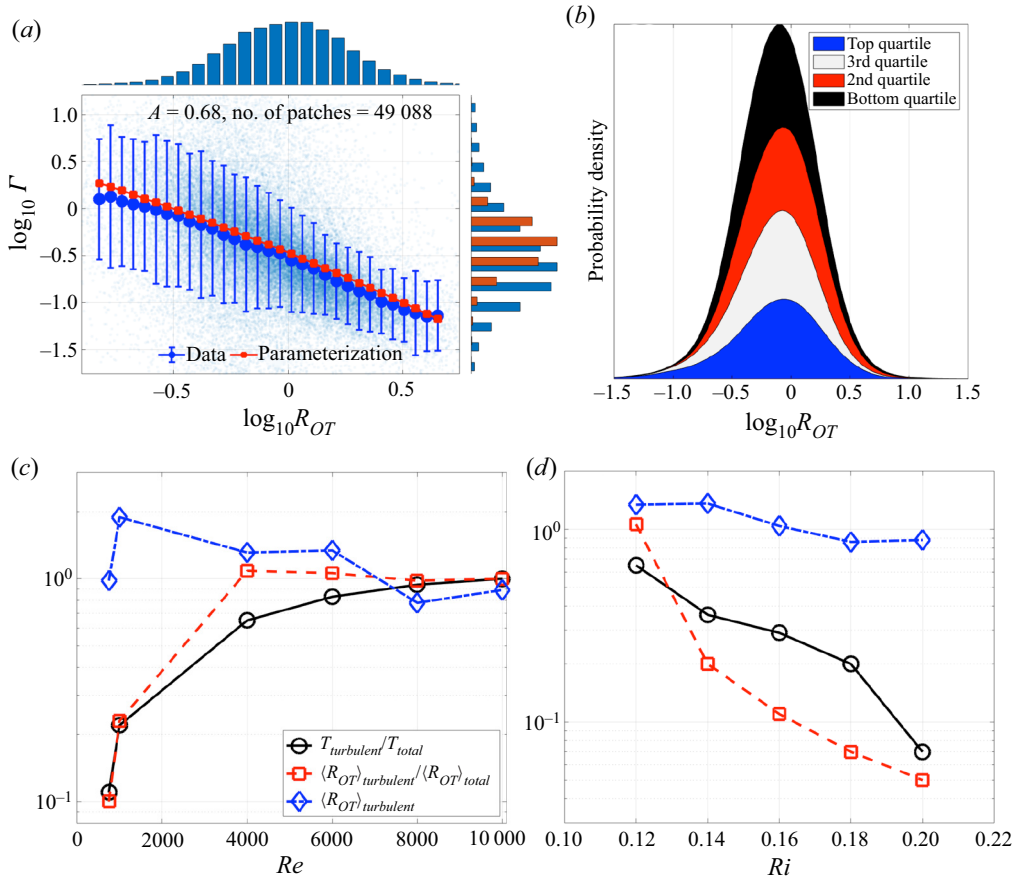
Figure 2. *(a) Agreement of* (2.2) *with ~50 000 turbulent patches from six different datasets covering a variety of turbulent regimes and processes. The bar plot insets show the histogram of R_{OT} (top axis) and Γ for the parametrization (in red) and data (in blue) along the right vertical axis. The value of A is obtained through regression of data to* (2.2). *Reproduced from Mashayek et al. (2021) – see supplementary materials for a brief description of data. Note that the data plotted here include observations in the Samoan Passage (shown in Figure 1b) and in the Brazil Basin (of relevance to Figures 5–7). (b) Probability density function (PDF) of $R_{OT} = L_O/L_T$ for the same data as in panel (a), compartmentalized in terms of the rate of dissipation of kinetic energy, ε. The bottom/second/third/top quartiles have increasing modes of 0.66/0.78/0.82/0.88, respectively. Note the negative skewness of the log-transformed PDF. (c,d) Temporal fraction of turbulence lifecycle, as well as the ratio of R_{OT} during the turbulent phase of the flow to its mean value over the whole life cycle. Each symbol represents a life-cycle-averaged quantity from a direct numerical simulation (such as the one in Figure 1d) for the corresponding Reynolds and Richardson numbers. All cases in panel (c) are for Ri = 0.12 while all cases in panel (d) are for Re = 6000. The turbulent phases of the life cycles are defined as the times when $Re_b > 20$ (Gibson, 1991; Smyth & Moum, 2000).*

less/non-turbulent regions, whereas historically, parametrizations have primarily focused on energetic turbulence. This will necessitate careful analysis of the statistics of turbulent patches.

We note that while we will employ the Goldilocks paradigm of MCA21 in the forthcoming recipe, in principle, it can be replaced by alternative parametrizations for Γ_i.

3. Statistics of turbulence

3.1. *Statistics of patches*

Figure 2b shows the histogram of the data used in Figure 2a separated into four quartiles in terms of ε. The distribution shows that most turbulent patches lie within a factor of 3 of $R_{OT} = 1$ and the larger the ε, the closer to 1 the peak of R_{OT} lies. This suggests the majority of turbulent patches are in this phase of optimal or Goldilocks mixing. Such a clustering of data has been widely reported in the past (Dillon, 1982; Mashayek et al., 2017; Mater, Venayagamoorthy, Laurent, & Moum, 2015; Thorpe, 2005) and suggests that out of the three phases of energetic ocean turbulence events, namely 'young' growth, intermediate Goldilocks mixing and 'old' decay, the intermediate phase spans the larger fraction of the turbulence life cycle. This is shown quantitatively in Figures 2c and 2d where life-cycle-averaged properties are plotted from turbulence life cycle simulations of shear instabilities by Mashayek and Peltier (2013), Mashayek et al. (2013). Figures 2c and 2d, together, show that for sufficiently energetic turbulence, a larger proportion of the turbulence life cycle corresponds to the Goldilocks phase rather than to the growth and decay phases.

The parametrization (2.2) describes individual turbulent patches. A turbulent region within the ocean, such as that shown in Figure 1b, hosts many patches in different stages of their evolution. MCA21 argued that an appropriate value for a bulk Γ may be constructed for such a region through

$$\Gamma_B = \frac{\mathcal{M}_B}{\varepsilon_B} \approx \frac{\sum\limits_{i=1}^{n} \Gamma_i \times \varepsilon_i}{\sum\limits_{i=1}^{n} \varepsilon_i}, \tag{3.1}$$

where n represents the number of patches in a region of interest and the subscript B denotes 'Bulk'. For example, typical resolution of climate models is $O(100 \text{ km})$ in the horizontal and $O(100 \text{ m})$ in the vertical direction. Models that employ a temporally evolving mixing parametrization, employ $\Gamma_B = 0.2$ to construct a diffusivity for each grid cell. Of course there is no obvious reason why such a constant should universally hold, and as we shall show, Γ_B relies on the statistical distribution of patches within the grid cell and their associated Γ_i values. By applying (3.1) to four datasets, MCA21 showed that Γ_B can be close to the Goldilocks mixing (i.e. $A/2 \approx 1/3$ in (2.2)) when the region of study has the right balance of power and stratification (where $R_{OT} \sim 1$), thereby comprising mostly Goldilocks mixing patches. However, regions that host a higher percentage of young or weak turbulent patches can have $\Gamma_B \sim O(1)$ since young patches have large Γ_i values, as shown in Figure 3 in the limit of small R_{OT} (Young patches have large Γ because of small ε NOT due to large $\mathcal{M}$. Thus, while they do not contribute much to the net turbulent flux $\sum \Gamma_i \varepsilon_i$, they bias Γ_B high).

3.2. *Statistics of continuous profiles*

Bulk mixing therefore depends on the statistics of turbulent patches. However, weakly/non-turbulent waters reside in between intermittent turbulent patches. Recently, CM21 showed that ε data from over 750 full depth microstructure profiles from 14 field experiments (covering a wide range of depths, geographical locations and turbulence-inducing processes; see CM21 for details), are well described by a log-skew-normal distribution (Figure 4a), which has the form

$$f(\varepsilon; \xi, \omega, \alpha) = \frac{2}{\omega \varepsilon} \phi\left(\frac{\log \varepsilon - \xi}{\omega}\right) \varphi\left(\alpha \frac{\log \varepsilon - \xi}{\omega}\right), \tag{3.2}$$

where f is a probability density function and ϕ and φ respectively are the probability and cumulative density functions (PDF and CDF) of a standard Gaussian random variable. We refer the reader to CM21
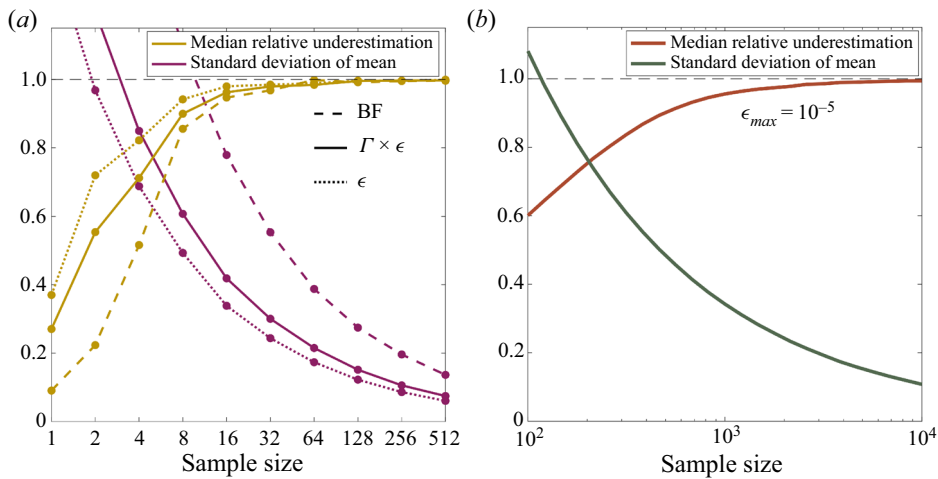
Figure 3. *(a) Sampling bias and uncertainty for buoyancy flux (dashed lines), $\Gamma \times \varepsilon$ (solid lines) and ε (dotted lines) for a single turbulent event (e.g. Figure 1c). Purple lines indicate the normalized standard deviation and yellow lines indicate the median relative underestimation, each as a function of sample size, estimated from bootstrap resampling a direct numerical simulation (shown in Figure 1c, from Mashayek et al., 2013). For example, relative uncertainty is >100 % for the time-averaged buoyancy flux of this event with fewer than ~16 samples, and the median time-averaged buoyancy flux estimate with four or fewer samples underestimates the buoyancy flux by a factor of two or more. (b) Sampling bias and uncertainty for the mean ε sampled from a log-skew-normal distribution with the parameters estimated from the dataset described in the text, discarding ε values >10^{-5} m^2 s^{-3}. The green line indicates the normalized standard deviation and the orange line indicates the median relative underestimation, each as a function of sample size, estimated from bootstrap resampling.*

for details, but here discuss the essential aspects of that paper for the problem at hand. As discussed in CM21, the relationship between the log-skewness (θ) of the distribution and the additional shape parameter α is one-to-one. The parameters (ξ, ω, α) are related to the log-mean, log-standard deviation and log-skewness (μ, σ, θ) of a log-skew-normal random variable according to the following equations:

$$\mu = \xi + \sqrt{\frac{2}{\pi}}\omega\delta, \tag{3.3}$$

where

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}, \quad \sigma = \omega\sqrt{1 - \frac{2}{\pi}\delta^2}, \quad \theta = (4-\pi)\left(\sqrt{\frac{2}{\pi}}\delta\right)^3 \bigg/ \left(2\left(1-\frac{2}{\pi}\delta^2\right)^{3/2}\right). \tag{3.4a–c}$$

The log-skew-normal distribution arises from the analogue of the Central Limit Theorem for log-normal variables: that is, the sum of log-normal variables converges to a log-skew-normal distribution (Wu et al., 2009). For turbulence, the log-skew-normal distribution is thought to arise because the total, measured ε results from a combination of multiple and/or not statistically steady turbulence-generating processes (Caldwell & Moum, 1995), which individually and/or instantaneously have log-normally distributed dissipation rates.
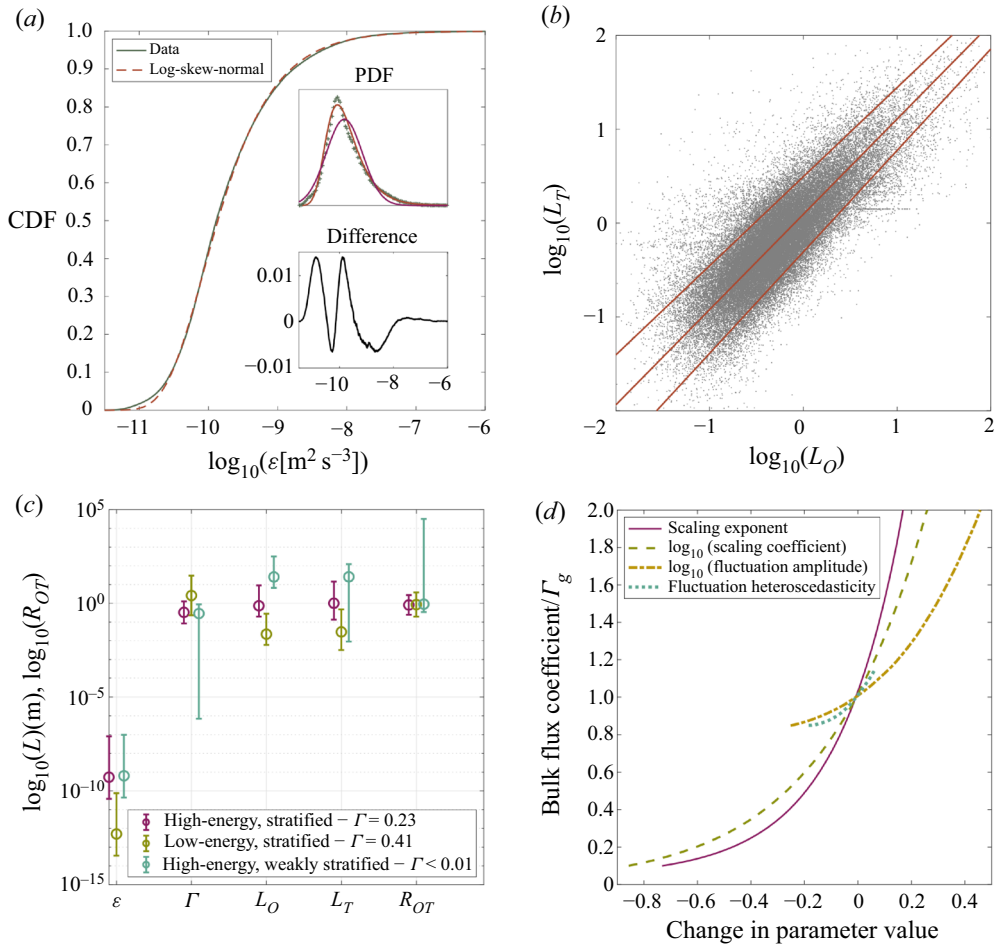
Figure 4. *(a) Reproduced from* CM21, *ε data from over 750 full depth microstructure profiles from 14 field experiments (see* Cael and Mashayek (2021) *for more details) are excellently characterized by a log-skew-normal distribution. Cumulative distribution functions (CDFs) are shown in the main plot; the upper inset shows the corresponding probability density functions and that of a log-normal distribution (purple line) for comparison; the lower inset shows the difference between the empirical and hypothesized log-skew-normal CDFs. (b) Regression of $\log_{10}(L_T)$ against $\log_{10}(L_O)$. The middle line captures the central scaling relationship and is estimated via model II regression as described in the text; the outside lines capture the heteroscedasticity of the residuals and is estimated via quartile regression as described in the text. (c) Mean quantities from the dataset in panel (b) (with error bars representing standard deviation), grouped into three categories: (I) energetic turbulence in weak stratification, (II) weak turbulence in strong stratification, and (III) energetic stratified turbulence; see main text for a discussion. (d) Sensitivity of the bulk flux coefficient to the different L_T–L_O scaling parameters, perturbed from a baseline Γ_g estimated from the combined global dataset described in the text. Larger fluctuations in L_T when L_O is large can increase the bulk flux coefficient, but it asymptotes to a constant value with decreasing fluctuations. Increasing either the scaling exponent or coefficient increases L_T values for large L_O, thus increasing the bulk flux coefficient for large ε; decreasing either the scaling exponent or coefficient drives bulk mixing to zero as $R_{OT} \gg 1$ when ε is large. See supplementary materials for information on data sources.*

3.3. Chronic undersampling: inevitability of a statistical approach to Γ_B

The difficulty of sampling intermittent turbulence has long been recognized (Gregg, Seim, & Percival, 1993), but the problem is perhaps even worse than conventionally thought, owing to the even more heavy-tailed nature of a log-skew-normal ε distribution than a lognormal one. This presents a severe challenge for constraining the average ε value for an ocean volume from a limited sample set, since many samples are required to resolve the disproportionately consequential tail. Direct numerical simulations have established that turbulent quantities, such as ε and Γ, vary substantially over the lifetime of an overturning event (Caulfield, 2021). This implies that a single measurement, or even a handful of measurements, of such an event can be a poor estimate of its time-integrated properties since they most likely miss the most energetic phase of turbulence for each event. This underscores the need for large measurement sets and a statistical approach to fill in the gaps and accurately estimate bulk turbulent properties. To highlight the sampling issue, in Figure 3, we consider two examples: one based on sampling of an individual turbulence event; and one based on sampling of the collective global dataset discussed in Figure 4*a*.

Direct numerical simulations, such as that shown in Figure 1*c*, clearly demonstrate that turbulent quantities such as $\mathcal{M}$, ε and Γ vary substantially over the lifetime of an overturning event. This implies that a single measurement, or even a handful of measurements, of such an event are highly likely to be a poor estimate of its time-integrated properties. Figure 3*a* shows the uncertainty and bias associated with random samples of one (other simulations yielded similar results) simulation's history ($Re = 6000, Ri = 0.16$) for different sample sizes. Uncertainty is quantified by the normalized standard deviation of the sample mean for collections of samples taken at random times along the temporal history of the overturning event. Bias is quantified by the ratio of the median sample mean from these collections of randomly timed samples to the true time-averaged mean of each property. Because all the properties shown are positively skewed quantities, the sample means tend to miss the peak values and thus underestimate the time-averaged value; because these skews are so large, means of different collections of random samples vary by e.g. more than a factor of two for fewer than $\sim$16 measurements. Turbulent measurements are snapshots of events; it is not possible to make tens of measurements of the same turbulent event. These appreciable biases and uncertainties for characterizing individual events thus underscore the importance of the statistics of turbulent properties.

The heavy-tailed nature of the ε distribution in Figure 4*a* presents a similar challenge for constraining the average ε value for an ocean volume from a limited sample set. Figure 3*b* shows the same as Figure 3*a*, but for random samples from a log-skew-normal distribution whose parameters are fit to match the ε data from the experiments described above ($\xi = -24.8$, $\omega = 3.91$, $\alpha = 5.89$) using the same procedure as in CM21. We discard values above a chosen ε_{max} threshold of 10^{-5} m^2 s^{-3}; as with any parametrization for the PDF of ε supported on $(0, \infty)$, the log-skew-normal allows for a non-zero probability density for unmeasurably and/or unphysically large ε values, which are either too large to be measured by sampling probes and/or yield impossibly large L_O values. Here, $\mathcal{O}(1000)$ samples are needed to make the underestimation bias less than $\sim$10 %, and $\mathcal{O}(100)$ samples are needed to make the standard deviation of sample means less than the true mean (i.e. <100 % relative uncertainty). (Larger/smaller values of ε_{max} increase/reduce these sample size numbers.) This further underscores the need for large measurement sets and a statistical approach to estimate bulk turbulent properties accurately.

4. A recipe for Γ_B

Ingredients: $\Gamma(R_{OT})$ for individual patches + statistics of L_O, L_T, and R_{OT}.

Here, we aim to exploit the log-skew-normality of ε and the statistics of R_{OT} to construct a parametrization for Γ_B based on the total power and the mean stratification for the region (or the grid cell) for which Γ_B is sought. To this end, first we note that $L_O \propto \varepsilon^{1/2}$ by definition; the former lies within the R_{OT}-based Γ parametrization in (2.2) and the latter is captured by (3.2) (historically, ε has often been described as log-normally distributed based on a simple argument from multi-stage subdivisions of

an initial flux for three-dimensional homogeneous isotropic turbulence (Gurvich & Yaglom, 1967), but the log-normal distribution has also been long recognized as a quantitatively inaccurate description of measured distributions (Yamazaki & Lueck, 1990)). Because L_O divides ε by another random variable and then takes a square root of their quotient, this distribution is also preserved for L_O (supplementary material, Figure S1, Table S1; also see below). Here, we pair the parametrizations for the statistical distribution of ε in (3.2) and $\Gamma_i = f(R_{OT})$ in (2.2) to obtain a parametrization for Γ_B.

If, as described above, ε_B is log-skew-normally distributed because it is the sum of log-normal random variables (call these ε_i), then L_O must also be so distributed. If we pass the N^3 to the summed-over ε_i, such that $\varepsilon_B/N^3 = \sum_i(\varepsilon_i/N^3)$, this will introduce a correlation into the ε_i/N^3 being summed over, which does not affect the log-skew-normality of their sum (Hcine & Bouallegue, 2015). For the ε_i, this introduces another variable multiplicative factor and these each should still then be log-normal, so altogether $\sum_i(\varepsilon_i/N^3)$ is still the sum of log-normal random variables. As the skew-normal distribution is insensitive to multiplicative transformations – we may write any skew-normal random variable s as $s = \mu + \sigma(\delta|n_1| + n_2\sqrt{1-\delta^2})$ (Pourahmadi, 2007) so multiplying by some factor m just changes $\mu \rightarrow m\mu$ and $\sigma \rightarrow m\sigma$ – the log-skew-normal must be equivalently insensitive to exponentiation. Thus, if ε_B is log-skew-normally distributed, then so is L_O. We indeed find that all the datasets used here, as well as their aggregate, are well described by a log-skew-normal distribution (Kuiper's statistic $V = 0.021$ for the combined dataset and as well as the median across the individual datasets; supplementary material, Figure S1, Table S1).

In contrast, there is not a clear description of the probability distribution of L_T. While calculating L_T is straightforward conceptually, it involves subjective choices that bias its distribution (Mater et al., 2015). Thus, it is less practical to identify the 'true' distribution for L_T from collections of L_T measurements, as is the equivalent identification of an underlying distribution for ε. It has recently been argued that the size of turbulent overturns, which are thought to have a correspondence with L_T, should be power-law distributed (Smyth, Nash, & Moum, 2019), but we find little to no evidence of this in the datasets we use here (supplementary material, Table S2). Either L_T is not proportional to patch size in the datasets examined here or the patch sizes in these datasets are not power-law distributed (Clauset, Shalizi, & Newman, 2009). The L_T probability distribution for all the datasets used here is unimodal in log-space, such that at least the bulk of the distributions are qualitatively much closer to log-normally distributed than power-law distributed (supplementary material, Figure S1).

Regardless, it is the probability distribution of $R_{OT} = L_O/L_T$ that is of interest here, for which a parametrization of the probability of L_T is not necessary. Instead, one needs a suitable description of the conditional distribution of L_T for a given L_O value, i.e. $P(L_T|L_O)$. Somewhat surprisingly, we find that the log-skew-normal distribution is again an excellent description of the probability distribution of R_{OT} ($V = 0.010$ for the combined dataset and the median $V = 0.027$ across the eight individual datasets; supplementary material, Table S1), but that in this case, the log-skewness is negative. The log-skewness of R_{OT} is in fact only positive for the IH18 dataset (skewness $\tilde{\mu}_3 = 0.48$). Thus, the log-skew-normal distribution is, similar to ε and L_O, a satisfactory parametrization of the probability distribution of R_{OT}, but by and large with a different sign in log-skewness. The question then becomes how L_O and L_T are related such that the distribution of R_{OT}: (i) is log-skew-normal; (ii) has a negative log-skewness and (iii) is peaked at $\mathcal{O}(1)$. In the absence of a theory for the probability distribution of L_T, we derive an empirical construction of $P(L_T|L_O)$ that recapitulates the probability distribution of R_{OT} with simulated L_T values. As L_T appears to be approximately log-normally distributed (supplementary material, Figure S1), it is justifiable to base such a construction on a statistical scaling relationship, with multiplicative fluctuations.

The conditions (i–iii) above can occur for R_{OT} if L_T scales heteroscedastically, i.e. the fluctuations around how L_T scales with L_O are themselves also a function of L_O, and the coefficients of the scaling relationship are close to unity. If $L_T \sim \zeta L_O^\beta$ with $\zeta \approx 1$ and $\beta \approx 1$, then most values of L_O/L_T will be $\mathcal{O}(1)$. If this scaling relationship is tight when L_O is large (i.e. small fluctuations and low probability of small L_T), this will make cases where L_O/L_T is very large unlikely. Conversely, if this scaling relationship is weaker when L_O is small (i.e. larger fluctuations and comparatively higher probability of larger L_T),

this will make cases where L_O/L_T is very small less unlikely. Together, these heteroscedastic effects can change the sign of the log-skewness.

As both L_O and L_T are random variables, their scaling relationship can be captured by model II regression; the heteroscedasticity of this relationship can then be captured by quantile regression, which estimates the conditional quantiles of the response variable (Koenker & Hallock, 2001), applied to the residuals. First, applying model II regression to the combined and log-transformed (L_O, L_T) dataset (i.e. that used in Figure 2a,b), we find the best-fit scaling for these data to be $L_T = 1.24L_O^{1.01}$ (see Figure 4b; we use least-squares cubic regression (York, 1966) but other standard methods such as bisector or major axis yield similar results). Then applying quantile regression to the residuals, we indeed find that the fluctuations in L_T around this scaling relationship increase as L_O decreases. This decrease is characterized by the relationship $r = r_o + r_1 \log_{10}(L_O)$, where r is the amplitude of the residuals around the best-fit scaling, r_o is the residual amplitude when $L_0 = 1$ m and r_1 captures how the residuals decrease as L_O increases. We calculate separate relationships for positive and negative residuals, via quantile regression. (We use the 10th and 90th percentiles to compute r_o and r_1 but our results are not sensitive to this choice.) We can then recover the joint (L_O, L_T) distribution and hence the R_{OT} distribution by simulating the L_T accordingly for a given L_O. The simulated L_T values yield good agreement with the empirical R_{OT} distribution ($V = 0.028$). We define $L_T = \eta \times 1.24L_O^{1.01}$, where η is a log-normal random variable with $\mu = 0$ and $\sigma \propto L_O$. This results in a conditional distribution for L_T that matches the estimated scaling relationships, and is thus an approximation of the joint distribution of the data shown in Figure 4b.

Physically, this scaling behaviour has an intuitive interpretation (as illustrated in Figure 4c): L_T scaling with L_O with an exponent close to unity, such that the peak in the probability density function for R_{OT} is $O(1)$, occurs because for most turbulent overturns, the displacement of fluid parcels in that patch (L_T) will be of the same order of magnitude as the vertical scale that they can be displaced given the background stratification (L_O), as discussed above. If an overturn has a small L_O, this could either be because it is a small overturn, meaning it has a similarly small L_T, or a young overturn with a larger L_T but still low turbulent dissipation. If L_O is large, however, it likely corresponds to the energetic phase of the flow (i.e. the optimal mixing phase in Figure 1c) and so L_T could be reasonably expected to be similarly large. This asymmetry in the life cycle of turbulent overturns produces this skewed, heteroscedastic scaling in Figure 4b. Figure 4c groups the data in Figure 4b (same as that in Figures 2a and 2b) into the three categories discussed above, and shows that (*a*) weak stratification cannot yield a high flux coefficient even at high energy levels as there is barely enough to mix, (*b*) low energy turbulence in the presence of strong stratification can result in a high flux coefficient, but importantly that does not imply a high flux (since flux $\approx \Gamma\varepsilon$); note that in the limit of no turbulence, $\Gamma \to \infty$ but flux tends to zero, and (*c*) for energetic stratified turbulence, the flux coefficient is within the conventional range (0.2–0.3). It is only the latter category (*c*) that corresponds to large effective turbulent flux.

4.1. The recipe

Altogether, these pieces can be assembled into a 'recipe' to compute a bulk flux coefficient as follows. To calculate Γ_B for a coarse resolution grid cell of a general circulation model, which is large enough to enclose sufficient statistics, we imagine a total power $\mathcal{P}$ is available for the grid cell (from one or more sources, such as tides, winds, etc.). Then, we follow the iterative procedure below.

(1) Assume an initial value of 0.2 for Γ_B and using $\mathcal{P} = \mathcal{M}_B + \varepsilon_B$, infer an initial value for ε_B; $\mathcal{M}_B$ and ε_B are the total mixing and dissipation in the cell.
(2) Distribute ε_B over a log-skew-normal distribution of ε for individual patches, so then $\Sigma\varepsilon_i = \varepsilon_B$ and individual ε_i values are random log-skew-normal sample draws.
(3) Calculate L_O for these patches by transforming their ε values and a specified background stratification into a distribution for L_O.
(4) Simulate L_T values corresponding to these L_O values according to the heteroscedastic scaling relationship between L_O and L_T.

(5) Calculate Γ for these patches based on the R_{OT} parametrization and their simulated L_O and L_T values (with a small background diffusivity added; see below).

(6) Calculate $\mathcal{M}_i$ for individual patches based on their Γ_i.

(7) Sum over $\mathcal{M}_i$ of all patches to compute $\mathcal{M}_B$.

(8) Calculate $\Gamma_B = \mathcal{M}_B / \varepsilon_B$.

(9) Back to step one and iterate until Γ_B converges (usually within a few iterations).

We repeat this procedure many times and take the average of the realizations to account for the stochasticity in the process; in practice, for oceanographically relevant values, the variability between realizations is negligible. In step 5, we also add a small (regularizing) background diffusivity of $\kappa_{background} = 10^{-6.5}$ m^2 s^{-1} to account for the background wave field processes not encapsulated in the R_{OT} parametrization. This value, which is close to the molecular diffusion coefficient of heat in seawater, was chosen according to the inflection point on the probability density function of κ from the combined dataset described above, which separates the low tail of the distribution from the bulk of the data (supplementary material, Figure S2). Including this $\kappa_{background}$, mixing for each patch becomes

$$\mathcal{M}_i = \mathcal{M}_{background} + \mathcal{M}_{turbulent} = \Gamma_i \varepsilon_i = \kappa_{background} N^2 + \Gamma_{turbulent} \varepsilon_i, \qquad (4.1)$$

and so the effective total flux coefficient for individual patches may be defined as

$$\Gamma_i = \kappa_{background} N^2 / \varepsilon_i + \Gamma_{turbulent}, \qquad (4.2)$$

where $\Gamma_{turbulent}$ is from (2.2). The advantage of adding the background mixing is that it allows for Γ to tend to large values in the limit of weak turbulence, i.e. when there is not much turbulence due to lack of power and/or overly strong stratification. From a physical perspective, in the limit of laminar flow, $\Gamma \to \infty$ since ε vanishes but $\mathcal{M}$ remains finite. For example, note that Figure 2a shows high values of Γ for observed weakly turbulent young patches.

Steps 1–5 in the above recipe involve a degree of computational expense that might prove impractical if applied to every grid cell of a coarse resolution climate model at every time step. It is, however, straightforward and computationally cheap to do the calculations *a priori* for ranges of power and stratification relevant to the oceans, and then employ a simple lookup table in the models. We employ this approach for the purpose of analysis to be discussed in the next section.

5. Regional/global implications

To highlight the importance of Γ_B for the larger scale circulation, we next apply our recipe to the South Atlantic Ocean. This choice is made for four reasons: (I) it is a region of tidal-shear dominance; (II) it is home to the iconic Brazil Basin Tracer Release Experiment (BBTRE), the first deep observation of bottom generated enhanced turbulent mixing (Ledwell et al., 2000; Polzin, Toole, Ledwell, & Schmitt, 1997; St. Laurent, Toole, & Schmitt, 2001); (III) there exist three-dimensional tidal power estimates which are constructed on theoretical grounds but have been verified to agree excellently with observations including BBTRE (de Lavergne et al., 2020) and (IV) the BBTRE data were shown to correspond to small-scale marginally unstable shear-induced turbulence, and hence consistent with the above-mentioned physical and statistical paradigms laid out in CM21, MCA21.

We construct a regional map from the sum of the tidally generated power and the contribution of the wind-induced near-inertial shear. The former, from de Lavergne et al. (2019, 2020), estimates the power in the internal tide field generated through the interaction of ocean tides with ocean bathymetry. The latter, constructed based on Alford (2020), estimates the power in the downward-propagating wind-induced internal wave field. Figure 5a–c shows the combined power plotted on three density layers (shallow, intermediate depth and deep) (neutral density is often used instead of *in situ* density as it removes the dynamically inconsequential compression of seawater due to increase in pressure with depth. It is also common to deduct 1000 when reporting neutral density, as in Figure 5). The deep layer, as will be shown,
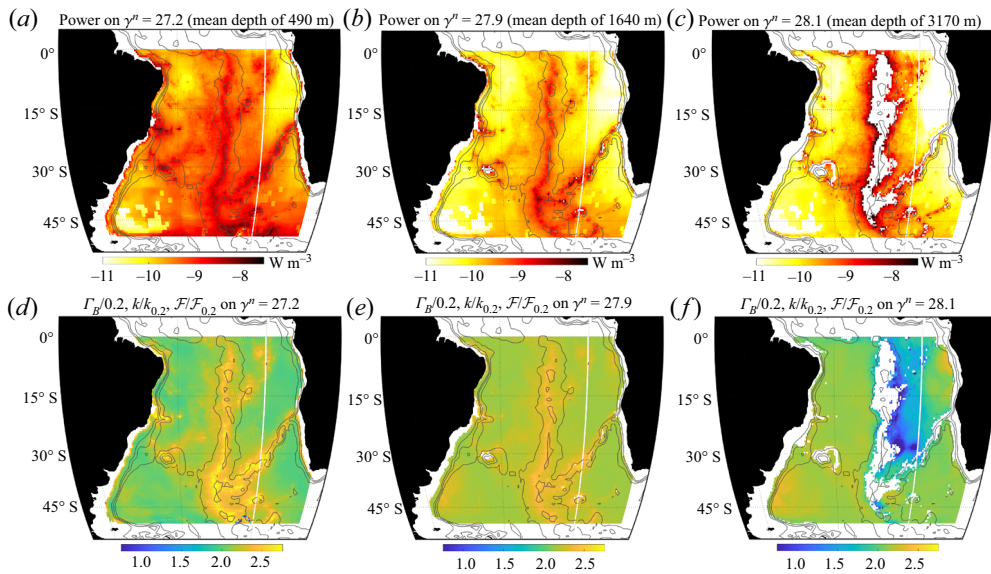
Figure 5. (a–c) Power in the internal wave field in the South Atlantic basin, induced by tides and winds, that is available for mixing and dissipation, plotted on various density levels. (d–f) Γ_B normalized by $\Gamma^* = A/2$ based on (2.2), plotted on the same density levels as in panels (a–c). Since $\kappa \approx \Gamma_B \varepsilon / N^2$ and the turbulent flux is $\mathcal{M} \approx \kappa N^2$, $\Gamma_B/0.2$ is equal to ratios of K and $\mathcal{M}$ based on Γ_B calculated using our recipe to their values when 0.2 is used.

corresponds to the peak cross-density (diapycnal) turbulence-induced upwelling that helps sustain the deep branch of the ocean circulation (de Lavergne, Madec, Roquet, Holmes, & McDougall, 2017). The shallow layer approximately corresponds to the base of the subtropical wind-driven gyres, across which turbulence mixes heat, carbon and nutrients between the upper ocean and the deep ocean (Talley, 2011). The patterns on the shallow layer reflect the seasonal atmospheric storm patterns at the surface and the rough topography at the bottom (from which internal tides radiate upwards). The patterns on the deep layer primarily represent the tidally generated turbulence, with little contribution from the surface generated downward propagating waves. The intermediate layer 'feels' both top and bottom generated wave fields.

The power maps, together with climatological hydrographic information (i.e. N^2) from the World Ocean Circulation Experiment (WOCE; see Gouretski & Koltermann, 2004), enable us to calculate Γ_B on the WOCE grid on which the power maps in Figure 5 are constructed (following de Lavergne et al., 2020). The resolution of the grid is half a degree in the horizontal ($\sim$50 km) and $\sim$110 m in the vertical. Applying our recipe to each grid cell with its local power and density stratification, we obtain the maps of Γ_B shown in Figure 5d–f, normalized by 0.2. Since the bulk turbulent diffusivity for each grid cell may be approximated as $\kappa \approx \Gamma_B N^2 / \varepsilon_B$ and the associated buoyancy flux is defined as $\mathcal{M}_B \approx \kappa N^2$, the ratio $\Gamma_B/0.2$ is exactly equal to the ratios of κ and $\mathcal{M}$ calculated based on our recipe to their values when calculated using $\Gamma_B = 0.2$. Thus, Figure 5d–f represents all three ratios and collectively show significant spatial variations to Γ_B. For example, for flux of a given quantity C between the surface ocean and the deep ocean (where C can be heat, carbon or nutrients for example), Figure 5d implies that the flux κC can be overestimated or underestimated by more than 200 % if $\Gamma_B = 0.2$ is used.

In the deep ocean, turbulent mixing leads to the irreversible transformation of water masses into lighter/denser waters (i.e. upwelling/downwelling), which facilitates a net upwelling of the dense abyssal waters, thereby closure of the deep branch of ocean meridional circulation (Ferrari, Mashayek, McDougall, Nikurashin, & Campin, 2016; de Lavergne, Madec, Le Sommer, Nurser, & Naveira Garabato, 2016; Mashayek, Ferrari, Nikurashin, & Peltier, 2015; Nikurashin & Vallis, 2011;
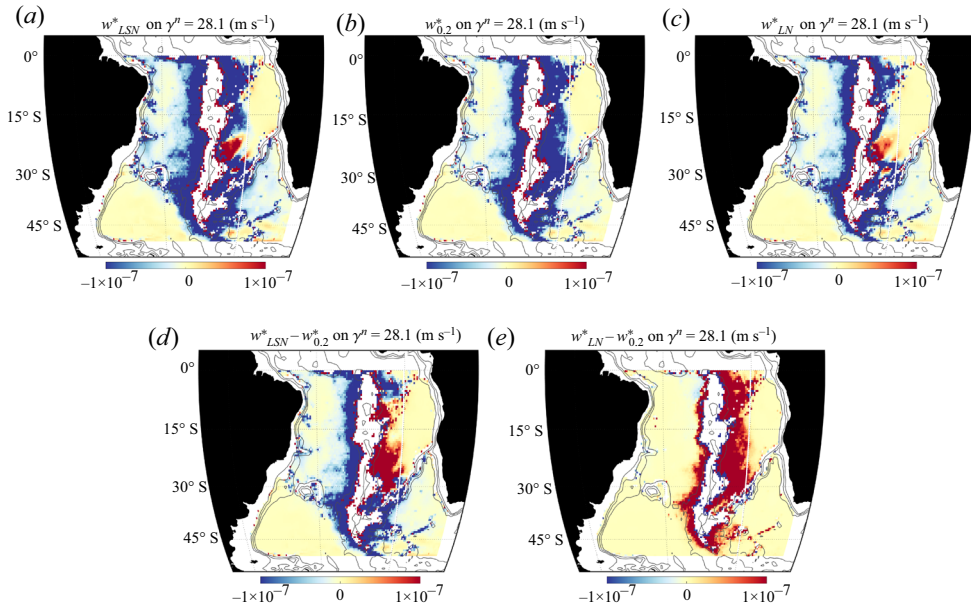
Figure 6. *(a) Local diapycnal velocity (see* (5.1)*) showing upwelling (in red) and downwelling (in blue) on the deep density level 28.1, calculated using the variable Γ_B recipe with a log-skew-normal (LSN; see* (3.2)*) probability function used for distributing the power in each grid cell over a statistically significant distribution of turbulent patches. (b) Same as panel (a), but when $\Gamma_B = 0.2$ is used everywhere. (c) Same as panel (a) but with a log-normal (LN) distribution used instead of LSN (i.e. $\alpha = 0$ in* (3.2)*). (d) Difference between panels (a) and (b). (e) Difference between panels (c) and (b).*

Wunsch & Ferrari, 2004). The net transformation across a density layer is an integral of complex upwelling and downwelling patterns, decided by topographic features, large-scale stratification and circulation, surface forcings and turbulence (Cimoli et al., 2019; de Lavergne, Madec, Capet, Maze, & Roquet, 2016; de Lavergne et al., 2017; Mashayek et al., 2017). The local diapycnal conversion of water masses by turbulent mixing is expressed as (Ferrari et al., 2016; Walin, 1982)

$$w^* \approx \frac{\partial_z \mathcal{M}_B}{N^2}. \tag{5.1}$$

This quantity, plotted in Figure 6a,b on the same deep density layer as in Figure 5c, is positive where waters become lighter and negative where waters become denser. Specifically, w^* is positive when $\mathcal{M}_B$ decreases with depth, for example when there is surface-intensified mixing, or in the bottom boundary layer where a one-dimensional model would predict $\mathcal{M}_B \to 0$ towards the ocean floor (Ferrari et al., 2016) (the nature of such transition is poorly understood and a topic of active research). Conversely, w^* is negative and waters become denser when mixing increases with depth, for example in the ocean interior near rough topography, where mixing is enhanced towards the bottom.

Figure 6a is based on the variable Γ_B shown in Figure 5e, obtained from our recipe using the log-skew-normal [LSN] patch distribution in (3.2). The net transformation is the sum of sharp near boundary upwelling squeezed between the bottom topography and regions of intense downwelling. Figure 6b shows the same map as Figure 6a but calculated with $\Gamma_B = 0.2$, while Figure 6d shows the difference between Figure 6a and Figure 6b. Red regions in Figure 6d imply either an increase in local upwelling or a decrease in local downwelling whereas blue implies enhanced downwelling or diminished upwelling (see Cimoli et al. (2019) for a more detailed discussion of these patterns). The change to w^* is of the same order of magnitude as w^* itself. Thus, the variations in Γ_B exert a leading order control over the rate
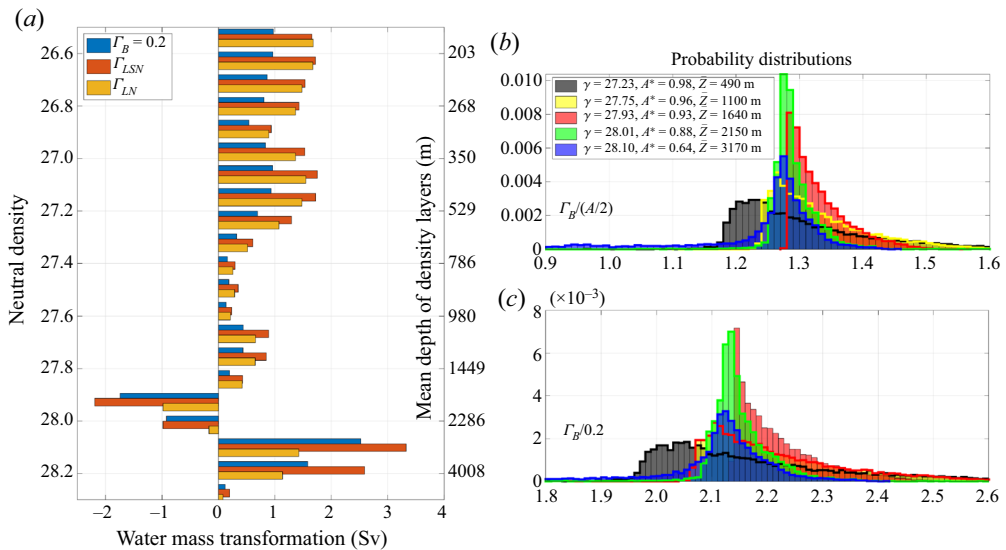
Figure 7. *(a) Water mass transformation (i.e. integral of w*, as per (5.2)) in the same domain as that in Figures 5,6, for $\Gamma_B = 0.2$ and variable Γ_B with LSN and LN patch distributions. (b,c) Histogram of the normalized Γ_B (calculated with LSN distribution) on various density levels. The legend includes information on the mean depths of density levels as well as their total area normalized by the area at the sea surface. Histograms are normalized by $\Gamma_{goldilocks} = A/2$ (see (2.2)) in panel (b) and by 0.2 in panel (c).*

and patterns of deep upwelling and downwelling in the ocean. Figure 6d shows asymmetric anomalies across the mid-Atlantic-ridge which can have important implications for mixing across the ridge and the broader inter-hemispheric Atlantic Meridional Circulation as well as inter-based exchange of water masses.

The net water mass transformation rate across a density surface may be defined as the area integral of the local transformation w^* through (Ferrari et al., 2016)

$$\mathcal{D}(\gamma^n) = - \iint_{A(\gamma^n)} w^* \cdot \hat{\boldsymbol{n}} \, \mathrm{d}A, \tag{5.2}$$

where $\hat{\boldsymbol{n}}$ is the unit vector normal to the neutral density surface γ^n and A is the surface area of the density surface. Here, $\mathcal{D}$ is measured in Sverdrups where $1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$. Figure 7a shows the net rate as a function of neutral density (with the mean depth of the density layers shown along the right vertical axis). In the deep ocean, Γ_B enhances the net transformation rate by over 50 % which is significant (although regionally the changes can be much higher, as shown in Figure 6d). In the upper ocean, where turbulence is key to air–sea tracer fluxes and primary productivity through supply of nutrients to the mixed layer, the change to the net upwelling from $\Gamma_B = 0.2$ to variable Γ_B is also significant.

Figure 7b shows the normalized histograms of $\Gamma_B/(A/2)$, where A is the coefficient in (2.2), for various density levels. The density levels range from a shallow layer with mean depth of 490 m and a total area of 0.98 of the sea surface area to a deep layer with mean depth of 3170 m and an area of 0.63 of the sea surface area. Recalling that $R_{OT} \sim 1$ corresponds to $\Gamma_{goldilocks} = A/2$, the fact that the histograms centre around $A/2$ implies that for most of the ocean, the statistical distribution of patches (within each grid cell of the climatological data) mostly comprises patches that are 'optimally' mixing, in the sense discussed in detail in MCA21. This implies an optimal balance between stratification and shear production. Or, in other words, regions of strong energy supply to turbulence appear to coincide really rather closely with the most 'desirable' stratification to yield optimal mixing. We use $A = 2/3$

in this study since it was justified on physical grounds in MCA21 and also was inferred through data regression as shown in Figure 3a and in MCA21. With this value, $\Gamma_{gold} = 1/3 \sim 1.7 \times 0.2$ which corresponds to green in Figure 7b. However, note that other values of A will not change the histograms and the optimal mixing argument that we make. Of course, built into this observation is the Goldilocks paradigm of (2.2) which is embedded within (but not essential to) our recipe. An alternate paradigm, not imagining energetic turbulence events that are in such an optimally mixing state, would yield another distribution. For completeness, Figure 7c shows the same distributions as in Figure 7b but normalized by 0.2.

6. On the sensitivity of our results to specific choices within the recipe

6.1. *A note on the choice of the log skew normal distribution for ε in (3.2)*

As was noted in § 1, the hybrid theoretical–statistical framework (i.e. the recipe) laid out in this work as a means for parametrization of small-scale mixing in climate models has merit in its own rights. Nevertheless, the choices for the ingredients, in the forms of the physical parametrization for patch-based Γ and the statistical distributions of turbulent patches, are critical to the final estimates of Γ_B and the local diffusivity inferred from it in coarse-grid models. Central to our estimates is the assumption that the rate of dissipation of kinetic energy has an LSN distribution, following the analysis of CM21 which was based on an extensive and diverse set of microstructure data. As discussed in CM21, (I) the skewness of the distribution may be contextual and (II) a log-normal distribution has traditionally been assumed. Since the tails of such distributions correspond to the largest turbulence overturns, one should expect the Γ_B estimates to be sensitive to the choice of the distribution within the recipe. To show that this is in fact the case, in Figures 6 and 7, we also consider estimates of w^* and $\mathcal{D}$ based on log-normal (LN) distribution. Comparisons of Figures 6d and 6e show the significant regional sensitivity to the choice of the distribution. The integral of w^*, as shown in Figure 7a, also shows that in the net, and specially in the abyssal ocean, the choice of the distribution function is of leading order importance. This highlights the need for further studying of statistical distributions of ε from field data with the goal of quantifying the regional, depth-wise and process-wise variations in the parameters of (3.2). Analogously to the Goldilocks paradigm embodied in (2.2), the LSN distribution is not an essential aspect of the recipe, but simply a base upon which to build.

6.2. *A note on the log skew normal scaling for R_{OT}*

The bulk flux coefficient inferred from our recipe depends significantly on the parameters involved in the L_T–L_O scaling described in § 4. Figure 4d highlights such sensitivity, by perturbing the parameters from a baseline Γ_g estimated from the combined global dataset used in Figure 4b. Larger fluctuations in L_T when L_O is large can increase the bulk flux coefficient, but it asymptotes to a constant value with decreasing fluctuations. Increasing either the scaling exponent or coefficient increases L_T values for large L_O, thus increasing the bulk flux coefficient for large ε; decreasing either the scaling exponent or the coefficient drives bulk mixing to zero as $R_{OT} \gg 1$ when ε is large. While there is significant evidence, as reviewed earlier, that R_{OT} possess a (seemingly universal) weakly skewed normal distribution (and so the parameters in our scaling are unlikely to vary over the range shown in Figure 4d), consideration of a much larger collection of observational data and deeper physical–statistical analyses are required to better parametrize this component of the recipe in the future.

6.3. *A note on our choice of parametrization for Γ_i*

As discussed earlier, our specific choices with regard to the physical and statistical components of the recipe are subjective choices which should be thought of as placeholders. We acknowledge that while our specific choice for the parametrization of Γ_i (as per (2.2)) is rooted in physical reasoning and seems

plausible based on a wide range of oceanic datasets (as discussed in MCA21), it is one of the existing paradigms. Alternative viewpoints, for example, have recently been expressed by Garanaik and Venayagamoorthy (2019) and Ivey, Bluteau, Gayen, Jones, and Sohail (2021) and can straightforwardly be implemented in our recipe.

7. Discussion

Climate model resolutions are orders of magnitude away from resolving small-scale ocean turbulence: while model resolution is $O(100$ km$)$, diapycnal mixing occurs on sub-meter scales. It is not even clear whether resolving such mixing is an ultimate goal of ocean modelling. Thus, parametrizations that connect the primary sources of power available to turbulence (e.g. from winds, tides, eddies, currents, etc.) to mixing are inevitable. Connecting power to mixing requires an assumption about how it partitions into mixing and viscous dissipation. Assuming a constant flux coefficient of 0.2, which is currently the norm in ocean modelling, for such partitioning can only hold in one of two ways: (I) each grid cell of a given model is always filled with turbulence processes all of which mix with $\Gamma_i = 0.2$, (II) each grid cell hosts spatio-temporally intermittent turbulence with variable Γ_i in such a way that a time-space average over the cell always results in $\Gamma_B = 0.2$. The first possibility is obviously unrealistic and there is no evidence to support the latter. So, in summary, the status quo in form of using $\Gamma_B = 0.2$ is unjustifiable. In this work, we showed that it also can lead to significant inaccuracies in predicting turbulence fluxes and diapycnally driven upwelling and downwelling in the ocean interior.

It was suggested herein that (somewhat) established and (seemingly) universal statistical characteristics of turbulent overturns in the ocean interior can be employed to connect the physics of small-scale mixing, in the form of parametrization of flux coefficient for individual patches, to an appropriately averaged bulk mixing on regional or coarse-model-grid scales. There are two aspects to our work. First, from a model parametrization perspective, we offered a recipe which connects power sources to small-scale mixing through a statistical–stochastical approach. Our choices of parametrization of the patch-based flux coefficient and statistical distribution of patches both may be substituted for by alternate paradigms or improved over time.

Second, with our specific choices of parametrization of Γ_i and statistical distributions, both of which were informed by extensive oceanic datasets, we reached a number of physical conclusions. We showed that in most energetically mixing regions, a given 'box' of the ocean is filled with turbulent patches which optimally mix with $\Gamma_i \sim 1/4 - 1/3$. As discussed in MCA21, this range is closely related to classical (marginal) instability criteria of stratified mixing layers. Such boxes tend to have a bulk $\Gamma_B \sim 0.4$. While larger Γ_B values may be found in the domain, such values typically correspond to weak turbulence activity (remembering that $\Gamma \to \infty$ in the limit of laminar flow). The identification of weakly and strongly mixing zones is key as it is the changes from one to another that sets the upwelling and downwelling of waters (we note that while high mixing zones typically correspond to mixing rates well above the effective numerical diffusion in coarse resolution models, that is not necessarily true for the weakly mixing zones. This issue, however, needs to be addressed on a case-by-case basis during implementation as different models use different advection schemes with varying implications for the implied numerical vertical diffusion).

In the deep ocean, for example, our recipe allows for the flux ($\Gamma_B \times \varepsilon_B$) to vary from small values away from topography, to high values in its vicinity, and from there to zero at the seafloor. In physical terms, the recipe allows for finite, non-trivial mixing where there is something to mix and (reassuringly) zero mixing where there is not anything to mix. Using a constant value will force the model to undermix where it should mix more effectively, and to overmix where there is not much to mix in reality. In climate models that employ a universally constant Γ, the diffusivity, which is proportional to Γ/N^2, blows up at the bottom where N^2 becomes too small – in models this is often set to 0 at the seafloor. Our recipe alleviates that problem through $\Gamma_B \to 0$ in that limit. Importantly, the transition from weak turbulent flux to strong flux marks the intensity and direction of diapycnal conversion (upwelling/downwelling

patterns) and thereby exerts a control over the closure of the diffusively driven abyssal circulation. Quantification of such patterns and the realization of the fact that the net abyssal upwelling is the residual of significant downwelling and upwelling is a new paradigm in physical oceanography (Powell, 2016; Voosen, 2022), and one which partially motivated this work.

The recipe proposed in this work, the code for which is available in the supplementary material, may be used to tackle these problems in climate models as a first step with great room for improvement as our understanding of the physics of mixing and statistical distributions of overturns improve with time. In our view, both improvements can benefit from a closer collaboration between the fluid mechanics and physical oceanography communities.

References

Alford, M.H. (2020). Global calculations of local and remote near-inertial-wave dissipation. *Journal of Physical Oceanography*, *50*(11), 3157–3164.

Alford, M.H., Girton, J.B., Voet, G., Carter, G.S., Mickett, J.B., & Klymak, J.M. (2013). Turbulent mixing and hydraulic control of abyssal water in the Samoan Passage. *Geophysical Research Letters*, *40*(17), 4668–4674.

Bouffard, D., & Boegman, L. (2013). A diapycnal diffusivity model for stratified environmental flows. *Dynamics of Atmospheres and Oceans*, *61–62*, 14–34.

Cael, B., & Mashayek, A. (2021). Log-skew-normality of ocean turbulence. *Physical Review Letters*, *126*(22), 224502.

Caldwell, D.R., & Moum, J.N. (1995). Turbulence and mixing in the ocean. *Reviews of Geophysics*, *33*(2 S), 1385–1394. https://doi.org/10.1029/95RG00123.

Carter, G.S., Voet, G., Alford, M.H., Girton, J.B., Mickett, J.B., Klymak, J.M., . . . Tan, S. (2019). A spatial geography of abyssal turbulent mixing in the samoan passage. *Oceanography*, *32*(4), 194–203.

Caulfield, C.P. (2021). Layering, instabilities, and mixing in turbulent stratified flows. *Annual Review of Fluid Mechanics*, *53*, 113–145. https://doi.org/10.1146/annurev-fluid-042320-100458.

Cimoli, L., Caulfield, C.P., Johnson, H.L., Marshall, D.P., Mashayek, A., Naveira Garabato, A.C., & Vic, C. (2019). Sensitivity of deep ocean mixing to local internal tide breaking and mixing efficiency. *Geophysical Research Letters*, *46*(24), 14622–14633. https://doi.org/10.1029/2019GL085056.

Clauset, A., Shalizi, C.R., & Newman, M.E. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.

Dillon, T.M. (1982). Vertical overturns: A comparison of Thorpe and Ozmidov length scales. *Journal of Geophysical Research*, *87*(C12), 9601. https://doi.org/10.1029/jc087ic12p09601.

Ferrari, R., Mashayek, A., McDougall, T.J., Nikurashin, M., & Campin, J.-M. (2016). Turning ocean mixing upside down. *Journal of Physical Oceanography*, *46*(7), 2239–2261.

Ferron, B., Mercier, H., Speer, K., Gargett, A., & Polzin, K. (1998). Mixing in the Romanche fracture zone. *Journal of Physical Oceanography*, *28*(10), 1929–1945.

Garabato, A.N., & Meredith, M. (2022). Ocean mixing: Oceanography at a watershed. In *Ocean Mixing* (pp. 1–4). Amsterdam, the Netherlands: Elsevier.

Garanaik, A., & Venayagamoorthy, S.K. (2019). On the inference of the state of turbulence and mixing efficiency in stably stratified flows. *Journal of Fluid Mechanics*, *867*, 323–333. https://doi.org/10.1017/jfm.2019.142.

Gargett, A.E., & Moum, J.N. (1995). Mixing efficiencies in turbulent tidal fronts: Results from direct and indirect measurements of density flux. *Journal of Physical Oceanography*, *25*(11), 2583–2608.

Garrett, C. (2001). Stirring and mixing: What are the rate-controlling processes? *Proceedings of the Hawaiian Winter Workshop 'From Stirring to Mixing in a Stratified Ocean', USA, 12*, 1–8.

Garrett, C., & Kunze, E. (2007). Internal tide generation in the deep ocean. *Annual Review of Fluid Mechanics*, *39*, 57–87.

Gibson, H. (1991). Turbulence, mixing, and heat flux in the ocean main thermocline. *Journal of Geophysical Research*, *96*(Cll), 403–420.

Gouretski, V., & Koltermann, K.P. (2004). WOCE global hydrographic climatology. *Berichte des BSH*, *35*, 1–52.

Gregg, M., D'Asaro, E., Riley, J., & Kunze, E. (2018). Mixing efficiency in the ocean. *Annual Review of Marine Science*, *10*(1), 443–473. https://doi.org/10.1146/annurev-marine-121916-063643.

Gregg, M.C., Seim, H.E., & Percival, D.B. (1993). Statistics of shear and dissipation. *Journal of Physical Oceanography*, *23*(8), 1777–1799.

Gurvich, A.S., & Yaglom, A.M. (1967). Breakdown of eddies and probability distributions for small-scale turbulence. *Physics of Fluids*, *10*(9), https://doi.org/10.1063/1.1762505.

Hcine, M.B., & Bouallegue, R. (2015). Highly accurate log skew normal approximation to the sum of correlated lognormals. Advance online publication. arXiv:1501.02347.

Ivey, G.N., Bluteau, C.E., Gayen, B., Jones, N.L., & Sohail, T. (2021). Roles of shear and convection in driving mixing in the ocean. *Geophysical Research Letters*, *48*(3), e2020GL089455.

Ivey, G.N., Winters, K.B., & Koseff, J.R. (2008). Density stratification, turbulence, but how much mixing? *Annual Review of Fluid Mechanics*, *40*, 169–184.

Koenker, R., & Hallock, K.F. (2001). Quantile regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

de Lavergne, C., Falahat, S., Madec, G., Roquet, F., Nycander, J., & Vic, C. (2019). Toward global maps of internal tide energy sinks. *Ocean Modelling*, *137*, 52–75. https://doi.org/10.1016/j.ocemod.2019.03.010.

de Lavergne, C., Madec, G., Capet, X., Maze, G., & Roquet, F. (2016). Getting to the bottom of the ocean. *Nature Geoscience*, *9*(12), 857–858. https://doi.org/10.1038/ngeo2850.

de Lavergne, C., Madec, G., Le Sommer, J., Nurser, A.J.G., & Naveira Garabato, A.C. (2015). The impact of a variable mixing efficiency on the abyssal overturning. *Journal of Physical Oceanography*, *46*(2), 663–681.

de Lavergne, C., Madec, G., Le Sommer, J., Nurser, A.J.G., & Naveira Garabato, A.C. (2016). On the consumption of antarctic bottom water in the Abyssal Ocean. *Journal of Physical Oceanography*, *46*(2), 635–661. https://doi.org/10.1175/JPO-D-14-0201.1.

de Lavergne, C., Madec, G., Roquet, F., Holmes, R.M., & McDougall, T.J. (2017). Abyssal ocean overturning shaped by seafloor distribution. *Nature*, *551*(7679), 181–186. https://doi.org/10.1038/nature24472.

de Lavergne, C., Vic, C., Madec, G., Roquet, F., Waterhouse, A.F., Whalen, C., . . . Hibiya, T. (2020). A parameterization of local and remote tidal mixing. *Journal of Advances in Modeling Earth Systems*, *12*(5), e2020MS002065.

Ledwell, J.R., Montgomery, E.T., Polzin, K.L., Laurent, L.C.S., Schmitt, R.W., & Toole, J.M. (2000). Evidence for enhanced mixing over rough topography in the abyssal ocean. *Nature*, *403*, 182–189.

Legg, S. (2021). Mixing by oceanic lee waves. *Annual Review of Fluid Mechanics*, *53*(1), 173–201.

Mashayek, A., Caulfield, C., & Alford, M. (2021). Goldilocks mixing in oceanic shear-induced turbulent overturns. *Journal of Fluid Mechanics*, *928*, A1.

Mashayek, A., Caulfield, C.P., & Peltier, W.R. (2013). Time-dependent, non-monotonic mixing in stratified turbulent shear flows: Implications for oceanographic estimates of buoyancy flux. *Journal of Fluid Mechanics*, *736*, 570–593.

Mashayek, A., Caulfield, C.P., & Peltier, W.R. (2017). Role of overturns in optimal mixing in stratified mixing layers. *Journal of Fluid Mechanics*, *826*, 522–552.

Mashayek, A., Ferrari, R., Nikurashin, M., & Peltier, W.R. (2015). Influence of enhanced Abyssal Diapycnal mixing on stratification and the ocean overturning circulation. *Journal of Physical Oceanography*, *45*(10), 2580–2597.

Mashayek, A., & Peltier, W.R. (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, W.R. (2012a). The 'zoo' of secondary instabilities precursory to stratified shear flow transition. Part 1 Shear aligned convection, pairing, and braid instabilities. *Journal of Fluid Mechanics*, *708*, 5–44.

Mashayek, A., & Peltier, W.R. (2012b). The 'zoo' of secondary instabilities precursory to stratified shear flow transition. Part 2. The influence of stratification. *Journal of Fluid Mechanics*, *708*, 45–70.

Mashayek, A., & Peltier, W.R. (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., Salehipour, H., Bouffard, D., Caulfield, C.P., Ferrari, R., Nikurashin, M., . . . Smyth, W.D. (2017). Efficiency of turbulent mixing in the Abyssal Ocean circulation. *Geophysical Research Letters*, *44*(12), 6296–6306.

Mater, B.D., & Venayagamoorthy, S.K. (2014). A unifying framework for parameterizing stably stratified shear-flow turbulence. *Physics of Fluids*, *26*(3), 36601.

Mater, B.D., Venayagamoorthy, S.K., Laurent, L.S., & Moum, J.N. (2015). Biases in thorpe-scale estimates of turbulence dissipation. Part I: Assessments from large-scale overturns in oceanographic data. *Journal of Physical Oceanography*, *45*(10), 2497–2521. https://doi.org/10.1175/JPO-D-14-0128.1.

Monismith, S.G., Koseff, J.R., & White, B.L. (2018). Mixing efficiency in the presence of stratification: When is it constant? *Geophysical Research Letters*, *45*(11), 5627–5634. https://doi.org/10.1029/2018GL077229.

Nikurashin, M., & Ferrari, R. (2013). Overturning circulation driven by breaking internal waves. *Journal of Physical Oceanography*, https://doi.org/10.1002/grl.50542.

Nikurashin, M., & Vallis, G. (2011). A theory of deep stratification and overturning circulation in the ocean. *Journal of Physical Oceanography*, *41*(3), 485–502.

Osborn, T.R. (1980). Estimates of the local rate of vertical diffusion from dissipation measurements. *Journal of Physical Oceanography*, *10*, 83–89.

Peltier, W.R., & Caulfield, C.P. (2003). Mixing efficiency in stratified shear flows. *Annual Review of Fluid Mechanics*, *35*, 135–167.

Polzin, K.L., Toole, J.M., Ledwell, J.R., & Schmitt, R.W. (1997). Spatial variability of turbulent mixing in the Abyssal Ocean. *Science*, *276*(5309), 93–96.

Pourahmadi, M. (2007). Construction of skew-normal random variables: Are they linear combinations of normal and half-normal? *Journal of Statistical Theory and Applications*, *3*, 314–328.

Powell, D. (2016). Role of chaos in deep ocean turned upside down: Reversal could change timeline of heat and carbon storage in the seas. *Nature*. https://doi.org/10.1038/nature.2016.19455.

Smyth, W.D., & Moum, J.N. (2000). Length scales of turbulence in stably stratified mixing layers. *Physics of Fluids*, *12*, 1327.

Smyth, W., Nash, J., & Moum, J. (2019). Self-organized criticality in geophysical turbulence. *Scientific Reports*, *9*(1), 1–8.

St. Laurent, L.C., Toole, J.M., & Schmitt, R.W. (2001). Buoyancy forcing by turbulence above rough topography in the Abyssal Brazil Basin*. *Journal of Physical Oceanography*, *31*(12), 3476–3495.

Talley, L.D. (2011). *Descriptive physical oceanography: An introduction*. Cambridge, MA: Academic Press.

Talley, L.D., Feely, R.A., Sloyan, B.M., Wanninkhof, R., Baringer, M.O., Bullister, J.L., . . . Zhang, J.Z. (2016). Changes in ocean heat, carbon content, and ventilation: A review of the first decade of GO-SHIP global repeat hydrography. *Annual Review of Marine Science*, *8*, 185–215. https://doi.org/10.1146/annurev-marine-052915-100829.

Thorpe, S.A. (2005). *The turbulent ocean*. Cambridge, UK: Cambridge University Press.

Voosen, P. (2022). Undersea mountains help stir up oceans. *Science*, *375*(6587), 1324–1325.

Walin, G. (1982). On the relation between sea-surface heat flow and thermal circulation in the ocean. *Tellus*, *34*(2), 187–195. https://doi.org/10.1111/j.2153-3490.1982.tb01806.x.

Whalen, C.B., De Lavergne, C., Naveira Garabato, A.C., Klymak, J.M., Mackinnon, J.A., & Sheen, K.L. (2020). Internal wave-driven mixing: Governing processes and consequences for climate. *Nature Reviews Earth & Environment*, *1*(11), 606–621.

Wu, Z., Li, X., Husnay, R., Chakravarthy, V., Wang, B., & Wu, Z. (2009). A novel highly accurate log skew normal approximation method to lognormal sum distributions. In *2009 IEEE Wireless Communications and Networking Conference* (pp. 1–6). IEEE.

Wunsch, C., & Ferrari, R. (2004). Vertical mixing, energy, and the general circulation of the oceans. *Annual Review of Fluid Mechanics*, *36*, 281–314.

Yamazaki, H., & Lueck, R. (1990). Why oceanic dissipation rates are not lognormal. *Journal of Physical Oceanography*, *20*(12), 1907–1918. https://doi.org/10.1175/1520-0485(1990)020<1907:WODRAN>2.0.CO;2.

York, D. (1966). Least-squares fitting of a straight line. *Canadian Journal of Physics*, *44*(5), 1079–1086.